

INTEGRATED CIRCUITS AND SYSTEMS

Aleksandr Tasić
Wouter A. Serdijn
Lawrence E. Larson
Gianluca Setti
Editors

Circuits and Systems for Future Generations of Wireless Communications

 Springer

Circuits and Systems for Future Generations of Wireless Communications

Series on Integrated Circuits and Systems

Series Editor: Anantha Chandrakasan
Massachusetts Institute of Technology
Cambridge, Massachusetts

For other titles published in this series, go to
<http://www.springer.com/series/7236>

Aleksandar Tasić • Wouter A. Serdijn
Lawrence E. Larson • Gianluca Setti
Editors

Circuits and Systems for Future Generations of Wireless Communications

 Springer

Editors

Aleksandar Tasić
Qualcomm Inc.
5775 Morehouse Dr.
San Diego CA 92121
USA
ataxic@qualcomm.com

Wouter A. Serdijn
Delft University of Technology
Electronics Research Lab.
Mekelweg 4
2628 CD Delft
ET Bldg.
Netherlands
w.a.serdijn@tudelft.nl

Lawrence E. Larson
University of California
San Diego
Dept. Electrical & Computer Engineering
9500 Gilman Drive
La Jolla CA 92093-0407
MS 0407
USA
larson@ece.ucsd.edu

Gianluca Setti
University of Ferrara
Dept. Engineering (ENDIF)
Via Saragat, 1
44100 Ferrara
and
University of Bologna
Ad. Research Center on Electronic Systems
(ARCES)
Via toffano 2/2
40136 Bologna
Italy
gianluca.setti@unife.it

Series Editor:

Anantha Chandrakasan
Department of Electrical Engineering and
Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139
USA

ISSN 1558-9412

ISBN 978-1-4020-9918-2

e-ISBN 978-1-4020-9917-5

DOI 10.1007/978-1-4020-9917-5

Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2009922216

© Springer Science+Business Media B.V. 2009

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The idea for this book originated from a Special Session on Circuits and Systems for Future Generations of Wireless Communications that was presented at the 2005 International Symposium on Circuits and Systems, which was then followed by two Special Issues bearing the same title that appeared in the March and April 2008 issues of the IEEE Transactions on Circuits and Systems – Part II: Express Briefs. Out of a large number of great contributions, we have selected those fitting best the book format based on their quality.

We would like to thank all the authors, the reviewers of the Transactions on Circuits and Systems – Part II, and the reviewers of the final book material for their efforts in creating this manuscript. We also thank the Springer Editorial Staff for their support in putting together all the good work. We hope that this book will provide you, the reader, with new insights into *Circuits and Systems for Future Generations of Wireless Communications*.

*Aleksandar
Wouter
Larry
and Gianluca*

Contents

1	Introduction	1
2	A Multimode Radio Transceiver for Cellular Applications	5
	John Groe	
3	Reconfigurable Multi-Band OFDM UWB Receivers: Circuits and System Considerations	27
	Luca Baldini, Danilo Manstretta, Tomaso Erseghe, Nicola Laurenti, Antonio Liscidini, and R. Castello	
4	Low Power UWB Circuits: Front-End Building Blocks	53
	Tommy Tsang, Kuan-Yu Lin, Karim Allidina, and Mourad El-Gamal	
5	CMOS IR-UWB Transceiver System Design for Contact-Less Chip Testing Applications	85
	Yanjie Wang, Ali M. Niknejad, Vincent Gaudet, and Kris Iniewski	
6	Multi-mode Power Amplifiers for Wireless Handset Applications	121
	Junxiong Deng and Lawrence E. Larson	
7	Polyphase Multipath Circuits for Cognitive Radio and Flexible Multi-phase Clock Generation	145
	Eric A.M. Klumperink, Xiang Gao, and Bram Nauta	
8	IIP2 Improvement Techniques for Multi-standard Mobile Radio	169
	Mohammad B. Vahidfar and Omid Shoaie	

9 Multi-standard Continuous-Time Sigma-Delta Converters for 4G Radios	203
Yi Ke, Jan Craninkx, and Georges Gielen	
10 Power Efficient Reconfigurable Baseband Filters for Multimode Radios	223
Pieter Crombez, Jan Craninckx, and Michiel Steyaert	
11 An Adaptive Digital Front-End for Multi-mode Wireless Receivers	249
Gernot Hueber, Rainer Stuhlberger, and Andreas Springer	
12 FEC Decoders for Future Wireless Devices: Scalability Issues and Multi-standard Capabilities	271
John Dielissen, Nur Engin, Sergei Sawitzki, and Kees van Berkel	
Index	299

Chapter 1

Introduction

The explosive demand in wireless-capable devices, especially with the proliferation of multiple standards, indicates a great opportunity for adoption of wireless technology at a mass-market level. The communication devices of both today and the future will have not only to allow for a variety of applications, supporting the transfer of characters, audio, graphics, and video data, but they will also have to maintain connection in a variety of environments with many other devices rather than with a single base station. Moreover, to provide various services from different wireless communication standards with higher capacities and higher data-rates, fully integrated and multifunctional wireless devices will be required.

Multifunctional circuits and systems can be made profitable by a large scale of integration, elimination of external components, reduction of silicon area, and extensive reuse of resources. Integration of (Bi)CMOS transceiver RF front-end and analog baseband circuits with computing CMOS circuits on the same silicon chip further reduces costs of multifunctional mobile devices.

However, as batteries continue to determine the lifetime and size of mobile equipment, further extension of capabilities of wearable and wireless devices will depend critically on the integrated circuits and systems design solutions.

The demand for multifunctional and multi-mode wireless-capable devices is accompanied by many significant challenges at system, circuit, and technology levels.

In this book, we discuss circuit and system design solutions for multiple communication standards and future generations of wireless communications:

In the following Chapter 2, J. Groe provides an overview of the design and performance of a multi-mode radio transceiver satisfying the requirements of GSM, EDGE, and WCDMA networks. This is accomplished using well-known architectures that previously fell short on performance. The radio receiver employs a sliding-IF architecture that takes advantage of both low-IF and direct-conversion concepts. The radio transmitter extends polar modulation, typical for GSM/EDGE, to WCDMA.

In Chapter 3, J. Baldini, D. Manstretta, T. Erseghe, N. Laurenti, A. Liscidini, and R. Castello present an analysis of receiver front-end architectures for multi-band

orthogonal frequency-division multiplexing ultra-wideband terminals. An interference analysis is carried out in order to derive the main linearity specifications of the receiver front-end. A reconfigurable narrowband architecture is introduced that can best cope with the main challenges of ultra-wideband receivers: broadband impedance matching and high out-of-band linearity. Measurement and simulations results show that linearity requirements can be met with good margins.

In Chapter 4, T. Tsang, K.-Y. Lin, K. Allindina, and M. El-Gamal first provide a brief overview of ultra-wideband communications, after which they describe some radio-frequency front-end building blocks. The authors present two ultra-wideband low-noise amplifiers based on a common-gate input stage for use in the 3.1–10.6 GHz band and a low-noise amplifier for use in the 0–900 MHz band. An implementation of a pulse-based transmitter for direct-sequence ultra-wideband systems concludes the chapter.

In Chapter 5, Y. Wang, A. Niknejad, V. Gaudet, and K. Iniewski present impulse-based ultra-wideband transceiver circuits for future contact-less chip testing applications using magnetic coupling transformers as wireless interconnects. An ultra-wideband architecture is proposed first and then transmit circuits, magnetic coupling transformers, and receive circuits described and simulations results provided.

In Chapter 6, J. Deng and L. E. Larson first introduce different process technologies used for power amplifiers and compare some published handset power amplifiers for cellular applications. Efficiency and linearity enhancement techniques are then detailed and complex-gain digital pre-distortion techniques elaborated from theory to implementation. Some measurement results are provided in support of the power-amplifier design procedure proposed in this chapter.

In Chapter 7, E. Klumperink, X. Gao, and B. Nauta discuss flexible cognitive radio circuits for dynamic access of unused spectrum. They review techniques to realize radios without resorting to frequency selective dedicated filters by exploiting a polyphase multipath technique canceling harmonics and sidebands. With this, a wideband and flexible power upconverter with a clean output spectrum can be realized on a CMOS chip, which allows implementing prototype chips transmitting at an arbitrary frequency between DC and 2.4 GHz. Unwanted harmonics and sidebands are more than 40 dB lower than the desired signal up to the 17th harmonic of the transmit frequency. The generation of the necessary multi-phase clock is done exploiting a Shift Register approach that is shown to be superior with respect to the one based on a Delay Locked Loop in terms of jitter performance at any power budget.

In Chapter 8, M. Vahidfar and O. Shoaie investigate the down-conversion mixer for a direct conversion multi-standard radio. This is a particularly challenging block, due to the stringent dynamic-range requirements set by cell-phone applications. A calibration technique for second-order intermodulation distortion and a circuit implementation are presented, leveraging the source and mechanisms of the distortion. The low-noise calibration circuitry improves second-order-input intercept point for 25 dB.

In Chapter 9, Y. Ke, J. Craninckx, and G. Gielen present a design approach for fully reconfigurable low-voltage Delta-Sigma analog-to-digital converters for next-generation wireless applications. Unlike classical multi-mode designs, this approach provides a better power efficiency and a better trade-off between power and performance. As a proof of concept, the system-level design of a digitally programmable Delta-Sigma modulator for 4G radios is presented.

In Chapter 10, P. Crombez, J. Craninckx, and M. Steyaert present a fully reconfigurable Gm-C biquadratic low-pass filter answering both the efficiency and flexibility demand of future mobile devices. Concentrating on linearity optimization and low-power consumption, a novel switching strategy inside Nauta's transconductor is implemented, which allows for a very large frequency-performance-power flexibility. Measurements of an implementation in 0.13 μm 1.2 V CMOS IC technology demonstrate a filter that can be tuned over more than two orders of magnitude, from 100 kHz up to 20 MHz, and with a performance scalable in terms of noise, power and linearity.

In Chapter 11, G. Hueber, R. Stuhlberger, and A. Springer give an overview of the main test cases for several third-generation standards. The authors describe architectures for adaptive multi-mode wireless receivers and then focus on the design of an adaptive multi-mode low-power digital front-end. Digital front-end circuits are described and key simulation results presented.

Finally, in Chapter 12, John Dielissen, Nur Engin, Sergei Sawitzki, and Kees van Berkel argue that implementing forward error correction decoders for wireless transmission standards in mobile handsets and other consumer devices is changing in nature, since the number of standards is increasing dramatically. They provide an overview of the multi-standard capabilities of forward error correction decoders from common decoder families, such Reed-Solomon, Viterbi, Turbo, and Low-Density Parity Check, and show that single-family multi-standard decoders are implementable with limited area overhead compared to reference designs for a single standard. Finally, they review the possibilities for combinations of decoders from different families within one hardware platform.

Chapter 2

A Multimode Radio Transceiver for Cellular Applications

John Groe

The explosive growth of wireless communications continues to drive the development of cellular networks with upgraded services. Many GSM networks now support EDGE and WCDMA features that allow high-speed data access. Future networks will add LTE capability for even faster data service. Consequently, multimode devices are becoming increasingly common and practically mandatory.

To fuel this growth, it becomes essential to provide these complex systems economically. This has led to a rebirth of radio technologies focused on efficient multimode architectures.

To satisfy this goal, a radio transceiver must be flexible enough to address the daunting and vastly different requirements associated with GSM, EDGE, and WCDMA networks. This is accomplished using well-known architectures that previously fell short on performance. The radio receiver employs a sliding-IF architecture that takes advantage of both low-IF [1–3] and direct-conversion [4–6] concepts. The radio transmitter extends polar modulation – the preferred solution for GSM/EDGE – to WCDMA [7]. The result is a streamlined multimode cellular radio [8].

2.1 Receiver

The radio receiver shown in Fig. 2.1 uses a pseudo-direct conversion architecture with an integrated notch filter to adapt to different requirements. In this approach, the down converter translates the receive band to either a low-IF frequency or to dc (zero-IF). This flexibility is the key to the multimode receiver.

GSM, EDGE, and WCDMA are three, very different, modulation schemes. GSM is a narrow-band, continuous-phase modulation approach. EDGE is another narrow-band approach, but with a more complex modulation format. In stark contrast,

J. Groe
Sequoia Communications, San Diego, Ca
e-mail: jgroe@sequoia-communications.com

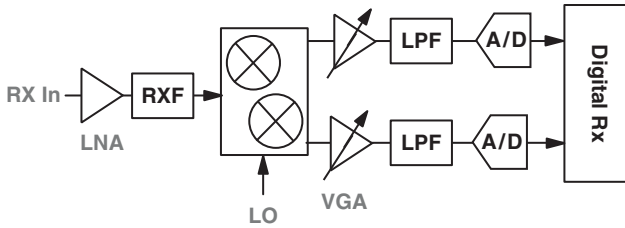


Fig. 2.1 Pseudo-direct conversion receiver with a sliding IF

Table 2.1 Comparison of GSM, EDGE, and WCDMA standards [8]

Parameter	GSM	EDGE	WCDMA
Modulation	GMSK	$3\pi/8$ -8PSK	Spread spectrum
Signal pk/ave	0 dB	3.4 dB	3.1–6.5 dB
Data rate	20 kbps	60 kbps	Up to 14.4 Mbps (DL) Up to 5.76 Mbps (UL)
Sample rate	270 kHz	270 kHz	3.84 Mcps
Multiple access	F/TDMA	F/TDMA	F/CDMA
Duplex	Half	Half	Full
Channel spacing	200 kHz	200 kHz	5 MHz
Tx power	33 dBm	27 dBm	24 dBm
Rx BER quality	$\leq 10^{-4}$	$\leq 10^{-4}$	$\leq 10^{-3}$

WCDMA is a wideband, highly-linear modulation scheme using spread spectrum techniques. Consequently, these three approaches affect the radio transceiver in different ways as shown in Table 2.1.

2.1.1 RF Front-End

The RF front-end consists of a multi-purpose low noise amplifier (LNA), notch filter, and down converter. The LNA shown in Fig. 2.2 provides multiple gain steps to maximize the dynamic range of the receiver. It uses cascode transistors N_3 – N_6 , controlled by bias voltages V_{b1} and V_{b2} , to steer the RF signal current to an R–2R ladder designed for 6 dB gain steps. This structure preserves the input/output impedances of the amplifier and maintains a low noise figure at low gain settings.

A differential LNA design ensures balanced drive to the notch filter and down converter mixers. As a result, any even order distortion appears as common mode signals which ideally cancel with symmetric design and layout.

The notch filter protects the down converter in full-duplex WCDMA mode. Full-duplex operation allows the receiver and transmitter to operate simultaneously, connecting to the antenna through a duplex filter. At high transmit power levels, the duplex filter's limited isolation allows a fairly strong signal to appear at the

Fig. 2.2 A variable gain LNA with two gain settings

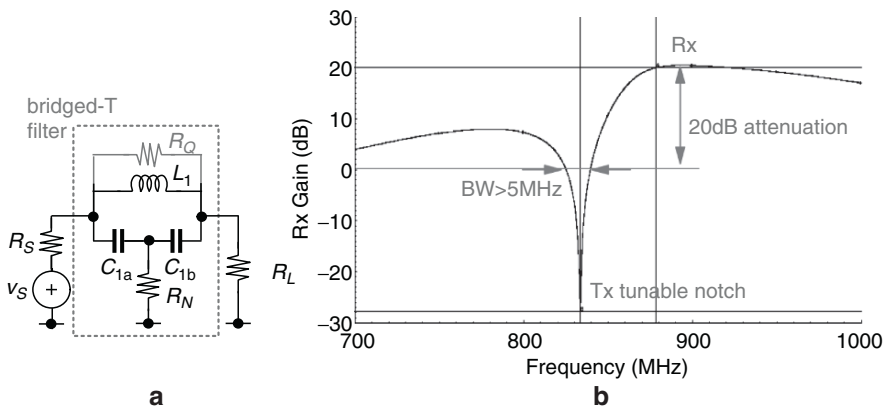
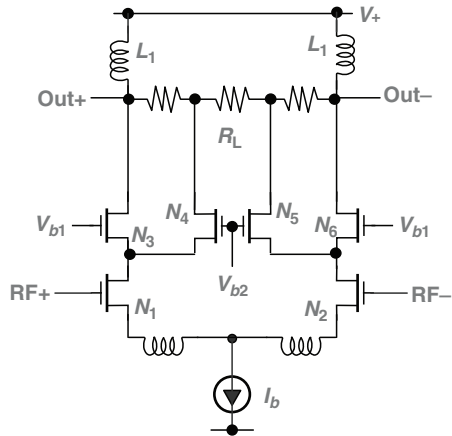


Fig. 2.3 Bridged-T filter used to implement receive notch filter – (a) equivalent model and (b) response

receiver’s input. As a result, the transmit leakage signal becomes potentially the strongest blocker signal seen by the receiver and can easily overwhelm the wanted signal¹ if untreated [9].

The tunable notch filter employs a modified version of the bridged-T network [10] shown in Fig. 2.3a. This network realizes a high- Q bandstop or notch response using only passive components. In general, the passive components and in particular the integrated inductor, limit the Q of the filter and ultimately the notch attenuation to a few dB. Even with thick Al-Cu metal and patterned ground shield technology, the Q of an integrated inductor peaks at about 25. This fails to meet the requirements of the notch filter ($Q > 100$).

¹ The act of degrading the receiver’s sensitivity is commonly known as desense.

The bridged-T network overcomes this limitation by transforming the real resistor R_N to a resistance equal and opposite to the resistance R_Q that models the losses of the LC network. The bridged-T network includes a Y-network formed by capacitors $C_{1a} - C_{1b}$ and resistor R_N . This network maps to an equivalent Δ -network where the branch replacing the two capacitors consists of an impedance with a negative real part. The resulting negative resistance cancels the losses modeled by resistor R_Q when

$$\omega = \sqrt{\frac{2}{LC}} \quad \text{and} \quad R_N = \frac{1}{2CR_Q}. \quad (2.1)$$

This in turn forces the output current to zero and creates a null at the resonance frequency. With less loss, the Q of the filter increases and the notch attenuation easily exceeds 20 dB over a 5 MHz bandwidth as shown in Fig. 2.3b.

Integrating the notch filter keeps the interface between the LNA and the down-converting mixers on-chip, which in turn allows for a flexible impedance level and ultimately lower power consumption. It also greatly simplifies multi-band operation. That's because traditionally, the transmit leakage signal has been attenuated using SAW filters. And since each radio band needs a dedicated SAW filter, this creates an untenable solution for multi-band radio receivers. In contrast, integrating the notch filters provides a fully-monolithic, low-cost receiver.

The down converter simply translates the entire receive band to baseband. The positive-frequency local oscillator (LO) signal described by $\cos \omega_c t + j \sin \omega_c t$ shifts the negative-frequency components of the receive band to a low-IF frequency as shown in Fig. 2.4a or directly to dc as shown in Fig. 2.4b. Oftentimes, the down converter is plagued by various artifacts.

Noise (especially $1/f$ noise), dc offsets, and even order distortion are some of the artifacts that degrade the downconverter. Narrowband GSM/EDGE signals are especially sensitive to these effects. That's because these modulated signals concentrate energy around dc where these artifacts generally lie.² Using a low-IF approach

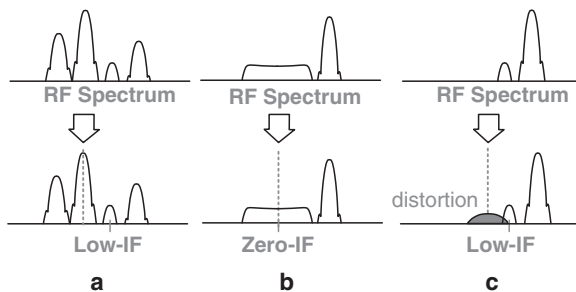


Fig. 2.4 Response of the pseudo-direct down conversion receiver

² For a GSM signal, 20% of its power lies below 12 kHz. Moreover, notching this low frequency energy degrades SNR 1 dB.

moves the wanted signal away from dc and mitigates the effects of these artifacts. It does however make the down converter sensitive to image signals. Any imbalance in the down converter and its mixers shift the image signals to the same IF frequency with

$$IRR = \frac{1 - \alpha(1 + \epsilon) \cos \theta + (1 + \epsilon)^2}{1 + \alpha(1 + \epsilon) \cos \theta + (1 + \epsilon)^2} \quad (2.2)$$

where the metric *IRR* corresponds to the image rejection ratio, ϵ represents the voltage gain mismatch, and θ represents the phase mismatch [5]. It's good practice to keep the image signal at least 20 dB below the wanted signal to avoid irreparable damage. GSM/EDGE networks restrict frequency re-use so that the adjacent channel interfering signals or blockers are at most only 9 dB stronger than the wanted signal – making the *IRR* requirements reasonable. This allows a low-IF approach to be chosen with a variable frequency of 100–140 kHz.

In contrast, the adjacent channel blockers found in WCDMA networks can be up to 44 dB stronger than the wanted signal. This increases the *IRR* requirements to an impractical level and makes the low-IF approach unattractive for WCDMA. Fortunately, the wide bandwidth of the WCDMA signal allows a dc notch (realized as a high pass filter with a 5 kHz corner³) to attenuate $1/f$ noise and dc offsets. This requires the down converter to slide the low-IF frequency to zero (or dc) for WCDMA.

But direct conversion receivers are especially sensitive to even order distortion. This type of distortion *demodulates* the amplitude modulation associated with the blocker signals and shifts the energy to dc – regardless of the carrier frequency. As a result, the spectrum of the distorted AM signal now overlaps the wanted down-converted signal. Using a low-IF approach however, doesn't completely mitigate this problem. That's because the spectrum of the distorted AM signal can still potentially extend into the wanted signal and cause problems as shown in Fig. 2.4c. Careful design and calibration is needed to suppress the even order distortion in both low-IF and zero-IF modes.

2.1.2 Baseband Receiver

The baseband receiver *selects* the received channel. It does this by processing the entire receive band, attenuating interfering signals, and isolating the wanted signal. The baseband receiver includes simple analog filters to reduce the strongest blocker signals. These filters use Butterworth structures and an all-pass equalizer [11] to minimize phase distortion and nonlinear group delay that otherwise causes intersymbol interference. The baseband receiver also includes variable gain amplifiers (VGAs) to adjust the signal level. An automatic gain control (AGC) loop controls these amplifiers so that the peaks of the aggregate input signal (wanted signal plus any remaining interfering signals) driving the A/D converters are within 6–10 dB of its full-scale level to allow for sudden increases in the received signal level [12].

³ The 5 kHz corner degrades the SNR of the WCDMA signal less than 0.1 dB.

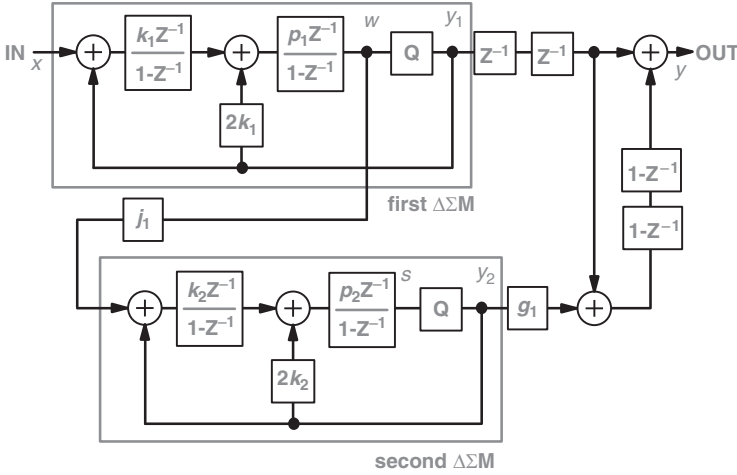


Fig. 2.5 Oversampled $\Delta\Sigma$ A/D converter based on 2-2 MASH structure (z-domain model)

The wireless propagation channel is unpredictable and is subject to small-scale effects that introduce frequency-selective fading [13]. In general, narrowband signals such as GSM/EDGE are more susceptible to frequency-selective fading than wideband signals like WCDMA. In addition, multi-slot GSM/EDGE operation compounds the situation and increases the probability of suddenly strong received signals appearing known as *up-fades* [12].

The A/D converters sample the analog radio signal and translate it to digital format for processing by the digital radio. A reconfigurable architecture based on the oversampled $\Delta\Sigma$ modulator shown in Fig. 2.5 is used to shape the quantization noise. It's based on a 2-2 cascaded MASH structure [14-16]. The output of the first $\Delta\Sigma$ modulator equals

$$y_1(z) = x(z)z^{-2} + e_1(z)(1 - z^{-1})^2 \quad (2.3)$$

where $e(z)$ represents the quantization noise from the single-bit quantizer and $k_1p_1 = 1$. The input to this quantizer $w(z)$ equals $y_1(z) - e_1(z)$ and drives the second $\Delta\Sigma$ modulator to produce the output

$$y_2(z) = j_1 \left[w(z)z^{-2} + e_2(z)(1 - z^{-1})^2 \right] \quad (2.4)$$

when $k_2p_2 = 1$. This then expands to

$$y_2(z) = j_1 \left[x(z)z^{-2} + e_1(z)(1 - z^{-1})^2 - e_1(z) \right] z^{-2} + e_2(z)(1 - z^{-1})^2. \quad (2.5)$$

The digital error cancellation logic combines $y_1(z)$ and $y_2(z)$ in a way that yields the desired fourth order noise shaping

$$y(z) = x(z)z^{-4} + e_1(z)(1 - z^{-1})^4 \quad (2.6)$$

when $j_1g_1 = 1$.

The fourth order response of the 2–2 cascaded MASH architecture shapes the power spectral noise density due to the single bit quantizer according to

$$n(f) = \sigma_n^2 2T [2 \sin(\pi f T)]^8 \quad (2.7)$$

where the variance of the quantization error equals

$$\sigma_n^2 = \frac{\Delta^2}{12}, \quad (2.8)$$

T corresponds to the sample period, and Δ represents the resolution of the single-bit quantizer embedded in the oversampled A/D converter [12]. Figure 2.6 shows the simulated noise power spectral density of this architecture and the theoretical curve for a fourth-order $\Delta\Sigma$ -modulated quantizer.

In practice, the dynamic range of the A/D converter depends on the sampling period T and its signal or equivalent noise bandwidth. A sampling clock of 26 MHz is chosen for GSM/EDGE. This results in an oversampling rate (OSR equals $f_s/2B$) of 96. By comparison, the sampling clock is increased to 156 MHz for WCDMA (OSR equals 40) to accommodate its wider signal bandwidth.

It's not uncommon for high dynamic range A/D converters to be ultimately limited by thermal noise. This sets up a compromise since the sampling capacitor at the A/D converter's input sets or influences the thermal noise level (kT/C),

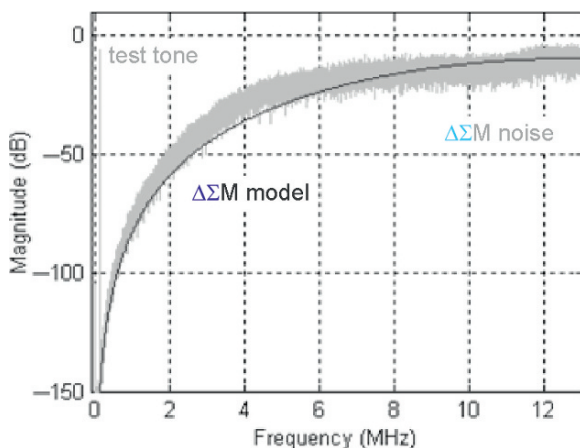


Fig. 2.6 Noise shaping response of oversampled $\Delta\Sigma$ A/D converter

minimum sampling period, power consumption, and gain accuracy (parameters k_1 , k_2 , ρ_1 , ρ_2 , j_1 , and g_1). Fortunately, this architecture is rather immune to gain errors.

An oversampled $\Delta\Sigma$ A/D converter generates a high-speed digital output. To transform the short words at high sampling rate to longer words at or near the Nyquist rate requires a sync filter. For this $\Delta\Sigma$ modulator, a fifth order *sinc* function is chosen [17]. In GSM/EDGE mode, the data stream is also digitally down converted to zero-IF. Additional digital filters remove any remaining blocker signals.

The digital receiver also includes algorithms to identify dc offsets, I/Q imbalance, and even order distortion so that adjustments and corrections can be implemented in the analog section. These nonidealities – if unattended – would otherwise add to the various noise sources and further degrade the quality of the received signal.

I/Q imbalance can be a serious problem. It distorts the received signal by *leaking* a portion of the signal $I(t)$ to the signal $Q(t)$ and vice-versa as shown in Fig. 2.7a. This in turn shifts the relative position of the received symbols compared to the ideal symbol points as Fig. 2.7b illustrates. The shift is measured using the metric error vector magnitude (EVM). A large enough shift results in decision errors. Moreover, I/Q imbalance can be characterized by the cross-correlation between the $I(t)$ and $Q(t)$ signals with

$$R_{IQ} \approx \frac{1}{M} \sum_{n=1}^M I(n)Q(n) \quad (2.9)$$

where the signals have been sampled at four times the symbol or chip rate. Higher modulation formats such as 16QAM and especially 64QAM require especially low EVM – less than 6.5% – for proper demodulation.

A digital approach to remove I/Q imbalance is shown in Fig. 2.8a [18]. It couples a fraction of the received in-phase signal $I(t)$ to the quadrature-phase signal $Q(t)$ and thereby cancels any leakage due to phase imbalance. A simple gain adjustment removes any amplitude imbalance. Ideally, the I/Q signals now represent the true orthogonal components of the received signal without any image signal energy. This

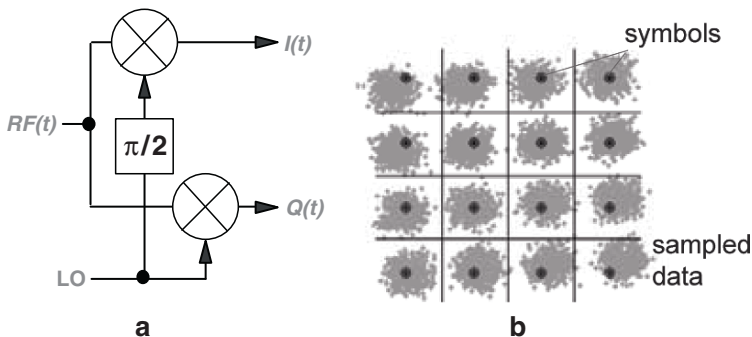


Fig. 2.7 I/Q down converter – (a) diagram and (b) effect of I/Q imbalance

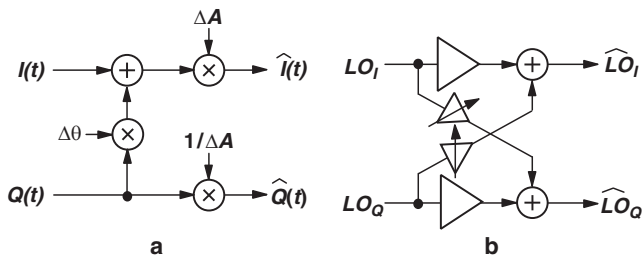
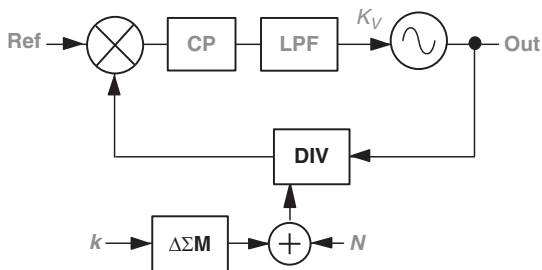


Fig. 2.8 Approaches to reduce I/Q imbalance – (a) digital and (b) analog

Fig. 2.9 Fractional-N PLL and integrated VCO



however only occurs if the digital I/Q signals are undistorted. Invariably the signals include intermodulation distortion, group delay distortion, and noise which reduce the effectiveness of digital I/Q balancing. As such, it’s better to make adjustments at the downconverting mixers. A simple and effective means to do this is the technique shown in Fig. 2.8b. It shifts the phase of each LO signal by adding to it a small amount of its orthogonal counterpart. If adjusted properly, this removes any phase error generated by the I/Q down converter.⁴ This is important since phase error is the primary cause of I/Q imbalance.

Nonlinear group delay can present another serious problem. Fortunately, digital FIR filters can be designed with symmetric coefficients and linear phase response. Still some demanding applications may require an equalizer [19] to mitigate the effects of the wireless channel as well as the analog receiver.

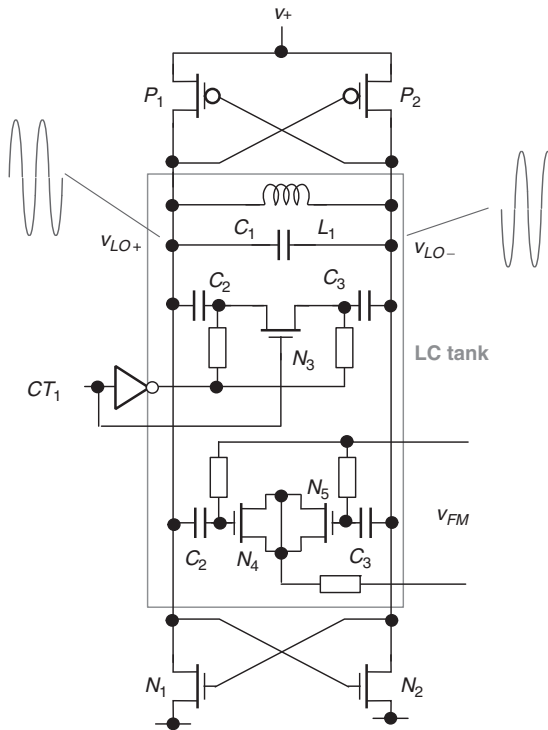
2.1.3 PLL

The local oscillator signals that drive the down converter are formed by the integrated voltage-controlled oscillator (VCO) and fractional-N phase-locked loop (PLL) shown in Fig. 2.9. The VCO operates near 4GHz at two or four times the RF carrier to allow a simple divider to generate the orthogonal signals needed by the I/Q downconverter. Table 2.2 lists the radio bands supported by the radio transceiver.

⁴ I/Q imbalance is similar to image rejection and is analyzed using Eq. (2.2).

Table 2.2 Radio bands supported by transceiver (Frequencies in MHz)

Band	Tx	Rx	Modes
I	1920–1980	2110–2170	WCDMA only
II	1850–1910	1930–1990	GSM, EDGE, WCDMA
III	1710–1785	1805–1880	GSM, EDGE, WCDMA
IV	1710–1755	2110–2155	WCDMA only
V	824–849	869–894	GSM, EDGE, WCDMA
VI	830–840	875–885	GSM, EDGE, WCDMA
VIII	880–915	925–960	GSM, EDGE, WCDMA
IX	1750–1785	1845–1880	WCDMA only
X	1710–1770	2110–2170	WCDMA only

**Fig. 2.10** Integrated VCO uses complimentary MOS devices to reduce current consumption and coarse-tuning to subdivide 1 GHz tuning range

The VCO uses complimentary MOS differential pairs as shown in Fig. 2.10 to reduce current consumption. In this design, the sustaining current flows through the full inductor (alternating direction each half cycle), unlike the single-type VCO structure that uses only PMOS or NMOS transistors. This VCO design also includes six coarse tuning bits [20] to cover the required 1 GHz frequency range and keep the sensitivity reasonable. In practice, VCO even-order distortion disturbs the orthogonal signals from the divide-by-2/4 circuit and produces phase error.

The fractional-N PLL includes a feedback counter controlled by a third order $\Delta\Sigma$ modulator ($\Delta\Sigma M$). Its output is 3 bits to guarantee stability, while its resolution (b) is 23 bits to provide fine frequency adjustment equal to

$$\Delta f = \frac{f_{VCO}}{2^b} \quad (2.10)$$

and suitable for automatic frequency control (AFC) requirements.

The $\Delta\Sigma$ modulator and fractional-N phase-locked loop architecture easily synthesize the required RF channel frequencies without dictating the reference frequency or constraining the loop bandwidth. This allows the loop bandwidth to be set to approximately 100 kHz which satisfies the settling time requirements for compressed mode operation ($\leq 200 \mu s$). For a type-II PLL, the settling time can be approximated by

$$t_s \approx \frac{1}{\zeta \omega_n} \ln \left(\frac{k}{M |\alpha| \sqrt{1 - \zeta^2}} \right) \quad (2.11)$$

where ζ and ω_n correspond to the PLL loop parameters for the damping factor and natural frequency, M represents the initial feedback divider setting, k equals the step change in the divider value, α corresponds to the settling accuracy [21].

The PLL loop bandwidth also shapes the phase noise profile of the local oscillator. It's set to suppress any high-frequency noise generated by the $\Delta\Sigma$ modulator. Normally, the output of the $\Delta\Sigma$ modulator is random. However, in some situations (such as when the input to the $\Delta\Sigma$ modulator is a rational fraction like $1/2, 1/4, \dots$), the output follows a repeating pattern. Unfortunately, any pattern maps to spectral lines in both the frequency spectrum of both the $\Delta\Sigma$ modulator output and the PLL output. If unattended, these spectral lines can potentially mix with strong interfering signals to deleteriously affect the receiver – a phenomena known as reciprocal mixing. It's therefore necessary to disrupt the output pattern by applying a *busy* input to the $\Delta\Sigma$ modulator. This is accomplished by a pseudo-random (pn) generator or dither signal [22]. Its amplitude is generally small to minimize noise. Note that if the pn signal is too small, nonlinearities in the phase detector and charge pump can cause the spectral lines to reappear as illustrated in Fig. 2.11.

2.2 Transmitter

The transmitter is shown in Fig. 2.12 and is architected to possess the same flexibility as the receiver. It uses polar modulation – the preferred approach for GSM/EDGE [23,24] – to efficiently form all three transmit waveforms. The polar transmitter consists of a digital processor, fractional-N PLL, divider, and VGA plus PA.

The transmitter translates the I/Q data to polar format using the mapping functions

$$AM(t) = \sqrt{I^2(t) + Q^2(t)} \quad PM(t) = \tan^{-1} \left[\frac{Q(t)}{I(t)} \right]. \quad (2.12)$$

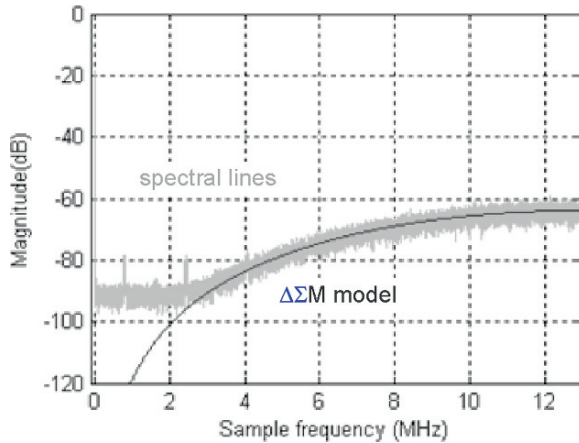


Fig. 2.11 Nonlinear PLL behavior produces spectral lines otherwise spread by the input dither signal

Fig. 2.12 The polar modulation concept

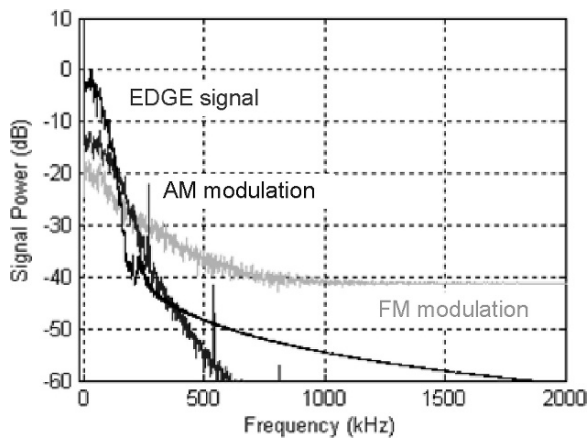
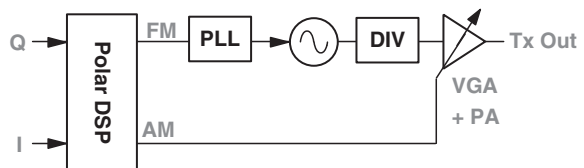


Fig. 2.13 An EDGE signal expands when separated into AM and FM components

This process, although simple, operates nonlinearly and tends to spread the power spectral density of the resulting modulation signals as shown in Fig. 2.13. The AM signal includes spectral lines at dc as well as multiples of the symbol rate. Its energy extends to about twice the symbol rate. By comparison, the power spectral density

of the FM signal falls slowly and never really dissipates.⁵ The spectrum expansion associated with the AM and FM signals complicates the system design. Moreover, this effect grows with wideband systems and presents a major obstacle to extending polar modulation to WCDMA operation.

2.2.1 Phase/Frequency Modulation

In a polar transmitter, the phase/frequency modulation is applied directly to the RF carrier using the modified fractional-N PLL shown in Fig. 2.14. The approach, sometimes referred to as “2-point modulation” [25], applies the FM signal at the feedback counter (low-frequency path) as well as the VCO (high-frequency path). This leads to the behavior shown in Fig. 2.15 and the following transfer functions

$$\Delta f = \frac{K_{PD}Z(s)K_V}{sN + K_{PD}Z(s)K_V} FM \tag{2.13}$$

$$\Delta f = \frac{sNK_{FM}}{sN + K_{PD}Z(s)K_V} \alpha FM$$

where K_{PD} is the charge pump’s gain, $Z(s)$ is the impedance presented by the loop filter, K_V is the VCO’s sensitivity at the tuning port, N is the value of the feedback counter, K_{FM} is the VCO’s gain at the modulation port, and α is a scaling parameter. Ideally, these two functions combine to realize a flat response – with the high-frequency analog path through the VCO enabling wide bandwidth WCDMA operation.

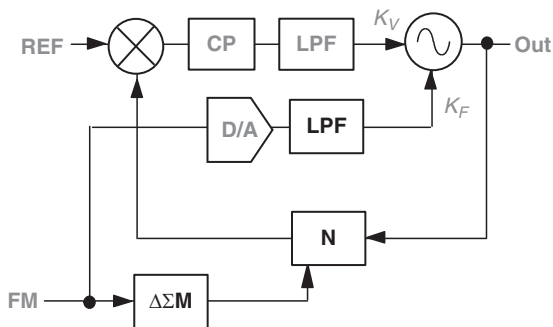


Fig. 2.14 A fractional-N PLL modified to include two modulation paths

⁵ By its nature, the transmitter better realizes frequency modulation than phase modulation. To form the FM signal requires differentiating the PM signal, which unfortunately further widens the modulation spectrum.

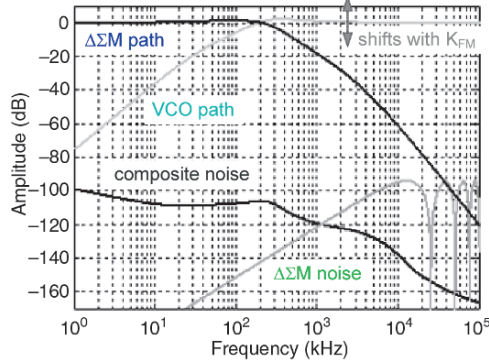


Fig. 2.15 Wideband response of the modified PLL with $\Delta\Sigma M$ and other noise sources attenuated

In practice, the response of the high-frequency analog path is extremely sensitive to the VCO gain K_{FM} – which changes with process as well as operating frequency. As a result, the analog FM signal must be properly scaled by α to achieve the desired flat response [26]. In contrast, the low-frequency digital path through the feedback counter is exact and is actually able to minimize errors within the PLL’s loop bandwidth caused by the analog path.

The PLL and VCO produce a constant-amplitude signal at two or four times the frequency of the RF carrier. Operating the VCO at a frequency different from the PA reduces the chances of injection pushing [27] by the high-power PA output. Still, careful design is needed to avoid *pushing* by the PA’s second (or fourth) harmonic output. Any pushing remodulates the VCO and results in significant phase error. The VCO phase/frequency-modulated signal is shifted on-frequency using a simple divider.

2.2.2 Amplitude Modulation

The amplitude modulation in a polar transmitter occurs at the VGA or PA. It’s rather straightforward to apply the modulation at the VGA using a commutating buffer. The VGA structure shown in Fig. 2.16 operates efficiently since its current level tracks the AM signal. The phase/frequency-modulated RF carrier drives transistors N_2 – N_3 to switch the signal current $AM(t)$ directly to the output. The signal $AM(t)$ is always positive since the RF carrier already includes any phase changes.⁶ This contrasts traditional I/Q mixer circuits that *need* to add a dc offset level to handle negative-going signals $I(t)$ and $Q(t)$. The dc offset consequently dictates that less-efficient, double-balanced mixer circuits be used in the I/Q upconverter to avoid

⁶ Since the AM signal is always positive, its average or mean value shows up as a spectral line at dc in its PSD.

Fig. 2.16 A commutating driver amplifier

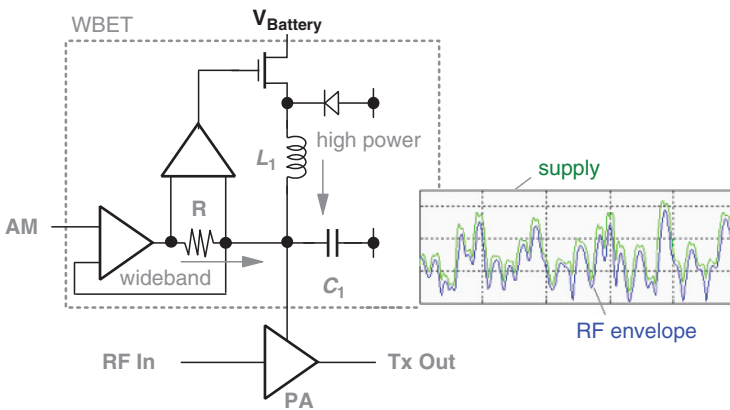
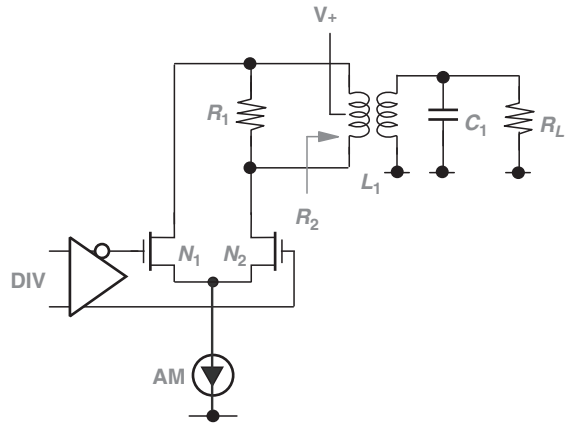


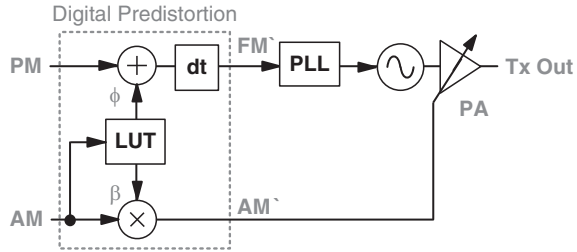
Fig. 2.17 WCDMA PA modulation based on wide bandwidth envelope tracking (WBET)

carrier leakage. As a result, the amplitude modulator operates much more efficiently than the I/Q upconverter [28].

Even better efficiency is possible if the modulation is applied at the PA, but this is more challenging due to the bandwidth of the AM signal, the dynamic range required of the transmitter, and wideband noise limits. At high output power levels, the PA can draw more than 1A from its supply during the signal peaks. This makes the efficiency of the PA critical. A linear PA offers poor efficiency while a more-efficient switched or saturated PA severely degrades EDGE performance and literally destroys WCDMA signal quality.

One promising way to apply wideband modulation at the PA is to use the wideband envelope tracking scheme (WBET) [29] depicted in Fig. 2.17. Originally developed for class D audio amplifiers, this approach relies on a switched, high-efficiency network to supply most of the PA bias and combines it with a linear

Fig. 2.18 Digital predistortion applied to polar transmitter



network to provide the wideband portion of the bias. The linear network simply acts to reduce the error between the switched network and the envelope level. As a result, the supply voltage *tracks* the envelope of the modulated signal; it rides slightly above the required level needed to keep the PA operating in its linear region. Since the PA operates linearly, the distortion is greatly reduced compared to switched-mode PA designs.

Any PA modulation approach invariably introduces AM–AM and AM–PM distortion. This distortion can potentially be corrected with the digital predistortion method shown in Fig. 2.18. The predistortion algorithm maps directly to the polar architecture using a look-up table (LUT) with minimal complexity according to

$$\begin{aligned} AM'(t) &= \beta(AM) \cdot AM(t) \\ PM'(t) &= PM(t) - \phi(AM) \end{aligned} \quad (2.14)$$

where β is the gain adjustment and ϕ is the phase shift associated with the PA at a given AM level. By comparison, gain and phase correction algorithms targeting I/Q upconverters generally require at least four to ten times more complexity [30].

The amount of distortion produced by the PA depends on its mode of operation. It can vary with operating frequency, temperature, and load impedance. Moreover, switched or saturated power amplifiers typically demonstrate memory phenomena that further complicate predistortion schemes [31, 32]. As such, it's common for PA modulation approaches using a switched PA to include feedback. In order to accurately track both amplitude and phase distortion, the feedback must include a dedicated receiver as illustrated in Fig. 2.19 [33–35]. This places a tremendous burden on the radio transceiver and highlights a key advantage of the WBET approach based on a linear PA.

2.2.3 Rx Band Noise

A key benefit of the multimode polar transmitter is that it eliminates all the external SAW filters – dramatically lowering the cost and shrinking the size of the radio. These filters are generally required to reduce noise in the receive bands. Although

Fig. 2.19 Transmitter feed-back using a dedicated receiver

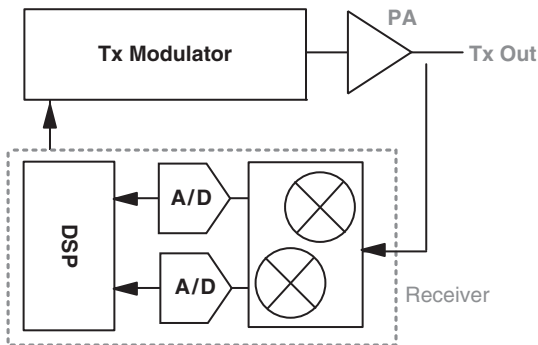
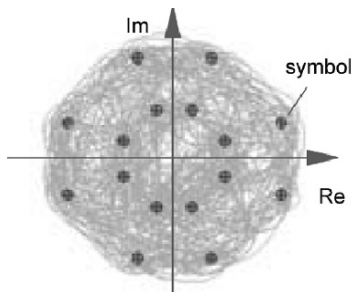


Fig. 2.20 Trajectory of a complex WCDMA signal



this is especially important in full duplex systems like WCDMA, the noise levels for all three modes are challenging.

The energy produced at the PA’s output in the receive band is a combination of its own circuit noise, amplified noise originally generated by the RF modulator, and noise folded into the receive band due to nonlinear PA operation. Noise folding due to intermodulation distortion shifts noise at the image frequency given by

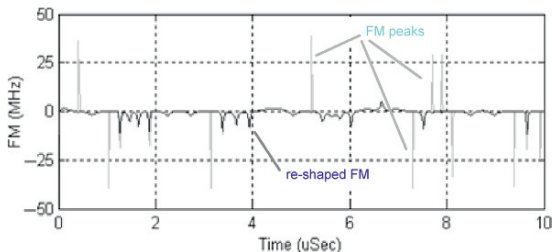
$$f_{image} = 2f_{Tx} - f_{Rx} \tag{2.15}$$

to the receive frequency f_{Rx} . Traditional SAW filters attenuate both the receive band and image noise. A notch filter such as the design used in this radio receiver does not attenuate the image noise.

For GSM/EDGE, the receive noise is measured at offsets of 10 and 20 MHz. With narrowband modulation, the main noise contributor is the VCO. Fortunately, the noise level of this circuit can approach -165 dBc/Hz with careful design and reasonable components (loaded tank $Q > 12$). In contrast, the wideband modulation associated with WCDMA presents a daunting problem. That’s because the power spectral density of the FM signal falls off slowly and cannot be filtered easily.

The spectrum of the composite WCDMA transmit signal (formed by the AM and PM/FM components) is defined by a root raised-cosine (RRC) pulse shape filter [13]. This filter dictates the trajectory of the complex time-domain transmit signal and it’s not uncommon for the trajectory to pass through or near the origin as depicted in Fig. 2.20. As the signal passes through or near the origin, its AM

Fig. 2.21 Re-shaping removes large FM signal peaks



component rapidly decreases and then increases while its PM component jumps as much as $\pm\pi$. The phase jump spikes the FM component as shown in Fig. 2.21 and expands the spectrum bandwidth of the FM signal.

The complex transmit signal is described in polar form by

$$S(t) = AM(t) \cdot \sin \left[\omega t + 2\pi \int FM(t) dt \right]. \tag{2.16}$$

This is equivalent to

$$S(f) = AM(f) * FM(f) \tag{2.17}$$

in the frequency domain, where $FM(f)$ is the spectrum resulting from the frequency modulation of the RF carrier and $*$ is the convolution operation. This process combines the polar modulation signals in a way that collapses the spectrum of the composite transmit signal back to its original shape. Unfortunately, noise and other unavoidable artifacts elevate the wideband energy. As a result, the effective bandwidth of the transmit signal $S(f)$ approaches the sum of the spectrum bandwidths for the two modulated AM and FM signals.

The spectrum bandwidth of the signal $FM(f)$ can be estimated using Carson’s rule [36]

$$BW \approx 2 (\Delta f + f_m) \tag{2.18}$$

where Δf represents the peak frequency deviation and f_m the associated bandwidth of the FM signal. Consequently, to minimize emissions and ultimately noise in the receive band, the sum of the $AM(f)$ and $FM(f)$ bandwidths must be less than 45 MHz for operation in radio bands V, VI, and VIII.

Filtering the AM and FM signals reduces the spectrum of these signals and their wideband energy, but it unavoidably degrades emission levels and the EVM of the composite transmit signal. That’s because any practical filter affects the FM component, alters the trajectory of the complex signal, and distorts the convolution process in a way that invariably results in spectral regrowth.

To avoid these problems, the AM and FM signals must be re-shaped in a way that essentially preserves as much of the trajectory of the complex signal as possible while smoothing out the discontinuities that widen the signal bandwidths [37]. This has led to the development of a set of DSP algorithms that reform the FM signal as illustrated in Fig. 2.21. The algorithms limit the peak frequency deviation Δf and

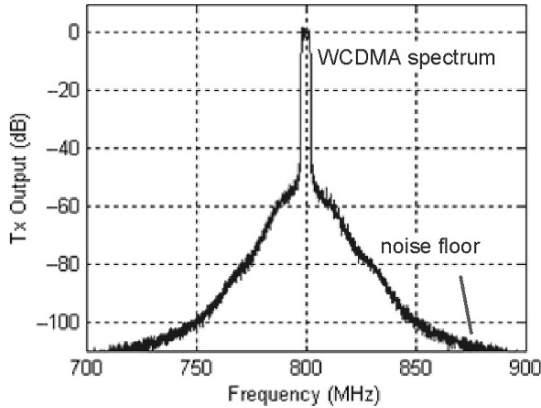


Fig. 2.22 Simulated WCDMA output spectrum with ultra-low Rx band noise

reduce the spectrum of the AM and FM signals, while they preserve the trajectory of the complex signal. The result is the WCDMA (HSUPA) transmit spectrum shown in Fig. 2.22. The simulated receive band noise drops to -163 dBc/Hz at 45 MHz offset while the EVM stays below 2.5%.

2.3 Summary

The radio transceiver presented provides an efficient, low-cost, solution to the complex, multimode problem. It achieves a single path for both the receiver as well as the transmitter with enough flexibility to adapt to very different signals. The receiver relies on integrated, tunable, notch filters to realize a completely monolithic approach. The polar transmitter includes digital algorithms to fully exploit amplitude modulation and to lower receive band noise. As a result, the external SAW filters traditionally needed by the radio have been eliminated.

In the future, flexible radio transmitters such as the one presented will directly drive multimode PA's and further simplify the transceiver. Moreover, the need for multimode radios will continue to grow as new systems, such as 3GPP Long Term Evolution (LTE) [38], develop.

References

1. J. Crols and M. S. J. Steyaert, "A Single-Chip 900 MHz CMOS Receiver Front-End with a High Performance Low-IF Topology", *IEEE Journal of Solid-State Circuits*, Dec., 1995, pp. 1483–1492.
2. J. Crols and M. S. J. Steyaert, "Low-IF Topologies for High-Performance Analog Front Ends of Fully Integrated Receivers", *IEEE Transactions on Circuits and Systems II*, Mar., 1998, pp. 269–282.

3. S. Cipriani, G. Sirna, P. Cusinato, L. Carpineto, F. Monchal, C. Sorace, and E. Duvivier, "Low-IF 90 nm CMOS Receiver for 2.5G Application", *Proceedings of the 12th IEEE Mediterranean Electrotechnical Conference*, May, 2004, pp. 151–154.
4. A. A. Abidi, "Direct-Conversion Radio transceivers for Digital Communications", *IEEE Journal of Solid-State Circuits*, Dec., 1995, pp. 1399–1410.
5. B. Razavi, "Design Considerations for Direct-Conversion Receivers", *IEEE Transactions on Circuits and Systems II*, June, 1997, pp. 428–435.
6. D. Manstretta, R. Castello, F. Gatta, P. Rossi, and F. Svelto, "A 018 μm CMOS Direct-Conversion Receiver Front-End for UMTS", *2002 IEEE International Solid-State Circuits Conference*, Feb., 2002, pp. 240–463.
7. J. Groe, "Polar Transmitters for Wireless Communications", *IEEE Communications Magazine*, Sept., 2007, pp. 58–63.
8. J. Groe, "A Multimode Cellular Radio", *IEEE Transactions on Circuits and Systems II*, Mar., 2008, pp. 269–273.
9. M. S. Khan and N. Yanduru, "Analysis of Self Mixing of Transmitter Interference in WCDMA Receivers", *2006 International Symposium on Circuits and Systems*, May, 2006, pp. 5451–5454.
10. B. Bauer, "Distortion Measuring Equipment", *HP Technical Journal*, Aug., 1951.
11. A. B. Williams, *Handbook on Electronic Filter Design*, McGraw-Hill, New York, 1980.
12. J. Groe and L. Larson, *CDMA Mobile Radio Design*, Artech House, 2000.
13. T. Rappaport, *Wireless Communications – Principles and Practice*, Prentice-Hall, New York, 1996.
14. A. Rusu, D. Rodriguez de Llera Gonzalez, and M. Ismail, "Reconfigurable ADCs Enable Smart Radios for 4G Wireless Connectivity", *IEEE Circuits & Device Magazine*, May, 2006, pp. 6–11.
15. T. Karema, T. Ritonieni, and H. Tenhunen, "An Oversampled Sigma-Delta A/D Converter Circuit Using Two-Stage Fourth Order Modulators", *1990 IEEE International Symposium on Circuits and Systems*, 1990, pp. 3279–3282.
16. H. Baher and E. Afifi, "A Novel Switched-Capacitor Cascade Structure for Sigma-Delta Converters", *1992 Proceedings of the 34th Midwest Symposium on Circuits and Systems*, 1992, pp. 1106–1107.
17. J. C. Candy, "Decimation for Sigma Delta Modulation", *IEEE Transactions on Communications*, Jan., 1986, pp. 72–76.
18. J. K. Cavers and M. W. Liao, "Adaptive Compensation for Imbalance and Offset Losses in Direct Conversion Transceivers", *IEEE Transactions on Vehicular Technology*, Nov., 1993, pp. 581–588.
19. J. Proakis, *Digital Communications*, McGraw-Hill, New York, 1995.
20. H. Sjoland, "Improved Switched Tuning of Differential CMOS VCOs", *IEEE Transactions on Circuits and Systems II*, May, 2002, pp. 352–355.
21. B. Razavi, *RF Microelectronics*, Prentice-Hall, Upper Saddle River, NJ, 1998.
22. I. Galton, "Granular Quantization in a Class of Delta-Sigma Modulators", *IEEE Trans. On Information Theory*, May, 1994, pp. 848–859.
23. M. R. Elliott, T. Montalvo, B. P. Jeffries, F. Murden, J. Strange, A. Hill, S. Nandipaku, and J. Harrebek, "A Polar Modulator Transmitter for GSM/EDGE", *IEEE Journal of Solid-State Circuits*, Dec., 2004, pp. 2190–2199.
24. A. W. Hietala, "A Quad-Band 8PSK/GMSK Polar Transceiver", *IEEE Journal of Solid-State Circuits*, May, 2006, pp. 1133–1141.
25. C. Durdodt et al., "A Low-IF Rx Two-Point $\Delta\Sigma$ -Modulation Tx CMOS Single-Chip Bluetooth Solution", *IEEE Transactions on Microwave Theory and Techniques*, Sept., 2001, pp. 1531–1537.
26. J. Groe, "Highly Linear Phase Modulation", US patent 10/420,952.
27. B. Razavi, "A Study of Injection Locking and Pulling in Oscillators", *IEEE Journal of Solid-State Circuits*, Sep., 2004, pp. 1415–1424.
28. J. Groe, "Low-Noise RF Modulators", submitted to *IEEE Transactions on Circuits and Systems II*.

29. D. Kimball et al., "High-Efficiency Envelope-Tracking WCDMA Base-Station Amplifier using GaN HFETs", *IEEE Transactions on Microwave Theory and Techniques*, Nov., 2006, pp. 3848–3856.
30. J. K. Cavers, "Amplifier Linearization Using a Digital Predistorter with Fast Adaptation and Low Memory Requirements", *IEEE Transactions on Vehicular Technology*, Nov., 1990, pp. 374–382.
31. P. M. Asbeck et al., "Augmented Behavioral Characterization for Modeling the Nonlinear Response of Power Amplifiers", *IEEE Microwave Theory and Techniques Symposium*, 2002, pp. 135–138.
32. W. Bosch and G. Gatti, "Measurement and Simulation of Memory Effects in Predistortion Linearizers", *IEEE Transactions on Microwave Theory and Techniques*, Dec., 1989, pp. 1885–1890.
33. T. Sowlati et al., "Quad-Band GSM/GPRS/EDGE Polar Loop Transmitter", *IEEE Journal of Solid-State Circuits*, Dec., 2004, pp. 2179–2189.
34. J. L. Dawson and T. H. Lee, "Automatic Phase Alignment for a Fully Integrated Cartesian Feedback Power Amplifier System", *IEEE Journal of Solid-State Circuits*, Sep., 2003, pp. 2269–2279.
35. W. Kim et al., "Digital Predistortion Linearizes Wireless Power Amplifiers", *IEEE Microwave Magazine*, Sep., 2005, pp. 54–61.
36. H. Taub and D. L. Schilling, *Principles of Communication Systems*, McGraw-Hill, New York, 1986.
37. J. Groe, "Spectrum Shaping of Polar Components and Composite Signal", US patent application 60/979,740.
38. H. Ekstrom et al., "Technical Solutions for the 3G Long-Term Evolution", *IEEE Communications Magazine*, Mar., 2006, pp. 38–45.

Chapter 3

Reconfigurable Multi-Band OFDM UWB Receivers: Circuits and System Considerations

Luca Baldini, Danilo Manstretta, Tomaso Erseghe, Nicola Laurenti,
Antonio Liscidini, and R. Castello

3.1 Introduction

ULTRA wideband (UWB) is intended to provide a standard for high-speed short range wireless communication [1, 2]. The ECMA 368 Standard [2] specifies the physical and medium access control layers for UWB networks using Multi-band Orthogonal Frequency Division Modulation (MB-OFDM). The spectrum from 3.1 to 10.6 GHz is divided into 14 bands of 528 MHz. Supported data rates range from 53.3 to 480 Mbps with a data-rate adaptation mechanism allowing each receiver to opt for the transmitter's data rate for the maximum throughput. The receiver RF front-end and the frequency synthesizer pose the highest design challenges. We will concentrate only on the former due to space limitations. Covering a broad bandwidth is quite challenging, especially for a CMOS implementation, where tuning out capacitive parasitics is the key to achieve low-power operation. An important aspect in UWB communication is the interference between different UWB devices and from other systems. Nearby wireless devices can produce in-band interference due to intermodulation and harmonic generation in the receiver front-end. In this chapter, we will evaluate the effects of interferers on the desired UWB signal, resulting from cross-modulation, intermodulation and harmonic distortion. Some of the existing receiver solutions will be reviewed and a receiver architecture with enhanced linearity performances will be described.

L. Baldini, D. Manstretta (✉), A. Liscidini, and R. Castello
Università degli Studi di Pavia, Pavia, Italy
e-mail: {luca.baldini; danilo.manstretta; antonio.liscidini; rinaldo.castello}@unipv.it

T. Erseghe and N. Laurenti
Università degli Studi di Padova, Padova, Italy
e-mail: {erseghe; nil}@dei.unipd.it

3.2 Receiver Specifications

Receiver sensitivity varies, depending on the data rate, from -80.8 dBm at 53.3 Mbps to -70.4 dBm at 480 Mbps. These figures are set by the standard in order to guarantee a packet error rate (PER) lower than 8% with a payload of 1,024 octets for each packet over an AWGN channel. The standard assumes a receiver noise figure (NF) of 6.6 dB referred to the antenna, 2.5 dB implementation loss and a 3-dB margin. In order to improve the robustness to interferences, two time-frequency codes can be used: one where the information is transmitted on a single band and one where the information is sent over three bands using time-frequency interleaving (TFI). Furthermore, each UWB terminal has the capability to dynamically change the channel in which it operates. The standard does not specify blockers tests; however, in-band and out-of-band interferers should be considered to ensure reliable operation in a real-world environment. We will first evaluate the effect of an in-band interferer on the system performance; then we will define an interferers' scenario and finally derive the receiver linearity specifications for each data rate. This extends the work presented in [3] by taking into account different link data rates and interferer bandwidths.

3.2.1 Interference Analysis

We begin our analysis by evaluating system robustness against in-band interferers. As customary for interference analysis [3, 4], we simulate transmission at different bit rates, from 53.3 to 480 Mbps, with the desired signal boosted by 6 dB with respect to the receiver sensitivity given by the standard [2]. In the simulation we assume ideal synchronization, both for the sake of a simpler analysis of the results and because synchronization algorithms tend to be more robust and less affected by interferers than the corresponding data detection [5]. As the band-hopping patterns offer some robustness against a fixed frequency interferer, we consider the constant band pattern for the UWB signal (corresponding to time-frequency codes 5, 6, 7 on all band groups) as a worst case. The interfering signal is modeled as a random process with either a rectangular, triangular or bell-shaped spectrum (the latter two would result from squaring or raising to the third power the rectangular spectrum signal). In particular, it can be effectively modeled as a Gaussian process for wide-band OFDM-based interferers. No scheme for interference cancellation is sought at the receiver, apart from the fact that the DFT output is limited according to the transmitted constellation in order not to load the Viterbi decoder with unrealistic values, coming from narrowband interferers. Similarly, the effect of quantization (equivalent to 8 bits at the Nyquist rate) at the receiver input is twofold: on the one hand it can limit impulsive interference, or the time distributed peaks of strong Gaussian interferer; on the other, the automatic gain control can be tricked into setting a quantization range that fits the strong interferer rather than the useful signal, thereby reducing the number of useful bits before demodulation. However, even quantization is seen to leave the system robustness to interference nearly unaffected.

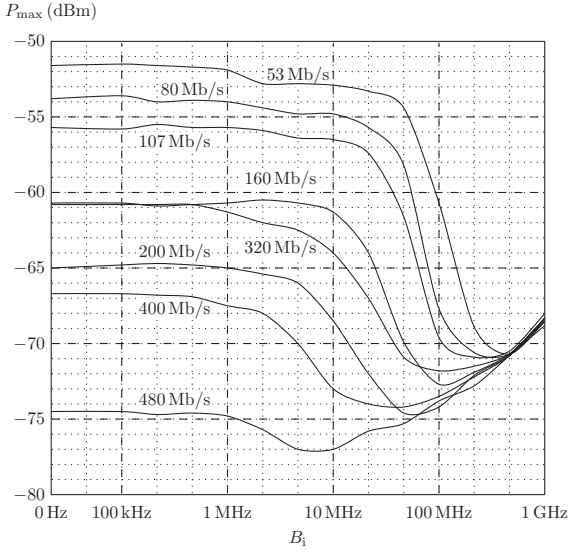


Fig. 3.1 Maximum interferer power allowed at the receiver versus interferer bandwidth (B_i) for in-band flat spectrum interferers. The useful signal is boosted by 6 dB over the receiver sensitivity, and the target PER is 8% with 1,024 octets per packet

We then seek to determine the maximum power for the interference at which the receiver can still comply with the required threshold of 8% in the PER. The results are shown in Fig. 3.1 for different bit rates and interferer bandwidth assuming a flat interference spectrum. We observe that the allowed interferer power is nearly constant for all bit rates as long as $B_i < 4$ MHz (i.e. the interferer band is not wider than one sub-carrier), then for the highest data rates it starts to decrease as B_i increases, until it hits a minimum around 20 MHz, then it increases again. For lower data rates, less dense constellations, time and frequency diversity and stronger coding allow the system to effectively contrast wideband interferers, so the decrease in allowed interferer power comes with bandwidths in excess of 100 MHz. Finally, as the interferer bandwidth approaches the signal bandwidth, all transmission rates converge to the same value. It might seem surprising that narrow band interferers still hinder transmission even if their nominal bandwidth takes only one (or just a few) subcarrier. However, one should consider that due to the sinc (T_f) shape of the sub-channel filters frequency response even a single tone interferer has some detrimental effect on all sub-channels. By comparing the results for the different bit rates we see that transmission at 200 Mbps offers a lower robustness to interference than transmission at 320 Mbps, thus showing the higher effectiveness of the dual carrier modulation used for higher rates with respect to the time-frequency spread QPSK used in lower rates. Moreover, slightly higher values are found for triangular and bell-shaped spectrum with respect to the flat spectrum case.

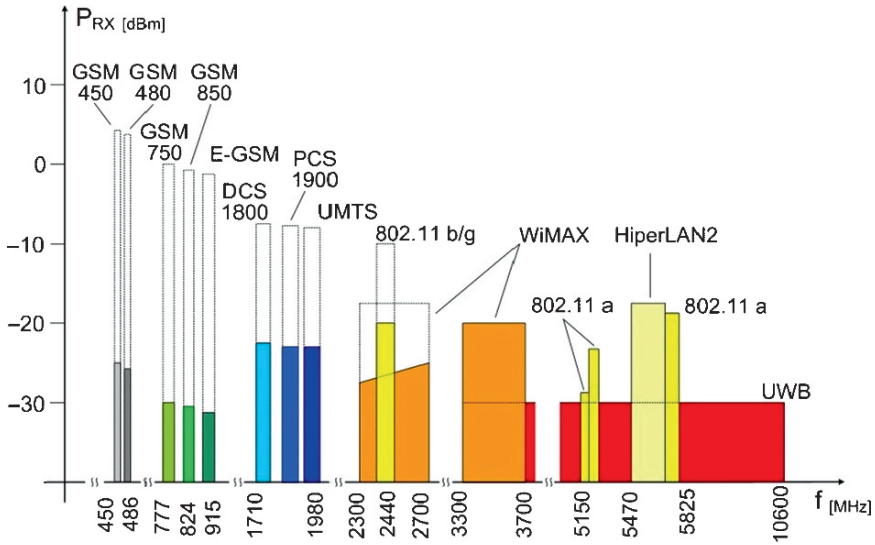


Fig. 3.2 Interferers spectrum profile at the receiver antenna (dotted blocks) and after the RF filter (colored solid blocks)

3.2.2 Interferers Scenario

In order to define the most realistic scenario we have considered a comprehensive set of wireless standards [2, 6–10], including UWB, second and third generation cellular, as well as wireless LAN systems. For each band we have considered the maximum power that each user terminal is allowed to transmit. Figure 3.2 shows the power received at an antenna standing at 1 m distance from the transmitter assuming a free space model for the path loss. The spectrum is dominated by signals generated by cellular systems. The power at the receiver input has been estimated assuming an RF filter is used [11]. In this case, WLAN signals that are located between 2.4 and 5.8 GHz become dominant as they pass through the RF filter with little or no attenuation.

3.2.3 Receiver Linearity Specifications

For each pair of interferers whose distortion product falls in the UWB spectrum we can derive a requirement in terms of IIP2 or IIP3. Tables 3.1–3.2 report the most stringent specifications for each UWB band in order to achieve a data rate of 480 Mbps. Specifications for lower data rates are in general more relaxed, due to the higher tolerance to in-band interferers, as can be seen in Figs. 3.3–3.4. The most stringent IIP3 requirement is set by the intermodulation of an 802.11 a/HiperLAN2

Table 3.1 Out-of-band IIP2 requirements for a data rate of 480 Mbps

Band #	Interferer 1 (f_{INT1})	Interferer 2 (f_{INT2})	IIP2
1	802.11a (5600 MHz)	UMTS (1920 MHz)	38.2 dBm
2	802.11a (5700 MHz)	UMTS (1920 MHz)	36.9 dBm
3	802.11b/g (2412 MHz)	UMTS (1920 MHz)	35.3 dBm
4	802.11b/g (2412 MHz)	802.11g (2412 MHz)	34.9 dBm
5	WiMAX (3500 MHz)	UMTS (1920 MHz)	35.2 dBm
6	WiMAX (3500 MHz)	802.11g (2412 MHz)	34.8 dBm
7	802.11a (5600 MHz)	GSM (785 MHz)	28.5 dBm
8	802.11a (5600 MHz)	DCS (1748 MHz)	36.5 dBm
9	802.11a (5600 MHz)	UMTS (1920 MHz)	38.2 dBm
10	802.11a (5600 MHz)	802.11g (2412 MHz)	37.9 dBm
11	WiMAX (5150 MHz)	WiMAX (3500 MHz)	31.3 dBm
12	802.11a (5600 MHz)	WiMAX (3500 MHz)	37.8 dBm
13	UWB (3960 MHz)	802.11a (5600 MHz)	23.1 dBm
14	WiMAX (5150 MHz)	WiMAX (5150 MHz)	27.9 dBm

Table 3.2 Out-of-band IIP3 requirements for a data rate of 480 Mbps

Band #	Interferer 1 (f_{INT1})	Interferer 2 (f_{INT2})	IIP3
1	WiMAX (3500 MHz)	WiMAX (3500 MHz)	7.2 dBm
2	UWB (7128 MHz)	802.11a (5600 MHz)	3.4 dBm
3	802.11b/g (2412 MHz)	WiMAX (3500 MHz)	7.2 dBm
4	UMTS (1920 MHz)	WiMAX (3500 MHz)	7.2 dBm
5	802.11a (5600 MHz)	802.11a (5600 MHz)	11.6 dBm
6	WiMAX (5150 MHz)	802.11a (5600 MHz)	8.4 dBm
7	802.11b/g (2412 MHz)	UMTS (1980 MHz)	7.7 dBm
8	WiMAX (3500 MHz)	UMTS (1920 MHz)	7.7 dBm
9	WiMAX (3500 MHz)	802.11a (5600 MHz)	10.1 dBm
10	WiMAX (3500 MHz)	802.11b/g (2412 MHz)	7.3 dBm
11	802.11b/g (2412 MHz)	802.11a (5600 MHz)	10.2 dBm
12	UMTS (1920 MHz)	802.11a (5600 MHz)	10.2 dBm
13	802.11a (5600 MHz)	UMTS (1980 MHz)	9.3 dBm
14	802.11a (5600 MHz)	802.11b/g (2412 MHz)	8.7 dBm

interferer with an 802.11b/g (or an UMTS) interferer, giving an intermodulation product in the 11th (12th) UWB band. By comparison, intermodulation between two adjacent UWB channels determines a minimum IIP3 requirement of only -10 dBm.

3.3 Design Challenges in UWB Receiver Front-Ends

A common challenge to all UWB receiver implementations is the design of the LNA and the down-conversion mixers. They limit the noise figure and the linearity of the overall receive chain and, therefore, they define system sensitivity and output SNR.

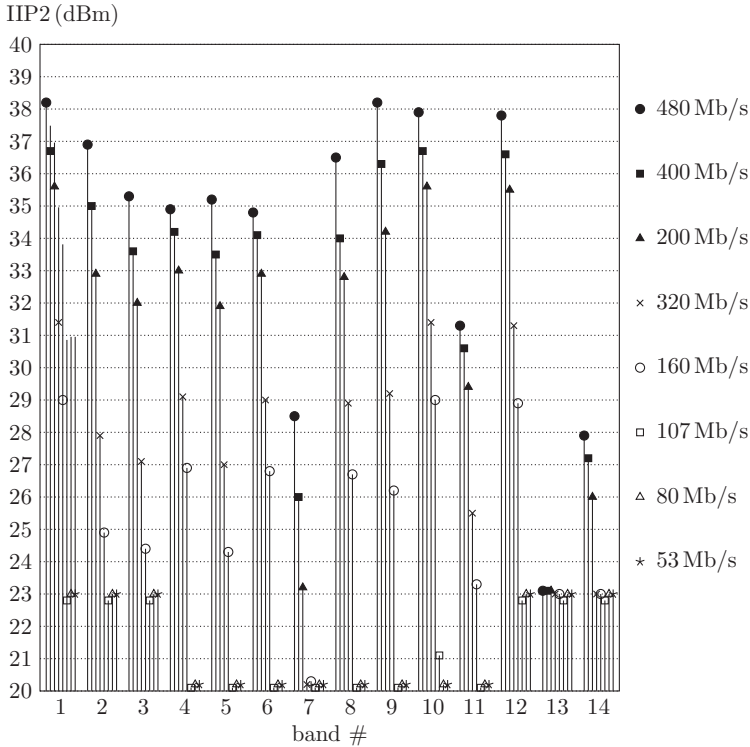


Fig. 3.3 IIP2 requirements for the different data rates

The LNA is the first element in the receive chain and therefore it usually defines the receiver noise figure. The primary issue in UWB LNA design is achieving low noise and impedance matching in a bandwidth spanning from 3.1 to 10.6 GHz. The classic solution adopted in narrow-band LNAs is the inductively-degenerated amplifier introduced by Van der Ziel [12]. To evaluate the limitations associated with a broad-band LNA we briefly revise the classic noise theory and derive the minimum achievable noise figure in an inductively-degenerated CMOS amplifier.

3.3.1 Noise and Power Matching in Wideband LNA Design

In general, noise minimization is achieved if an optimum input impedance is presented at the input of the amplifier. In order to find this optimum input impedance and the associated minimum NF, the amplifier is characterized with an equivalent circuit model as shown in Fig. 3.5, where all the internal noise sources of the amplifier are represented by two input-referred noise generators (current generator i_n

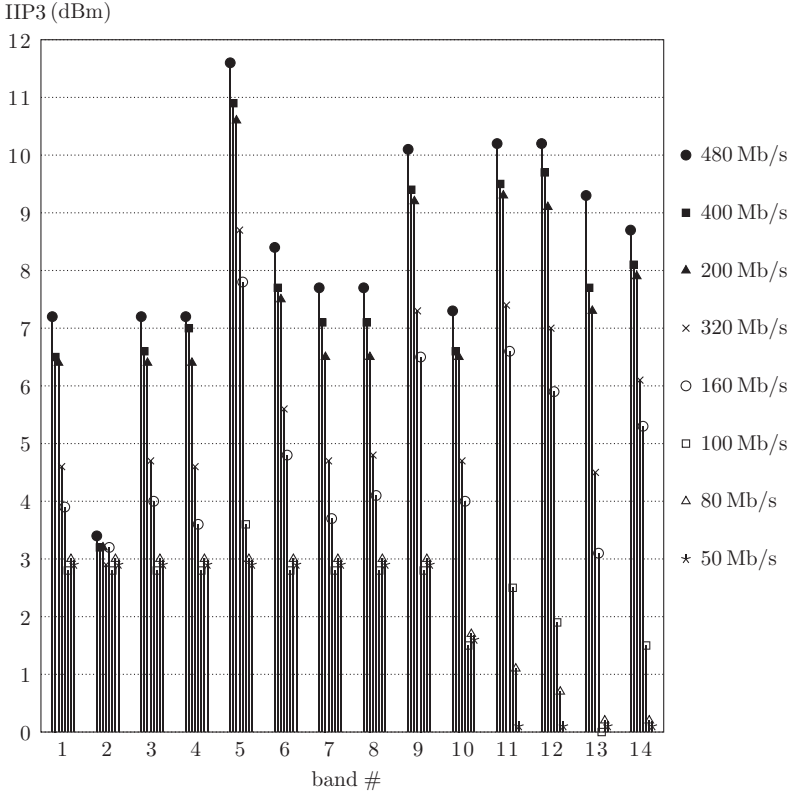
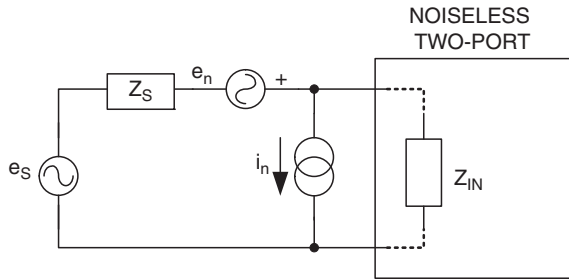


Fig. 3.4 Out-of-band IIP3 requirements for the different data rates

Fig. 3.5 Standard noise model for a two-port network with two input-referred noise generators and the thevenin equivalent of the source generator



and voltage generator e_n) and the amplifier is ideal (noiseless). The noise factor as a function of the input impedance Z_s is given by:

$$F = 1 + \frac{\overline{|e_n + Z_s i_n|^2}}{\overline{e_s^2}} \tag{3.1}$$

Notice that the two equivalent input noise generators in (Eq. 3.1) are in general correlated; in fact any noise source internal to the amplifier may in general give rise to (fully correlated) noise components both in the voltage and in the current noise generators. To take this into account more easily, one of the two generators (e.g. the voltage generator) is decomposed in two components, one (e_c) fully correlated with the noise current generator and the other one (e_u) fully uncorrelated. The correlated noise voltage e_c is proportional to the current noise generator i_n through a correlation impedance $Z_c = R_c + jX_c$, defined as follows:

$$e_c = Z_c i_n \quad Z_c \triangleq \frac{\overline{|i_n e_n^*|^2}}{i_n^2} \quad (3.2)$$

The three remaining independent noise generators can also be expressed in terms of equivalent noise resistance or conductance:

$$G_n = \frac{\overline{i_n^2}}{4kT\Delta f} \quad R_u = \frac{\overline{e_u^2}}{4kT\Delta f} \quad R_s = \frac{\overline{e_s^2}}{4kT\Delta f} \quad (3.3)$$

From this general formulation, classic noise theory results give the minimum noise factor (F_{\min}) and the optimal (“noise matching”) source impedance (Z_{opt}) [13, 14]; the expressions are shown in Table 3.3. The optimum source impedance and the associated minimum noise factor depend in general on the noise parameters of the amplifier and therefore they may vary considerably with frequency. Therefore it is intuitive that achieving minimum noise in a wideband amplifier is more difficult as compared to a narrow-band one. We will show this for the simplest possible case: an amplifier made of a single active element, such as a MOS transistor. In order to derive Z_{opt} and F_{\min} , the transistor is characterized with its small-signal equivalent model, including the noise generators (Fig. 3.6) as reported in [14].

Referring to the model in Fig. 3.6, the noise current generators (i_{nd} and i_{ng}) for the MOS transistor are given by [14]:

$$\overline{i_{nd}^2} = 4kT\gamma g_{d0}\Delta f \quad \overline{i_{ng}^2} = 4kT \frac{\delta}{5} \frac{\omega^2 C_{gs}^2}{g_{d0}} \Delta f \quad (3.4)$$

Table 3.3 Noise parameters for a generic amplifier and for a MOS transistor

Parameter	Generic amplifier	MOS transistor
F_{\min}	$1 + 2G_n \left[\sqrt{R_c^2 + \frac{R_u}{G_n}} + R_c \right]$	$1 + \frac{2}{\sqrt{5}} \frac{\omega}{\omega_T} \sqrt{\gamma \delta (1 - c ^2)}$
X_{opt}	$-X_c$	$\frac{-(1 + \alpha c \sqrt{\delta/5\gamma})}{\omega C_{gs} (1 + 2\alpha c \sqrt{\delta/5\gamma} + \alpha^2 \delta/5\gamma)}$
R_{opt}	$\sqrt{R_c^2 + \frac{R_u}{G_n}}$	$\frac{\alpha \sqrt{\delta/5\gamma} \sqrt{1 - c ^2}}{\omega C_{gs} (1 + 2\alpha c \sqrt{\delta/5\gamma} + \alpha^2 \delta/5\gamma)}$

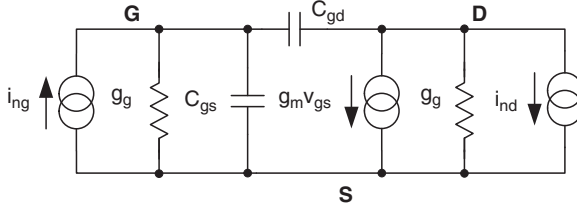


Fig. 3.6 MOS noise equivalent model proposed by Van der Ziel [14]

The two noise sources are generated by the same physical phenomenon and therefore present a finite correlation coefficient:

$$c = \frac{\overline{i_{ng} i_{nd}^*}}{\sqrt{\overline{i_{ng}^2}} \sqrt{\overline{i_{nd}^2}}} \cong j0.395 \quad (3.5)$$

The two current generators can be transformed into a pair of equivalent input noise generators following the procedure shown in [12]. Minimization of noise gives the optimum source impedance and the minimum noise factor [14, 15]. For the reader convenience, the results are reported in Table 3.3, where $\alpha \triangleq g_m/g_{d0}$.

In a narrow-band receiver, it is relatively easy to come close to this minimum noise while also approaching a power matching condition using the inductively-degenerated amplifier. An inductance is added to the source of the transistor, creating a noiseless input resistance $\omega_\tau L_s$ equal to the optimum noise resistance R_{opt} . In this way, the input power match and the input noise match conditions are made very close. Notice that the two conditions are not exactly coincident since the power match would require an input reactance that exactly cancels the input gate-source capacitance C_{gs} , while the noise match requires an input reactance that is smaller [15]. Nevertheless, low noise figures have been demonstrated up to several gigahertz using standard 90 and 65 nm CMOS technologies adopting the inductive-degeneration topology.

In a wideband inductively-degenerated amplifier, it is even harder to make the noise and power match conditions of the device to coincide over the entire bandwidth. In fact, R_{opt} is inversely proportional to frequency, while the input resistance $\omega_\tau L_s$ remains constant over the amplifier bandwidth. As a result, the inductively-degenerated amplifier achieves the MOS transistor minimum noise factor F_{min} only at a single frequency ω_{opt} , where the input impedance for power matching and the optimum noise impedance coincide. In general, even assuming that the reactive part of the input impedance equals X_{opt} over the entire bandwidth, the minimum achievable noise factor will be given by:

$$F_{L-DEG,min} = F_{min} + \sqrt{\frac{\delta\gamma}{5}} (1 - |c|^2) \frac{|\omega - \omega_{opt}|^2}{\omega_\tau \omega_{opt}} \quad (3.6)$$

The intrinsic noise penalty associated with wideband operation is directly proportional to the square of the fractional bandwidth and reduces α linearly with the device

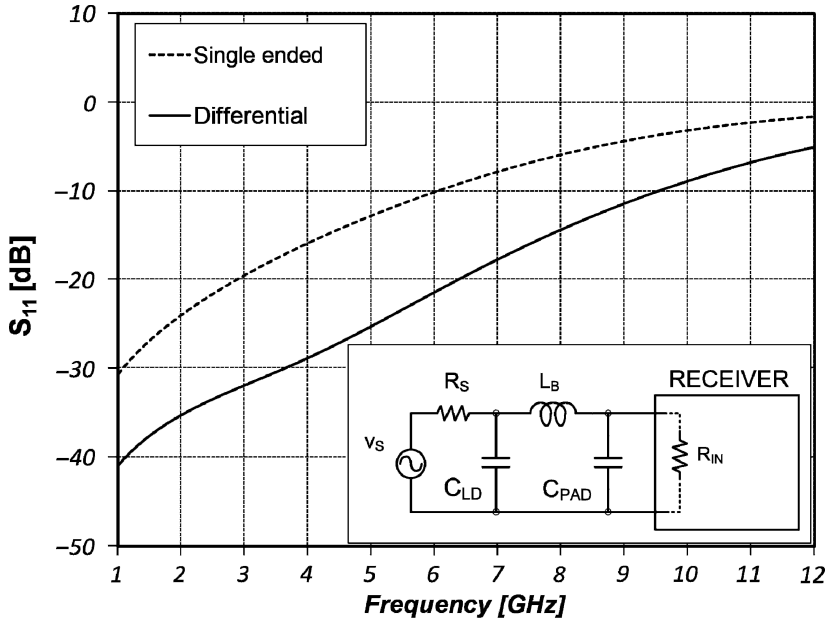


Fig. 3.7 Simulated S_{11} using the simplified package model in the dotted box

cut-off frequency. Notice that, as shown in Eq. (3.6), as the frequency increases, the noise factor (F_{\min}) of the device increases linearly.

In order to derive this minimum noise factor, an ideal power matching network was assumed at the receiver input. In practice an additional limitation arises when the circuit is placed into a low-cost package. In Fig. 3.7, a simplified model of the package is shown, where R_S represents the source impedance, R_{IN} the input resistance of the LNA, C_{PAD} the parasitic pad capacitance, and L_B and C_{LD} are the bondwire and package lead parasitic inductance and capacitance. The bandwidth limitations associated with the package parasitics are shown in Fig. 3.7, reporting the S_{11} across the UWB frequency band for an ideal receiver with 50Ω input impedance. If a differential LNA is chosen, mutual coupling between the two input bond-wires and metal leads can significantly mitigate the bandwidth limitation associated with these parasitic elements. The comparison is shown in Fig. 3.7, where a coupling coefficient of 0.35 is assumed. The input return loss remains below 10 dB up to 6 GHz for the single-ended case and up to 9.5 GHz for the differential case.

3.3.2 Receiver Linearity and LNA Selectivity

Robustness to interferers is one of the most critical requirement in UWB receivers. In a generic receiver, linearity is usually limited by the down-conversion mixers and the following blocks in the receive chain. This holds for UWB receivers as well, but

due to the limited selectivity of the external RF filters, a large number of interferers hits the LNA making its linearity requirements more stringent compared to cellular or wireless LAN receivers. Moreover, due to the wideband input, several second-order intermodulation products may fall within the UWB band. As a result, the LNA has to meet stringent out-of-band IIP2 specifications, that are much more stringent compared with in-band ones. This suggests that the use of a selective RF front-end can significantly improve receiver interference immunity. Strong interferers in the cellular and lower WLAN bands falling outside the UWB spectrum can be filtered out using an external RF filter [11], relaxing some of the linearity requirements of the receiver IC.

3.4 Evolution and State-of-the-Art in UWB Receivers

Several CMOS and BiCMOS UWB receivers have been published recently [4, 16–21]. Covering the whole UWB band from 3.1 to 10.6 GHz has proved quite challenging, especially in CMOS, where narrowband approaches are generally the key to achieve low-power operation. Although it is not easy to precisely categorize existing UWB transceivers, focusing only on the receiver front-end, we have identified three approaches: narrowband multi-path, broadband direct-conversion and broadband heterodyne (Fig. 3.8). In the multi-path approach a separate signal path is used for each UWB band. In this way, each parallel path can be optimized independently using narrowband approaches. This potentially gives the best performances in each band but at the price of a large silicon area. A CMOS multi-path narrowband direct conversion receiver covering the three lower bands from 3 to 5 GHz is reported in [16]. The common-gate stage at the low-noise amplifier input provides broadband impedance matching, while a multiple-gate cascode device is used to rapidly switch between the three resonant loads, each using a different spiral inductor. Three direct down-conversion mixers, driven by three independent phase-locked loops (PPLs), down-convert the signal to baseband. The main drawback of this approach is the limited UWB bands coverage capability. Extending this architecture to cover all 14 bands of UWB would require a very large area. Besides, even when separate RF paths are used for each UWB band, programmability is not implemented for low-noise broadband impedance matching.

A number of reported receiver solutions [4, 17–20] are based on a broadband direct conversion architecture. The receiver RF front-end consists of a broadband LNA followed by a direct down-conversion quadrature mixer, as shown in Fig. 3.8b. In [4, 17–19] the LNA is based on an inductively-degenerated common-emitter input stage. Inductive degeneration results in a broadband noiseless real input impedance component $\omega_T L_S$. This impedance can be matched to the source impedance, typically 50 Ω , using a band-pass doubly-terminated ladder filter. The advantage of this approach is that the reactive part of the input impedance of the amplifier can be embedded in the filter structure. The main disadvantage is that, since the signal current at the input of the active device is constant over the input bandwidth, the effective

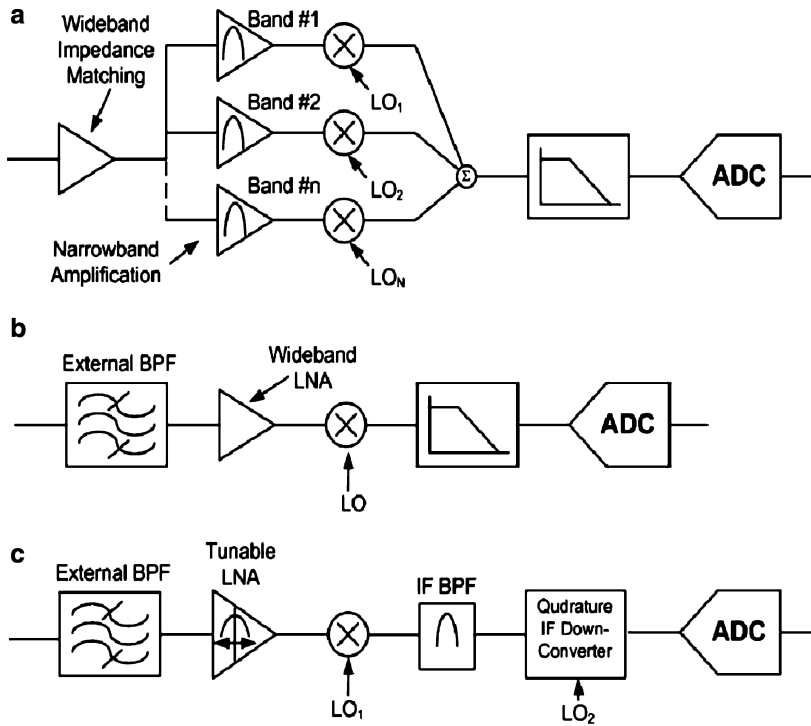


Fig. 3.8 UWB receiver architectures: (a) narrow-band multipath, (b) wideband direct conversion, (c) wideband super-heterodyne

voltage is inversely proportional to frequency. The transconductance gain roll-off at high frequencies can be compensated using an inductive-peaking load. On the other hand, the noise factor degrades with the square of the frequency and the robustness to interferers, usually stronger in the lower part of the spectrum, is reduced. With this solution, using bipolar transistors, low noise figure and good impedance match have been demonstrated up to 8 GHz.

An alternative solution for the implementation of the wideband LNA is the noise cancellation technique. In [20] a noise-cancelling CMOS LNA is presented covering the whole UWB band, from 3.1 to 10.6 GHz. Flat $50\ \Omega$ input impedance is achieved over the band using a common-gate transistor. The noise associated with this device is cancelled exploiting a parallel common-source stage and combining the two outputs with properly scaled gains. In principle, the circuit bandwidth can be extended while keeping noise cancellation using shunt peaking inductors [20]. The advantage of this solution compared with the inductively-degenerated common-source amplifier is that it avoids the high frequency roll-off in the input transconductance and can therefore achieve, at least in principle, lower noise over a wide bandwidth. In practice, achieving exact noise cancellation of the common-gate input device requires good impedance matching and is highly dependent on the parasitic elements

(parallel C and series L) at the receiver input. In [20], test measurements show good matching and noise performances from 3.1 to 10.6-GHz but using wafer probes.

Linearity performances in broadband direct conversion architectures is limited by the mixer since the LNA has a high and wideband gain profile. Moreover, when a single-ended input LNA is used, IIP2 in the LNA itself may severely limit the robustness to interferers.

The use of Q-enhanced on-chip notch filters has been proposed to improve immunity to in-band interferers (e.g. 5-GHz WLAN) [18, 19]. In [18] a minimum suppression of 10-dB in the band between 5.15–5.35 GHz is demonstrated with a moderate increase in noise. However, the low dynamic range of the Q-enhancement active circuits limits the achievable attenuation when the interferers power increases [19].

Most receivers employing either multi-path narrowband or broadband architectures use direct conversion, i.e. the most compact and low-cost solution. On the other hand, an external RF pre-filter is usually required [4, 17, 18] to attenuate signals below 3 GHz. The idea behind the wideband heterodyne architecture [21] is to exploit the RF pre-filter for image-rejection purposes, avoiding issues with time-varying DC offsets, typical of direct conversion receivers. In [21], the signal is down-converted to a fixed IF of 2.64 GHz, where an on-chip LC third-order band-pass filter provides strong filtering, relaxing linearity requirements for the following down-converter chain. With this choice of IF, the RF filter provides about 10 dB image rejection for all channels below 7.6 GHz. Additional image rejection is provided on-chip by dynamically tuning the LNA load. The main limitation of this architecture is the limited image rejection offered by the RF filter when receiving the higher sub-bands. In fact, for UWB sub-bands above 8.6 GHz, the image frequency falls within the UWB spectrum and the RF filter ceases to provide image rejection. As a consequence, a significant portion of the spectrum is subject to interference from other UWB transmission. Furthermore, coping with the package parasitic still remains an issue.

3.5 UWB Architecture

From the above discussion, we need an architecture capable of coping with the two main challenges in UWB systems: the input parasitics limiting the impedance matching, and the large interferers limiting the linearity. The proposed architecture stems from the fact that of the 7 GHz bandwidth available only 528 MHz are utilized at any given time. It therefore becomes feasible to use a 528 MHz front-end provided that its center frequency can be programmed. In this way the advantages of a selective topology can be exploited even in a very broad band system like UWB. This idea extends the principle of multi-path architectures (also used in the wideband heterodyne receiver) to its limit, i.e. 14 different optimized channels. In order to suit UWB, such a reconfigurable/selective architecture should have the following characteristics. First it should provide re-configurability over a total of 14 bands, each 528 MHz wide. Second it should be able to reconfigure itself in a short time

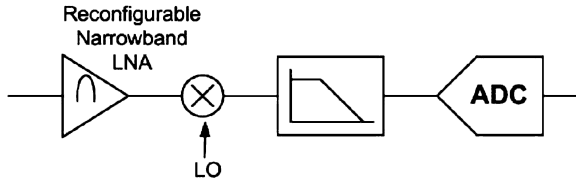


Fig. 3.9 Re-configurable narrow-band direct conversion receiver architecture

when hopping from one band to another. Third, it should provide enough filtering in front of the mixer to make it compatible with the interference scenario. For the first point the characteristics (selectivity, gain, input matching, noise, etc.) of the 14 bands must be equalized as they are swept across the 7 GHz frequency span. Furthermore programming should be done in the most area efficient way. For the second point, the 528 MHz bandwidth turns out to be compatible with the 9.5-ns maximum reconfiguration time. For the third point, the LNA interference filtering has to be realized without degrading receiver linearity. In addition, even when using a filtering LNA, the mixer linearity must still be optimized (Fig. 3.9).

3.5.1 Selective and Reconfigurable LNA

From a general point of view, the voltage gain of the LNA is the product of the input device transconductance times the load impedance. To achieve maximum results both the input transconductance and the load impedance should be made selective and programmable. This can be achieved for the load without significant noise penalty by using a tank whose L and C are programmed by switches. On the other hand doing the same for the transconductance of the LNA is a lot more challenging. One possibility is to use inductive degeneration in a common emitter topology. In this case the selectivity is controlled by the Q of the input network, which, in general, includes the passives of the package. Unfortunately reconfiguring the frequency response of such a circuit using MOS switches is unpractical since it compromises noise. On the other hand, the input impedance can be reconfigured without excess noise by using feedback. By using shunt positive feedback, as shown conceptually in Fig. 3.10, the input impedance can be chosen to attain impedance matching in the band of interest while, at the same time, out-of-band impedance mismatch can be exploited to achieve selectivity in the transconductance. In addition it is possible to tune out the input parasitic over the programmed band. The idea is to use a common base device (M1) with a transconductance larger than $1/R_s$ where R_s is the source impedance and to achieve impedance matching by injecting in the source of M1 an additional signal I_{FB} in phase with the one coming from the source. The in-band transconductance is $1/2g_m$, while out of band, due to the reduced loop gain, the transconductance approaches $g_m/(g_m R_s + 1)$. Using positive feedback results in an enhanced selectivity of the LNA response with respect to that of the load as shown

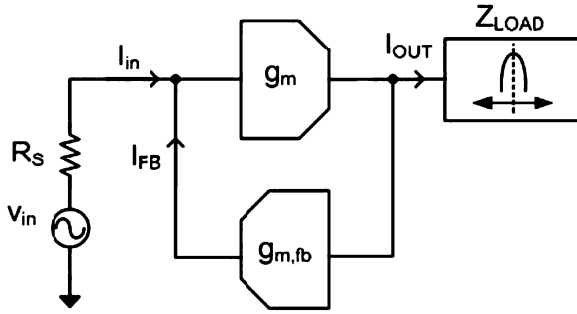


Fig. 3.10 Block diagram of the re-configurable narrow-band LNA using shunt positive feedback

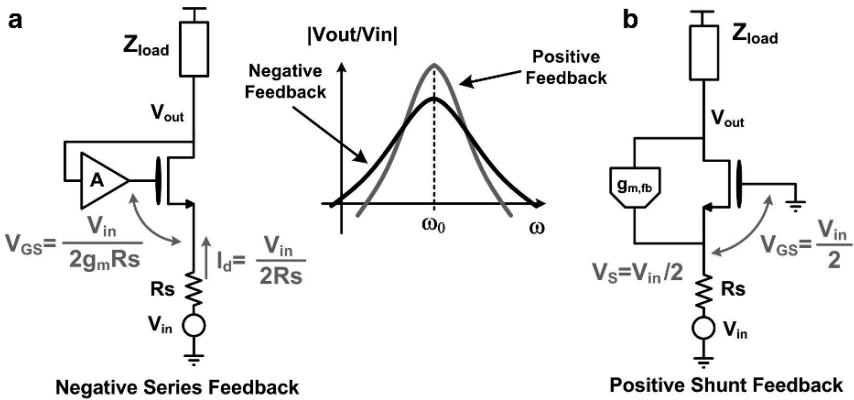


Fig. 3.11 Effect of series and shunt feedback networks on LNA linearity

in Fig. 3.11. The enhanced selectivity is achieved by reducing the voltage seen by the input device for out-of-band signals and, as a consequence, the out-of-band linearity of the LNA is better than the in-band one, consistently with the interference scenario. Not all types of feedback that give impedance matching in a narrow band give also a corresponding enhanced selectivity in the transconductance. As an example, a series negative feedback will result in a transconductance that is increasing outside the pass band, actually reducing the LNA selectivity with respect to a broadband termination.

3.5.1.1 Linearity Considerations for Shunt (Positive) and Series (Negative) Feedback LNAs

The feedback network affects the LNA linearity by acting on the gate-source voltage swing of the input stage and modifying the filtering profile. Figure 3.11 shows the effect of the two possible feedback networks (series and shunt) on the small signal gate-source voltage V_{GS} and on the LNA transfer function.

For in band signals, a series negative feedback reduces the amount of signal between gate and source of the input device, as compared with a simple common gate. This occurs in order to satisfy input matching when $g_{m1} \gg 1/R_s$. This results in a direct improvement of the in-band linearity compared to a simple common gate amplifier. On the contrary, the proposed shunt feedback forces the V_{GS} to be $V_{in}/2$ for any value of $g_{m,fb}$ guaranteeing roughly the same linearity of a common gate stage [22] for a given overdrive.

For out-of-band signals, both series negative feedback and shunt positive feedback tend to behave similarly: the input impedance reduces to $1/g_m$ (which is lower than R_s) and the modulation of V_{GS} is lower than $V_{in}/2$, resulting in an improved LNA linearity compared to a broadband common-gate stage. Another important effect of feedback on receiver linearity is associated with the LNA selectivity as shown in Fig. 3.11. Using a positive shunt feedback, the current injection provided by $g_{m,fb}$ boosts the quality factor of the resonant load, resulting in a sharper frequency profile. On the contrary, the negative feedback reduces the LNA selectivity compared to a common gate amplifier since the transconductance of the stage increases when the load impedance decreases (i.e. out-of-band). Enhanced selectivity is highly desirable because it allows to relax the dynamic range required by the mixers to handle out-of-band interferers.

3.5.2 Highly Linear Mixer

A block diagram of the RX architecture used in this paper is shown in Fig. 3.12. The reconfigurable narrow band amplifier is followed by two transconductors that split the signal in the I and Q paths. They drive two current-mode passive mixers terminated on a low impedance which is provided by a base-band second-order channel-select filter. Despite the LNA's selectivity, the mixer still remains the limiting element for the overall receiver linearity. We have chosen a passive mixer topology for the reasons outlined below.

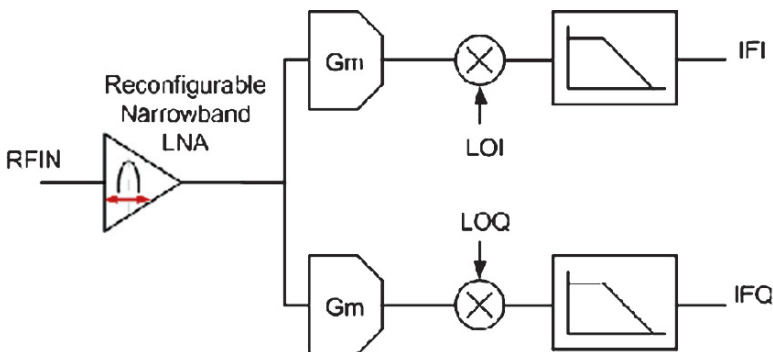


Fig. 3.12 Complete receiver front-end block diagram

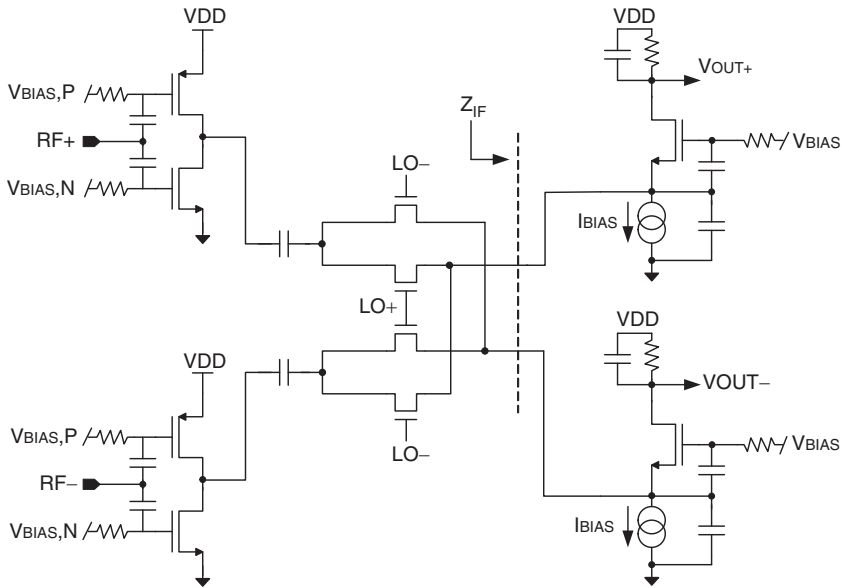


Fig. 3.13 Passive current-mode mixer schematic

As shown in Fig. 3.13, a fully differential passive mixer consists of four minimum-channel switching devices. The mixer performance (in terms of noise, linearity and power consumption) is determined by the characteristics of its IF port and, to some extent, its driving stage. In general, the mixer tends to behave in a linear fashion when the impedance of its driver is very different from that of its load. This is because distortion components can't be generated if neither current nor voltage division take place. We have found that the best performance in terms of linearity and $1/f$ noise is provided by a current mode mixer (See [23, 24]).

The main challenge to implement a truly current mode structure is to design the IF port in such a way that it maintains a sufficiently low impedance level over a very large bandwidth, i.e. up to the frequency where the furthest blockers are located. If a closed loop circuit is used, e.g. a trans-impedance amplifier, a very low in band impedance is easily achieved but power consumption and stability constraints severely limit the maximum frequency where current mode operation can be guaranteed. Furthermore, stability and area constraints limit the value of the shunt capacitance that can be placed at the IF port which is the easiest way to keep the impedance of such a node low at high frequency. In this design a common gate circuit is used at the IF port. In this way, there are no stability issue when a shunt capacitor is placed at the source of the common gate device. Actually, such a circuit configuration has a low pass transfer function which could be used, at least in principle, as the first stage of the channel select base-band filter. In general for a given capacitor value, i.e. a certain silicon area, there is a trade-off between the level of the DC impedance at the mixer output node (i.e. the transconductance of the common source device) and the bandwidth of the baseband filter. In general, for applications

like cellular phones and WLAN, to guarantee low impedance at the mixer output, a very large capacitance value must be used to give the required bandwidth (approximately half the channel bandwidth) in the baseband filter. However, for the UWB case, such a bandwidth is about 250 MHz, which requires a relatively small shunt capacitor (in the order of a few picofarads).

The other degree of freedom that can be used in a passive mixer to optimize its performance is the DC bias point of the four switching nMOS devices, relative to the crossover point between the raising and the falling edges of the LO signal. By changing such a DC bias point it is possible to move from a situation where both the switches that are turning off and the ones that are turning on are simultaneously off for a significant portion of the rise (fall) time (off-overlap), toward a situation where all switches are on for the same significant time interval (on-overlap). These two extreme situations correspond to a better noise figure of the mixer but a less linear behavior (the first one) or to a worse noise figure but a more linear situation (the second one).

The last parameter that can be changed by the designer is the size of the switching devices. Increasing the device size tends to improve linearity up to the point where their on-resistance becomes significantly smaller than the impedance level of the mixer IF port. On the other hand, larger size switches give a larger parasitic capacitance lowering the equivalent driving impedance (due to switched capacitor effect), which in turn increases the noise contribution of the IF port.

3.6 Measurement Results

A prototype implementing the UWB front-end described in this chapter was realized in a 90 nm CMOS process from TSMC. The chip microphotograph is shown in Fig. 3.14. Its active area is $1 \times 0.9 \text{ mm}^2$.

3.6.1 Broadband Measurements and De-embedding Techniques

The measurements setup used for the characterization of the UWB receiver front-end is shown in Fig. 3.15. The chip was mounted on the printed circuit board using a chip-on board technique and the signal pads were connected to the board traces using bondwires. The differential RF signal at the LNA input is connected to a filtering structure, implemented using the board traces, that helps matching the chip input impedance with the 50Ω impedance of the coplanar transmission lines over the entire UWB band. This approach avoids the use of surface-mount components with their associated parasitics that would make it difficult to achieve broadband impedance matching over the whole band. The coplanar impedance lines are connected through SMA connectors to hybrid couplers that are used to convert the single-ended signal into a differential one. The main issue with such a measurement

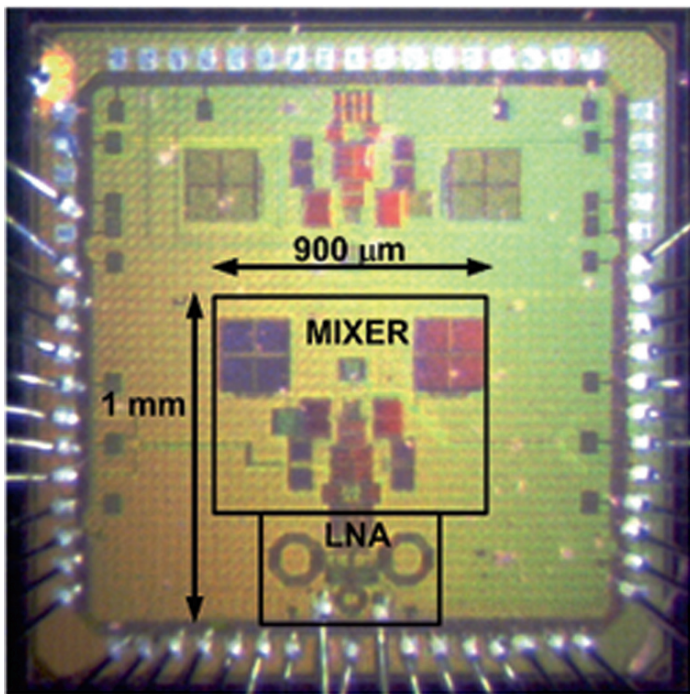


Fig. 3.14 UWB receiver front-end chip photograph

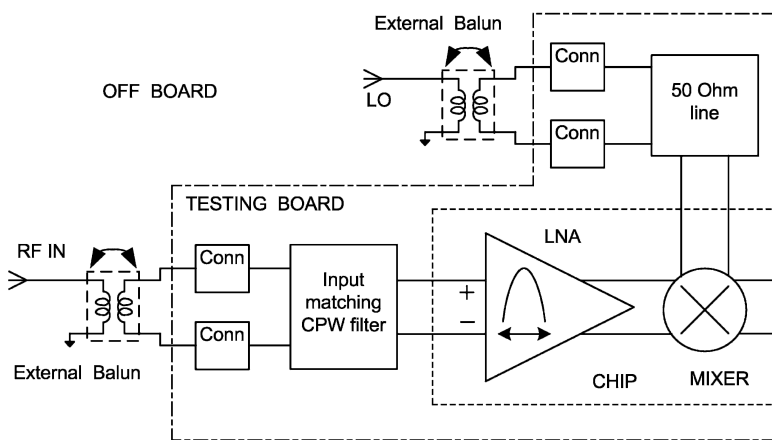


Fig. 3.15 Measurements setup

setup was in the imperfect de-embedding of the balun and the interconnects between the balun and the coplanar transmission lines, which masked the performances of the device under test (DUT). Using a series of two-port S-parameters measurements, a four-port S-parameters matrix for the hybrid coupler was derived. From

the four-port matrix, a two-port matrix that accounts for the hybrid used as a single-ended to differential converter was derived. The SMA-coplanar transmission line transition was modeled using the parameters extracted from two different calibrations of the vector-network analyzer: one short-open-load calibration carried out using standard SMA terminations and one thru-reflect-line calibration carried out using transmission lines of the same type used for the DUT test board.

3.6.2 Measurement Results

From a qualitative point of view the chip behaves as expected and all his programmability is fully functional, including in particular the ability to simultaneously change the center frequency, the gain and the bandwidth (selectivity) of each band-pass response. However, significant quantitative deviations from the expected values are observed: i.e. the centre frequency is shifted down while the in-band gain and quality factor are reduced (Fig. 3.16). This effect is quantitatively shown in Fig. 3.16, that compares the measured RF to base-band transfer function for two programmable bands with the corresponding curves obtained by simulation.

A detailed analysis of the chip layout has pointed out two major problems that explain the observed discrepancy between measurements and schematic simulation results. First, very long and narrow lines were used to connect the LNA output (that includes a fairly complicate tuning capacitor/inductor array) to the mixer input. The effect of this resulted in a significant variation of the LNA resonant load with

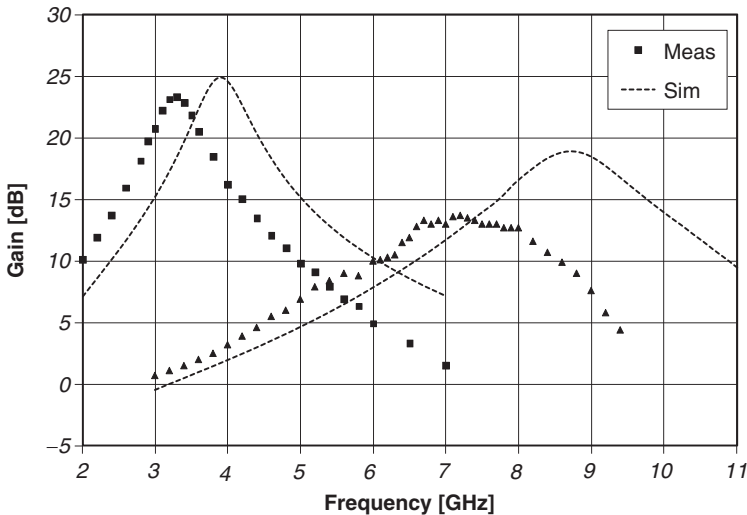


Fig. 3.16 Comparison between measured and circuit simulation results: down-conversion gain at 10 MHz IF frequency

respect to the schematic simulations. This effect can be modeled with a symmetrical π equivalent circuit with a total capacitance of 220 fF and a series resistance of 6 Ω . Second, the loading of the mixer, including the bottom-plate parasitics of the bypass capacitor, was underestimated by about 160 fF; furthermore the Q of this additional capacitance was fairly low due to the mixer input device overlap capacitance (enhanced by the poor layout choices). Figures 3.17–3.18 and Tables 3.4–3.7 show the comparison between measurement and post-layout manually extracted simulations. In the characterization of the chip, we exploited the available programmability to center the frequency response of the first four bands to the correct values, as required by the UWB standard. An additional band, centered at 2.8 GHz, was also

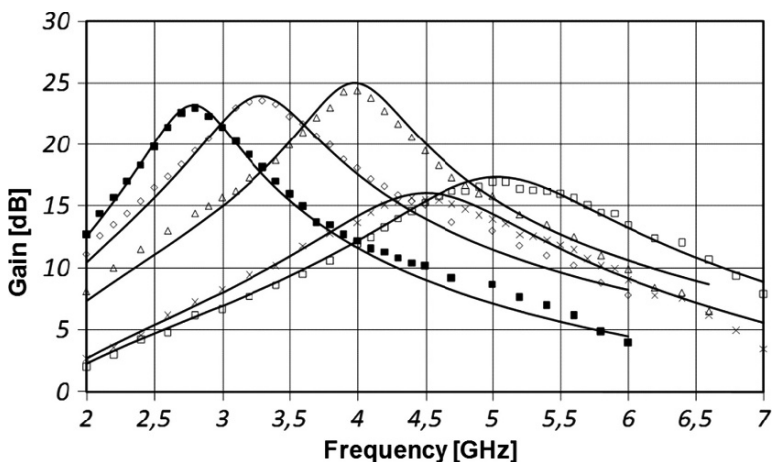


Fig. 3.17 Comparison between measurements (symbols) and post-layout circuit schematic simulations (continuous lines): down-conversion gain at 10 MHz IF frequency

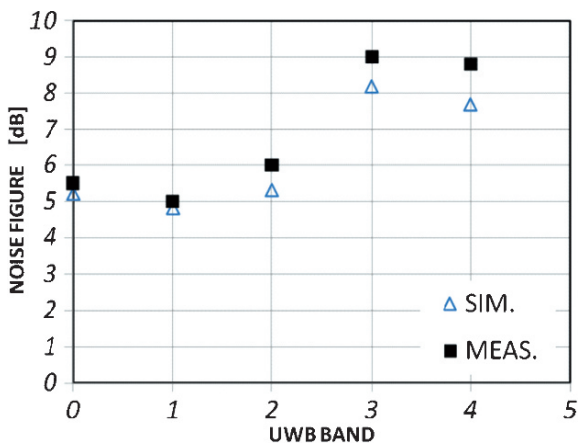


Fig. 3.18 Measured and simulated NF of the receiver front-end over the four lower UWB bands (Band #0 corresponds to a center frequency of 2.8 GHz)

Table 3.4 IIP2 measurement and circuit simulation results comparison

BAND #	Fc [MHz]	Freq. Int1 [MHz]	Freq. Int2 [MHz]	IIP2 meas. [dBm]
1	3,432	1,747	1,747	46.3
2	3,960	1,950	1,950	52.8
3	4,488	1,950	2,448	44.8
4	5,016	2,448	2,448	46.1

Table 3.5 IIP3 measurement and circuit simulation results comparison

BAND #	Fc [MHz]	Freq. Int1 [MHz]	Freq. Int2 [MHz]	IIP3 sim. [dBm]	IIP3 meas. [dBm]
1	3,432	454	1,950	18.3	19.3
2	3,960	454	1,748	19.1	20
3	4,488	2,448	3,500	3	4
4	5,016	1,950	3,500	3.2	4.3

Table 3.6 Out-of-band IIP2 measurements on the first UWB band (Fc = 3.3 GHz)

Freq. Int1 [MHz]	Freq. Int2 [MHz]	IIP2 spec. (480 Mbps) [dBm]	IIP2 meas. [dBm]
454	2,978	15.4	39.7
784	2,647	18.3	40
784	4,216	10.7	39.3
897	2,534	17.1	42
897	4,329	9.5	47.3
1,747	1,684	29.4	46.3
1,747	5,179	22.9	41.2
2,448	984	24.3	49.7
2,448	5,880	37.1	40
3,960	7,392	11	44.2
5,016	8,448	11	41.6
6,072	2,640	13.6	36.4
6,072	9,504	11	54.2
7,128	3,696	20.7	40.3
7,656	4,224	11	41.7
8,712	5,280	17.1	41.9

characterized. As can be seen, a good agreement between the measured and simulated frequency responses is obtained in all cases. Measured noise figure and IIP3 also show a good agreement with the simulations.

In order to verify the effectiveness of the selective LNA topology in terms of robustness to interferers, an extensive set of two-tones linearity measurements were carried out on the first UWB band which is the least affected by the layout. The experimental results show that both IIP2 and IIP3 requirements for 480 Mbps are met with a good margin and that the LNA selectivity improves receiver linearity as the frequency spacing between the interferes and desired signal increases.

Table 3.7 Out-of-band IIP3 measurements on the first UWB band ($F_c = 3.3$ GHz)

Freq. Int1 [MHz]	Freq. Int2 [MHz]	IIP3 spec. (480 Mbps) [dBm]	IIP3 sim. [dBm]	IIP3 meas. [dBm]
454	1,943	4.5	17.2	17.7
482	2,467	4.8	21.6	22.9
785	2,108	2.1	13.2	13
898	1,267	-1	15.5	18.8
898	2,165	1.5	12.2	12.5
1,748	2,590	-0.5	4	5
2,448	2,940	0.5	-4	-3.2
3,960	3,696	-6.9	-4.7	-4.7
4,488	528	-5.6	23.7	21.9
4,488	5,544	-9.5	5.5	5
5,016	792	-9.8	18	17.9
5,016	6,600	-9.5	7.5	8
5,300	7,168	-6.5	7.8	8.7
5,598	1,083	-2.2	14	16.1
6,072	1,320	-11	12.8	15.1
6,072	4,752	-9.5	7	8.2
6,072	8,712	-9.5	8.2	12.7
7,656	2,112	0.3	11.2	12.3
7,656	5,544	-9.5	8.2	8
9,240	2,904	-6.9	6	14
9,240	6,336	-9.5	9	13.8

3.6.3 Improved Design Simulations Results

Although the experimental measurements show acceptable performance for the lower bands, better results can be obtained if the circuit layout is improved and parasitics are taken care of. A modified version of the chip has been re-designed/laid-out acting primarily at the LNA-mixer coupling interface. First, the layout has been re-organized to reduce the parasitic resistance of the metal traces between the LNA and the mixer; second, the input stage of the mixer has been scaled down to reduce the loading effect of the mixer input devices; third, the bypass coupling capacitor between LNA and mixer has been re-scaled in order to reduce the parasitic bottom-plate capacitance. In Figs. 3.19–3.21, the simulated performances for the receiver front-end is shown, assuming all of the above changes have been implemented. In Fig. 3.19, the simulated transfer function and input return loss for the receiver front-end for a selected number of programmable band settings is shown. As it can be seen, correct operation in terms of gain and S_{11} is recovered up to the highest band of the standard. Receiver noise figure simulations show that a NF between 4 and 6 dB is achieved across the 14 bands. Linearity exceeds the IIP3 requirements for 480 Mbps operation for all bands, with the exception of band #9, where requirements are met for data rates up to 320 Mbps.

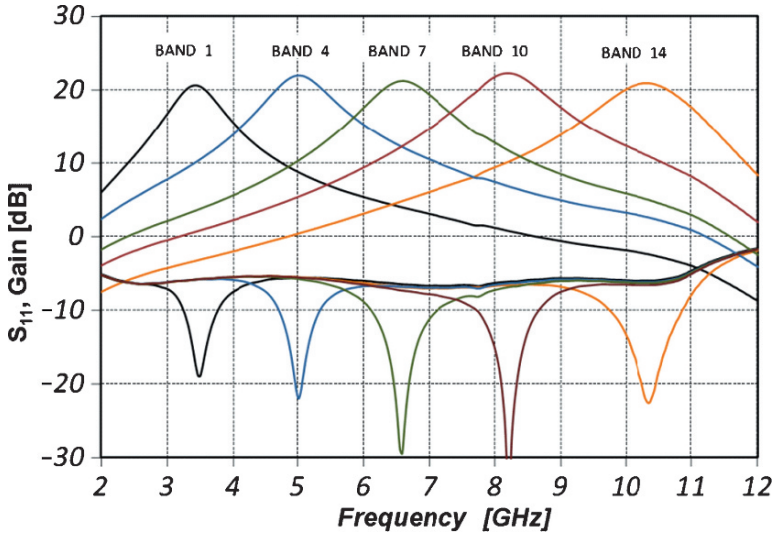


Fig. 3.19 Gain and S_{11} simulation of the improved circuit schematic

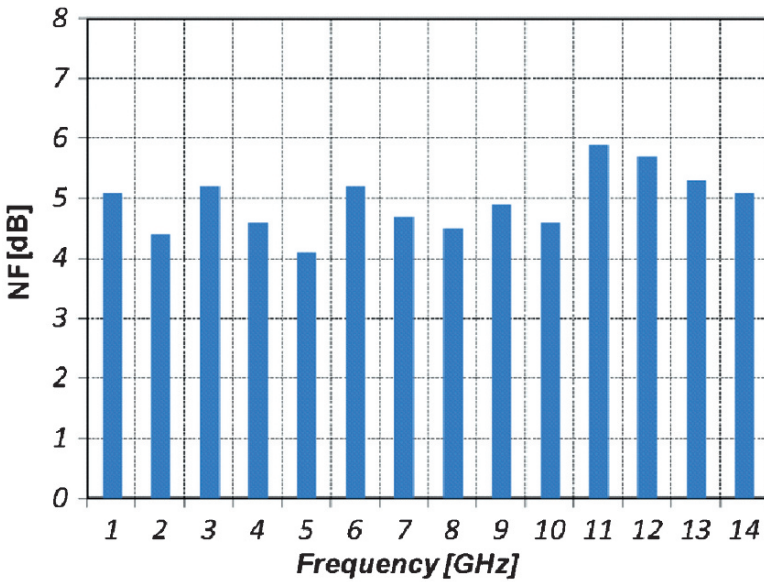


Fig. 3.20 Simulated NF for the improved design

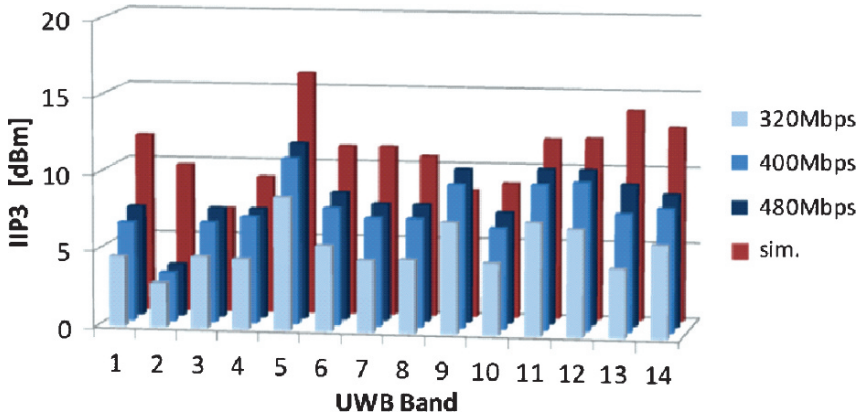


Fig. 3.21 Comparison between the simulated IIP3 of the improved design and the specifications for the highest data rates: 320, 400 and 480 Mbps

3.6.4 Conclusion

A reconfigurable narrowband receiver architecture tailored to MB-OFDM UWB receivers has been presented in this chapter. Linearity requirements for this system are predominantly set by narrowband out-of-band interferers (WLAN, cellular, etc). The proposed architecture is more resilient to these interferers compared to broadband architectures commonly used in UWB systems. A reconfigurable narrowband receiver front-end based on a positive shunt-feedback LNA allows to simultaneously achieve reconfigurable impedance matching in the band of interest and improved out-of-band linearity of the receiver.

Acknowledgment The authors wish to thank F. De Paola for help with chip characterization. This work has been partially supported by the Italian Basic Research Program FIRB under contract nr. RBAP06L4S5.

References

1. First Report and Order, Revision of Part 15 of the Commission's Rules Regarding Ultra-Wideband Transmission Systems FCC, 2002, ET Docket 98–153.
2. High Rate Ultra Wideband PHY and MAC Standard, ECMA standard 368, Dec. 2005.
3. M. Ranjan, and L. E. Larson, "Distortion Analysis of Ultra-Wideband OFDM Receiver Front-Ends," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 12, Dec. 2006, pp. 4422–4431.
4. R. Roovers, et al., "An Interference-Robust Receiver for Ultra-Wideband Radio in SiGe BiCMOS Technology", *IEEE Journal of Solid-State Circuits*, vol. 40, no. 12, Dec. 2005, pp. 2563–2572.

5. N. Laurenti, F. Renna, "Estimation of Carrier and Sampling Frequency Offset for Ultra Wide Band Multiband OFDM Systems", in Proceedings of IEEE International Conference on Ultra-Wideband, ICUWB, Sept. 2008, vol. 2.
6. Digital Cellular Telecommunications System (Phase 2); Radio Transmission and Reception, GSM 05.05 (ETS 300 577), ETSI, 1996.
7. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications—High-Speed Physical Layer in the 5 GHz Band, ANSI/IEEE Standard 802.11a, 1999.
8. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications—Further Higher-Speed Physical Layer Extension in the 2.4 GHz Band, IEEE Standard 802.11g/D1.1, 2002.
9. Broadband Radio Access Network (BRAN); HiperLAN type 2; physical (PHY) layer, ETSI, Sophia Antipolis Cedex, France, TS 101 475, ver. 1.3.1, 2001.
10. Air Interface for Fixed Broadband Wireless Access Systems, IEEE Standard 802.16, 2004.
11. L. Zhu, S. Sun, W. Menzel, "A Ultra-Wideband (UWB) Bandpass Filters Using Multiple-Mode Resonator", IEEE Microwave and Wireless Components Letters, vol. 15, no. 11, Nov. 2005, pp. 796–798.
12. M. J. O. Strutt, A. Van Der Ziel, "Suppression of spontaneous fluctuations in amplifiers and receivers for electrical communication and for measuring devices", Physica, vol. 9, no. 6, Jun. 1942, pp. 513–527.
13. H. Rothe, W. Dahlke, "Theory of Noisy Fourpoles," Proceedings of the IRE, vol. 44, no. 6, Jun. 1956, pp. 811–818.
14. A. Van Der Ziel, "Noise in Solid State Devices and Circuits", Wiley, New York, 1986.
15. T. H. Lee, "The Design of CMOS Radio-Frequency Integrated Circuits", 2nd ed., Cambridge University Press, Cambridge, UK, 2004.
16. B. Razavi, et al., "A UWB CMOS transceiver," IEEE Journal of Solid-State Circuits, vol. 40, pp. 2555–2562, Dec. 2005.
17. A. Ismail and A. Abidi, "A 3.1- to 8.2-GHz zero-IF Receiver and Direct Frequency Synthesizer in 0.18- μm SiGe BiCMOS for Mode-2 MB-OFDM UWB communication," IEEE Journal of Solid-State Circuits, vol. 40, pp. 2573–2582, Dec. 2005.
18. A. Valdes-Garcia, C. Mishra, F. Bahmani, J. Silva-Martinez, and E. Sánchez-Sinencio, "An 11-Band 3–10 GHz Receiver in SiGe BiCMOS for Multiband OFDM UWB Communication," IEEE Journal of Solid-State Circuits, vol. 42, no. 4, Apr. 2007, pp. 935–948.
19. A. Bevilacqua, A. Vallese, C. Sandner, M. Tiebout, A. Gerosa, A. Neviani, "A 0.13mm CMOS LNA with Integrated Balun and Notch Filter for 3–5GHz UWB Receivers", ISSCC Digest of Technical Papers, Feb. 2007, pp. 420–421.
20. C.-F. Liao, and S.-I. Liu, "A Broadband Noise-Canceling CMOS LNA for 3.1–10.6 GHz UWB Receivers", IEEE Journal of Solid-State Circuits, vol. 42, no. 2, Feb. 2007, pp. 329–339.
21. M. Ranjan, and L. E. Larson, "A Low-Cost and Low-Power CMOS Receiver Front-End for MB-OFDM Ultra-Wideband Systems," IEEE Journal of Solid-State Circuits, vol. 42, no. 3, Mar. 2007, pp. 592–601.
22. A. Liscidini, G. Martini, D. Mastantuono, R. Castello, "Analysis and Design of Configurable LNAs in Feedback Common-Gate Topologies," IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 55, no. 8, pp. 733–737, Aug. 2008.
23. E. Sacchi, I. Bietti, S. Erba, L. Tee, P. Vilmercati, and R. Castello, "A 15 mW, 70 kHz 1/f corner direct conversion CMOS receiver," in Proceedings of IEEE Custom IC Conference, Sep. 2003, pp. 459–462.
24. A. A. Abidi, "The Path to the Software-Defined Radio Receiver", IEEE Journal of Solid-State Circuits, vol. 42, no. 5, May 2007, pp. 954–966.

Chapter 4

Low Power UWB Circuits: Front-End Building Blocks

Tommy Tsang, Kuan-Yu Lin, Karim Allidina, and Mourad El-Gamal

4.1 Introduction

Ultra wideband (UWB) communications is a rapidly developing field that possesses desirable characteristics when compared to traditional narrowband systems. These characteristics stem from Shannon's channel capacity formula, which states that the capacity of a channel increases linearly with bandwidth and only logarithmically with an increased signal to noise ratio (SNR). As such, UWB communications offer both high data-rate communications over short distances and low data-rate communications over long distances. Additionally, UWB signals are localized in time, which enables precise location and ranging capabilities.

These characteristics make UWB systems desirable for a number of applications, including:

- Wireless high speed peripherals (e.g., USB drives, video interconnects, speakers)
- Low speed sensors (e.g., wireless sensor networks, RFID tags)
- Ranging (e.g., vehicle detections, object tracking devices)

This chapter will first provide a brief overview of UWB communications, after which some radio-frequency front-end building blocks will be explored in detail.

4.2 Ultra Wideband Technology

In 2002, the Federal Communications Commission (FCC) authorized the use of UWB systems in the frequency ranges of 0–960 MHz and 3.1–10.6 GHz [1]. Since these UWB systems occupy the same spectrum as pre-existing radio systems, the allowed power levels are limited to the same order as unintentional radiators.

T. Tsang, K.-Y. Lin, K. Allidina, and M. El-Gamal (✉)
Department of Electrical and Computer Engineering, McGill University
e-mail: {tommy.tsang; kuan-yu.lin; karim.allidina; mourad.el-gamal}@mail.mcgill.ca

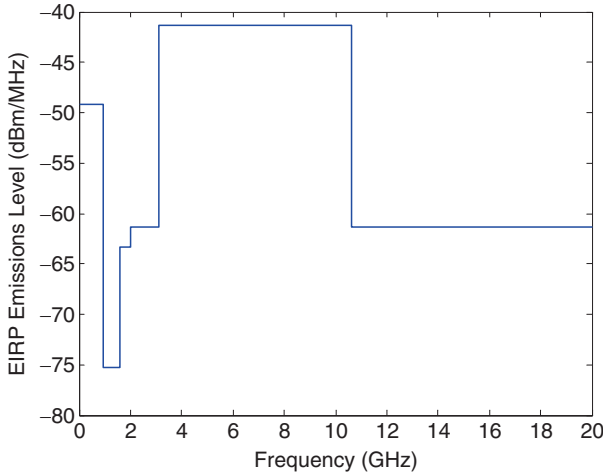


Fig. 4.1 FCC emissions mask for ultra wideband communications

The Equivalent Isotropically Radiated Power (EIRP) limit for the lower frequency band is -49.2 , and -41.3 dBm/MHz for the upper frequency band. The emissions mask also includes a stopband to avoid interference with sensitive GPS systems, and is shown in Fig. 4.1.

The low power spectral density levels required for transmission do have an advantage in that UWB communications appears as noise to other radio systems, which results in a low probability of interception and detection [2].

To be classified as a UWB system, the signal must either have a -10 dB bandwidth that is greater than 500 MHz, or one that is greater than 20% of the center frequency. There are currently two main approaches to create UWB systems under these definitions: one based on Orthogonal Frequency Division Multiplexing (OFDM) technology, and one that uses impulse-based communications.

4.2.1 Orthogonal Frequency Division Multiplexing Proposal

This method has been adopted by the WiMedia Alliance, and it expands upon the OFDM techniques that are currently used in wireless local area network (WLAN) standards [3]. The UWB spectrum from 3.1 to 10.6 GHz is divided into five groups, each containing two to three 528 MHz sub-bands, as shown in Fig. 4.2. Rapid frequency hopping between the sub-bands in each group is implemented, meaning that the transmitter can operate at instantaneous power levels that are three times the FCC limit while still maintaining compliance with the average emission levels. However, this does require band hopping in less than 10 ns, which is less than the settling time of most standard phase-locked loops (PLLs). Techniques to mitigate this problem will be discussed in Section 4.5. The transceiver architecture for a

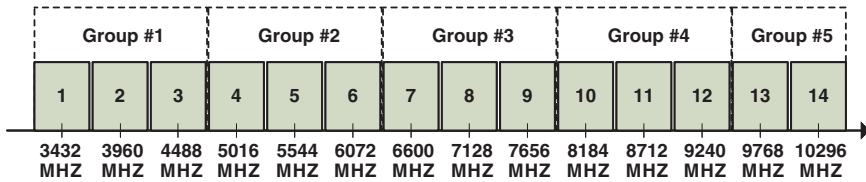


Fig. 4.2 Band Allocation for OFDM UWB Systems

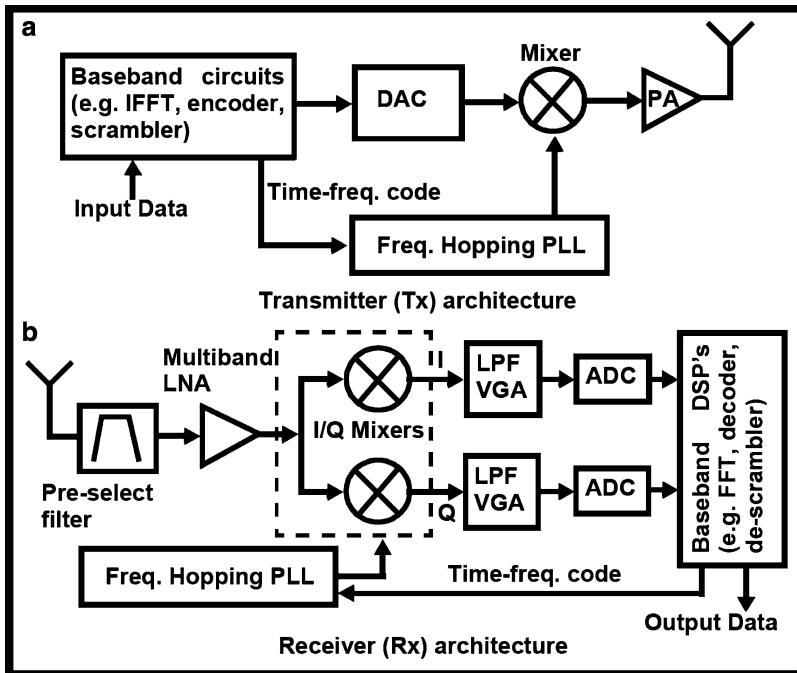


Fig. 4.3 Sample (a) transmitter and (b) receiver architectures for an OFDM UWB transceiver

multi-band OFDM system is shown in Fig. 4.3, and it is very similar to existing narrowband architectures. Since the OFDM signal is synthesized using a high-speed digital to analog converter (DAC), the transmitted signal can make efficient use of the spectrum allocated by the FCC.

4.2.2 Direct-Sequence (DS) Ultra Wideband

This method uses a dual-band pulse-based approach to transmit short duration pulses on the order of nanoseconds or less with bandwidths in the gigahertz range [3]. It utilizes almost all of the UWB spectrum allocation, with the exception

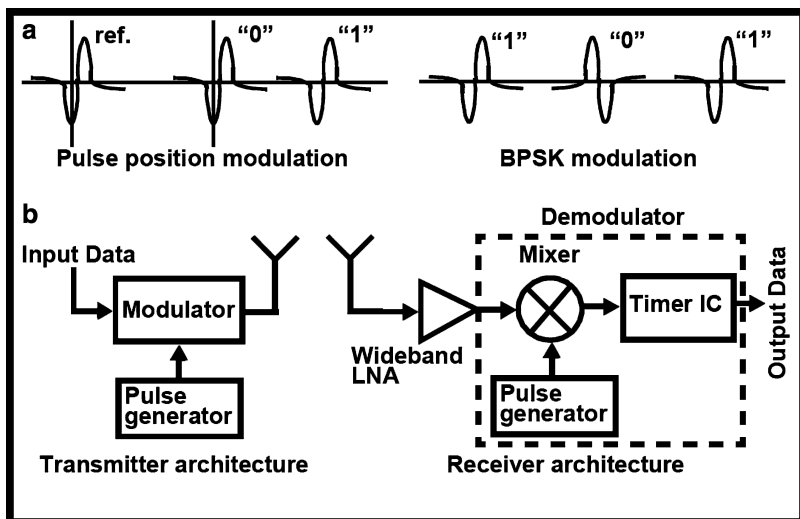


Fig. 4.4 Sample (a) modulation schemes and (b) architecture for a DS-UWB transceiver

of a null from 5.2 to 5.8 GHz to avoid the existing WLAN band. This approach offers digitally modulated signals that are essentially carrier-free, which means simplified transmitter and receiver architectures can be used for communications. An example of this transceiver architecture is shown in Fig. 4.4, along with possible modulated pulse-based waveforms. These waveforms indicate one key advantage of DS-UWB over OFDM systems. The narrow pulses mean that the transceiver only needs to be active for the short period of time in which a pulse is expected. This characteristic is especially attractive for low data-rate communications, in which the transceiver will have a very small duty-cycle and large power savings.

4.3 Low Noise Amplifiers for Ultra Wideband Systems

The previously discussed approaches to UWB communications both require a wideband low-noise amplifier (LNA) to help regenerate the signal while injecting as little noise as possible. In narrowband LNA design, the goal is to minimize the noise figure (NF) while achieving a moderate gain and presenting a $50\ \Omega$ match to the antenna in order to minimize signal reflections [3]. The main challenge in wideband LNAs is to extend this $50\ \Omega$ match over a wide bandwidth while still maintaining other performance characteristics. However, it should be noted that the noise figure specification in an UWB LNA can be relaxed as compared to the narrowband case. This is because the large operating bandwidth prohibits the filtering of interferers, whose average power can be on the order or greater than that of the ambient noise floor [4]. Generally, a noise figure of 3 to 6 dB is sufficient to achieve receiver performance targets.

Recent UWB amplifiers in the literature have used both non-feedback [5–11] and feedback topologies [4, 12–16] to meet different UWB specifications. One example of non-feedback UWB amplifiers are distributed amplifiers (DA), which use parallel transistor amplifiers and artificial transmission lines to periodically combine the gain of each stage [5–8]. These amplifiers offer a wideband impedance match, as well as a relatively flat gain, good linearity, and a constant group delay over a large frequency range. However, CMOS DAs often have high power consumption due to the need for multiple stages. With careful design optimization, UWB amplifiers with fairly good performance (e.g. Gain = 8 dB, NF = 2.9 dB) and moderate power consumption (<20 mW) can be implemented, as demonstrated in [8].

A non-feedback wideband LNA can also be designed using a multi-resonant matching network at the input of the amplifier [9, 10]. Since this type of LNA can be implemented with passive components and only a single-stage amplifier, the power consumption can be fairly low.

However, in both of these previous approaches, the low quality and relatively large size of CMOS on-chip passive components limit the noise performance and cause a large area penalty. One non-feedback topology that does not need on-chip passive components to create a $50\ \Omega$ match is a simple common-gate amplifier. Unfortunately, this topology requires a transconductance (g_m) of around 20 mA/V to present a $50\ \Omega$ impedance to the antenna. This makes it unsuitable for low power applications, especially in deep-submicron technologies.

Feedback topologies can also be used to realize the wideband characteristics needed in a UWB LNA. Variations of wideband resistive feedback LNAs have been implemented in [4, 13], and this technique offers high stability and bandwidth enhancement. However, the noise performance is limited as compared to reactive feedback, a technique used in [14]. Reactive feedback also provides an increased linearity at the expense of chip area. Active feedback techniques can also be used to obtain LNAs suitable for ultra wideband applications [12, 16], however, the extra transistors may cause an increase in power dissipation and limit high frequency performance.

The rest of this section will be devoted to UWB LNA circuit examples for use in both the 3.1–10.6 GHz band and the 0–960 MHz band. The focus will be on simple, compact structures that satisfy common UWB specifications while operating at relatively low power levels.

4.3.1 UWB LNAs for the 3.1–10.6 GHz Band

In this section, we present two UWB LNAs for use in the 3.1–10.6 GHz band that are both based on a common-gate input stage amplifier [12]. As mentioned earlier, a common-gate amplifier can achieve a wideband match to $50\ \Omega$ by setting the g_m of the transistor equal to 20 mA/V, but this may require too much current for low power applications. In order to reduce the input impedance under low power/current conditions, a local feedback stage can be added at the input [17]. Figure 4.5b shows

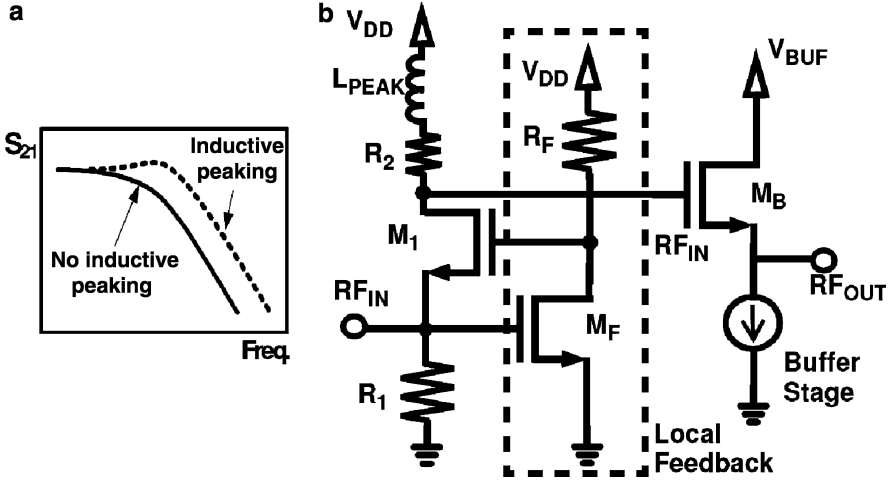


Fig. 4.5 (a) Conceptual view of the inductive peaking technique. (b) Schematic of the inductive peaking common-gate amplifier with local feedback (CGF) for UWB applications

the schematic of a common-gate amplifier, composed of M_1 , R_1 , and R_2 , with local feedback provided by M_F and R_F . L_{PEAK} is used for bandwidth enhancement, which will be described later. We will refer to this circuit as the CGF topology.

According to small-signal analysis, the input impedance of the CG amplifier with local feedback is given by:

$$Z_{in} \approx \frac{1}{g_{m1} \cdot (1 + g_{mF} R_F)}, \quad (4.1)$$

where $(1 + g_{mF} R_F)$ is the voltage gain of the local feedback stage.

The addition of local feedback reduces the input impedance by the amount of its own voltage gain (i.e. $1 + g_{mF} R_F$), when compared to the CG stage. Qualitatively, it can be viewed as a CG stage with a boosted transconductance of $g_{m1}(1 + g_{mF} R_F)$.

One important note here is that the local feedback stage inherently adds a zero, which causes a peaking in the frequency response of the system. The peaking frequency can be approximated by:

$$f_{peak} \approx \frac{1}{2\pi R_F \cdot (C_{gs1} + C_{gdF})}, \quad (4.2)$$

where C_{gs1} and C_{gdF} are the parasitic gate-source and gate-drain capacitances of M_1 and M_F , respectively.

To avoid excessive peaking in the frequency response, f_{peak} should be set such that a flat band gain is achieved. Since the contribution of C_{gdF} is much smaller than those of the other two terms (i.e. R_F and C_{gs1}), the resistance R_F and the sizing of M_1 should be designed such as to place the zero at the frequency which would achieve the desired maximally flat bandwidth extension. Note that decreasing R_F has

a negative impact on the voltage gain of the local feedback, as well as on the input impedance (Eq. 4.1), which has to be compensated for by increasing the transconductance (g_{mF}) of M_F . This is clearly not desirable from a power consumption perspective. Hence, reducing C_{gs1} (i.e. the sizing of M_1) is the preferred method to push the zero up in frequency. However, excessive reduction of the size of M_1 (resulting in a smaller g_{m1}) would lead to an unacceptable high input impedance and excessive channel thermal noise. Therefore, the sizing of M_1 has to be chosen carefully, to meet both the noise figure and power consumption specifications.

Combined with parasitic capacitances, purely resistive loads would result in limited high frequency performance. Simulations have shown that the gain roll-off starts as early as 4 GHz, due to significant nodal parasitic capacitances, contributed by both the gain and buffer transistors. This bandwidth is clearly insufficient, and bandwidth extension techniques are needed. In the first design presented here, a simple inductive shunt peaking approach is used (shown in Fig. 4.5b). This technique enhances the bandwidth of the amplifier by transforming the frequency response from a single pole system to one with two poles and a zero, where the zero is determined primarily by the L/R time constant for bandwidth enhancement. The shunt peaking inductor (L_{peak}) and the resistor (R_2) in Fig. 4.5 are designed to achieve a 60% bandwidth extension with an optimum group delay, which is desirable for optimizing pulse fidelity in broadband systems [18]. The final design shown in Fig. 4.5 has a flat band gain above 6 GHz, which is sufficient to cover both the WLAN and the lower band of the UWB standards.

The second bandwidth extension technique explained here utilizes a two-gain stage approach, where a wideband first gain stage is followed by a narrowband second stage [19]. The conceptual view of this technique and the design schematic are shown in Fig. 4.6.

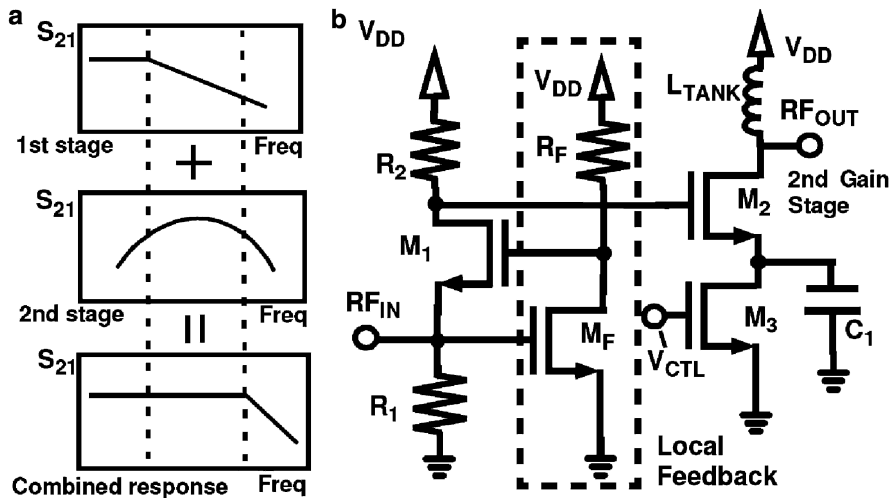


Fig. 4.6 (a) Conceptual view of the bandwidth extension technique in a two-stage amplifier design. (b) Schematic of a gain controllable two-stage amplifier design for multi-standard applications

In this design, the first stage is implemented by a wideband CGF amplifier with a -3 dB cutoff frequency at around 5 GHz. Note that this cutoff frequency is higher than the one in the single stage design, because the second gain stage in Fig. 4.6b contributes less parasitic capacitances than the buffer stage in Fig. 4.5b. The narrowband second gain stage is designed to have the LC elements resonate at 7 GHz. The combination of both frequency responses results in bandwidth extension.

In order to ensure a flat bandwidth extension, the peak gain of the narrowband stage should be equal to the gain of the wideband stage. This can be achieved by properly sizing the transistor of the second stage and its bias current level.

To satisfy the DC bias point of the second stage, a DC shifting transistor M_3 is needed (Fig. 4.6b). By controlling the gate voltage of M_3 , the bias current and gain of the second stage can be tuned. This provides an added gain-control feature for this topology.

With a power budget of less than 10 mW, an upper limit for the current, and thus the available transconductance, in the CGF amplifier is set. Based on the earlier discussion, a minimum sizing of M_1 that satisfies the gain requirement is chosen, while the maximum allowable sizing of M_F , within the power budget limit, is used. The final sizing of M_F is approximately three times larger than M_1 . The voltage gain of the local feedback is set to be six, which is again constrained by the power budget, as well as the desired input impedance level. With a supply voltage of 1.8 V, the core power consumption is 5.8 mW. This CGF amplifier is used for both the first and second UWB amplifiers in this section.

The second gain stage is a narrowband common source amplifier with a nominal bias current of 1 mA at 1.8 V. An AC ground coupling capacitor C_1 is connected to the source of transistor M_2 , as shown in Fig. 4.6b. The buffer stage, which is designed to drive a 50Ω external load for testing purposes, is independently biased by a current mirror. This results in a 6 dB difference between the measured power gain at the output of the test setup and the actual voltage gain of the LNA core.

Figures 4.7 and 4.8 show the microphotographs and the S-parameters (S_{21} and S_{11}) of the two UWB amplifiers, respectively. Both designs are implemented in a standard CMOS 0.18 μm process. Including the buffer, but excluding all testing pads, the single stage inductive peaking amplifier occupies an active area of 0.14 mm^2 , while the two-stage CGF amplifier consumes an area of 0.17 mm^2 . A ratioed design approach, with the same unit transistor and resistor fingers, is employed to minimize the effect of mismatch and process variations on the amplifiers performances.

The measured S_{11} and S_{21} plots of the single stage inductive peaking CGF amplifier are shown in Fig. 4.8a. A good wideband input matching (i.e. $S_{11} < -10$ dB) is achieved across the 1–10 GHz band. With a power consumption of 5.8 mW at 1.8 V, a 6 GHz flat-band gain of 12 dB is achieved, with a -3 dB cutoff frequency above 7 GHz. The amplifier continues to provide a wideband gain of higher than 7 dB with good input matching at 1.4 V supply. This demonstrates the effectiveness of this approach for multi-standard applications.

Figure 4.8b shows the measured S_{11} and S_{21} plots of the two-stage UWB amplifier. At a 2.5 V supply, the amplifier has a flat band gain of 13 dB over a 7 GHz

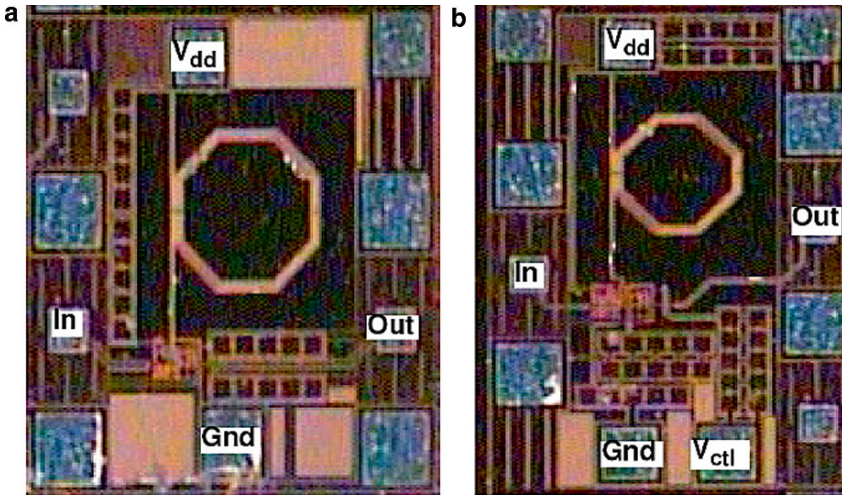


Fig. 4.7 Micrographs of the UWB amplifiers: (a) the single stage inductive peaking CGF amplifier, and (b) the two-stage gain controllable CGF amplifier

bandwidth. The -3 dB cutoff frequency is at 8.5 GHz. By controlling the gate voltage of transistor M_3 in the second gain stage (Fig. 4.6b), a 5 dB control of the overall amplifier gain is achieved, without affecting the quality of the input matching. An excellent input reflection coefficient of $S_{11} < -17$ dB is achieved across the full 1–10 GHz frequency band.

The measured noise figure for both amplifiers is shown in Fig. 4.9a. For the single-stage CGF, the noise figure across the 2–6 GHz band ranges from 4.4 to 6.7 dB. The noise figure of the two-stage CGF is 4.1–4.8 dB between 2–7 GHz. It is approximately 0.6–1 dB higher than the expected value, and this likely due to a slight reduction in the overall gain of the amplifier.

Two-tone tests at 3 GHz with a tone spacing of 1 MHz are performed to measure intermodulation (IM) distortions of the amplifier. A center frequency of 3 GHz is chosen because both the IM2 at 6 GHz and IM3 at 3 GHz fall within the flat-band region of the amplifier. From the IM measurements on the single-stage CGF amplifier, the second-order (IIP2) and the third-order (IIP3) intermodulation intercept points are -7 and -13.5 dBm, respectively. The measured 1 dB compression gain (P_{1dB}) is -23 dBm. From the two-tone test performed on the two-stage CGF amplifier, the measured P_{1dB} , IIP2 and IIP3 are -23.7 , -7.5 , and -13 dBm, respectively, as shown in Fig. 4.9b.

Both designs have good reverse isolation (S_{12}) lower than -32 dB and output reflection coefficient (S_{22}) lower than -13 dB, across the band of interest. The performance of these designs, as well as a comparison to a recent LNA in the literature, is given in Table 4.1.

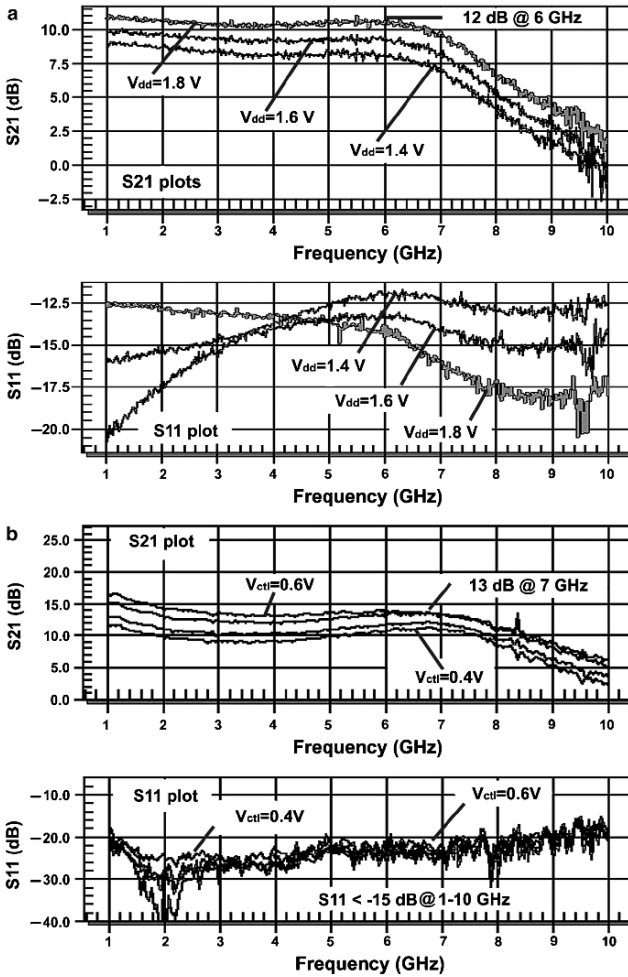


Fig. 4.8 Measured S_{21} and S_{11} plots of (a) the single stage inductive peaking CGF amplifier, and (b) the two-stage gain controllable CGF amplifier

4.3.2 UWB LNAs for the 0–960 MHz Band

The challenges in designing an LNA in the 0–960 MHz UWB frequency range are slightly different than for those operating from 3.1–10.6 GHz. Since the bandwidth of this band is smaller, the allowed signal power is reduced, and this limits the achievable data-rates. However, since this band is at a lower frequency, the power consumption of the transistors can be lower. This fact can be exploited to create very low power transceivers for applications such as wireless sensor networks, which only require data-rates in the 10–100 kbit range. And, as mentioned earlier, the

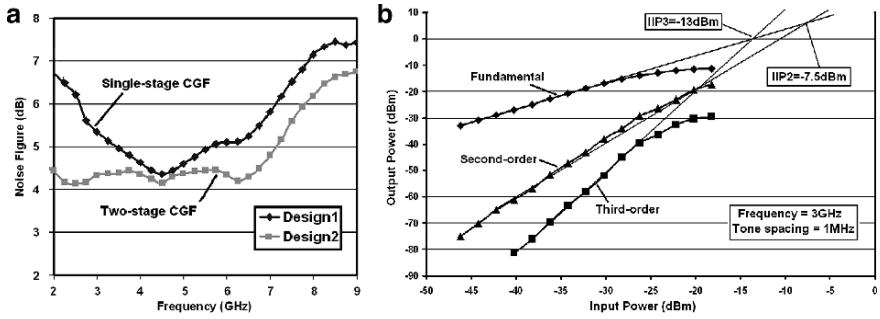


Fig. 4.9 (a) Measured noise figure of the two UWB amplifiers, and (b) measured IIP2 and IIP3 of the two-stage controllable CGF amplifier

Table 4.1 Performance summary of the two CMOS UWB amplifiers in this work, and a comparison with [17]

	Design 1	Design 2	CICC'06 [17]
Technology	CMOS 0.18 um	CMOS 0.18 um	CMOS 0.18 um
Topology	One stage CGF with inductive peaking	CGF first stage, with a narrowband second stage	Common-gate with inductor current sink, gm enhancement
V_{dd}/P_{dd} (V/mW)	1.8/5.8	2.5/9.3	1.8/7.9
S_{11} (dB)	<-10 (1-10 GHz)	<-17 (1-10 GHz)	<-6 (1.7-13.6 GHz)
S_{21} -flat (dB)	12 (2-6 GHz)	13 (2-7 GHz)	15.4 (3-10.5 GHz)
S_{22} (dB)	<-15 (2-6 GHz)	<-13 (2-7 GHz)	-
S_{12} (dB)	<-32 (1-10 GHz)	<-40 (1-10 GHz)	-
-3 dB freq. (GHz)	7	8.5	13.6
NF_{flat} (dB)	4.4-6.7	4.1-4.8	3.1-4.0*
P_{1dB} (dB)	-23	-23.7	-
IIP ₂ (dB)	-7 at 6 GHz	-7.5 at 6 GHz	-
IIP ₃ (dB)	-13.5 at 3 GHz	-13 at 3 GHz	-11.2* at -5 GHz
Variation in group delay (ps)	30	30	-
Area (mm ²)	0.14 (incl. buffer)	0.17 (incl. buffer)	0.22
Other feature	Operational down to 1.4 V	5 dB gain-control	4.25 kV ESD-protected

*Simulation results

narrow pulse durations used in UWB communications also offer enhanced power savings at low data rates.

This section will present two LNAs for use in the 0-960 MHz UWB band. The primary focus for the LNAs is to minimize power consumption while still maintaining reasonable performance.

The topology used for the first LNA is shown in Fig. 4.10a [16]. It is based on a current amplifier that can be constructed from a flipped voltage follower [20],

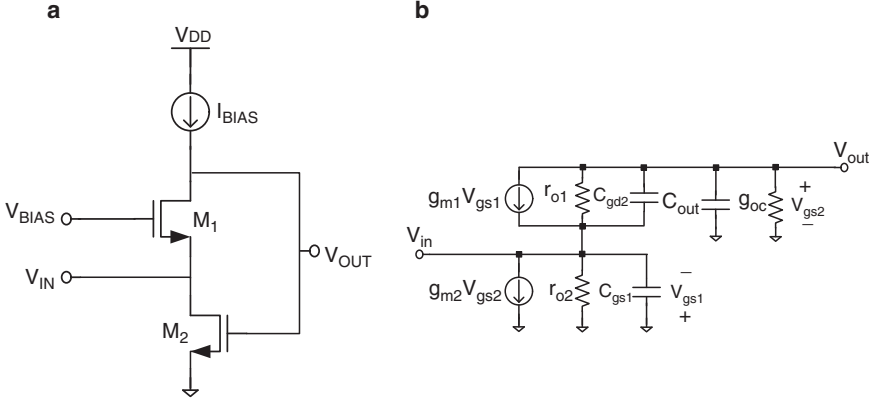


Fig. 4.10 (a) LNA topology based on a common-gate amplifier with active feedback and (b) the related small-signal model

and additional analyses relevant to LNA design will be presented. The circuit uses the common-gate transistor M_1 as the core amplifier, with shunt-feedback provided by the common-source transistor M_2 . The fact that both transistors share the same bias current makes this structure very attractive for low power operation. The small-signal model is shown in Fig. 4.10b, where C_{out} includes C_{gs2} , C_{gd1} , and the capacitance of the current source, and g_{oc} is the output conductance of the current source. The additional transconductance caused by the body effect in transistor M_1 is included in g_{m1} .

Analysis shows that the output impedance of M_2 can be ignored as long as it is larger than $2\text{ k}\Omega$, which is possible even in deep submicron technologies. Intuitively, this makes sense since the output impedance of M_2 is connected to the input node, and it will be much larger than the $50\ \Omega$ source impedance, R_S .

The low frequency input impedance of the LNA can be derived as:

$$Z_{in}(\omega = 0) \approx \frac{g_{o1} + g_{oc}}{g_{m1}g_{m2}}. \quad (4.3)$$

Since the input impedance is multiplied by the combined conductances of M_1 and the current source, this equation shows that an input impedance of $50\ \Omega$ can be obtained quite easily with low transconductance values for M_1 and M_2 .

Including the capacitances in the analysis of the input impedance yields:

$$Z_{in} \approx \frac{sC_{out} + g_{o1} + g_{oc}}{s^2C_{gs1}C_{out} + sg_{m1}C_{out} + g_{m1}g_{m2}}. \quad (4.4)$$

This equation shows that there is an inductive component to the input impedance that is caused by the feedback loop, which forms a gyrator [21]. The inductive portion of the input impedance can be represented as:

$$L \approx \frac{C_{out}}{g_{m1}g_{m2}}. \quad (4.5)$$

This means that the output capacitance must be decreased or the transconductances of M_1 and M_2 must be increased to reduce the value of the inductance, so it does not affect the circuit below 960 MHz.

The low-frequency gain of the circuit is given by:

$$\frac{V_{out}}{V_{in}}(\omega = 0) \approx \frac{g_{m1} + g_{o1}}{R_s g_{m2}(g_{o1} + g_{m1}) + g_{o1}} \approx \frac{1}{R_s g_{m2}}, \quad (4.6)$$

which shows that the gain can be maximized by increasing the transconductance of M_1 , but mainly by decreasing the transconductance of M_2 .

The noise factor can be analyzed by including the drain noises of M_1 and M_2 , and the noise due to the current source. The gate noise has been neglected, since a careful layout can reduce the gate resistances, rendering them insignificant. After simplification, the noise factor equation is:

$$F \approx 1 + \gamma/\alpha \left(\frac{1}{R_s g_{m1}} + g_{m2} R_s \right) + \frac{g_{oc}}{R_s g_{m1}^2}, \quad (4.7)$$

where γ/α accounts for short channel effects, and is assumed to be the same for both M_1 and M_2 . This equation shows that the noise factor can be minimized by increasing the transconductance of M_1 and decreasing that of M_2 (i.e. increasing the gain).

The circuit is designed using an ST Microelectronics 90 nm CMOS process, and operates from a 1 V supply. In accordance with the theoretical analysis presented, the transistors have been sized to obtain a large transconductance from M_1 and a small one from M_2 . Placing a current sink in parallel with M_2 would serve to further reduce its transconductance, however the decrease in bias current would also decrease the linearity of the amplifier. The body of M_1 is tied to ground in order to increase its transconductance from g_m to $g_m + g_{mb}$, which serves to reduce the power consumption of the LNA. A simple common-drain buffer was also added to the LNA to drive the 50 Ω impedance of measurement equipment.

The circuit was designed using the Cadence design environment and simulated using Spectre-RF. This particular LNA is currently under fabrication, so only simulation results are available.

Figure 4.11 shows the S-parameter responses of the circuit. These results indicate that the LNA is input and output matched over the entire bandwidth. The inductive portion of the input impedance within the bandwidth is around 1 nH, which is small enough to have almost no effect on input matching at the frequencies of interest. The gain is 10.5 dB, and the 3 dB bandwidth of the circuit is 1.66 GHz. S-parameter simulations also show that the amplifier is unconditionally stable, and has a reverse isolation of over 35 dB.

Figure 4.12 shows the noise figure response of the circuit. At low frequencies, the flicker noise caused by the small transistor sizes increases the noise figure, whereas

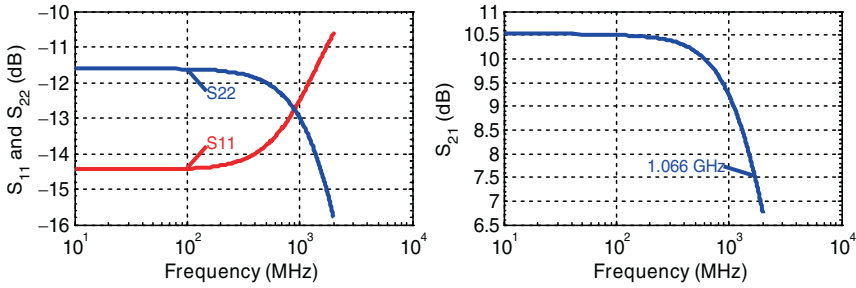


Fig. 4.11 S_{11} , S_{22} , and S_{21} responses of the amplifier

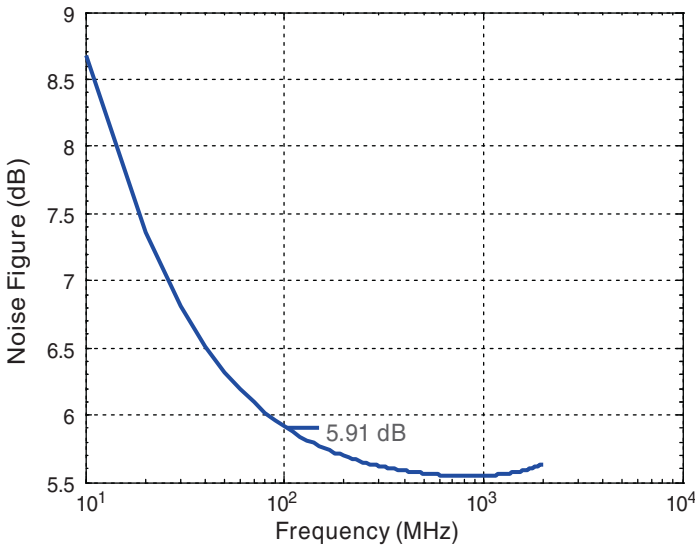


Fig. 4.12 Noise figure response of the amplifier

beyond 100 MHz the noise figure decreases below 6 dB. The minimum noise figure is 5.55 dB, and the average over the entire band (starting at 100 MHz) is 5.64 dB.

Two-tone tests are applied to determine the IIP3 of the circuit. Figure 4.13a shows the response at a frequency of around 500 MHz, and the IIP3 is -4.2 dBm. The input referred 1 dB compression point at this frequency is -14.5 dBm. Figure 4.13b shows how the IIP3 varies from 100 to 900 MHz. The IIP3 stays fairly constant around the average value of -4.4 dBm, dipping just below -5 dBm near the upper edge of the band.

The DC current consumption of the circuit is $425 \mu\text{A}$ from a 1 V supply, yielding a power consumption of $425 \mu\text{W}$. The turn-on time of the LNA is also tested since it is an important parameter for duty cycled operation. Using the PMOS current source as the control mechanism, the LNA is able to turn on in less than 1 ns. The leakage current with the LNA turned off and the input signal still active is under $3 \mu\text{A}$. The variation in group delay for this amplifier is less than 30 ps.

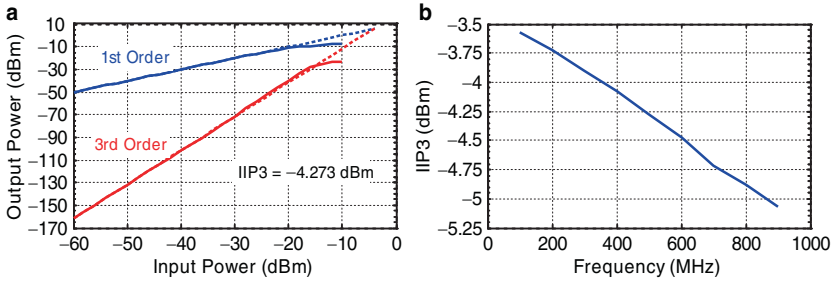


Fig. 4.13 IIP3 simulation results at (a) 500 MHz and (b) across the band of interest

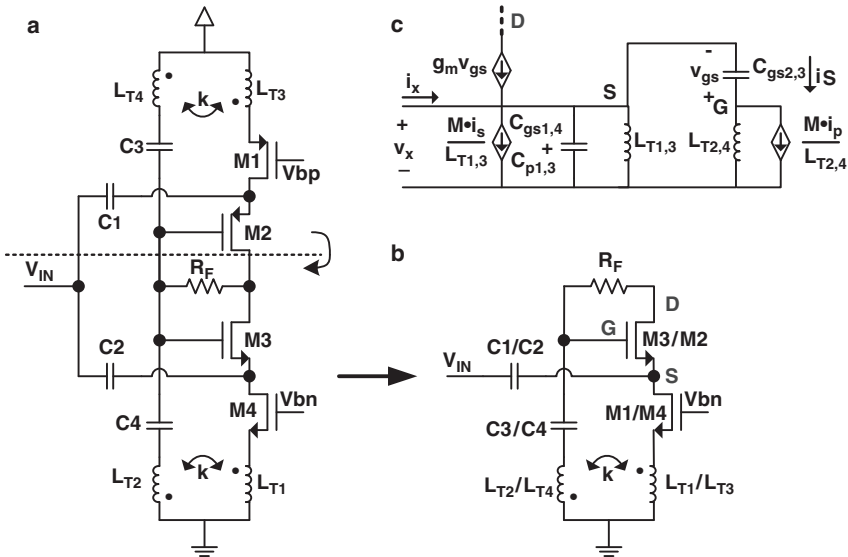


Fig. 4.14 (a) Schematic, (b) simplified schematic, and (c) small-signal equivalent circuit of the LNA

The analysis of the previous LNA showed that shunt-feedback is an effective way to lower the input impedance of a common-gate amplifier and achieve a low power match to 50 Ω. The second LNA presented here uses this same principle, but reactive components are used to implement the feedback as opposed to transistors. This approach offers the possibility of lower noise and power consumption, and in this work, a transformer provides the desired feedback. Unfortunately, inductors at these low frequencies are too large to be implemented on-chip in a reasonable amount of area. To solve this problem, the inductive properties of bond-wires were used to re-realize the transformer [22]. This novel approach allows a high quality transformer to be implemented without sacrificing chip area, while still maintaining the goal of a compact solution.

The full and simplified schematics of the LNA are shown in Fig. 4.14, along with the small signal model. The topology is based on the two common-gate transistors

M_2 and M_3 , which are in a current reuse configuration to conserve power. The input is connected to the source nodes of both gain transistors to double the overall effective transconductance of the amplifier (i.e. $G_m = g_{m2} + g_{m3} = 2g_m$). Since both transistors are stacked, the current consumption to achieve the G_m of $1/50 \Omega$ is reduced to half that of a single transistor CG amplifier, at the cost of a reduced voltage headroom and linearity.

A feedback resistor R_F is placed between the output of the LNA core and the gate nodes of transistors M_2 and M_3 . As mentioned earlier, the resistive feedback R_F improves the stability and gain bandwidth of the amplifier. In addition, it supplies the necessary DC biasing to transistors M_2 and M_3 .

The anti-phase transformers composed of L_{T1} to L_{T4} implement the negative shunt feedback mentioned earlier to further decrease the current needed for a 50Ω input match.

Transistors M_1 and M_4 act as duty-cycle on/off switches to turn off the LNA and to save power during idle periods. However, the extra parasitics introduced by those transistors affect the input impedance matching and the quality factor of the transformers, thus these transistors must be carefully sized.

The input impedance of this structure can be found from the small-signal model, and the simplified expression is:

$$Z_{in} \approx \left[2 \left(\frac{1}{sL_{T1,3}} + (1+k)g_{m1,2} + (2+2k)sC_{gs2,3} + s(C_{gs1,4} + C_{P1,3}) \right) \right]^{-1} \quad (4.8)$$

where k is the coupling factor of the transformer. The above equation holds for the case where the each transformer has a 1:1 turns ratio. It can be seen that the input impedance decreases as the transconductance of the common-gate transistors increases, which is expected. It also decreases as the coupling factor increases, since this leads to an increase in the shunt feedback.

The simplified noise factor expression for the case where each transformer has a 1:1 turn ratio is given by:

$$F = 1 + \frac{1}{((1+k)g_m R_S)^2} \frac{\gamma}{\alpha} g_m R_S + 4\delta\alpha \frac{\omega^2 C_{gs}^2}{5g_m} R_S, \quad (4.9)$$

where R_S is the source resistance, γ and α are short channel MOSFET noise parameters, and δ is the gate noise coefficient.

For the case where the input impedance is matched to the source, i.e. $(1+k)g_m R_S = 1$, this expression can be simplified to:

$$F = 1 + \frac{1}{1+k} \left(\frac{\gamma}{\alpha} + \frac{4}{5}\delta\alpha \frac{\omega^2 C_{gs}^2}{g_m^2} \right). \quad (4.10)$$

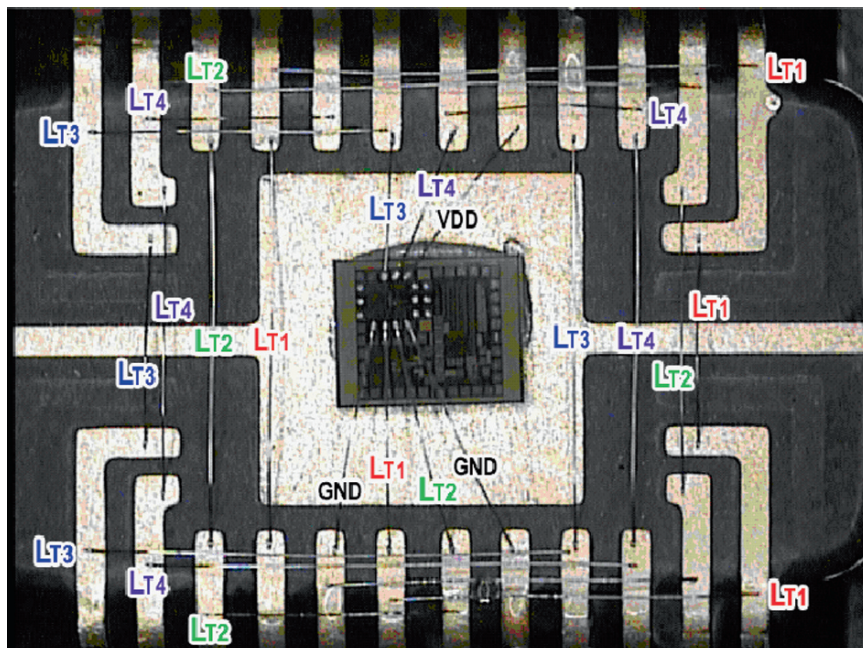


Fig. 4.15 Photograph of the LNA in a 24-pin package with bonding wire transformers

It can be seen that an increase in the coupling factor, and thus the feedback, reduces the noise factor of the amplifier.

This amplifier was fabricated in a TSMC $0.18\ \mu\text{m}$ CMOS process, and the die size, including pads and a buffer to drive measurement equipment, is $0.67 \times 0.63\ \text{mm}$. Figure 4.15 shows a picture of the die in a 24-pin glass-ceramic bottom package, along with the bond-wire transformers. The transformers are bonded around the chip using high precision automated ultrasonic wedge bonding to achieve repeatable, measurable, and precise positioning. The bonding wires are composed of Aluminum (1% silicon) with a diameter of $25.4\ \mu\text{m}$. The loop height of all the bonding wires is set to $350\ \mu\text{m}$ at their mid-points.

Four parallel bonding wires around the package implement the two bonding wire transformers. One input of the transformer is formed by connecting in series four bonding wires around the package, using several package leads as connection sites. To enhance the coupling coefficient k between the bonding wires, the wires are bonded with an average spacing of $300\ \mu\text{m}$. The total length of the bonding wire transformer representing inductors L_{T1} to L_{T4} is approximately 25 mm. Figure 4.16 shows the measured coupling coefficient and inductance value for the primary inputs of the transformers after de-embedding the test setup. The bonding wire transformers have a measured input inductance $L = 35\ \text{nH}$, and a coupling coefficient $k = 0.25$.

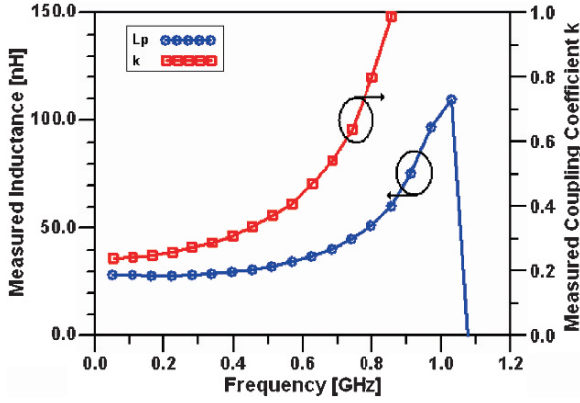


Fig. 4.16 The measured inductance and coupling factor of a bonding wire transformer

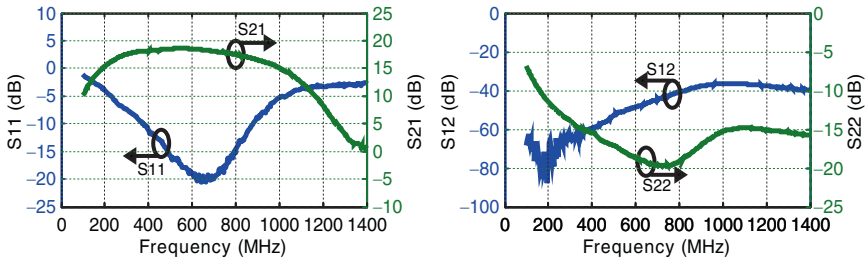


Fig. 4.17 S-Parameter responses of the LNA with bonding wire transformers

The measured S-parameter responses of the LNA with bonding wire transformers are plotted in Fig. 4.17. The design achieves an S_{11} lower than -10 dB and an S_{21} better than 17 dB from 375 to 875 MHz with an LNA core power consumption of $698.5 \mu\text{W}$ at 1.5 V. The maximum gain and minimum S_{11} are 18.5 dB at 525 MHz and -20 dB at 650 MHz, respectively. The -3 dB bandwidth is 770 MHz, spanning from 200 to 970 MHz. The design has a good reverse isolation S_{12} lower than -37 dB, and an output reflection coefficient S_{22} lower than -13 dB across the band of interest.

The measured noise figure and group delay of the LNA are shown in Fig. 4.18. The noise figure ranges from 4 to 6.5 dB across the 400 to 930 MHz band, which is 1 dB higher than expected from simulation. This discrepancy is mainly caused by a slight reduction in the expected overall gain of the amplifier, and probably due to the inaccurate modeling of the transistors' gate and channel noises. The average group delay is 510 ps across the band of interest with variations less than 100 ps. This low variation group delay curve without peaking is desirable for optimal pulse fidelity in pulse-based UWB systems.

A two tone test is performed at 400 MHz to measure the second and third-order intermodulation distortions (IM2, IM3). Both the IM2 at 800 MHz and the IM3 at

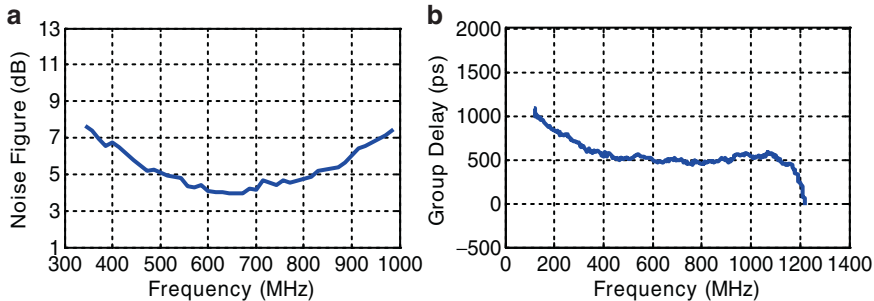


Fig. 4.18 (a) Noise figure and (b) group delay of the LNA with bonding wire transformers

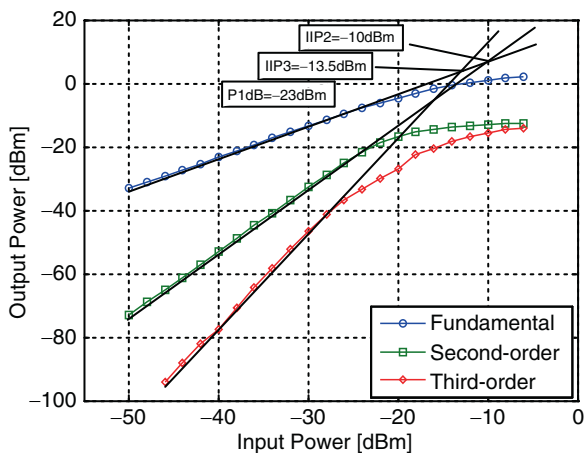


Fig. 4.19 Measured second and third order intermodulation distortions of the LAN with bonding wire transformers

400 MHz fall within the operating band of the LNA. The measured intermodulations are plotted in Fig. 4.19. From the measurements, the 1-dB compression point (P_{1dB}), second-order (IIP_2), and third-order (IIP_3) intermodulation intercept points are -23 , -10 , -13.5 dBm, respectively. Over the operating band, the intermodulation distortions are found to be relatively constant.

A summary of the results obtained with the two LNAs described in this section, as well as a comparison with other recently published LNAs, is shown in Table 4.2.

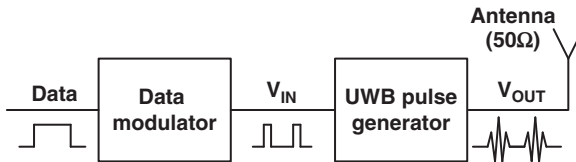
4.4 Pulse Based Transmitters for DS-UWB Systems

Unlike LNAs, OFDM and DS-UWB systems require very different types of transmitters. The topologies used for OFDM UWB modulation are very similar to those of existing narrowband systems. Therefore, this section will focus on the implementation of pulse-based transmitters.

Table 4.2 Summary of the LNAs for the 0–960 MHz band, as well as a comparison to other recent works

	This work*	This work	JSSC'06 [4]	RFIC'07 [23]
Topology	Common-gate LNA with active feedback	Common-gate LNA with feedback through bonding wire transformers	Common-gate with resistive shunt feedback	Multi-stage LNA with resistive feedback
Technology	90 nm CMOS	0.18 μm CMOS	0.13 μm CMOS	90 nm CMOS
Bandwidth (MHz)	100–1,660	200–970	100–930	400–1,000
Gain (dB)	10.5	17	13	16
Average NF (dB)	5.4	5.8	4	4.4
IIP3 (dBm)	−4.5	−13.5	−10.2	−17
Supply voltage (V)	1	1.5	1.2	1.2
Power consumption (mW)	0.425	0.699	0.72	16.8
Differential amplifier?	No	No	Yes	No

*Simulation results

**Fig. 4.20** Block diagram of a typical DS-UWB transmitter

A typical pulse-based transmitter is shown in Figure 4.20. It consists of a data modulator, which activates the UWB pulse generator in different ways depending on the modulation type (e.g., OOK, PPM, BPSK). The UWB pulse generator then produces modulated pulses with precise frequency characteristics that satisfy the FCC UWB spectral mask. These circuits are generally categorized into analog and digital implementations [24–27].

The transmitter in [24], based on an analog approach, generates UWB pulses using a multiplier circuit. A triangular signal is multiplied with a carrier oscillating signal to produce an up-converted triangular pulse. By controlling the width of the triangular signal and the frequency of the carrier signal, the center frequency and bandwidth of the UWB pulse can be tuned to satisfy the FCC UWB spectral mask. This design consumes 2 mW from a 1.8 V supply for a pulse repetition frequency (PRF) of 40 MHz in a 0.18 μm CMOS process.

An alternative analog-based approach utilizes a passive on-chip filter to shape the spectrum of the pulse into the FCC UWB spectral mask [25]. The resulting pulse is then driven by a wideband amplifier to the 50 Ω antenna. In this approach, the driver amplifier consumes constant biasing current. Therefore, the transmitter

consumes moderate power of 10.8 mW from a 1.8 V supply in a 0.18 μm CMOS process with a large silicon area of 0.96 mm^2 .

For low power applications, the transmitter is better implemented using a digital approach, due to the low duty cycles involved. The transmitter in [26] utilizes a NOR gate and inverters to generate a 1st order pulse. This pulse is then shaped into a second order Gaussian pulse by a switching amplifier with a passive pulse shaping network. This technique can significantly reduce the power consumption to below 100 μW from a 1.8 V supply for a PRF of 100 MHz. However, this transmitter generates only a second order Gaussian pulse, which theoretically cannot efficiently satisfy the FCC UWB spectral mask [28].

The transmitter in [27], also based on a digital approach, utilizes a set of digital triangular pulse generators. Each generator creates a triangular pulse with a different pre-determined delay time. All the pulses are combined in a specific order and driven to the 50 Ω antenna. This technique generates digitally specific pulse shapes, which eliminates the need for analog filtering. However, this design is sensitive to component mismatches, which can disrupt the pulse shapes. Since the power is only consumed during switching transients, this transmitter, fabricated in a 0.18 μm CMOS technology, consumes only 675 μW from a 1.8 V supply for a PRF of 1 MHz.

This section will present two UWB pulse generators: one whose output covers the entire UWB band for high data-rate communications, and another that emits pulses with a smaller bandwidth for use in low power, low data-rate sub-band systems. These different pulse-based transmitters present different challenges. For example, a full-band UWB transmitter should have an output power spectral density (PSD) that covers as much area as possible under the FCC UWB spectral mask. This maximizes the total energy transmitted and leads to a higher probability of detection in the receiver. For sub-band UWB transmitters, the PSD of the UWB pulse should have low sidelobes to minimize the interference with adjacent sub-bands. To compensate for mismatch and process variations, the transmitters should also be able to control the pulse shape and, consequently, the output PSD.

The pulse generators presented in this section will meet the above requirements while minimizing the power consumption to less than 300 μW . To achieve this goal, the pulse generators are implemented using a digital approach, in order to eliminate continuous power consuming analog circuits. For simplicity, both transmitters use OOK modulation, however, it is straightforward to modify the transmitters for other modulation schemes.

4.4.1 Full-Band 3.1–10.6 GHz UWB Pulse Generator

The schematic of the full-band UWB pulse generator is shown in Fig. 4.21a [29]. First, a voltage-controlled current-starved inverter and a NOR gate produce a variable width rectangular pulse, V_{REC} , to power on the fast start-up tunable ring oscillator. The delay of the inverter is controlled by the amount of current supplied to it, which is set by the voltage V_{DELAY} in a current mirror, as shown in Fig. 4.21c.

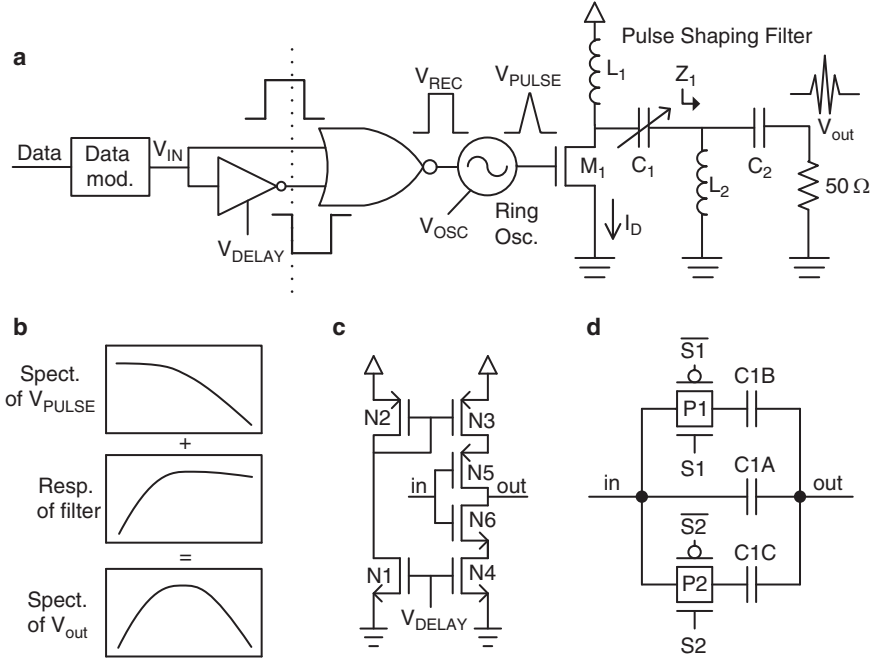


Fig. 4.21 Schematic of (a) the full 3.1–10.6 GHz UWB transmitter, (b) the conceptual view of the operation of the UWB pulse generator, (c) the voltage-controlled current-starving inverter, and (d) the tunable capacitor C_1

This delay establishes the pulse width of the signal V_{REC} , which then sets the number of the oscillation cycles of V_{PULSE} and, consequently, the width of the UWB pulse. The voltage V_{OSC} , applied to the ring oscillator, controls its frequency of oscillation. Therefore, the bandwidth and the center frequency of the UWB output pulse can be tuned by the voltages V_{OSC} and V_{DELAY} to compensate for mismatch and process variations.

The V_{PULSE} signal generated by the ring oscillator however does not meet the FCC spectral mask in the low frequency range, as described in Fig. 4.21b. To attenuate these frequencies, a high-order high pass filter is needed. The pulse shaping high pass filter in this work is shown on the right hand side of Fig. 4.21a. An NMOS switch M_1 acts as a switching transconductance amplifier, whose current (I_D) is filtered by the pulse shaping filter and sent to the 50Ω antenna. The pulse shaping filter consists of a fourth order high pass network, implemented by cascading two second order LC high pass filters (i.e. a total of two inductors and two capacitors). Since the filter is passive, no power is consumed. MIM capacitors and standard on-chip inductors are used. The transfer function of the pulse shaping filter in the S-domain is written as:

$$H(s) = \frac{V_{OUT}}{I_D} = \left(\frac{s^2 L_1 C_1}{s^2 L_1 C_1 + s Z_1 C_1 + 1} \right) \left(\frac{s^2 L_2 C_2}{s^2 L_2 C_2 + s R_L C_2 + 1} \right). \quad (4.11)$$

The fourth order high pass filter offers abrupt roll-off due to the four zeros at $s = 0$. The LC values (i.e. pole locations) are chosen to meet the stop band edge at 3.1 GHz. To control the corner frequency of the filter for spectral mask fitting purposes, the design utilizes a variable capacitor C_1 , consisting of three capacitors and two pass transistors, as shown in 4.21d. This variable capacitor C_1 achieves four different capacitances: 60, 120, 140, and 200 fF.

The fast start-up tunable ring oscillator is shown in Fig. 4.22. It consists of three stages, and the frequency of oscillation is tuned with transistors M_5 and M_8 . The biasing voltages v_{OSC1} and v_{OSC2} are set by a current mirror circuit with a single control voltage v_{OSC} . Transistors M_1 and M_4 force initial oscillation conditions when turned off, therefore, allowing almost instantaneous oscillation start-up time. At the output, a buffer, consisting of inverter stages, is used to drive the NMOS switch.

With a voltage supply of 1.8 V, the power consumption of the full-band transmitter is 237.4 μ W. Figure 4.23a shows the time domain response of the FCC compliant

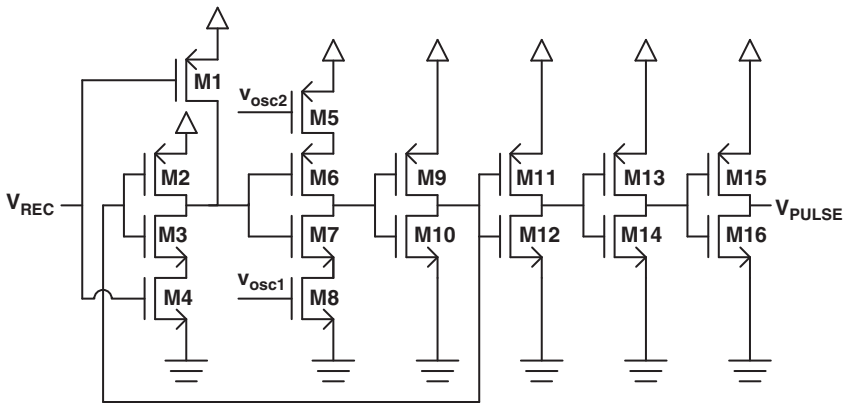


Fig. 4.22 Schematic of the fast start-up tunable ring oscillator

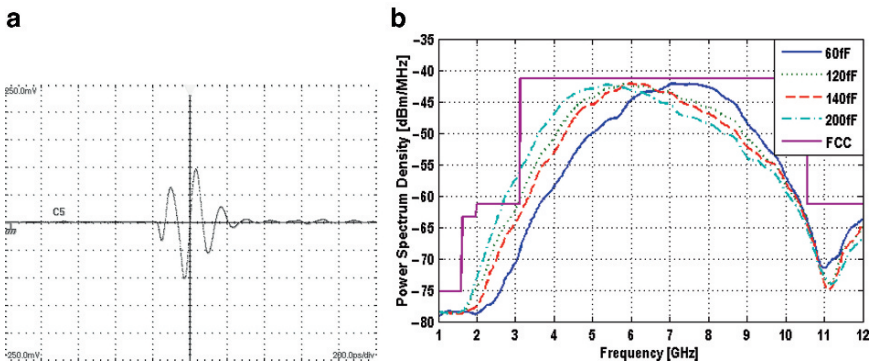


Fig. 4.23 (a) Time and (b) frequency domain measurements of the full-band pulse transmitter

full-band pulse, which has a peak to peak voltage of 200 mV and a pulse width of 455 ps. Figure 4.23b shows the envelope of the measured output PSD for different capacitance values C_1 in the high-pass pulse shaping filter. The outputs are generated by the full-band UWB transmitter at a PRF of 110 MHz. As shown, a smaller capacitance C_1 shifts the lower corner frequency of the filter to a higher frequency. This corner frequency shifting is used to compensate for variations. The signal meets the FCC UWB spectral mask when the 120 fF capacitance C_1 is selected. This UWB signal has a -10 dB bandwidth of 5 GHz and achieves a spectral power efficiency of 30%, which is the ratio between the pulse radiation power to the maximum FCC radiation power limit of 556 μ W, when the full 7.5 GHz bandwidth is considered.

4.4.2 Low Sidelobe Sub-Band 3.1–10.6 GHz UWB Transmitter

Low bandwidth UWB pulses can easily be generated using the full-band UWB transmitter presented earlier (Fig. 4.21) by changing the control voltage V_{DELAY} to lengthen its oscillation time. However, the generated gated sine wave signal possesses high sidelobes, which can cause interference. To suppress these sidelobes, the gated sine wave can be multiplied by a triangular signal, forming an up-converted triangular pulse. This multiplication technique has been explored in [24]. However, in this design, instead of using a power consuming differential multiplier circuit, two NMOS switches, M_1 and M_2 , are used. This significantly reduces power consumption and complexity.

Figure 4.24 shows the schematic of the low sidelobe sub-band UWB transmitter [29]. Similar to the first design, a high pass pulse shaping filter is used to suppress the low frequency content of the pulses. A fast start-up tunable ring oscillator generates the carrier signal, V_{PULSE} , which is connected to the top NMOS

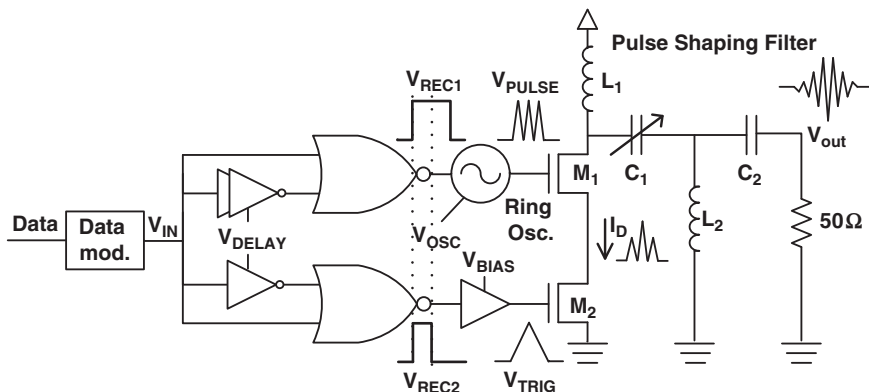


Fig. 4.24 Schematic of the low sidelobe sub-band UWB transmitter

switch, M_1 . The triangular signal, V_{TRIG} , is generated by charging and discharging the gate capacitance of M_2 . The charging and discharging current is controlled by a voltage-controlled current-starved buffer. It charges when V_{REC2} is high, and discharges when V_{REC2} is low. Since the NMOS switches M_1 and M_2 are in series, the resulting drain current, I_D , has the same shape as the voltage multiplication of the inputs.

To synchronize the multiplication of the carrier and the triangular pulse, two rectangular pulse generators are used. Both rectangular pulses, V_{REC1} and V_{REC2} , are generated at the same starting time. Since the triangular pulse consists of charging and discharging periods, the pulse width of the bottom signal V_{REC2} is two times shorter than that of the top signal, V_{REC1} , to make sure that both signals V_{TRIG} and V_{PULSE} overlap in time.

With a voltage supply of 1.8 V, the power consumption of the sub-band transmitter is 254.9 μ W. Figure 4.25a shows the time domain response of the sub-band 528 MHz bandwidth pulse, which has a peak to peak voltage of 89 mV and pulse width of 4 ns. A higher peak voltage is obtained from the first transmitter ($V_{pk-pk} = 200$ mV) due to the larger bandwidth, which not only allows more energy to be transmitted, but also condenses this energy over a shorter time span.

Figure 4.25b shows four 528 MHz bandwidth pulses at different center frequencies generated by this transmitter at a PRF of 30 MHz. By varying the bias voltage V_{DELAY} , the bandwidth of the pulse can be changed from 500 to 5,000 MHz. Similarly, the voltage V_{OSC} can tune the center frequency from 4.8 to 6.8 GHz. This up-converted triangular pulse transmitter provides a sidelobe rejection of more than 20 dB compared to the main lobe, as shown in Fig. 4.25b.

Table 4.3 summarizes the performance of the two transmitters presented here, and compares their performance to a recent transmitter in the literature. It should be noted that the modulation scheme used in [24] is PPM, which transmits one pulse every cycle, regardless of the input data. For a fair power comparison, the OOK transmitters described here were also set to transmit a pulse every cycle (i.e., the input data was composed of all 1's).

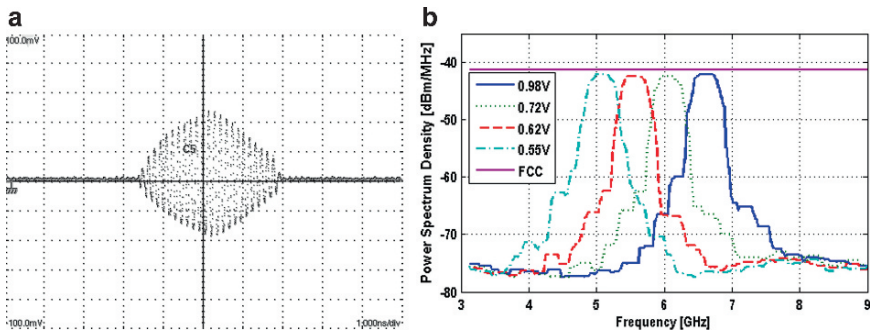


Fig. 4.25 (a) Time and (b) frequency domain measurements of the sub-band pulse transmitter

Table 4.3 Summary of the transmitters presented here along with a recent transmitter from the literature

	Full-band transmitter	Sub-band transmitter	TCAS'05 [24]
Technology	CMOS 0.18 μm	CMOS 0.18 μm	CMOS 0.18 μm
Topology	Pulse generator and pulse shaping filter	Multiplier and pulse shaping filter	Multiplier for triangular pulse, and carrier
Vdd (V)	1.8	1.8	1.8
Modulation	OOK	OOK	PPM
Pulse repetition frequency (MHz)	110	30	40
Pulse bandwidth (MHz)	500–5,000	500–5,000	200–2,000
Power (mW)	0.237	0.255	2
Area (mm^2)	0.57	0.63	0.36
FCC compliant	Yes	Yes	Yes
Other feature	Low power	Low sidelobes	Low sidelobes

4.5 Frequency Synthesis for OFDM UWB

As discussed in the Section 4.2.1, one challenging specification to meet in multi-band OFDM systems is the requirement to hop from one sub-band to another (as shown in Fig. 4.2) within 9.5 ns. The band switching time is determined by the settling time of the frequency synthesizer. Therefore, the main challenge for PLLs in multiband OFDM systems is to achieve very fast band switching times, while still meeting all of the other requirements (e.g. phase noise and spurious tones).

Typical PLLs with a programmable divider in feedback cannot meet the band hopping requirement due to settling times larger than 9.5 ns. One simple approach to achieve fast band hopping is to have a PLL for each frequency sub-band in the multi-band UWB systems (e.g. three PLLs for band group #1, as shown in Fig. 4.26a) [30]. However, this simple approach is costly in term of area and power. Furthermore, the coupling and carrier leakage between PLLs degrade the overall performance.

Another approach [31] uses a single-sideband (SSB) mixer and a multiplexer with two quadrature PLLs to synthesize all of the LO frequencies needed in band group #1. As shown in Fig. 4.26b, two quadrature PLLs lock at start-up and produce fixed frequencies of 3,960 and ± 528 MHz. The 3,960 MHz signal is at the center frequency of sub-band 2, and by quadrature mixing it with the ± 528 MHz signals, the center frequencies of sub-bands 1 and 3 are generated. This work demonstrated a fast band switching of 1 ns with a power consumption of 73.4 mW in a 0.25 μm SiGe BiCMOS process. The drawbacks of this approach are that the SSB mixer requires accurate quadrature inputs, and consumes relatively high power.

For compact and low cost UWB applications, a single PLL with a delay-locked loop (DLL)-based frequency multiplier (shown in Fig. 4.27) can generate all three carrier frequencies in band group #1 within the band hopping requirement [32]. In this frequency synthesizer, a non-switching PLL generates a 528 MHz reference

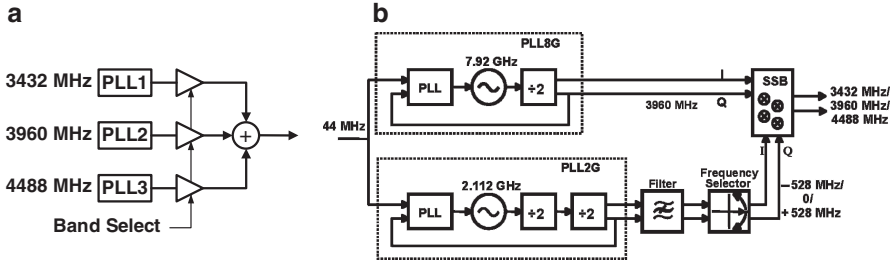


Fig. 4.26 UWB frequency hopping synthesizer architectures using (a) three PLLs [30] and (b) two PLLs with SSB mixing [31]

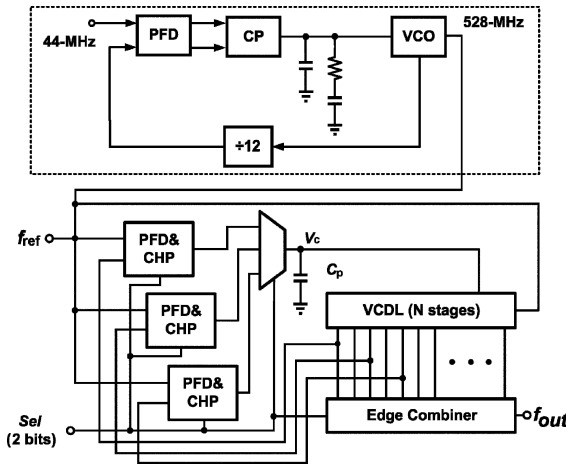


Fig. 4.27 DLL-based frequency hopping synthesizer architecture [32]

frequency for the DLL-based frequency multiplier consisting of a DLL and an edge combiner. The DLL determines the multiplication factor of the reference frequency by switching the feedback clock between three delay blocks. The phase shifted clocks generated by the DLL are then combined together by an edge combiner to produce any of the three carrier frequencies in band group #1. A band switching time of less than 8 ns can be achieved with a power consumption of 54 mW and an area of 0.52 mm² in a 0.18 μm CMOS technology.

For high data rate applications, the frequency synthesizer should cover the entire UWB frequency allocation (i.e., all five band groups). The frequency synthesizer in [33], consisting of two PLLs, three SSB mixers and two multipliers, generates all of the necessary 14 carriers with sufficient sideband rejection. This scheme is shown in Fig. 4.28. One PLL generates the center frequency of 3,960 MHz in group #1, and the second PLL utilizes multiple divider circuits to generate the spacing frequencies of 6,336, 3,168, 1,584, and 528 MHz. Through different combinations (up/down conversions) of the 3,960 MHz signal and the spacing frequencies, all the 14 carriers can be obtained. A quadrature divider circuit in the PLL is used to

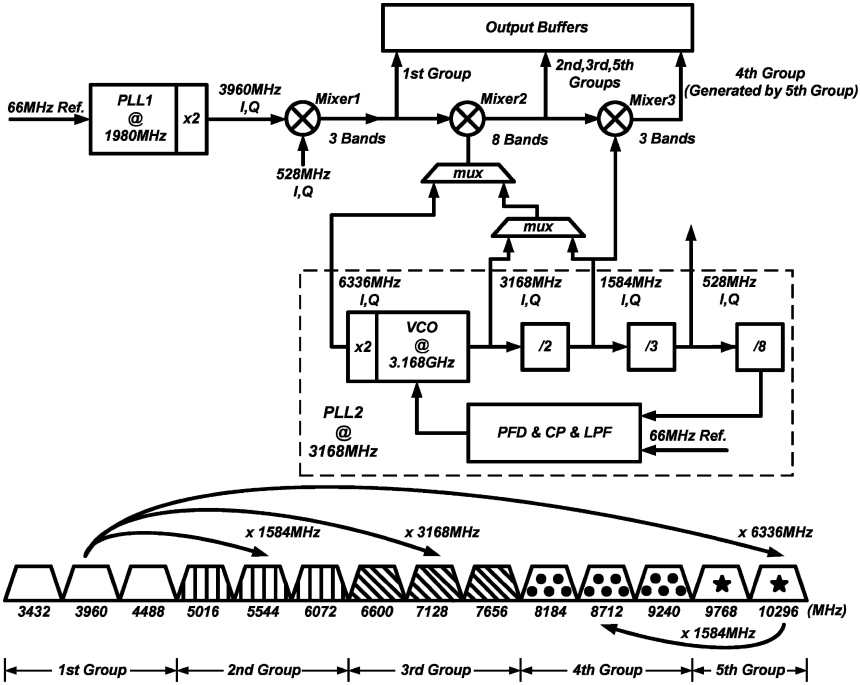


Fig. 4.28 A 14-band frequency synthesizer architecture for MB-OFDM UWB applications [33]

solve distortion and spur issues for the SSB mixers. A more detailed analysis can be found in [33]. This work demonstrated a band switching time of less than 3 ns with power consumption of 162 mW and an area of 1.53 mm² in a 0.18 um CMOS technology.

4.6 Summary

This chapter has provided a brief overview of ultra wideband technology for wireless communications systems. Basic operations and features of the two competing UWB proposals were summarized, along with sample transceiver architectures. Examples of low noise amplifiers and pulse-based transmitters, which are two key building blocks for low power UWB wireless systems, were presented.

A common-gate amplifier topology with local feedback was employed in the presented 3.1–10.6GHz band UWB amplifiers. Robust wideband input impedance matching from 1–10GHz with a flatband gain from 2–7 GHz was achieved with less than 10 mW of power dissipation.

The challenges of designing LNAs in the 0–960MHz UWB frequency range were also addressed. The presented designs focused on minimizing power

consumption while still maintaining reasonable performance. Two common-gate amplifiers with different shunt-feedback elements (transistors and bondwire transformers) were designed, and the power consumption of each circuit was less than 0.7 mW.

Two UWB pulse generators for DS-UWB systems were also presented. The first one covered the entire UWB band for high data-rate communications, while the second one transmitted pulses with a smaller bandwidth for use in low power, low data-rate sub-band systems. Both pulse generator designs met the UWB spectral mask requirements while consuming less than 300 μ W of power.

Finally, an overview of suitable PLL architectures for frequency synthesis in multi-band OFDM UWB systems was presented. These architectures were capable of meeting the stringent 9.5 ns band-hopping specification required by the standard.

References

1. FCC, "Part 15 - Radio Frequency Devices," Available at: <http://www.fcc.gov/oet/info/rules/part15>, July 2008.
2. R. J. Fontana, "Recent System Applications of Short-Pulse Ultra-Wideband (UWB) Technology," *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, pp. 2087–2104, September 2004.
3. T. K. K. Tsang and M. N. El-Gamal, "Ultra-Wideband (UWB) Communications Systems: An Overview," *IEEE Northeast Workshop on Circuits and Systems*, pp. 381–386, 2005.
4. S. B. T. Wang, A. M. Niknejad, and R. W. Brodersen, "Design of a Sub-mW 960-MHz UWB CMOS LNA," *IEEE Journal of Solid-State Circuits*, vol. 41, pp. 2449–2456, November 2006.
5. X. Guan and C. Nguyen, "Low-Power-Consumption and High-Gain CMOS Distributed Amplifiers Using Cascade of Inductively Coupled Common-Source Gain Cells for UWB Systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, pp. 3278–3283, August 2006.
6. A. Safarian, L. Zhou, and P. Heydari, "A Distributed RF Front-End for UWB Receivers," *IEEE Custom Integrated Circuits Conference*, pp. 805–808, September 2006.
7. H.-T. Ahn and D. J. Allstot, "A 0.5–8.5 GHz Fully Differential CMOS Distributed Amplifier," *IEEE Journal of Solid-State Circuits*, vol. 37, pp. 985–993, August 2002.
8. P. Heydari, D. Lin, A. Shameli, and A. Yazdi, "Design of CMOS Distributed Circuits for Multi-band UWB Wireless Receiver [LNA and Mixer]," *IEEE Radio Frequency Integrated Circuits Symposium*, pp. 695–698, June 2005.
9. A. Bevilacqua and A. M. Niknejad, "An Ultra-Wideband CMOS LNA for 3.1–10.6-GHz Wireless Receivers," *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 2259–2268, December 2004.
10. A. Ismail and A. A. Abidi, "A 3–10-GHz Low Noise Amplifier with Wideband LC-Ladder Matching Network," *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 2269–2277, December 2004.
11. F. Bruccoleri, E. A. M. Klumperink, and B. Nauta, "Generating All Two-MOS-Transistor Amplifiers Leads to New Wide-Band LNAs," *IEEE Journal of Solid-State Circuits*, vol. 36, pp. 1032–1040, July 2001.
12. T. K. K. Tsang, K.-Y. Lin, and M. N. El-Gamal, "Design Techniques of CMOS Ultra-Wideband Amplifiers for Multi-Standard Communications," *IEEE Transactions on Circuits and Systems II*, Accepted for Publication, 2008.
13. J. C. Zhan and S. S. Taylor, "A 5GHz Resistive-Feedback CMOS LNA for Low-Cost Multi-Standard Applications," *IEEE International Solid-State Circuits Conference*, pp. 721–730, February 2006.

14. M. T. Reiha and J. Long, "A 1.2V Reactive-Feedback 3.1–10.6 GHz Low-Noise Amplifier in 0.13 μm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 42, pp. 1023–1033, May 2007.
15. X. Li, S. Shekhar, and D. J. Allstot, "Gm-Boosted Common-Gate LNA and Differential Colpitts VCO/QVCO in 0.18- μm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 40, pp. 2609–2619, December 2005.
16. K. Allidina and M. N. El-Gamal, "A 1V CMOS LNA for Low Power Ultra-Wideband Systems," *IEEE International Conference on Electronics, Circuits, and Systems*, pp. 165–168, 2008.
17. K. Bhatia, S. Hyvonen, and E. Rosenbaum, "An 8-mW, ESD-protected, CMOS LNA for Ultra-Wideband Applications," *IEEE Custom Integrated Circuits Conference*, pp. 385–388, September 2006.
18. S. Mohan, M. Hershenson, S. Boyd, and T. Lee, "Bandwidth Extension in CMOS with Optimized On-Chip Inductors," *IEEE Journal of Solid-State Circuits*, vol. 35, pp. 346–355, March 2000.
19. K.-H. Chen, J.-H. Lu, B.-J. Chen, and S.-I. Liu, "An Ultra-Wide-Band 0.4–10-GHz LNA in 0.18- μm CMOS," *IEEE Transactions on Circuits and Systems II*, vol. 54, pp. 217–221, March 2007.
20. R. G. Carvajal, J. Ramirez-Angulo, A. J. Lopez-Martin, A. Torralba, J. A. G. Galan, A. Carlosena, and F. M. Chavero, "The Flipped Voltage Follower: A Useful Cell for Low-Voltage Low-Power Circuit Design," *IEEE Journal of Solid-State Circuits*, vol. 52, pp. 1276–1291, July 2005.
21. K. Allidina and S. Mirabbasi, "A Widely Tunable Active RF Filter Topology," *IEEE International Symposium on Circuits and Systems*, pp. 879–882, May 2006.
22. K.-Y. Lin and M. N. El-Gamal, "Performance and Modeling of Bonding Wire Transformers in a Package for RF ICs," *IEEE Proceedings of International Conference Microelectronics*, pp. 347–350, December 2007.
23. M. Vidojkovic, M. Sanduleanu, M. v. d. Tang, P. Baltus, and A. v. Roermund, "A 1.2V, Inductorless, Broadband LNA in 90 nm CMOS LP," *IEEE Radio Frequency Integrated Circuits Symposium*, pp. 53–56, 2007.
24. J. Ryckaert, C. Desset, A. Fort, M. Badaroglu, V. D. Heyn, P. Wambacq, G. V. d. Plas, and S. Donnay, "Ultra-Wide-Band Transmitter for Low-Power Wireless Body Area Networks: Design and Evaluation," *IEEE Transactions on Circuits and Systems I*, vol. 52, pp. 2515–2525, December 2005.
25. K. W. Wong, S. R. Karri, and Y. Zheng, "Low-Power Full-Band UWB Active Pulse Shaping Circuit Using 0.18- μm CMOS Technology," *IEEE Radio Frequency Integrated Circuits Symposium*, June 2006.
26. T. K. K. Tsang and M. N. El-Gamal, "Fully Integrated Sub-MicroWatt CMOS Ultra Wideband Pulse-Based Transmitter for Wireless Sensors Networks," *IEEE Proceedings of International Symposium on Circuits and Systems*, pp. 670–673, May 2006.
27. T. Norimatsu, R. Fujiwara, M. Kokubo, M. Miyazaki, Y. Ookuma, M. Hayakawa, S. Kobayashi, N. Koshizuka, and K. Sakamura, "A Novel UWB Impulse-Radio Transmitter with All-Digitally-Controlled Pulse Generator," *IEEE Proceedings of International Solid-State Circuits Conference*, pp. 267–270, September 2005.
28. H. Sheng, O. Orlik, A. M. Haimovich, L. J. Cimini, and J. Zhang, "On the Spectral and Power Requirements for Ultra-wideband Transmission," *Proceedings of IEEE Conference on Communications*, pp. 738–742, May 2003.
29. K.-Y. Lin and M. N. El-Gamal, "Design of Low Power CMOS Ultra-Wideband 3.1–10.6 GHz Pulse-Based Transmitters," *IEEE Custom Integrated Circuits Conference*, September 2008.
30. B. Razavi, T. Aytur, C. Lam, F. Yang, K. Li, R. Yan, H. Kang, C. Hsu, and C. Lee, "A UWB CMOS Transceiver," *IEEE Journal of Solid-State Circuits*, vol. 40, pp. 2555–2562, December 2005.

31. D. Leenaerts, R. V. D. Beek, J. Bergervoet, K. S. Harish, H. Waite, Y. Zhang, C. Razzell, and R. Roovers, "A SiGe BiCMOS 1ns Fast Frequency Hopping Synthesizer for UWB Radio," *IEEE International Solid-State Circuits Conference*, pp. 202–203, February 2005.
32. T. Lee and K. Hsiao, "The Design and Analysis of a DLL-based Frequency Synthesizer for UWB Applications," *IEEE Journal of Solid-State Circuits*, vol. 41, pp. 1245–1252, June 2006.
33. C. Liang, S. Liu, Y. Chen, T. Yang, and G. Ma, "A 14-band Frequency Synthesizer for MB-OFDM UWB Application," *IEEE International Solid-State Circuits Conference*, pp. 428–429, February 2006.

Chapter 5

CMOS IR-UWB Transceiver System Design for Contact-Less Chip Testing Applications

Yanjie Wang, Ali M. Niknejad, Vincent Gaudet, and Kris Iniewski

5.1 Introduction

Today's semiconductor products are more complex and highly integrated due to the increasing demands for system-on-chip solution, which results in a significant increase in time, cost and complexity of testing. Testing of semiconductors has become a significant and growing problem in the very-large-scale-integration (VLSI) circuit manufacturing industry [1, 2]. Semiconductor testing issues such as smaller pad size, increased pad density, increased signal input/output (I/O) frequencies, longer test times, and probe card contact and alignment are restricting the progress towards smaller, faster, and more economical integrated circuits [3].

Conventional wafer probing techniques utilize probe tips to contact the device-under-test (DUT) physically and have the limitations of the number of pads, pitch sizes, operating frequency, parallel testing capability and risk of damage to the DUT. Moreover, the calibration of probe tips and silicon substrate especially for high-speed RF circuits have made testing more complicated and may affect the accuracy of testing.

The steady downscaling of semiconductor device dimensions has become the main stimulus to achieve higher speed and performance in integrated circuits. However, scaling may increase the delay associated with the parasitic resistance,

Y. Wang (✉) and A.M. Niknejad
Berkeley Wireless Research Center, Department of Electrical Engineering and Computer Sciences,
University of California at Berkeley, Berkeley, CA 94720, USA
e-mail: {yanjie; niknejad}@eecs.berkeley.edu

Y. Wang and V. Gaudet
Department of Electrical and Computer Engineering, University of Alberta, Edmonton,
AB T6G 2V4, Canada
e-mail: {yanjie; vgaudet}@ece.ualberta.ca

K. Iniewski
CMOS Emerging Technologies Inc., 2865 Stanley Pl. Coquitlam, BC V3B 7L7, Canada
e-mail: kri.iniewski@gmail.com

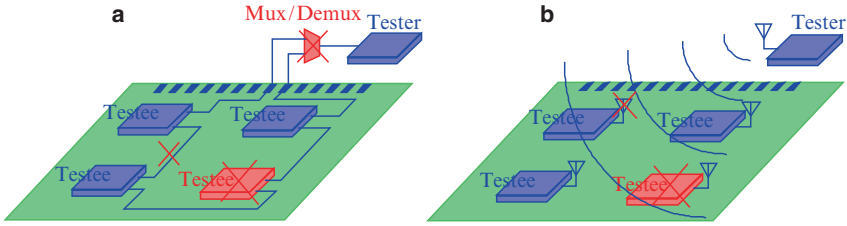


Fig. 5.1 Diagram of testing methodologies (a) wires and (b) wireless

capacitance and inductance of wiring interconnects, and those parasitics have become the main obstacle for high-speed data transmission [1].

Wireless, contact-less chip testing structures alleviate those dependencies in conventional wired testing structures, and offer many advantages, such as higher operating frequencies, parallel testing capability, increased throughput, decreased testing time [2], and most importantly, the testee and tester are completely decoupled offering little risk of damage to the DUT during testing process. Figure 5.1a gives an example of a scan path such as Joint Test Action Group (JTAG) [4] using a traditional wired testing structure that connects components serially in the form of a daisy chain. It shows that a fault of a shared component or wiring would affect the accuracy of the overall test results even if other components are properly working, while in the wireless testing structure as shown in Fig. 5.1b, the testing is conducted in parallel, and the faulty components and other properly working components are independent from each other [5].

In this chapter, we present an impulse-radio ultra-wideband (IR-UWB) transceiver system and some circuits for future contact-less chip testing applications using magnetic coupling transformers as wireless transmission interconnects. The chapter is organized as follows, Section 5.2 overviews the wireless links and introduces the ultra-wideband technology. Section 5.3 illustrates a proposed impulse-based UWB architecture and transmitter circuits. Section 5.4 discusses wireless data transmission using on-chip magnetic coupling transformers and Section 5.5 describes two Low Noise Amplifier (LNA) topologies for UWB receivers. Simulation results and analysis are also discussed, followed by the conclusion in Section 5.6.

5.2 Wireless Links and UWB Wireless Overview

Wireless interconnects that form the essential components of wireless testing systems can be realized by on-chip antennas or magnetic coupling transformers. Figure 5.2 shows the conceptual diagram of a wireless interface for sensor networks [6] and body area networks [3] respectively. Integration of on-chip antennas eliminates the need for external transmission line connections and sophisticated packaging, which can dramatically reduce the cost of a test system operating at a high speed. In [6], a 15-GHz sinusoid wave is transmitted from an on-chip

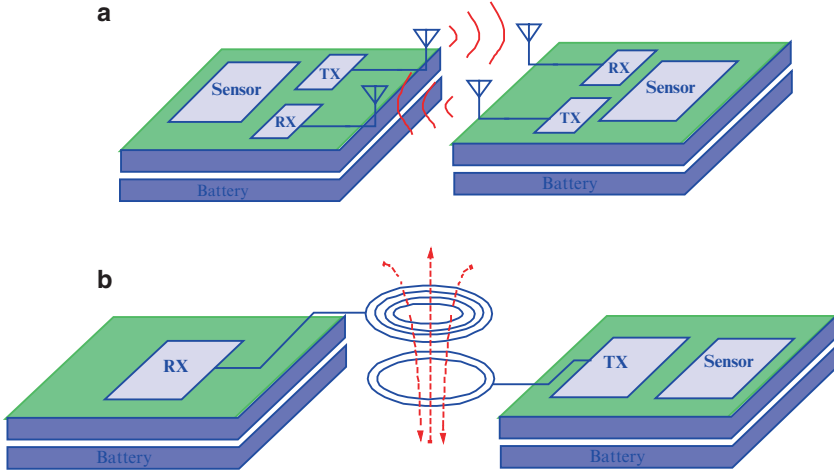


Fig. 5.2 Conceptual diagram of wireless chip communication link using (a) on-chip antenna, (b) magnetic coupling transformer

2 mm-long zigzag dipole antenna and picked up by a receiver located 2.2 cm away. However, for chip testing application, the relatively small magnetic coupling transformer is preferred as the area is more critical than transmission range and power consumption. The idea of having a simple wireless interface based on a magnetic coupling transformer was originally motivated by the principle of Radio Frequency Identifier (RFID) technology and body area networks [3, 7].

Such a wireless interconnect testing system as shown in Fig. 5.2 requires a transmitter, receiver and on-chip antenna/transformer. For high data transmission rates and multiple access capability, this wireless testing system requires wideband characteristics in the integrated transmitter, receiver and antenna. According to the channel capacity theorem, the data transmission rate or channel capacity grows linearly with channel bandwidth and logarithmically with signal to noise ratio. This is clearly seen in Shannon's equation for the channel capacity [8]:

$$C = BW \cdot \log \left(1 + \frac{S}{N} \right) \quad (5.1)$$

where C is the channel capacity, BW is the bandwidth in Hz, S is the signal power spectral density (PSD) in W/Hz and N is the noise single-sided PSD in W/Hz. Thus, an ultra-wideband system appears to have a great potential for implementation of wireless testing system for future VLSI.

Due to the demanding performance requirements for fourth-generation wireless systems, UWB systems are particularly attractive due to their high data rate, low power and robustness to harsh multipath environments. From a communications theory perspective, a UWB system's capability of operating in power-limited regimes is the most important characteristic. From Eq. (5.1), channel capacity is linear to channel bandwidth and logarithmic to signal to noise ratio. Therefore, for a UWB wireless network, the system can operate at very low signal to noise ratios (SNR)

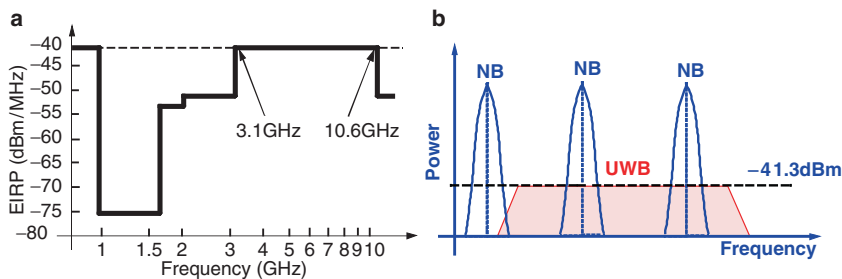


Fig. 5.3 (a) FCC spectral mask, (b) UWB power spectrum vs narrow band

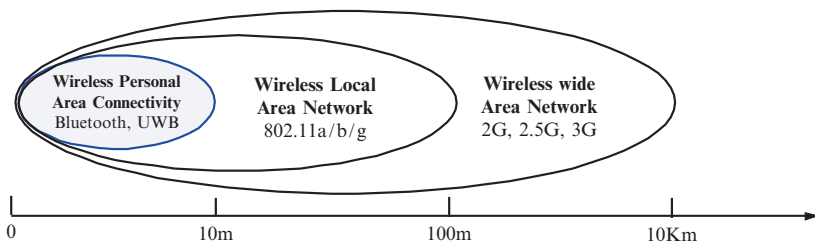


Fig. 5.4 Wireless data communications coverage range

with high bandwidth. This means that a UWB wireless network is able to achieve high data rates with relatively low transmit power.

In February 2002, the Federal Communications Commission (FCC) issued a report and giving users permission to deploy low power UWB systems within the 3.1–10.6 GHz spectrum [9] as shown in Fig. 5.3a. Such a wide bandwidth (7.5 GHz) results in a low power spectral density, making UWB operating at very low power (−41.3 dBm) for short ranges (30 m or less), without interfering with the spectrum of existing radio frequency (RF) systems as shown in Fig. 5.3b [10]. The application and coverage range of UWB is presented in Fig. 5.4, UWB communication standard provides opportunities for meeting the increasing channel capacity demands of the wireless world and becomes the suitable candidate for low-power and short-distance wireless communication system such as wireless personal area networks (WPANs) as well as short range wireless chip testing.

Aside from low power consumption and reliability, cost is also important for these applications. Full integration of digital and RF circuitry on a single die had been desired to reduce cost, and over the past few years this has become a reality. Owing to recent speed improvements in the standard complementary MOSFET (CMOS) technology, especially for new nanometer CMOS technologies (e.g. 180, 90 nm CMOS processes), highly integrated System-On-Chip (SOC), with both RF transceivers and complex digital functions on the same CMOS die, can be readily acquired on the semiconductor market. As wireless usage proliferates, the cost pressures will continue to steer designers to use CMOS for UWB circuit implementations.

5.2.1 Design Approaches for UWB

The UWB definition by FCC has given system designers the opportunity to employ two approaches to UWB system design: (1) impulse-based (I-UWB) [11], (2) multicarrier-based (MC-UWB) [12]. I-UWB proponents claim less complex hardware implementation, less demanding digital processing, more resilience to multipath fading effects [13]. The relative simplicity of this implementation compared to the conventional narrow-band superheterodyne receivers manifests itself by low power consumption. The MC-UWB proponents claim greater spectral efficiency and flexibility, more efficient energy capture, and easier coexistence. And it is improved through adaptive band selection from well understood schemes such as multi-band code division multiple access (CDMA) and orthogonal frequency-division multiplexing (OFDM) [13]. In following sections I-UWB and MC-UWB are described in more technical detail.

5.2.1.1 Impulse-Based UWB

The impulse-based UWB communication is established by modulating an impulse-like waveform with sharp rise/fall times and wide bandwidth. Typically, the pulses last for hundreds of picoseconds and have a bandwidth of several gigahertz. The Gaussian monocycle has been widely used due to its mathematical tractability and good approximation to actual measurement [13, 14]. The pulse name implies the number of zero-crossing points in the time domain, e.g. Gaussian pulses do not cross the x-axis, the Gaussian monocycle crosses the x-axis once. Figure 5.5 shows time domain response and frequency domain power spectral density representations of Gaussian monocycle pulses. A train of Gaussian pulses without modulation causes discrete spectral lines with high peak power [15]. The lines appear as narrowband interferers, so some spectral smoothing is necessary. The choice of a modulation

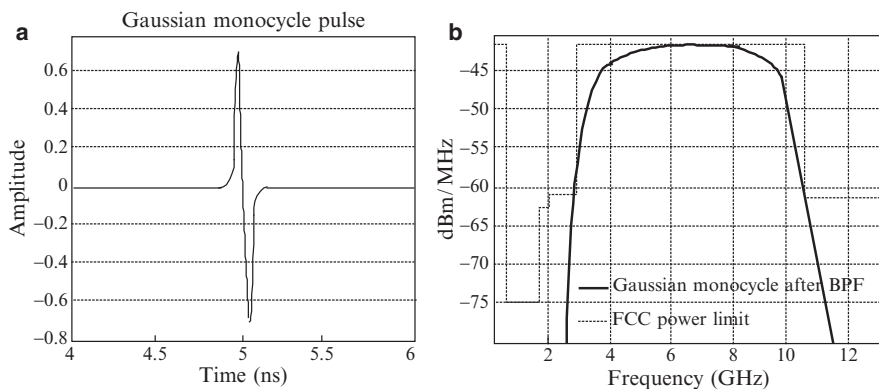


Fig. 5.5 Gaussian monocycle pulse (a) pulse shape (b) power spectral density [13]

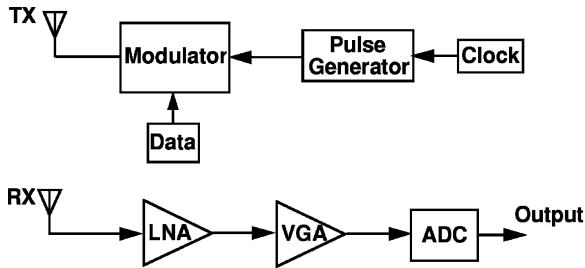


Fig. 5.6 Block diagram of the impulse-radio UWB transceiver [16]

scheme affects the amount of smoothing. Some possible modulation schemes are pulse amplitude modulation (PAM), pulse position modulation (PPM), on-off keying (OOK), and binary phase shift keying (BPSK) [13].

Figure 5.6 shows a prototypical impulse-radio UWB transceiver [16], where the pulse has sharp response and its shape is designed to concentrate energy over the broad range of 3.1–10.6 GHz. Because the output power level of UWB transmitter is limited to -41 dBm/MHz, the pulse generator needs only produce a voltage swing on the order of 100 mV. Therefore, an important attribute of the impulse-based transmitter is that the power amplifier is not required. The pulse signals are often filtered by a bandpass filter (BPF) before sending to antenna to more efficiently meet frequency band regulatory limits. In addition, an I-UWB system may use notch filtering in order to prevent overloading of the receiver from narrowband interference and also decrease interference to narrowband systems.

In an impulse-radio UWB receiver, the main function of the low noise amplifier is to achieve input matching to impedance of antenna ($50\ \Omega$) for noise optimization and in order to filter out-of-band interferers. In addition, it must show flat gain over the entire bandwidth, minimum possible noise figure (NF) and low power consumption. The variable-gain control amplifier (VGA) stabilizes the signal to a constant gain level. The analog-to-digital converter (ADC) transfers signal processing to the digital domain and recovers the information data for baseband digital signal processing.

The design challenges for impulse-radio transceiver are: (1) impulse generator with sharp rise/fall times, which occupies several GHz of bandwidth, (2) low noise amplifier that operates in the whole 7.5 GHz frequency band, (3) analog to digital converter (ADC) with high sampling rate and low power consumption.

5.2.1.2 Multicarrier-Based UWB

Figure 5.7 shows the pulse shape and spectral density of a multicarrier UWB system. In an MC-UWB transceiver, the whole 3.1–10.6-GHz bandwidth is split into 14 of 528-MHz sub-bands, which are also grouped into five channels; Channel 1 (3.1–4.8 GHz), and Channel 2,3,4,5 (4.8–10.6 GHz), as depicted in Fig. 5.8. Either

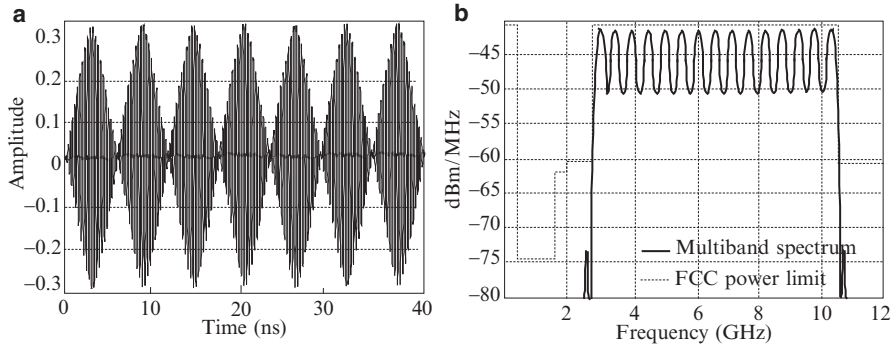


Fig. 5.7 The multicarrier UWB signal (a) pulse shape (b) power spectral density

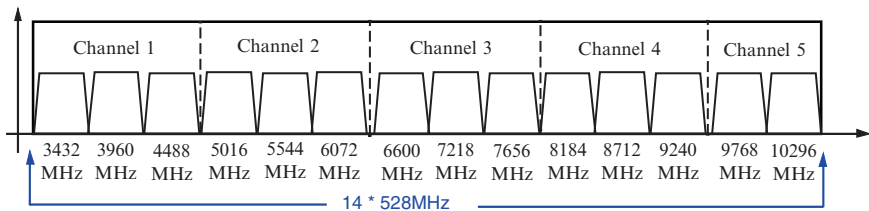


Fig. 5.8 The multiband UWB frequency band plan

single-carrier or multi-carrier modulation may be employed in each sub-band. Single-carrier modulation facilitates the design of inexpensive transmitters, however, at the cost of more complicated receivers. Multi-carrier modulation, also known as OFDM, widely used in the implementation of IEEE 802.11.a/g standards, [16] performs well in dispersive channels, and enables high-rate communication with inexpensive low-power receivers.

The multi-band approach accommodates base-band processing over smaller bandwidths (528 MHz), thereby relaxing the design constraints on the key components of the UWB transceiver, mostly the data conversion modules. Conventional circuit techniques can be employed to implement the data conversion circuits [17]. The ADC is now digitizing the 528-MHz down-converted signal. Designing a power-efficient flash ADC with a sampling rate of 1.1G samples/s is quite achievable in standard CMOS processes [17]. In spite of simplifying the ADC design, the receiver front-end LNA/mixer still entails many design challenges. On the transmitter side, the efficient power amplifier is the main challenge for low power systems.

Based on the advantages and disadvantages between the two design approaches, the impulse-based UWB system is chosen for our application with requirements of a ultra wide bandwidth, a flat gain, a low noise figure, good input/output matching, low circuit complexity and power consumption.

5.3 IR-UWB Transceiver Systems Architecture for Short-Range Wireless Chip Testing

Figure 5.9 shows the proposed wireless non-contact chip testing methodology using magnetic coupling transformers for wireless interconnects. There are two advantages of this method: Firstly, the wired DC power delivery provides sufficient power for high power applications ($>\mu\text{W}$) and saves significant chip size than using extra circuitry to deliver DC power supply wirelessly. Secondly, there is no impact on testing circuits performance as the transceiver circuits are disconnected from the application circuits when the chip is in normal operating mode.

Figure 5.10 depicts the block diagram of the proposed IR-UWB transceiver testing system. The transmitter consists of a base-band modulation unit, an impulse

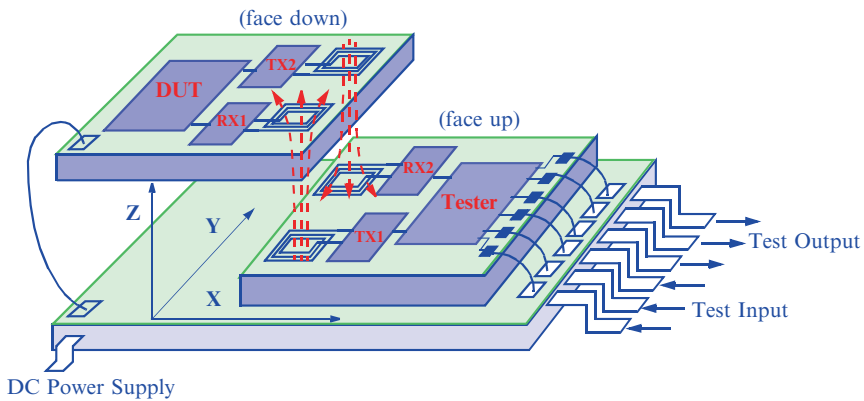


Fig. 5.9 3-D view of the proposed wireless non-contact testing system with wired DC power delivery

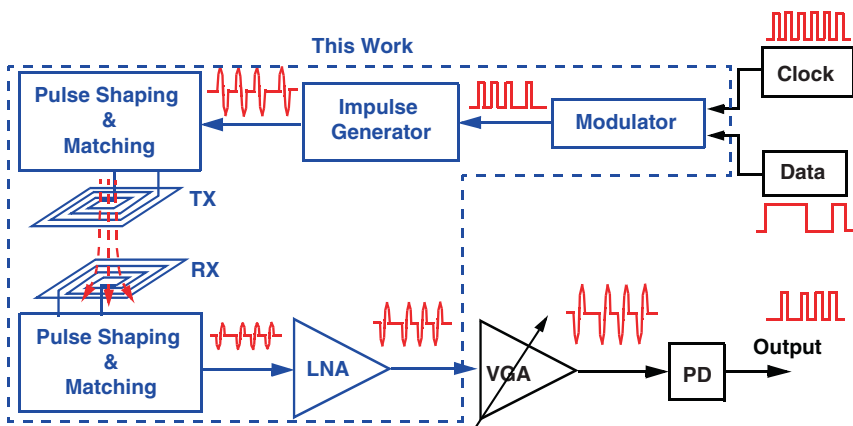


Fig. 5.10 Block diagram of the proposed UWB wireless testing transceiver

generator, and impedance matching to an on-chip transmitter coil. The receiver is formed by the LNA, VGA and peak detection. This architecture is the first reported system for wireless chip testing applications and it offers several advantages:

1. Low power transmission – A power amplifier is not required for the TX side as the transmitted power level is as low as -41 dBm/MHz. The pulse generator needs only to produce a voltage swing on the orders of 100 mV [16].
2. Lower circuit complexity – The impulse-based transmitter has lower circuit complexity and lower power dissipation compared to conventional narrowband and OFDM systems, which saves energy and chip space.
3. Matching flexibility – The input impedance matching for an integrated LNA is more flexible, so it is not necessary to match input to 50Ω . The matching depends on the input impedance of TX and output impedance of RX coils.

Several design challenges for impulse-based transceivers must be considered such as: (a) an impulse generator should generate sharp rise/fall times and occupies 7.5 GHz of bandwidth; (b) impulse shape has to the FCC spectrum mask which operates in the whole 7.5 GHz frequency band, should have flat gain over the entire bandwidth and good input and output impedance matching to the magnetic coupling transformer and the following stage respectively; (d) a peak detector with high speed operation and low power consumption is required. In the following sections, we will only focus on the RF front-end components shown in the dashed box of Fig. 5.10.

5.4 IR-UWB Transmitter Designs

There are several modulation schemes for data transmission, which include pulse amplitude modulation, pulse position modulation, binary phase shift keying and on-off keying [13]. In this design, the OOK modulation scheme is chosen as it can be easily realized by a simple D-latch device. Figure 5.11 shows the block diagram of the proposed impulse-based UWB transmitter using OOK modulation.

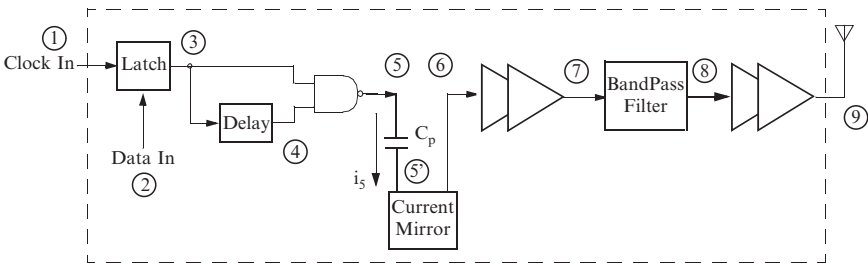


Fig. 5.11 Block diagram of the impulse-based UWB transmitter

5.4.1 Impulse Generation

In this section, we propose simple and robust design methods for Gaussian monocycle impulse generators in UWB transmitters. The idea is based on the fundamental theory on capacitive behavior to generate the first derivative of the Gaussian pulse. Simulation results will be presented using TSMC 180 nm CMOS technology.

The desired impulse shape must be determined before generation. The Gaussian monocycle has been chosen for this design due to its mathematical tractability and good approximation to actual measurement [13, 14]. To reduce the circuit complexity, the OOK modulator is placed in front of the pulse generator [18]. The block diagram and schematic of the proposed impulse generator are illustrated in Figs. 5.11 and 5.12 respectively. It consists of a D-latch, a delay element, a NAND gate, an on-chip capacitor and a current mirror. The data pattern is set to “0110111” in this case. The clock and data signals are modulated by a D-latch and are split into two delay element branches, namely Branch 1 (node 3–3d) and Branch 2 (node 3–4). In Branch 1, the modulated signal is delayed using two inverters, whereas it is delayed and inverted using three inverters in Branch 2. This causes the signal in Branch 2 to be inverted when in parallel with the signal in Branch 1. The NAND gate combines the rising edge of signal from node 3d and falling edge from node 4 to form a triangular pulse signal to approximate an impulse-like waveform.

The duration of the generated impulse is determined by the delay between the rising and falling edges. The load capacitor $C_d = C_{ds1} + C_{gd1} + C_{gs1}$, which is presented by the capacitors of NMOS transistor M_1 as indicated in Fig. 5.12, is used to control this delay. The delay depends on the time used by the current to charge and discharge the load capacitance. With different load capacitance values (namely the different width of transistor M_1), the current charging/discharging time can be varied, consequently, changing the delay, and producing the desired impulse. Moreover, the control voltage V_c controls the charging and discharging current to the parasitic capacitor C_d . Figure 5.13 displays different impulse widths with various width of transistor M_1 at fixed control voltage $V_c = 0.8$ V and Fig. 5.14 shows the impulse width with changes of control voltage V_c .

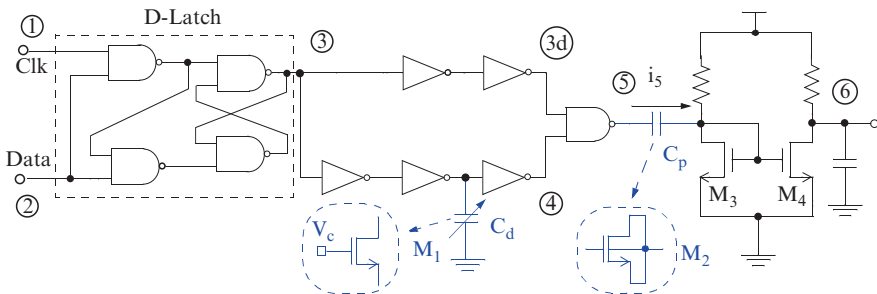


Fig. 5.12 Schematic of the impulse generator

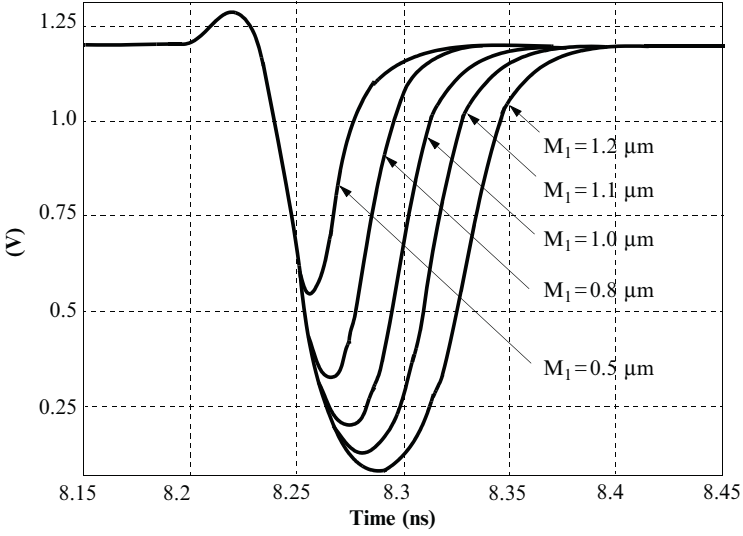


Fig. 5.13 Simulated pulse widths with $V_c = 0.8\text{ V}$ and various M_1 widths

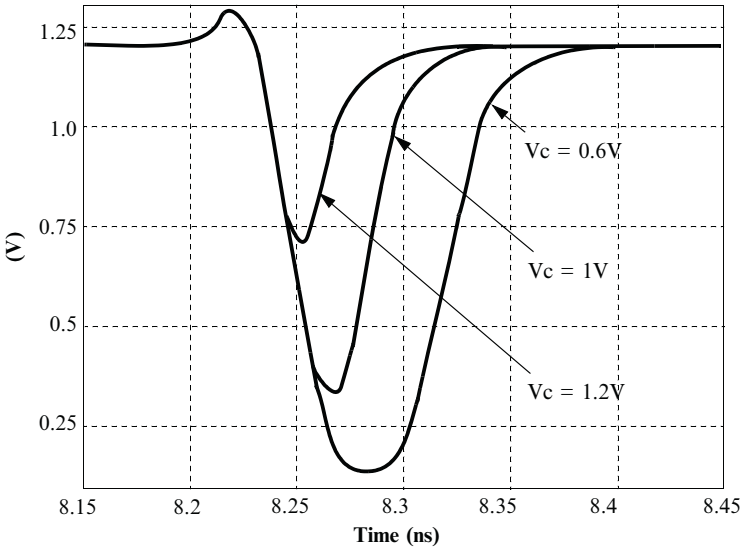


Fig. 5.14 Simulated pulse widths with various V_c and fixed M_1 widths

The output of the NAND gate (node 5) is a train of inverted voltage triangular pulses that can be approximated by:

$$V_5(t) = A \cdot e^{-(2/t)^2} \tag{5.2}$$

where A is the voltage amplitude and τ is the pulse shape parameter. The resulting voltage pulse is applied across an on-chip capacitor C_p , which produces a current according to the simple and robust first-derivative capacitive current–voltage relationship, which is given by:

$$i_5 = C_p \left(\frac{dv_5}{dt} \right) = C_p \left(-\frac{2At}{\tau^2} \right) e^{-(t/\tau)^2} \tag{5.3}$$

The current flowing through the capacitor is copied using a current mirror. The output voltage taken across the load (Z_L) of the current mirror represents the first derivative of the Gaussian pulse, which is similar to the Gaussian monocycle and is approximately given by:

$$V_6(t) = Z_L C_p \left(-\frac{2At}{\tau^2} \right) e^{-(t/\tau)^2} \tag{5.4}$$

Figure 5.15 illustrates the timing waveform at each node of the impulse generator in Fig. 5.12. From Eqs. (5.3) and (5.4), the bigger the C_p (namely larger size of

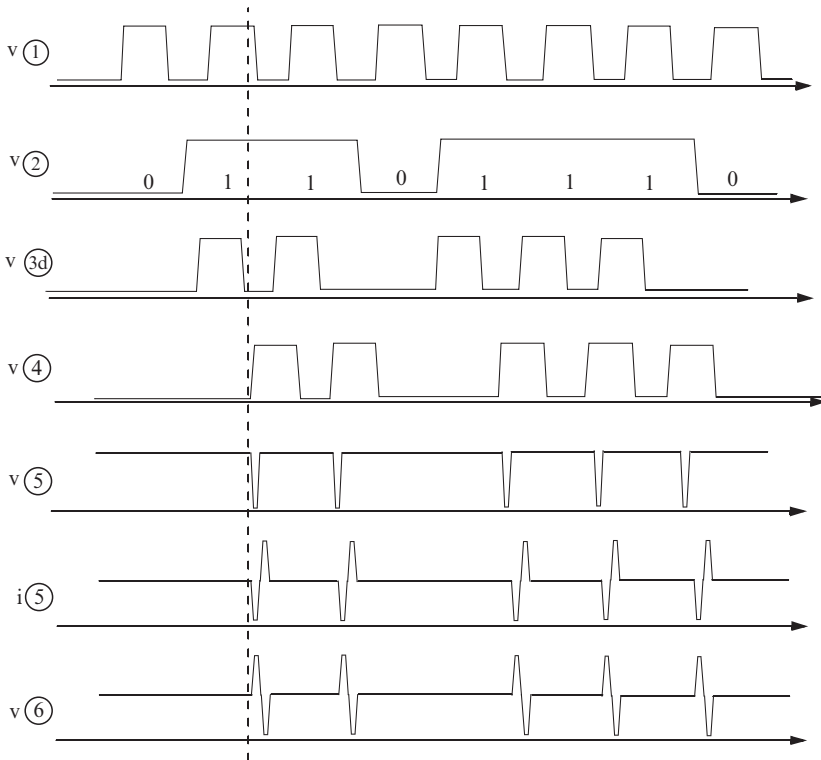


Fig. 5.15 Timing diagram of Fig. 5.12 from node 1 to node 6

M_2), the higher amplitude of V_6 can be obtained, but, there are trade-offs among the output amplitude, settling time, and impulse waveform symmetry. Figure 5.16 shows the output waveforms at node 6 with different width of transistor M_2 (1, 5, 10 μm). A value of 5 μm is chosen for the design with the consideration of the trade-offs described previously.

5.4.2 Impulse Shaping Using LC BPF

As shown in Fig. 5.17, the spectrum of the generated impulse monocycle does not meet the frequency mask of the FCC. In order to meet the FCC spectral mask at the output of the antenna, the pulses generated by the current mirror have to be shaped. A third-order π -type Chebychev LC bandpass filter is designed to filter out all the frequency components except the frequencies ranging from 3.1 to 10.6 GHz. A third order filter is designed in order to have a sharp transition band which makes efficient use of the FCC frequency spectrum mask [19]. Figure 5.18a shows the schematic of the BPF with the component values given in Table 5.1.

The input and output termination of the BPF is designed to be 50 Ω . In order to match the input and output impedance of the filter, a source follower is connected to the input of the filter with a source degeneration resistance of 50 Ω . In addition, a 50 Ω load resistor is added between the filter and the output amplifier thereby

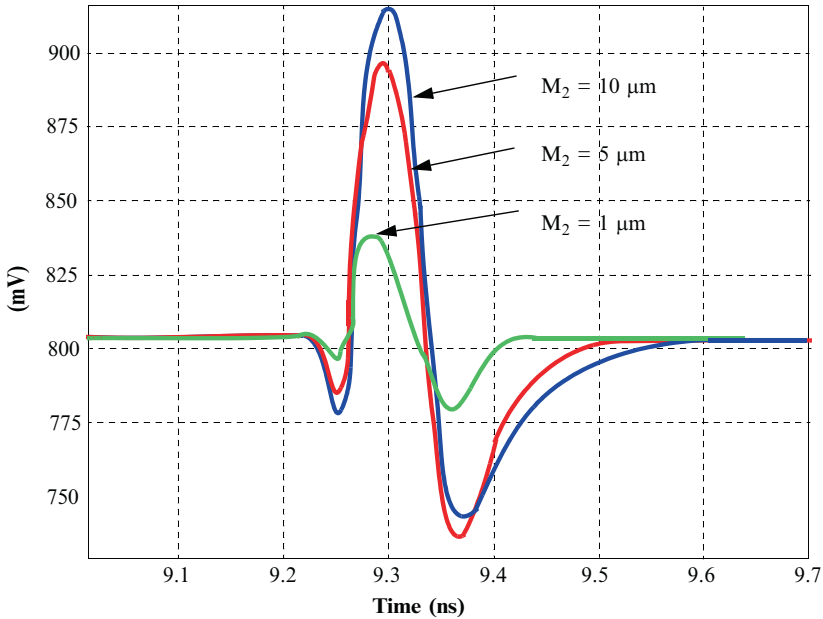


Fig. 5.16 Simulated pulse widths with various M_2 widths at $V_c = 1\text{ V}$

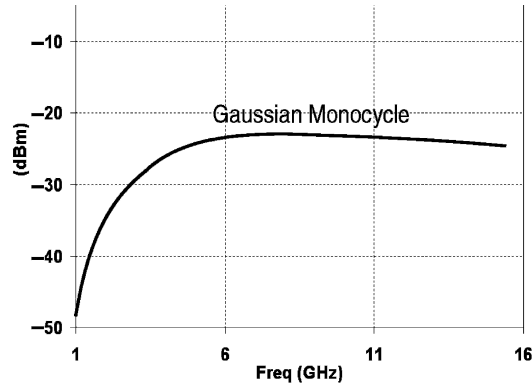


Fig. 5.17 Simulated PSD at the output of the impulse generator

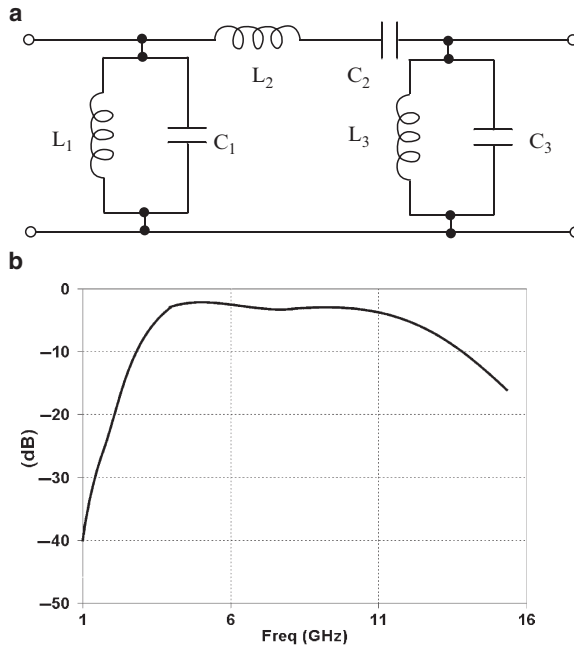


Fig. 5.18 LC BPF schematic (a) and corresponding transfer function (b)

Table 5.1 Third order Chebychev LC BPF components

Type	L_1	C_1	L_2	C_2	L_3	C_3
π -Chebychev	1.3 nH	470 fF	1.3 nH	460 fF	1.3 nH	470 fF

allowing maximum power transmission efficiency between the input and output of the BPF. Figure 5.18b shows the corresponding simulated transfer function of the BPF after adding the input and output impedance matching network described

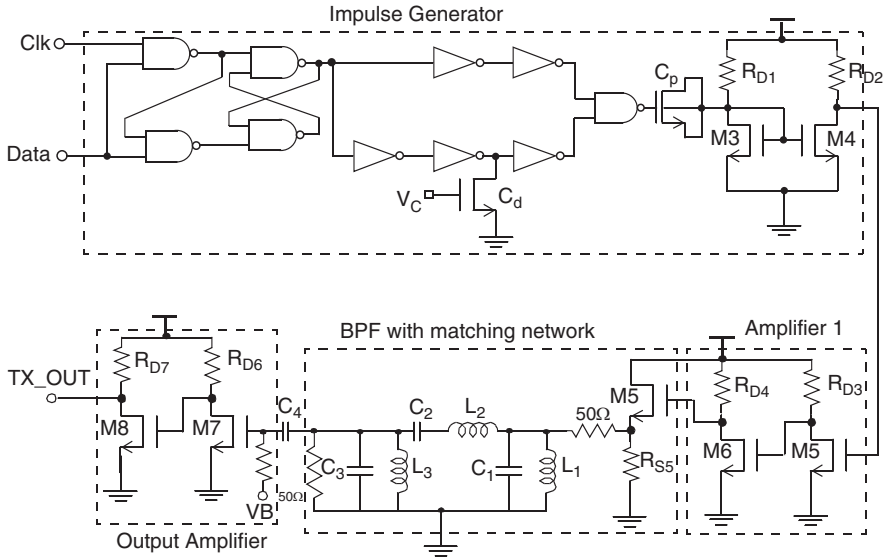


Fig. 5.19 Complete schematic of the UWB transmitter

above. The complete schematic of the UWB transmitter with impulse shaping circuit is shown in Fig. 5.19. The Gaussian monocycle generated at the output of the current mirror is amplified using a two-stage common-source amplifier. This amplifier uses a simple resistive load without any inductive shunt peaking, saving chip area considerably. The amplified Gaussian monocycle pulse is then shaped using an LC BPF.

The input and output termination of the BPF is designed to be $50\ \Omega$. In order to match the input and output impedance of the filter, a source follower is connected to the input of the filter with a source degeneration resistance of $50\ \Omega$. In addition, a $50\ \Omega$ load resistor is added between the filter and the output amplifier thereby allowing maximum power transmission efficiency between the input and output of the BPF. Figure 5.18b shows the corresponding simulated transfer function of the BPF after adding the input and output impedance matching network. The FCC matched pulse is further amplified using an output amplifier as a buffer before being sent to an antenna for transmission. The transmitter is laid out using a standard 180 nm CMOS technology. Post-layout simulations are conducted to validate the circuit performance.

Figure 5.20a shows the transient response of the impulse generator in Figure 5.12 from node 1 to node 6, respectively. The clock input signal is set to 500 MHz and a data pattern of “0110111” is used at the rate of 500 Mb/s. The voltage signal represents the first derivative of the Gaussian pulse with an amplitude of $130\ mV_{pp}$ and a rising/falling time of less than 100 ps as shown in Fig. 5.20b.

The Gaussian monocycle pulse is shaped using an LC BPF in order to meet the FCC spectrum specifications, which is optional for wireless testing applications. The

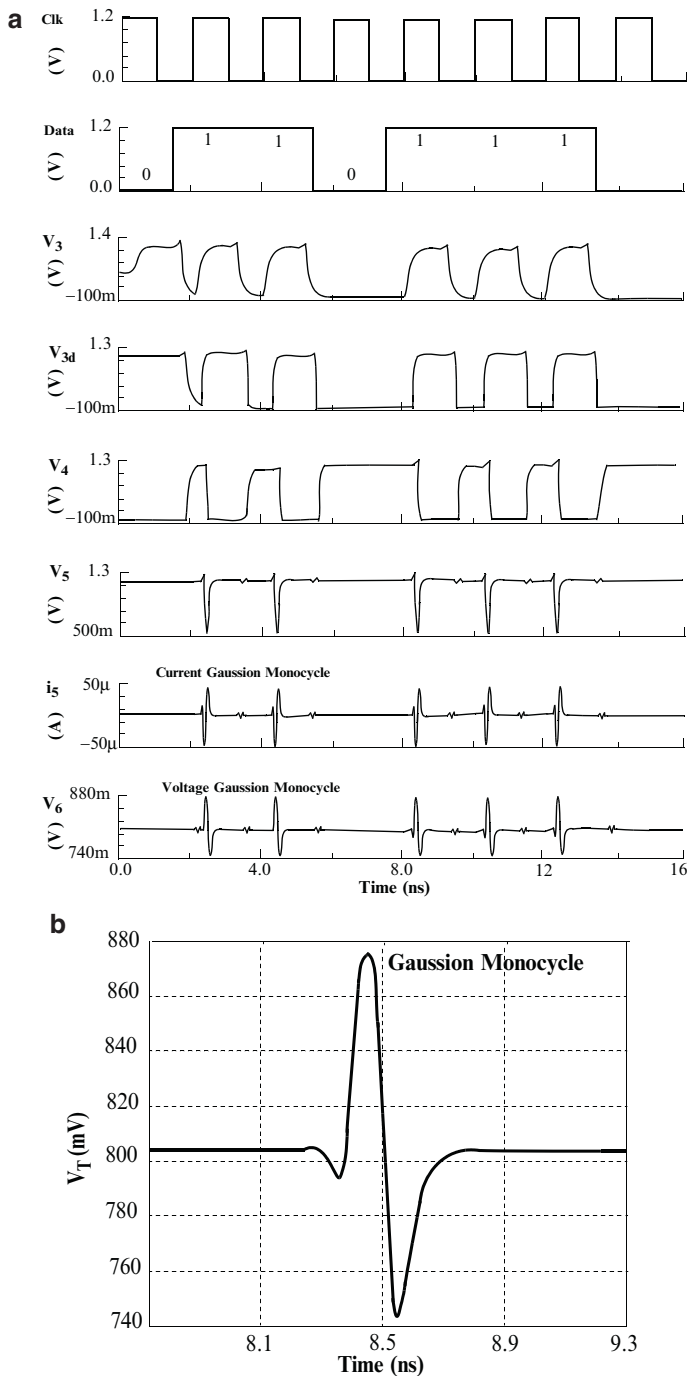


Fig. 5.20 Transient response of the impulse generator in Fig. 5.12 at each node (a) and Gaussian monocycle zoom in (b)

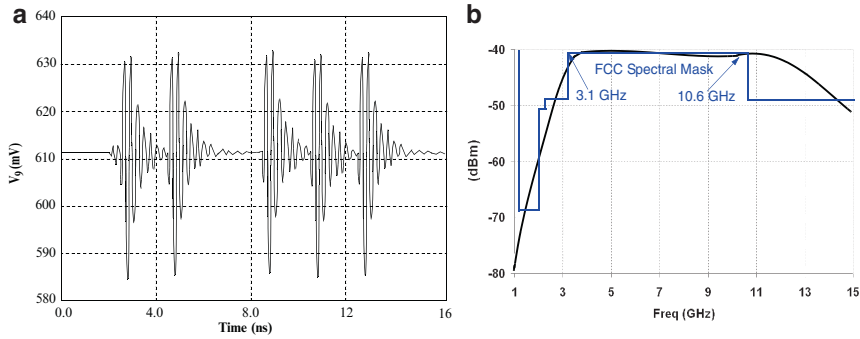


Fig. 5.21 (a) Transient response, (b) AC response of the transmitter output

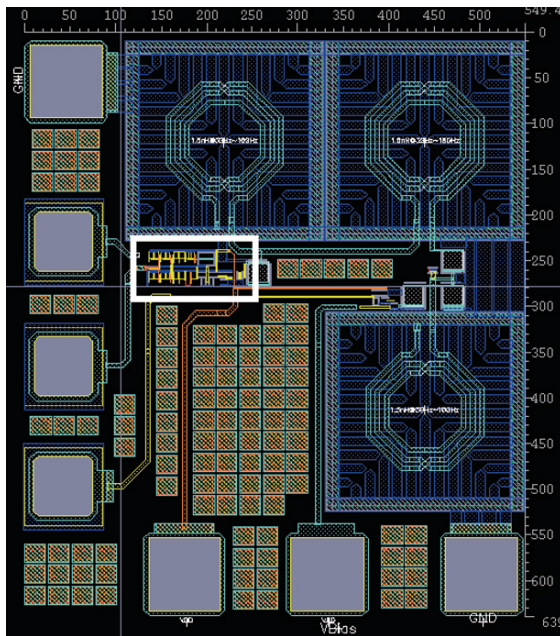


Fig. 5.22 Layout of the transmitter with LC BPF

time domain transient response at the output of the transmitter at node 9 in Fig. 5.11 is shown in Fig. 5.21a and its corresponding frequency response Fig. 5.21b. Higher order derivatives of the Gaussian pulse is observed as a side effect of the pulse shaping as shown in Fig. 5.21a. The amplitude of the signal at the output of the transmitter is 48 mV_{pp} and the duration is approximately 2 ns . The total power dissipation of the transmitter is 1.15 mW while the impulse generator only consumes $316 \mu\text{W}$ with a 1.2 V power supply. The layout of the transmitter is shown in Fig. 5.22, which takes $550 \times 650 \mu\text{m}$ including pads. The impulse generator itself without the inductor takes $45 \times 95 \mu\text{m}$.

5.4.3 Summary

A fully integrated CMOS impulse-based transmitter with on-off keying modulation scheme, for impulse radio system has been designed in a standard 180 nm CMOS technology. A simple and robust design method for a Gaussian mono-pulse generator which is based on the simple first-derivative capacitive current–voltage relationship has been presented. On-chip pulse shaping using a third-order π -type Chebychev LC band-pass filter is developed to meet the Federal Communications Commission spectrum requirement. The UWB transmitter was simulated and results analyzed. A Gaussian mono-pulse of less than 100 ps falling/rising time with a 130 mV_{pp} amplitude was observed, at a clock frequency of 500 MHz. The impulse generator consumed only 316 μ W with a total power dissipation of 1.2 mW. The post-layout simulations show that third order LC BPF shapes the output of the monocycle impulse at the cost of extra chip area and higher-order derivatives of the Gaussian pulse generated.

5.5 Magnetic Coupling Transformer for Wireless Interconnect

5.5.1 Transformer Design Considerations

The two inductors overlay each other to create H-field magnetic coupling and form wireless interconnect. The strength of the magnetic coupling between primary and secondary coils is indicated by the coupling coefficient K_m , as:

$$K_m = \frac{M}{\sqrt{L_{TX}L_{RX}}} \quad (5.5)$$

The K_m is determined by the mutual inductance (M) and self-inductance, which depend primarily upon the vertical distance between primary and secondary coils, the width (W) and spacing (S) of the metal traces, winding turns (N_t), and outer dimension (OD) as well as the substrate thickness [20]. The typical K_m value is $0.75 < K_m < 0.9$ and depends on the vertical distance and horizontal shifting between the inductors. The characteristics of the energy, voltage, current, inductivity between transmitter and receiver coils are given by [21]:

$$E_{TX} = \frac{1}{2}L_{TX}I^2_{TX} \approx \frac{1}{2}L_{TX} \left(\frac{V^2_{TX}}{\omega^2 L^2_{TX}} \right) \quad (5.6)$$

$$V_{RX} = j\omega L_{RX}I_{RX} + j\omega MI_{TX} \quad (5.7)$$

where E_{TX} is the radiated energy, V_{RX} is the induced voltage, I_{TX} , L_{TX} , I_{RX} , and L_{RX} are the current and inductances of the transmitter/primary and receiver/secondary coils respectively, and ω is the resonant frequency.

According to Eq. (5.6), current flowing through the transmitter coil is the dominant factor for transmitted energy E_{TX} . With a fixed voltage across the TX coil, more current can be obtained for a smaller resistance of the inductor. Thus, on the transmitter side, wider metal width with less turns are desired for low power consumption and maximum connectivity. On the receiver side, the magnetic flux should be picked up as much as possible to maximize the induced voltage. Therefore, from Eq. (5.7), larger L_{RX} should be chosen namely with more turns. But, the trade-offs must be considered as more secondary turns introduce extra parasitic capacitances and resistances, which cause a shift in the resonance frequency and may have a big impact on the operational bandwidth of the transformer.

A four-port S-parameter network is shown in Fig. 5.23a with simple RC input/output sources presenting the matching network on the primary and secondary ports. Eq. (5.8) illustrates how the voltage gain is calculated and converted from AC simulation and S-parameter test-bench as shown in Fig. 5.23a,

$$A_V = \frac{V_{RX}}{V_{TX}} = \frac{N_{RX}^f}{N_{TX}^f} = \frac{N_{RX}}{N_{TX}} \times \frac{\sqrt{Z_{RX}}}{\sqrt{Z_{TX}}} = S_{21} \times \frac{\sqrt{Z_{RX}}}{\sqrt{Z_{TX}}} \quad (5.8)$$

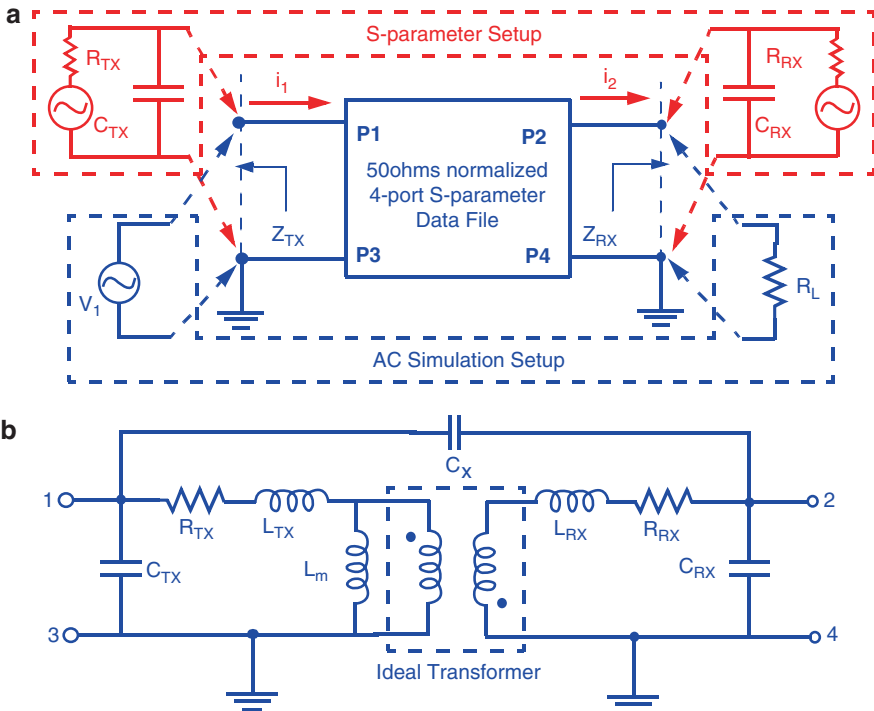


Fig. 5.23 (a) 4-port to 2-port AC, S-parameter simulation test-bench, (b) transformer simplified equivalent model

where V_{RX}^f , V_{TX}^f present the forward incident voltage waves, N_{RX} , N_{TX} present the normalized incident waves, and Z_{TX} , Z_{RX} are impedance of TX and RX ports respectively. When $Z_{TX} = Z_{RX}$, $A_V = S_{21}$, and thus, identical spiral inductors result in identical matching networks on both the primary and secondary ports, $R_{TX} = R_{RX}$, $C_{TX} = C_{RX}$. The S-parameter simulations are carried out at the center frequency of 6.85 GHz in the UWB band. The impedance networks are adjusted so that the reflection data, S_{11} and S_{22} have zero values, which represent the input/output impedance values ($R_{TX,RX}$, $C_{TX,RX}$) of the simplified equivalent model of the transformer as shown in Fig. 5.23b. L_{TX} , L_{RX} and L_m are calculated to match to the S_{21} performance from S-parameter response and can be verified by the calculated voltage gain from the AC simulation.

The octagon shape inductor is chosen over the square shape for higher Q value. The layout of the octagon shape transformer in 3-D is shown in Fig. 5.24. The transformers with different metal width (W), spacing (S), winding turns (Nt), and outer dimension (OD) have been laid out and simulated. The simulation results demonstrate that the performance of the transformer is influenced by many factors. The magnetic coupling coefficient is mainly determined by the vertical air distance (Z) between two transformer coils. The cross-section and corresponding S-parameters of the transformers are shown in Fig. 5.25. The simulated S-parameter results with $Z = 15, 30, 60 \mu\text{m}$ are shown in Fig. 5.26. The transformer with $W = 10 \mu\text{m}$, $S = 5 \mu\text{m}$, $OD = 200 \mu\text{m}$, and $N_t = 3 \times 3$ is selected among several combinations as it provides a 3-dB bandwidth from 2.5 to 11 GHz based on the S_{21} response as shown in Fig. 5.26b. The impact of horizontal misalignment (Y) on coupling is relatively small. As shown in Fig. 5.27a, the coupling is decreased less than 20% with a $30 \mu\text{m}$ misalignment. The cross-talk S_{31} , S_{41} shown in Fig. 5.25 effect from neighboring transformers is shown in Fig. 5.27b. The cross-talk between adjacent

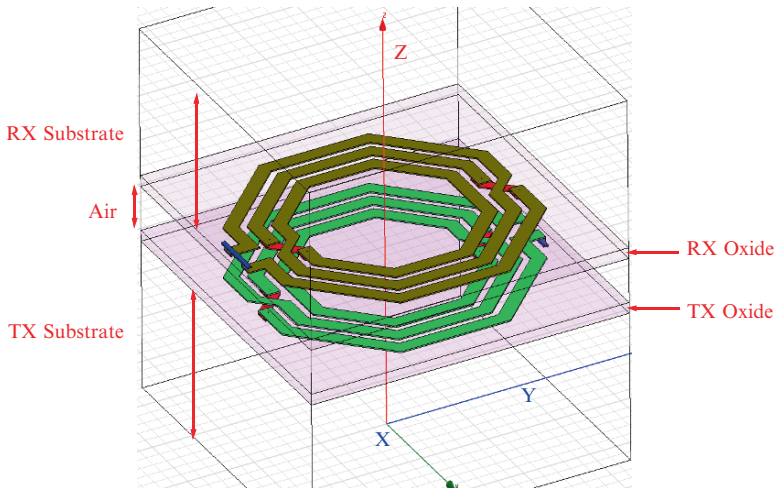


Fig. 5.24 3-D view of the transformer

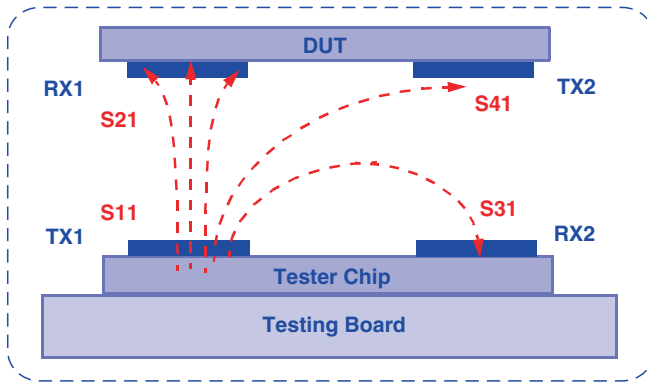


Fig. 5.25 Cross-section of transformer and corresponding S-parameter

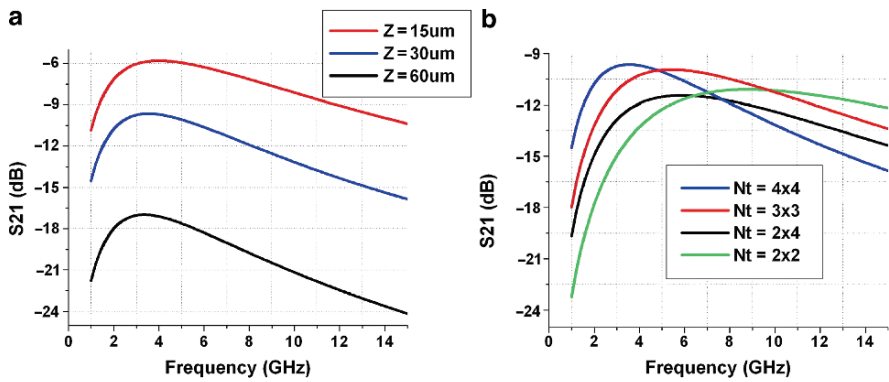


Fig. 5.26 S_{21} response (a) with various Z values, (b) with different turns

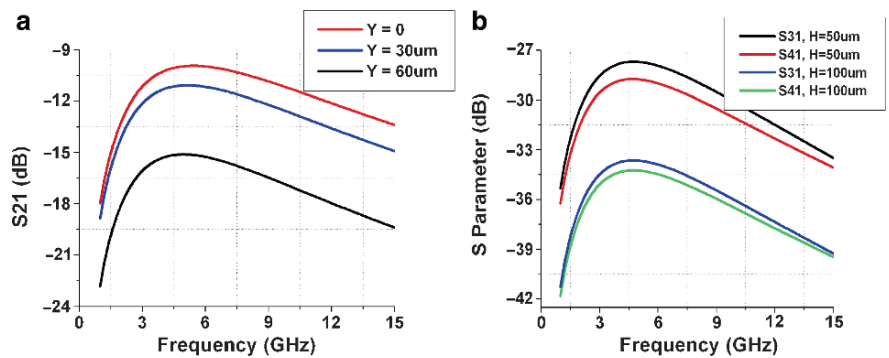


Fig. 5.27 (a) S_{21} response Y-axis misalignment, (b) cross-talks

coils is slightly greater than the coils in an echelon when there is $30\ \mu\text{m}$ vertical air distance and $50\ \mu\text{m}$ horizontal separation. It can be reduced when two channels are placed $100\ \mu\text{m}$ away horizontally.

5.5.2 Summary

In this section, the simplified equivalent model and test-bench for the magnetic coupling transformers are described. Some design constrains for octagon shape inductors are also discussed. The optimum width, space, outer dimension and turn ratio for 2.5–11 GHz UWB bandwidth is selected to provide bandpass filtering based on simulation results. The impact of transformers misalignment and cross-talk from adjacent channels on magnetic coupling performance are also discussed.

5.6 UWB LNA Design Approaches

Some low noise amplifiers for UWB receivers have been reported recently [22–30]. The main challenges for UWB LNA designs include keeping a flat high gain throughout the whole bandwidth with good input and output matching, a minimum possible noise figure and low power dissipation. The reported UWB LNA input-stage topologies can be classified as: shunt-feedback, common-gate, passive LC bandpass filter and G_m-boosted as shown in Fig. 5.28. The drawbacks of the classified LNAs are: (a) needs high feedback resistance and power to keep high gain, which results in high noise figure, (b) to keep high gain more current is needed, which results in high power consumption, (c) has large chip size because of on-chip spiral inductors of the input LC BPF, (d) need extra bias circuits for feedback components.

In the following sections we will propose a primary–secondary inductive electrostatic discharge (ESD) protected UWB LNA based on LC BPF topology in Fig. 5.28c and a modified G_m-boosted UWB LNA based on Fig. 5.28d in 180 and 90 nm CMOS technology respectively.

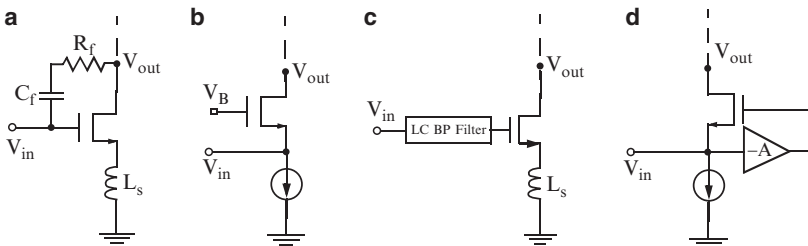


Fig. 5.28 LNA input-stage topologies: (a) shunt-feedback, (b) common-gate, (c) LC BPF, (d) G_m-boosted

5.6.1 Primary–Secondary Inductive ESD Protected LNA Design

As CMOS technologies scale down, transistor gate-oxides become thinner and RF CMOS circuits become more sensitive to electrostatic discharge (ESD) damage. The LNA is the first component of a wireless receiver system which connects to the outside world through an antenna and is readily exposed to ESD. The ESD protection capability is one of the main concerns for LNA design. There are several ESD protection strategies for RF circuits most are diode based [31–33], or inductor based [34].

Diode based ESD protection structures are commonly used in ICs. However, they are made large to handle large ESD currents and result in parasitic capacitances that are especially detrimental to RF circuits. To avoid these effects, inductor-based ESD protection devices were used. A suitable inductor works as a low-pass filter to provide a low-impedance path to ground at low frequencies (i.e. ESD pulse frequencies) and a high impedance element to feed through high RF signals.

A UWB LNA can be made by the combination of narrow-band (NB) LNA in series with an input matching network. But after ESD is inserted, extra design work is needed to compensate the ESD effect in both NB LNA and the input matching network. Figure 5.29 presents a novel idea of merging the ESD protection into the

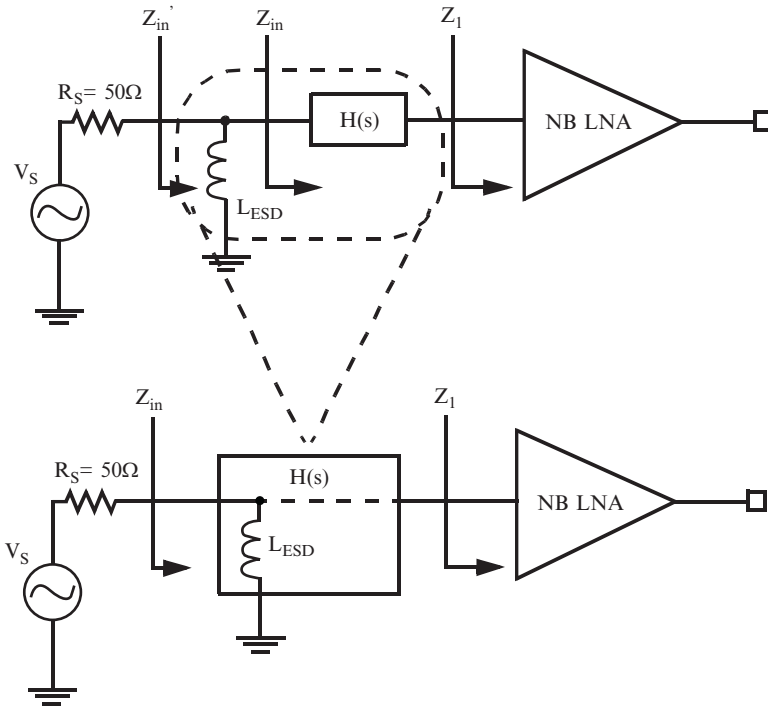


Fig. 5.29 Merging ESD protection into input matching network

input matching network of a UWB LNA. In this case, the designer needs not re-design the input matching network and NB LNA to compensate for ESD effects. Thus, it allows for more design flexibility in both ESD protection and input matching in a single UWB LNA design. The schematic of the proposed UWB LNA and its simplified small signal model are shown in Fig. 5.30a and b. The novelty of the circuit is not only because of the combination of ESD protection and input matching network (π -LC BP filter) but the inductor components L_1 and L_3 act as a primary–secondary ESD protection network which further reduces the ESD voltage stress at the gate of the LNA as shown in Fig. 5.30. The third-order π -type Chebyshev LC filter is implemented in the input matching network, with selection of the first grounded inductor which is used as the primary ESD protection component. The second ground inductor is used for secondary ESD protection as shown in Fig. 5.31.

As shown in Fig. 5.32, with 2 kV human-body model (HBM) ESD, a reported non-ESD protected LNA [22] even with a ‘plug-play’ gate-grounded NMOS (100 μ m), the ESD pad will fail as the voltage stress at the RF input port and gate of the amplifier are 125 and 28 V, while for the proposed LNA, those ESD stresses drop significantly to 4.3 and 1.1 V with any practical on-chip inductor values (1, 3, 5 nH). The gate voltage stress of 1.1 V is small enough to avoid the

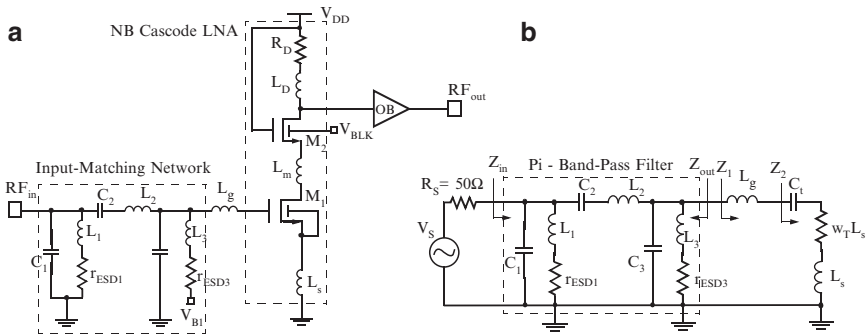


Fig. 5.30 (a) Schematic of the input matching network and LNA, (b) simplified small signal model of input matching network and LNA

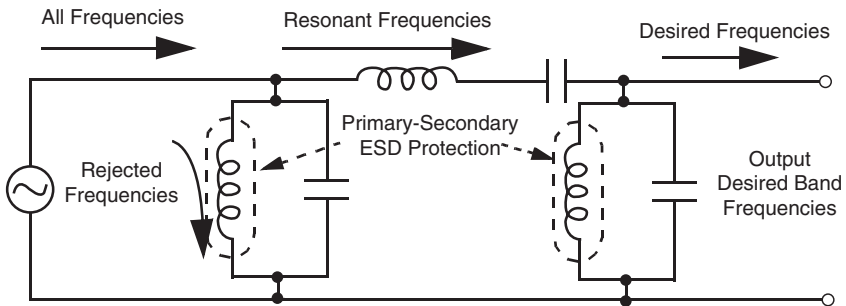


Fig. 5.31 Merging primary–secondary ESD protection into input matching network

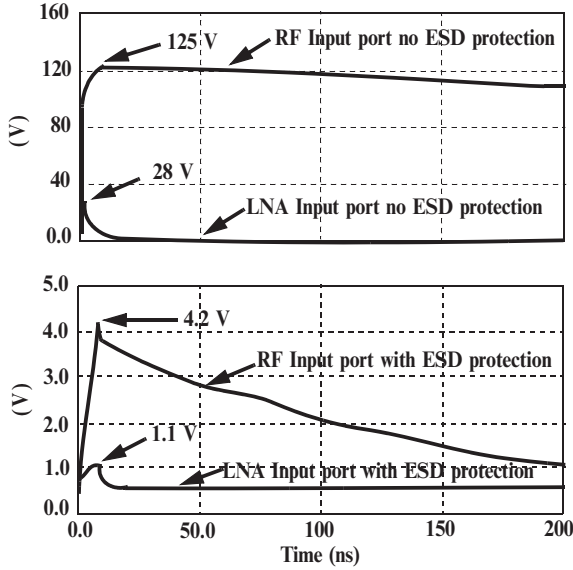


Fig. 5.32 Transient voltage at the input of LNAs with a 2KV HBM

LNA input transistor gate-oxide failure even with 90 nm thin-gate-oxide CMOS processes.

The input impedance of the LNA as shown in Fig. 5.30 is given by:

$$Z_{in}(s) = \frac{Z_1(s)}{H(s)} = \frac{sL_g + \frac{1}{sC_{gs1}} + sL_s + \omega_T L_s}{H(s)} \approx \frac{R_s}{H(s)} \quad (5.9)$$

where $H(s)$ is the transfer function of the input matching network which has a unity gain in-band and zero out-of-band, L_g , L_s is the gate inductance and source degeneration inductance at the transistor M_1 , C_{gs1} is the gate source capacitance of the transistor M_1 , ω_T is the resonant frequency. Thus, the input impedance of the LNA has 50Ω in-band and very large out-of-band. Assuming the voltage gain of the output buffer is unity, the overall voltage gain of the proposed LNA with ESD can be derived in a similar way as in [22]:

$$\frac{V_{out}(s)}{V_{in}(s)} \approx \left(\frac{g_{m1} H(s)}{sR_s C_{gs1}} \right) \cdot \left(\frac{R_D(1 + (sL_D/R_D))}{S^2 L_D C_L + sR_D C_L + 1} \right) \quad (5.10)$$

where C_L is the input load capacitance of the output buffer, R_D and L_D is the load resistance and inductance the drain of the cascode transistor M_2 in Fig. 5.30a. The normalized components of the third-order π -type Chebychev filter with < 1 dB ripple can be found in [35]. The components of the filter can be obtained by denormalization, for the given bandwidth (3–9 GHz) and center frequency (5.7 GHz). Table 5.2 shows the component values for third-order π -type Chebychev filter assuming input and output are matched to 50Ω . The optimized ESD inductor value for L_1 , which

Table 5.2 Input matching components

Type	L_1	C_1	L_2	C_2	L_3	C_3
π -Chebyshev	2 nH	270 fF	1 nH	524 fF	2 nH	270 fF

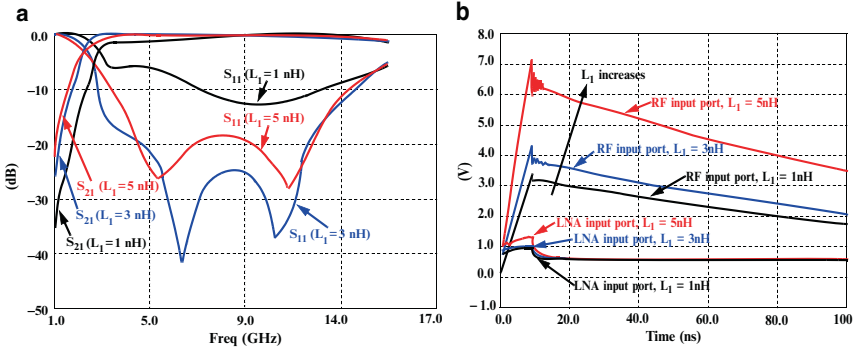


Fig. 5.33 (a) S_{21} and S_{11} response of UWB LNA with various L_1 , (b) transient voltages with various L_1 at 2KV HBM

is the primary inductor exposed to ESD, can be obtained using S-parameter simulations with various L_1 inductor values as shown in Fig. 5.33a. From simulations 3 nH is the optimum value to choose.

The LNA with inductive ESD protection was implemented in a 180 nm CMOS process. Post layout simulation were conducted to validate the design. A transient simulation with 2 kV HBM ESD in Fig. 5.33b shows that 4.3 and 1.1 V voltages were built-up at the proposed LNA’s RF input node and gate node respectively, for the corresponding current levels of 1.3 A and 0.2 mA. At up to 4 kV HBM, the voltage at the gate of the LNA is less than 4 V with peak current of 1 mA. S-parameter simulation results indicated that a flat power gain of 10 dB from 2.7 to 9 GHz and an input matching of less than -10 dB are obtained as shown in Fig. 5.34a. Output matching of less than -20 dB is obtained but is not effected by ESD inductor because of the output buffer. A minimum noise figure of 3.2 dB is obtained for the proposed LNA as shown in Fig. 5.34b. A two-tone signal of 5.6 and 5.8 GHz has been applied to LNAs input ports and the input-referred third-order intercept points (IIP3) are around -10 dBm simulated. The proposed amplifier is unconditionally stable with stability factor, $K_f \gg 1$. The die photo of the circuit is shown in Fig. 5.35, which takes an area of $800 \times 990 \mu\text{m}$.

5.6.2 G_m -Boosted LNA Design

Figure 5.36a shows a schematic of the proposed UWB LNA without DC biasing circuits. It consists of a G_m -boosted common-gate input stage amplifier, a cascode

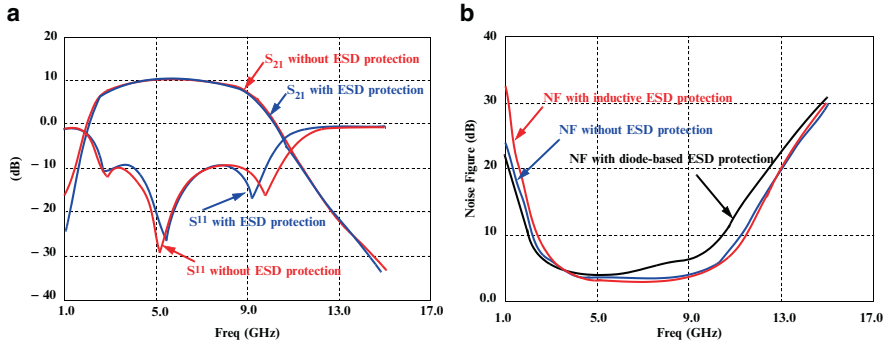


Fig. 5.34 (a) S_{21} and S_{11} response of LNA and [23], (b) noise figure of LNAs

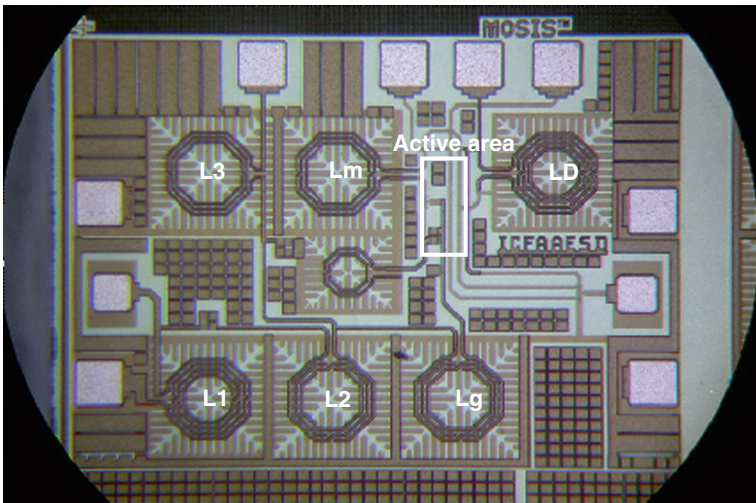


Fig. 5.35 Die micrograph of ESD protected LNA

LNA stage for gain-bandwidth (GBW) extension, and a common-drain output buffer for testing purpose.

5.6.2.1 Input Impedance

To simplify the analysis, the input stage is considered as a conventional common-gate stage M_1 with gate connecting to AC ground and an enhanced g_m . The enhanced transconductance (G_{m1B}) of the CG transistor M_1 is achieved through a negative feedback network formed by a common-source transistor M_2 with a NMOS active load transistor M_3 as:

$$G_{m1B} = (1 - A_V)g_{m1} = [1 + g_{m2}/(g_{m3} + g_{o2} + g_{o3})] \cdot g_{m1} \quad (5.11)$$

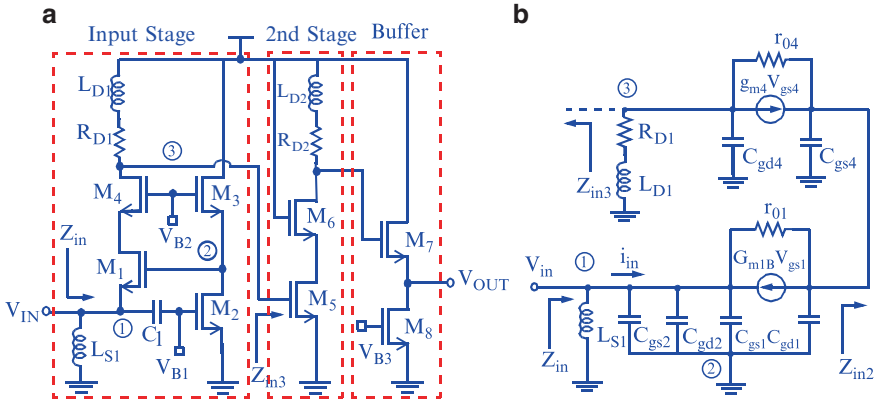


Fig. 5.36 (a) Schematic of G_m -boosted UWB LNA, (b) simplified small signal model of the input stage

where A_v is gain of the feedback network and g_{m1} , g_{m2} , g_{m3} are the transconductance of transistor M_1 , M_2 and M_3 respectively, g_{o2} , g_{o3} are the output transconductance of transistor M_2 and M_3 . The input admittance of the LNA can be approximated from the simplified small signal model of CG input stage as shown in Fig. 5.34b, where M_2 and M_3 are absorbed by M_1 with an enhanced G_{m1B} :

$$Y_{in}(\omega) = G_{m1B} + Y_1(\omega) + \frac{Y_2(\omega) - G_{m1B}}{Y_1(\omega)r_{01} + 1} \quad (5.12)$$

where $Y_1(\omega)$ and $Y_2(\omega)$ are given by:

$$Y_1(\omega) = j\omega(C_{gs1} + C_{gs2} + C_{gd2}) + (1/(j\omega L_{s1})) \quad (5.13)$$

$$Y_2(\omega) = g_{m4} + j\omega C_{gs4} + \frac{Y_3(\omega) - g_{m4}}{Y_3(\omega)r_{04} + 1} \quad (5.14)$$

$$Y_3(\omega) = Y_{in3} + \frac{1}{j\omega C_{gs4}} + j\omega L_{D1} + R_{D1} \quad (5.15)$$

where Y_{in3} is the input admittance of the next stage. A DC blocking capacitor $C_1 \gg C_{gs2}$ is neglected in the small signal model.

With G_m boosting, the high transconductance ($G_{m1B} = 20 \text{ ms}$) can be easily obtained without large size of CG transistor M_1 as long as the ratio of $g_{m2}/(g_{m3} + g_{o2} + g_{o3})$ is high. Therefore, the value of parasitic capacitance (C_{gs1} , C_{gd1}) of M_1 can be kept small to maintain wide bandwidth. However, it is not necessary for the LNA to achieve 50Ω input matching when an RF input signal directly comes through wireless interconnect where large voltage swing is one of the main concerns.

5.6.2.2 Gain Bandwidth and Noise Analysis

The gain of the input stage G_m -boosted CG amplifier is determined by the ratio of R_{D1}/R_S where R_S is 50Ω in this case. As R_{D1} also sets the dominant pole frequency, it cannot be set too high. The load inductor of L_{D1} resonates with the total output capacitance of the input stage CG amplifier, but it is not enough to achieve a wide bandwidth. To extend the bandwidth, a cascode amplifier stage with inductive load L_{D2} is added to resonate with the output capacitance of the previous stage and enhance both the gain and bandwidth within the 3.1–10.6 GHz band [24]. Figure 5.37a shows S-parameter simulation results of how the gain bandwidth is extended by a second-stage cascode amplifier with different L_{D2} values.

In general, the noise figure of the system mainly depends on the first stage. Due to the relatively low gain of the common-gate amplifier, the noise figure of the second stage of cascode LNA cannot be neglected. With the input G_m -boosted stage, the noise factor of the boosted CG input stage is improved by a factor of $(1 - A_V)^2$ compared to a CG stage as given by [30]:

$$F_1 = 1 + \frac{\gamma}{\alpha(1 - A_V)^2 g_{m1} R_S} \tag{5.16}$$

where R_S is the source resistance, and α and γ are FET noise parameters. The noise factor introduced by the next stage common-source transistor M5 is given by [24]:

$$F_2 = 1 + \frac{\gamma \cdot R_{D1}}{\alpha \cdot g_{m5} R_{D2}^2} \tag{5.17}$$

The total noise factor of the system contributed by the CG and next stage amplifier is $F_1 + F_2$ or given by follows with matched stages:

$$F_{tot} = F_1 + [(F_2 - 1)/(R_{D1}/R_S)] \tag{5.18}$$

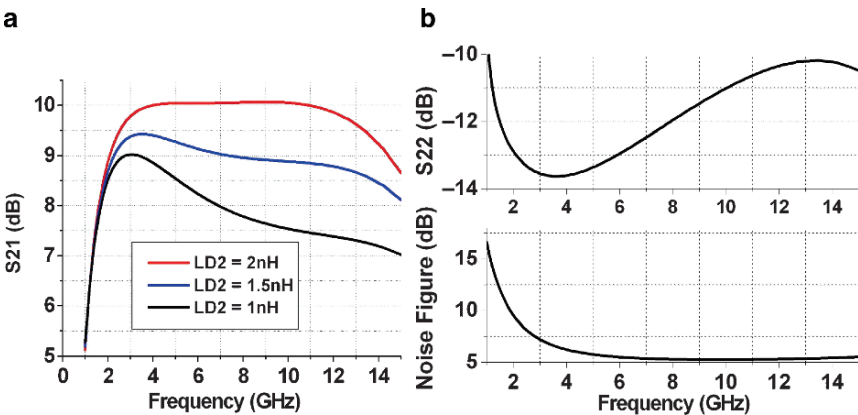


Fig. 5.37 (a) S₂₁ response with different L_{D2} values, (b) S₂₂ and noise figure

From Eqs. (5.16) and (5.17), A_v , g_{m1} , R_s , and g_{m5} appear in the denominators of all noise components and therefore by increasing them the overall noise can be reduced. However, the input matching degrades as g_{m1} and A_v increase. Similarly, to minimize the noise contribution of the common-source stage, g_{m5} can be increased to minimize the noise contribution from the second stage. The S-parameter response of S_{22} (<-10 dB) and noise figure (<6 dB) are obtained as shown in Fig. 5.37b. Another advantage of wireless interconnect is that the input matching is determined by the output impedance of the receiver coil which can be designed larger than 50Ω to relax the matching and noise requirement.

5.6.3 Summary

A novel idea for inductive primary–secondary ESD protected LNA is presented. By merging inductive ESD protection into input LC bandpass matching network, the LNA performance is not degraded by ESD circuits. The LNA can be protected by the ESD up to 2 KV HBM and prevents the damage from the testing environment. A modified G_m -boosted LNA design is also discussed. The G_m enhancement techniques relax the input matching requirement and reduce the noise figure of the circuit. The second cascode stage resonates out the parasitic capacitances and extends the gain bandwidth of the LNA.

5.7 Proposed UWB Transceiver Circuit Implementations

In Section 5.4, we presented a Gaussian monocycle generator and an LC BPF for pulse shaping. However, it is not necessary to meet FCC mask for chip testing applications as long as the transmission power is low enough and can be detected by the receiver. To save chip space the LC BPF is not implemented in the system and we only employ wireless interconnect transformers to provide partial pulse shaping and ESD protection.

The complete schematic of the UWB transceiver system for contactless chip testing is shown in Fig. 5.38 and the layout of the circuit with testing pads in a 90 nm CMOS process is shown in Fig. 5.39 [36]. The dimension of the UWB transmitter is 540×650 and $670 \times 730 \mu\text{m}$ for the receiver. The active areas are 60×10 , and $63 \times 30 \mu\text{m}$ for receiver and transmitter respectively.

Figure 5.40 shows the extracted transient response from node 1 to node 5 and node 6 to node 10 in Fig. 5.38. The clock signal is set to 5 GHz frequency and a data pattern of “0111001100010...” is used at the rate of 5 Gb/s. Both of their rising/falling times are set to 10 ps. The amplitude of the signal at the output of the impulse generator is about 480 mV_{pp} with 50 ps rising/falling time. At the TX coil the signal is 200 mV_{pp} with 45 ps rising/falling time. At the RX coil, the signal drops to 90 mV_{pp} and the output is amplified to 210 mV_{pp} after the LNA. Figure 5.41

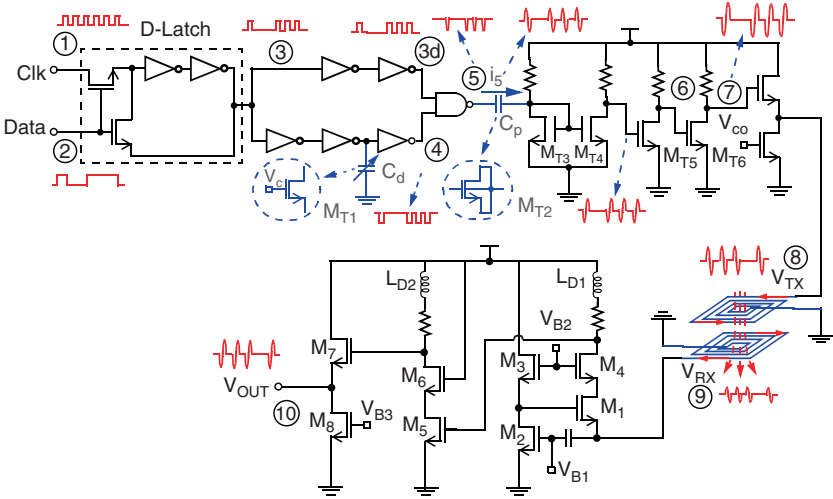


Fig. 5.38 Complete schematic of the UWB transceiver system

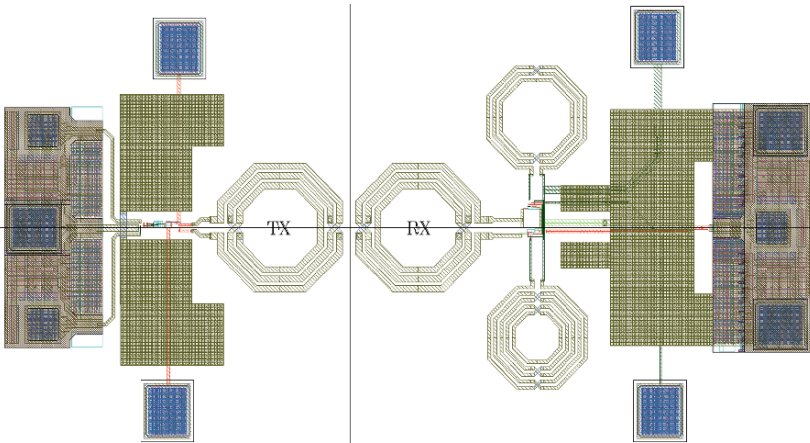


Fig. 5.39 Layout of the UWB transceiver system

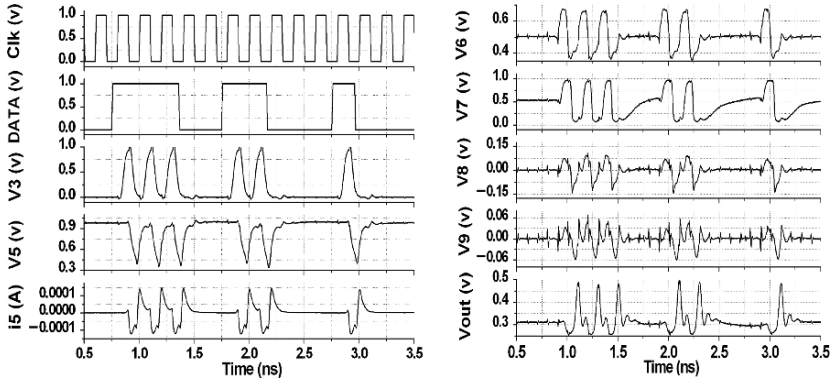


Fig. 5.40 Post-layout transient response of the UWB transceiver system at each node

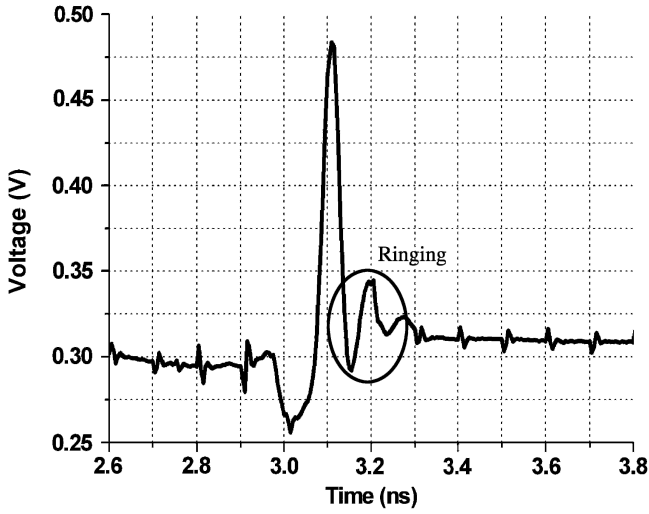


Fig. 5.41 Zoom-in of V_{out} single impulse

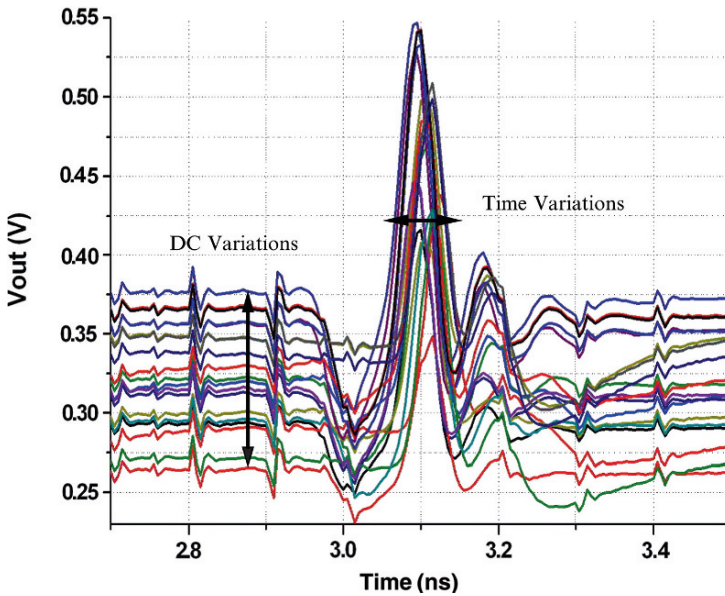


Fig. 5.42 Monte-Carlo analysis

shows the detail of the output with 90 ps rising times, 40 ps falling times, and less than 100 ps of ringing. A Monte Carlo analysis is run to estimate the system's sensitivity. From Fig. 5.42 the output signal shows a reasonable discrepancy as a result of process and mismatch variations (i.e., a typical of 20 iterations). Variations in dc level and time are around 100 mV and 60 ps respectively. The total power

consumption of the transceiver is 9 mW with less than 3 mW consumed in the transmitter, 6 mW in the LNA excluding 3 mW in the output buffer under a 1 V power supply using a 90 nm CMOS technology.

5.8 Conclusions

Conventional wired testing methodologies have limitations on the number of pads, pitch sizes, operating frequency and parallel testing capability, which makes testing a growing problem for highly integrated circuits. This has become the main driving force behind recent research on wireless testing methods. One of the main challenges is to provide high data rates and low power consumption. The primary goal of this work was to investigate wireless links and UWB transceiver systems to select a suitable platform for wireless testing.

An overview of the wireless links and UWB systems shown the advantages of magnetic coupling for short-range communications and characteristic of UWB systems for higher operating speed with low power transmission. Two design approaches for UWB systems have been studied. The IR-UWB system is chosen over the OFDM for its low circuit complexity, and low power consumption.

A fully integrated CMOS impulse-based transmitter with on-off keying modulation scheme, for ultra-wideband systems has been designed in a standard 180 nm CMOS technology. A simple and robust design method for a Gaussian mono-pulse generator has been presented. On-chip pulse shaping using a third-order π -type Chebyshev LC BPF is developed to meet the FCC spectrum requirement.

The equivalent model and design considerations for on-chip transformers with more efficient power transmission were discussed. Based on HFSS 3-D EM simulations with 90 nm CMOS technology substrate mapping files, optimum metal width, space, outer dimension and turn ratio of the transformer are found. It has 2.5–11 GHz UWB bandwidth and provides a bandpass filtering for the transmission signals. The impact of misalignment and cross talk from adjacent channel on transformer coupling is also described.

Two topologies have been proposed for UWB LNAs. A novel inductive ESD protection LNA employing input matching network for the ESD devices is presented and studied. A modified G_m -boosted LNA design to relax high G_m requirement and noise figure is also discussed. The G_m enhancement techniques relax the input matching requirement and reduce the noise figure of the circuit.

A fully integrated CMOS impulse-based transmitter with on-off keying modulation scheme, for ultra-wideband impulse radio system has been designed in a standard 90 nm CMOS technology. Instead of a third-order π -type Chebyshev LC Band-Pass Filter for pulse shaping, the magnetic coupling transformer provides both wireless data transmission and bandpass filtering. At the receiver side, we have employed G_m -boosted LNAs, with an inductive GBW enhancement technique to achieve a higher gain bandwidth and relieve the trade-offs among input matching, noise performance, flat gain response and high gain. The overall transceiver

system can operate at a data rate of 5 Gb/s, with an output of 90 ps rising times, 40 ps falling times, and less than 100 ps of ringing. The total power consumption of the transceiver is 9 mW with less than 3 mW consumed in the transmitter, 6 mW in the LNA excluding 3 mW in the output buffer under a 1 V power supply.

References

1. B. A. Floyd et al., "Intra-Chip Wireless Interconnection for Clock Distribution Implemented with Integrated Antennas", *IEEE Journal of Solid-State Circuits*, Vol. 37, No. 5, pp. 543–552, May 2002.
2. A. Valdes-Garcia, J. Silva-Martinez, and E. Sanchez-Sinencio, "On-Chip Testing Techniques for RF Wireless Transceivers", *IEEE Design & Test of Computers*, Vol. 23, pp. 268–277, April 2006.
3. M. Quirk and J. Serda, *Semiconductor Manufacturing Technology*, Prentice-Hall, Columbus, NJ, 2001.
4. IEEE Standards Committee. IEEE standard test access port and boundary-scan architecture, July 1990. IEEE Std 1149.1-1990, IEE, 345 East 47th street, New York, NY, 10017-2349.
5. H. Eberle and A. Wander, "Testing Systems Wirelessly", *Proceedings of the 22nd IEEE VLSI Test Symposium*, pp. 335–340, 2004.
6. F. Carrez et al., "A Low-Cost Active Antenna for Short-Range Communication Applications", *IEEE Microwave and Guided Wave Letters*, Vol. 8, No. 6, pp. 215–217, June 1998.
7. P. K. Saha, N. T. Sasaki, Kikkawa, "A CMOS UWB Transmitter for Intra/Inter-Chip Wireless Communication", *IEEE Eighth International Symposium on Spread Spectrum Techniques and Applications*, pp. 962–966, September 2004.
8. C. E. Shannon, "A Mathematical Theory of Communications," *Proceedings of the IRE*, Vol. 37, pp. 10–21, January 1949.
9. Federal Communications Commission, "First Report and Order, Revision of Part 15 of the Commission's Rules Regarding Ultra-Wideband Transmission Systems", *ET Docket 98-153*, February 14, 2002.
10. S. Roy et al., "Ultrawideband Radio Design: The Promise of High-Speed, Short-Range Wireless Connectivity" *Proceedings of the IEEE*, Vol. 92, No. 2, pp. 295–311, April 2004.
11. M. Z. Win and R. A. Scholtz, "Ultra-Wide Bandwidth Time-Hopping Spread-Spectrum Impulse Radio for Wireless Multiple-Access Communications", *IEEE Trans. Communications*, Vol. 48, pp. 669–679, April 2000.
12. http://grouper.ieee.org/groups/802/15/pub/TG3a_CFP.html
13. W. Chung et al., "Signaling and Multiple Access Techniques for Ultra Wideband 4G Wireless Communication Systems" *IEEE Wireless Communications*, pp. 46–55, April 2005.
14. X. Chen and S. Kiaei, "Monocycle Shapes for Ultra Wideband System" *IEEE Symposium on Circuits and Systems*, pp. 597–600, April 2002.
15. M. Z. Win, "A Unified Spectral Analysis of Generalized Time-hopping Spread-Spectrum Signals in the Presence of Timing Jitter", *IEEE Journal on Selected Areas in Communications*, Vol. 2, No. 9, pp. 1664–1676, December 2002.
16. P. Heydari, "Design Considerations for Low-Power Ultra Wideband Receivers", *IEEE 6th Symposium on Quality Electronic Design*, pp. 1–6, 2005.
17. M. Choi and A. A. Abidi, "A 6 b 1.3 GSample/s A/D Converter in 0.35 μm CMOS", *IEEE Solid-State Circuits Conference*, pp. 126–127, February 2001.
18. S. Bagga, G. Vita, S. A. P. Haddad, W. A. Serdijn and J. R. Long, "A PPM Gaussian Pulse Generator for Ultra-Wideband Communications" *IEEE Symposium on Circuits and Systems*, pp. 109–112, May 2004.
19. Y. Wang, A. Ho, K. Iniewski and V. Gaudet, "Inductive ESD Protection For Narrow Band and Ultra-Wideband CMOS LNAs" *IEEE Symposium on Circuits and Systems*, pp. 3920–3923, May 2007.

20. J. Long, "Monolithic Transformers for Silicon RFIC Design", *IEEE Journal of Solid-State Circuits*, Vol. 35, No. 9, pp. 1368–1382, 2000.
21. A. M. Niknejad. "Electromagnetics for High-Speed Analog and Digital Communication Circuits", Cambridge University Press, Cambridge, 2007.
22. A. Bevilacqua and A. M. Niknejad, "An ultra-wideband CMOS LNA for 3.1 to 10.6 GHz wireless receiver", *IEEE Journal of Solid-State Circuits*, Vol. 39, No. 12, pp. 2259–2268, December 2004.
23. A. Ismail and A. A. Abidi, "A 3–10-GHz Low-Noise Amplifier with Wideband LC-Ladder Matching Network", *IEEE Journal of Solid-State Circuits*, Vol. 39, No. 12, pp. 2269–2277, December 2004.
24. Y. Lu et al. "A Novel CMOS Low-Noise Amplifier Design for 3.1 to 10.6 GHz Ultra-Wide-Band Wireless Receivers", *IEEE Trans. on Circuits and Systems I*, Vol. 53, No. 8, pp. 1683–1692, 2006.
25. S. B. T. Wang, A. M. Niknejad and R. W. Broderson, "A Sub-mW 960-MHz Ultra-Wideband CMOS LNA", *IEEE Radio Frequency Integrated Circuits Symposium*, pp. 35–38, 2005.
26. S. Vishwakarma, S. Jung and Y. Joo, "Ultra Wideband CMOS Low Noise Amplifier with Active Input Matching", *Ultra Wideband Systems and Technology Conference*, pp. 415–429, 2004.
27. Y. Wang and Z. Khan, "A Very Low Voltage Design for Different CMOS Low-Noise Amplifier Topologies at 5GHz," *IEEE Midwest Symposium on Circuits and Systems 2005*, pp. 42–45, August 2005.
28. B. Analui and A. Hajimiri, "Bandwidth Enhancement for Transimpedance Amplifiers" *IEEE Journal of Solid-State Circuits*, Vol. 39, No. 8, pp. 1263–1270, August 2004.
29. T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, 2nd Ed., Cambridge University Press, Cambridge, pp. 348–351, 2004.
30. D. J. Allstot et al., "Design Considerations for CMOS low noise amplifiers" *IEEE Radio Frequency Integrated Circuits Symposium*, pp. 97–100, June 2004.
31. Z. H. Wang, "On-Chip ESD Protection for Integrated Circuits", Kluwer, Dordrecht, 2002.
32. Liu et al., "A 6.5kV ESD Protected 3–5GHz Ultra-wideband BiCMOS Low Noise Amplifier Using Interstage Gain Roll-off Compensation", *Ultra Wideband Conference*, pp. 525–529, 2005.
33. M. I. Natarajan et al., "RFCMOS ESD Protection and Reliability", *Proceedings of the 12th International Symposium on Physical and Failure Analysis*, pp. 59–66, 2005.
34. P. Leroux and M. Steyaert, "High-performance 5.2GHz LNA with On-chip Inductor to Provide ESD Protection", *Electronic Letters*, Vol. 37, pp. 467–469, 2001.
35. Y. Wang, A. Ho, K. Iniewski and V. Gaudet, "Inductive ESD Protection For Narrowband and UWB CMOS LNAs", *IEEE Symposium on Circuits and Systems*, pp. 3920–3923, 2007.
36. Y. Wang, A. M. Niknejad, V. Gaudet and K. Iniewski, "A CMOS IR-UWB Transceiver Design for Contact-less Chip Testing Applications", *IEEE TCASII*, Vol. 55, No. 4, pp. 334–338, 2008.

Chapter 6

Multi-mode Power Amplifiers for Wireless Handset Applications

Junxiong Deng and Lawrence E. Larson

6.1 Introduction

With the increasing demand for integrating more standards in a single handset, multiple parallel transmit paths are normally implemented. To further reduce die size and cost, an adaptive multi-mode solution is more plausible. Accordingly a multi-mode power amplifier is highly desired. Efficiency and linearity are the most critical design parameters in the multi-mode PA design. Power amplifiers (PAs) in wireless handsets are key components that consume a significant portion of the DC power budget in the transmitter. For those with constant envelope signals, like GSM with Gaussian Minimum Shift Keying (GMSK), power amplifiers are mainly designed for high efficiency and there is no specific linearity requirement. Therefore, nonlinear but high efficiency PA designs, such as Class B [1] and Class E [2], are typically employed in these systems. For third-generation cellular systems with non-constant envelope signals, such as Wideband CDMA [3], stringent linearity requirements are posed in the design of power amplifiers for higher spectral efficiency. Class AB operation is typically selected [4, 5], because it has the best tradeoff between efficiency and linearity. For multi-mode handset power amplifiers, amplifier bias operation may vary with different system modes: Class AB is typically employed in high-linearity modes and Class B is more suitable in high-efficiency modes.

Average power efficiency (over the full range of output powers), instead of peak power-added efficiency (PAE), is the key factor determining the talk time and battery

J. Deng (✉)

Qualcomm Inc, WT-540H, 5775 Morehouse Drive, San Diego, CA 92121
e-mail: jdeng@qualcomm.com

L.E. Larson

Department of Electrical and Computer Engineering, University of California at San Diego,
9500 Gilman Drive, La Jolla, CA 92093 USA
e-mail: larson@ece.ucsd.edu

life for portable wireless applications [6]. Linear PAs are typically operated in a class AB mode and are often backed off from the maximum output power to achieve the desired linearity. Even though peak power efficiency may be high at high output powers, it drops quickly with the power back-off, resulting in a poor average power efficiency. To increase power efficiencies in the low power region, different dynamic biasing techniques [7–10] have been developed. Altering the DC bias *current* in response to changing power requirements – also known as Dynamic Current Biasing (DCB) – often results in significant amplifier gain variation. Although this gain-variation issue may be alleviated through the power control loop of the CDMA handset, it adds an extra burden to the algorithm. Furthermore, considering the power amplifier as a standalone block, this gain variation can dramatically degrade the Error Vector Magnitude (a system-level figure of merit for the accuracy of digitally-modulated signals). Altering the DC bias collector *voltage* in response to changing power requirements – also known as Dynamic Voltage Biasing (DVB) – typically requires the use of DC-DC converters with their associated large off-chip components and extra chip area and cost. Therefore, it is not attractive for small feature size handset PAs.

Switched dual dynamic biasing (SDDB) [11, 12] that does not require any additional off-chip components is proposed to provide an *integrated* solution with substantially improved average power efficiency while keeping the power gain roughly constant.

Linearity improvement of power amplifiers increases the maximum output power satisfying the linearity requirements. Consequently, the peak power efficiency will be also boosted. Different linearization techniques, such as feedback, feedforward, and predistortion, carry on distinct features and limitations [13–18]. To accommodate the high-linearity requirement in multi-mode handset power amplifiers, digital predistortion is chiefly presented because it can be conveniently integrated, flexibly controlled, and economically implemented without sacrificing other performance, such as power gain and efficiency.

The chapter is organized as follows. Section 6.2 introduces different process technologies popular in power amplifier design and compares published CDMA/WCDMA handset power amplifiers in recent years. Section 6.3 is concerned with efficiency enhancement techniques in power amplifier design. SDDB is mainly covered with operation principle and design considerations. Section 6.4 deals with linearization techniques in power amplifiers. Complex-gain digital predistortion is elaborated from theory to implementation. Measurement results are presented in Section V and conclusions are given in last section.

6.2 Power Amplifier Technologies

It is vital to select the matched power amplifier technology depending on various requirements, such as performance, cost, and time to market. There are a variety of technologies for RF power amplifiers [19]. RF power amplifiers are normally

Table 6.1 Comparison of typical technologies for RF power amplifiers [19]

Technology	GaAs HBT	Si BJT	SiGe HBT	CMOS	LDMOS
f_T	46 GHz	27 GHz (HF ^a) 22 GHz (HV ^b)	44 GHz (HF) 25 GHz (HV)	<20 GHz	<20 GHz
BV_{ceo}	14.3 V	3.3 V (HF) 6.2 V (HV)	3.0 V (HF) 6.0 V (HV)	5 V	15 V
Thermal conductivity	0.49 W/cm-C	1.5 W/cm-C	1.5 W/cm-C	1.5 W/cm-C	1.5 W/cm-C
Normalized cost 1mm²	1	0.3	0.3	<0.2	0.3
Wafer size	6'' wafer	8'' wafer	8'' wafer	12'' wafer	8'' wafer
PAE	>40%	>30%	>30%	>30%	>35%
Linearity	Great	Good	Good	Good	Good
Power gain	Great	Good	Great	Poor	Poor
Integration level	Poor	Good	Good	Great	Good

^aHigh frequency npn transistor.

^bHigh breakdown npn transistor.

characterized in device physics by cutoff frequency, breakdown voltage, thermal conductivity, integration level, size and cost. Furthermore, they are assessed with efficiency or power added efficiency (PAE), power gain, and linearity in terms of inter-modulation distortion (IMD) or adjacent channel power ratio (ACPR). Table 6.1 summarizes several popular technologies for RF power amplifiers. Their devices are characterized with their schematic cross-sections, as shown in Fig. 6.1.

The GaAs HBT is the preferred technology [20] in the current commercial handset PA market. Its efficiency and linearity are so great that fast engineering development becomes feasible to meet the design specifications on the first design iteration. High cost and low integration level are the two drawbacks of the GaAs HBT.

From the standpoints of fabrication cost and integration level, CMOS is the best among all technologies. However, there are a number of issues limiting its application in the PA market. For the same functionality, CMOS PAs are larger than their GaAs or SiGe counterparts. Unlike GaAs or SiGe HBT technology, which incorporates vertical transistors, the footprint of the MOSFETs in CMOS increases rapidly when scaled up to accommodate high output powers. This tends to cancel out the advantage of having a low-cost manufacturing process. Also, the low breakdown voltage in CMOS [21] limits the maximum voltage swing and its low transconductance-to-current ratio results in relatively poor power gain.

LDMOS has a low fabrication cost and good integration level. However, the bias voltage in LDMOS PAs is quite high, which makes this technology typically suitable for base-station PAs, instead of handset applications. Similar to CMOS, the application of LDMOS in PA market is also limited due to its poor power gain.

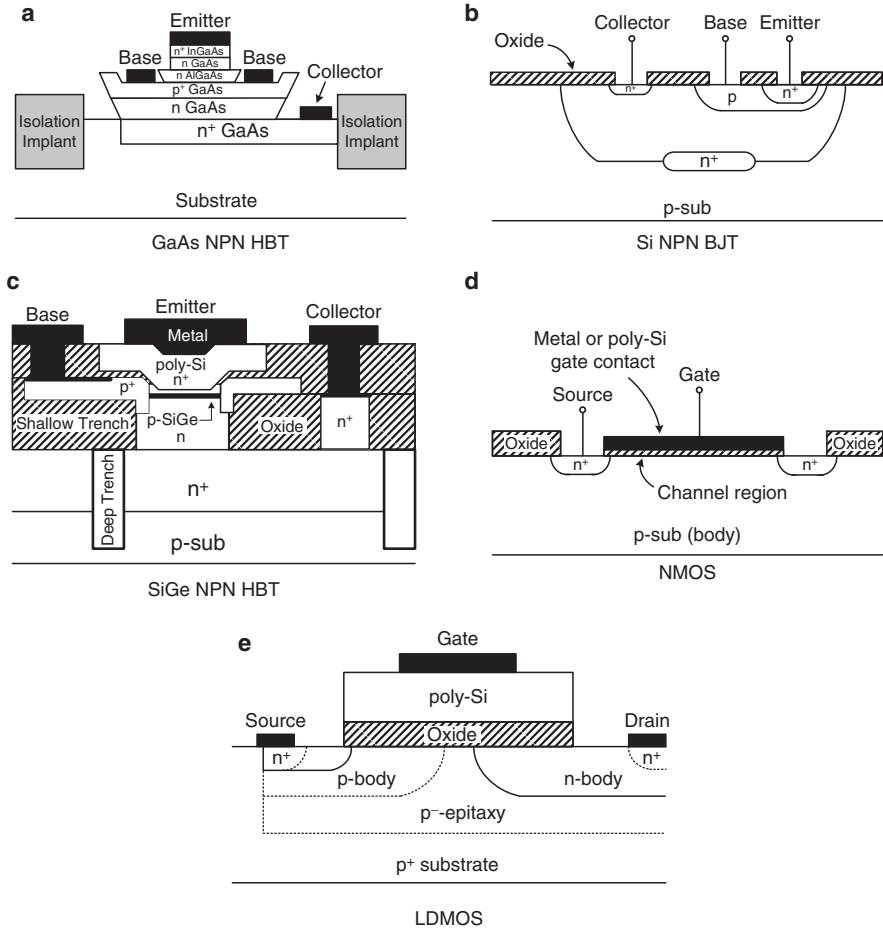


Fig. 6.1 Schematic cross-sections of different devices. (a) GaAs NPN HBT. (b) Si NPN BJT. (c) SiGe NPN HBT. (d) NMOS. (e) LDMOS

In recent years, the SiGe HBT has become a competitive candidate [22] for the development of cellular handset power amplifiers, since the SiGe HBT exhibits good linearity, low-cost and compatibility with BiCMOS technology [23], even though the SiGe HBT has a lower breakdown voltage and efficiency than its GaAs counterpart and is also affected by the thermal runaway issue [24].

Table 6.2 compares the performance of recently reported CDMA/WCDMA handset power amplifiers in different technologies. To the authors' knowledge, SiGe power amplifiers are becoming increasingly popular in the market for third-generation handset power amplifiers.

Table 6.2 Comparison of recently reported CDMA/WCDMA handset power amplifiers

Ref.	Technology	P_{out} (dBm)	PAE @ P_{out}	Gain (dB)	Bias Voltage	ACPR @ P_{out}	Comments
Yamamoto [25]	InGaP/GaAs HBT	27.5	40%	26.5	3.5 V	-50 dBc @ 1.25 MHz	CDMA
Kim [26]	InGaP/GaAs HBT	27	33%	25	3.4 V	-30 dBc @ 5 MHz	WCDMA
Wang [27]	CMOS	24	29%	23.9	3.3 V	-35 dBc @ 5 MHz	WCDMA
Jeon [28]	InGaP/GaAs HBT	28	37%	27	3.5 V	-47 dBc @ 1.25 MHz	CDMA
Noh [29]	InGaP/GaAs HBT	28.3	52.4%	11.5	3.4 V	-33 dBc @ 5 MHz	WCDMA
Staudinger [30]	PHEMT	30	38%	27	3.5 V	-42 dBc @ 5 MHz	WCDMA
Hau [31]	HJFET	26	57.4%	n/a	3.5 V	-40 dBc @ 5 MHz	WCDMA
Luo [32]	Si BiCMOS	28.2	30%	21.5	3.6 V	-45 dBc @ 1.25 MHz	CDMA
Vintola [33]	AlGaAs/GaAs HBT	24	27%	30	3.5 V	-36 dBc @ 5 MHz	WCDMA
Tseng [34]	SiGe HBT	28	36%	22	4 V	-44 dBc @ 1.25 MHz	CDMA
Iwai [35]	InGaP/GaAs HBT	27	42%	30.5	3.5 V	-38 dBc @ 5 MHz	WCDMA
Kawamura [36]	GaAs HBT	27.6	44%	21	3.4 V	-37 dBc @ 5 MHz	WCDMA
Hanington [37]	GaAs MESFET	24	20%	12	Dynamic biasing	-26 dBc @ 1.25 MHz	CDMA

6.3 Efficiency Enhancement

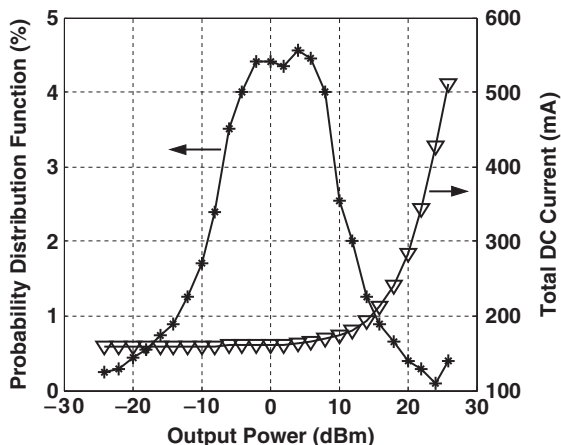
To improve the average power efficiency, it is necessary to reduce the DC current when the PA delivers a low output power. Figure 6.2 shows the DC current consumption of a two-stage class AB power amplifier and a representative CDMA probability distribution function (PDF) as a function of output power.

At low output powers, the power amplifier is operated at a fixed bias current, and therefore, the efficiency is degraded; more than 90% of the output power occurs between -15 dBm and +15 dBm, where the efficiency is low. The average power efficiency is defined as:

$$\langle \eta \rangle = \frac{\langle P_{out} \rangle}{\langle P_{dc} \rangle} = \frac{\int P_{out} p(P_{out}) dP_{out}}{\int \frac{P_{out} p(P_{out}) dP_{out}}{\eta(P_{out})}} \quad (6.1)$$

where P_{out} is the output power, $p(P_{out})$ the probability of output power P_{out} and $\eta(P_{out})$ the efficiency at output power P_{out} . For example, using the PDF above, the

Fig. 6.2 DC current consumption of a two-stage class AB PA with fixed bias and a representative CDMA probability distribution function (PDF)



average power efficiency of a Class AB amplifier is very low – below 2%. One of the main design objectives in this work is to achieve an average power efficiency far better than 2%.

6.3.1 Principle of Dual Dynamic Biasing

As shown in Fig. 6.3, we propose an integrated switched dual dynamic bias (DDB) control that reduces both the DC bias current and voltage without the use of external DC–DC converters. Switched DDB adopts a one-step approach with current re-use in the “low-power” group to reduce the DC voltage across the transistors.

Switched DDB includes dynamic current biasing and dynamic voltage biasing. DCB is achieved by varying the bias current in response to the output power requirements. As shown in Fig. 6.4, there are two groups of SiGe HBT transistors: a high-power group (of 100 devices) and a low-power group (of 20 devices). The high-power group is biased at a current of 110 mA, whereas the low-power group is biased at 22 mA. The switching between different power groups is controlled by low-loss NFET switches on the bases of the HBT transistors. When the PA enters the low power region, the high-power group is switched off and the low-power group is switched on. Therefore, the total bias current is reduced.

On the other hand, DVB is realized by reusing the bias current. As shown in Fig. 6.5, the high-power group is the same as that in Fig. 6.4. The low-power group has two sub-groups and each sub-group has 20 devices. In the high-power group, each transistor is connected in parallel and biased at V_{CC} ; whereas in the low-power group, two sub-groups are series-connected to reuse the common bias current. Each transistor in the sub-group is biased at $V_{CC}/2$, and hence the DC bias

Fig. 6.3 BJT current versus voltage, demonstrating different dynamic biasing strategies (DC/VB: dynamic current/voltage biasing)

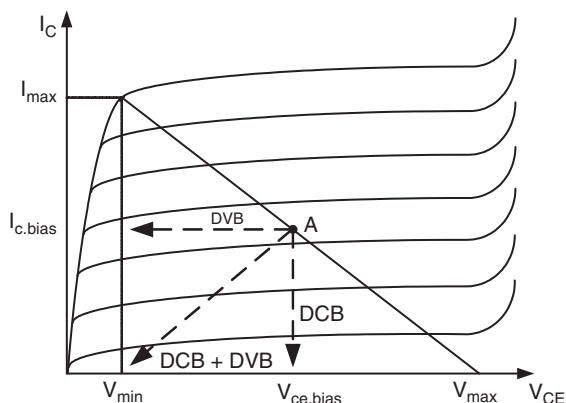
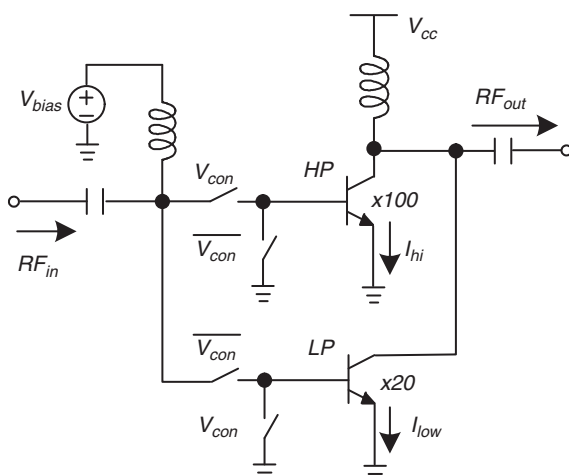


Fig. 6.4 Simplified schematic of a dynamic current bias. $V_{cc} = 3\text{ V}$



voltage is lowered. Combining DCB and DVB, the DC power consumption of the power amplifier is significantly reduced.

Switched DDB technique can be expanded to more than the one power control step used here. Several factors influence the optimum number of power control steps, including the added complexity and parasitics associated with additional steps, the expected gain variation, and the reduced power consumption at the lowest output power levels. For example, based on the simulation results for a SiGe PA shown in Fig. 6.6, a single output stage with a two-step DDB achieves better average power efficiency (8.9%) than its one-step counterpart (7.7%); however, the corresponding circuit complexity increases dramatically and the achievable power gain also drops due to the added parasitics. More steps result in further diminishing returns, so the one-step topology was finally adopted in our prototype PA.

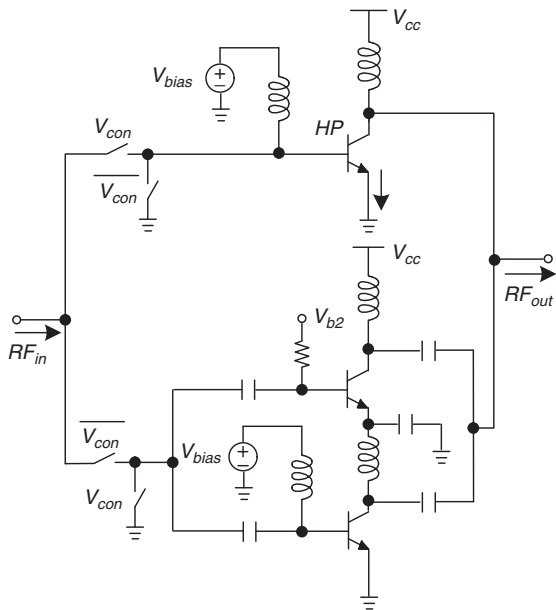


Fig. 6.5 Simplified schematic of a dynamic voltage bias. $V_{cc} = 3\text{ V}$

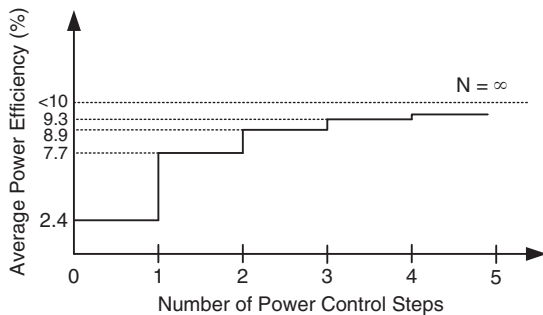


Fig. 6.6 Comparison of average power efficiencies versus the number of power control steps for a SiGe power amplifier

6.3.2 Gain Variation with Power Control

The gain of the PA should remain roughly constant as it switches from low-power to high-power operation¹. To achieve this goal, the following techniques have been employed:

¹ Besides the gain variation, the phase change is another important parameter for EVM. There is a phase jump when the amplifier is switched from the high-power mode to the low-power mode in the DDB operation. Since this phase change is fixed, it is readily compensated during the handset calibration period.

- A constant collector current density is essential for achieving constant gain [11].
- The connection between two power groups adds parasitics to each individual group, which degrades the overall gain. However, this connection decreases the *difference* between the gains in the two power modes since the input impedance of the output stage changes little between high-power and low-power operation.
- The routing line inductance L_{line} from the output of the low-power group to that of the high-power group is optimized to boost the gain in the low-power mode, by partially resonating the output capacitance of the high-power group. Since the output impedance of the low-power device is higher than that of the high-power device, the parasitic capacitance at the output has a larger degrading effect on the low-power gain. By partially resonating the parasitic capacitance in the low-power operation, the gain difference between the two power modes is further reduced.

6.3.3 Circuit Design Considerations

6.3.3.1 Two-Stage Design

A simplified schematic of the prototype two-stage PA with switched DDB is shown in Fig. 6.7. The driver stage, low-power group, and high-power group are all biased in class AB mode to achieve the best trade-off between efficiency and linearity. The driver stage has 20 HBTs and the low-power group has 20 switches and 2×20 HBTs

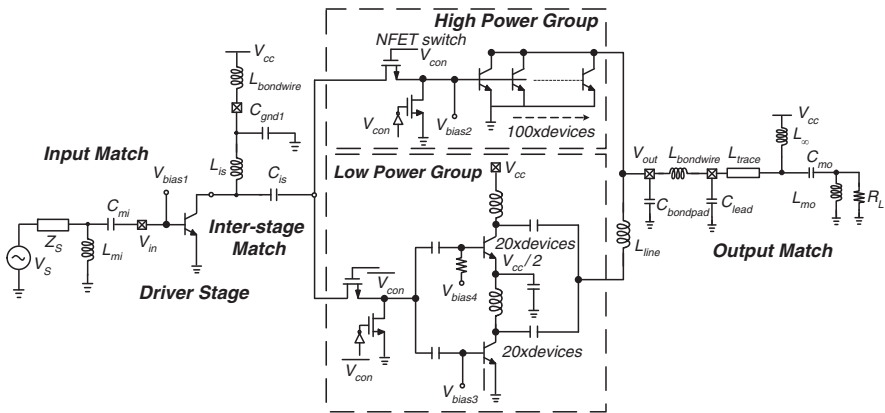


Fig. 6.7 Simplified schematic of a two-stage prototype power amplifier with the dual dynamic biasing in the output stage. $V_{cc} = 3V$

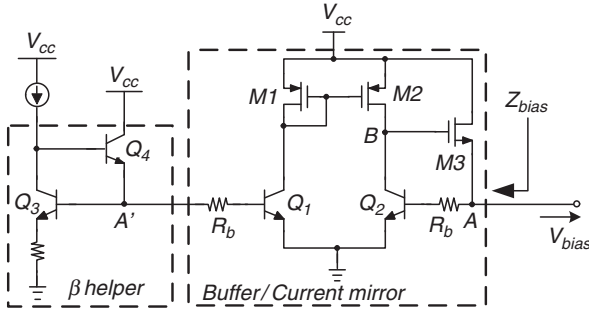


Fig. 6.8 Simplified schematic of the constant voltage bias network

in series. The high-power group uses 100 HBTs and 100 NFET switches. Each HBT emitter is $48 \times 0.44 \mu\text{m}$ and each NFET switch is $45 \mu\text{m}/0.25 \mu\text{m}$. The optimal design of the NFET switches for power gain and 1 dB compression point has been analyzed in [11].

6.3.3.2 Bias Network

To achieve better linearity [38], a low DC base impedance is employed for the driver stage and the high-power group in the output stage. As shown in Fig. 6.8, the bias network is composed of a β helper and a low impedance buffer, to provide a constant voltage bias and to terminate the low-frequency components for improved linearity [39]. The DC impedance is approximately

$$Z_{bias}(\Delta\omega) \approx \frac{1}{g_{m,M3} g_{m,Q2} r_{o,Q2}} \quad (6.2)$$

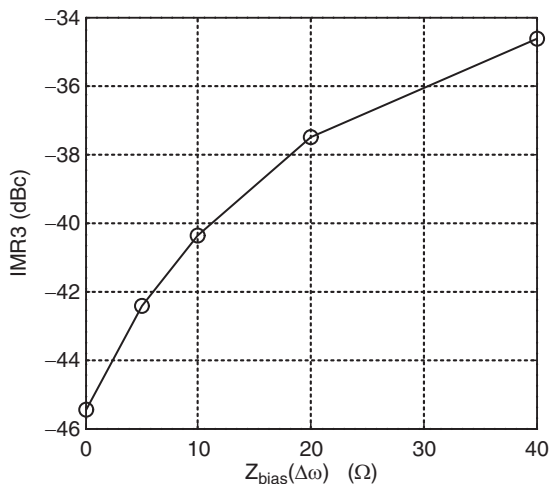
Based on our simulations, $Z_{bias}(\Delta\omega) \approx 0$ for $\Delta\omega \leq 5$ MHz (i.e. the channel bandwidth for WCDMA handset PAs). Figure 6.9 shows the simulated third-order intermodulation ratio IMR3 (i.e. the ratio between the third-order intermodulation product and the fundamental tone) at 12 dBm output power and 10 MHz offset frequency. Clearly, the low frequency termination improves the linearity of the power amplifier.

6.3.3.3 Miscellaneous Details

Ballasting resistors are inserted in the emitter of each transistor to prevent thermal runaway. From [24], a lower bound on R_E to prevent thermal runaway is given by

$$R_E \geq \frac{kT}{qI_C} [(0.05I_C) \theta_{th} V_C - 1] \quad (6.3)$$

Fig. 6.9 Simulated third-order intermodulation ratio versus the bias network impedance



where I_C and V_C are the DC collector bias current and voltage respectively, and $\theta_{th} = \Delta T / I_C V_C$ is the thermal resistance. Assuming that $\theta_{th} = 0.33\text{C/mW}$, $I_C = 110\text{mA}$ and $V_C = 3\text{V}$, we find $R_E \geq 1.1\Omega$ to insure thermal stability.

6.4 Linearity Improvement

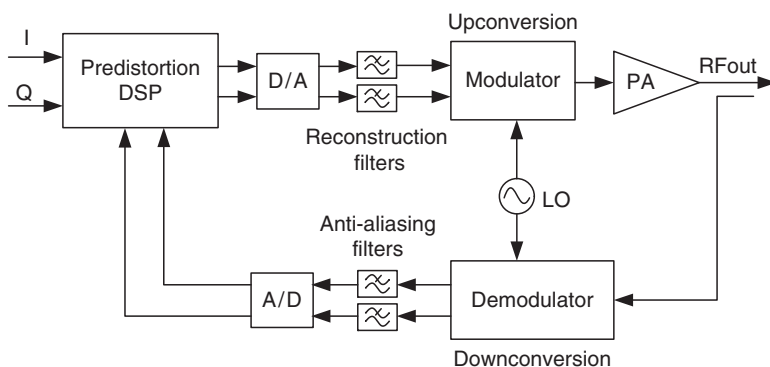
Different PA linearization techniques are compared in Table 6.3. According to their individual features, they are employed in different scenarios.

Memoryless digital predistortion (DP) techniques [18] are attractive for a number of reasons:

- DP can be more economical than feed-forward techniques, which are widely applied in base-station PAs [40]. Even though the cost to implement DP in commercial handsets is still high, the demand for wireless high-data-rate applications should keep lowering the cost.
- DP does not result in the gain loss that is common in analog predistortion techniques [41].
- Memory effects are not strong in handset PAs compared to base-station PAs, so only memoryless distortion, including gain compression (AM/AM distortion) and phase deviation (AM/PM distortion), needs to be considered and the corresponding predistortion algorithm is straightforward.
- Predistortion is implemented with digital signal processing (DSP); therefore it can be performed accurately and flexibly. Predistortion algorithm can be adaptive for multi-mode operation. Therefore, DP is employed to improve the linearity of the prototype WCDMA PA.

Table 6.3 Comparison of different PA linearization techniques

Linearization technique	Linearization performance	Compensation bandwidth	Cost	Comments
Feedback	Moderate	Narrow	Moderate	Stability; reduced gain
Feedforward	Good	Wide	High	Not for handset PAs
Analog/RF predistortion	Low	Wide	Low	Simplest form; reduced gain
Digital predistortion	Moderate	Wide	Moderate	Easy to integrate and control

**Fig. 6.10** Block diagram of memoryless digital predistortion

Therefore, DP is employed to improve the linearity of the prototype PA with switched DDB.

6.4.1 System Topology

A block diagram of the DP system is shown in Fig. 6.10. The DP system down-converts a portion of the RF output signal to an analog IF signal, which is then converted to a digital second IF and fed to the DSP. After comparison with the corresponding input signal during a “training” period, the resulting error amplitude and phase signal are used to adaptively predistort subsequent input I/Q signals. The predistorted digital signal is converted to an analog IF signal and then upconverted to the final RF output. This digital IF approach was recently demonstrated in a low power WCDMA upconverter IC [42].

6.4.2 Predistortion Algorithm

The amount of predistortion is controlled by a look-up table (LUT) in the DSP that characterizes the nonlinearities of the RF PA. For the LUT implementation, a number of predistortion algorithms have been proposed [43–45]. A straightforward solution is mapping predistortion. As shown in Fig. 6.11a, a two-dimensional table in a random access memory (RAM) contains complex predistorting signals. The mapping table accepts the input signal S in Cartesian form as the table address and outputs the corresponding correction value $C(S)$ to predistort the signal S . The predistorted signal \bar{S} is obtained with the sum of S and $C(S)$. The major drawback of this approach is the size of the two-dimensional table (e.g. 2 Mwords in [43]).

To overcome the size issue of the mapping predistortion, Faulkner proposed a polar predistortion [44], which contains two one-dimensional tables controlling gain and phase compensations respectively. As shown in Fig. 6.11b, the input signal is predistorted after the gain factor multiplication and phase rotation. In order to find the gain factor in the gain table, the Cartesian input signals need to be converted into

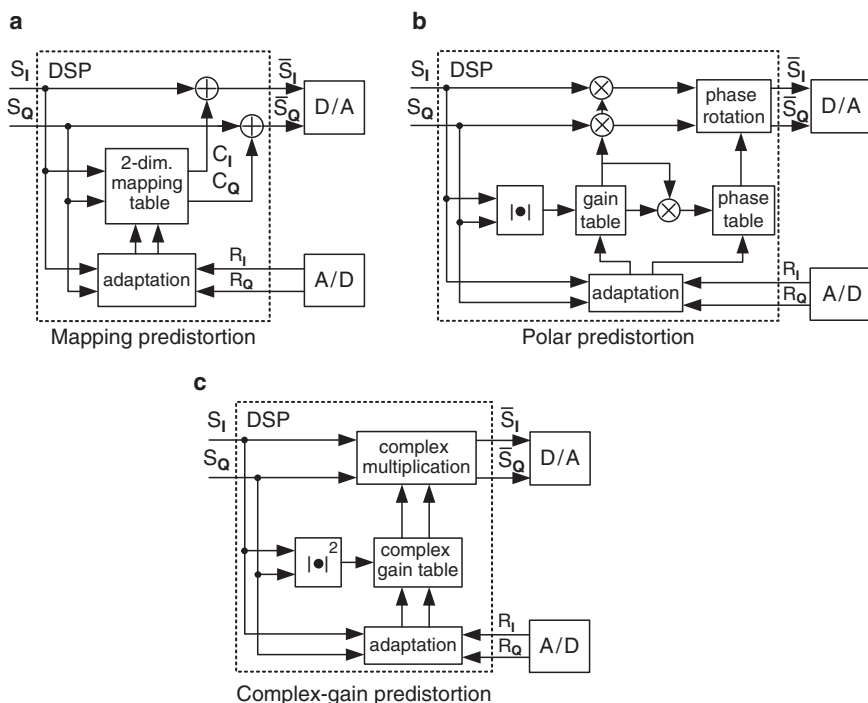


Fig. 6.11 Predistortion algorithm comparison. (a) Mapping predistortion. (b) Polar predistortion. (c) Complex-gain predistortion

the one-dimensional amplitude signal as the address in the gain table. The address of the phase table is obtained by multiplying the amplitude signal with the gain factor. The adaptive correction terms for the gain and phase tables are obtained by comparing the gain and phase differences respectively between the input signal and the sampled output signal. Compared with the mapping counterpart, the polar digital predistortion reduces the table size considerably. The size of the polar table is three orders of magnitude less than that of the mapping counterpart [43].

Cavers [45] presented another approach using a complex-gain table instead of two separate gain and phase tables. As shown in Fig. 6.11c, the input signal is pre-distorted by a complex multiplication between the input signal and the complex gain factor. The address of the complex gain factor is obtained by calculating the squared magnitude of the input signal. The complex-gain approach requires more than three orders of magnitude less memory than its mapping counterpart [43] and reduces convergence time.

6.4.3 Complex-Gain Digital Predistortion

Typically, the baseband AM–AM and AM–PM behavior of the PA can be modeled by a complex polynomial of the form

$$y_n = x_n \sum_{k=1}^m a_k |x_n|^{k-1} = g(x_n) \quad (6.4)$$

where y is the instantaneous complex baseband output, x is the instantaneous complex baseband input to the power amplifier, and the a coefficients are the complex gain of the amplifier. The function $g(x_n)$ denotes the transfer function of the nonlinear PA. Once these a_k coefficients are known, the nonlinearity can be inverted in the digital baseband through a complex series reversion and the nonlinearity will be eliminated. The magnitude correction can be found as [12]:

$$f_A(\alpha) = \frac{g_A^{-1}(G\alpha)}{\alpha} \quad (6.5)$$

where α is the input signal at certain time instant and G is the amplifier gain in the ideal amplification. Consequently, calculating the magnitude correction of the predistortion function is a procedure of rescaling the input magnitude. The final amplifier is capable of yielding the same output power, but at the scaled input power. The phase correction is simply the inverse of the phase error at the rescaled magnitude:

$$f_\theta(\alpha) = -g_\theta(\alpha f_A(\alpha)) \quad (6.6)$$

6.4.4 Practical Issues

From the output of the DSP to the input of the PA, the predistorted signal passes through the transmitter IC (TxIC), which consists of a D/A converter, a reconstruction filter, and a modulator. To maximize the overall system performance, it is critical to analyze and correct the non-idealities of the TxIC, including phase deviation of the reconstruction filter [46] and I/Q imbalance of the modulator [47].

6.4.4.1 Phase Deviation

To simplify the analysis, a two-tone test in a transmitter will be analyzed here. Figure 6.12a shows the case without digital predistortion and with an ideal TxIC, where a third-order intermodulation product (IMD3) occurs at the output of the nonlinear PA. Figure 6.12b demonstrates the case with digital predistortion and an ideal TxIC. Two pre-generated IMD3 products ($2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$) together with two fundamental tones (ω_1 and ω_2) are transmitted to the input of the nonlinear PA. Without any dispersion by the ideal TxIC, the resultant IMD3 becomes zero at the output of the nonlinear power amplifier, due to the cancellation of the digitally generated products and the nonlinear products of the amplifier.

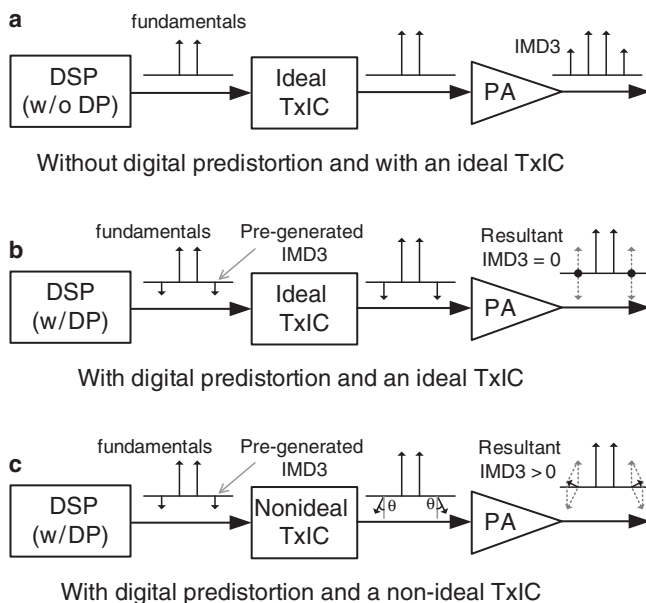


Fig. 6.12 Visualization of a two-tone test in a transmitter. (a) Without digital predistortion and with an ideal TxIC. (b) With digital predistortion and an ideal TxIC. (c) With digital predistortion and a non-ideal TxIC

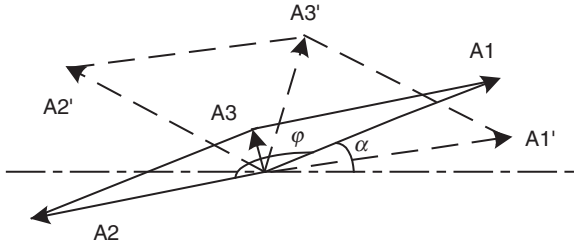


Fig. 6.13 Vector analysis of IMD3 cancellation

In the case with a *non-ideal* TxIC, the reconstruction filter yields the major phase deviation in the TxIC. As shown in Fig. 6.12c, the resultant IMD3 will be greater than zero and hence the linearization performance is degraded.

A simple vector analysis can be used to illustrate this cancellation degradation problem [48]. In Fig. 6.13, vectors A1 and A2 are the IMD3 and anti-IMD3 sidebands generated by the PA and the PD respectively for the case without a filter. The initial phase of A1 is α and the phase difference of A1 and A2 is φ . The amplitude of the IMD3 after cancellation is A3:

$$|A_3| = \sqrt{|A_1|^2 + |A_2|^2 + 2|A_1||A_2|\cos\varphi} \quad (6.7)$$

The filter can alter both the amplitude and phase of the fundamental and predistorted sideband components. For simplicity, we make several assumptions. First, the base-band amplitude variation is ignored, which is reasonable since most filters for this application will have small passband ripple – usually less than 0.5 dB. Second, there are no PM to AM and PM to PM distortions in the PA, which is also reasonable, if the PA is represented by a quasi-memoryless model. With these assumptions, the filtered IMD3 and anti-IMD3 vectors A1' and A2' have the same amplitude as A1 and A2, but the phase difference between them is changed to $(\varphi - \theta)$ instead of φ due to the different phase shift θ within the passband of the filter. The amplitude of the new IMD3 after cancellation is A3':

$$|A'_3| = \sqrt{|A'_1|^2 + |A'_2|^2 + 2|A'_1||A'_2|\cos(\varphi - \theta)} \quad (6.8)$$

Using (6.7) and (6.8), the cancellation degeneration due to the phase dispersion can be quantified.

To maximize the predistortion system performance, it is necessary to correct the phase deviation due to the reconstruction filter in the predistortion algorithm. In [49], this “memory effect” has been effectively compensated with an iterative technique of successive approximations.

6.4.4.2 I/Q Imbalance

It is important to maintain equal gains and exactly 90° phase difference between the in-phase (I) and quadrature (Q) paths of the upconversion chain. Any disparity in gain and phase will respectively yield amplitude- and phase-dependent amplitude (AM–AM and PM–AM) distortion. Since an amplitude-based predistortion algorithm will *not* correct phase-dependent errors, I/Q imbalance needs to be separately corrected in the DSP.

I/Q imbalance can be measured by transmitting a sine wave on the I path ($A_{LO} \sin \omega_{LO} t$) and a cosine wave on the Q path ($(A_{LO} + \Delta A_{LO}) \cos(\omega_{LO} t + \Delta \phi)$), where ΔA_{LO} and $\Delta \phi$ are the amplitude and phase errors of the Q path LO signal. Therefore, at the output the upconverter, the output amplitude at the desired frequency $\omega_{LO} - \omega_{IF}$ is found to be

$$\begin{aligned} A_{LO-IF}^2 &= \left(\frac{A_{LO}}{2} + \frac{A_{LO} + \Delta A_{LO}}{2} \cos \Delta \phi \right)^2 + \left(\frac{A_{LO} + \Delta A_{LO}}{2} \sin \Delta \phi \right)^2 \quad (6.9) \\ &= A_{LO}^2 (1 + \Delta A_{LO-IF}) \end{aligned}$$

where the output amplitude error ΔA_{LO-IF} is

$$\Delta A_{LO-IF} = \frac{\cos \Delta \phi - 1}{2} + \frac{\Delta A_{LO}}{2A_{LO}} (1 + \cos \Delta \phi) \quad (6.10)$$

which includes PM–AM distortion as well as AM–AM distortion. Furthermore, the output phase ϕ_{LO-IF} is given by

$$\phi_{LO-IF} = \tan^{-1} \left(\frac{-\frac{A_{LO} + \Delta A_{LO}}{2} \sin \Delta \phi}{\frac{A_{LO}}{2} + \frac{A_{LO} + \Delta A_{LO}}{2} \cos \Delta \phi} \right) \quad (6.11)$$

Consequently, the output phase error is nonlinear and comprises both AM–PM and PM–PM distortions. In general, I/Q imbalance is compensated digitally [50] with less expense and higher precision. A simple criterion for the compensation algorithm is employed by observing the output amplitude at the image frequency $\omega_{LO} + \omega_{IF}$, which is readily found to be

$$A_{LO+IF}^2 = \left(-\frac{A_{LO}}{2} + \frac{A_{LO} + \Delta A_{LO}}{2} \cos \Delta \phi \right)^2 + \left(\frac{A_{LO} + \Delta A_{LO}}{2} \sin \Delta \phi \right)^2 \quad (6.12)$$

As shown in Fig. 6.14, there are two tones at $\omega_{LO} - \omega_{IF}$ and $\omega_{LO} + \omega_{IF}$ respectively on the resulting output spectrum, with the first being much larger than the second. Assuming that I/Q imbalance has been compensated perfectly (i.e. both ΔA_{LO} and $\Delta \phi$ are zero) and there are no other modulator imperfections, the second tone at $\omega_{LO} + \omega_{IF}$ will be completely suppressed.

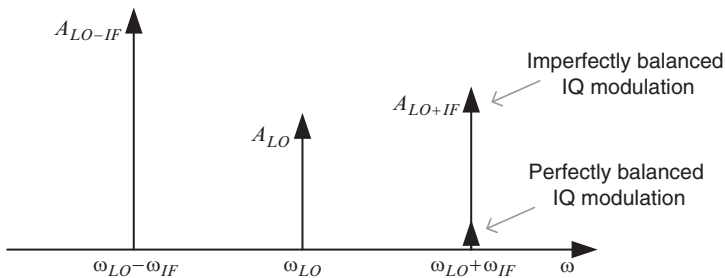


Fig. 6.14 Output spectrum showing imperfectly and perfectly balanced IQ modulations

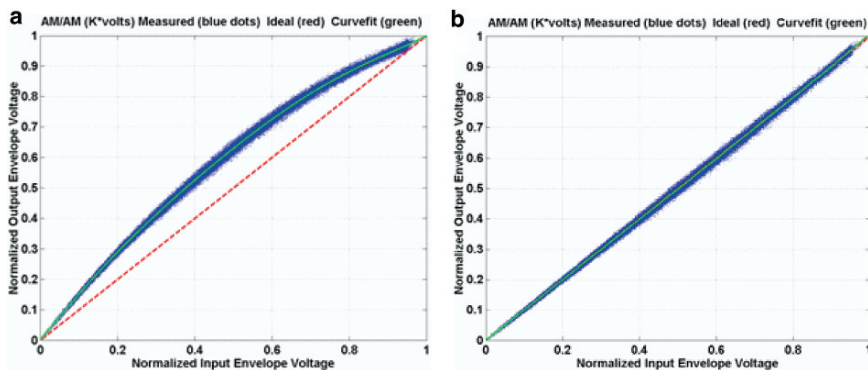


Fig. 6.15 Measured AM–AM response of CDMA power amplifier (a) before memoryless predistortion (b) after memoryless predistortion

6.4.5 Experimental Validation

The effectiveness of digital predistortion is checked by comparing the AM–AM relationship between input and output envelope amplitudes. Figure 6.15 shows that with digital predistortion, the output amplitude has been thoroughly linearized. Based on the above discussions, digital predistortion is an effective linearization technique for handset PAs.

6.5 Measurement Results

The prototype two-stage PA with the dual dynamic bias control was fabricated in a 0.25 μm SiGe BiCMOS process [51]. The die area is 1 × 1.8 mm. A packaged die with bonding wires is shown in Fig. 6.16. The Micro Lead Frame (MLF20) package

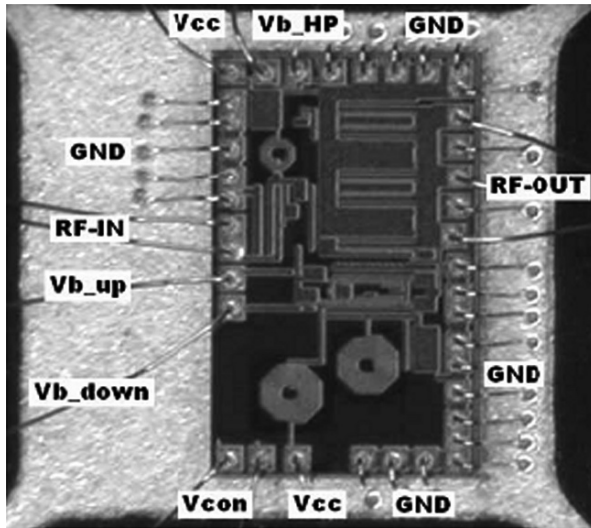


Fig. 6.16 Photograph of the packaged die of the prototype power amplifier. The die area is 1.8 mm^2

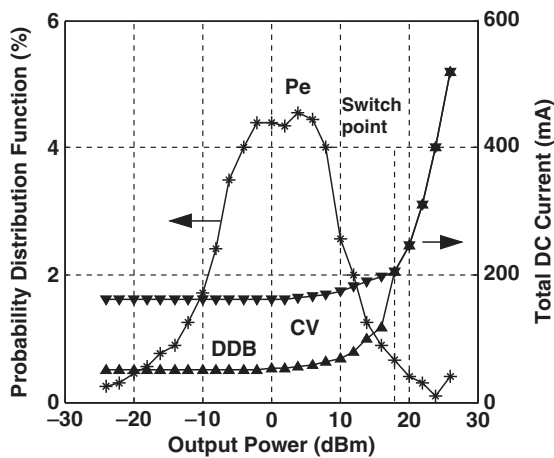


Fig. 6.17 Output power probability distribution function and measured DC current comparison for different biasing techniques (CV: constant voltage biasing; DDB: dual dynamic biasing). The switch point from high-power mode to low-power mode is 16 dBm. $V_{cc} = 3 \text{ V}$

is used due to its small body size and short bond-wires along with excellent electrical and thermal performance. Figure 6.17 compares the measured DC currents for different biasing approaches for the two-stage PA. These approaches include constant base voltage (CV) biasing with a fixed number of parallel transistors (the traditional Class-AB approach) and switched DDB proposed here. Obviously, DDB reduces the DC current at the low-power region significantly compared to the traditional approach.

As shown in Fig. 6.18, average power efficiencies are calculated from (6.1) to be: 1.9% for CV biasing, 3.8% for the Dynamic Current Biasing (DCB) approach proposed earlier [11]² and 5.0% for DDB, all in two-stage PAs. This verifies that DDB does achieve dramatically improved average efficiency for handset applications.

For multi-mode applications, digital predistortion is adopted to improve the linearity of the power amplifier. WCDMA reverse-link signal is taken at Tx baseband input and the linearity of the power amplifier is measured in terms of ACPR for two cases: with and without digital predistortion. The measured results at the output power of 22.4 dBm are shown in Fig. 6.19, where the ACPR performance of the power amplifier is improved by 10 dB with digital predistortion.

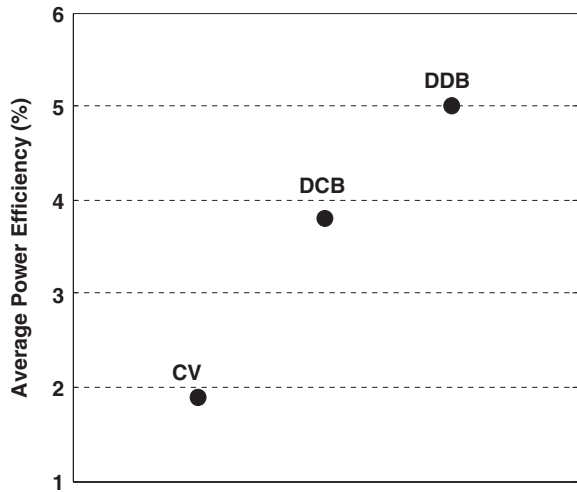


Fig. 6.18 Average power efficiency comparison (CV: constant base voltage; DCB: dynamic current biasing; DDB: dual dynamic bias) in a two-stage power amplifier

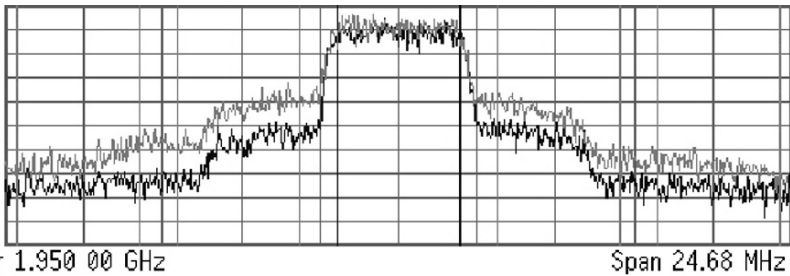


Fig. 6.19 Comparison of measured spectrum with DP (dark line) and without DP (light line)

² Note that the average power efficiency reported in [16] is based on a single output stage; therefore it is higher than what is reported here based on a two-stage power amplifier, where the driver stage adopts the classical CV bias considering the gain degradation and circuit complexity.

The ACPR measurement at 5 MHz is shown in Fig. 6.20. Note that it is only necessary to utilize the predistortion in the high-power mode, so DP is not applied in the low-power mode; this is the reason for the large discontinuity in ACPR in the “after DP” curve. The ACPR is improved by at least 8 dB with digital predistortion, and the maximum output power satisfying the WCDMA linearity specification is improved from 22.4 dBm to 26 dBm. This satisfies the WCDMA Class 3 requirement of maximum output power. Correspondingly, the peak PAE is improved by 60% (from 17% to 27%).

Figure 6.21 shows the measured gain of the DDB PA with DP. The gain change for DDB is less than 1.8 dB, which is much more constant than the case if the power amplifier is operated with dynamic bias without changing the device size.

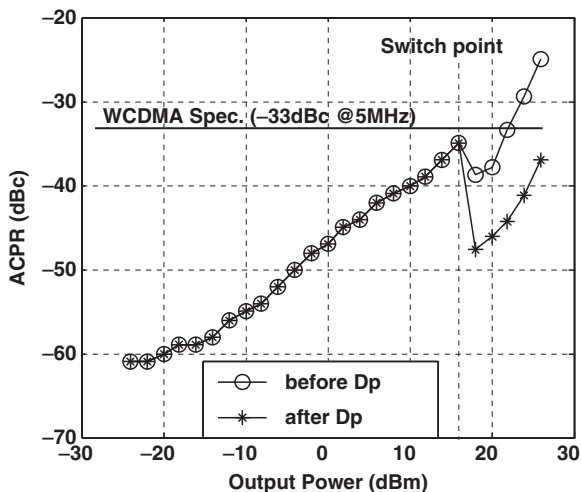


Fig. 6.20 Measured ACPRs of the DDB power amplifier with digital predistortion (DP) (before DP and after DP)

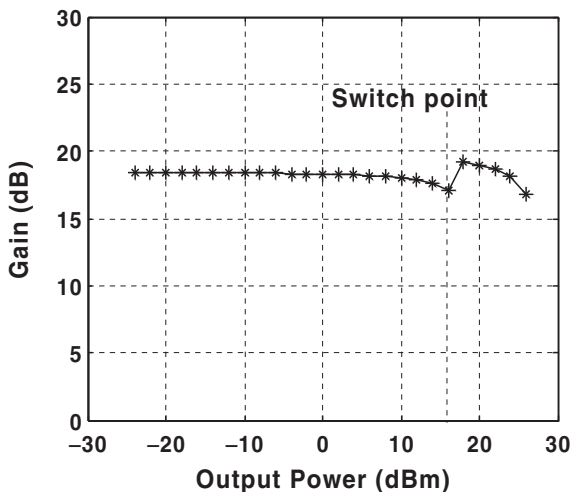


Fig. 6.21 Measured gain of the DDB power amplifier with DP

6.6 Summary

In this chapter we compared different power amplifier technologies with various requirements and characteristics and reviewed recently reported CDMA/WCDMA handset power amplifiers in different technologies. We proposed the switched dual dynamic bias technique for power amplifier efficiency enhancement. We also presented different linearization techniques in power amplifiers and the complex-gain digital predistortion was employed to improve the linearity of the power amplifier. Measurement results demonstrated the feasibility to achieve high-efficiency and high-linearity power amplifiers for multi-mode handset applications.

Acknowledgment The authors would like to acknowledge valuable discussions with Professor Peter Asbeck of UCSD, Dr. Prasad Gudem of Qualcomm and Dr. Mingyuan Li of Broadcom, and the support of the Member Companies of the UCSD Center for Wireless Communications, the California Institute for Telecommunications and Information Technology and the University of California Discovery Grant Program.

References

1. T. Shimizu, Y. Nunogawa, T. Furuya, S. Yamada, I. Yoshida, and H. Masao, "A small GSM power amplifier module using Si-LDMOS driver MMIC," *IEEE International Solid-State Circuits Conference*, pp. 196–522, Feb. 2004.
2. K. Mertens and M. Steyaert, "A 700-MHz 1-W fully differential CMOS class-E power amplifier," *IEEE Journal of Solid-State Circuits*, vol. 37, pp. 137–141, Feb. 2002.
3. *Technical Specification 3GPP TS 25.101 V6.1.0*, June 2003.
4. S. Maeng, S. Chun, J. Lee, C. Lee, K. Youn, and H. Park, "A GaAs power amplifier for 3.3 V CDMA/AMPS dual-mode cellular phones," *IEEE Transactions on Microwave Theory and Techniques*, vol. 43, pp. 2839–2844, Dec. 1995.
5. T. Iwai, K. Kebayashi, Y. Nakasha, T. Miyashita, S. Ohara, and K. Joshin, "42% high-efficiency two-stage HBT power-amplifier MMIC for W-CDMA cellular phone systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 48, pp. 2567–2572, Dec. 2000.
6. J.F. Sevic, "Statistical characterization of RF power amplifier efficiency for CDMA wireless communication systems," in *1997 Wireless Communications Conference*, pp. 110–113.
7. D. Dening, Setting Bias Points for Linear RF Amplifiers. *Microwaves & RF* [online]. Available: <http://www.mwrf.com>, June 2002.
8. T. Fowler, K. Burger, N.S. Cheng, A. Samelis, E. Enobakhare, and S. Rohlfing, "Efficiency Improvement Techniques at Low Power Levels for Linear CDMA and WCDMA Power Amplifiers," in *2002 IEEE Radio Frequency Integrated Circuits Symposium*, pp. 41–44.
9. G. Hanington, P. Chen, P.M. Asbeck, and L.E. Larson, "High-efficiency power amplifier using dynamic power-supply voltage for CDMA applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, pp. 1471–1476, Aug. 1999.
10. F. Wang A. Yang, D. Kimball, L.E. Larson, and P.M. Asbeck, "Design of wide-bandwidth envelope-tracking power amplifiers for OFDM applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, pp. 1244–1255, April 2005.
11. J. Deng, P.S. Gudem, L.E. Larson, and P.M. Asbeck, "A high average-efficiency SiGe HBT power amplifier for WCDMA handset applications" *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, pp. 529–537, Feb. 2005.

12. J. Deng, P.S. Gudem, L.E. Larson, D.F. Kimball, and P.M. Asbeck, "A SiGe PA with dual dynamic bias control and memoryless digital predistortion for WCDMA handset applications" *IEEE Journal of Solid-State Circuits*, vol. 41, pp. 1210–1221, May 2006.
13. M.R. Moazzam and C.S. Aitchison, "A low third order intermodulation amplifier with harmonic feedback circuitry," *IEEE MTT-S International Microwave Symposium*, pp. 827–830, June 1996.
14. V. Petrovic, "Reduction of spurious emission from radio transmitters by means of modulation feedback," *IEE Conference on Radio Spectrum Conservation Techniques*, pp. 44–49, Sept. 1983.
15. H.S. Black, "Translating system," U.S. Patent 1,686,792, Oct. 9, 1928.
16. J. Presa, J. Legarda, H. Solar, J. Melendez, A. Munoz, and A. Garcia-Alonso, "An adaptive feedforward power amplifier for UMTS transmitters," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 2715–2719, Sept. 2004.
17. P.B. Kenington, *High-Linearity RF Amplifier Design*, Artech House, 2000.
18. A. Mansell and A. Bateman, "Practical Implementation Issues For Adaptive Predistortion Transmitter Linearisation," *IEE Colloquium on Linear RF Amplifiers and Transmitters*, pp. 5/1–7, April 1994.
19. K. Nellis and P.J. Zampardi, "A comparison of linear handset power amplifiers in different bipolar technologies," *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 1746–1754, Oct. 2004.
20. R. Jos, "Technology developments driving an evolution of cellular phone power amplifiers to integrated RF front-end modules," *IEEE Journal of Solid-State Circuits*, vol. 36, pp. 1382–1389, Sept. 2001.
21. T. Sowlati and D. Leenaerts, "A 2.4-GHz 0.18 μ m CMOS self-biased cascode power amplifier," *IEEE Journal of Solid-State Circuits*, vol. 38, pp. 1318–1324, Aug. 2003.
22. W. Bakalski, W. Simburger, R. Thuringer, A. Vasylyev, and A.L. Scholtz, "A fully integrated 5.3-GHz 2.4-V 0.3-W SiGe bipolar power amplifier with 50- Ω output," *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 1006–1014, July 2004.
23. J.B. Johnson, A.J. Joseph, D.C. Sheridan, R.M. Maladi, P.O. Brandt, J. Persson, J. Andersson, A. Bjorneklett, U. Persson, F. Abasi, and L. Tilly, "Silicon-germanium BiCMOS HBT technology for wireless power amplifier applications," *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 1605–1614, Oct. 2004.
24. M.G. Adlerstein, "Thermal stability of emitter ballasted HBT's," *IEEE Transactions on Electron Devices*, vol. 45, pp. 1653–1655, Aug. 1998.
25. K. Yamamoto, T. Moriwaki, H. Otsuka, N. Ogawa, K. Maemura, and T. Shimura, "A CDMA InGaP/GaAs-HBT MMIC Power Amplifier Module Operating With a Low Reference Voltage of 2.4 V," *IEEE Journal of Solid-State Circuits*, vol. 42, pp. 1282–1290, June 2007.
26. J.H. Kim, K.Y. Kim, Y.H. Choi, and C.S. Park, "A power efficient W-CDMA smart power amplifier with emitter area adjusted for output power levels," *IEEE MTT-S International Microwave Symposium*, vol. 2, pp. 1163–1166, June 2004.
27. C. Wang, M. Vaidyanathan, and L.E. Larson, "A Capacitance-Compensation Technique for Improved Linearity in CMOS Class-AB Power Amplifiers," *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 1927–1937, Nov. 2004.
28. Y.J. Jeon, H.W. Kim, H.T. Kim, G.H. Ryu, J.Y. Choi, K. Kim, S.E. Sung, and B. Oh, "A highly efficient CDMA power amplifier based on parallel amplification architecture," *IEEE Microwave and Wireless Components Letters*, vol. 14, pp. 401–403, Sept. 2004.
29. Y.S. Noh and C.S. Park, "An intelligent power amplifier MMIC using a new adaptive bias control circuit for W-CDMA applications," *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 967–970, June 2004.
30. J. Staudinger, "An overview of efficiency enhancements with application to linear handset power amplifiers," *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*, pp. 45–48, June 2002.
31. G. Hau, T.B. Nishimura, and N. Iwata, "A highly efficient linearized wide-band CDMA handset power amplifier based on predistortion under various bias conditions," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, pp. 1194–1201, June 2001.

32. S. Luo and T. Sowlati, "A monolithic Si PCS-CDMA power amplifier with 30% PAE at 1.9 GHz using a novel biasing scheme," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, pp. 1552–1557, Sept. 2001.
33. V.T.S. Vintola, M.J. Matilainen, S.J.K. Kalajo, and E.A. Jarvinen, "Variable-gain power amplifier for mobile WCDMA applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, pp. 2464–2471, Dec. 2001.
34. P.D. Tseng, L. Zhang, G. Gao, and M. Chang, "A 3-V monolithic SiGe HBT power amplifier for dual-mode (CDMA/AMPS) cellular handset applications," *IEEE Journal of Solid-State Circuits*, vol. 35, pp. 1338–1344, Sept. 2000.
35. T. Iwai, K. Kebayashi, Y. Nakasha, T. Miyashita, S. Ohara, and K. Joshin, "42% high-efficiency two-stage HBT power-amplifier MMIC for W-CDMA cellular phone systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 48, pp. 2567–2572, Dec. 2000.
36. H. Kawamura, K. Sakuno, T. Hasegawa, M. Hasegawa, H. Koh, and H. Sato, "A miniature 44% efficiency GaAs HBT power amplifier MMIC for the W-CDMA application," *Gallium Arsenide Integrated Circuit (GaAs IC) Symposium*, pp. 25–28, Nov. 2000.
37. G. Hanington, P.F. Chen, P.M. Asbeck, and L.E. Larson, "High-efficiency power amplifier using dynamic power-supply voltage for CDMA applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, pp. 1471–1476, Aug. 1999.
38. P. Gudem and L.E. Larson, UCSD Wireless Communication Circuit Design, Course Materials, 2002.
39. V. Aparin and C. Persico, "Effect of out-of-band terminations on intermodulation distortion in common-emitter circuits," in *1999 IEEE MTT-S International Microwave Symposium*, vol. 3, pp. 977–980.
40. K. Cho, J. Kim, and S. Stapleton, "A highly efficient Doherty feedforward linear power amplifier for W-CDMA base-station applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, pp. 292–300, Jan. 2005.
41. K. Yamauchi, M. Nakayama, Y. Ikeda, H. Nakaguro, N. Kadowaki, and T. Araki, "An 18 GHz-band MMIC linearizer using a parallel diode with a bias feed resistance and a parallel capacitor," in *2000 IEEE MTT-S International Microwave Symposium*, vol. 3, pp. 1507–1510.
42. V. Leung, L.E. Larson, and P.S. Gudem, "Digital-IF WCDMA handset transmitter IC in 0.25μm SiGe BiCMOS," *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 2215–2225, Dec. 2004.
43. Y. Nagata, "Linear amplification technique for digital mobile communications," in *1989 IEEE Vehicular Technology Conference*, pp. 159–164.
44. M. Faulkner and M. Johansson, "Adaptive linearization using predistortion-experimental results," *IEEE Transactions on Vehicular Technology*, vol. 43, pp. 323–332, May 1994.
45. J.K. Cavers, "Amplifier linearization using a digital predistorter with fast adaptation and low memory requirements," *IEEE Transactions on Vehicular Technology*, vol. 39, pp. 374–382, Nov. 1990.
46. L. Sundstrom, M. Faulkner, and M. Johansson, "Effects of reconstruction filters in digital predistortion linearizers for RF power amplifiers," *IEEE Transactions on Vehicular Technology*, vol. 44, pp. 131–139, Feb. 1995.
47. L. Sundstrom, *Digital RF Power Amplifier Linearizers – Analysis and Design*, PhD Thesis, Lund University, Aug. 1995.
48. M. Li, J. Deng, L. Larson, and P. Asbeck, "Nonideal Effects of Reconstruction Filter and I/Q Imbalance in Digital Predistortion," *2006 IEEE Radio and Wireless Symposium*, pp. 259–262.
49. P. Draxler, J. Deng, D. Kimball, I. Langmore, and P.M. Asbeck, "Memory Effect Evaluation and Predistortion of Power Amplifiers," in *2005 IEEE MTT-S International Microwave Symposium*.
50. J. Glas, "Digital I/Q imbalance compensation in a low-IF receiver," in *1998 IEEE Global Telecommunications Conference*, pp. 1461–1466.
51. IBM 6HP BiCMOS process, http://www-3.ibm.com/chips/techlib/techlib.nsf/products/BiCMOS_6HP.

Chapter 7

Polyphase Multipath Circuits for Cognitive Radio and Flexible Multi-phase Clock Generation

Eric A.M. Klumperink, Xiang Gao, and Bram Nauta

7.1 Introduction

Cognitive radios aim at exploiting the scarcely available radio spectrum in a smart flexible way. Traditional TV bands between 50 and 900 MHz are currently being freed for new applications. New licensed users are planned (e.g. DVB-H), but in addition new ideas for more flexible use of the spectrum are explored [1]. For higher frequencies similar ideas are developed. In general, regulatory organizations seem to move in the direction of providing more freedom to new standards, where only a minimum set of requirements are enforced. E.g. regulations might allow to exploit white spectrum, where “Detect And Avoid” rules are defined (e.g. response times, maximum interference levels to incumbent services). This will lead to new radio systems with different requirements on the radio software and hardware. In this chapter we will mainly focus on the impact of cognitive radio system requirements on the physical layer (PHY), and especially the radio frequency hardware. Flexible multi-phase clocking will turn out to play a crucial role, and will be discussed in detail.

To allow for flexible spectrum access, a flexible radio hardware platform is desired, allowing for flexible choice of the radio frequency depending on free available spectrum. Traditional radio hardware is primarily optimized for cost and low power, but not for flexibility. Low power is often achieved using inductors and capacitors in resonating circuits with a high quality factor, dissipating only a fraction of the maximum energy stored in the reactive components. However, such circuits only work effectively in a narrow band around their resonance frequency, and are hence application specific for a certain band. Micro-Electrical-Mechanical system (MEMS) technology may help to relax this problem; however for reasons of cost and form factor fully integrated solutions in mainstream CMOS technology are preferred if feasible. Thus we focus in this chapter on CMOS circuits and IC architectures.

E.A.M. Klumperink (✉), X. Gao, and B. Nauta
IC-Design Group, CTIT, University of Twente, Enschede, The Netherlands
e-mail: {e.a.m.klumperink; x.gao; b.nauta}@utwente.nl

We will analyze the desired functionality of the radio interface for dynamic spectrum access, and look at some feasibility bottlenecks induced by CMOS circuit properties, like timing jitter, nonlinearity and time-variance. Some possible solution directions are reviewed, especially a recently proposed polyphase multipath technique. This technique allows for realizing a highly flexible radio transmitter for the DC-2.4 GHz range on a CMOS chip without dedicated filters. It requires multi-phase clocks for which the phase-accuracy is critical. Two competing techniques to realize such clocks, one based on a Shift Register (SR) and the other on a Delay Locked Loop (DLL), are discussed in the second half of this chapter, to show that SR-based clocking has fundamental advantages.

7.2 Flexible RX/RFS: Not Just an ADC

Figure 7.1 shows a high level functional block schematic of a cognitive radio. It consists of an antenna connected to a radio receiver (RX), a radio transmitter (TX) and a Radio Frequency Scanner (RFS). A Baseband Processing and Control unit processes the spectral information, and decides which frequency is free for use. It controls the frequency synthesizer to generate the desired radio frequency carrier, sends bits to the TX and receives bits from the RX.

Ideally, a cognitive radio should be able to communicate wherever free spectrum is available, i.e. be very flexible in terms of the transmit frequency. This suggests a wideband radio receiver should be used for detecting free spectrum and receiving data, in contrast to traditional narrowband radio systems. For maximum flexibility, radio signal processing should be done in the digital domain. On a high abstraction

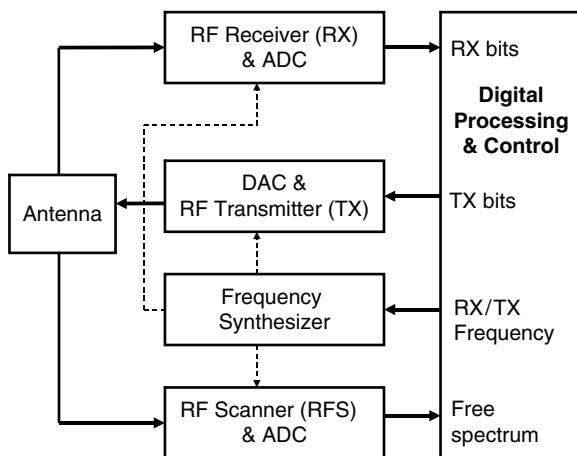


Fig. 7.1 Block diagram of a cognitive radio system for dynamic spectrum access

level, a cognitive radio can then be considered as an A/D Converter (ADC) for the RX and RF Scanner blocks, and a D/A Converter (DAC) for the TX block.

To judge the feasibility of a wideband ADC based receiver, data from Walden's overview paper on ADCs is useful [2]. Consider a mobile radio communication receiver operating at popular radio frequencies between 0.05–6 GHz. Typical transmit power levels for mobile radio standards are in the range of 10 mW up to more than 1 W. The radio path-loss strongly varies from case to case, but it is quite common to receive radio antenna voltages in the range from 1 μ V up to 100 mV. To detect a weak 1 μ V signal, in the presence of a 100 mV interferer, we need an ADC with more than $100 \text{ mV} / 1 \mu\text{V} = 100,000$ detection levels, i.e. roughly 2^{16} levels (16 bits). To observe 5 GHz signals, the ADC should at least take ten Giga samples every second. Assuming for a moment this is technically feasible, at a (rather optimistic) energy of 1 pJ per conversion [2], this leads to a power consumption of $10^{10} \text{ samples/second} \times 2^{16} \text{ levels} \times 10^{-12} \text{ J} \approx 1 \text{ kW}$! The energy per conversion decreases only slowly over time because analog accuracy requirements are involved, which do not benefit much from Moore's law. Note also that the actual radio bandwidth of interest is typically orders of magnitude lower than the radio-carrier frequency. This makes "full-Nyquist" A/D conversion really overkill, and a waste of power, even if it would become technically feasible. Thus we feel there is a need for architectural innovations to make highly flexible cognitive radio systems feasible.

A more realistic and still reasonably flexible approach is to down-convert an RF signal of interest to DC ("zero-IF architecture"), reduce its bandwidth and dynamic range by low-pass filtering and then do the A/D conversion at a rate and a resolution which are feasible at 10–100 mW A/D converter power. Recently a software defined front-end using this approach for the 500 MHz–5 GHz band has been proposed [3]. It uses a wideband low noise amplifier exploiting thermal noise cancellation [4], followed by a highly linear passive down-conversion mixer. However, as there is hardly any RF pre-filtering, the linearity requirements on the RF front-end are very high. Moreover, wideband down-converters using hard-switched mixers are plagued by spurious responses, i.e. they do not only down-convert the wanted RF-band, but also its harmonics. Thus harmonic rejection mixers are needed, e.g. as proposed in [3, 5]. We will address this harmonic rejection mixing later in this chapter when dealing with upconversion mixers.

7.3 Sampling Clock Jitter Requirements

Instead of a mixer, a sampler can also be used for frequency down-conversion. Whereas full Nyquist rate A/D conversion of GHz signals is currently far from feasible, sampling at GHz rates *without* high resolution *quantization* is practical, as demonstrated for a Bluetooth and GSM receiver [6]. These receivers sample the antenna signal at RF and then process it in the charge domain via passive switched

capacitor circuits. Via decimation with internal anti-alias filtering, the sample rate is reduced to a sufficiently low rate to do A/D conversion at acceptable power consumption [6].

The sampling at RF might surprise people who work on low jitter sampling clocks for high-speed ADCs, where clock jitter requirements are increasingly becoming a feasibility bottleneck. This is because timing uncertainty shifts the sampling moments, introducing significant amplitude errors especially for high-amplitude high-frequency signals. To keep these errors from degrading the resolution of the ADC, an extremely low RMS-jitter of less than 11 fs would be needed for an 11 bit ADC sampling a 6 GHz full swing sine wave signal [7].

Fortunately, for radio receiver applications, sampling jitter turns out to be much less harmful. This is because radio signals are narrowband in nature, so *only the noise level in the wanted channel band* is relevant. Jitter in a sampling clock introduces noise at the output of the sampler which strongly varies with frequency and is mainly concentrated around strong high-frequency interferers [7]. The roll-off with frequency distance from the interferer depends on the shape of the phase noise spectrum of the sampling clock. Overall, the requirement on the sampling clock jitter is close to what is needed for traditional mixer based receiver systems limited by reciprocal mixing [7]. Calculation for a Bluetooth receiver shows that 1.3 ps RMS-jitter can be accepted, which is more than *two orders of magnitude easier than corresponding ADC clock jitter specs* [7]. Thus jitter is not as big a problem as often thought, opening the door for radio architectures exploiting high-speed sampling like in [6]. Still, if no or not enough RF-filtering is used, RF signals at harmonics of the sampling clock will again be downconverted and will interfere with the desired signal. Thus harmonic rejection techniques are needed, e.g. as proposed in [8].

7.4 Flexible TX: Not Just a DAC

Realizing a flexible transmitter using a DAC seems possible in principle, as the dynamic range of a transmitted signal is typically significantly lower than the dynamic range of a received signal. However, apart from the useful TX-signal, many other spurious components may be produced. As an integrated radio transmitter should produce significant output power, typically in the range of milli-Watts up to a few Watt, power drivers and power amplifier circuits with transistors working at large signal swings are used. Thus non-linearity of the transistors plays an important role, resulting in harmonics (see Fig. 7.2) and intermodulation distortion products at many unwanted frequencies [9]. As the power efficiency of most amplifiers increases for higher signal swings, it is desirable to drive the amplifiers to a level close to their compression point. However, in practice significant “back off” is needed [10] to suppress distortion products sufficiently at the cost of efficiency.

Apart from nonlinearity, a time-variant transfer function can also introduce many unwanted frequency components. Ideal DACs and hard-switched mixers can be

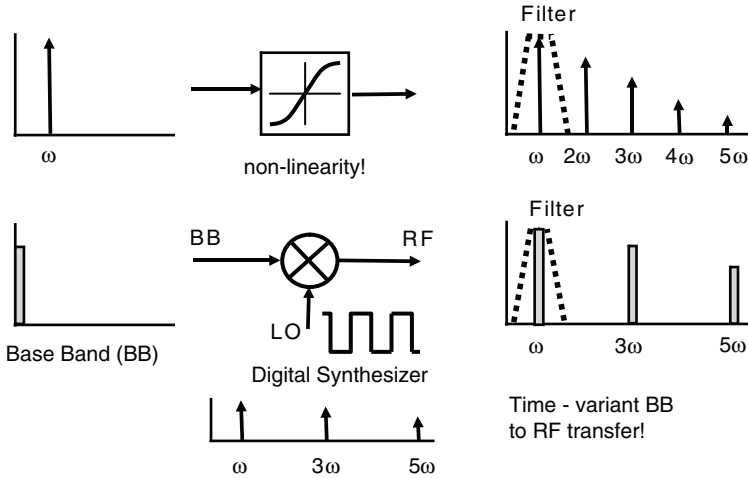


Fig. 7.2 Nonlinearity and time-variance due to switched mixers generate unwanted spectral components, which are traditionally removed by dedicated band-pass filters

modeled as linear time-variant circuits, with a linear transfer from input to output, which changes instantaneously with the state of the clock signal. For simplicity, we only discuss the case of an upconversion TX-mixer here, but similar conclusions hold for a DAC. The mixer is shown in Fig. 7.2, where an ideal 50% square wave switching between $+1$ and -1 models the hard-switching mixer operation. This square waveform has odd harmonics with a relative strength of $1/3$, $1/5$, $1/7$, etc. compared to the fundamental. Thus the ninth harmonic is still stronger than -20 dB compared to the fundamental.

In order to avoid harmonic mixing, the input signal could be multiplied by a sine wave signal using a highly linear multiplier. However, realizing a linear multiplier is much more difficult than a hard-switched mixer, and the generation of a clean sine wave is problematic, especially when a large frequency range is involved. Typical sine-wave oscillators, e.g. LC oscillators have only a limited tuning range in the order of 5–50%. If a larger tuning range is needed, digital dividers are commonly used to divide the VCO frequency to an appropriate value. As digital circuits benefit from Moore's law, we strongly prefer flexible digital synthesizer techniques over analog sine wave generation. However, this means we have to find a solution to suppress unwanted harmonics.

In traditional radio transmitters, these unwanted products are rejected using dedicated band-pass filters typically implemented using inductors and capacitors (LC filters). We like to avoid such filters on CMOS chips, as they require high quality inductors which are difficult to implement and/or take large chip area. For dynamic spectrum access, such filters are even more problematic as LC band-pass filters work at a fixed frequency related to the LC-resonance frequency, which limits the

flexibility in choosing a TX-frequency. The next section discusses a recently proposed polyphase multipath technique to eliminate these filters or relax their requirements significantly.

7.5 Polyphase Multipath Circuits for Spectral Purity Enhancement

Figure 7.2 shows a nonlinear circuit excited by a single sine wave at ω , producing a wanted output signal at ω but also unwanted harmonic distortion at 2ω , 3ω , 4ω , etc.. Figure 7.3 shows a polyphase three-path circuit, cancelling many harmonics of ω [11]. The basic idea is to divide a nonlinear circuit of Fig. 7.2 into ‘n’ equal smaller pieces, and apply an equal but opposite phase shift before and after each nonlinear circuit. If the phase shift in path ‘i’ is $(i - 1) \times \varphi$, where φ is a phase shift constant satisfying $n \times \varphi = 360^\circ$, the circuit will produce the same wanted harmonic as Fig. 7.2, but cancel many higher harmonics. Mathematically this can easily be shown using a power series expansion, assuming a memory-less weakly nonlinear system. If the signal $x(t) = A \cos(\omega t)$ is applied to the input, the output of the nonlinear circuit of the i^{th} path can be written as:

$$p_i(t) = a_0 + a_1 \cos(\omega t + (i - 1)\varphi) + a_2 \cos(2\omega t + 2(i - 1)\varphi) + a_3 \cos(3\omega t + 3(i - 1)\varphi) + \dots \tag{7.1}$$

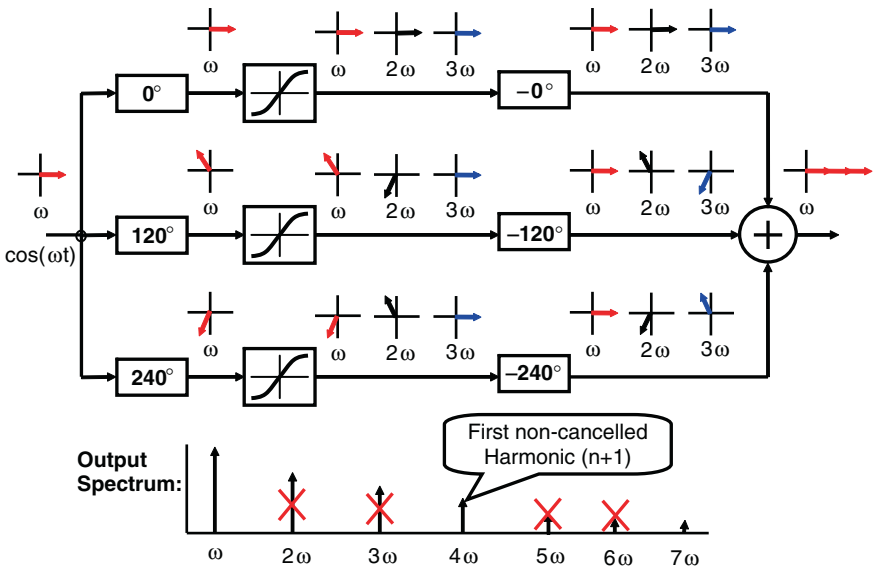


Fig. 7.3 Polyphase three-path circuit with harmonic cancellation except for harmonics $j \times n + 1$ (in this case $n = 3$, so harmonics 1, 4, 7,.. are not cancelled)

Where $a_0, a_1, a_2, a_3 \dots$ are Taylor series constants characterizing the nonlinearity [9]. From Eq. 7.1, it can be seen that the phase of the ‘ k^{th} ’ harmonic at the output of the nonlinear circuit rotates by ‘ k ’ times the input phase $(i - 1)\varphi$. The phase shifters, $-(i - 1)\varphi$, after the nonlinear blocks are required to align the fundamental components at ω in phase again.

The signals at the output of these phase shifters can be written as:

$$y_i(t) = a_0 + a_1 \cos(\omega t) + a_2 \cos(2\omega t + (i - 1)\varphi) + a_3 \cos(3\omega t + 2(i - 1)\varphi) + \dots \quad (7.2)$$

In Eq. 7.2, the phase of the fundamental component is identical for all the paths, but the phases of the harmonics are different for each path. If the phase φ is chosen such that $\varphi = 360^\circ/n$, then all the higher harmonics are cancelled [11], except for the k^{th} harmonics for which k equals $j \times n + 1$ ($j = 0, 1, 2, 3, \dots$).

The simplest example of a polyphase multipath circuit is a well-known differential circuit driven with balanced (anti-phase) input signals. It cancels all even harmonics (no cancellation of $k = j \times 2 + 1$, i.e. odd harmonics).

A system with three paths is shown in Fig. 7.3. In this case, phase shifts of $0^\circ, 120^\circ$ and 240° are added before the nonlinear block to path 1, 2 and 3 respectively, and equal but opposite phases $-0^\circ, -120^\circ$ and -240° behind the block. Due to the nonlinearity, the phase rotation for the k^{th} harmonic is k times the input phase. Thus the respective phases at the output of the nonlinear block for path [1–3] are $[0^\circ, 120^\circ, 240^\circ]$ for ω , $[0^\circ, 240^\circ, 120^\circ]$ for 2ω and $[0^\circ, 0^\circ, 0^\circ]$ for 3ω products. Figure 7.3 also shows how the phases of the harmonics at the output of each path combine. Only the fundamental components add up in phase (red arrows), while the black and blue vectors for the second and third harmonics create a “balanced structure” at the output, resulting in a zero sum (cancellation). However, the fourth harmonic components will align in phase again, and will add up like the fundamental. The output spectrum in the lower part of Fig. 7.3 shows that the second, third, fifth, sixth etc. harmonics are cancelled and the first non-cancelled is the fourth for a three-path system. Similarly for a four-path system the first non-cancelled harmonic will be the fifth harmonic and in general for an n -path system the $(n + 1)^{\text{th}}$ harmonic is the first non-cancelled harmonic. Theoretically, an infinite number of paths is needed to cancel all the harmonics. However, in practice higher order harmonics are weaker than low order harmonics and need not all be cancelled. Also, some filtering will in practice always be present, e.g. due to the limited bandwidth of an antenna or the speed limitations in a circuit. Moreover mismatches will put a practical limit on what is feasible [11].

If the non-linear system is excited by a two-tone input signal $x(t) = A_1 \cos \omega_1 t + A_2 \cos \omega_2 t$, besides harmonics the output will also contain intermodulation products at new frequencies $p\omega_1 + q\omega_2$, where p and q identify harmonics of ω_1 and ω_2 respectively, and can be positive or negative integer numbers. It can be shown easily that many intermodulation products are cancelled, except if $p + q$ equals $j \times n + 1$ (where $j = 0, 1, 2, 3, \dots$).

7.6 Mixer: Phase and Frequency Shifter

To realize wideband harmonic rejection using a polyphase multipath system, we need very wideband phase shifters before and after the nonlinearity. This is because all phase shifters need to have a constant phase shift over all relevant frequencies involved in the cancellation process. In a DSP intensive radio transmitter, digital signal processing techniques can be exploited to realize phase shifters before D/A conversion and nonlinear power amplification. Therefore, a good solution can be to shift this polyphase generation problem to the digital domain, and use a DSP followed by multiple DACs to generate multi-phase baseband signals. However, behind the nonlinear element we are in the analog domain, and there can be many harmonics. In that case cancellation of a multitude of harmonics requires a constant phase shift over many octaves of frequency.

A very wideband phase shifter can be implemented with a mixer, since a mixer as shown in Fig. 7.2 transfers phase information of both the “baseband” (BB) and “Local Oscillator” (LO) port to the output. Whatever phase is added to the LO signal will appear at the output of the mixer. So by replacing the second set of phase shifters in Fig. 7.3 with mixers, as shown in Fig. 7.4, we can achieve a wideband phase shift but simultaneously we will get frequency conversion. As upconversion is desired in a transmitter circuit anyway, this fits nicely to our goal. However, a mixer produces not only a sum frequency but also a difference frequency. Usually only one of these is the wanted signal, while the other (“the image”) needs to be

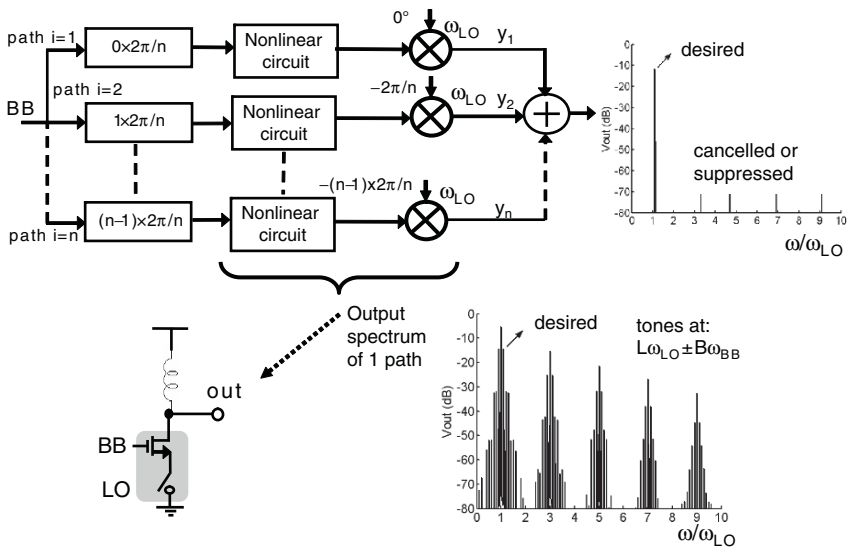


Fig. 7.4 Polyphase n-path transmitter with mixers as second phase shifters. Each path can be as simple as a switch and transistor, but produces many harmonics and sidebands due to time-variance and nonlinearity. The polyphase n-path system cancels most of these terms

suppressed. Moreover, the LO-signal usually is a square wave containing many harmonics, because flexible frequency synthesizers rely on digital dividers, as discussed in the previous section. For power efficiency reasons it is also highly desired to use a switching mixer and a large BB-signal swing, e.g. a single transistor with switch as shown in Fig. 7.4. Thus, the output spectrum for one path will now contain a forest of harmonics and sidebands as shown in the lower part of Fig. 7.4 for the case with a single-tone BB-signal. Spectral components occur at frequencies $L\omega_{LO} \pm B\omega_{BB}$, where L and B are integers, due to the multiplication of the square wave LO with the baseband input signal BB, and also the nonlinearity of the circuit. In the next section we will see how we can exploit the polyphase multipath technique to cancel almost all the unwanted components.

7.7 Filter-Less Power Up-Converter

A power upconverter combines the functionality of a power amplifier and upconversion mixer. The PA and mixer can be as simple as shown in Fig. 7.4, which is equivalent to first amplification and then mixing. Here the PA is a single transistor operating as transconductor (V-I converter), which is switched on and off by the LO signal via a switch (NMOS transistor driven by a digital inverter). Thus the V-I conversion and upconversion is done in the same circuit, via a switched transconductor mixer [12]. With respect to efficiency this circuit resembles a single transistor (class A) power amplifier. However, due to the polyphase multipath technique distortion products are cancelled and larger signal swings can be tolerated, improving efficiency.

Unfortunately, a few problematic products still remain present at the output. Since we have two input ports now (BB and LO), and mixing produces several sum and difference frequencies, a slightly different condition for non-cancelled products is found [11, 13] ($L = j \times n + B$ where $j = \dots - 2, -1, 0, 1, 2 \dots$, and B is a positive or negative integer number).

Especially the $3\omega_{LO} + 3\omega_{BB}$ is troublesome because the third order distortion term is usually much stronger than higher order distortion components [9] and is also close to the desired signal. It cannot be cancelled with any number of paths as all products for which $L = B$ are not cancelled ($j = 0$ case, so independent of n). To eliminate the strong $3\omega_{LO} + 3\omega_{BB}$ terms, the duty cycle of the LO was chosen to be $1/3$ [13]. By doing so, the third, sixth, ninth, etc. harmonic terms disappear from the Fourier series expansion, however some even order terms appear. Fortunately, it is quite easy to cancel even order products by using a differential baseband input (balancing).

To demonstrate the feasibility of a highly flexible multipath transmitter, we designed a power upconverter in a $0.13 \mu\text{m}$ CMOS process, covering all frequencies up to 2.4 GHz [13]. To show wideband spectral cleaning we designed an 18-path system, which can clean-up the spectrum up to the 17th harmonic. Figure 7.5 shows the 18-path power upconverter. Each path consists of a switched transconductor

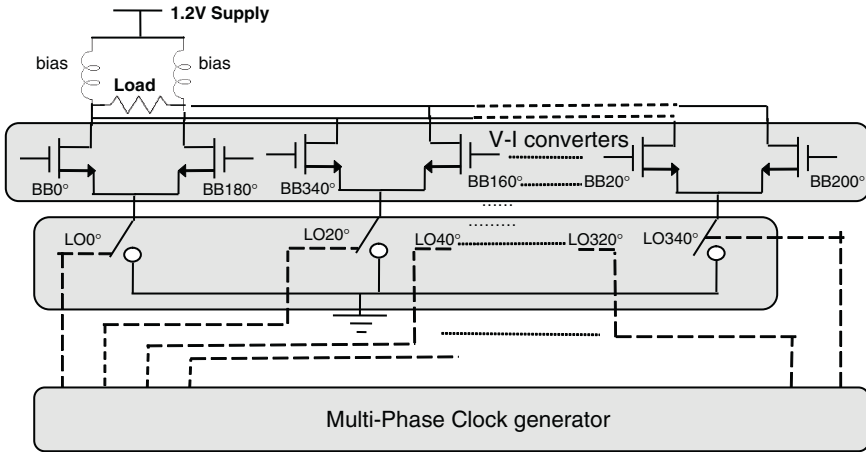


Fig. 7.5 Circuit concept of an 18-path power upconverter [13]

mixer [12] with a baseband signal applied to a differential pair, acting as a differential transconductor (V-I converter), and an LO signal driving a grounded switch. The output currents of the V-I converters are easily added by connecting them together, and the wanted output signals from all paths add up in phase. Thus the total area and power of the power upconverter core is not increased by splitting it into 18 paths.

The V-I converter transistors are biased at the supply voltage via two large inductors (see Fig. 7.5) to increase the output swing and efficiency, as commonly done in power amplifier design. The inductance and the load resistance constitutes a high-pass AC-coupling, which puts a lower limit to the RF frequency, but the chip itself can work at arbitrarily low frequency.

Operating each individual switched transconductor mixer at the 1 dB compression point, the upconverter is designed for a large output swing of about 2.5 V differential peak-to-peak voltage, to maximize efficiency. This is close to the maximum swing that can be achieved from a 1.2 V supply while keeping the output transistors in strong inversion and saturation, to maintain V-I converter functionality. For a 100 Ω load, the 2.5 V swing corresponds to roughly 8 mW output power. To further increase the output power without adding an external power amplifier, a transformer could be added for broadband impedance transformation while scaling up the output current via wider transistors. To maximize the flexibility and frequency range, we implemented the LO phase generation via a current mode logic shift register running at nine times the LO frequency. This enabled us to evaluate the circuit for an arbitrary LO-frequency between DC and a maximum given by the speed limitation of the logic used to realize the shift register. For 18 paths we need LO signals of 18 different phases (0° , 20° , 40° ... 340°) with one third duty cycle. Applying a positive and a negative clock edge alternately to successive latches in a chain of 18 D latches (see Fig. 7.5), 18 different phases are produced. The feedback through the NOR gate is used to make the duty cycle 1/3.

In our experimental setup, the nine differential baseband voltages with different phases are generated off-chip. More work has to be done to explore the most effective way to generate multi phase baseband signals on-chip via DSP techniques and multiple DACs.

The multipath technique cleans the output spectrum from unwanted harmonics, which result from the hard-switching mixer, but also from non-linearity in the switched transconductor. Simulations and measurements show that we can drive the power upconverter close to its 1 dB compression point with harmonics well below <-40 dBc and realize the high 2.5 V output voltage swing directly over the load (e.g. antenna). Note that the two inductors are only used for biasing, and not for (dedicated) band-pass filtering.

The proposed upconverter has been fabricated in a 0.13 μm CMOS process and takes an active area of only 0.14 mm^2 . It delivers 8 mW output power to a 100 Ω off-chip load [13]. Figure 7.6 shows the output frequency spectrum for a transmit frequency of 350 MHz for one path (no cancellation) and for the complete 18-path system (lower part of Fig. 7.6). Clearly all problematic products are suppressed significantly. Please note that the unfortunate FM-radio spurs that are modulated with our output signal are caused by a 100 MHz high power FM-radio broadcast transmitter on the roof of our building. Overall, ten chips were measured with spurious

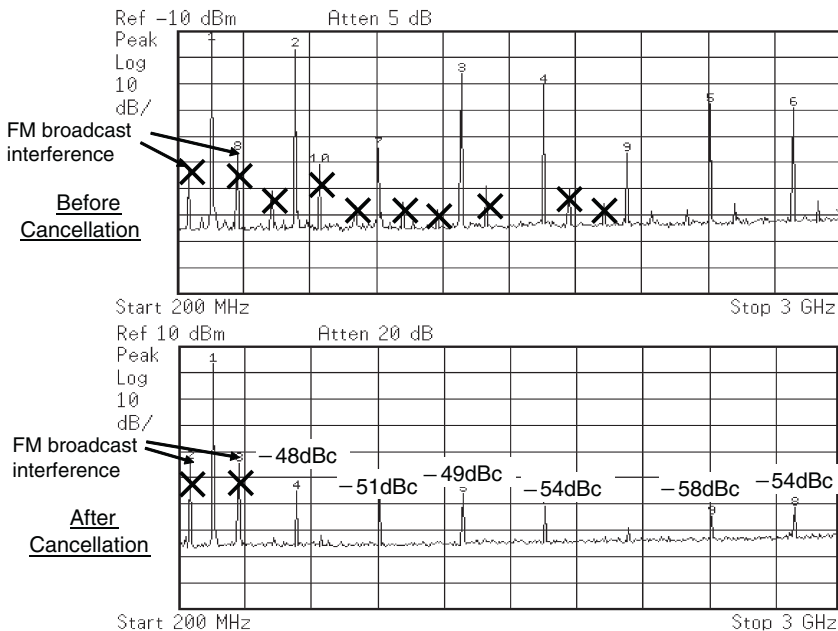


Fig. 7.6 Output spectra of the 18-path Power Upconverter (PU) chip [13], with out-of-band power <-40 dBc up to the 17th harmonic (LO = 350 MHz)

emissions < -40 dBc for all harmonics up to the 17th harmonic of the LO, for an LO-frequency from 30–800 MHz. For higher frequencies the chip has a six-path mode which was measured for 30 MHz–2.4 GHz with similar rejection up to the fifth harmonic of the LO. The rejection of products related to the fundamental of the LO, like the LO-feedthrough and image component, can be a few decibel worse, but requirements on in-band products are usually less strict than for out-of-band spurious emissions.

The (drain) efficiency of the core of the power upconverter is 11%, which is good compared to other power upconverters, given the low harmonics. However, we used current-mode logic circuits biased at high currents at 8 GHz LO frequency. As a result the power consumption of the digital part currently dominates (~ 150 mW). In the following sections we examine alternative architectures for multi-phase clock generation, and will look at possibilities to reduce the power consumption while still achieving a low phase error.

7.8 Multi-phase Clock Requirements

As discussed in the previous sections, polyphase multipath circuits require multi-phase clocks. Such clocks are also useful in many other applications. For quadrature down-conversion mixers, two differential clocks with 90° phase-separation are needed (or four single ended clocks with phases 0° , 90° , 180° and 270°). A popular harmonic rejection mixer architecture [3, 5, 8] needs eight equidistant phases with 45° separation. For high-speed serial links multi-phase clocks are used [14] to process data streams at a bit rate higher than the clock frequency, and in time-interleaved ADCs to realize a conversion rate higher than feasible with individual quantizers [15]. Aiming for multi-functionality (e.g. software defined radio), we would like a flexible Multi-phase Clock Generator (MPCG) to adapt to largely different data rates, sampling rates or radio frequencies.

To implement a MPCG, both delay-locked loops (DLLs) and shift registers (SRs) have been used. A SR MPCG also functions as a divide-by- N divider for N -phase clock generation. Although a SR MPCG seems more attractive due to its wide working frequency range (flexibility), it requires an N times higher clock-frequency and at first glance seems to consume more power. However, a SR MPCG doesn't have jitter accumulation from one clock phase to the other as in a DLL equivalent, which should be taken into account for a fair comparison. In the following sections, we aim to make a solid comparison between these two MPCGs, primarily based on their power and absolute output jitter performance [16]. Furthermore, flexibility aspects relevant for multi-functionality will be discussed.

We will start with a DLL MPCG, discuss its architecture and analyze its jitter performance, and then addresses the SR MPCG. Later in the chapter we will make a comparison and verify the analysis via simulation results.

7.9 DLL MPCG Jitter

7.9.1 DLL MPCG Architecture

The architecture of a DLL MPCG is shown in Fig. 7.7a. It consists of a voltage controlled delay line (VCDL) which has N identical delay units (DUs) and a control loop consisting of a phase detector (PD), a charge pump (CP) and a loop filter (LF). In the DLL, a reference clock CLK_{ref} , generated by a VCO with a frequency of f , is propagated through the VCDL. The loop compares the phase of the last output of the VCDL with CLK_{ref} and controls the VCDL so that its total delay time is one reference clock period. Once locking is achieved, the N outputs $CLK_1 \sim CLK_N$ are multi-phase clocks with $2\pi/N$ phase spacing.

7.9.2 DLL MPCG Output Jitter

The DLL MPCG output jitter can be divided into three parts: 1) jitter transferred from the reference clock, 2) jitter generated by the VCDL and 3) jitter from the control loop. The jitter of the reference clock is transferred to the DLL outputs with some jitter peaking [17, 18]. The DLL cannot decrease reference clock jitter, but jitter peaking can be made very small by choosing a low DLL loop bandwidth [17, 18]. For an optimal DLL design, the jitter contribution of the control loop is negligible [17] and hence ignored hereafter. Thus, VCDL jitter is our main worry.

In a DLL MPCG, the VCDL generates two types of jitter: random noise jitter caused by *thermal noise* and deterministic mismatch jitter due to *mismatch* of the delay units. The DLL renders no improvement of VCDL noise jitter. Again, the VCDL noise jitter is lowest for low values of the loop bandwidth, in which case it would be almost equal to that of a free-running VCDL [17]. The jitter will thus accumulate from one delay unit to the other. If the noise jitter variance of one delay unit is $\sigma_{i,DU,noise}^2$, and we assume uncorrelated white noise, the noise jitter variance on the output of the n^{th} delay unit will be n times bigger. For multi-phase clock applications like the software defined radio transmitter discussed in the beginning

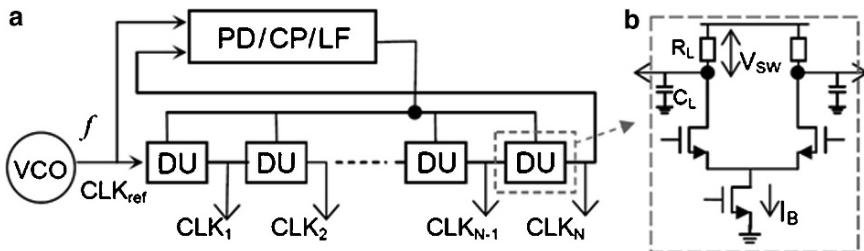


Fig. 7.7 (a) DLL MPCG architecture (b) CML delay unit schematic

of this chapter [13], the jitter of every clock phase is equally relevant. To quantify the jitter of a set of N -phase clocks, the averaged jitter variance of the N clocks is a meaningful quantity. The average noise jitter variance generated by the DLL can be calculated as:

$$(\sigma_{t,DLL,noise}^2)_{avgN} = \frac{1}{N} \cdot \sum_{n=1}^N n \cdot \sigma_{t,DU,noise}^2 = \frac{N+1}{2} \sigma_{t,DU,noise}^2 \quad (7.3)$$

Different from noise jitter, the DLL loop *can* improve the deterministic mismatch jitter. The start and end of the VCDL are both aligned to the reference clock and thus have zero deterministic time error. The maximum mismatch jitter appears at the middle of the VCDL. If we define the mismatch jitter variance of one delay unit as $\sigma_{t,DU,mis}^2$, the jitter variance on the output of the n^{th} delay unit can be calculated as [17]:

$$\sigma_{t,DU_n,mis}^2 = \frac{n(N-n)}{N} \sigma_{t,DU,mis}^2 \quad (7.4)$$

The average mismatch jitter variance generated is then:

$$(\sigma_{t,DLL,mis}^2)_{avgN} = \frac{N^2-1}{6N} \sigma_{t,DU,mis}^2 \approx \frac{N^2 \gg 1}{6} \sigma_{t,DU,mis}^2 \quad (7.5)$$

7.10 SR MPCG Jitter

7.10.1 SR MPCG Architecture

The architecture of a SR MPCG is shown in Fig. 7.8a. It consists of a D flip-flop (DFF) chain with N identical DFFs. A reference clock CLK_{ref} , generated by a VCO with a frequency $N \cdot f$, is fed into the DFF chain. A flip logic (FL) circuit monitors the N outputs of the DFF chain and flips the logic value at the D input of the first DFF twice every N reference clock cycles. In other words, the outputs of the DFF

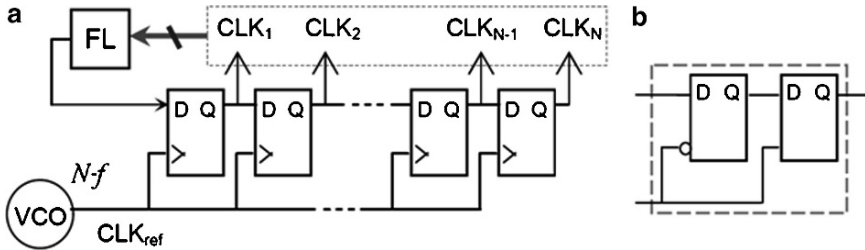


Fig. 7.8 (a) SR MPCG architecture (b) DFF block schematic

chain run at a frequency of f and the SR based MPCG also functions as a divide-by- N divider. Since a DFF is sensitive to rising edges, the Q output of each DFF is delayed from the previous DFF's output by one reference clock period, which is equivalently a $2\pi/N$ phase delay. In this way, N -phase clocks $CLK_1 \sim CLK_N$ are generated. Depending on different implementations of the flip logic, the duty cycle of the N -phase clocks can theoretically vary from $1/N$ to $(N-1)/N$. For example, if 18-phase clocks with a $1/3$ duty cycle are wanted, the flip logic can simply be a NOR-gate with CLK_6 and CLK_{12} as its inputs [13]. This gives the SR based MPCG extra flexibility.

7.10.2 SR MPCG Output Jitter

The SR MPCG output jitter can be divided into two parts: jitter transferred from the reference clock and jitter generated by the DFF chain. The flip logic is simply a logical “enabler” for the first DFF and will not contribute to jitter.

For the jitter transferred from the reference clock, the SR MPCG renders no improvement. Any timing error at the reference clock will be transferred to the DFF chain outputs.

Similar to the VCDL, the DFF chain also generates two types of jitter: noise jitter and mismatch jitter. However, there is *no jitter accumulation* from one DFF to the other, since each DFF output only acts as an “enabler” for the next DFF, while the VCO defines the timing. A DFF can be designed with two master/slave latches as shown in Fig. 7.8b. For a proper design, only the second latch contributes to jitter since the first is just an “enabler”. If we define the rms noise and mismatch jitter variance of one latch as $\sigma_{t,Latch,noise}^2$ and $\sigma_{t,Latch,mis}^2$ respectively, the average jitter variance for the set of N -phase clocks generated by the SR can be easily calculated as:

$$(\sigma_{t,SR,noise}^2)_{avgN} = \frac{1}{N} \cdot \sum_{n=1}^N \sigma_{t,Latch,noise}^2 = \sigma_{t,Latch,noise}^2 \quad (7.6)$$

$$(\sigma_{t,SR,mis}^2)_{avgN} = \frac{1}{N} \cdot \sum_{n=1}^N \sigma_{t,Latch,mis}^2 = \sigma_{t,Latch,mis}^2 \quad (7.7)$$

7.11 Comparison Between DLL and SR Jitter

7.11.1 Jitter Transferred from the Reference Clock

From the analysis above, we see that both the DLL and SR MPCGs render no improvement on the reference clock jitter. However, the SR MPCG needs a reference clock with N times higher frequency than the DLL. If both clocks are generated

by a VCO¹, the VCO for the SR should work at N times higher frequency, raising the question how this impacts power consumption. Assuming the VCO has an f^{-2} power spectrum and its quality of design is adequately assessed via the often used figure of merit FOM [19], the single sideband phase noise to carrier ratio at an offset frequency f_m can be expressed as:

$$L(f_m) = \frac{10^{FOM/10}}{P_{VCO}} \cdot \frac{f_{VCO}^2}{f_m^2} \quad (7.8)$$

where f_{VCO} is the frequency and P_{VCO} is the power dissipation in [mW]. It is well-known that the variance for stationary absolute jitter is related to the total area of its power spectrum, i.e. the reference clock jitter variance $\sigma_{t,ref}^2$ becomes:

$$\sigma_{t,ref}^2 = \frac{2 \times \int_{f_l}^{f_h} L(f_m) d(f_m)}{(2\pi f_{VCO})^2} = \frac{10^{FOM/10}}{2\pi^2 \cdot P_{VCO}} \cdot \left(\frac{1}{f_l} - \frac{1}{f_h} \right) \quad (7.9)$$

where $[f_l, f_h]$ is the specified integration region. Equation 7.9 indicates that although the VCO in the SR MPCG runs at N times higher frequency, it outputs the same jitter, given the same power and the same quality of design. If an LC VCO is used, higher working frequency may even be preferred, since the quality factor of an inductor ($\omega L/R$) increases with frequency and smaller inductors are needed (less chip area). On the other hand there are limits to increasing the frequency, and also clock buffer power consumption can become an issue.

7.11.2 Comparing Jitter Generated due to Thermal Noise

Because of better supply noise rejection, current mode logic (CML) circuits are often used in low jitter designs. To compare the jitter generated by the two MPCGs, we assume that they both use CML circuits. The simplified schematic of a CML delay unit is shown in Fig. 7.7b. It is based on an NMOS source coupled differential pair driving the resistive load R_L and biased by a current source I_B . As the loads are RC circuits, the propagation delay t_d can be approximated as:

$$t_d = \ln 2 \cdot R_L C_L = \ln 2 \cdot (V_{SW}/I_B) \cdot C_L \quad (7.10)$$

where V_{SW} is the differential output swing and is determined by R_L and I_B due to the full switching of the tail current.

¹ The VCO can be part of a synthesizer, e.g., a PLL. We didn't discuss the effect of the PLL loop on the reference clock phase noise since it's the same for the SR and DLL. The PLL for the SR does not require an extra divide-by- N since the SR itself functions as a divide-by- N and can be re-used.

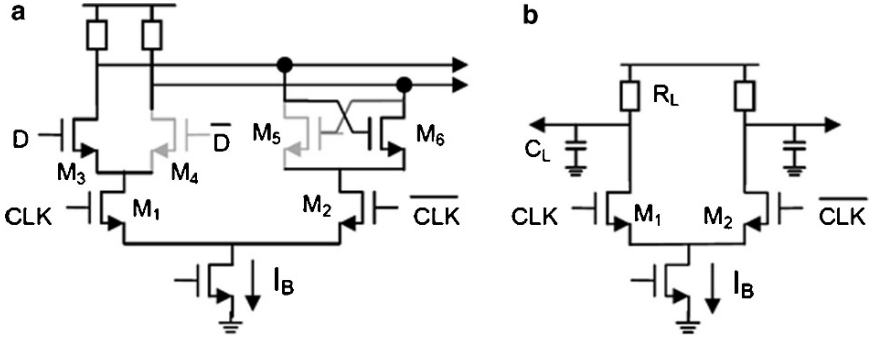


Fig. 7.9 (a) Schematic of a CML latch at the switching instant. (b) Simplified schematic for jitter analysis

The CML implementation of a latch is shown in Fig. 7.9a. For a proper operation, the D inputs of the latch should be already stable before the CLK starts to switch. For example, D is high and \overline{D} is low. Therefore, at the switching moment, transistors M_4 and M_5 are off. M_3 and M_6 are in their saturation region and work as cascode transistors on top of the differential pair. The noise contribution of M_3 – M_6 can thus be neglected. The schematic of the latch can be simplified to Fig. 7.9b which is exactly the same as the schematic of the CML delay unit in Fig. 7.7b. Therefore, we can apply the same noise jitter analysis for the delay unit and the latch.

The noise jitter variance of a CML delay unit can be predicted using the analysis presented in [20] as:

$$\sigma_{t,noise}^2 = (1 + \gamma + \gamma_T \cdot \frac{2I_B}{V_{OV,T}} \cdot \frac{R_L}{2}) \cdot \frac{2kTC_L}{I_B^2} \quad (7.11)$$

where γ and γ_T are respectively the noise factor of the differential pair transistors and the tail bias transistor, $V_{OV,T}$ is overdrive voltage of the tail bias transistor and $2I_B/V_{OV,T}$ represents its transconductance assuming a square-law model.

In most of the clock generator designs, jitter and power are two important parameters. Via admittance level scaling [21], both noise and mismatch jitter can always be reduced at the cost of increasing the power consumption P . In order to take this tradeoff into account and make a fair comparison, jitter variance is normalized to power, with 1mW as reference:

$$(\sigma_t^2)_{NorP} = \sigma_t^2 \cdot (P/1mW) \quad (7.12)$$

For a given circuit, applying admittance level scaling will not change the value of $(\sigma_t^2)_{NorP}$. Smaller $(\sigma_t^2)_{NorP}$ means generating less jitter for a given amount of power. For a CML circuit, the power consumption is dominated by the static power $I_B \cdot V_{DD}$. With Eqs. 7.11 and 7.12, we find for both a CML delay unit and latch:

$$(\sigma_{t,\text{noise}}^2)_{\text{NorP}} = (1 + \gamma + \gamma_T \cdot \frac{I_B R_L}{V_{OV,T}}) \cdot \frac{2kT \cdot V_{DD}}{1mW} \cdot \frac{C_L}{I_B} \quad (7.13)$$

Substituting Eq. 7.10 into Eq. 7.13 yields:

$$(\sigma_{t,\text{noise}}^2)_{\text{NorP}} = \{(1 + \gamma + \gamma_T \cdot \frac{V_{SW}}{V_{OV,T}}) \cdot \frac{2kT \cdot V_{DD}}{\ln 2 V_{SW} \cdot 1mW}\} \times t_d \quad (7.14)$$

Equation 7.14 indicates that the *normalized noise jitter variance is proportional to t_d* for a given power budget.

In a DLL, if t_d is tuned by tuning R_L while keep V_{SW} constant, I_B and thus $V_{OV,T}$ in Eq. 7.14 will vary with t_d . Here to simplify the comparison, we ignore this second order effect and assume the delay unit and the latch have the same V_{SW} and $V_{OV,T}$. We will see the effect of this simplification in Section 7.12 (simulation results). A DLL has N delay units contributing to jitter and power while a SR has N latches contributing to jitter and $2N$ latches dissipating power. The average noise jitter variance generated by the DLL and the SR MPCGs can then be compared using Eqs. 7.3, 7.6 and 7.14, as:

$$\frac{(\sigma_{t,\text{SR},\text{noise}}^2)_{\text{avg}N,\text{NorP}}}{(\sigma_{t,\text{DLL},\text{noise}}^2)_{\text{avg}N,\text{NorP}}} = \frac{(\sigma_{t,\text{Latch},\text{noise}}^2)_{\text{NorP}} \times 2N}{\frac{N+1}{2} \times (\sigma_{t,\text{DU},\text{noise}}^2)_{\text{NorP}} \times N} = \frac{4}{N+1} \cdot \frac{t_{d,\text{Latch}}}{t_{d,\text{DU}}} \quad (7.15)$$

The comparison result thus depends on the amount of delay of the delay unit $t_{d,\text{DU}}$ and that of the latch $t_{d,\text{Latch}}$. In a DLL MPCG, the VCO defines the frequency and the VCDL defines the delay in between the N output clocks. Both the VCO and the delay line need to be tuned for the DLL MPCG to work at a frequency f , where the delay of each delay unit should satisfy:

$$t_{d,\text{DU}} = \frac{T}{N} = \frac{1}{N \cdot f} \quad (7.16)$$

In contrast, the SR MPCG is more flexible. For different f , only the VCO needs to be tuned since both the frequency and the delay in between the N output clocks are defined by the clock period of the VCO. The only concern is that the DFFs should operate correctly, which requires [22]:

$$t_{d,\text{Latch}} + t_{su} \leq \frac{1}{N \cdot f} \quad (7.17)$$

where t_{su} is the setup time required by the DFF. Defining the maximum working frequency of a SR MPCG for N -phase clock generation in a certain technology as $f_{\text{max},\text{SR}}$, the latch delay will have its minimum value $t_{d,\text{Latch},\text{min}}$ at $f_{\text{max},\text{SR}}$ given by:

$$t_{d,\text{Latch},\text{min}} = \frac{1}{1 + \alpha_{su}} \cdot \frac{1}{N \cdot f_{\text{max},\text{SR}}} \quad (7.18)$$

with α_{su} the ratio between t_{su} and $t_{d,Latch,min}$. As a small delay is preferred for a small $(\sigma^2_{t,noise})_{NorP}$, the latch delay can be equal to its minimum in Eq. 7.18. For a delay unit, the delay is limited by Eq. 7.16. Taking this factor into account, Eq. 7.15 can be re-written as:

$$\frac{(\sigma^2_{t,noise,SR})_{avgN,NorP}}{(\sigma^2_{t,noise,DLL})_{avgN,NorP}} = \frac{1}{1 + \alpha_{su}} \cdot \frac{f}{f_{max,SR}} \cdot \frac{4}{N + 1} \quad (7.19)$$

As soon as the wanted number of clock phases is larger than three ($N > 3$), Eq. 7.19 is smaller than one since the DFF needs a finite setup time ($\alpha_{su} > 0$) and the working frequency of the SR can't surpass the technology limit ($f \leq f_{max,SR}$). This means that the SR based MPCG generates less noise jitter than the DLL counterpart for a given power budget. Equation 7.19 also indicates that the advantage of the SR based MPCG will be larger if more advanced technologies are used and in applications where clocks with a larger number of phases at lower frequencies are needed.

7.11.3 Comparing Jitter Generated due to Mismatch

Based on similar reasoning as for the noise jitter analysis, the latch can be simplified as shown in Fig. 7.9b for mismatch jitter analysis and we can apply a similar analysis. In a CML delay unit, there are two mismatch jitter sources: one is the RC load which contributes to RC delay mismatch $\sigma^2_{t,RC,mis}$ and the other is the differential pair input referred offset voltage σ^2_{Voff} which makes the switching moment deviate from the actual crossing point of the input clocks. The tail bias transistor mismatch does not lead to jitter since it's a common mode error and we are interested in the crossing points.

Using Eq. 7.10, the jitter due to the RC load mismatch becomes:

$$\left(\frac{\sigma_{t,RC,mis}}{t_d}\right)^2 = \sigma_{\Delta R_L/R_L}^2 + \sigma_{\Delta C_L/C_L}^2 \quad (7.20)$$

with ΔR_L and ΔC_L the absolute error in the value of R_L and C_L .

In a DLL, the RC delay must be tunable. For simplicity, we assume that C_L is tuned by putting less or more capacitors in parallel and R_L is tuned by putting less or more resistors in parallel.² Since the matching improves with area [21], Eq. 7.20 can be rewritten as:

$$\sigma^2_{t,RC,mis} = [(A_R \cdot \sqrt{R_L})^2 + (A_C / \sqrt{C_L})^2] \times t_d^2 \quad (7.21)$$

² If R_L is realized with a MOS transistor in linear region and tuned by tuning the gate voltage, it can be shown that the matching property of R_L in a DLL DU is even worse.

where A_R and A_C are IC process constants for the matching property of the load resistance and capacitance, respectively.

The input referred offset voltage of a differential pair can be calculated using the method presented in [23] as:

$$\sigma_{V_{off}}^2 = \sigma_{\Delta V_t}^2 + \frac{I_B}{4K} \times \sigma_{\Delta R'_L/R_L}^2 + \frac{I_B}{4K} \times \sigma_{\Delta K/K}^2 \quad (7.22)$$

where $\sigma_{\Delta V_t}^2$ is the differential pair threshold voltage mismatch variance, $\Delta R'_L$ is the relative error between the two R_L loads, K is the transconductance parameter of the differential pair with $\sigma_{\Delta K/K}^2$ describing its mismatch.

The total mismatch jitter variance $\sigma_{t,mis}^2$ can be found by adding $\sigma_{t,RC,mis}^2$ and the jitter variance caused by $\sigma_{V_{off}}^2$ which is $\sigma_{V_{off}}^2$ divided by $(I_B/C_L)^2$, the square of the slope of the differential switching voltage at the zero crossing.

$$\sigma_{t,mis}^2 = A_R^2 \cdot R_L \cdot t_d^2 + \frac{A_C^2 \cdot t_d^2}{C_L} + \frac{\sigma_{\Delta V_t}^2 + \frac{I_B}{4K} \times A_R^2 \cdot R_L + \frac{I_B}{4K} \times \sigma_{\Delta \beta/\beta}^2}{(I_B/C_L)^2} \quad (7.23)$$

The power normalized mismatch jitter variance can be derived with Eq. 7.12 and Eq. 7.23 as:

$$\begin{aligned} (\sigma_{t,mis}^2)_{NorP} = & \frac{V_{DD}}{1mW} \cdot \{V_{SW} \cdot A_R^2 \times t_d^2 + \ln 2 \cdot V_{SW} \cdot A_C^2 \times t_d \\ & + \frac{\sigma_{\Delta V_t}^2}{\ln 2 \cdot V_{SW}} \times C_L \cdot t_d + \frac{A_R^2}{\ln 2 \times 4K} \times C_L \cdot t_d + \frac{\sigma_{\Delta K/K}^2}{4K} \times C_L^2\} \end{aligned} \quad (7.24)$$

Equation 7.24 shows that the delay unit and latch generates less mismatch jitter for a smaller delay, with a given power. It also suggests that with a constant V_{SW} , it's better for a DLL to tune up R_L instead of C_L when larger delay is needed.

Assuming the terms with t_d proportionality in Eq. 7.24 which include the threshold voltage mismatch are the dominating mismatch jitter sources and setting the other initial conditions the same for a fair comparison, the mismatch jitter generated by the DLL and SR can be compared with Eqs. 7.5, 7.7 and 7.24 as:

$$\frac{(\sigma_{t,SR,mis}^2)_{avgN,NorP}}{(\sigma_{t,DLL,mis}^2)_{avgN,NorP}} \approx \frac{12}{N} \cdot \frac{t_{d,Latch}}{t_{d,DU}} \quad (7.25)$$

Substituting Eq. 7.16 and Eq. 7.18 into Eq. 7.25 yields:

$$\frac{(\sigma_{t,SR,mis}^2)_{avgN,NorP}}{(\sigma_{t,DLL,mis}^2)_{avgN,NorP}} = \frac{1}{1 + \alpha_{su}} \cdot \frac{f}{f_{max,SR}} \cdot \frac{12}{N} \quad (7.26)$$

The situation where Eq. 7.26 is larger than one only occurs when the wanted number of clock phases N is smaller than 12 together with a high frequency f close to $f_{max,SR}$. In other cases, Eq. 7.26 is smaller than one, which means that the SR MPCG generates less mismatch jitter than the DLL counterpart for a given power budget. Equation 7.26 also indicates that the advantage of the SR based MPCG will be larger if more advanced technologies are used and a larger number of clock phases at lower frequencies are needed.

7.11.4 Discussion

The analysis above shows that a SR MPCG transfers the same jitter from the reference clock and almost always generates less jitter³ than a DLL MPCG for a given power consumption. For mismatch jitter, the DLL MPCG may have a slight advantage in some high frequency cases.⁴ Although we assumed that current mode logic circuits are used to implement the MPCG, the way of analysis developed can also be applied when other logic families like CMOS logic, true single phase clocking or dynamic transmission gate logic are used. Note that the advantage of a SR MPCG comes from its features like no jitter accumulation from one clock phase to the other and the flexibility of setting small latch delay time. These features are independent of the logic family used.

From an implementation point of view, the SR MPCG is easier since it does not require a phase detector, loop filter and analog tuned delays. However, it can be difficult to implement in applications where N is large and f is high since the SR works at $N \cdot f$. Still, speed improves as technology advances. Another concern is that the loading of the VCO is more severe in the SR MPCG, since it needs to drive N DFFs. This problem can be alleviated by downscaling the DFFs by admittance scaling [21], which is acceptable because they generate less jitter than the delay units, thus saving power and chip area.

From a multi-functionality point of view, the SR MPCG is clearly more attractive: it is basically a digital circuit which can operate from arbitrarily low frequency up to $f_{max,SR}$, while a DLL requires tuning of an “analog” delay. Also, a SR can basically instantaneously change its output frequency, while a DLL settles slowly, due to the preferred low loop bandwidth. Finally, a SR MPCG has the flexibility to generate clocks with different duty cycle.

³ In case phase noise is important, the SR is also better as both the SR and DLL generate white phase noise, while the reference clock has the same spectrum shape for both cases.

⁴ If 50% reference clock duty cycle is guaranteed, both edges can be used. The N DFFs in the SR can be replaced with N latches as in [12]. The previous analysis then overestimates the SR MPCG power consumption by two times.

7.12 Simulation Results

In order to verify the calculations, simulations were done for a DLL and a SR for $N = 8$ in 0.13- μm CMOS. The reference clocks are voltage sources with 1 k Ω source resistance. The VCDL delay is tuned up by tuning up the load resistance as suggested by Eq. 7.24 while keep V_{SW} to be 0.6 V. For the DFFs, α_{su} is about 0.5. The load capacitance is 100 fF, which is comparable to the parasitic capacitances. In this implementation, $f_{max,SR}$ is about 1.5 GHz for 8-phase clock generation. Figure 7.10 shows the strobed PNoise analysis results for noise jitter. The simulated values coarsely fit the estimated curve. The larger deviation when t_d is larger relates to the simplification we made below Eq. 7.14. We see this simplification is in favor of the DLL which normally has a larger t_d . Therefore, it does not affect the conclusion. Figure 7.11 shows the Monte Carlo analysis results for mismatch jitter. The bent shape of the simulated values when t_d is tuned from low to high is predicted by Eq. 7.24. The simulated values fit the estimated curve well which means the threshold voltage mismatch dominates in this design.

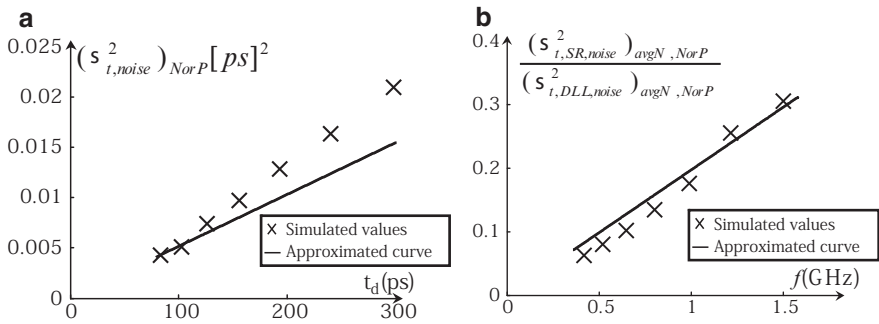


Fig. 7.10 Noise jitter simulation results in 0.13 μm CMOS with $N = 8$ for (a) a CML delay unit (b) DLL and SR comparison

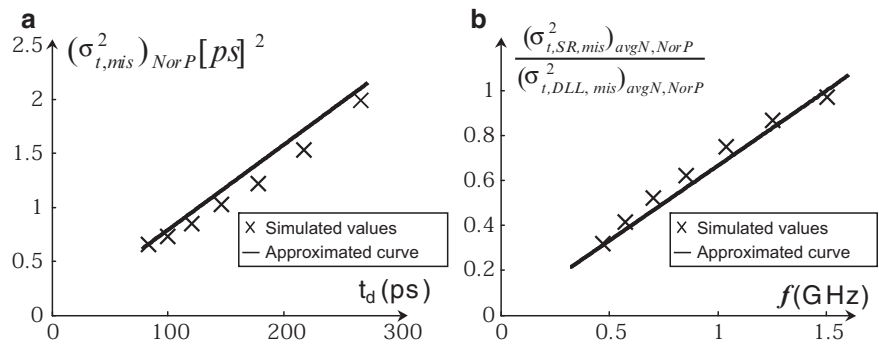


Fig. 7.11 Mismatch jitter simulation results in 0.13 μm CMOS with $N = 8$ for (a) a CML delay unit (b) DLL and SR comparison

7.13 Conclusion

In this chapter we reviewed some recent research results relevant for the feasibility of fully integrated CMOS cognitive radio transceivers. We motivated why an ADC and DAC are not sufficient to realize the radio interface. Coarse power estimates show that A/D conversion of high dynamic range radio signals at the antenna is not realistic for gigahertz radio signal. However, RF sampling is feasible and the sampling clock jitter requirements are not as difficult as often thought, but are similar to those of traditional mixer based RF receivers. A key fundamental problem in radio circuits is their nonlinear and/or time-variant nature. As a result they produce not only a wanted output signal, but also many unwanted harmonics and sidebands. We presented a polyphase multipath technique that addresses this problem without using any dedicated filters [24]. Using this technique, a highly flexible power up-converter has been realized on in CMOS, operating at an arbitrary transmit frequency between DC and 2.4 GHz, with unwanted harmonics and sideband lower than <-40 dBc.

Flexible multi-phase clock generation is at the heart of multipath polyphase transceivers. This chapter motivates why a SR MPCG is more attractive for flexible multi-functional circuits than a DLL MPCG as it is easier to change its frequency and duty cycle. Furthermore, analysis shows that a SR MPCG almost always generates less jitter than a DLL equivalent when both are realized with CML circuits, at a given power budget. This is partly because a SR MPCG has no jitter accumulation from one clock phase to the other as in a DLL counterpart. In addition, a SR MPCG can use latches with very small delay time, while jitter generation of a CML circuit is proportional to its (functionally required) delay time. A SR MPCG requires a reference clock with higher frequency, which can be realized in a power neutral way provided that the VCO core determines power consumption. The advantages of a SR MPCG will be larger as technology advances.

Acknowledgements The authors would like to thank Eisse Mensink and Rameswor Shresta for contributions to this work, Henk de Vries and Gerard Wienk for practical assistance during design and measurements and Fokke Hoeksema and Jaap Haartsen for useful discussions. Philips Research is acknowledged for providing the silicon.

References

1. M. J. Marcus, "Unlicensed cognitive sharing of TV spectrum: the controversy at the Federal Communications Commission," *IEEE Communications Magazine*, vol. 43, pp. 24–25, 2005.
2. R. H. Walden, "Performance trends for analog to digital converters," *IEEE Communications Magazine*, vol. 37, pp. 96–101, 1999.
3. R. Bagheri, A. Mirzaei, M. E. Heidari, S. Chehrazi, L. Minjae, M. Mikhemar, W. K. Tang, and A. A. Abidi, "Software-defined radio receiver: dream to reality," *IEEE Communications Magazine*, vol. 44, pp. 111–118, 2006.

4. F. Bruccoleri, E. A. M. Klumperink, and B. Nauta, "Wide-band CMOS low-noise amplifier exploiting thermal noise canceling," *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 275–282, 2004.
5. J. A. Weldon, R. S. Narayanaswami, J. C. Rudell, L. Li, M. Otsuka, S. Dedieu, T. Luns, T. King-Chun, L. Cheol-Woong, and P. R. Gray, "A 1.75-GHz highly integrated narrow-band CMOS transmitter with harmonic-rejection mixers," *IEEE Journal of Solid-State Circuits*, vol. 36, pp. 2003–2015, 2001.
6. K. Muhammad, R. B. Staszewski, and D. Leipold, "Digital RF processing: toward low-cost reconfigurable radios," *IEEE Communications Magazine*, vol. 43, pp. 105–113, 2005.
7. V. J. Arkesteijn, E. A. M. Klumperink, and B. Nauta, "Jitter requirements of the sampling clock in software radio receivers," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 53, pp. 90–94, 2006.
8. Z. Ru, E.A.M. Klumperink, B. Nauta, "A Discrete-Time Mixing Receiver Architecture with Wideband Harmonic Rejection", *2008 IEEE International Solid-State Circuits Conference (ISSCC), Digest of Technical Papers*, pp. 322–323/616, 2008.
9. W. Sansen, "Distortion in elementary transistor circuits," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 46, pp. 315–325, 1999.
10. P. B. Kenington, "Linearized transmitters: an enabling technology for software defined radio," *IEEE Communications Magazine*, vol. 40, pp. 156–162, 2002.
11. E. Mensink, E. A. M. Klumperink, and B. Nauta, "Distortion cancellation by polyphase multipath circuits," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, pp. 1785–1794, 2005.
12. E. A. M. Klumperink, S. M. Louwsma, G. J. M. Wienk, and B. Nauta, "A CMOS switched transconductor mixer," *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 1231–1240, 2004.
13. R. Shrestha, E. Mensink, E. A. M. Klumperink, G. J. M. Wienk, and B. Nauta, "A Polyphase Multipath Technique for Software Defined Radio Transmitters", *IEEE Journal of Solid State Circuits*, Vol. 41, No. 12, pp. 2681–2692, Dec. 2006.
14. C. K. Yang and M. A. Horowitz, "A 0.8-/spl mu/m CMOS 2.5 Gb/s oversampling receiver and transmitter for serial links," *IEEE Journal of Solid-State Circuits*, vol. 31, pp. 2015–2023, Dec. 1996.
15. W. C. Black, and D. A. Hodges, "Time interleaved converter arrays", *IEEE Journal of Solid-State Circuits*, vol. 15, no. 6, pp. 1022–1029, Dec. 1980.
16. X. Gao, E. Klumperink and B. Nauta, "Advantages of Shift Registers Over DLLs for Flexible Low Jitter Multiphase Clock Generation", *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 55, no. 3, pp. 244–248, Mar. 2008.
17. R. van de Beek, E. Klumperink, C. Vaucher, and B. Nauta, "Low-jitter clock multiplication: a comparison between PLLs and DLLs," *IEEE Transactions on Circuits and Systems II*, vol. 49, pp. 555–566, Aug. 2002.
18. M.-J. Edward Lee, et al., "Jitter transfer characteristics of delay-locked loops-theories and design techniques," *IEEE Journal of Solid-State Circuits*, vol. 38, pp. 614–621, Apr. 2003.
19. P. Kinget, "Integrated GHz voltage controlled oscillators," *Analog Circuit Design: (X)DSL and Other Communication Systems; RF MOST Models; Integrated Filters and Oscillators*, W. Sansen, J. Huijsing, and R. van de Plassche, Ed. Boston, MA: Kluwer, 1999, pp. 353–381.
20. S. Levantino, et al., "Phase noise in digital frequency dividers", *IEEE Journal of Solid-State Circuits*, vol. 39, no. 5, pp. 775–784, May 2004.
21. E. Klumperink, B. Nauta, "Systematic Comparison of HF CMOS Transconductors", *IEEE Transactions on Circuits and Systems II*, vol. 50, no.10, pp. 728–741, Oct. 2003.
22. J. M. Rabaey, *Digital Integrated Circuits, A Design Perspective*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
23. P. Gray, P. Hurst, S. Lewis, and R. Meyer, *Analysis and Design of Analog Integrated Circuits*, 4th Edition. New York: Wiley, 2001, pp. 236–237.
24. E. A. M. Klumperink, R. Shrestha, E. Mensink, V. J. Arkesteijn, B. Nauta, "Cognitive radios for dynamic spectrum access - Polyphase Multipath Radio Circuits for Dynamic Spectrum Access", *IEEE Communications Magazine*, Volume 45, Issue 5, pp. 104–112, May 2007.

Chapter 8

IIP2 Improvement Techniques for Multi-standard Mobile Radio

Mohammad B. Vahidfar and Omid Shoaie

8.1 Introduction

An effort toward realization of universal radios, i.e. radios able to tune a carrier frequency over a wide RF range and supporting high data rate modulations is underway [1–4]. This is a trend toward the integration of multiple functions (phone, video-game console, personal digital assistant, digital camera, web-browser, e-mail, etc.) into a wireless device that can be used anywhere in the world. Enabling factor of this evolution is the increasing availability of multi-standard terminals, integrated in low-cost silicon technologies that can communicate efficiently, using many different standards, for voice and data, depending on the availability and convenience. The ultimate solution for the radio of such a multi-functional terminal would be a multi-standard radio, built in a very low-cost CMOS technology, capable of being programmed to operate according to all major communications standards [1–5].

While progress has been made in DSP and base-band functions of such a terminal, the low voltage low power radio front-end has remained intangible. Hardware sharing between different applications and programmability are key to save silicon costs. Direct conversion architecture lends itself to a compact solution, with minimum number of external components and a simpler reconfigurable base-band [5–7]. However, several challenging issues including second order intermodulation (IM2) appears in this kind of receivers. Moreover, the dynamic range requirements set by cell-phones in particular make the RF radio front-end extremely challenging [8–10].

In this manuscript, we focus the attention on the down-conversion mixer, the major responsible one for the limited dynamic range in the RF front-end, investigating a high second order input intercept-point (IIP2) reconfigurable Gilbert type solution.

M.B. Vahidfar (✉)
Universita degli studi di Pavia, 27100 Pavia, Italy
e-mail: mohammad.vahidfar@unipv.it

O. Shoaie
University of Tehran, Tehran, Iran
e-mail: oshoaie@ut.ac.ir

In a perfectly balanced down converter, the even-order distortion caused by device nonlinearity would not appear in the signal path. However, in a practical situation where mismatch in the load and switching transistors exist, the even-order intermodulation appears in the signal path. The even-order nonlinearity can be reduced by using differential topology and symmetric layout, but the required performance especially for cellular phone applications can not be met.

An effective solution for improving IIP2 is using an external SAW filter between LNA and mixer in order to make the mixer linearity requirement relaxed by attenuating the out of bands blockers [11–13]. This solution is expensive and also for multi-standard application realizing a multi-standard (MS) tunable SAW is necessary. There are already MS SAWs for the main GSM bands, with low insertion loss, while electronically tunable MS SAWs are in their research phase [14].

Therefore, to meet the required IIP2 performance, especially for cellular phone applications in sub scaled CMOS technologies with low voltage supplies, using some IIP2 improvement techniques including on-chip calibration circuits [15–18] or analog techniques [10, 19, 20] is necessary. The core competency of analog techniques is being simple and consuming lower power, while generally using several techniques is needed to be considered in order to cancel the IM2 components generated by different sources. However the calibration techniques, which are usually complex, are less sensitive in PVT variation and are capable of cancelling IM2 components generated by different sources and mechanisms. In this manuscript, both analog techniques and calibration circuits are considered for enhancing mixer IIP2 performance in multi-standard applications. Considering multi-standard, the main standards for personal communications currently deployed: GSM and UMTS, for cellular telephony and IEEE802.11b-g-a for wireless local area network access are assumed. These standard are in 1–6 GHz frequency range which is today crowded of a variety of standards for wireless communications (Fig. 8.1).

The manuscript is organized as follows. The IIP2 performance required by MS-radio is discussed in Section 8.2. The IM2 sources in an active mixer are introduced in Section 8.3. The proposed IIP2 improvement techniques are presented in Sections 8.4–8.6 while Section 8.7 draws conclusions.

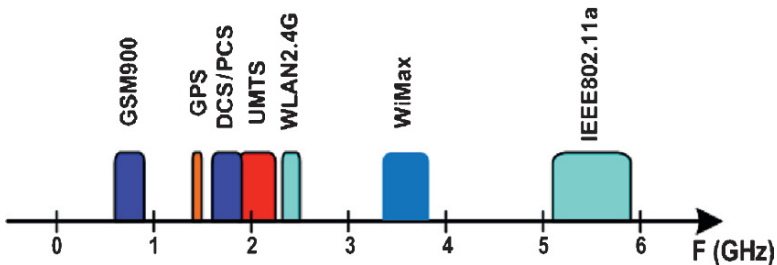


Fig. 8.1 Multi-standard frequency spectrum

8.2 IIP2 Requirements of MS-Radio

The cellular phone standards set a strict dynamic range requirement which make the design of CMOS RX front-end for these applications more challenging. In order to have an IIP2 specification, required by MS radio, the IIP2 requirement of GSM and UMTS RX front-end is calculated

8.2.1 GSM

GSM is the leading actor of the wireless telecommunications market since more than 10 years. Several frequency bands have been allocated in order to spread the GSM all over the world, enabling the global roaming. The most diffused bands are the 850,900, 1,800 and 1,900 MHz bands. Enhanced-GSM (E-GSM) and Digital Cellular Communication System (DCS) are both used in Europe, while Personal Communication Services (PCS) has been deployed in the United States. Quad-band GSM cell-phones operating in all the most diffused bands are today widespread on the market.

Although the constant-envelope GMSK modulation would not generate wide-band second-order intermodulation products by its own, the standard specifies an AM suppression test [21] which sets stringent IIP2 specification. This test was introduced in order to avoid receiver desensitization in presence of a GMSK pulse jammer, as produced by the on/off-switching signal on another carrier. Both second order intermodulation product generated in the AM suppression test and $1/f$ noise are superimposed onto the down-converted GSM signal in a zero-IF receiver. According to the standard, a -99 dBm desired signal has to be correctly demodulated in presence of a -31 dBm AM interferer (Fig. 8.2). The IIP2 requirement can be calculated as follows [3]:

$$IIP2 > 2P_{AM} - NoiseFloor \tag{8.1}$$

Where P_{AM} is the AM interferer power. As a result, an IIP2 higher than $+49$ dBm has to be guaranteed. This is a very challenging specification, because with a

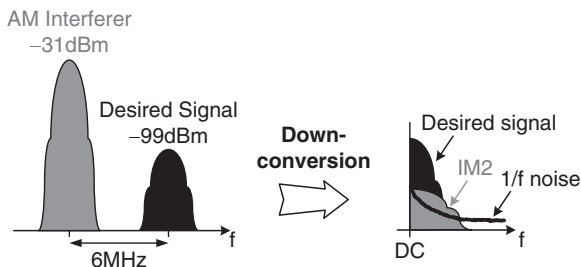


Fig. 8.2 AM suppression test in GSM

typical 16–18 dB low noise amplifier (LNA) gain, the IIP2 requirement for the down-converter exceeds +65 dBm. The very narrow-band GSM signal (200 kHz) is extremely susceptible to $1/f$ noise and DC offsets. This is the foremost reason why the earliest chipset solutions for GSM were based on super-heterodyne [22] or, on single [23] or double [24] conversion to a very low intermediate frequency (~ 100 kHz).

8.2.2 UMTS

UMTS is a continuously transmitting and receiving frequency division duplexing (FDD) system. At the user equipment side, the TX band is located between 1,920 and 1,980 MHz while the RX band is between 2,110 and 2,170 MHz. The minimum frequency TX–RX spacing is thus 135 MHz. Because of the duplexing technique adopted by UMTS, a highly linear receiver is required to reject the transmitter leakage into the RX section of a zero-IF architecture. Since the receiver specifications heavily depend on commercial duplexers performance, in order to obtain the receiver requirements a typical duplexer with 1.8 dB insertion loss (IL) and 54 dB isolation between TX and RX sections, is considered.

In the sensitivity test, considering only antenna and receiver noise, a 9 dB-NF and -99 dBm antenna-referred maximum noise power are needed not to exceed the 0.001 BER as required [25]. Second-order intermodulation products, due to transmitter (IM_{2TX}) and reciprocal mixing of TX leakage due to oscillator phase noise (RM_{TX}), need also to be considered in the sensitivity test and added to the receiver noise [26]. Consequently, the following relation holds [3]:

$$-99\text{dBm} > N_R + IM_{2TX} + RM_{TX} \quad (8.2)$$

Where N_R is the antenna referred receiver thermal noise. Considering a class III UMTS transmitter, the TX leakage is -30 dBm. The reciprocal mixing product is made negligible setting the phase noise specification at -150 dBc/Hz at 135 MHz offset. In order to keep the NF specification within acceptable limits an IIP2 requirement of at least +46 dBm has to be met.

For this high IIP2 requirement and because of large TX leakage, the earliest UMTS chip solutions employ modified direct conversion architecture with an external filter (Fig. 8.3) [11–13]. This external inter-stage band-pass RF filter placed between two LNAs in the front-end section attenuates the TX leakage in the receive path strongly relaxing the second order intermodulation requirements [27, 28].

8.3 Second Order Nonlinearity Sources

Generally a low voltage active mixer is made by cascading of pseudo differential transconductor, switching pairs and load. Although a conventional Gilbert cell allows IIP2 around 90 dBm up to few tens of megahertz, the achievable values drop

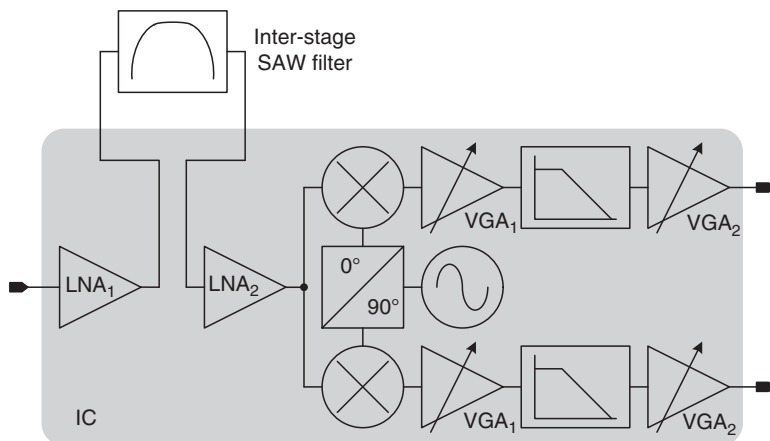


Fig. 8.3 Inter-stage filtering in UMTS RX front-end for relaxing mixer IIP2

dramatically, in the GHz range. Several mechanisms are main reason of second-order intermodulation distortion in CMOS downconversion mixers: self-mixing, transconductor nonlinearity, switching pairs nonlinearity, and mismatch in load resistors which are briefly reviewed in this section [29].

Self-mixing of radio frequency signals due to coupling into the local oscillator port can be significantly reduced by means of layout counter-measures: metal lines carrying RF and LO signals should never cross each other, or rigorously kept orthogonal if crossing is unavoidable. In this way, coupling can be kept below 80 dB even at RF frequency, allowing IIP2 values in excess of 90 dBm and making this mechanism not the dominant one [19].

In gigahertz frequencies the IIP2 performance of the mixer is degraded mainly due to the limited bandwidth of the switching core caused by the parasitic capacitor of the source node of the switching transistors (C_p). The random mismatch between switching transistors can be modeled by an offset voltage [10, 29]. The parasitic capacitor C_p is charged and discharged by this offset voltage in each f_{LO} period in which f_{LO} is the frequency of local oscillator (LO). The low pass behavior of the source node of the switching transistor causes a nonlinear current, which is modulated by transconductor intermodulated components, injected to the switching transistors [8]. This IM2 current is appeared at the mixer output after down-converting by the switching transistors. Detailed analysis of these mechanisms is found in the literature [29].

Switching pair IIP2 values increase rapidly with the biasing current. However, increasing the biasing current also raises both thermal and flicker noise contributions [19], i.e., at biasing levels allowing relatively large IIP2, the switching pair noise performance would be irremediably compromised.

Moreover, in a perfectly tuned switching stage, the IIP2 of the mixer is usually lower than the cellular phone requirement; because of nonlinearities in the I-V characteristic, active devices are responsible for generation of second-order

inter-modulation distortion components [30, 31]. The analysis of inter-modulation distortion in the input transconductor can be developed based on Volterra series expansion of the I–V characteristic [29]:

$$I_{diff} = G_1^{diff}(\omega) \circ V_{in} + G_2^{diff}(\omega, \omega) \circ V_{in}^2 + G_3^{diff}(\omega, \omega, \omega) \circ V_{in}^3 \quad (8.3)$$

$$I_{CM} = G_1^{CM}(\omega) \circ V_{in} + G_2^{CM}(\omega, \omega) \circ V_{in}^2 + G_3^{CM}(\omega, \omega, \omega) \circ V_{in}^3 \quad (8.4)$$

Where, I_{diff} and I_{CM} are the differential and the common mode output current. G_1 , G_2 and G_3 show the transconductance gain, the second-order and third order common-mode and differential conductance, respectively, and V_{in} presents the RF signal applied to the mixer.

The major reason of IIP2 degeradation in mixer input stage is due to the large common mode IM2 ($I_{IM2,CM}$) current generated in the pseudo differential transconductor. As it is shown in Eq. (8.5), this CM current ($I_{IM2,CM}$) is converted to differential products by the LO switching pairs asymmetry and load mismatch and is added to the other differential IM2 current (the first term of Eq. 8.5) and appeared in the output IM2 current of the mixer ($I_{IM2,out}$).

$$I_{IM2,out}^2 = L^2 I_{IM2,diff}^2 + \left(L^2 + \left(\frac{\Delta R_L}{R_L} \right)^2 \right) I_{IM2,CM}^2 \quad (8.5)$$

In which $I_{IM2,diff}$, R_L , ΔR_L and L are the differential IM2 current generated in the transconductor, the mixer output resistance, mismatch in the output load and low frequency leakage of the switching transistors, respectively.

Using (8.5) the input intercept second-order voltage point (IIV2) is defined:

$$IIV2 = \frac{2}{\pi} \frac{g m_{RF} \cdot v_{RF}^2}{\sqrt{L^2 \left(I_{IM2,diff}^2 + I_{IM2,CM}^2 \right) + \left(\frac{\Delta R_L}{R_L} \right)^2 I_{IM2,CM}^2}} \quad (8.6)$$

In which $g m_{RF}$ is the input mixer transconductance and is equal to G_1^{diff} . By using the Eqs. (8.3)–(8.6), IIV2 can be rewritten in [29]:

$$IIV2 = \frac{2}{\pi} \frac{g m_{RF}}{\sqrt{L^2 \left[\left(G_2^{diff} \right)^2 + \left(G_2^{CM} \right)^2 \right] + \left(G_2^{CM} \frac{\Delta R_L}{R_L} \right)^2}} \quad (8.7)$$

By leveraging in-depth analysis of second-order intermodulation mechanisms in active downconversion mixers, some IIP2 improvement techniques is reviewed in the next sections which are organized as follows:

The second order nonlinearity improvement of mixer input stage is discussed in Section 8.4. An IIP2 calibration technique is presented in Section 8.5. A multi-band solution for the switching pairs is discussed in Section 8.6.

8.4 Second Order Nonlinearity of Input Stage

As discussed, the main IIP2 degradation in the input stage is caused by IM_{2CM} . This problem is addressed in [10] by a RC degenerative input stage in which the input transconductance is reduced in IF frequencies comparing to RF frequencies. However, this solution can not work with low voltage supplies, because of using stacked transistors and consuming high overdrive voltage in the cascade transistors for achieving high gm degeneration. This technique is used in [5] for a low voltage application by using a folded architecture. A low voltage solution in which a feedback loop is employed for attenuating IM_{2CM} is presented in [19] for GSM receivers. However it is not applicable for high bandwidth applications like UMTS, because the bandwidth of the feedback loop is low (the bandwidth of the loop should be equal to the channel bandwidth in order to remove all in-bands IM_2 components).

This loop bandwidth is dominated by several poles such as the two poles of the compensated amplifier and the current injecting transistor and the pole generated by the CM part of the mixer load. To meet the required IIP2 performance by using this approach, the gain bandwidth product (GBW) of the feedback loop which is 10 MHz in GSM should be greater than 200 and 250 MHz for UMTS and IEEE802.11, respectively.

The poor IIP2 performance of the mixer input stage can be improved by employing a low voltage and high bandwidth feedback loop, as presented in the next sub-section. The IM_{2CM} products of the input stage are attenuated by injecting a nonlinear current which is controlled in the feedback loop.

8.4.1 *IM2 Cancellation Mechanism*

A simple model of this IM_2 cancellation technique is shown in Fig. 8.4. An IM_2 generator block senses the current of the transconductor and regenerates the required IM_{2CM} . For simplicity, suppose this current sensing is done by a small resistor r in the source of input transistors. The generated current is amplified, filtered out and then is injected to the output of M1 and M2 transistors. The amplitude of the injected current should be controlled in order to cancel the IM_{2CM} current of transconductor.

Referring to Fig. 8.5, it is supposed that M3 and M4 transistors, connected to the input RF signals, are similar. Since the output of these two transistors are connected together, the differential components of their currents including signal

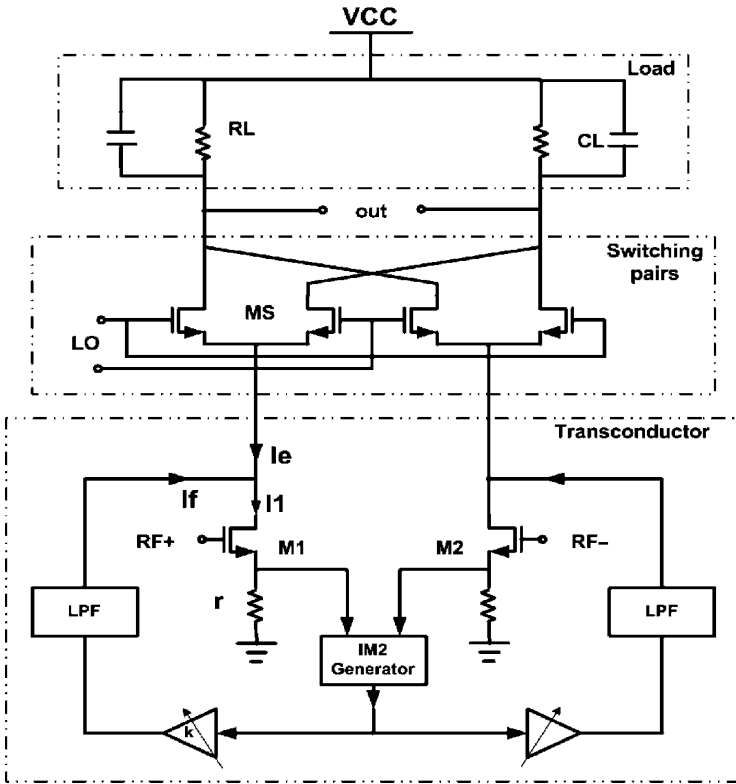


Fig. 8.4 Block diagram of the IM2 cancellation technique

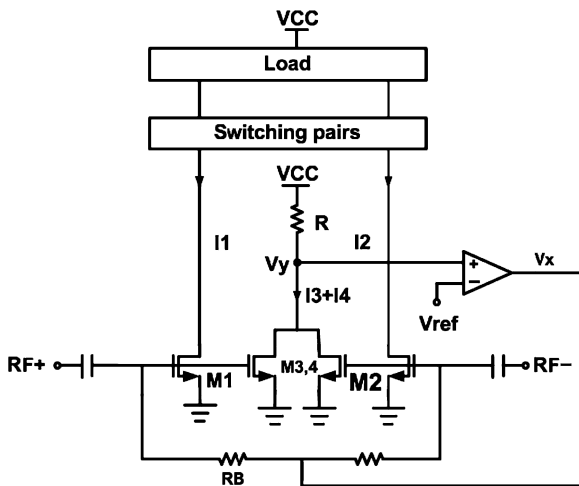


Fig. 8.5 Mixer enhanced by the IM2 cancellation loop

current, $IM_{2\text{diff}}$ and other differential intermodulated components do not appear in the sum of their output currents (I_{OL}). This is shown by the following equation:

$$I_{OL} = I_3 + I_4 = I_{IM2,CM,OL}(M3) + I_{IM2,CM,OL}(M4) + I_{IMy,CM} \quad (8.8)$$

Where $I_{IM2,CM,OL}(M3)$ and $I_{IM2,CM,OL}(M4)$ show the IM_{2CM} current of transistors M3 and M4, respectively, and $I_{IMy,CM}$ represents all CM even-order distortions except the second order one.

The first two terms of the Taylor series expansion of the I–V characteristic of a MOS transistor is expressed as below:

$$i = gm \cdot v_{gs} + g_2 \cdot v_{gs}^2 \quad (8.9)$$

By using Eqs. (8.8) and (8.9) and considering the second order distortion:

$$I_{OL} = 2I_{IM2,CM,OL}(M3) = 2g_2 \cdot (v_{in})^2 \quad (8.10)$$

In which, v_{in} is the RF signal applied to the input RF+ and RF– nodes by different polarity.

Now consider a negative feedback loop around M3 and M4 transistors made by an amplifier and a RC filter, as it is shown in Fig. 8.5. The Eq. (8.8) in the closed loop condition is:

$$I_{CL} = I_3 + I_4 = I_{IM2,CM,CL}(M3) + I_{IM2,CM,CL}(M4) + I_{IMz,CM} \quad (8.11)$$

Where, $I_{IM2,CM,CL}(M3)$ [$I_{IM2,CM,CL}(M4)$] shows the IM_{2CM} current of M3 (M4) transistor in closed loop condition.

By considering Eqs. (8.9) and (8.11):

$$I_{CL} = gm_3 \cdot v_{gs,M3} + gm_4 \cdot v_{gs,M4} + g_2 \cdot (v_{gs,M3} + v_{gs,M4})^2 \quad (8.12)$$

Where, gm_3 and gm_4 are the transconductance of transistors M3 and M4, respectively. The v_{gs} of transistors M3 and M4 are calculated by the small signal analysis of the loop:

$$\begin{cases} v_{gs,M3}(s) = \frac{1}{1+R_B \cdot C \cdot s} (R_B \cdot C \cdot s \cdot v_{in} + v_x) \\ v_{gs,M4}(s) = \frac{1}{1+R_B \cdot C \cdot s} (-R_B \cdot C \cdot s \cdot v_{in} + v_x) \end{cases} \quad (8.13)$$

In which, v_x is the voltage at the output of amplifier.

Using Eqs. (8.10)–(8.13) and assuming that 3 dB frequency of RC filter is higher than IM_2 components frequency and also is one decade lower than RF frequencies, the following relationship between IM_{2CM} current of transistor M3 (M4) in open and closed loop conditions is derived:

$$I_{IM2,CM,CL}(M3) = I_{IM2,CM,OL}(M3) + gm_3 \cdot v_x \quad (8.14)$$

Moreover; referring to Eq. (8.11) and Fig. 8.5 and neglecting CM intermodulated components with orders higher than 2, the following equation is derived:

$$I_{IM2,CM,CL}(M3) + I_{IM2,CM,CL}(M4) = -\frac{v_x}{A \cdot R} \quad (8.15)$$

Where, A is the gain of the amplifier.

From Eqs. (8.14) and (8.15) it can be concluded that:

$$I_{IM2,CM,CL}(M3) = \frac{I_{IM2,CM,OL}(M3)}{1 + A \cdot (gm_3 + gm_4) \cdot R} \quad (8.16)$$

Equation (8.16) shows that the $IM2_{CM}$ current of M3 transistor is attenuated in the feedback loop by the loop gain (LG) as it was expected from the feedback theory. Also the LG is

$$LG = A \cdot (gm_3 + gm_4) \cdot R = 2A \cdot gm_3 \cdot R \quad (8.17)$$

Equation (8.15) can be simplified as below:

$$v_x = (-2A \cdot R) I_{IM2,CM,CL}(M3) \quad (8.18)$$

Moreover; using Eqs. (8.16) and (8.18) and assuming high loop gain condition ($LG \gg 1$):

$$v_x \approx -\frac{I_{IM2,CM,OL}(M3)}{gm_3} \quad (8.19)$$

In other words, this voltage which is made by the feedback loop is responsible for attenuating the $IM2_{CM}$ current in M3 and M4 transistors.

Considering that M1 and M3 transistors are in parallel and also v_x acts as a common mode voltage for all M1, M2, M3 and M4 transistors then:

$$I_{IM2,CM,CL}(M1) = I_{IM2,CM,OL}(M1) + gm_1 \cdot v_x \quad (8.20)$$

Where; $I_{IM2,CM,CL}(M1)$ and $I_{IM2,CM,OL}(M1)$ are the $IM2_{CM}$ current of M1 transistor when the mixer is with and without the feedback loop, respectively.

Considering Eqs. (8.19) and (8.20) then:

$$\frac{I_{IM2,CM,CL}(M1)}{I_{IM2,CM,OL}(M1)} = 1 - \frac{gm_1}{gm_3} \cdot \frac{I_{IM2,CM,OL}(M3)}{I_{IM2,CM,OL}(M1)} \quad (8.21)$$

As it is shown in [29], the $IM2_{CM}$ of pseudo differential transconductor is proportional to g_2 which, as explained in Eq. (8.9), is the second order coefficient of the MOS I-V characteristic.

$$I_{IM2,CM} \propto g_2 \quad (8.22)$$

Therefore considering Eqs. (8.21) and (8.22):

$$\frac{I_{IM2,CM,CL}(M1)}{I_{IM2,CM,OL}(M1)} = 1 - \frac{gm_1}{gm_3} \cdot \frac{g_{2,M3}}{g_{2,M1}} \quad (8.23)$$

From this equation it is concluded that $IM2_{CM}$ current of M1 transistor can be canceled if the following equation is satisfied.

$$\frac{gm_3}{g_{2,M3}} = \frac{gm_1}{g_{2,M1}} \quad (8.24)$$

For the CMOS transistors in the strong inversion region the g_2 is:

$$\frac{g_2}{gm} = \frac{gm}{I} \cdot \frac{2}{(1 + \theta \cdot v_{od})^2} \quad (8.25)$$

Where; v_{od} and θ are overdrive voltage and fitting parameter for mobility, respectively [32].

Therefore; it can be concluded that the above condition (Eq. 8.24) can be satisfied by proper sizing and biasing of transistor M3.

In conclusion, the $IM2_{CM}$ components generated in M1 and M2 transistors are canceled by employing a feedback loop which attenuates the $IM2_{CM}$ current of M3 and M4 transistors. However referring to Eqs. (8.23)–(8.25), the mismatch between M1 and M3 transistors degrades this $IM2$ cancellation.

8.4.2 Circuit Implementation

Figure 8.6 shows the implementation of the proposed $IM2$ cancellation technique. The required LG is provided by two stages of the gain. The first one is made by M5, M3 and M4 transistors and the second one is implemented by a differential common gate (M6 and M7 transistors). To ensure the loop stability, the loop is compensated by a LPF made by R_B and C. The second pole of the loop is in the output of M5 transistor which is very far from the first pole because of low capacitance load of this node. Moreover the bias current of M3 and M4 are chosen high enough to ensure high bandwidth in this node. The bias current of M5 transistor is set to 1.8 mA and consequently that of M3 transistor is 0.9 mA.

The loop gain in DC is 41 dB, while the unity gain frequency and phase margin of the loop are 250 MHz and 70° , respectively. The $IM2$ cancellation circuit bandwidth should be higher than half of the signal bandwidth in a Zero-IF receiver. Therefore, the proposed circuit can work for high channel bandwidth applications like UMTS with 4 MHz channel bandwidth as well as GSM with 200 KHz.

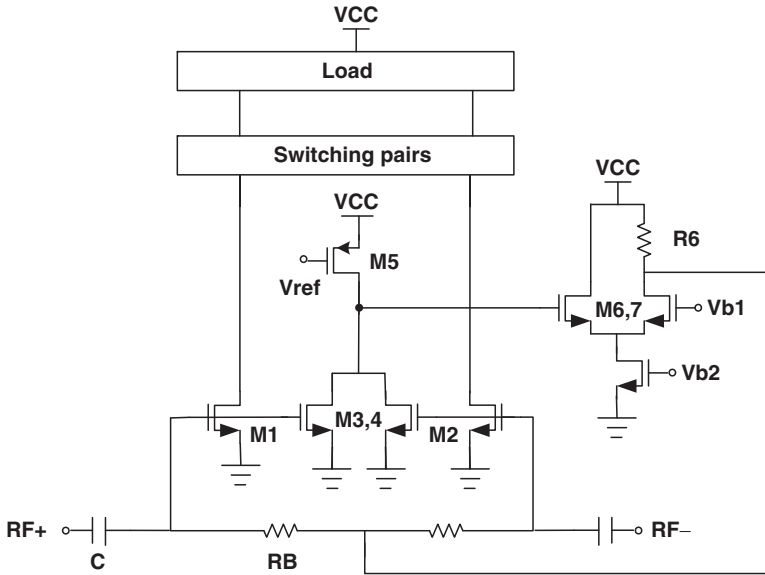


Fig. 8.6 The proposed IM2 cancellation circuit

The proposed IM2 cancellation circuit does not affect the mixer NF because both thermal and flicker noises generated in the circuit are injected to the mixer as a common mode noise and consequently are rejected in the mixer output.

The minimum supply voltage of the proposed circuit is equal to $V_{TH} + 3V_{od}$, in which V_{TH} is the threshold voltage of NMOS Transistors. This minimum voltage is less than 1 V if V_{TH} is less than 500 mV and V_{od} is considered 150 mV.

8.4.3 The Proposed Mixer

The performance of the proposed IM2 cancellation circuit is demonstrated in a mixer, designed for UMTS application, and presented in Fig. 8.7.

The M1 and M2 transistors, sized by 120/0.15 μm , are biased by 4.5 mA, which 1.5 mA of this bias current is sourced by the Mp and the rest comes from the switching pairs. In this way, there is a degree of freedom for choosing the transconductor bias current for the optimized IIP3 performance, independent of the bias current of the switching pairs which highly affects the mixer NF.

The switching transistors (MS) are sized by 400/0.24 μm in order to achieve the optimum flicker noise and IIP2 performance. To maximize the mixer dynamic range the parasitic capacitor of the switching transistors is tuned in the LO frequency by a LC filter [10]. The quality factor of the differential inductor (Q) is about 10 in the working frequency. This inductor is grounded by a large capacitor C_p (25 pf) in RF

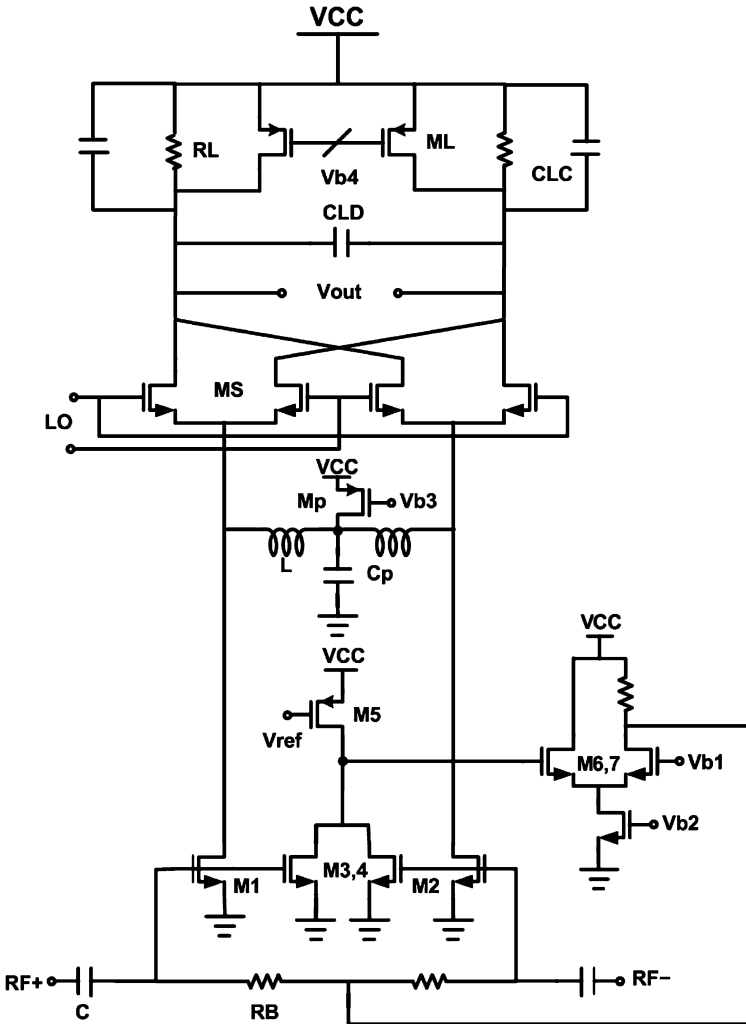


Fig. 8.7 The proposed mixer

frequencies. Additionally, the differential part of IM2 generated in the mixer input transistors is also re-circulated by the inductor by making the output of M1 and M2 transistors shorted in IF frequencies.

8.4.4 The Simulation Results

The design was done in a 65 nm CMOS technology. The IIP2 simulation is done in the worst case condition by applying offset voltages equal to three times of

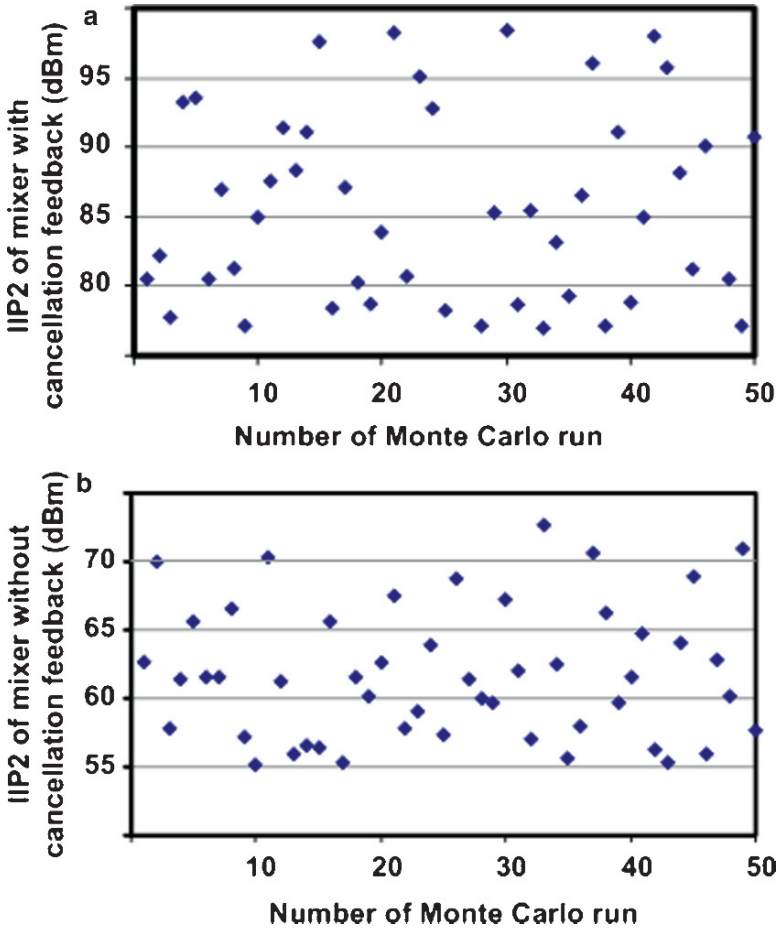


Fig. 8.8 (a) The result of Monte Carlo analysis of the enhanced mixer, (b) The result of Monte Carlo analysis of the conventional mixer

the standard deviation mismatch (σ) to all transistors [10, 29]. These mismatches are calculated based on the process information and are verified by Monte Carlo simulations.

The IIP2 simulation is done by two-tone test at 1.98 GHz with different spacing less than 10 MHz which emulates the TX leakage as the main blocker while the LO frequency is 2.11 GHz. The Monte Carlo simulation of the mixer with and without cancellation loop is depicted in Fig. 8.8. It shows that the IIP2 of the mixer enhanced by the cancellation feedback is higher than 75 dBm which is required in cellular phones, while in the conventional mixer it is between 55 and 73 dBm. Moreover it shows that the mixer achieves about 22 dB improvement in IIP2 in the expense of about 40% current consumption increase.

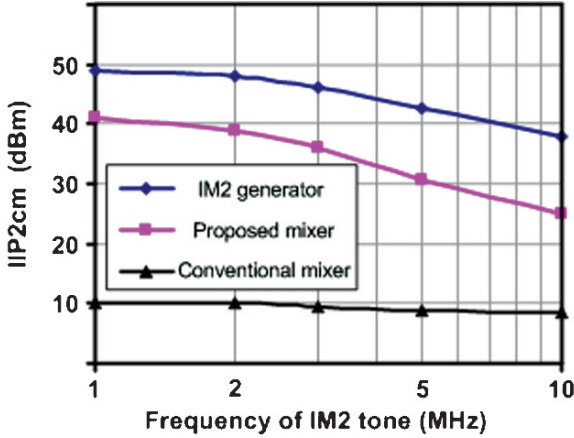


Fig. 8.9 The IIP2_{CM} of the mixer with and without cancellation loop versus the frequency of the IM2 tone besides the IIP2_{CM} of IM2 generator

Figure 8.9 shows the IIP2_{CM} of the mixer with and without cancellation feedback which is compared with IIP2_{CM} of the IM2 generator. It can be concluded that the IIP2_{CM} improvement in GSM, UMTS and IEEE802.11 are about 31, 29 and 17 dB respectively. The improvement of IIP2_{CM} in IEEE802.11 is lower than UMTS because of the lower LG of the cancellation loop in 10MHz comparing to 2MHz. However, the IIP2 requirement of IEEE802.11 can also be met because it is more relaxed than that of cell-phone mixers [3]. Moreover this figure shows that the attenuation of IM2_{CM} current of M1 transistor is less than that of M3 transistor (IM2 generator), because the IM2 cancellation performance in M1 transistor is degraded by mismatch between M1 and M3 transistors.

The input referred noise, which is in-band averaged noise (10 kHz–1.92 MHz), is less than 3.4nV/sqrt(Hz).

The third-order nonlinearity is simulated with the transmitter signal leakage inter-modulating with a blocker which is 67.5 MHz away from the receive band [5]. The mixer IIP3, conversion gain and power consumption are 8 dBm, 12 dB and 8.5 mW, respectively.

8.5 IIP2 Calibration Technique

The IM2_{diff} current of pseudo differential transconductor is mainly due to the variation in the threshold voltage [8]:

$$I_{IM2,diff} \propto \frac{3g_2}{4(V_{gs} - V_{TH})} \sigma_{V_{TH}} \tag{8.26}$$

In which; V_{gs} , V_{TH} and $\sigma_{V_{TH}}$ show the gate-source voltage, threshold voltage and the standard deviation of V_{TH} , respectively.

By using Eqs. (8.22) and (8.26):

$$\frac{I_{IM2,diff}}{I_{IM2,CM}} = \frac{3\sigma_{V_{TH}}}{2(V_{gs} - V_{TH})} \quad (8.27)$$

Assuming $V_{gs} - V_{TH} = 100$ mV and $\sigma_{V_{TH}} = 2$ mV, this ratio is about 0.03.

Since, IIP2 is reversely proportional with IM2 current, it can be concluded that in pseudo differential transconductor the $IM2_{CM}$ current is more powerful than $IM2_{diff}$ current by more than one order of magnitude. This can be written as below:

$$\left(\frac{I_{IM2,diff}}{I_{IM2,CM}} \right)^2 = \alpha^2 \ll 1 \quad (8.28)$$

The CM and differential IIP2 of a pseudo differential transconductor versus biasing current is shown in Fig. 8.10.

Moreover, considering a pessimistic value for L and $\frac{\Delta R_L}{R_L}$ such as -40 dB in Eq. (8.5), it can be found that $IM2_{diff}$ current at mixer output is much lower than $IM2_{diff}$ current at transconductor output. Therefore, a good idea for canceling differential IM2 current at the mixer output is converting the $IM2_{CM}$ current of transconductor to the differential part and then injecting it to the mixer output after attenuation. As the proposed calibration system is shown in Fig. 8.11, an IM2 generator block senses the input RF blockers, generates the $IM2_{CM}$ current and converts it to the required differential IM2 components which is controlled by a programmable OTA and is injected to the mixer output. Referring to Fig. 8.12, The IM2 generator is made by M3 and M4 transistors which are in parallel with mixer input transistors (M1, M2).

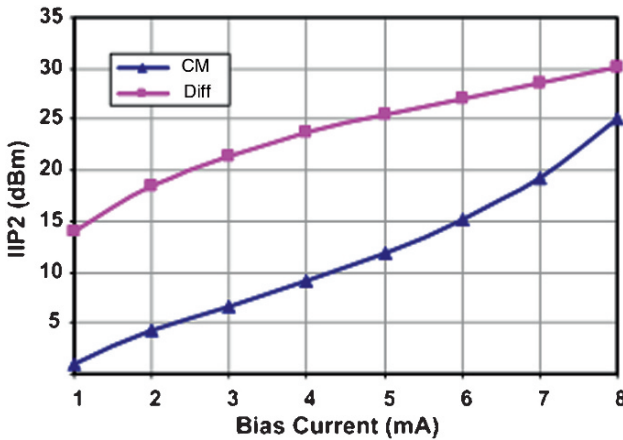


Fig. 8.10 The CM and differential IIP2 of a pseudo differential transconductor versus biasing current ($W = 120$ μ m, $L = 0.15$ μ m)

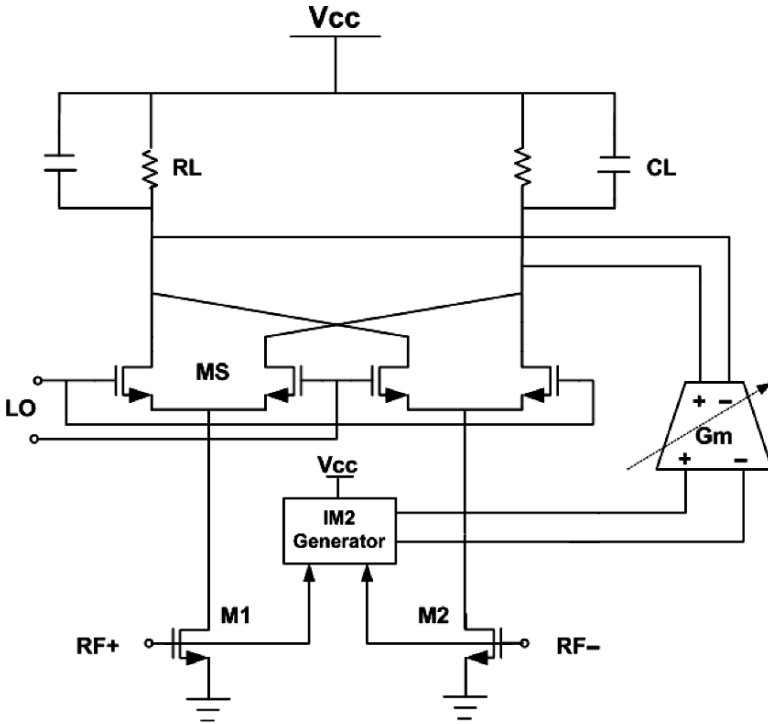


Fig. 8.11 Block diagram of the IIP2 calibration technique

8.5.1 The Proposed Calibration Technique

The analytical analysis of the proposed calibration technique is as follows:

Considering Eqs. (8.5) and (8.28), the relationship between differential IM2 current at the uncalibrated mixer output ($I_{IM2,out}$) with $IM2_{CM}$ current generated in the mixer input stage can be simplified as below (σ_R is the standard deviation of ΔR_L):

$$I_{IM2,out} = I_{IM2,CM} \sqrt{L^2(1 + \alpha^2) + \frac{\sigma_R^2}{R_L^2}} \tag{8.29}$$

Since transistor M3 and M1 are in parallel in the input but with different biasing and sizing, by using Eqs. (8.22) and (8.25) it can be found that the $IM2_{CM}$ current generated in M3 transistor is proportional to that of M1 transistor. This is shown as below:

$$I_{IM2,CM,M3} = \beta I_{IM2,CM,M1} \tag{8.30}$$

Suppose that $IM2_{CM}$ current of M3 and M4 transistors is converted to differential and is injected to the mixer output:

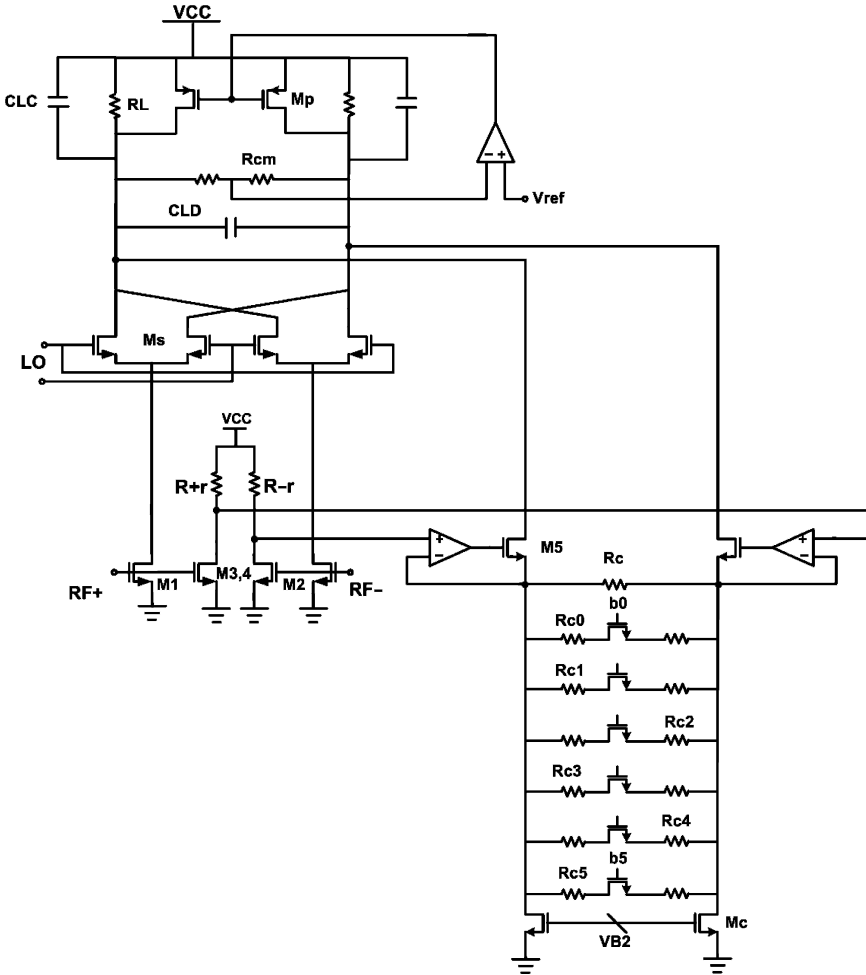


Fig. 8.12 The proposed calibrated mixer

$$I_{IM2,inj} = -(g_{m_{Cal}} \cdot r) I_{IM2,CM,M3} \tag{8.31}$$

In which $I_{IM2,inj}$ is the injected current and other terms show the CM to differential conversion gain. The $g_{m_{Cal}}$ is the OTA transconductance and r shows the desired mismatch between load resistance of M3 and M4 transistors.

Considering Eqs. (8.29) and (8.31), the $IM2_{diff}$ current in the output of the calibrated mixer ($I_{IM2,out,Cal}$) is:

$$I_{IM2,out,Cal} = I_{IM2,CM} \left(\sqrt{L^2(1 + \alpha^2) + \frac{\sigma_R^2}{R_L^2}} - \beta \cdot g_{m_{Cal}} \cdot r \right) \tag{8.31}$$

Therefore, it can be concluded that the IM2 cancellation in the output of the mixer can happen if the following condition is met.

$$g_{m_{Cal}} = \frac{\sqrt{L^2(1 + \alpha^2) + \left(\frac{\sigma R}{R_L}\right)^2}}{\beta \cdot r} \tag{8.32}$$

8.5.2 Circuit Implementation

The proposed calibration technique is implemented as it is shown in Fig. 8.12. The required IM2_{CM} current, generated by an IM2 generator, is converted to differential by the desired mismatch between the load resistors of IM2 generator. The output of the IM2 generator is connected to a highly linear OTA stage.

Since resistors are more linear than MOS transistors biased in the triode region, a resistor is chosen as the voltage to current converter in the OTA. This resistor determines the value of transconductance of the OTA as below:

$$g_{m_c} = \frac{1}{R_c} \tag{8.33}$$

To achieve programmability in the OTA, this resistor is implemented by a fixed resistor (R_c) shunted with some switched resistors controlled by a 6-bit control word (b₀–b₅). These control bits provide enough margins to fix PVT variations.

To avoid the nonlinearity of the OTA input transistors two Opamps (Fig. 8.12), which are two stage miller compensated amplifiers, are employed as it is shown in Fig. 8.13.

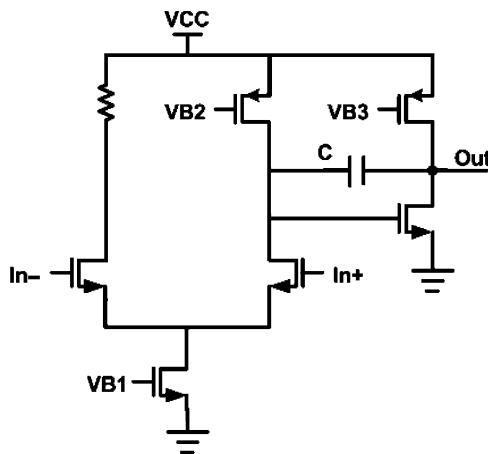


Fig. 8.13 The two stage amplifier used in the input of OTA

The bandwidth of the calibration circuit is dominated by the input stage of the OTA. The GBW of the OTA input stage is higher than 10 MHz because of using opamp with about 20 MHz GBW in a unity gain feedback. The second pole of the loop is in the source node of M5 transistor which is very far from the first pole.

To cancel the all IM2 components in the signal channel, the calibration circuit bandwidth should be higher than half of the signal bandwidth in a Zero-IF receiver. Therefore; the proposed calibration circuit can work for high channel bandwidth applications like UMTS and IEEE802.11 as well as GSM.

The temperature variation is an issue for all calibration circuits. A general solution for this kind of variations is based on sensing the temperature effect in the calibration circuit and adjusting the proper calibration code by considering the temperature variation effect. This adjustment is done in [16] by varying the temperature coefficient of the resistors. Since a LMS feedback algorithm in which the calibration is updated periodically should be employed to adjust the calibration [15], therefore, the temperature and other low speed variations can be compensated if the calibration code setting is updated every few seconds.

8.5.3 Noise in the Calibrated Mixer

The main problem of the differential calibration is the flicker and thermal noises of the calibration circuit which are injected directly to the mixer output. Moreover, in cellular phone applications and especially GSM there is tough requirement on the noise performance which is mainly degraded by flicker noise of the mixer. Therefore, the noise of the calibration circuit should be highly lower than mixer noise in order not to degrade the mixer noise figure.

Considering the thermal noise, the noise of calibration circuit seen in the output of the mixer is:

$$v_{n1,Cal}^2 \approx 8KTR_L^2 \left(\gamma \cdot gm_3 \frac{R^2}{R_C^2} + \frac{1}{R_C} + \gamma \cdot gm_5 \right) + v_{n,op}^2 \frac{R_L^2}{R_C^2} \quad (8.34)$$

In which, the terms from the first to the last are the noise of M3 transistor, R_C resistor, M5 transistor and Opamp, respectively. Moreover, γ is the channel noise factor which is 2/3 for long channel MOS transistors. The noise of resistor R (the load of M3 and M4 transistors) can be neglected because:

$$R \ll R_C \quad (8.35)$$

Moreover, the thermal noise of R_C is also reduced by choosing R_C highly larger than R_L .

$$R_C^2 \gg R_L^2 \quad (8.36)$$

Increasing R_C leads to lower CM to differential conversion gain in IM2 generator and OTA which is equal to $\frac{r}{R_C}$. However this can be compensated by increasing r.

By using Eqs. (8.34)–(8.36), it can be shown that the M5 transistor has the main contribution in the thermal noise of calibrating circuitry.

The thermal noise of the uncalibrated mixer which mainly is dominated by input transistors is [10]:

$$v_{n,mix}^2 \approx 8KT \left(\frac{1}{R_L} + \gamma gm_1 + \gamma \frac{2}{\pi} \frac{I}{A} \right) R_L^2 \quad (8.37)$$

In which A is the LO amplitude and I is the bias current of input transistors.

Therefore, considering main noise sources in Eqs. (8.34) and (8.37), the thermal noise in calibrating circuitry is negligible comparing to uncalibrated mixer, if the following equation is met.

$$gm_5 \ll gm_1 \quad (8.38)$$

Which it can easily be satisfied because, firstly the bias current of M5 transistor is much lower than that of the input RF transistors and secondly the channel length of M5 transistor is highly larger than the channel length of M1 transistor which acts as a RF transistor and its size is limited due to low requirement in input capacitive load of the mixer.

Regarding the flicker noise, the noise of the calibration circuit seen at the mixer output is as below:

$$v_{n2,Cal}^2 \approx 2(v_{n,M3}^2 gm_3^2 \frac{R^2}{R_C^2} + v_{n,M5}^2 gm_5^2 + \frac{v_{n,Op}^2}{2R_C^2}) R_L^2 \quad (8.39)$$

In which, the main flicker noise contributors including M3 transistor, M5 transistor and Opamp ($v_{n,M3}^2$, $v_{n,M5}^2$, $v_{n,Op}^2$) are considered.

The flicker noise of M3 transistor is reduced by lowering the biasing current and using non minimum size channel length, while in M5 and Opamp it is achieved by lowering bias currents and increasing the area of transistors. By this strategy, the low noise performance in Opamp can be achieved without affecting the gain and bandwidth requirements.

8.5.4 Proposed Mixer

A mixer, enhanced by the proposed cancellation technique, for UMTS application is designed and is shown in Fig. 8.12. The M1 and M2 transistors are biased each by 2 mA and are sized by 120/0.15 μm . The input stage of the IM2 generator, including M3 and M4 transistors, are sized by 10/0.15 μm which is lower than that of M1 and M2 transistors. This is because of reducing noise of the calibration circuitry (as it is mentioned in part C of Section 8.5) and also providing less capacitive load in the input stage. M3 and M4 transistors are biased each by 0.15 mA and connected to 100 and 200 Ohms resistors. The OTA resistor R_C is chosen 2K Ohms. The OTA input stage transistor is biased by 0.3 mA and is sized as 160/1 μm .

The switching transistors are derived by a local oscillator with 5 dBm power and are sized by 400/0.24 μm .

Two PMOS transistors (M_p) with long channels are responsible for providing biasing current in the mixer output. Two RL resistors providing required conversion gain are also shunted with ML transistors for reduction of their current and achieving lower voltage headroom. The common and differential mode filtering in IF is done by CLC and CLD capacitors.

8.5.5 Simulation Results

The simulation is done by the conditions mentioned in part D of section IV. The Monte Carlo simulation of the calibrated and uncalibrated mixer is depicted in Fig. 8.14. This figure shows that the IIP2 of the proposed mixer is more than

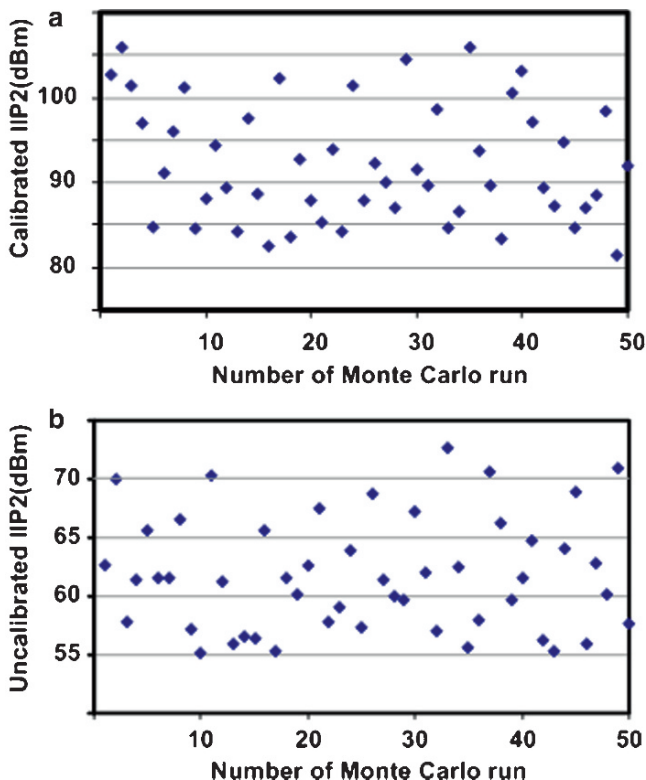


Fig. 8.14 (a) The result of Monte Carlo analysis of the enhanced mixer, (b) The result of Monte Carlo analysis of the conventional mixer

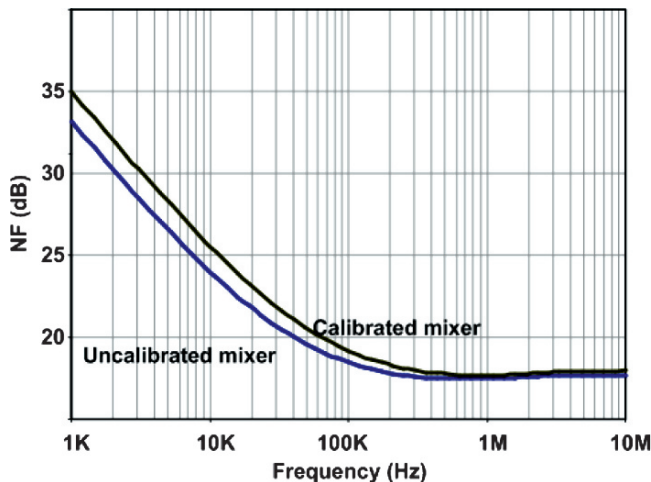


Fig. 8.15 NF in calibrated and uncalibrated mixers

80 dBm which is higher than the cellular phones requirement (75 dBm), while in the conventional mixer it less than 75 dBm. Moreover it shows that more than 25 dB improvement in IIP2 is achieved.

The input referred noise which is in-band averaged noise (10 kHz–1.92 MHz) is less than $3.8\text{nV}/\sqrt{\text{Hz}}$. The mixer NF performance in both calibrated and uncalibrated modes is shown in Fig. 8.15. The mixer IIP3, conversion gain and power consumption are 7 dBm, 12 dB and 6 mW, respectively.

8.6 Second Order Nonlinearity of Switching Pairs

As it is explained, the limited dynamic range of mixer when direct down-converting an RF signal to DC or close to DC, is firstly caused by the limited bandwidth of the switching core. In particular, the parasitic capacitor, loading the pair source, is charged and discharged during each LO period. Noise components and low frequency inter-modulation tones modulate the charging and discharging capacitor current, being finally down-converted around DC at mixer output. A very effective counter-measure consists in tuning out the parasitic capacitor by means of an inductor as it is shown in Fig. 8.16. The node impedance raises by Q , the inductor quality factor, and by the same amount increases the IIP2 of the switching core [3, 10]. A significant improvement both in $1/f$ and white noise comes along.

Because this technique is intrinsically narrow-band, it does not directly lend itself to address a universal radio, where the number of area hungry inductors would equal the number of standards to be tuned. The idea, proposed here, is to exploit the properties of a transformer in order to realize a programmable inductor. Covering a band as wide as 900 MHz to 6 GHz would require an inductance change by ~ 16 ,

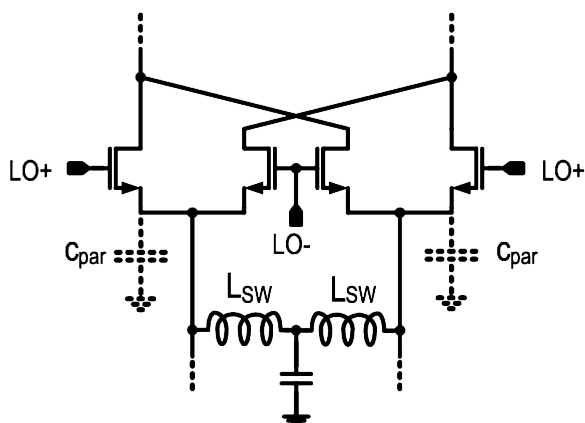


Fig. 8.16 Mixer switching pair set with LC filter for maximum dynamic range

which is not feasible. The strategy is then employing the programmable inductor in the 1.7–2.4 GHz, the spectrum portion most crowded of wireless standards, while using dedicated inductors at 900 MHz and between 5–6 GHz for wireless LANs.

8.6.1 Programmable Inductor

The most straightforward way to tune the filter of Fig. 8.16 is by means of a varactor in parallel with the resonating inductor. In ultra-scaled CMOS nodes, varactors prove very high quality factors at RF frequency thus not degrading the filter Q . But, any capacitor added to the source node has a detrimental effect on second order inter-modulation performance. While this is intuitive because the charging and discharging transients become longer and any tone injected into the pair is more effective in producing inter-modulation products, quantification is not. In fact the circuit is non linear, cyclo-stationary and with memory. For an accurate model and discussion, refer to [29]. The switching pair IIP2 strongly depends on design and biasing conditions, e.g. device dimensions, biasing current, operating frequency. To gain insight, let us assume a varactor is used to tune the filter between 1.7 and 2.2 GHz and the switching devices are 440/0.28 μm . Considering 0.6 pF of differential parasitic node capacitance (including parasitics to substrate of switching and input transistor), this means the differential varactor maximum capacitance is 0.45 pF. Figure 8.17 plots the simulated switching pair IIP2 versus current assuming a source capacitance, made of parasitics only (curve a), parasitics plus 0.2 pF (curve b) and parasitics plus maximum varactor capacitance (curve c).

In this simulation, the capacitors are not tuned out by the inductor. Though variable in the proposed current range, the degradation (from a–c) is >12 dB for $I > 2.5$ mA. When set with the tuning inductor, the switching pair improves IIP2 performance by roughly a factor Q , the inductor quality factor. Nonetheless, this

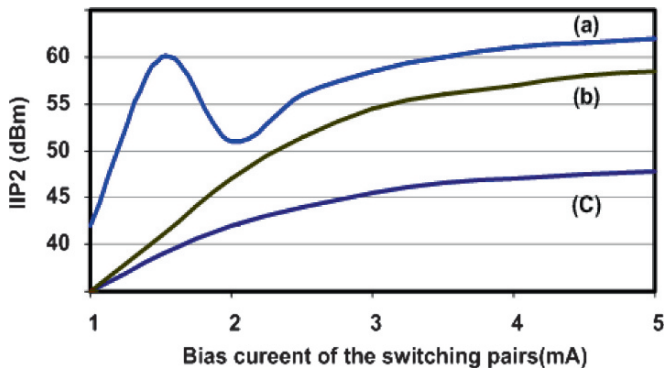


Fig. 8.17 Switching pairs IIP2 versus biasing current when the source capacitance is not resonated out. (a) parasitic only, (b) parasitic plus 0.2 pF capacitance, (c) parasitic plus 0.45 pF

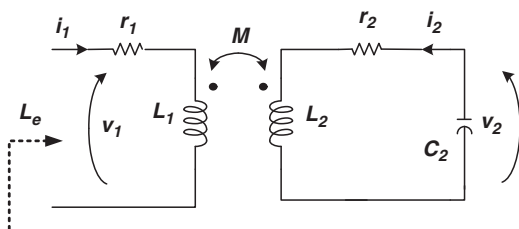


Fig. 8.18 Simplified model of the programmable inductor

tuning method does not lend itself to a highly linear solution, as required by cell-phone applications and for this reason disregarded.

Instead of varying the capacitance, the filter can be tuned varying the inductance. Active inductors have been proposed for RF applications, though mainly limited to tune the frequency in Voltage Controlled Oscillators and filters [33]. On the contrary, they did not make in-roads in the RF front-end blocks, mainly because the active devices, making up the inductor, introduce non-linearities and noise, besides increasing power consumption.

As an alternative, an inductor can be implemented using passive components only, i.e. based on an impedance transformation concept. Referring to Fig. 8.18, if the secondary winding of a transformer is loaded by means of a capacitor, the equivalent reactance at the primary winding is inductive [34]. In fact, capacitor C_2 forces V_2 and i_2 to be orthogonal and, as a consequence, i_1 and i_2 in-phase. If the capacitor is made variable, also the reactance at the primary varies, thus realizing a variable inductor. The system of equation governing the circuit of Fig. 8.18 is found by inspection:

$$\begin{cases} v_1 = (j\omega \cdot L_1 + r_1) \cdot i_1 + j\omega \cdot M \cdot i_2 \\ v_2 = j\omega \cdot M \cdot i_1 + (j\omega \cdot L_2 + r_2) \cdot i_2 \\ i_2 = -j\omega \cdot C_2 \cdot v_2 \end{cases} \quad (8.40)$$

Where L_1 , L_2 and M are self-inductance of primary and secondary windings and their mutual inductance, respectively. The ohmic losses are taken into account by means of series resistors (r_1 and r_2) for sake of model simplicity.

Solving (8.40) for the equivalent inductance (L_e) and resistance (r_e) seen at transformer primary, provides:

$$L_e(\omega) = L_1 + \frac{i_2}{i_1}M = L_1 + \frac{L_2 C_2 \omega^2}{1 - L_2 C_2 \omega^2} k^2 L_1 \quad (8.41)$$

And

$$r_e = r_1 + \left(\frac{\Delta L}{M} \right)^2 \cdot r_2 \quad (8.42)$$

where k is the magnetic coupling coefficient and L is:

$$k = \frac{M}{\sqrt{L_1 \cdot L_2}} \quad (8.43)$$

$$\Delta L = \frac{L_2 C_2 \omega^2}{1 - L_2 C_2 \omega^2} k^2 L_1 \quad (8.44)$$

Equations (8.41) and (8.42) hold provided the operation frequency is much lower than the secondary self-resonance frequency and Q_2 , the secondary winding quality factor, is much higher than 1. Both these conditions are verified in practice.

By definition, the equivalent quality factor of the programmable inductor (Q_e) is given by:

$$Q_e = \frac{L_e \cdot \omega}{r_e} = \frac{(L_1 + \Delta L) \cdot \omega}{r_e} \quad (8.45)$$

By substituting the Eqs. (8.40)–(8.43) in (8.45) and assuming Q_1 as the Q of primary winding, the following equation is derived for Q_e :

$$Q_e = Q_1 \cdot \frac{\left(1 + \frac{\Delta L}{L_1}\right)}{1 + \left(\frac{\Delta L}{k \cdot L_1}\right)^2 \cdot \frac{Q_1}{Q_2}} \quad (8.46)$$

For small ΔL ($\Delta L \ll L_1$) the equivalent Q is equal to the quality factor of the primary winding, as expected. For large ΔL the equivalent Q tends to reduce because the power dissipation in the secondary winding increases.

To gain quantitative insight, Fig. 8.19 plots Eqs. (8.41) and (8.45) versus C_2 at $\omega = 2$ GHz, for $L_1 = 4.3$ nH, $L_2 = 2.1$ nH, $k = 0.75$ and assuming both inductors have a Q of 12.

The inductance variation is large, 2x in the 0–2 pF capacitance range. Notice that at higher capacitance values, the inductance increases faster because the $L_2 C_2$ network approaches resonance. On the other hand, Q_e decreases from 12 to about 8 in the same range. A compromise between inductance range, i.e. filter frequency

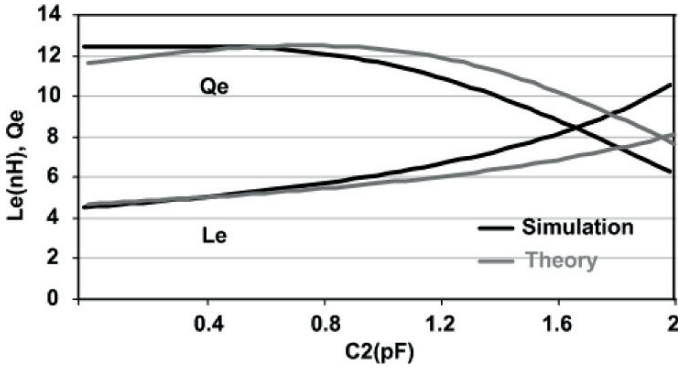


Fig. 8.19 Equivalent inductance (L_e) and quality factor (Q_e) versus capacitance C_2

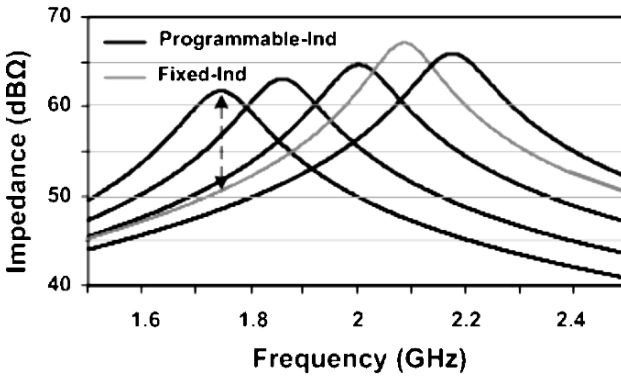


Fig. 8.20 Impedance of the LC network tuning out 0.6 pF switch pair parasitic capacitance: (a) the programmable inductor implements L , (b) a spiral fixed inductor implements L

tuning range, and Q exists. Because the model assumed for analytical calculation is simplified, we have also plotted a simulation where an accurate transformer model in 65 nm CMOS is included. The two curves do not significantly deviate.

This programmable inductor lends itself to tune out source node parasitics in a wide-band range. The expected IIP2 improvement is proportional to the source impedance increase. Because both Q and L vary in band, so does the network impedance. Simulations are plotted versus frequency in Fig. 8.20, with the programmable inductor tuning out 0.6 pF parasitic capacitor. The maximum variation in the overall range is 4 dB .

For the sake of comparison, the impedance of a fixed LC, tuning out the same capacitance at 2.1 GHz , has been simulated and plotted in the same figure. As expected the latter impedance quickly falls over frequency (roughly 12 dB at $f = 1.8\text{ GHz}$), thus being inadequate as tuning element in a wide frequency range.

8.6.2 Proposed Implementation

The core of the resonant network, employed to tune out switched pair parasitics, is reported in the dashed box of Fig. 8.21, where the complete mixer is detailed.

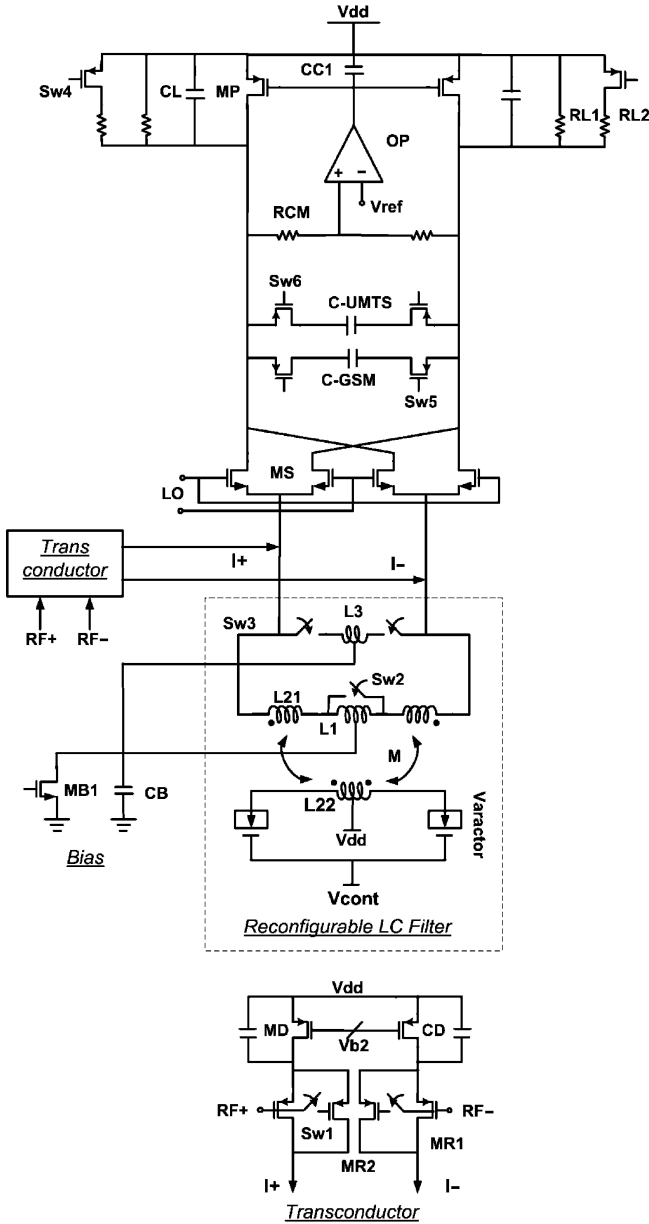


Fig. 8.21 Schematic of the proposed multi-standard mixer

Inductors L_1 and L_3 are used in GSM and IEEE 802.11a bands, respectively. L_3 is a compact, small inductor with relatively low Q , used primarily to reduce switch white noise and minimize gain reduction. L_1 is series connected to the transformer primary and its Q is of primary concern. Moreover switch S_{W2} is critical because its on resistance degrades the programmable inductor Q . The switches are sized as 160/60 nm, the best compromise between low on-resistance and off parasitic capacitance.

The transformer-based programmable inductor has been customized by means of the full-wave electromagnetic solver Agilent Momentum, in order to maximize magnetic coupling k , primary winding quality factor and self-resonance frequency. Three top metal layers shorted together are employed for maximum Q . Inter-wound conductors, on the same plane, have been adopted. The primary and secondary windings are 7 and 9 μm wide, respectively. The two windings use a common center tap.

From simulations, $L_{21} = 2 * 4.3\text{nH}$, $L_{22} = 2 * 2.1\text{nH}$ and k equals 0.75. Inductors L_1 and L_3 are 16 and 1 nH, respectively.

While the differential inductor also filters out differential second order inter-modulation components injected by the transconductor, it is not effective against common mode ones. Selection of a fully differential transconductor assures the best IIP2 common mode performance versus a pseudo-differential alternative [30, 31]. On the contrary, to maximize IIP3 the latter is preferable [29]. An RC degenerated topology compromises the two parameters, by virtue of its high impedance at low frequency and low impedance at RF frequency. In particular, an RC degenerated topology makes the transconductor contribution to mixer IIP2 negligible while improving IIP3 with respect to a fully differential implementation [10]. The transconductor implementation, highlighted in Fig. 8.21, makes use of pMOS devices to allow a folded mixer topology stacked within 1.2 V.

The mixer load is made of pMOS transistors (M_P) with long channels, shunted by resistors R_L , in order to maximize mixer gain for given voltage room.

Considering the different requirements set by different standards, e.g. higher IIP3 in UMTS, lower flicker noise in GSM, different output bandwidths, several features tailored to the specific application have been implemented [26, 35]. For example, transistors MR2 in the transconductor are shunt connected to MR1 in GSM and DCS in order to maximize gain thus reducing the switching pair $1/f$ noise impact [36]. On the contrary, the stringent IIP3 requirement set by UMTS asks for a lower gain. The output bandwidths are set changing the output capacitance.

Transistor M_{B1} , connected to the centre tap of the differential inductors L_3 and L_1 and bypassed by C_B capacitor at RF, biases the mixer. Notice that M_{B1} does not contribute to mixer NF because its thermal and flicker noises are common mode noise and consequently rejected at differential output.

8.6.3 Experimental Results

The proposed mixer has been integrated in a standard 65 nm CMOS technology from STMicroelectronics. The chip is encapsulated in a TQFP48 plastic package and soldered on a dedicated double-sided RF board with $\tan \delta = 0.027$ and $\epsilon_r = 3.38$ [37]. External 180° hybrid couplers (BD0810 for GSM, 3W525 for 1.8–2.4 GHz band and BD4859 for IEEE802.11a mode, all from Anaren) were used both at the RF and LO ports to implement the single-ended-to-differential conversion. 50 ohms strip lines, optimized by means of EM simulations, carry the differential signals from the SMD connectors to the package inputs. The mixer draws 7.5 mA from a 1.2 V supply.

The IIP2 performance was evaluated, injecting a double side-band with suppressed carrier tone according to standard specifications, i.e. at 6 MHz from all GSM versions, 130 MHz for UMTS, and 20 MHz for 802.11. The tones spacing is chosen in a way that the second order intermodulation components stay in-band. The IIP2 measurement setup and results for UMTS mode are shown in Figs. 8.22 and 8.23.

Due to the narrow channel spacing, low frequency noise is particularly critical in GSM. The $1/f$ noise corner is 110 and 140 kHz in 900 and 1,800 MHz frequencies, respectively. The input referred noise of GSM900 is $5 \text{ nV}/\sqrt{\text{Hz}}$ when integrated from 100–200 kHz. The conversion gain versus frequency is plotted in Fig. 8.24.

Measurement results are summarized in Table 8.1. The die photo is presented in Fig. 8.25 and the total area including pads is 2.5 mm^2 .

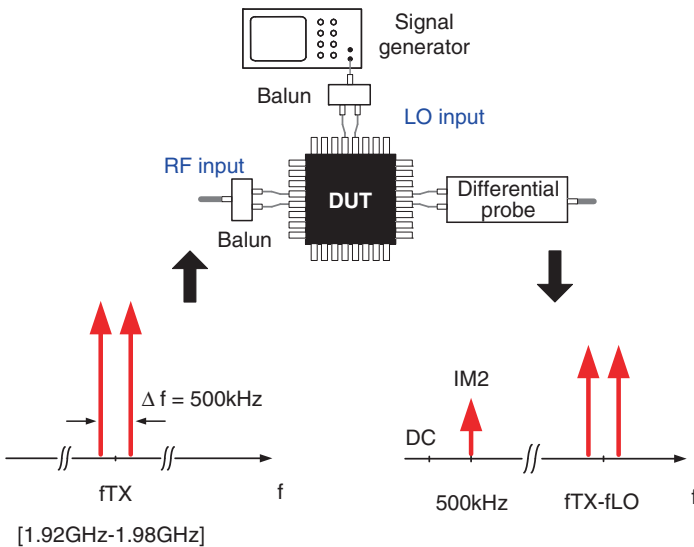


Fig. 8.22 IIP2 measurement setup (UMTS Mode)

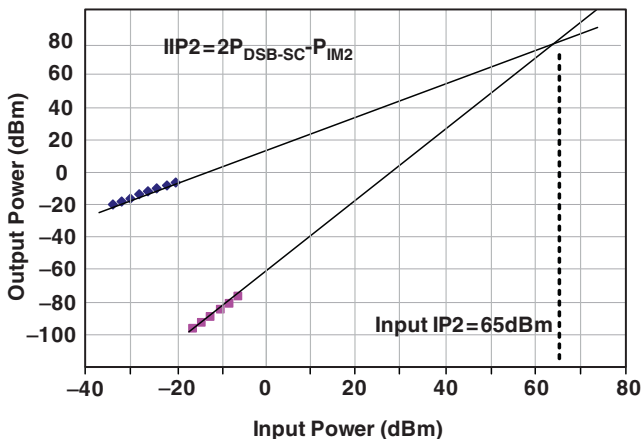


Fig. 8.23 IIP2 measurement results in UMTS mode

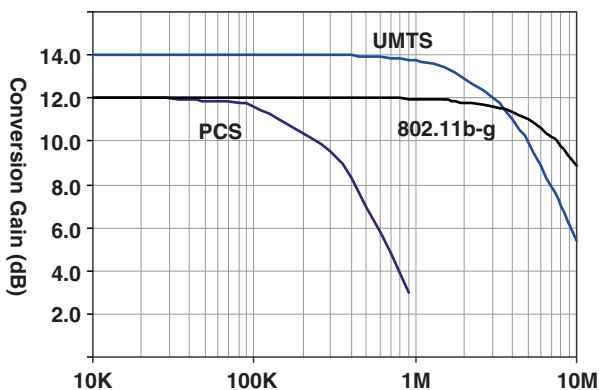


Fig. 8.24 Mixer conversion gain in different bands

Table 8.1 Measurements summary

Specification	GSM	DCS/PCS	UMTS	IEEE 802.11b-g	IEEE 802.11a
Process	STM 65 nm CMOS				
Supply (V)	1.2				
Current consumption (mA)	7.5				
Frequency (GHz)	0.9	1.8, 1.9	2.1	2.4	5.15–5.35
Conversion gain (dB)	12	12	14	12	11
IIP2 (dBm)	71	65	65	60	54
Input ref. noise (nV/sqrt(Hz))	5	5.4	3.8	3.6	4
	1–200 kHz				
IIP3 (dBm)	4	4	7	5	3

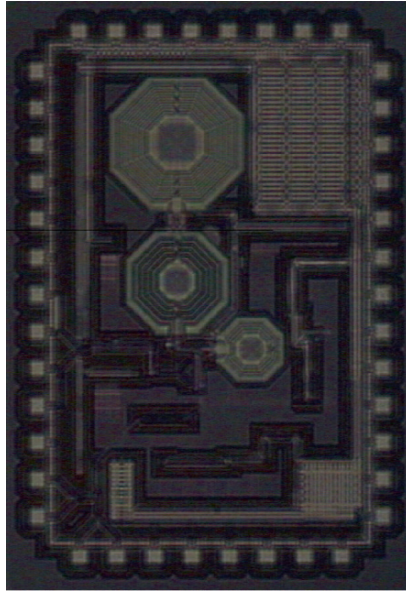


Fig. 8.25 Die photo of proposed mixer

8.7 Conclusion

The design of a CMOS mixer for cellular phone and 3G applications is challenging because of stringent linearity and noise requirements. Some IIP2 improvement circuits including both calibration techniques and simple analog techniques are reviewed in order to achieve the required dynamic range. Several mixers, supporting GSM, DCS, PCS, UMTS and IEEE802.11b-g-a applications, enhanced by these techniques and designed in 65 nm CMOS technology are presented.

The $IM2_{CM}$ generated in the pseudo differential input stage of the mixer is attenuated by employing a feedback loop. Making use of this feedback loop about 22 dB improvement in the mixer IIP2 is achieved. The mixer meets the tough cellular phone IIP2 requirement while it does not affect other mixer parameters like NF.

An IIP2 calibration technique based on injecting a nonlinear current to the mixer output is presented. This current is made in an $IM2$ generator and is controlled by an OTA. Making use of this technique more than 25 dB improvement in the mixer IIP2 is achieved.

Considering the switching pairs, outstanding performance from Gilbert based mixers are achieved in narrow-band applications by resonating out the switching pair parasitics. In order to extend to wide-band applications, a transformer based tunable inductor is introduced. This tunable inductor has been employed in the 1.7–2.4 GHz band while dedicated inductors at 900 MHz and between 5–6 GHz for wireless LANs has been used.

The mixer enhanced by above mentioned techniques meets the tough cellular phone IIP2 requirement. The design works with low voltage supplies and

has enough bandwidth for high channel bandwidth applications like UMTS and IEEE802.11 as well as GSM; therefore it is a good candidate for high IIP2 multi-standard mixers. The linearity performance of this mixer goes toward the demand of fully integrated, hardware shared universal mobile terminal.

Acknowledgement The authors would like to thank Prof. Francesco Svelto from universita degli studi di Pavia.

References

1. R. Bagheri, A. Mirzaei, S. Chehrazi, M. E. Heidari and A. A. Abidi, "An 800MHz–6GHz software defined wireless receiver in 90nm CMOS," *IEEE J. Solid-State Circuits*, vol. 41, no. 12, pp. 2860–2876, Dec. 2006.
2. N. Poobuanheun, W. chen, Z. Boos and A. M. Niknejad, "A 1.5-V 0.7–2.5-GHz CMOS quadrature demodulator for multiband direct-conversion receivers," *IEEE J. Solid-State Circuits*, vol. 42, no. 8, pp. 1669–1677, Aug. 2006.
3. M. Brandolini, P. Rossi, D. Manstretta and F. Svelto, "Toward multistandard mobile terminals-fully integrated receivers requirements and architectures," *IEEE Trans. on Microwave Theory and Architecture*, vol. 53, no. 3, pp.1026–1035, Mar. 2005.
4. J. Mitola, "The software radio architecture," *IEEE Commun. Mag.*, vol. 33, no. 5, pp. 26–38, May 1995.
5. A. Liscidini, M. Brandolini, D. Sanzogni and R. Castello, "A 0.13 μ m CMOS front-end, for DCS1800/UMTS/802.11b-g with multi-band positive feedback low noise amplifier," *IEEE J. Solid-State Circuits*, vol. 41, no. 4, pp. 981–989, Apr. 2006.
6. E. Götz et al., "A quad-band low power single chip direct conversion CMOS transceiver with Delta sigma modulation loop for GSM," in *Proc. ESSCIRC'03*, Sep. 2003, pp. 217–220.
7. E. Duviolier, G. Puccio, S. Cipriani, L. Carpineto, P. Cusinato, B. Bisanti, F. Galant, F. Chalet, F. Coppola, S. Cercelaru, N. Vallespin, J.-C. Jiguet, and G. Sirna, "A fully integrated zero-IF transceiver for GSM-GPRS quad-band application," *IEEE J. Solid-State Circuits*, vol. 38, no. 12, pp. 2249–2257, Dec. 2003.
8. M. Terrovitis and R. G. Meyer, "Intermodulation distortion in current-commutating CMOS mixers," *IEEE J. Solid-State Circuits*, vol. 35, no. 10, pp. 1461–1473, Oct. 2000.
9. I. Elahi, K. Muhammad and P. T. Balsara, "IIP2 and DC offsets in the presence of leakage at LO frequency," *IEEE Trans. Circuits Syst. II*, vol. 53, no. 8, pp. 647–651, Aug. 2006.
10. M. Brandolini, P. Rossi, D. Sanzogni and F. Svelto, "A CMOS direct down-converter with +78dBm minimum IIP2 for 3G cell-phones," *International Solid State Circuit Conference, IEEE Proc.of*, pp. 320–322, Feb. 2005.
11. J. Rogin, I. Koucev, G. Brenna, D. Tschopp, and Q. Huang, "A 1.5 V 45 mW direct conversion WCDMA receiver IC in 0.13 μ m CMOS," *IEEE J. Solid-State Circuits*, vol. 38, no. 12, pp. 2239–2248, Dec. 2003.
12. S. K. Reynolds, B. A. Floyd, T. Beukema, T. Zwick, U. Pfeiffer, and H. Ainspan, "A direct-conversion receiver IC for WCDMA mobile systems," *IEEE J. Solid-State Circuits*, vol. 36, no. 9, pp. 1555–1560, Sep. 2003.
13. H. Waite et al., "A CDMA2000 zero IF receiver with low-leakage integrated front-end," in *Proc. Eur. Solid-State Circuits Conf.*, Sep. 2003, pp. 433–436.
14. A.R. Brown, and G.M.Rebeiz, "A Varactor-Tuned RF Filter", *IEEE Transactions on Microwave Theory and Techniques*, Vol. 48, n.7, pp. 1157–1160, July 2000.
15. H. Darabi, H. Joung Kim, J. Chiu, B. Ibrahim and L. Serano, "An IP2 improvement technique for Zero-If down-Converters," *International SolidState Circuit Conference, IEEE Proc.of*, pp. 464–465, Feb. 2006.

16. M. Chen, Y. Wu and M. Chang, "Active 2nd order intermodulation calibration for direct conversion receivers," *International SolidState Circuit Conference, IEEE Proc. of*, pp. 458–459, Feb. 2006.
17. S. Dow et al., "A dual-band, direct-conversion/VLIF transceiver for 50 GSM/GSM/DCS/PCS," in *Int. Solid-State Circuit Conf. Tech. Dig.*, vol. 1, Feb. 2002, pp. 230–462.
18. E. Duviolier et al., "A fully integrated zero-IF transceiver for GSM–GPRS quad-band application," *IEEE J. Solid-State Circuits*, vol. 38, no. 12, pp. 2249–2257, Dec. 2003.
19. M. Brandolini, M. Sosio and F. Svelto, "A 750mV, 15kHz 1/f noise corner, 51dBm IIP2, direct conversion front-end for GSM in 90nm CMOS," *International SolidState Circuit Conference, IEEE Proc. of*, pp. 374–376, Feb. 2006.
20. K. Kivekas, A. Parssinen, and K. Halonen, "Characterization of IIP2 and DC-offset in transconductance mixers," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 48, no. 11, pp. 1028–1038, Nov. 2001.
21. Digital Cellular Telecommunications System (Phase 2); Radio Transmission and Reception, GSM Standard 05 05, 1999.
22. J. Sevenhans, A. Vanwelsenaers, J. Wenin, and J. Baro, "An integrated Si bipolar transceiver for a zero IF 900 MHz GSM digital mobile radio front-end of a hand portable phone," in *Proc. Custom Integrated Circuits Conf.*, May 1991, pp. 771–774.
23. S. Tadjipour, E. Cijvat, E. Hegazi, and A. Abidi, "A 900-MHz dualconversion low-IF GSM receiver in 0.35- μ m CMOS," *IEEE J. Solid-State Circuits*, vol. 36, no. 12, pp. 1992–2002, Dec. 2001.
24. M. Steyaert, J. Janssens, B. De Muer, M. Borremans, and N. Itoh, "A 2-V CMOS cellular transceiver front-end," *IEEE J. Solid-State Circuits*, vol. 35, no. 12, pp. 1895–1907, Dec. 2000.
25. A. Springer, L. Maurer, and R. Weigel, "RF system concepts for highly integrated RFICs for W-CDMA mobile radio terminals," *IEEE Trans. Microw. Theory Tech.*, vol. 50, no. 1, pp. 254–267, Jan. 2002.
26. F. Gatta, D. Manstretta, P. Rossi, and F. Svelto, "A fully integrated 0.18- μ m CMOS direct conversion receiver front-end with on chip LO for UMTS," *IEEE J. Solid-State Circuits*, vol. 39, no. 1, pp. 15–23, Jan. 2004.
27. D. Brunel, C. Caron, C. Cordier, and E. Soudée, "A highly integrated 0.25 μ m BiCMOS chipset for 3G UMTS/WCDMA handset RF subsystem," in *IEEE Radio Frequency Integrated Circuits Symp.*, Jun. 2002, pp. 191–194.
28. J. C. Haartsen and S. Mattisson, "A new low-power radio interface providing short-range connectivity," *Proc. IEEE*, vol. 88, no. 10, pp. 1651–1661, Oct. 2000.
29. D. Manstretta, M. Brandolini and F. Svelto, "Second-order inter-modulation mechanisms in CMOS downconverters," *IEEE J. Solid-State Circuits*, vol. 38, no. 3, pp. 394–406, Mar. 2003.
30. M. B. Vahidfar and O. Shoaiei, "A new IIP2 enhancement technique for CMOS down-converter mixers," *IEEE Trans. On Circuits and Syst. II*, vol. 54, no. 12, pp. 1062–1066, Dec. 2007.
31. M. B. Vahidfar and O. Shoaiei, "A high IIP2 mixer enhanced by a new calibration technique for Zero-IF receivers," *IEEE Trans. On Circuits and Syst. II*, vol. 55, no. 3, pp. 219–223, March 2008.
32. B. Razavi, *Design of analog integrated circuits*, McGraw-Hill, 2001.
33. R. Mukhopadhyay et al., "Reconfigurable RFICs in Si-based technologies for a compact intelligent RF frontend," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 1, pp. 81–93, Jan. 2005.
34. D. Ham and A.T Nakamura, T. Masuda, A. Kodama and K. Washio, "A variable inductor using mutual current control and application to a SiGe 4.5GHz VCO for wide tuning range," *AMPC, IEEE Proc. Of.*, June 2005.
35. H. Sjoland, A. Karimi-Sanjaani, A. A. Abidi, "A merged CMOS LNA and mixer for a WCDMA receiver", *IEEE Journal of Solid State Circuits*, vol.38, no.6, pp.1045–1050, June 2003.
36. H. Darabi and A. A. Abidi, "Noise in RF-CMOS mixers: a simple physical model," *IEEE J. Solid-State Circuits*, vol. 35, no. 1, pp. 15–25, Jan. 2000.
37. M. B. Vahidfar, O. Shoaiei and F. Svelto, "A High Dynamic Range Multi-Standard CMOS Mixer for GSM, UMTS and IEEE802.11b-g-a Applications," *RFIC Symposium, IEEE Proc. of*, pp. 193–196, June 2008.

Chapter 9

Multi-standard Continuous-Time Sigma–Delta Converters for 4G Radios

Yi Ke, Jan Craninkx, and Georges Gielen

9.1 Introduction

In recent years, there has been an explosive demand for wireless and portable applications, resulting in more flexible communication systems which can handle various standards in different environments. The 4G systems that are expected to appear in the market around 2010 aim to seamlessly integrate the existing and future wireless technologies on a single handset, with high data rate and more functions. A Software Defined Radio (SDR) adopting a fully reconfigurable front-end is believed to be the right answer to realize such systems. Therefore, a fully reconfigurable analog-to-digital converter (ADC) is needed for the different modes in 4G radios. This ADC switches its resolution and bandwidth depending on the communication mode, but it can also relax its specifications with fine granularity within a given mode to save power. $\Delta\Sigma$ modulators (DSMs) are normally favored in multi-mode designs due to their low power consumption and inherent trade-off between speed and accuracy. In the last 5 years, several semi-flexible DSMs [1,2] were presented, which can handle up to three fixed modes. A more flexible continuous-time (CT) DSM was presented in [3] which can be switched to 121 modes by using combinations of 11 resistor and 11 capacitor values. All these previous multi-mode design approaches employ a fixed topology with reconfigurable passive-component arrays such as capacitor and resistor arrays (shown in Fig. 9.1) in order to adapt to different wireless standards. Our work presents a more flexible design with a fully reconfigurable structure to further save power in every mode.

Y. Ke (✉) and G. Gielen
Katholieke Universiteit of Leuven, Department of Elektrotechniek, ESAT-MICAS,
B-3001 Leuven, Belgium
e-mail: {yke; georges.gielen}@esat.kuleuven.be

J. Craninkx
IMEC, Kapeldreef 75, B-3001 Leuven, Belgium
e-mail: cranincj@imec.be

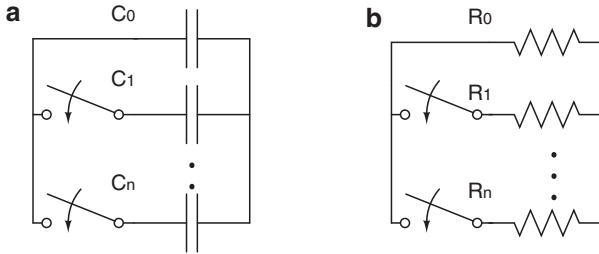


Fig. 9.1 Reconfigurable passive-component array: (a) capacitor array (b) resistor array

9.2 Specifications for ADCs in 4G Radios

Compared to the previous generations of wireless communication systems, the next generation system of so called 4G radios has two major features. One feature which is an evolution from previous generations is the faster speed and larger bandwidth. The bandwidth of the input signals for the new standards such as Digital Video Broadcasting – Handheld (DVB-H) and wireless local area network (WLAN) exceed 5 MHz, which is 25 times that of the GSM standard. The second feature of the 4G radios which is the revolution part compared to the previous generations is called universal communication. The target is to make one mobile set adapt to different standards without changing the hardware. This becomes the driving force of using reconfigurable circuit.

9.2.1 Major Standards for 4G Radios

The input signals of the ADC in 4G radios cover a wide range of standards ranging from the newest WLAN 802.11 n to the widely used GSM mode. Table 9.1 summarizes the main foreseen specifications in 4G radios. In order to meet all these specifications, a flexible baseband ADC needs to cover a wide range of signal bandwidths from a few hundreds of kilohertz up to 40 MHz, and the dynamic range (DR) performance also needs to be scaled from 85 dB for GSM down to 55 dB for WLAN. It is notable from Table 9.1 that for the WLAN and DVB-H modes the DR requirements can also be scaled depending on the modulation methods. Thus, it is reasonable to make the performance of the ADC scalable for a certain mode to further save power when different modulation schemes are available.

9.2.2 Design Challenges for the Reconfigurable ADC

A power-optimal fully reconfigurable solution can only be found when flexibility is taken into account at the system level. To meet the requirements described above,

Table 9.1 Specifications for major standards in 4G radios

Standards	Signal BW (MHz)	Modulation	Dynamic range (dB)
GSM	0.2	GMSK	85
UMTS TDD	2	BPSK	74
		64 QAM	74
UMTS FDD	4	BPSK	72
		64 QAM	72
DVB-H	7.6	BPSK	54
		64 QAM	63.5
WLAN 802.11	20	BPSK	42
a		64 QAM	55
WLAN	40	BPSK	42
802.11n		64 QAM	55

there are several design challenges to be tackled. The first one is the programmability. Due to the wide range of both the signal bandwidth and DR requirements, the sampling frequency can be 100 times different from WLAN mode to GSM mode. Large passive component arrays are needed to cover this wide range, and these arrays, which can result in serious parasitic effects, should be minimized. Secondly, the design complexity normally goes up quickly when programmability is introduced. Therefore, a simple and robust basic structure is important for the implementation, where the basic components should be reused. Thirdly, power consumption should be scalable. We need the power to be scaled down from wide-band modes to narrow-band modes to make the power consumption in each mode comparable to dedicated single-mode design solutions. In addition, when a low-interferer environment exists or different modulation modes are available, a Quality of Service (QoS) [4] manager can configure the best power/performance trade-off according to the user/environment requirement. Finally, the total chip area should be well controlled to minimize the cost, so the area overhead of the reconfigurability should be minimal.

9.3 Topology Exploration Based on Power Considerations

Most of the previous multi-mode designs [1–3] chose the single-loop DSM topology because of its robustness to non-idealities of circuit components. Furthermore the relatively simple structure makes it more suitable for a complex reconfigurable circuit design. Consequently, the single-loop DSM becomes the basic choice for our design.

The ideal performance of the single-loop DSM, when only taking the quantization noise into account is given by [5]:

$$DR = \frac{3}{2} \cdot \frac{2n + 1}{\pi^{2n}} \cdot (2^N - 1)^2 \cdot OSR^{2n+1} \quad (9.1)$$

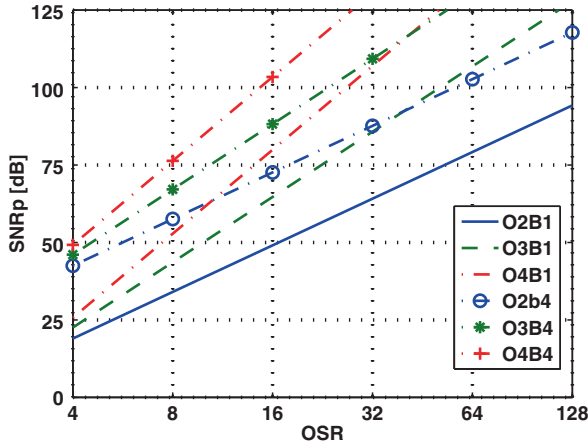


Fig. 9.2 Ideal DSMs SNR performance. OxBy means DSMs with order- x and y -bit quantizer

with N the number of quantizer bits, n the modulator order, and OSR the oversampling ratio. Figure 9.2 shows the ideal SNR performance with different parameters. It can be seen from the plot and (Eq. 9.1) that different combinations of number of quantizer bits, filter order and OSR can be used to achieve the same dynamic range. Each combination of these parameters forms a design solution. However, which combinations give the best power efficiency for various specifications has not been thoroughly discussed in the literature in the past. Thus, a power minimization methodology has been developed, as described in the following subsections.

9.3.1 Power Considerations of System Level

The actual performance of a DSM is normally quite different from the one given by (Eq. 9.1) due to stability issues. For the single-bit quantizer case, the infinity norm $\|H\|_{\infty}$ of the noise transfer function H is traditionally chosen to be 1.5 as rule of thumb to ensure stability [6]. In the case of a multi-bit quantizer, $\|H\|_{\infty}$ can be increased to improve the Signal-to-Quantization Noise Ratio (SQNR). However, this can not be pushed too far, because the Overload Level (OL) then starts to decrease. The maximum input voltage swing is the product of the OL and V_{ref} (feedback reference voltage), where V_{ref} is limited by the supply voltage and the specific circuit implementation. Therefore, the optimal $\|H\|_{\infty}$ should be chosen to maximize the SQNR and at the same time to minimize the decrease of the OL . As example, Fig. 9.3 shows the SNR performance of a third-order modulator with a 2-bit quantizer ($OSR = 128$). The optimal $\|H\|_{\infty}$ is around two here since the SQNR increases very slowly when $\|H\|_{\infty}$ is larger than two. Further increase of the $\|H\|_{\infty}$ leads to the rapid decrease of the signal power, which in turn increases the power consumption budget used to reduce the thermal noise for the same DR. Similarly,

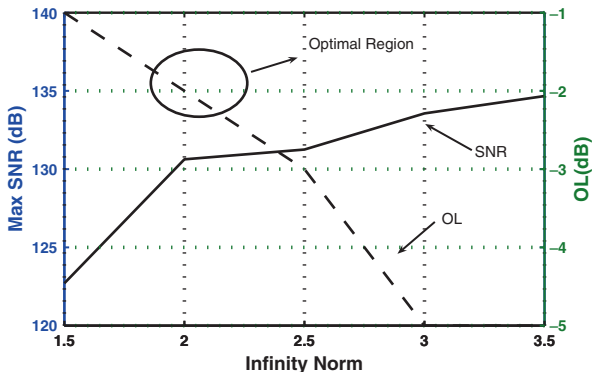


Fig. 9.3 SNR and OL vs. the $\|H\|_\infty$ of the NTF for third-order 2-bit DSM case

Table 9.2 Optimal infinity norm of the NTF for DSM from second to fifth order with number of quantizer bits from 1 to 5

	1 bit	2 bits	3 bits	4 bits	5 bits
Order 2	2	2.5	3	3.5	3.5
Order 3	1.5	2	3	3.5	4
Order 4	1.5	2	2.5	3.5	4
Order 5	1.5	2	2.5	4	5

Table 9.3 OL level for DSM from second to fifth order with number of quantizer bits from 1 to 5 corresponding to $\|x\|_\infty$ values in Table 9.2

	1 bit	2 bits	3 bits	4 bits	5 bits
Order 2	-1 dB	-1 dB	-1 dB	-1 dB	-1 dB
Order 3	-2 dB	-2 dB	-2 dB	-1 dB	-1 dB
Order 4	-4 dB	-3 dB	-3 dB	-2 dB	-1 dB
Order 5	-5 dB	-5 dB	-3 dB	-2 dB	-1 dB

balancing between OL and SNR, the optimal $\|H\|_\infty$ for different filter orders and quantizer bits has been determined as listed in Table 9.2. With these values, the optimal Noise Transfer Function (NTF) can be determined using the *synthesizeNTF* function in [7]. Furthermore, the corresponding OL values can be found and are listed in Table 9.3. It is notable that the OL also depends on the input test frequency, so only worst-case OL values (when a near-DC test tone is applied [6]) are taken into account here. Finally, long-term transient simulations are used to verify the stability. Besides the *synthesizeNTF* function, the CLANS method in [7] is another way to design NTFs. This method determines the poles of the NTF optimally and achieves a better SNR performance compared to that of the *synthesizeNTF* function. However, no power considerations are involved in the CLANS method and much higher power consumption is needed for the optimal performance. By using *synthesizeNTF*

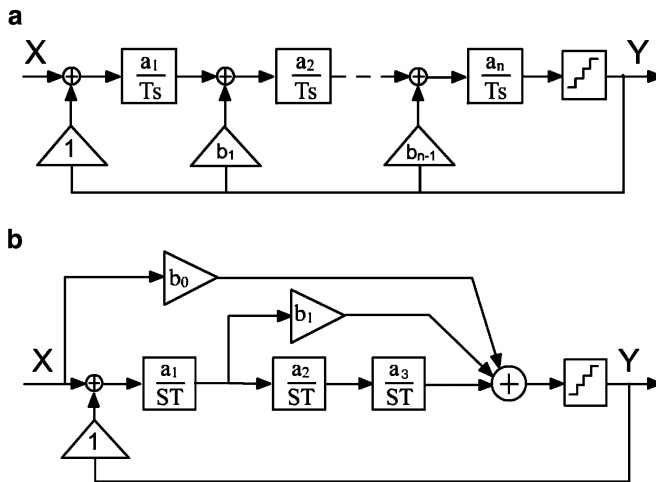


Fig. 9.4 Topology with (a) feedback compensation, and (b) feedforward compensation

function, more flexibility is allowed in power-performance trade-off by tuning the $\|H\|_{\infty}$ value. This explains why we chose to use the *synthesizeNTF* function for NTF design.

Except for the *OL* level, the integrator gain coefficient is another factor which influences the power consumption. Normally, the coefficient values are directly related to the filter structure. For single-loop DSMs, there are mainly two types: the feedback and the feedforward topologies (shown in Fig. 9.4). Both these types can be used to implement the same NTF, whereas the STF of the feedback (FB) topology has a higher-order anti-aliasing filtering property compared to the STF of the feedforward (FF) topology. Therefore, normally the anti-aliasing filter can be omitted when the feedback topology is used. For the FF topology, a reasonable anti-aliasing filtering property can also be guaranteed when the direct path b_0 (from the input of the DSM to the input of the quantizer) is absent.

Another major difference between the FF and FB topologies is the output voltage swings of the integrators, especially the first integrator. Due to the feedforward paths, the output swing of the first integrator of the FF topology is much smaller than that of the FB topology when the same integrator coefficients are used. In other words, a much larger coefficient can be used for the first integrator to achieve the same output swing in the FF topology. On the other hand, a fast amplifier is usually needed for the summing stage before the quantizer of the FF topology. Luckily, this power-hungry amplifier can be omitted by using capacitor feedforward summation at the last integrator [8], or current summation at the resistor ladder in the quantizer [9]. To allow a fair comparison both CT FF and FB topologies. The swings of the first integrators are scaled to be less than 0.6 (normalized to the reference voltage) for linearity reasons when a full-swing input signal is applied. The final coefficients of the first integrator for FB and FF topologies with different modulator order and

Table 9.4 Coefficient of the first integrator for second to fifth order FB DSMs with quantizer bits from 1 to 5

	1 bit	2 bits	3 bits	4 bits	5 bits
Order 2	0.3	0.45	0.55	0.65	0.65
Order 3	0.15	0.3	0.35	0.43	0.46
Order 4	0.1	0.15	0.2	0.25	0.28
Order 5	0.07	0.1	0.15	0.2	0.2

Table 9.5 Coefficients of the first integrator for second to fifth order FF DSMs with quantizer bits from 1 to 5

	1 bit	2 bits	3 bits	4 bits	5 bits
Order 2	0.3	0.5	0.9	1.6	2.5
Order 3	0.3	0.5	0.9	1.6	2.5
Order 4	0.3	0.5	0.9	1.6	2.5
Order 5	0.3	0.5	0.9	1.6	2.5

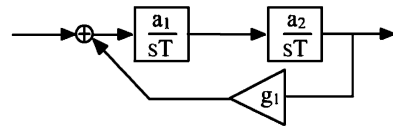


Fig. 9.5 Resonator formed by local feedback

number of quantizer bits are listed in Tables 9.4 and 9.5, respectively. It should be mentioned that the zeros of the NTF are spread over the signal bandwidth to achieve optimal performance. This is done by introducing a small negative local feedback, as shown in Fig. 9.5, from the output of one integrator to the input of the previous integrator. The resonator frequency is determined by

$$f_{resonator1} = 2 \cdot \sqrt{a_1 a_2 g_1} \cdot BW \cdot OSR \tag{9.2}$$

As indicated in (Eq. 9.2), the positions of the zeros depend on the value of the *OSR*. To maintain the output swing at the same level, normally the scaling coefficients a_1 and a_2 should be kept the same for different *OSR*s. Thus, the change of *OSR* mainly influences the g_1 value while the scaling coefficient of the integrators is unchanged. As indicated in (Eq. 9.1), the quantization noise level can be determined once the order of the modulator, the number of bits in the quantizer and the *OSR* are fixed. To achieve the required SNR performance for each standard, a proper *OSR* should be chosen for a certain topology. Here, the *OSR* is chosen to make the quantization noise level around 10 dB lower than the required noise floor as thermal noise dominates the noise budget.

It can be seen from the tables that the coefficient of the first integrator for the FB case is scaled down as the modulator order increases to guarantee stability and a reasonable output swing. For the FF case, the coefficient of the first integrator

is mainly determined by the number of quantizer bits, since the quantization noise dominates the output in this case. Due to the coefficient difference, the FF topology shows better power and area efficiency, as will be analyzed in detail in the following sections.

9.3.2 Power Considerations of Circuit Level

In this subsection, the performance and power trade-offs at circuit level are discussed in detail for low supply voltages. There are mainly three parts in a DSM: the loop filter which is composed of integrators, the quantizer and the DAC in the feedback path.

9.3.2.1 Integrator Power

The Gm-C integrator is more power-efficient for wide-band applications with medium resolution while the RC integrator shows better power efficiency for higher linearity. When the Gm-C integrator is used as the front-end of the DSM, the inputs of the DSM are connected directly to the input transistors of the differential pair as shown in Fig. 9.6. To make both M1/M2 and M3 work in saturation, the input range of the Gm-C integrator is limited by:

$$V_{in_range} = V_{DD} - V_{GS} - V_{SAT} \quad (9.3)$$

where V_{GS} is the gate-source voltage of the input transistor and V_{SAT} is the saturation voltage of M3. For supply voltages smaller than 1 V, the headroom left for the input range is seriously limited, which directly degrades the DR. In an RC integrator, the input of the amplifier is fixed to virtual ground, and the input signal only faces a resistor. Hence, a larger input range can be used. Linearity is another factor in choosing the integrator topology. The nonlinearity of the integrator can be modeled as a nonlinear function before a linear integrator. For fully differential circuit, this function can be given by $f(x) = x - d_3x^3$. d_3 is the third-order coefficient which can

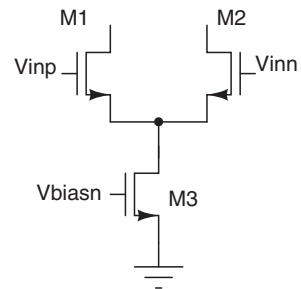


Fig. 9.6 Input differential pair

be derived from the integrator and amplifier topology. For Gm-C integrator using a single-stage transconductor, d_3 equals $g_3/4g_m$ [10], where g_3 is the third-order coefficient of the transconductor and g_m is the transconductance. For RC integrator using the same transconductor, d_3 is $2g_3/(g_m(2 + g_m R)^3)$ [10], which is much smaller. To achieve the same level of linearity, a larger transconductance is needed for the Gm-C integrator, leading to a higher power consumption. Since we want to reuse the integrator for accuracies ranging from 9 bits to 14 bits, according to Table 9.1, the RC integrator becomes a better choice for our reconfigurable low-voltage and low-power application.

The Dynamic Range (DR) calculation takes into account both quantization noise and thermal noise:

$$DR \approx \frac{1}{2} \frac{V_{in_range}^2}{N_q + N_{th}} = \frac{1}{2} \frac{(OL \cdot V_{ref})^2}{N_q + N_{th}} \tag{9.4}$$

$$N_{th} = (2 \times 4kTR_{in} + N_{th_DAC}) \cdot BW \tag{9.5}$$

where N_q is the quantization noise; N_{th} is the input equivalent thermal noise of the integrator; V_{ref} is the equivalent feedback reference voltage; R_{in} is the resistance of the RC integrator (shown in Fig. 9.7); k is Boltzmann’s constant; T is the temperature; N_{th_DAC} is the thermal noise from the feedback DAC (current-steering DAC is used for high speed). The thermal noise of the amplifier itself is neglected here, since it usually contributes little to the overall thermal noise compared to the noise from the resistor and DAC. The DAC consists of N current-source transistor and two shared PMOS current sources which are used to cancel out the common-mode current. Assuming the overdrive voltage of the PMOS current transistors and NMOS current transistors are the same, the noise contribution from the PMOS and NMOS current-source transistors are the same. Thus, the input-referred noise power spectral density (PSD) of the DAC can be expressed as:

$$N_{th_DAC} = 2 \cdot \gamma \cdot 4kTg_{m_DAC} \cdot R_{in}^2 BW = 16kT \cdot \gamma \cdot \frac{I_{DAC}}{V_{gst_DAC}} R_{in}^2 BW \tag{9.6}$$

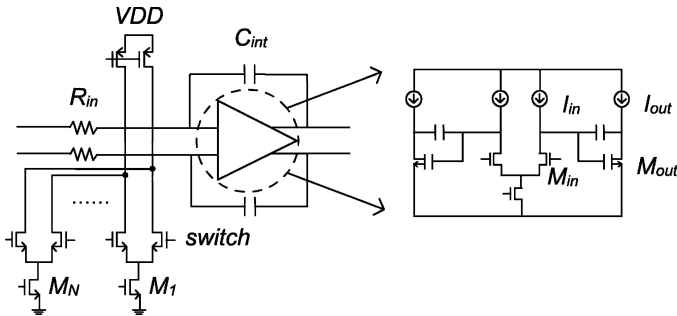


Fig. 9.7 OTA-RC integrator with current-steering DAC and Miller OTA

where BW is the signal bandwidth; γ is a coefficient which is determined by CMOS technology; gm_{DAC} is the sum of the transconductances of the transistors $M_I - M_N$ of the DAC (shown in Fig. 9.7). By using the relationship between V_{ref} and I_{DAC} (total current of all the current sources in the current-steering DAC):

$$V_{ref} = I_{DAC} \cdot R_{in} \quad (9.7)$$

Eq. (9.5) can be written as:

$$N_{th} = (8kT + 16kT \cdot \gamma \cdot \frac{V_{ref}}{V_{gst_DAC}}) R_{in} BW \quad (9.8)$$

where V_{gst_DAC} is the overdrive voltage of the transistors $M_I - M_N$ of the DAC. In a low-voltage design, thermal noise dominates the noise budget, so the R_{in} can be found from (Eqs. 9.4 and 9.8) as:

$$R_{in} \approx \frac{1}{16kT \cdot DR \cdot BW (V_{gst_DAC} + 2 \cdot \gamma \cdot V_{ref})} (OL \cdot V_{ref})^2 V_{gst_DAC} \quad (9.9)$$

The corresponding integration capacitance C_{int} is given by

$$C_{int} = 8kT \frac{DR (V_{gst_DAC} + 2 \cdot \gamma \cdot V_{ref})}{V_{gst_DAC} (OL \cdot V_{ref})^2 \cdot a \cdot OSR} \quad (9.10)$$

where a is the scaling coefficient defined in Tables 9.4 and 9.5. As indicated in (Eq. 9.10), C_{int} increases as a scales down, which makes the FF topology more area efficient than the FB type for the same DR. On the other hand, C_{int} is one of the major contributors to the load capacitance of the amplifier:

$$C_{load} = \alpha C_{int} + C_{mos} + C_{wire} \quad (9.11)$$

where C_{mos} and C_{wire} are the parasitic capacitance from the MOS transistors and connection wires respectively. α is a coefficient which is related to the type of capacitors or in other words to the way of implementation of the capacitors.

For low supply voltage, the two-stage amplifier is preferred for larger output swing. The standard Miller two-stage OTA shown in Fig. 9.7 is used here to estimate the power consumption. The power consumption of the integrator can be found to be:

$$P_{int} = 2(I_{in} + I_{out})V_{DD} \quad (9.12)$$

where I_{in} and I_{out} are the bias quiescent currents of the input and output branch, respectively. Normally the I_{in} value is a small fraction of I_{out} . I_{out} in the output branch should be selected to meet both the stability and slew rate requirements:

$$\frac{2I_{out}}{V_{gst_out} 2\pi C_{load}} = \frac{gm_{out}}{2\pi C_{load}} \geq 2 \cdot GBW \text{ and } GBW \geq c \cdot Fs \quad (9.13)$$

$$I_{out} \geq 2I_{DAC} \quad (9.14)$$

where the gm_{out} is the transconductance of M_{out} in Fig. 9.7. c is normally less than two in the CT DSM case [11]. Fs is the sampling frequency. The first inequality in (Eq. 9.13) guarantees the stability of the two-stage operational transconductance amplifier (OTA). The GBW represents the open-loop gain bandwidth of the OTA. The second inequality is used to guarantee a stable DSM [11]. Besides, the OTA in the integrator needs to sink both the input current and feedback DAC current, and it is quite likely that both the input current and the DAC current will go to the wrong side during the start-up of circuit. Thus, I_{out} needs to be at least twice I_{DAC} as shown in (Eq. 9.14) to provide enough slew rate. By putting (Eqs. 9.10, 9.11) into (Eq. 9.13) and (Eqs. 9.7, 9.9) into (Eq. 9.14), we get

$$I_{out} \geq 2\pi cFs(8kT \frac{\alpha DR(V_{gst_DAC} + 2 \cdot \gamma \cdot V_{ref})}{a \cdot (OL \cdot V_{ref})^2 \cdot V_{gst_DAC} OSR} + C_{mos} + C_{wire})V_{gst_out} \quad (9.15)$$

$$I_{out} \geq 32kT \frac{(V_{gst_DAC} + 2 \cdot \gamma \cdot V_{ref})DR}{OL^2 \cdot V_{ref} \cdot V_{gst_DAC}} BW \quad (9.16)$$

As discussed in the previous subsection, the optimal OL and a can be determined once the corresponding system-level filter topology is fixed. To minimize the power consumption, parameters V_{ref} and V_{gst_DAC} can be increased while V_{gst_out} should be minimized for a certain topology at circuit level. The final I_{out} depends on the larger value of (Eq. 9.15) and (Eq. 9.16). By using Tables 9.2 to 9.5 and Eqs. (9.15–9.16), the power consumption of the first integrator for both the FB and FF DSMs with different filter order and number of quantizer bits can be estimated for a given DR. To make the estimation more accurate, at least 6-dB margin for the DR should be taken into account.

Compared to the first integrator, the other integrators can be scaled to save power since both their noise and nonlinearity are suppressed by the gain of the first stage. In order not to degrade the total performance, the scaling factor should be smaller than the gain of the first integrator:

$$f_{scale} \leq \left(\frac{aFs}{2\pi BW} \right)^2 = \left(\frac{a \cdot OSR}{\pi} \right)^2 \quad (9.17)$$

As shown in (Eq. 9.17), due to the larger a , a larger scaling factor for the other integrators can be used in the FF topology to save more power. Unlike in discrete-time (DT) DSMs, large capacitor arrays are used not only to reconfigure between different modes but also to compensate the process variations in CT DSMs. In order not to be influenced by parasitic capacitances, the minimum capacitor in the array

is lower bounded which is one of the fundamental limitations in the reconfigurable design. It directly limits the maximum f_{scale} .

9.3.2.2 DAC

The first feedback DAC's power is given by

$$P_{DAC} = V_{DD} \cdot I_{DAC} \quad (9.18)$$

where I_{DAC} is the total current of all the current sources in the first feedback DAC. By putting (Eqs. 9.7 and 9.9) into (Eq. 9.18), we have

$$P_{DAC} = 16kT \frac{DR \cdot BW \cdot (V_{gst_DAC} + 2 \cdot \gamma \cdot V_{ref})}{V_{ref} \cdot OL^2 \cdot V_{gst_DAC}} V_{DD} \quad (9.19)$$

which shows that a larger OL also helps to reduce the power consumption of the DAC. In the FB topology, the non-idealities such as the thermal noise and non-linearity are suppressed by the first integrator, thus the power of the other DACs can be scaled in the same way as the integrators.

9.3.2.3 Quantizer

The power consumption of the quantizer is estimated by using the power estimator proposed in [12]:

$$P_{quantizer} = (2^b - 1) \frac{0.5V_{DD} - 0.3V_{swing}}{32.1 \times 10^3} V_{DD} L_{min} F_s + P_X \quad (9.20)$$

where b is the quantizer resolution, V_{swing} is the input signal swing, L_{min} is the gate length for the used technology, P_X is the encoder power (This part is not taken into account here). So, to save power, V_{swing} should be maximized for a given supply voltage and V_{ref} .

9.3.3 Power Estimation

Based on the calculations of the previous section, once the standard specifications are fixed, the power consumption of the integrator and DAC for a certain topology is closely related to the system-level design parameters like scaling coefficients of the integrators and OSR , and these parameters can be chosen when the optimal NTF is determined. At circuit level, the power consumption can be optimized by maximizing the V_{ref} and V_{gst_DAC} while minimizing the V_{gst_out} .

The total power of a DSM solution can be estimated as the sum of the three parts: integrator, DAC and quantizer. By optimizing the power consumption both at system and circuit level, the minimal power consumption can be found for a certain design solution.

For a 90 nm CMOS technology with 1 V V_{DD} , the estimated power consumptions for both the FB and FF topologies with second to fifth-order loop filter and number of quantizer bits from 1 to 5 are shown in Fig. 9.8. The results show that, in GSM mode, second- and third-order DSMs consume less power both in FB and FF topologies. This is because the total power budget is dominated by the integrator part, especially the first integrator. The power consumption of the quantizer is much smaller compared to the analog part as the sampling frequency is still low even when the OSR value is high in the GSM mode. In this case, the total power is mainly determined by (Eqs. 9.15 and 9.16), so DSMs with higher OL value have better power efficiency. As shown in Table 9.3, the OL values for low-order modulators are much higher than for high-order modulators in both FF and FB cases when a 1-bit quantizer is used. As the number of quantizer bits is increased, the OL value of the high-order modulators can be increased resulting in lower power consumption. From Fig. 9.8(a), we can see that when increasing the number of quantizer bits, the power consumption of the higher-order DSMs goes down due to the increased OL level. For the fourth-order and fifth-order modulator, the power-optimal points are achieved when a 5-bit quantizer is used. For second- and third-order modulators, the power consumption level is almost flat when the number of quantization bits is increasing. As indicated in (Eq. 9.14), the scaling coefficient of the integrators is another factor which influences the power consumption. This explains why the power of the fifth-order single-bit FB design is much higher than the other cases since its scaling coefficient of the first integrator is the smallest. It should be mentioned that the power estimations here do not take into account the power used in the linearity enhancement technique needed when multi-bit is used. However, this influences all the points simultaneously and doesn't change the relative position of the curves. When switching gradually from the GSM to the WLAN mode, the power of the second-order DSMs increases faster compared to the other orders. The reason is the rapid increase of the quantizer power for increasing signal bandwidth, while the integrator power increases at a slower rate due to the lowered DR specification. As shown in Fig. 9.8, the relative position of the power-consumption curve for the second-order modulator is moving upwards while those of the higher order modulators are moving downwards. In the WLAN mode, the power consumption of the second-order modulator becomes higher than the other designs. The above estimation results have not taken into account the large reconfigurable capacitor array. For linearity reasons, the switches of the capacitors should be connected to the virtual ground side and the output node of the OTA is directly connected to the capacitor array. Therefore, the load capacitance should include the total parasitic capacitance of the capacitor arrays. From (Eq. 9.10), the largest capacitor in a reconfigurable DSM is determined by the specification with the lowest sampling frequency and the highest DR. Thus, the GSM mode determines the area budget and the major parasitic capacitance. When taking into account the large capacitor for the GSM mode, the

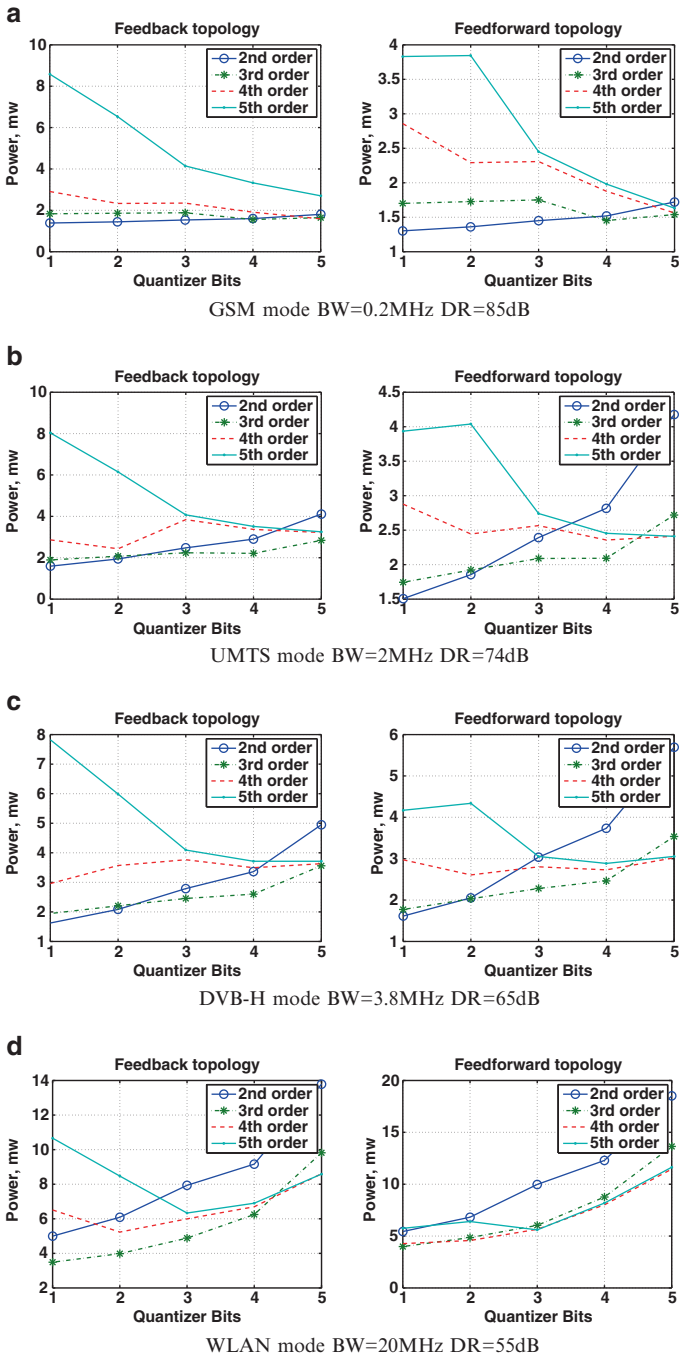


Fig. 9.8 Power estimation results for both FB and FF topologies for the different standards

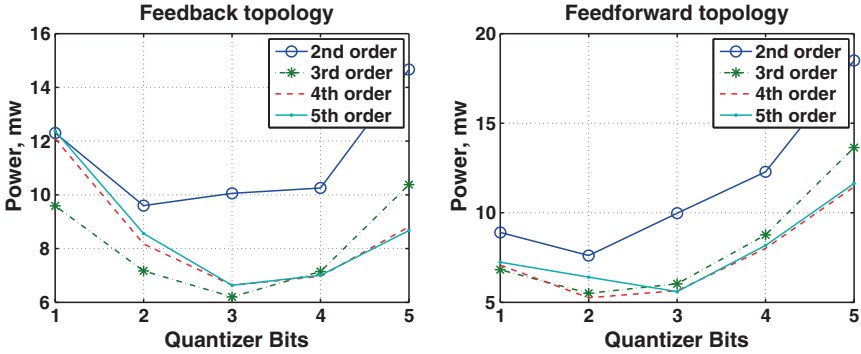


Fig. 9.9 Power estimation for WLAN with reconfigurability consideration

plots for the WLAN case are different as they are influenced dramatically by the new consideration due to the increased load capacitance. The major change in Fig. 9.9 is that the power-optimal points shift from the single-bit to 2-bit solutions, due to the reduced OTA GBW. However, further increasing the number of quantizer bits cannot save more power, as the quantizer power goes up quickly when the number of quantizer bits is larger than three. Therefore, it is advisable to choose a third- to fifth-order DSM with 2–3 bit quantizer for WLAN.

A similar conclusion can be drawn for the results of the DVB-H and UMTS mode. For the GSM mode, the extra consideration doesn't influence the result, since the large capacitor array has already been taken into account in the single-mode case. The second- to third-order DSMs with 2- to 3-bit quantizer consume less power. Especially when a single bit is adopted, capacitor area can be saved due to the increased *OSR* as indicated in (Eq. 9.9). A smaller capacitor array also helps to reduce the parasitic capacitance for the WLAN mode. Furthermore, a single-bit quantizer and DAC provide perfect linearity to meet the high DR specification.

9.4 Proposed Reconfigurable DSM

Inspired by the above observations, we found that a power- and area-optimal reconfigurable DSM for 4G radios can be realized by reconfiguring its filter order and the number of quantizer bits together with modifying the *OSR*. It seems that both the FB and FF topologies can be used to achieve this goal. However, for comparable power consumption, a FB topology needs a larger signal swing which decreases the linearity. Furthermore, as indicated in (Eq. 9.9), the corresponding capacitor would be much larger compared to the FF topology and this directly increases the power consumption for the WLAN case, as shown in Fig. 9.9. All these considerations result in the power-optimal fully reconfigurable DSM shown in Fig. 9.10. The extra DAC2 is added to compensate the quantizer delay [9]. To ease the design of the high-GBW OTA in the WLAN mode, the flexible DSM uses all four integrators and two lo-

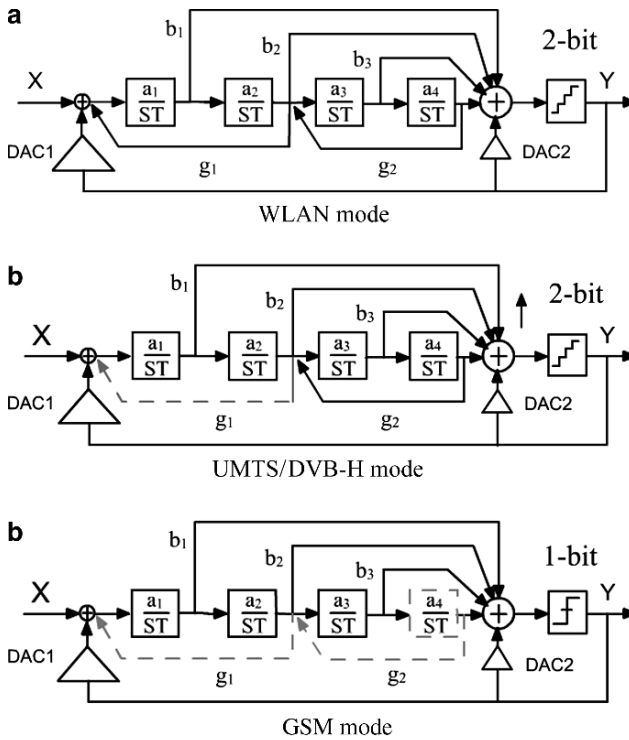


Fig. 9.10 Architecture configurations for different modes

cal feedback paths to create two resonators for further suppressing the quantization noise [6]. The 2-bit quantizer is used in this mode for better power efficiency according to Fig. 9.9. Meanwhile, the linearity can be guaranteed by increasing the transistor area. Since a 2-bit DAC only needs three current-source transistors, there is no need to use dynamic element matching to save area. Besides, if 1 bit would be used in the WLAN mode, the sampling frequency would have to be higher than 1 GHz to achieve the same performance, which is unpractical for the OTA design in 90 nm CMOS. For the DVB-H and UMTS mode, the fourth-order loop filter and 2-bit quantizer are still used, but only one resonator is used to save the large resistor in the local feedback path, while still providing enough suppression of the quantization noise (see Fig. 9.10.). In GSM mode, both g_1 and g_2 are set to zero for the same reasons. Besides, the fourth integrator is powered off and the coefficients are adapted correspondingly. Thanks to the FF topology, no extra bypass circuits are needed. In the quantizer, only one comparator is used as a single-bit quantizer to provide high linearity and to reduce the largest integration capacitor. The binary-scaled resistor arrays and capacitor arrays are used to reconfigure the gain of each integrator according to the specifications of the different standards. The g_1 and g_2 coefficients are tuned by the binary-scaled resistor array in the local feedback paths and can be set to zero by switching off the whole resistor array. The feedforward coefficients

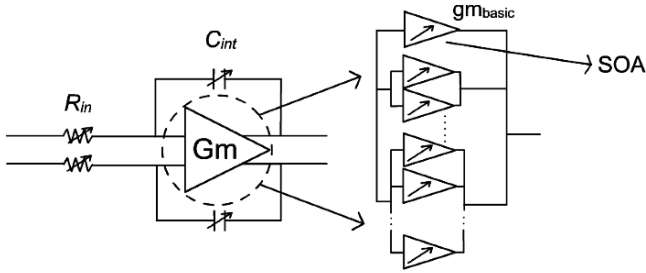


Fig. 9.11 Flexible OTA array

b_i stays the same in the WLAN/UMTS/DVB-H modes, but need to be reconfigured for the GSM mode. DAC2 is also made tunable to compensate the process variation and a wide range of quantizer delays.

The optimal power efficiency can only be achieved by using a flexible OTA which can adapt its GBW to the different specifications. In [13], the input transistors are biased in the weak inversion region to maintain a linear relationship between the GBW and the biasing current. This method decreases the speed and only a limited tuning range is available. To allow fully digital control, the concept of Switchable Op-Amp (SOA) [14] is used here to implement the OTA in the RC integrator. Each SOA can be powered on or off by switching the voltage at the gate of all the transistors in the two-stage Miller OTA simultaneously. By reusing a basic robust unit, the design complexity is reduced and fully digital control is available. By connecting the SOAs in a binary scale as shown in Fig. 9.11, a fully reconfigurable OTA is obtained.

As mentioned in Section 2.2, the DR requirement for different modulation modes and different environments are different. So besides switching between the different modes, it is also better to scale the power consumption of the reconfigurable DSM depending on the environmental conditions (presence of interferers, etc). There are two ways to lower down the performance in the proposed modulator. The first way is to reduce the OSR . This is done by either increasing the corresponding capacitors or increasing the input resistors of the integrators. The first way is more suitable for the WLAN case, as the small nominal capacitance can't be further reduced. Due to the decreased sampling frequency, the power of the digital part can be saved in the first approach. When the resistance of the input resistors is increased, the integration capacitance can be kept the same, which means that the load capacitance seen at the virtual ground node is fixed. Thus, the number of unit OTAs which is turned on can be scaled down in the same manner as the sampling frequency. In this way, the power consumption of the whole modulator is almost linearly scaled with the sampling frequency. However, as indicated in (Eq. 9.1), only rough tuning is allowed when a high-order DSM is used. The second approach is to increase the resistors of the integrators and simultaneously decrease the integration capacitances which results in a fixed R-C time constant for all the integrators. The sampling frequency is kept the same when using this approach. Unlike the first approach, the second method only increases the thermal noise, and less than a 3 dB tuning step

Table 9.6 Simulated performance of the reconfigurable DSM for different standards

	BW (MHz)	OSR	SNR (dB)	Estimated Power ^a
WLAN	20	16	59	8 mW
DVB-H	3.8	24	68.5	3.5 mW
UMTS	2	32	77	3 mW
GSM	0.2	128	84.5	2 mW

^a CMFB power is included.

Table 9.7 Reported state of art of the DSMs' performance

	BW (MHz)	OSR	SNR (dB)	Reported power	CMOS technology
WLAN	20	16	76	20 mW [15]	0.13 μ m
DVB	4	25	73	19 mW [16]	0.18 μ m
UMTS	1.92	12	65	3.5 mW [13]	0.13 μ m
GSM	0.2	65	82	1.44 mW [3]	90 nm

is available in a binary-scaled resistor and capacitor array. In the second approach, the transconductance of the OTA can be reduced proportional to the decrease of the integration capacitor value.

To verify the presented system solution, the finite gain and GBW of the OTA, parasitic capacitance, loop delay, clock jitter, etc., have been taken into account. The simulated performances using Matlab are summarized in Table 9.6. Compared to the multi-mode designs in [2, 3, 13], the results listed in Table 9.6 show better power efficiency especially in the WLAN mode. This verifies the effectiveness of the design approach described above. The state of the art of reported DSMs' performance (including single-mode DSMs reported in literature) for different standards is listed in Table 9.7 for comparison. Our solution compares very favorably for the WLAN mode.

9.5 Conclusions

A low-power digitally controllable, fully reconfigurable continuous-time Delta-Sigma modulator is presented at the system level for low-voltage 4G Software-Defined-Radio applications. Both the FF and FB topologies have been explored towards low power with different combinations of design parameters in the design space leveraging reconfigurability for different specifications. The power minimization methodology has been described. The simulation results have confirmed that power is saved in every mode compared to previously published multi-mode designs by reconfiguring the filter order, the number of quantizer bits at topology level and the OTA transconductance, component values at the circuit level. This DSM is fully compatible with an SDR front-end, where a QoS Manager can configure the best power/performance trade-off according to the user/environment requirements.

References

1. R. H. M. van Veldhoven, "A Triple-Mode Continuous-Time $\Sigma\Delta$ Modulator with Switched-Capacitor Feedback DAC for a GSM-EDGE/CDMA2K/UMTS Receiver," *IEEE Journal of Solid State Circuits*, vol. 38, pp. 2069–2076, Dec., 2003.
2. J. Arias, P. Kiss, V. Prodanov, et al., "A 32-mW 320-MHz Continuous-Time Complex Delta-Sigma ADC for Multi-Mode Wireless-LAN Receivers," *IEEE Journal of Solid State Circuits*, vol. 41, pp. 339–351, Feb., 2006.
3. S. Ouzounov, R. H. M. van Veldhoven, et al., "A 1.2v 121-Mode CT $\Delta\Sigma$ Modulator for Wireless Receivers in 90nm CMOS". *ISSCC Digest of Technical Papers*, pp. 238–239, Feb., 2007.
4. L. Guojun, "Quality of service management in distributed multimedia systems", *IEEE International Conference on Systems, Man, and Cybernetics*, Volume 2, Issue 14–17, Oct. 1996, pp. 1321–1326.
5. R. L. Carley, R. Schreier and G. C. Temes, " $\Delta\Sigma$ ADCs with Multi-bit internal Converters". In " $\Delta\Sigma$ Data Converters: Theory, Design, and Simulation," Chapter 8. IEEE, New York, 1997.
6. R. Schreier, "An empirical study of high-order single-bit delta-sigma modulators" *IEEE Transaction and circuit system II*, vol. 40, pp. 461–466, Aug., 1993.
7. R. Schreier, "The Delta-Sigma Toolbox Version 7.1", <http://mathworks.com/matlabcentral/fileexchange>.
8. M. Schimper, L. Dorrer, et al., "A 3mW Continuous-Time $\Delta\Sigma$ -Modulator for EDGE/GSM with High Adjacent Channel Tolerance," In *Proceedings of European Solid State Circuits Conference*, 2004, pp. 183–186.
9. S. Paton, A. Di Giandomenico, L. Hernandez, et al., "A 70mw 300MHz CMOS Continuous-Time Sigma-Delta ADC with 15MHz Bandwidth and 11-bit of Resolution," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 7, pp. 1056–1063, July 2004.
10. P. Sankar and S. Pavan, "Analysis of Integrator Nonlinearity in a Class of Continuous-Time Delta-Sigma Modulators," *IEEE Transactions on Circuits and Systems II*, vol. 54, pp. 1125–1129, Dec., 2007.
11. M. Ortmanns, F. Gersfers, and Y. Manoli, "Compensation of Finite Gain-Bandwidth induced Errors in Continuous-Time Sigma-Delta Modulators," *IEEE Transactions on Circuits and Systems II*, vol. 51, pp. 1088–1100, June 2004.
12. H. Zhaohui, Z. Peixin, "An Architectural Power Estimation for Analog-to-Digital Converters," *IEEE Proceedings of ICCD*, pp. 1063–6404/04, 2004.
13. T. Christen, T. Burger, H. Qiting, "A 0.13um CMOS EDGE/UMTS/WLAN Tri-Mode $\Delta\Sigma$ ADC with -92 THD". *ISSCC Digest of Technical Papers*, pp. 240–241, Feb., 2007.
14. V. Giannini, J. Craninckx, et al., "Fully reconfigurable Active-Gm-RC biquadratic cells for Software Defined Radio applications," *Proceedings of ISCAS*, pp. 1047–1050, May, 2006.
15. M. Gerhard, E. Christian, et al., "A 20mw 640MHz CMOS Continuous-time $\Delta\Sigma$ ADC with 20MHz Signal Bandwidth, 80dB Dynamic Range and 12 bit ENOB". *JSSC*, volume 41, Issue 12, Dec., 2006, pp. 2461–2469.
16. Y. Fujimoto, Y. Kanazawa, et al., "An 80/100MHz/s 76.3/70.1dB SNDR $\Delta\Sigma$ ADC for Digital TV Receivers". *ISSCC Digest of Technical Papers*, pp. 201–210, Feb., 2006.

Chapter 10

Power Efficient Reconfigurable Baseband Filters for Multimode Radios

Pieter Crombez, Jan Craninckx, and Michiel Steyaert

10.1 Introduction

Recent developments in wireless applications and systems have lead to a significant increase in wireless standards. Therefore, the communication devices of tomorrow should not only fulfill traditional requirements such as low power, low area and low cost but must also be capable of complying with a multitude of standards [1]. This combination requires well developed design strategies so that a short time to market and the low cost demand remain guaranteed in the highly competitive world of electronics and wireless communications.

Apart from efficient design methods, multimode terminals become unavoidable. Multimode terminals are required to provide various services from different wireless standards and to satisfy multiple levels of desired quality of service (QoS). This last item is one of the biggest advantages of flexible systems; their power consumption can be made scalable with respect to the channel conditions and the level of quality. In this way, the flexibility which is required to enable multimode systems is also exploited within one given standard to adapt the power consumption to the environment. This leads to an increased efficiency when for example no interferers are present or when the distance to the base station is small. Therefore, multimode solutions outperform dedicated solutions which are designed for the most stringent mode and thus are over-designed when operating in other modes. Fully reconfigurable and integrated transceivers are preferable to answer this demand. They avoid using multiple chipsets and can be made programmable which makes them more efficient in terms of cost, area and power consumption.

P. Crombez (✉) and J. Craninckx
IMEC, Leuven, Belgium
e-mail: {pieter.crombez; jan.craninckx}@imec.be

P. Crombez and M. Steyaert
K.U.Leuven - Dep. ESAT-MICAS, Leuven, Belgium
e-mail: michiel.steyaert@esat.kuleuven.be

An efficient design strategy for $g_m - C$ biquadratic filters. In a flexible receiver front-end, analog baseband filtering is a key task as it offers channel selection – the ability to pick out the wanted signal from a frequency band, while rejecting the others – and anti-aliasing. The first part of this chapter illustrates how a good design strategy leads to a fast and well optimized filter design. Depending on the wireless standard, the dynamic range and bandwidth vary a lot and the specifications on noise and linearity of the filter also differ accordingly. Clearly, when no proper design strategy is available, it is likely that many costly and time consuming iterations will be required to optimize the system for different standards. Therefore, this chapter presents an efficient methodology to optimize power and linearity in the design of biquadratic sections for a low-pass baseband filter based on the $g_m - C$ architecture that can be applied for all current wireless standards below 10 GHz. The biquad will be optimized on an architectural level such that the presented results remain valid for every chosen transistor. Often published work [2, 3] and also filter synthesis literature [4, 5] start from choosing the g_m and C values based on the desired frequency response. Our design strategy however, takes linearity, noise and power into account immediately at architectural level such that the biquad is completely determined early in the design flow. The proposed design approach is demonstrated for a 10 MHz Butterworth biquad and will be used as a starting point for a fully reconfigurable $g_m - C$ filter.

A 100 kHz–20 MHz reconfigurable Nauta $g_m - C$ biquadratic low-pass filter. The second part of this chapter presents a fully reconfigurable $g_m - C$ filter with a large degree of performance-power flexibility which has been design based on the proposed design strategy. Nauta's transistor [6] is chosen since its performance is guaranteed up till high frequencies. In this way, reconfiguring the filter for a higher bandwidth forms no obstacle as further explained in Section 10.2.2. Moreover, the typical common mode feedback loop is avoided which turns out to be useful in controlling the complexity. Finally, this choice [7] allows low supply voltage applications as it is based on an inverter making the flexibility approach extendable to future technologies.

This filter design outperforms the tunable bandwidth (100 kHz–20 MHz) with at least a factor three compared to previously reported work where flexibility is often limited to dedicated modes [8] or where noise and power can not be scaled [9]. Crucial for reconfigurable receivers is that the power consumption can be adapted to the desired performance as imposed by the selected standard or QoS. Next to the bandwidth, also the quality factor, noise level and linearity of this design can be tuned to the desired performance, always aiming for the optimal power trade-off. Therefore the proposed biquad is an excellent candidate to be used as a building block in more selective, higher order filters for Software-Defined Radio (SDR) applications.

The large degree of flexibility is reached thanks to a novel switching technique. This technique allows independent transconductance and capacitance tuning inside the chosen transistor and uses only gate transistor capacitance. By improving Nauta's transistor [6], very low noise levels are achieved while the overall performance is still comparable to recent filter design [10–13]. At the end of this chapter, measurements results on a 0.13 μm CMOS prototype for different supply

voltages are presented which validate our design strategy and illustrates both frequency and performance flexibility proportional to the power consumption.

Outline. Section 10.2 introduces the $g_m - C$ biquad architecture and the chosen Nauta’s transconductor. In Section 10.3, the design strategy is described in detail taking into account all filter parameters directly at architectural level with main focus on the linearity. Design rules guaranteeing optimal performance are proposed. The novel switching technique is elaborated in Section 10.4 and it is explained how the flexible filter is built up starting from a small unit cell up to full control of the filters performance. Finally, Section 10.5 presents measurements results and the performance is validated.

10.2 The $g_m - C$ Filter Architecture and Circuit Discussion

Important parameters in filter design are the filter characteristic (bandwidth, pass and stop band attenuation, the accuracy,), noise, linearity and power consumption [14–16]. These parameters depend not only on the design strategy but also on the chosen topology. The two main topologies with respect to filter design are active- RC and $g_m - C$ implementations. $g_m - C$ solutions are typically used to achieve a high frequency flexibility which is why they are chosen here. However, $g_m - C$ filters are known to be less linear than active- RC solutions. Hence it is crucial that the $g_m - C$ filter is optimally designed with respect to linearity. Special attention is therefore given to linearity and power optimization throughout the next two sections.

10.2.1 Architectural Study of the $g_m - C$ Biquad

Figure 10.1 represents the $g_m - C$ biquad architecture. Six filter parameters (4 g_m ’s, 2 C ’s) need to be optimized. The values of these parameters not only determine the

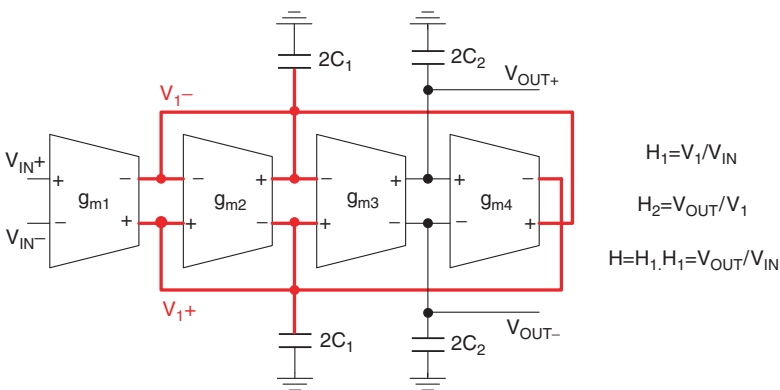


Fig. 10.1 The g_m - C biquad architecture. Only one internal node can be identified which is marked in bold (V_1). The node to node transfer functions H_1 and H_2 are also shown (10.5)

frequency response but also noise, power and distortion performance of the biquad and should therefore be taken into account early in the design flow leading to an optimal $g_m - C$ biquad sizing strategy. The traditional frequency response is determined by three independent equations: DC gain (A_0), bandwidth (f_0) and quality factor (Q).

$$A_0 = \frac{g_{m4}}{g_{m1}} \quad \omega_0^2 = \frac{g_{m3}g_{m4}}{C_1C_2} \quad Q = \frac{C_1}{g_{m2}}\omega_0^2 \quad (10.1)$$

Substituting those equations in the typical second order transfer function gives us the frequency response in function of the biquads design parameters.

$$H(s) = \frac{A_0 \omega_0^2}{s^2 + \omega_0/Q s + \omega_0^2} = \frac{\frac{g_{m1}}{g_{m4}} \frac{g_{m3}g_{m4}}{C_1C_2}}{s^2 + \frac{g_{m2}}{C_2}s + \frac{g_{m3}g_{m4}}{C_1C_2}} \quad (10.2)$$

Consequently three more design conditions are required to uniquely define the biquadratic section. These are derived from a distortion and noise analysis applying a design strategy which optimizes the filter already at architectural level [15, 16]. Clearly, choosing proper transconductance values is a key issue as they influence the signal swing on the internal nodes. As the circuit is most sensitive to distortion where large signal levels exist, amplification to these nodes must be avoided. An in depth linearity study will be performed in the next section and design rules will be proposed.

10.2.2 Nauta’s Transconductor

Nauta’s transconductor is based on six inverters (Fig. 10.2) and is very well suited for high-frequency and low-voltage applications [6]. The absence of internal nodes and the associated extra poles and zeros allow a high degree of reconfigurability

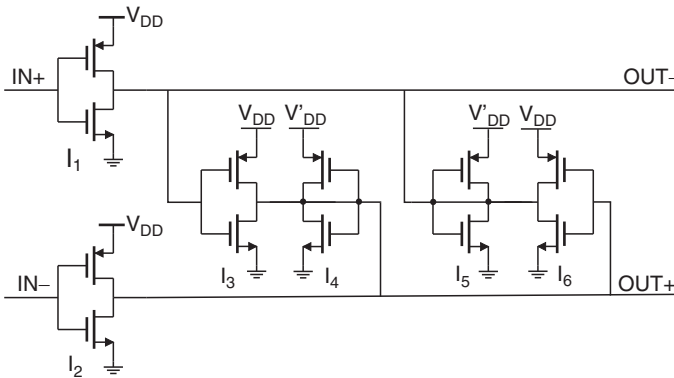


Fig. 10.2 Nauta’s transconductor

which is required when switching between different modes. A differential voltage is turned into a current by the transconductance of only a single inverter. This means that if a higher bandwidth is wanted, only the inverter needs to be resized (reconfigured) to obtain a larger transconductance. This is in contrast with e.g. two-stage OTA's (operational transconductance amplifier) where attention has to be paid to the internal stability and the gain bandwidth. Both are limited by the location of higher order poles. Additionally, the typical common mode feedback loop is avoided. Another property is that the gate capacitance significantly contributes to C_1 and C_2 which opens an opportunity to use only intrinsic capacitance which increases the chip density. Clearly, Nauta's transconductor helps to reduce the complexity in highly flexible systems.

I_1 and I_2 generate the transconductance function (differentially) while I_3 , I_4 , I_5 and I_6 guarantee common mode (CM) stability. The principle is to generate a high-ohmic load for differential signals (resulting in a high DC gain) and a low-ohmic load for common mode signals (being less sensitive to CM perturbations). Gain expressions for common mode and differential gain are:

$$A_{cm} = \frac{g_{m1}}{g_{ds1} + g_{ds5} + g_{ds6} + g_{m5} + g_{m6}} \approx \frac{g_m}{3g_{ds} + 2g_m} \quad (10.3)$$

$$A_{dif} = \frac{g_{m1}}{g_{ds1} + g_{ds5} + g_{ds6} + g_{m5} - g_{m6}} \approx \frac{g_m}{3g_{ds} + \delta g_m} \quad (10.4)$$

Equation (10.4) learns that if I_5 and I_6 are perfectly matched, they have a negligible effect on the DC gain. By tuning g_{m6} , it is even possible to achieve a very high DC gain because the output conductances can be tuned away (DC enhancement). In our design this feature is not exploited since sufficient gain was already realized. Sufficient gain and matching is achieved by choosing the dimensions large enough.

Equation (10.3) learns that common mode stability is always guaranteed but it also shows that there is a large safety margin as $A_{cm} \approx 0.5$. As long as the common mode gain remains smaller than one, common mode signals will be suppressed. How close A_{cm} can approximate 1 is dependent on the amount of common mode signal present at the input. In well designed systems this amount is kept very low such that A_{cm} can be increased from 0.5 closer to 1. This will be exploited to make the filter more power efficient.

In a traditional Nauta, this power efficiency is only 33% (2/6 useful inverters). By resizing the middle inverters, the power consumption can be significantly reduced. As a consequence A_{cm} will come closer to 1. Recall that the maximum A_{cm} will depend on the system specification and allowed sensitivity to common mode input signals. The maximum improvement (at transconductance level) is to resize the middle inverters such that their g_m value is halved. In reality we need to take some larger values otherwise process variations could jeopardize the stability ($A_{cm} \ll 1$). The improved Nauta improves the power efficiency from 33% to almost 50%. Note that the CM inverters also have less contribution to the total noise.

To further improve the efficiency of this transconductor, new CM stability criteria are proposed taking into account the full biquad topology. In the filter biquad

topology, the output of the first, second and fourth transconductor are shared. (Highlighted in Fig. 10.1.) This means that the CM inverters of those three transconductors can be shared and two of them can be left out. The power efficiency is then further improved by introducing stability criteria per filter node instead of per transconductor separately, still guaranteeing a node to node common mode gain smaller than one. By rescaling the two remaining CM inverters, the power efficiency is improved to about 70%. This leads to a power reduction of 50% compared to a traditional based Nauta.

10.3 Design Optimization and Strategy

10.3.1 Architecture Optimization

In a $g_m - C$ filter, high linearity is not easily obtained. Therefore this section mainly focuses on the linearity performance. Good linearity is achieved if the distortion level is minimized in the entire filter bandwidth. Therefore the biquad architecture presented in Section 10.2 is first studied in more detail allowing us to present linearity guidelines further on. Finally, the architecture is optimized to obtain the best noise-power trade-off.

10.3.1.1 Optimal in band linearity performance

To optimize the linearity performance, the signal behavior on all nodes of the filter needs to be analyzed to identify the amplifying stages. The chosen topology contains only one internal node (Fig. 10.1). Simplified expressions for the transfer function from input to internal node (H_1) and from that node to the output (H_2) are given by:

$$H_1(s) = \frac{\frac{g_{m1}g_{ds3}}{C_1C_2} + \frac{g_{m1}}{C_1}s}{s^2 + \frac{g_{m2}}{C_2}s + \frac{g_{m3}g_{m4}}{C_1C_2}} \quad H_2(s) = \frac{\frac{g_{m1}g_{m3}}{C_1C_2}s}{\frac{g_{m1}}{C_1}s + \frac{g_{m1}g_{ds3}}{C_1C_2}} \quad (10.5)$$

These transfer functions are plotted in Fig. 10.3 for a 10 MHz example where the g_m values are randomly chosen only based on the desired frequency response. Equations (10.5) indicate a bandpass characteristic from the input to the internal node. This means that the signal is first attenuated and then again amplified to its original level. However this is not the case for signals close to the cut-off frequency (f_0). If the transconductance and capacitor values are not properly chosen, it may happen that for signals close to f_0 , an amplification to the internal node occurs due to this bandpass behavior (Fig. 10.3a). It is primordial that this behavior is avoided as it degrades the linearity performance. Unfortunately, this is not always taken into account resulting in a suboptimal design. Moreover, linearity measurements are mostly performed at frequencies far enough from f_0 hiding this effect.

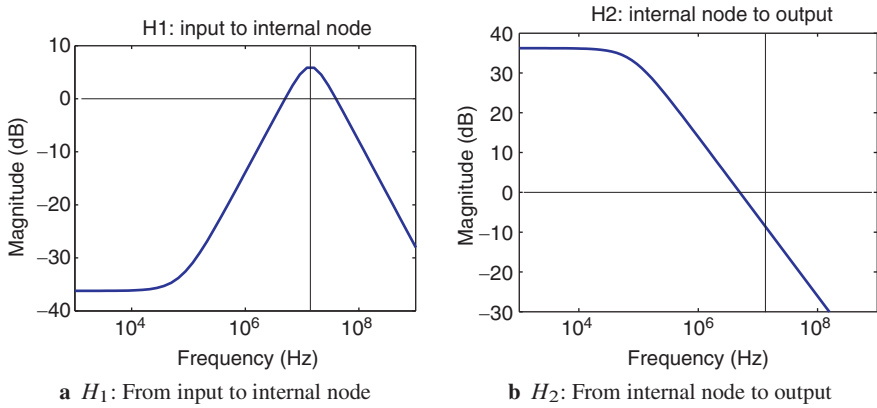


Fig. 10.3 Transfer functions of the $g_m - C$ biquadratic filter: a band-pass characteristic is identified to the internal node

Design rule for optimal linearity in the full bandwidth. To avoid an amplification of the signal to the internal node and to obtain an acceptable level of nonlinearity, also near f_0 , the maximum allowable amplification has to be defined. The peak value of H_1 (10.5) is given by g_{m1}/g_{m2} . For linearity, this ratio must be lowered without changing the filter characteristic. As discussed at the end of this subsection, this is comparable with a technique which in literature is referred to as state scaling [5, 14]. Since the signal is spread over the entire band, L_2 scaling is preferred. The best way to lower the peak value is to increase g_{m2} . Multiplying g_{m2} with a factor x leads to a multiplication with x for g_{m3} and C_1 to maintain the same frequency response.

$$g_{m2} \cdot x \rightarrow g_{m3} \cdot x \rightarrow C_1 \cdot x \tag{10.6}$$

This multiplication results in an increase of g_{m3} leading to an increase of g_{ds3} . The latter results in a shift of the zero frequency of H_1 ($s = g_{ds3}/C_2$) closer to the pole frequencies at f_0 thus limiting the rising part of H_1 (10.5). Shifting this zero of H_1 has no effect on the total frequency response as it is cancelled by the denominator of H_2 (10.5). Note that the proposed scaling keeps the g_{m3}/g_{m2} ratio constant while scaling g_{m1} would change it. Further on it is demonstrated that the noise-power trade off imposes the g_{m3}/g_{m2} ratio. Therefore (10.6) is the most suited way of scaling to determine g_{m2}/g_{m1} . To derive its optimum value, a trade-off between linearity at lower and higher frequencies is necessary (Fig. 10.4).

Linearity performance is expressed in terms of the third order intermodulation product. This can easily be simulated by applying a two-tone test. By changing the input frequencies of the input tones while keeping the input amplitude fixed, the frequency dependent linearity performance can be studied. Moreover, by modifying the g_{m2}/g_{m1} ratio, the relationship between the height of the bandpass peak and the linearity can be monitored. This approach is applied on a 10 MHz Butterworth biquad section from which the results are shown in Fig. 10.4 in terms of $|IM3|$ for a frequency range from 1 to 10 MHz. The simulations with SpectreRF confirm

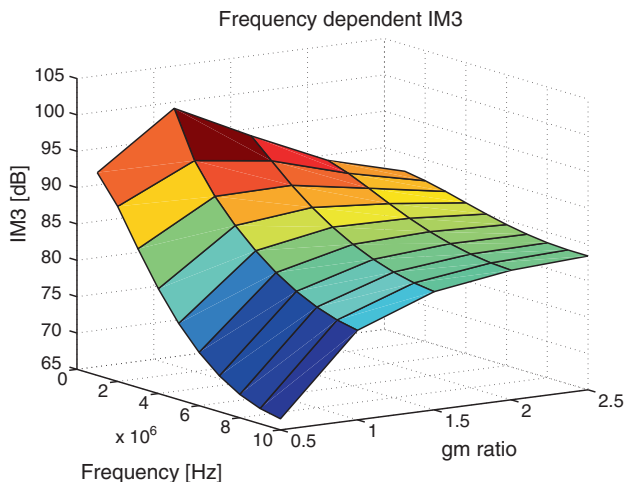


Fig. 10.4 Frequency dependent $IM3$ simulation as function of g_{m2}/g_{m1} : $|IM3| \approx 1.4$ gives the optimal distortion

the decreasing linearity performance very close to f_0 . In addition, they prove that lowering the bandpass peak by increasing the g_{m2}/g_{m1} ratio indeed optimizes the linearity (Fig. 10.4). From a certain value, the linearity improvement near f_0 is limited, so further scaling would lead to performance loss at lower frequencies. It is concluded that an acceptable level of linearity in the full bandwidth is obtained for

$$g_{m2} \approx \sqrt{2} g_{m1}. \quad (10.7)$$

Calculations indicate that this corresponds to a maximum internal signal level of -3 dB. This level is now defined as the minimum attenuation for signals to the internal node to achieve a good linearity.

Standard filter design guidelines [5, 14] propose to balance the node signal levels to optimize the dynamic range. In this sense, we propose a similar method by deriving the maximum signal level on the internal node. For higher order filters, this leads to transconductance scaling (state scaling). For a single biquad, a common seen design results in $g_{m1} = g_{m3} = g_{m4}$ and $g_{m2} = g_{m1}/Q$ where the ratio is quality factor dependent. This can possibly show a linearity drop near f_0 which is avoided in our approach.

Out-of-band interferer characterization. A final note is needed with respect to the out-of-band linearity performance. Typical for this type of $g_m - C$ filter is the increasing distortion for out-of-band interferers. Since the transfer function shows an attenuation to the internal node, the next transconductors are of less importance due to the low signal level. Therefore, the out-of-band $IIP3$ converges to the $IIP3$ of the input transconductor. As a consequence, the out-of-band linearity will only be determined by the implementation choice of the transconductor. One must guarantee that this choice, together with the applied circuit level techniques is sufficient to fulfill the systems specifications.

10.3.1.2 Optimal noise-power trade-off

Given the g_m ratio for optimal linearity (10.7), the remaining degree of freedom is fixed by minimizing the power consumption of the filter for a wanted noise specification. A general expression is derived based on analytical expressions taking into account the imposed g_{m2}/g_{m1} ratio (10.7) for linearity reasons. The noise contributions of the four transconductors are given by their equivalent input referred noise source. The noise transfer functions are then calculated, followed by the equivalent input voltage noise source for the full biquadratic section. Finally one obtains an expression for the integrated input referred noise (IRN):

$$IRN^2 = \frac{4kT\alpha \cdot BW}{g_{m1}} \left(1 + \frac{g_{m2}}{g_{m1}} + \frac{g_{m2}^2(3+Q)}{3g_{m1}g_{m3}} + \frac{g_{m4}}{g_{m1}} \right) \quad (10.8)$$

Design rule for optimal noise-power trade-off. Now it is possible to optimize the noise given by (10.8) by minimizing the power consumption for a given noise specification. A measure for power consumption is given by:

$$\sum g_{mi} \quad (10.9)$$

Then we substitute:

$$\begin{cases} g_{m1} = G_M & \text{relative value} \\ g_{m2} = a \cdot G_M & \text{linearity rule} \\ g_{m3} = x \cdot G_M & x \text{ to be optimized} \\ g_{m4} = G_M & \text{DC gain} \end{cases} \quad (10.10)$$

The optimal value for x is found by minimizing (10.9) given the noise specification (10.8). This optimum is translated in design rule (10.11). Note that this confirms that high Q sections are noisier than their low Q counterparts.

$$g_{m3}/g_{m2} = \sqrt{1 + Q^2/3} \quad (10.11)$$

With this last design rule, the six design equations are defined and the $g_m - C$ biquad is uniquely defined.

10.3.2 Circuit Level Implementation and Validation

10.3.2.1 Bottom Up Circuit Optimization

Linearity of a MOS transistor. In order to perform an elementary linearity study, the basic textbook formula of a MOS transistor in saturation is extended. The main causes of third order distortion are due to mobility reduction and velocity saturation.

These effects will be included in the models and equations. Only the effect of a nonlinear transconductance will be studied. The output conductance is assumed to be linear. This assumption will prove to be accurate as long as the output is a low impedant point. For a high impedant point, the contributions due to a non linear output conductance are important as well and may start dominating. To allow hand calculations, first the harmonic distortion $HD3$ has been calculated to identify the main parameters influencing the third order distortion. Afterwards simulations will be performed to obtain more accurate results needed in submicron technologies. The intermodulation product $IM3$ based on a two-tone test is then analyzed. We find for $HD3$ [17]:

$$\frac{V_{in}^2 \left(\theta + \frac{\mu_0}{v_{sat}L} \right)}{4 \left(1 + \left(\theta + \frac{\mu_0}{v_{sat}L} \right) (V_{GS} - V_T) \right)^2 \left(2 + \left(\theta + \frac{\mu_0}{v_{sat}L} \right) (V_{GS} - V_T) \right) (V_{GS} - V_T)} \quad (10.12)$$

with θ the mobility reduction coefficient, μ_0 the surface mobility and v_{sat} the saturation velocity. Equation (10.12) learns that the most important parameter to limit distortion is the overdrive voltage $V_{GS} - V_T$. It is of extreme importance to maximize its value to obtain minimum distortion. Secondly, the length has to be large enough such that these contributions are non dominant; the linearity becomes worse for minimal length transistors. The absolute values differ for NMOS and PMOS because of their physical differences. Because the mobility degradation and velocity saturation are less pronounced in the PMOS, it is expected that a PMOS transistor in a $0.13 \mu\text{m}$ technology is more linear than the NMOS for a given overdrive voltage.

Simulations results, measured on the output current of a MOS loaded with an AC short circuit, indicate a difference around 5 dB in saturation and strong inversion regime (Fig. 10.5). Therefore a slightly higher overdrive should be given to the NMOS when sizing the inverter. Furthermore, if the length is chosen higher than $1 \mu\text{m}$ the amount of distortion remains almost constant.

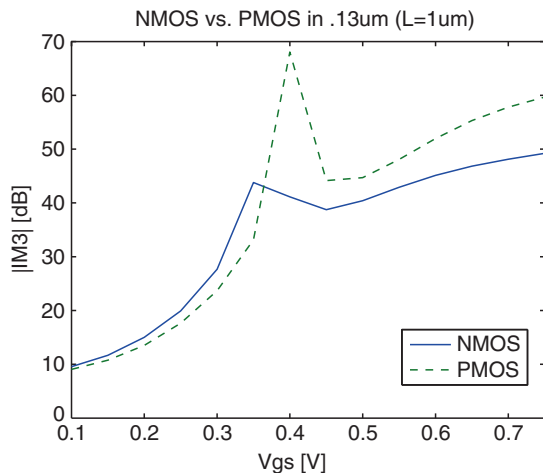


Fig. 10.5 Linearity of a MOS transistor in a $0.13 \mu\text{m}$ technology illustrating that a PMOS is more linear (large L)

Optimal bias point of the transconductor. To derive the optimum bias point of Nauta's transconductor, circuit simulations have been performed with the input common mode voltage (V_{cm}) as variable for a length of $1\ \mu\text{m}$ for both NMOS and PMOS. Because V_{cm} affects the V_{gs} of both transistors oppositely, V_{cm} determines the width ratio as the same current has to flow through both transistors. The optimum common mode level is found around $0.625\ \text{V}$ for a $1.2\ \text{V}$ supply giving a ratio R of $(W_p/L_p)/(W_n/L_n) \sim 8$. This means that indeed a little more overdrive is given to the NMOS to compensate its inferior linearity performance against the PMOS. Moreover, this result proves to be robust against process variations and mismatch which is proven by measurements in Section 10.5.2. Simulations prove that a smaller R leads to better immunity against second order distortion originating from mismatches and large blocking as the $IIP2$ optimum for the single ended filter is found for $R = 5$. $IIP2$ measurements and mismatch are further considered in Section 10.5.2. Since large sizes are required (e.g. L_{min}), mismatch is not expected to give problems. Therefore the optimal $R = 8$ to minimize third order distortion has been chosen.

The final biquad is then constructed as represented in Fig. 10.1. The relative values for the four transconductors and capacitors are chosen based on all derived design rules. The absolute numbers are function of the bandwidth. The performance of a $10\ \text{MHz}$ example is discussed next.

10.3.2.2 Simulated performance summary of the biquad $g_m - C$ filter

The proposed methodology has been applied on a biquad low-pass filter design. The filter's frequency response (Fig. 10.6) shows a maximum flat ($Q = 0.7$) $10\ \text{MHz}$

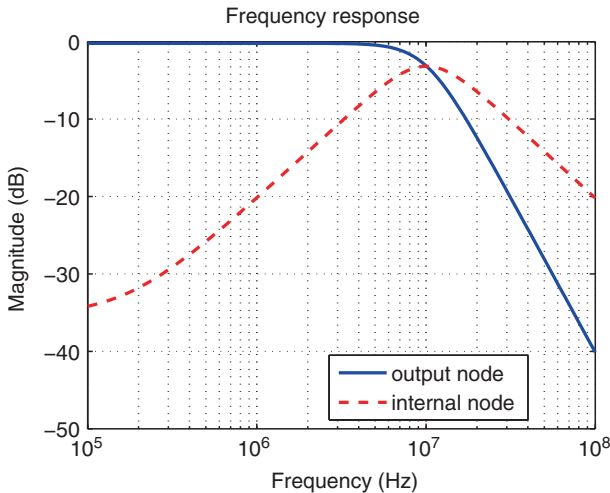


Fig. 10.6 Frequency response: A maximum flat Butterworth filter with $f_0=10\ \text{MHz}$

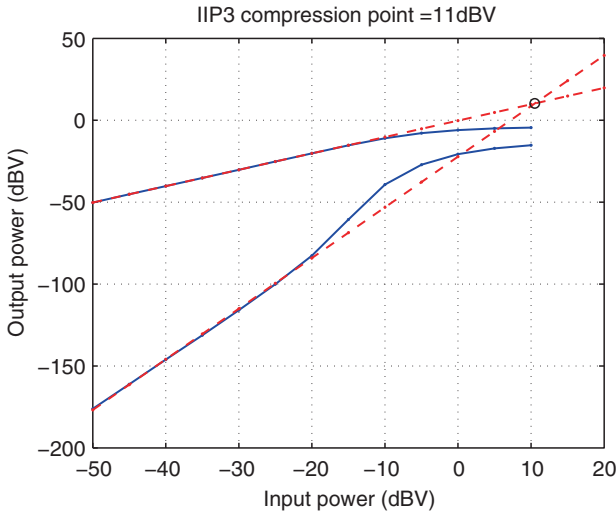


Fig. 10.7 IIP3 simulation where the input tones are well in band (3–4 MHz)

cut-off Butterworth shape. The peak on the internal node is 3 dB lower than the DC gain as imposed by design Eq. (10.7). Without using special circuit techniques such as source degeneration, the achieved $IIP3$ for a Nauta transconductor is 11 dBVp (Fig. 10.7). Furthermore, noise and power consumption are optimized with respect to the results of the linearity. This noise-power trade-off (10.11) leads to an input referred noise level of $33.75 \mu\text{V}_{\text{rms}}$ for a current consumption of 2.5 mA from a 1.2 V supply.

10.4 Multimode Reconfigurable $g_m - C$ Filters

10.4.1 A Novel Switching Technique to Enable High Flexibility

High flexibility is understood as the ability to reconfigure the desired filter response, noise level, bias point etc. This is achievable if the physical parameters of each transistor (W and L) can be chosen at runtime. By splitting up each transistor into a matrix of unit cells (Fig. 10.8) and by using both parallel and series switching in that unit cell, independent control of the total transistor length and width becomes possible [18].

10.4.1.1 Width control using parallel switching

In the case of parallel switching of inverter unit cells, the g_m scales linearly with the number of active cells as this corresponds to changing the total width of the

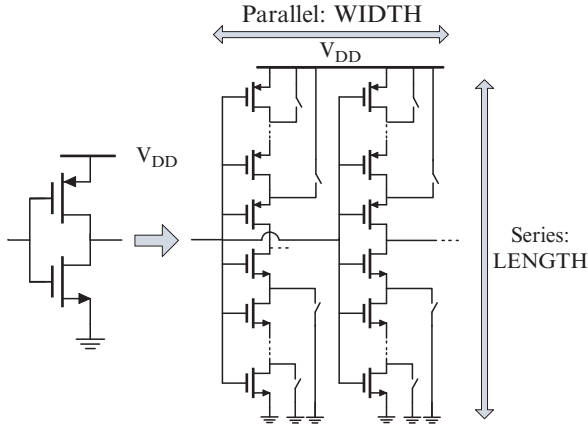


Fig. 10.8 Each transconductor is split into a matrix of inverter unit cells allowing to tune the width and length at runtime

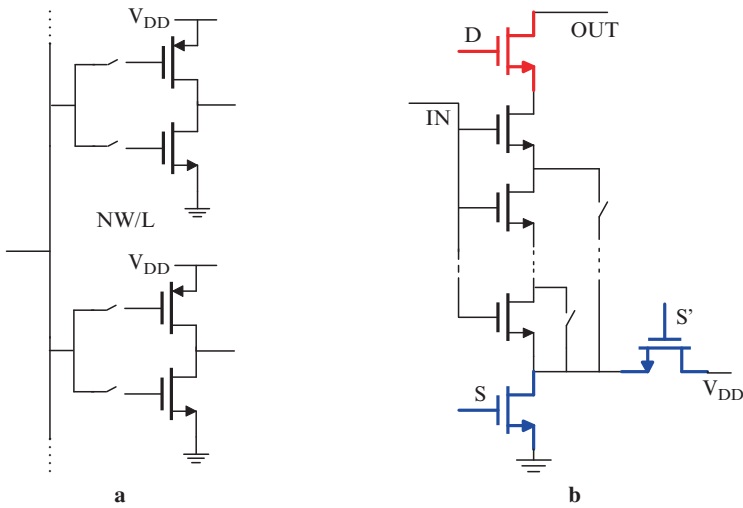


Fig. 10.9 (a) Parallel switching at the gate allows no independent capacitance tuning. (b) Parallel switching at drain and source allows both g_m and C tuning

transistors. Parallel switching can be done in three different ways: at the gate, drain and source. Analyzing their properties results in the following statements.

1. *Gate switching does not provide independent capacitance tuning*
 Placing a switch in series with the gate turns out to be impractical (Fig. 10.9a). When transistors are switched off, they are completely invisible which prevents using them for independent C -tuning. Moreover, the switch needs to be sized too large because of the parasitic pole formed by the on resistance and the large input capacitance.

2. Drain switching for ON capacitance

Opening the top switch (D in Fig. 10.9b) breaks the current and therefore the unit cell does not contribute to the total g_m . If the switch is on, then the g_m of the cell is determined by the length control. In both cases, the channel of the regular transistors is still active such that their gate oxide capacitance is always seen at the input. In the switched off case this capacitance is even little higher because now all transistors are in triode region.

3. Source switching for OFF capacitance

Switching off the bottom transistor (S in Fig. 10.9b) also breaks the current which means that the unit cell does not contribute to the total g_m . At the same time, signal S' is high, such that the source of the bottom transistor is brought to V_{DD} . This yields a negative overdrive voltage for the regular transistors, resulting in a considerably lower input capacitance.

4. Combination of drain and source switching allows independent g_m and C tuning

Implementing a unit cell with switches at drain and source results in three configuration modes.

- $D = 1$ & $S = 1 \rightarrow g_m$ ON & C ON
- $D = 0$ & $S = 1 \rightarrow g_m$ OFF & C ON
- $D = 1$ & $S = 0 \rightarrow g_m$ OFF & C OFF

This *independent* configuration of g_m and C of a transconductor offers us the required flexibility for every filter design. The fact that gate capacitance is used for the filter functionality offers a high density implementation in a pure digital process.

10.4.1.2 Length control using series switching

The principle of series switching or length control is shown in Fig. 10.10. The transistor closest to the output always operates in saturation as $V_d = V_g$. The others

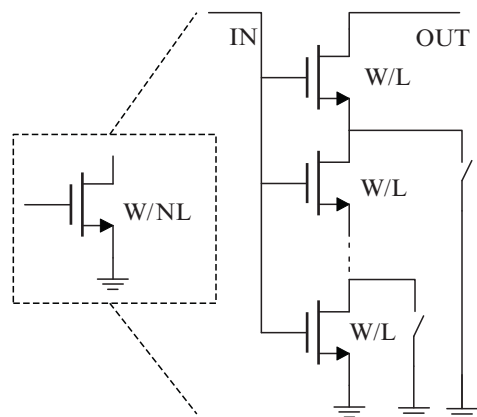


Fig. 10.10 Series switching provides length control. Series switching influences both transconductance and input capacitance

work in the triode region; it can be compared to a common source amplifier with source degeneration formed by the switch resistance. The situation where all N transistors are active is equivalent to a single transistor of length equal to NL . Note that if all transistors are chosen equal, the total transconductance does not scale linearly. The choice of the length is limited by f_T of the cell which is proportional to $1/L^2$. A larger length corresponds to a larger C_{gs} which limits the bandwidth of the cell. Transistors that are bypassed by a closed switch still contribute to the input capacitance. Therefore the transconductance can be programmed independently of that input capacitance which is controlled by the way of parallel switching as explained above.

10.4.2 From Flexible Transconductor to Flexible Filter

10.4.2.1 Design of the Unit Cell

Sizing for optimal linearity. The design guidelines for dedicated biquad design as discussed in Section 10.3 are applied here to optimally design the unit cell. Recall that a slightly higher overdrive is given to the NMOS to obtain the optimal ratio $R = (W_p/L_p)/(W_n/L_n)$ of 8 which sets the common mode input voltage. The minimum length has been defined as well ($\geq 1\ \mu\text{m}$) and remains valid for series switching.

Tunable frequency range and performance: a trade-off. The minimum length has been determined by linearity and mismatch considerations (see next paragraph). On the other hand, the frequency response will set a constraint on the maximum equivalent length. If the total length becomes too large, the WL product increases which in turn increases the input capacitance. A larger C_{gs} means a decrease of f_T as explained previously. Moreover, a length increase results in a larger off capacitance which can never be tuned away completely. Therefore a larger g_m is required for a high bandwidth mode which is unwanted due to power consumption reasons. A design of a unit cell which provides sufficient g_m for not too much C is required or in other words: the most efficient design is a unit cell with the highest C_{on}/C_{off} . Note that a higher capacitance ratio indeed lowers the power consumption as less active cells are required to achieve the same bandwidth.

In theory, the Nauta architecture can handle very high frequencies. However, practically C_{on}/C_{off} is limited to about 5, which limits the highest possible bandwidth if the tuning range starts in the 100 kHz region (GSM standard). Therefore the tuning range is set from a signal bandwidth of 100 kHz up to 20 MHz which still covers all modern standards of today and tomorrow (up to WLAN).

Mismatch. If the transconductor is split in unit cells, the mismatch should be controlled. Obviously the minimum size of the unit cell will therefore be limited.

Next to mismatch, also the linearity has put a boundary on the minimum size as explained before. It turns out, using the Pelgrom formulas, that with the minimum sizes ($W.L > 1 \mu\text{m}^2$) for linearity, the filters performance does not suffer from mismatch.

With respect to the intercell mismatch and the influence on the accuracy, two cases are considered. First, for high cut-off frequencies where a lot of cells are active, the equivalent dimensions are so high that mismatch is not a problem at all. Therefore, the accuracy or fine tuning of f_0 by switching on a unit cell is very high as well. Nevertheless, a layout technique is used such that the active cells are always optimally spread reducing the influence of process variations. For low cut-off frequencies, only few unit cells are switched on and therefore the relative fine tuning step becomes much smaller. To avoid mismatch problems in this case, the transconductance value is reduced not by scaling the width (by switching off cells) but by increasing the equivalent length of the unit cell, thus increasing its area. The flexible cell with its dimensions is presented below.

The flexible unit cell. Cells of equal width ($W = 1.12 \mu\text{m}$) have been chosen for layout and mismatch reasons. The best C_{on}/C_{off} ratio has been found when the length of the NMOS is twice the length of the PMOS. Therefore four times more PMOS cells are needed as the width is set by switching on unit cells in parallel. Moreover, too large $(W_p/L_p)/(W_n/L_n)$ ratios are avoided. The final lengths for the PMOS are: $L_p = [1 ; 0.2 ; 0.3 ; 0.5 ; 1] \mu\text{m}$ such that $L_{min} = 1 \mu\text{m}$ and $L_{max} = 3 \mu\text{m}$. For NMOS they are a factor two higher. The length steps are chosen to approximate a linear g_m variation.

10.4.2.2 Selecting the number of unit cells per transconductor

The final transconductor is built from unit cells with the above described switching properties. Switches are not in the critical signal path and proper design guarantees a negligible performance impact. The chosen number of implemented cells is based on a trade-off between area and flexibility. To meet the wanted tuning range, the width ratio W_{max}/W_{min} is around 1,024. However implementing 1,024 unit cells would be a waste of area since the total amount of capacitance present, even in the off mode, is still very high. Since this capacitance is higher than the one required for the low noise modes, scaling can be performed while the ratio of 1,024 is maintained. Area can be saved by using 255 NMOS transistor cells plus one 1/2 and one 1/4 cell to achieve 10 bits of resolution. Since those latter cells are only used for fine tuning, mismatch will not be affected as all other cells have the original dimensions. Additionally the number of cells in the first transconductor and in the common mode inverters is further reduced since those cells do not require any input capacitance and all cells are never used to contribute to the total transconductance for power reasons.

10.4.2.3 Control of the flexibility

With a smart coding scheme that allows tuning to every g_m/C ratio, it is possible to control the filters response and performance with a manageable number of control lines.

Concerning the series switching, all unit cells will be configured in the same mode so only $2.4 = 8$ control lines are needed for length control.

With respect to control of the parallel switching, just implementing a binary control method for the 255 cells is not possible as it would lead to a serious degradation of flexibility. For example, if 126 g_m cells are needed, only the MSB bit is zero and the others are set to 1, this means that all C flexibility is lost because $\#g_{m.on} = \#C_{on}$. Therefore the other 127 C cells have only one control line left so it is impossible to create all g_m/C ratios. Which bit is connected to which cell would differ for drain and source control and should be changed depending on the configuration. To solve this, the following scheme is proposed. A conceptual diagram is depicted in Fig. 10.11.

The unit cells are divided into two blocks, one for g_m control and one for C control. Only $N/2$ cells are able to provide g_m and only $N/2$ can provide C_{on} or

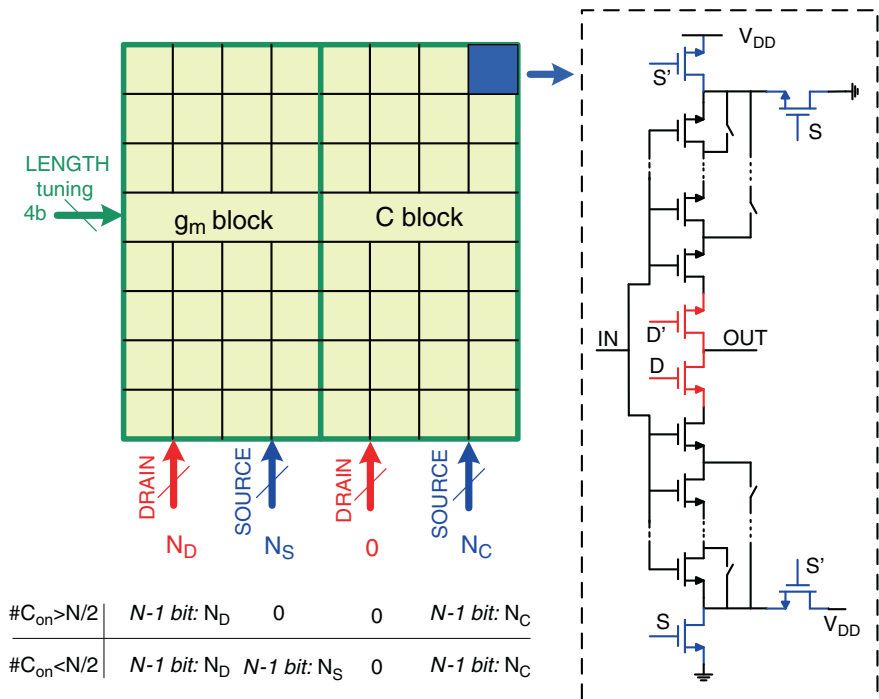


Fig. 10.11 Control mechanism for the matrix of switchable unit cells allowing every g_m/C ratio. Every transconductor is split into a g_m block and a C block for which specific control bits are valid controlling the drain and source switches of a unit cell

C_{off} . The switch matrix is controlled as follows: N_D is the number of drain switches on and N_S the number of source switches on in the g_m block; N_C is the number of C_{on} cells in the C block and u stands for unit.

C block. $N-1$ bits control the source, the drain control is set to zero.

g_m block. The configuration depends on the total capacitance.

- If more than $N/2 C_{on}$ cells are needed for small bandwidths: $N - 1$ bits control the drain, the source is set to zero. All cells contribute to C . If still more capacitance is needed, it can be provided by the C block.

$$\frac{g_m}{C} = \frac{N_D \cdot g_{m,u}}{\frac{N}{2} C_{on,u} + N_C C_{on,u} + (\frac{N}{2} - N_C) C_{off,u}} \quad (10.13)$$

- If less than $N/2 C_{on}$ cells are needed for larger bandwidths: $N - 1$ bits control both drain and source. Only the cells providing g_m , contribute to C , the other part is provided by the C block.

$$\frac{g_m}{C} = \frac{N_D \cdot g_{m,u}}{N_D C_{on,u} + (\frac{N}{2} - N_D) C_{off,u} + N_C C_{on,u} + (\frac{N}{2} - N_C) C_{off,u}} \quad (10.14)$$

With the implementation of this coding scheme, the full flexibility is programmable on a transparent manner.

10.5 Measurements Results

The $g_m - C$ biquad has been processed in a $0.13 \mu\text{m}$ 1.2V CMOS technology. The core area of the die shown in Fig. 10.12 occupies 0.5mm^2 . Note that different areas are occupied for different transconductors in correspondence with the required number of cells. Also, the common mode area has been significantly reduced as

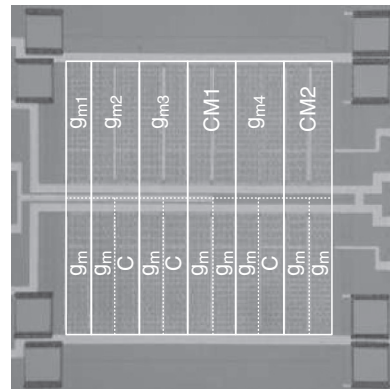


Fig. 10.12 Photograph of the die with core area of 0.5mm^2 : The top labels indicate to which transconductor the unit cells belongs. The labels at the bottom indicate the function of those unit cells (see Section 10.4.2)

discussed in Section 10.2.2. As this is a prototype, no internal supply regulator is present, however a clean supply is guaranteed (lab environment) to avoid supply rejection issues.

10.5.1 Flexible Frequency Response vs. Noise and Power Performance

The required noise level determines the total C and by selecting the correct g_m values, a very large range of bandwidths is achievable as it is proportional to g_m/C (10.1). Figure 10.13 shows a few of the many possible cut-off frequencies of a Butterworth frequency response ($Q = 0.707$). The bandwidth can be tuned from 100 kHz up to frequencies larger than 20 MHz achieving a tuning range of more than two orders of magnitude. Also different quality factors can be set (Fig. 10.13) which allows extension to higher order sections requiring biquads with various Q -factors.

The key feature for reconfigurable building blocks is power scalability with respect to the desired performance. Only then, the filter will be competitive with dedicated solutions as reconfigurability should have no power penalty. This implementation has a large number of possible power/performance trade-offs. For a constant integrated noise level (IRN), Fig. 10.14 shows how the power consumption

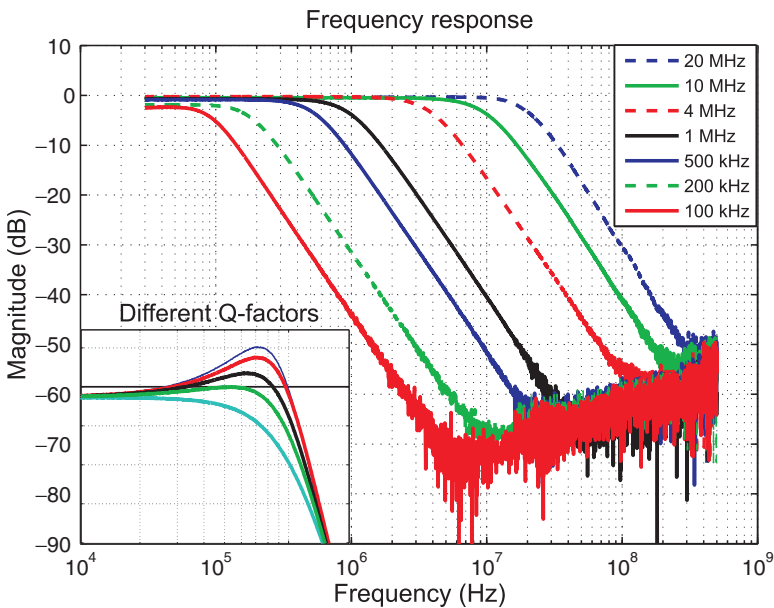


Fig. 10.13 Measurement results show a tunable bandwidth over more than two orders of magnitude. Also different Q -factors are possible which allows extension to higher orders sections

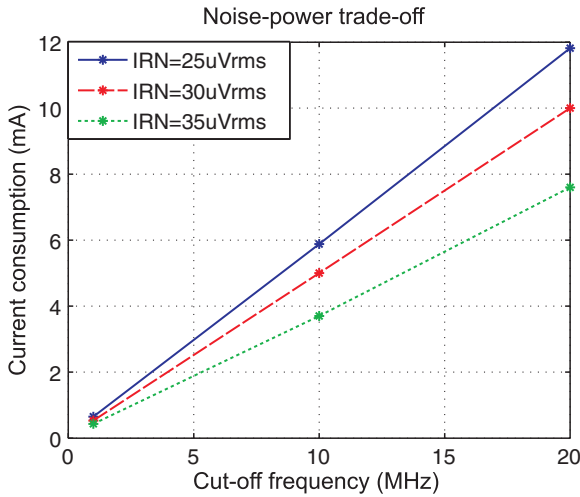


Fig. 10.14 Measurement results illustrating the noise-power-bandwidth trade-off: The power consumption is proportional to the performance

linearly scales with the bandwidth for three examples; clearly, noise and power can be traded. The current consumption scales from $103 \mu\text{A}$ (100 kHz, $35 \mu\text{Vrms}$ IRN) up to 11.85 mA (20 MHz, $25 \mu\text{Vrms}$ IRN) which is very efficient for such low IRN levels [10, 12].

10.5.2 Linearity Performance

Figure 10.15 shows an $IIP3$ measurement using two in-band tones that are close in frequency with $f_0 = 10 \text{ MHz}$. An $IIP3$ of 10 dBVp is achieved. Repeating this test near f_0 , the $IIP3$ drops to 7 dBVp as simulated. By following (10.7), this drop is limited with an almost negligible loss in linearity at lower frequencies. For out of band tones near $3f_0$, 6 dBVp is measured.

Figure 10.16 shows $|IM3|$ measurements (input amplitude = 0.1 V) for a 10 MHz filter with different g_{m2}/g_{m1} ratios. Measurements are presented where the two tones are well in-band and secondly near f_0 or close to the band-pass peak at the internal node. Increasing the g_{m2}/g_{m1} ratio indeed improves the linearity near f_0 as the maximum amplitude on the internal node is limited. A larger ratio than 1.5 does not make sense as the linearity improvement is limited in contrast with the rise in power consumption. Moreover, the linearity at lower frequencies drops. Measurements indicate that an acceptable level of linearity near f_0 is obtained for an internal signal level of -3 dB corresponding to a g_{m2}/g_{m1} ratio of 1.4 which confirms the earlier derived design rule (10.7).

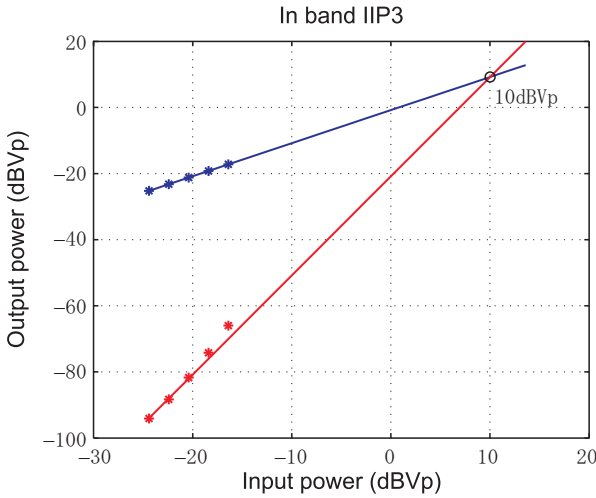


Fig. 10.15 IIP3 measurement for a 10 MHz filter (in-band excitation)

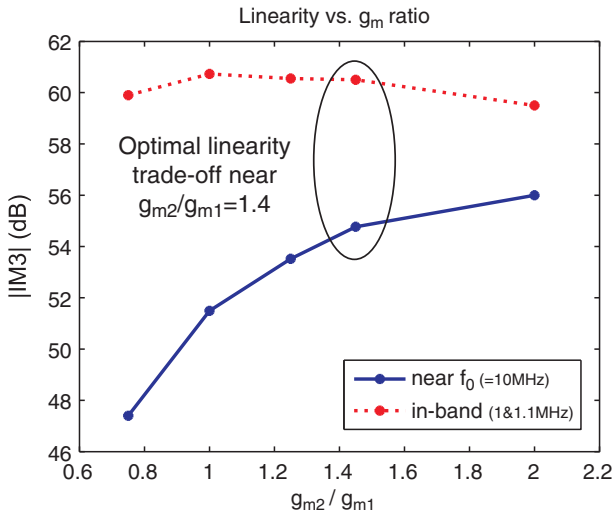


Fig. 10.16 $|IM3|$ measurements: Optimal g_{m2}/g_{m1} ratio for linearity in the full bandwidth is found near 1.4 ($A = 0.1$ V; in-band tones @ 1&1.1 MHz)

The effect of mismatch is characterized by $IIP2$. Differentially, values larger than 40 dBVp have been measured, and 24 dBVp at the single ended output. This high $IIP2$ value indicates that mismatch is limited in this design as a result of the large transistor sizes used. This also holds for the low bandwidth modes where the length is made considerably larger up to 6 μm for NMOS transistors which is also beneficial for $1/f$ noise.

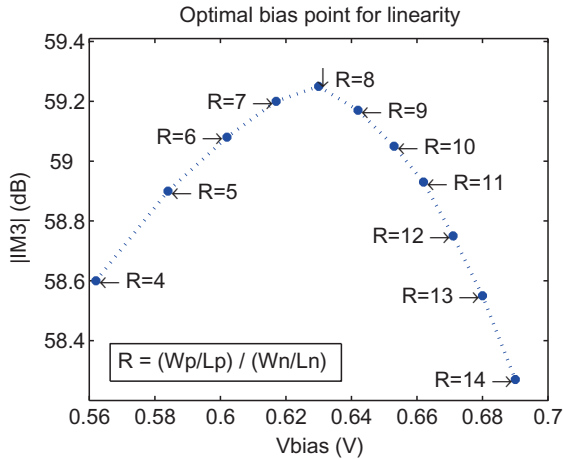


Fig. 10.17 Measurements validate the optimal bias point at $R = 8$

Bias point. By changing the number of PMOS cells, it is possible to change the $(W_p/L_p)/(W_n/L_n)$ ratio R and study the effect of a changing bias point (Fig. 10.17). This optimal ratio R is indeed found around 8 corresponding to a bias point around 0.625 V. However, these measurements indicate that the differences in linearity are limited despite a larger difference between PMOS and NMOS respectively. As long as the inverter is biased between 0.60 and 0.65 V, the linearity remains almost equally good. Therefore it makes sense to bias the inverter at 0.6 V corresponding to a ratio R of 6 instead of 8, leading to a significant area reduction of the PMOS part in the filter. Additionally, the immunity against second-order distortion originating from mismatch and blocking would improve.

10.5.3 Low Supply Voltage Performance

To illustrate the extension towards low supply voltage, the filter operation with a supply voltage of 1 and 0.8 V has been verified. Lowering the supply voltage reduces the g_m value and thus the bandwidth. This is compensated by turning on more g_m cells; sufficient cells are available even for 20 MHz. For the same bandwidth and noise level, the g_m/I increases and therefore the power consumption significantly reduces. Because of the lower overdrive however, the linearity becomes worse. For a $V_{DD} = 1$ V, one third of the power is saved for a drop in $IP3$ of about 2 dB. A V_{DD} of 0.8 V gives a drop around 4 dB for 66% less power. All flexibility is preserved for these low supply voltages.

10.5.4 Performance Summary and Comparison

The measurements results have indicated that this design not only has a large tuning range but it also obtains a very nice performance. How nice can be evaluated by comparing the results with previous work. However it is hard to make a fair comparison as filters vary in order, bandwidth, performance and tuning range. In literature, different figures of merit (*FoM*) exist but they all use the spurious free dynamic range (*SFDR*) as a performance measure taking into account linearity and noise. *SFDR* is defined as:

$$SFDR = \left(\frac{P_{IIP3}}{P_N} \right)^{2/3} \quad (10.15)$$

where P_{IIP3} is the input power of the third order intercept point and P_N is the input-referred noise power. One *FoM* has been proposed by [12]:

$$FoM = 10 \log_{10} \left(\frac{SFDR \cdot f_0 \cdot TR}{PpP} \right) \quad (10.16)$$

where f_0 is the cut-off frequency, TR the tuning range defined as f_{max}/f_{min} and PpP the power per pole. However, the weight of the tuning range is quite large compared to the actual performance making it easy for this design to have the best *FoM* thanks to its large tuning range of 200.

Another *FoM* has been described in [11, 13]. This *FoM* looks similar but the *SFDR* is normalized with a factor $N^{4/3}$ resulting in:

$$SFDR_N = SFDR \cdot N^{4/3} \quad (10.17)$$

where N is the filter order. The higher the filter order, the more the obtained *SFDR* is appreciated. In this way, the *SFDR* has a more equal importance in the calculation of the figure of merit compared to the tuning range. Therefore, the authors believe that a *FoM* which allows a fair comparison can be defined as:

$$FoM = \frac{PpP}{SFDR_N \cdot f_0 \cdot TR} = \frac{P_{tot}}{N \cdot SFDR \cdot N^{4/3} \cdot f_0 \cdot TR} [J] \quad (10.18)$$

This design is now compared with previous works on low-pass filters. Table 10.1 gives an overview of the most important parameters required to calculate the *FoM*. Figure 10.18 plots the *FoM* defined in [11, 13] ((10.18) without TR) in function of the tuning range. It confirms that this design has the largest tuning range together with a very competitive performance.

Table 10.1 Performance comparison for recent published works on low-pass filters (normalized to 0 dB gain)

	Topology	Technology	Order	Power-supply voltage	$f_{co,min}$ – $f_{co,max}$	Noise [μ Vrms]	IIP3 [dBm]
(a) Pav00 [9]	$G_m C$	0.25 μ m	4th	70 mW @ 3.3 V	60–350 MHz	257	n.r.
(b) Beh00 [8]	$G_m C$	0.6 μ m	5th	18–184 mW @ 3.3 V	4–18 MHz	40–64.5	18.5
(c) Hor03 [10]	$G_m C$	0.18 μ m	6th	10–15 mW @ 1.8 V	1.5–12 MHz	150	7.2–9.3
(d) Hor04 [11]	$G_m C$	0.18 μ m	4th	1.1–4.5 mW @ 1.8 V	0.5–12 MHz	n.r.	9.4–12.5
(e) Cha05 [12]	$G_m C$	0.25 μ m	3th	2.5–7.3 mW @ 2.5 V	50 kHz–2.2 MHz	13.6–520	22–28
(f) Gia07 [13]	$G_m RC$	0.13 μ m	6th	0.72–21.6 mW @ 1.2 V	0.35–23.5 MHz	85–163	19.8
This work	$G_m C$	0.13 μ m	2nd	0.12–14.2 mW @ 1.2 V	100 kHz–20 MHz	25–35	20

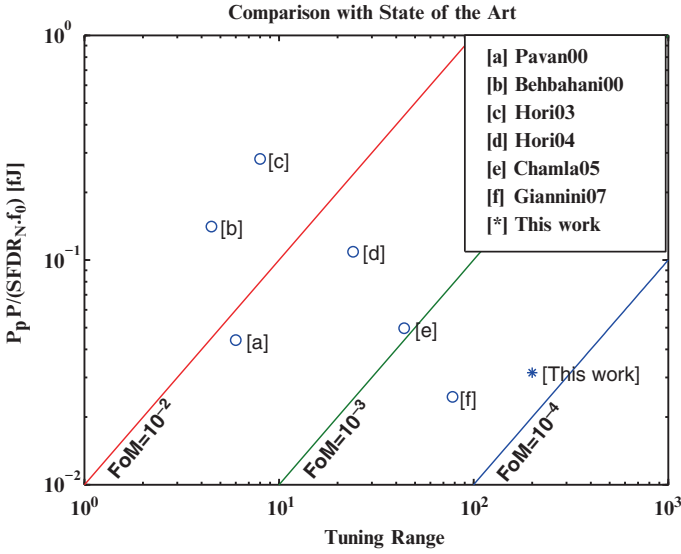


Fig. 10.18 Comparison with SoA: Performance in function of the tuning range

10.6 Conclusion

This chapter discusses a fully reconfigurable $g_m - C$ biquadratic low-pass filter which answers both the efficiency and flexibility demand of future mobile devices. In order to meet these challenging requirements, an efficient design strategy is followed with the main focus on linearity optimization and low power consumption. Furthermore, a novel switching strategy inside Nauta’s transconductor is implemented which allows a very large frequency-performance-power flexibility. This feature can be exploited not only to be compatible with the specifications of multiple standards but also to optimize power consumption according to varying channel conditions. Measurements of an implementation in 0.13 μm 1.2 V CMOS demonstrate a filter with a bandwidth tunable over more than two orders of magnitude starting from 100 kHz up to 20 MHz and with a performance scalable in terms of noise, power and linearity. On top of that, state of the art performance of dedicated designs is achieved. Also the extension to lower supply voltages has been validated. With this outstanding flexibility, the filter shows application potential for mobile devices of the next generation.

References

1. J. Craninckx et al., “A fully reconfigurable Software-Defined Radio transceiver in 0.13- μm CMOS”, *IEEE International Solid-State Circuits Conference*, San Francisco, CA, February 2007.
2. R. Chawla, G. Serrano, D. Allen, P. Hasler, “Programmable floating-gate second-order sections for Gm-C filter applications”, *Proc. IEEE MWSCAS*, vol. 2, pp. 1649–1652, August 2005.

3. U. Stehr, F. Henkel, L. Dallge, P. Waldow, "A Fully Differential CMOS Integrated 4th order Reconfigurable Gm-C Lowpass Filter for Mobile Communication", *International Conference on Electronics, Circuits and Systems*, vol 1, pp. 144–147, December 2003.
4. T. Deliyannis, Y. Sun, T. Fidler, *Continuous-Time Active Filter Design*. Boca Raton, FL: CRC Press, 1999.
5. J. E. Kardontchik, *Introduction to the Design of Transconductor-Capacitor Filters*. Boston, MA: Kluwer, 1992.
6. B. Nauta, "A CMOS Transconductance-C Filter for Very High Frequencies", *IEEE J. Solid-State Circuits*, vol. 27, pp. 142–152, February 1992.
7. E.A.M. Klumperink, B. Nauta, "Systematic Comparison of HF CMOS Transconductors", *IEEE TCAS-II: Analog and Digital Signal Processing*, vol. 50, no. 10, pp. 728–741, October 2003.
8. F. Behbahani, W. Tan, A. Karimi, A. Roithmeier, and A. Abidi, "A broad band tunable CMOS channel-select filter for a low-if wireless receiver", *IEEE J. Solid-State circuits*, vol. 35, no. 4, pp. 476–489, April 2000.
9. S. Pavan, Y. P. Tsvividis, and K. Nagaraj, "Widely programmable high-frequency Continuous-Time filters in digital CMOS Technology", *IEEE J. Solid-State circuits*, vol. 35, no. 4, pp. 503–511, April 2000.
10. S. Hori et al., "A widely tunable CMOS Gm-C filter with a negative source degeneration resistor transconductor", *Proc. IEEE ESSCIRC*, pp. 449–452, September 2003.
11. S. Hori et al., "Low-power widely tunable Gm-C filter with an adaptive DC-blocking triode-based MOSFET transconductor", *Proc. IEEE ESSCIRC*, pp. 99–102, September 2004.
12. D. Chamla, A. Kaiser, A. Cathelin, and D. Belot, "A Gm-C low-pass filter for Zero-IF mobile applications with a very wide tuning range", *IEEE J. Solid-State circuits*, vol. 40, no. 7, pp. 1443–1450, July 2005.
13. V. Giannini, J. Craninckx, S. D'Amico, and A. Baschiroto, "Flexible Baseband Analog Circuits for Software Defined Radio Front Ends", *IEEE J. Solid-State Circuits*, vol. 42, no. 7, pp. 1501–1512, July 2007.
14. Y. Palaskas and Y. Tsvividis, "Dynamic Range Optimization of Weakly Nonlinear Fully Balanced Gm-C Filters with Power Dissipation Constraints", *IEEE TCAS-II*, vol. 50, no. 10, pp. 714–727, October 2003.
15. P. Crombez, J. Craninckx, M. Steyaert, "A Linearity and Power Efficient Design Strategy for Architecture Optimization of Gm-C Biquadratic Filters", *Proc. IEEE PRIME*, pp. 229–232, July 2007.
16. P. Crombez, J. Craninckx, P. Wambacq, M. Steyaert, "A 100kHz - 20MHz Reconfigurable Power-Linearity Optimized Gm-C Biquad in 0.13- μm CMOS", *IEEE TCAS-II*, vol. 55, no. 3, pp. 224–228, March 2008.
17. P. Wambacq, W. Sansen, *Distortion Analysis of Analog Circuits*. Boston, MA: Kluwer, 1998.
18. P. Crombez, J. Craninckx, M. Steyaert, "A 100kHz - 20MHz Reconfigurable Nauta Gm-C Biquad Low-Pass Filter in 0.13 μm CMOS", *Proc. IEEE A-SSCC*, pp. 444–447, November 2007.

Chapter 11

An Adaptive Digital Front-End for Multi-mode Wireless Receivers

Gernot Hueber, Rainer Stuhlberger, and Andreas Springer

11.1 Introduction

Today, customers are demanding handhelds with small form factors and increased functionality for multiple applications ranging from text messaging and telephony to receiving audio and video broadcast services. Furthermore, the ongoing evolution from 2G to 3G and beyond requires multi-mode operation of wireless transceivers. The increasing number of cellular standards together with the variety of frequency bands these standards use in different regions of the world require a high degree of reconfigurability of the baseband processing and the RF front-end. As a result, reconfiguration has become the key issue in the design of wireless terminals. Although complexity increases by implementing multi-standard and multi-functional terminals and by integrating previously separated building blocks on a single die, the current consumption of the terminal needs to be reduced at the same time.

Making the RF part of the cellular transceiver capable of supporting different cellular standards is a difficult task as the reconfiguration of RF building blocks such as filters is usually only achievable in a very limited way at reasonable costs in terms of chip area and power consumption [1–4]. A different approach is the enhancement of analog front-end circuitry by adding dedicated on-chip digital signal-processing capabilities. This advanced digital circuitry is commonly called a digital front-end (DFE) [5–7]. Usually this DFE should account for most of the reconfigurability built

G. Hueber (✉)
DICE GmbH & Co KG, Austria
e-mail: gernot.hueber@infineon.com

R. Stuhlberger
DICE GmbH & Co KG, Austria
e-mail: rainer.stuhlberger@infineon.com

A. Springer
University of Linz, Austria
e-mail: a.springer@icie.jku.at

into the receiver by simply adapting the programming of its signal-processing building blocks in it. Depending on the wireless standards that are supported, the analog front-end (AFE) must fulfill a variety of different requirements. To save power and chip area, the AFE usually has a generic architecture, typically a zero-IF receiver, and is only slightly reconfigurable. The advantage of this approach is that shareable digital circuits can be used for the respective communication standards [8]. Additionally, the functional blocks can be adapted depending on instantaneous channel conditions. For example, the order of a filter can be reduced, or it can be bypassed completely if interfering signals are weak or absent. This technique can significantly save power.

The work presented by Yeung [6] only describes the concept of a DFE for a receiver focused on multiplier-less filter architectures. In this approach, static coefficients for the filter stages are foreseen, without any programmability for arbitrary matched filters. The DFE published by Martelli [7] is designed for cellular communication standards (GSM, UMTS) and WLAN, though its configuration options are limited. For instance, the sample-rate conversion is integer only, so the ADC rates have to be adapted to each standard, and phase or I/Q impairments are not considered. Both DFEs are static in a certain mode. In this chapter, we discuss the design of a DFE-based adaptive wireless receiver for multi-mode operation supporting GSM, EDGE, CDMA2000, UMTS, and Global Navigation Satellite System (GNSS). The reconfigurability is achieved by sharing common but programmable digital signal-processing stages for all operation modes and adapting them to the respective communications or navigation standards and to the current interference situation for minimum power consumption.

Based on the experience from an IC design [9] and investigations [5, 10, 11], in which some of the aspects of multi-mode receivers and the enhancement of RF receivers with a DFE have been presented, the current work extends this concept by adding adaptability to support multiple standards and to reduce power consumption. This can be achieved by using adaptive front-ends that tradeoff some RF performance (e.g. SNR, linearity, selectivity, inband magnitude ripple, group delay) for power consumption on the fly. This tradeoff is reasonable, in part because the adjacent channel selectivity (ACS) test case for UMTS specifies a challenging environment, whose probability of occurrence is small, as the simulated distribution of power levels predicts [12].

11.2 Interferer Scenarios

This section gives an overview of the main test cases for UMTS [13] and CDMA 2000 [14]. The ACS test case is a measure of a receiver's ability to receive a UMTS signal at its assigned channel frequency in the presence of an adjacent channel signal. ACS is the ratio of the receive filter attenuation on the assigned channel frequency to the receive filter attenuation on the adjacent channel(s). In UMTS, the

Table 11.1 The main test cases for UMTS and CDMA2000 with the power of the wanted signal at the antenna \hat{I}_{or} , the power of the the adjacent channel at the antenna \hat{I}_{oac} , the frequency offset f_{off} of the adjacent channel from the carrier frequency, the difference in power between wanted and adjacent channel ΔP , and the power of a continuous wave or narrowband blocker \hat{I}_{ibl}

Test case	Parameter	UMTS	CDMA2000
Ref Sens	\hat{I}_{or}	−106.7 dBm/3.84 MHz	−104 dBm/1.23 MHz
Max Input	\hat{I}_{or}	−25 dBm/3.84 MHz	−25 dBm/1.23 MHz
ACS	\hat{I}_{or}	−92.7 dBm/3.84 MHz	−101 dBm/1.23 MHz
	\hat{I}_{oac}	−52 dBm	−37 dBm
	f_{off}	5 MHz	2.5 MHz
	ΔP	40.7 dB	64 dB
CW/	\hat{I}_{or}	−96.7 dBm/3.84 MHz	−101 dBm/1.23 MHz
Narrowband	\hat{I}_{ibl}	−56 dBm	−30 dBm
	f_{off}	2.7 MHz	900 kHz

total received power of the wanted channel is −92.7 dBm/3.84 MHz. The power of the modulated interference signal on the adjacent channel (spaced 5 MHz apart) is specified as −52 dBm/3.84 MHz.

Additional important test cases are the reference sensitivity test case and the maximum input level test case. The reference sensitivity level is the minimum mean power of a UMTS signal received at the antenna so that the receiver is able to detect and to process it within predefined limits. The maximum input level test case defines the maximum mean power of a signal received at the antenna that the receiver is able to process. A summary of the test cases is given in Table 11.1.

11.3 Architecture of the Adaptive Multi-Mode Wireless Receiver

Figure 11.1 depicts the block diagram of zero-IF (ZIF) (also called direct-conversion) receiver architecture, which is currently the dominant architecture in the cellular terminal market [15]. One of the well-known main advantages of the ZIF architecture is its ability to support different channel bandwidths and, therefore, different standards, by reconfiguration of the analog baseband filters. This architecture is enhanced with an ADC for I- and Q-path and a subsequent DFE. Within the DFE, digital signal processing functions are performed for handling the diverse multi-standard requirements to facilitate the design of the AFE. The reconfigurability of digital signal processing in the DFE offers the extensive flexibility required for multi-mode operation. Additionally, problems of the ZIF architecture such as DC offsets and IQ-mismatch can be overcome effectively by appropriate digital signal processing in the DFE. However, optimum system performance can be achieved only by co-designing the DFE and the AFE in a closely matched process.

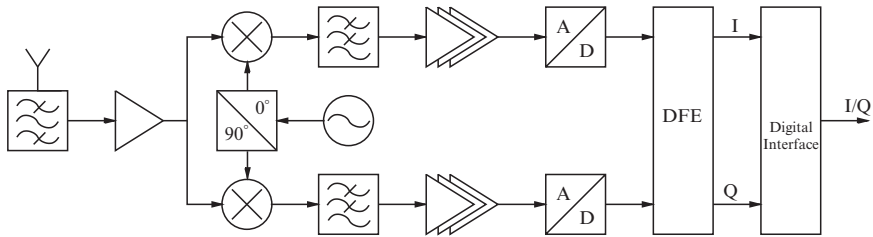


Fig. 11.1 Receiver architecture based on zero-IF topology enhanced with a DFE

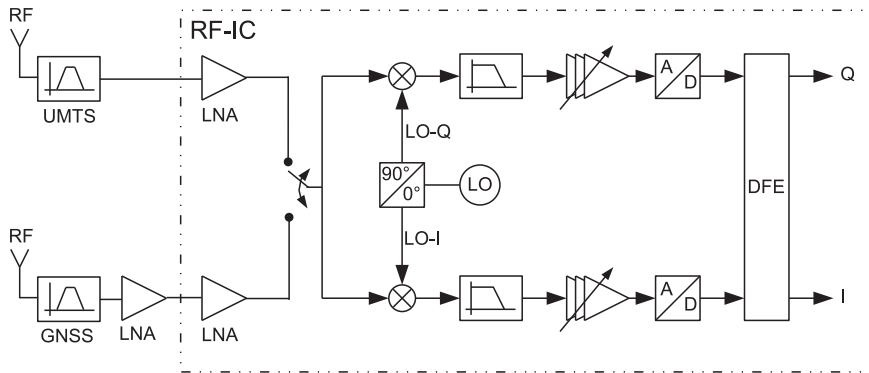


Fig. 11.2 Zero IF/low IF receiver front-end

Examples can be found in [9] and [16]. The receiver IC interfaces with the front-end module, which includes a first LNA and a set of RF filters for the various standards supported. The data is transferred to the baseband IC via a digital interface, which is the natural choice as the conversion to the digital domain is already accomplished in the receiver IC.

Furthermore, a configurable direct-conversion/low-IF receiver as depicted in Fig. 11.2 is investigated for cellular and GNSS dual-mode receivers. It is apparent from the differences in the Noise Figure (NF) specifications between GNSS and UMTS that the GNSS configuration requires an additional external LNA to meet the tight GNSS-NF specification. This concept enables UMTS-signal reception with the DCR configuration, whereas a low-IF architecture is used for GNSS. Both modes can be built with only one reconfigurable RF receiver. The choice of the low-IF mode for GNSS circumvents the flicker noise issue common for the DCR architecture, which would otherwise exceed the tolerable receiver NF. The image rejection is implemented in the digital domain. The DCR architecture, on the other hand, has much better receiver selectivity and lower power consumption, which are essential features for UMTS. The mode selection is mainly done in the DFE. Depending on the operation mode (UMTS or GNSS), the analog and digital signal processing building blocks are properly configured.

The main functions of the DFE in GNSS mode comprise:

- Channel selection filtering
- Decimation to reduce the sampling rate of the oversampling ADC
- Phase and group delay equalization of the analog anti-aliasing filter
- Coordinate-Rotation-Digital-Computer (CORDIC) based digital downconversion of the IF signal to baseband
- Hilbert transformer-based image suppression

The ADC is based on an oversampling $\Delta\Sigma$ architecture operating at a sampling frequency of more than 100 MHz. The channel selection filtering and the final integer decimation are accomplished with similar digital filter architectures as in the telecommunication mode. For the reception of GNSS signals, the DFE must include a digital down-conversion from low-IF to baseband, which is achieved by a CORDIC. To minimize impairments due to images folded into the wanted signal band, image rejection must be sufficient. Due to phase and amplitude imbalances, the Image Rejection Ratio (IRR) of classical analog polyphase filter-based implementations is limited to 30–40 dB at most. The IRR can be calculated by

$$\text{IRR} = \frac{1 - 2\epsilon \cos \phi + \epsilon^2}{1 + 2\epsilon \cos \phi + \epsilon^2} \quad (11.1)$$

In Eq. 11.1, ϵ denotes the I/Q gain mismatch and ϕ the I/Q phase error. A value for ϵ of 0.1 dB and for ϕ of 1° results in an IRR of approximately 40 dB. However, these specifications are extremely challenging for mass-market receiver designs due to unavoidable parameter variations over temperature and production tolerances. Higher IRR values require some means of calibrating the analog signal-processing chain.

11.3.1 Analog Receiver Architecture

The receive analog front-end (Rx-AFE) is based on the direct conversion architecture (see Fig. 11.2). A fractional-N phase-locked-loop (PLL) can be fully integrated on the chip. Digital tuning can be used for the loop filter, the voltage-controlled oscillator (VCO), and the charge pump to compensate for process and temperature variations. The analog baseband filter in front of the ADC only removes anti-aliasing components while channel selection filtering is performed in the DFE. Thus a simple low-order Butterworth filter with two different cut-off frequencies (e.g. 1.15 and 4.4 MHz) for small and wideband systems can be employed. Due to the high cutoff frequency settings with respect to the signal bandwidth, requirements on the calibration are low, and chip area is reduced.

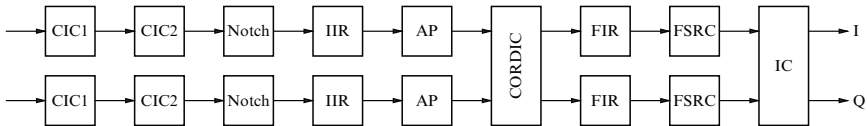


Fig. 11.3 Rx-DFE Architecture consisting of CIC-, notch-, IIR-, Allpass-, CORDIC, and FIR filters, a fractional sample rate conversion, and a building block for impairment correction

11.3.2 Digital Front-End

The digital signal-processing (DSP) blocks of the Rx-DFE depicted in Fig. 11.3 are the enabling enhancements for the multi-mode capability of this approach. As in most cellular applications, continuous-time $\Delta\Sigma$ ADCs are employed. They must be designed with sufficient dynamic range for sampling the desired signal accompanied with unwanted channel interferers, because these are suppressed only in the DFE. Thus the ADCs are clocked by a constant system clock with high oversampling ratios (OSRs) for all modes with adaptation to the signal bandwidth by tuning of the loop filter. Within the DFE, the high sample rate of the ADCs is transformed to a sample-rate as low as twice or four times the bit-, chip-, or symbol-rate of the supported standard. Due to the required different bit-, chip-, or symbol-rates, fractional sample-rate decimation is a must in the DFE. Further tasks of the DFE include channel selection filtering, gain control, integer sample-rate decimation, and matched filtering, which are highly dependent on the respective standards requirements. Furthermore, the DSP allows for the compensation of signal impairments introduced by the analog front-end, e.g. the correction of I/Q imbalances by means of blind signal-processing techniques [17]. In any practical implementation, the DFE functionality is fully configurable (e.g. via a three-wire bus-type control interface, with configuration options including all decimation factors, filter coefficients, gain, and correction parameters).

11.4 Concept and Design of an Adaptive Multi-Mode Low-Power DFE

In this section we describe the building blocks of the DFE-based receiver concept and highlight their reconfiguration capabilities. The line-up of the building blocks is depicted in Fig. 11.3. The purpose of the DFE is three-fold:

- Most important is the extraction of the wanted channel and suppression of interfering channels. Therefore the DFE consists of a cascade of different types of filters. Some filters include an integer sample rate decimation.
- This contributes to the second task of the DFE, which is the downsampling from the high sample rate delivered by the $\Delta\Sigma$ ADC to the bit-, chip-, or symbol rate

of any of the supported standards. Because the ADC is clocked at a fixed rate, this requires in some cases a fractional sample rate conversion (FSRC).

- Finally, the correction of impairments introduced by the analog front-end is an important feature of the DFE and makes the overall architectural approach increasingly attractive for future semiconductor technology nodes, as shrinking feature size usually degrades the analog performance of the transistors.

11.4.1 Cascaded-Integrator Comb Filters

Cascaded-integrator-comb (CIC) filters are frequently used for efficient sample rate decimation by an integer factor [18]. The decimation in our DFE is achieved by a series of two CIC filters, with the second CIC providing a configurable decimation rate adjustment with respect to the bit-, chip-, or symbol-rate of the received standard. The structure of a CIC filter is depicted in Fig. 11.4.

It consists of three parts. The first part is built from a series of N integrators operating at the input sample rate. It is followed by a decimator with the integer decimation factor R . The third part is composed of N comb elements, each of which comprise a delay of M samples and a subtraction operating on the decimated-by- R input sample rate.

The transfer function of an N -th order CIC is described by

$$H_{CIC}(z) = \left(\frac{1 - z^{-RM}}{1 - z^{-1}} \right)^N . \tag{11.2}$$

CIC filters are reconfigurable by adjusting the filter order N , the decimation factor R , and the number of delays M in the comb stage. For example, integrators and comb stages can be deactivated if they are not required. Due to the simplicity of the filter and the fact that the integrators and comb stages can be implemented entirely with adders, sample rate conversion (SRC) can be performed at minimum cost in silicon area and with minimum power consumption although the activity is high in the DFE, because the input sample rate is determined by the high OSR of the $\Delta\Sigma$ ADC.

A major drawback is the passband droop in the transfer function and the loss in anti-aliasing attenuation if the signal is decimated to twice or four times bit-, chip-, or symbol rate. In the context of the DFE, it is therefore necessary that additional integer and/or fractional decimation stages follow and that the passband droop is

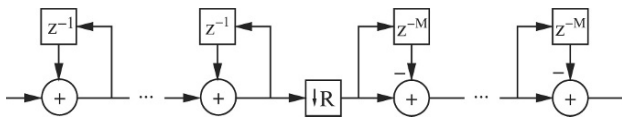


Fig. 11.4 Architecture of a CIC filter

corrected in a compensation stage (e.g. by an FIR filter). This will be described in the following sections.

The purpose of the cascade of the two CIC filters in our DFE is mainly the efficient sample-rate decimation [19]. The first CIC performs a fixed decimation by a factor of two, while the decimation factor of the second CIC is adaptable and depends on the actual standard to be processed. The use of two CIC filters greatly reduces the clock rate and therefore the current consumption of the second CIC.

11.4.2 Coordinate Rotation Digital Computer (CORDIC)

In the ZIF-architecture, a CORDIC is implemented in the DFE to compensate for fine frequency and phase shifts, whereas in the low-IF architecture, the CORDIC shown in Fig. 11.5 is used for final down-conversion to baseband and to achieve sufficient image rejection. A CORDIC is a simple and efficient algorithm to calculate hyperbolic and trigonometric functions. The algorithm can be run without any hardware multipliers; it requires only shift-and-add operations along with a lookup table. The CORDIC algorithm provides an iterative method of performing vector rotations by arbitrary angles [20]. During the n -th CORDIC iteration, the input vector (I_0, Q_0) is rotated by successively decreasing elementary rotations with the predefined basic angle

$$\alpha_n = \arctan(2^{-n}) \quad n = (0, 1, \dots, N - 1). \tag{11.3}$$

The resulting iterative process to accomplish all micro-rotations by simple shift-and-add operations is described by the following equations

$$I_{n+1} = I_n + d_n \cdot Q_n \cdot 2^{-n} \tag{11.4}$$

$$Q_{n+1} = Q_n + d_n \cdot I_n \cdot 2^{-n} \tag{11.5}$$

$$z_{n+1} = z_n - d_n \cdot \alpha_n \tag{11.6}$$

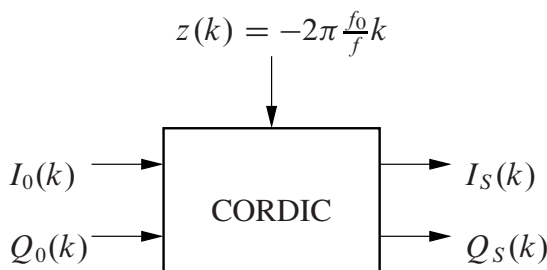


Fig. 11.5 CORDIC-based frequency correction

with

$$d_n = \begin{cases} -1 & \text{if } z_n < 0 \\ +1 & \text{otherwise} \end{cases} \quad (11.7)$$

being the direction for each rotation. The transfer characteristic of a CORDIC after N iterations results in

$$\begin{bmatrix} I_N(k) \\ Q_N(k) \end{bmatrix} = K \begin{bmatrix} I_0(k) \\ Q_0(k) \end{bmatrix} \begin{bmatrix} \cos(z_0(k)) & -\sin(z_0(k)) \\ \sin(z_0(k)) & \cos(z_0(k)) \end{bmatrix} \quad (11.8)$$

with $z_N \approx 0$ and the constant scaling factor

$$K = \prod_{n=0}^{N-1} \sqrt{1 + 2^{-2n}} \approx 1.6, \quad (11.9)$$

which is independent of the rotation angle z_n . Let f_0 be the frequency offset to be corrected, and f_s the sampling frequency of the CORDIC, then the complex-valued baseband signal $I(k) + jQ(k)$ is multiplied by $e^{-jz(k)}$ with

$$z_0(k) = -2\pi \frac{f_0}{f_s} k. \quad (11.10)$$

Figure 11.6 shows the digital image rejection of the DFE-based down-converter depending on the number of iterations and the simulated bit-widths. Apparently, the image rejection can be up to 100 dB when using 16 bits and at least 11 iterations. This digital solution clearly outperforms a conventional analog down-converter implementation.

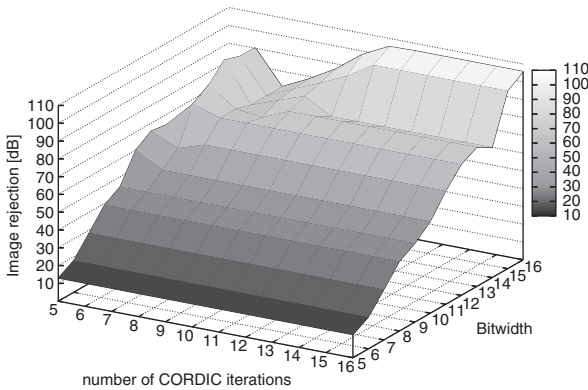


Fig. 11.6 Image rejection depending on the number of CORDIC iterations and the simulated bit widths

11.4.3 Infinite Impulse Response Filter

The channel selection filtering is done mainly by an infinite impulse response (IIR)-type lowpass filter. This filter offers a high stopband attenuation with an acceptably low filter order. Its adaptability is implemented by reconfiguration of coefficients and order to the respective standards' requirements. The transfer function $H_{\text{IIR}}(z)$ of the adaptive IIR filter of order 7 is given by

$$H_{\text{IIR}}(z) = \frac{a_0 + a_1z^{-1} + a_2z^{-2} + a_3z^{-3} + a_4z^{-4} + a_5z^{-5} + a_6z^{-6} + a_7z^{-7}}{1 + b_1z^{-1} + b_2z^{-2} + b_3z^{-3} + b_4z^{-4} + b_5z^{-5} + b_6z^{-6} + b_7z^{-7}} \quad (11.11)$$

$$= \left(\frac{\alpha_1 + \alpha_2z^{-1} + \alpha_2z^{-2} + \alpha_3z^{-3}}{\beta_1 + \beta_2z^{-1} + \beta_2z^{-2} + \beta_3z^{-3}} \right)^{x_0} \\ \left(\frac{\alpha_4 + \alpha_5z^{-1} + \alpha_6z^{-2}}{1 + \beta_5z^{-1} + \beta_6z^{-2}} \right)^{x_1} \left(\frac{\alpha_7 + \alpha_8z^{-1} + \alpha_9z^{-2}}{1 + \beta_8z^{-1} + \beta_9z^{-2}} \right)^{x_2} \quad (11.12)$$

where a_i , b_i are the coefficients of the fractional polynomial representation. For the implementation, the transfer function $H_{\text{IIR}}(z)$ is split into three parts with the coefficients α_i and β_i and with $x_i \in 0, 1$ which makes it possible to selectively adjust filter order and, as a consequence, selectivity.

If the IIR-type lowpass filter is built of a cascade of third- and second-order sections, an efficient implementation in terms of power consumption can be achieved. The benefit of high stopband attenuation of the IIR-type lowpass filter, required for the suppression of strong adjacent channel signals, comes at the cost of a non-constant group delay that introduces distortions to the wanted signal.

Figure 11.7 depicts the magnitude plots for the low-pass filter with the stop-band attenuation increased from 33 dB to 43 dB and 60 dB for third, fifth, and seventh order, respectively.

11.4.4 Allpass Filter

The nonlinear group delay introduced by the IIR and the analog anti-aliasing filter is compensated for in an allpass stage, whereby the allpass order correlates with the channel selection stage.

Similar to the IIR case, an efficient implementation of an I/Q multiplexed allpass comprises only one multiplier per lattice and allows for reducing the filter order by deactivating single allpass adaptor stages. Figure 11.8 depicts group delay of the IIR-type lowpass filter after equalization with the AP in the UMTS mode of our DFE implementation.

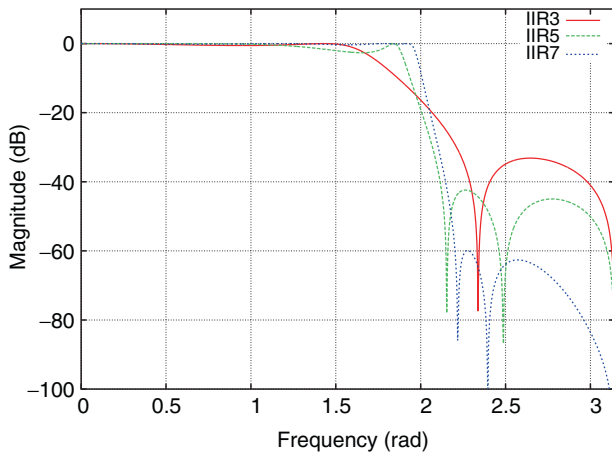


Fig. 11.7 Magnitude response of the low-pass IIR filter in third (red), fifth (green), and seventh (blue) order mode

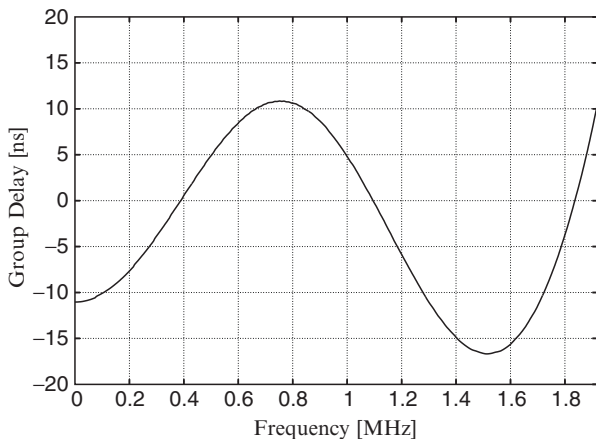


Fig. 11.8 Group delay of the lowpass IIR-type filter compensated by an allpass filter of order 2 in UMTS mode

11.4.5 DC Notch Filter

The AFE of ZIF receivers generates a DC offset that can heavily distort the wanted signal. This offset signal is mainly generated due to the leakage of the local oscillator signal of the mixer to its RF input and the resulting self-mixing [21]. Due to the required high gain provided for the baseband signal in a ZIF receiver, special care has to be taken in the analog circuit design to keep the DC offset values small. However, a certain offset value always remains before the digitization of the wanted signal. The DFE therefore contains a DC notch to eliminate the residual DC offset.

Analog gain-switching strategies are often applied in cellular systems to optimize the tradeoff between performance and power efficiency. This switching causes transients in the DC offset, which in turn can severely degrade the system performance. Therefore, the DC notch has to be designed in a way that it can filter out most of the DC offset, even in the presence of transients.

11.4.6 Finite Impulse Response Filter

The function of the FIR filter in the DFE of the multi-mode receiver is twofold. Its primary task is the matched filtering according to each communication standards specification (e.g. root-raised cosine for UMTS). Its secondary task is compensation of the passband ripple and droop that is induced by the preceding filter stages, i.e. the two CIC filters and the IIR-type lowpass filter. It also compensates for passband droop of the following fractional sample rate conversion stage.

The support of all standards requires full reconfigurability of the taps, where the maximum number of taps is chosen to satisfy all standards' pulse-filter lengths. For optimization of the filter taps, semi-definite programming can be employed [22]. Choosing a polyphase decomposition for the implementation of the FIR filter can significantly reduce the effective number of hardware resources needed. For instance, the number of multipliers for a symmetric FIR is half of the number of taps per channel. Furthermore, it is easily possible to bypass tap stages if not all FIR coefficients are required. This adapts the filter characteristic as needed and reduces power consumption.

11.4.7 Fractional Sample-Rate Converter

The decimation factor in the CIC and IIR filter stages can be integer only. The target sample rate at the output of the DFE is exactly two times or four times the bit-, chip-, or symbol-rate of the supported standards. Due to the fact that the ADC is clocked at a constant rate for all modes, a fractional sample-rate conversion (FSRC) stage is necessary. Possible sample rate partitionings for the DFE in terms of decimation factors and fractional conversion factors for the main investigated communication standards are listed in Table 11.2.

The exact bit-, chip-, or symbol-rate is obtained by placing an FSRC after channel filtering. This drastically reduces the demand on the FSRC internal image reject anti-aliasing filter, because no adjacent channel power can fold into the wanted signal band. Moreover, the magnitude of the transfer function of the FSRC is not required to be constant in the passband, because the preceding FIR filter is designed to compensate for this magnitude ripple.

If the oversampling ratio at the DFE input is high, which will be the case if a $\Delta\Sigma$ ADC is used, large integer decimation is processed in separate stages (CIC1

Table 11.2 Sample-rate partitioning for UMTS, CDMA2000, GSM/GPRS/EDGE (G/G/E), and GNSS for an output rate of twice the bit-, chip-, or symbol-rate of the respective standard at the DFE output

Parameter	UMTS	CDMA2000	G/G/E	GNSS
CIC1 dec	2	2	2	2
CIC2 dec	6	20	96	8
IIR dec	1	1	2	1
FSRC numer	288	1,536	—	1,024
FSRC denom	325	1,625	—	1,625

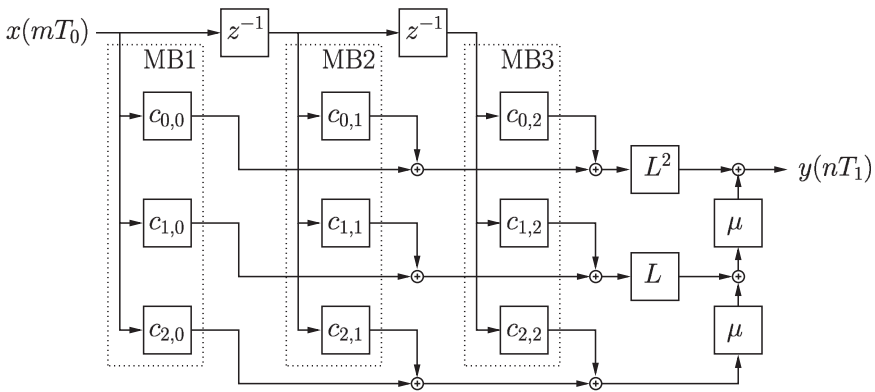


Fig. 11.9 Block diagram of the Farrow FSRC. The multiplications with the coefficient matrix c are combined into multiplier blocks (MB)

and CIC2). Thus it is possible to limit the range of the FSRC to between 1 and approximately 1.3. FSRC is employed to achieve exactly two-times bit-, chip-, or symbol-rate for the respective standards while the input sample rate is fixed to a single system clock rate for all modes.

The FSRC is based on the interpolating Farrow structure shown in Fig. 11.9, where dedicated polyphases are calculated to form an FRSC. The conversion factor is separated into an integer numerator L and an integer denominator M . The polyphase μ used in the FSRC multiplication stage selects (and thus calculates) only the necessary output samples, whereas the polyphase μ_n is defined by

$$\mu_n = (\mu_{n-1} + M) \bmod L. \tag{11.13}$$

With the depicted FSRC, an impulse response of three second-order polynomials is interpolated. When using a $\text{sinc}(x)^3$ as a model filter (see [5]), the coefficients can be implemented by fixed shift-and-add multiplications combined into multiplier blocks [23].

Table 11.3 Coefficients for the Farrow fractional sample-rate converter

Coefficient	Value
$c_{0,0}$	0
$c_{1,0}$	0
$c_{2,0}$	0.5
$c_{0,1}$	0.5
$c_{1,1}$	1
$c_{2,1}$	-1
$c_{0,2}$	0.5
$c_{1,2}$	-1
$c_{2,2}$	0.5

Because the FSRC is located after all channel-selection filter stages, an efficient filter with low order and complexity can be applied. Thus, the number of multiplications and filter length and thus filter complexity and hardware resources are kept low.

The viable transfer function $H(f)$ can be expressed by

$$H(f) = \left(\frac{\sin(f)}{f} \right)^3 \quad (11.14)$$

in the frequency domain, whereas a time-domain representation can be expressed exactly by following three second-order polynomials

$$p_0(x) = \frac{1}{2}x^2 = c_{2,0}x^2 \quad (11.15)$$

$$p_1(x) = \frac{1}{2}(-2x^2 + 2x + 1) = c_{0,1} + c_{1,1}x + c_{2,1}x^2 \quad (11.16)$$

$$p_2(x) = \frac{1}{2}(x^2 - 2x + 1) = c_{0,2} + c_{1,2}x + c_{2,2}x^2. \quad (11.17)$$

The coefficients $c_{x,y}$ of the Farrow structure are powers of two as shown in Table 11.3, and thus the multiplications are performed by shift operations in the hardware design at no cost, whereas the number of multiplications is 4 per channel that can be simplified to two distinct multiplier coefficients for L and μ , respectively.

11.4.8 ADC Leveling

For satisfactory performance of the ADC, the ADC input signal level must be adequate. In cellular systems, the received signal power can vary over several orders of magnitude, and the variation can occur fast. The effects of a signal power level that

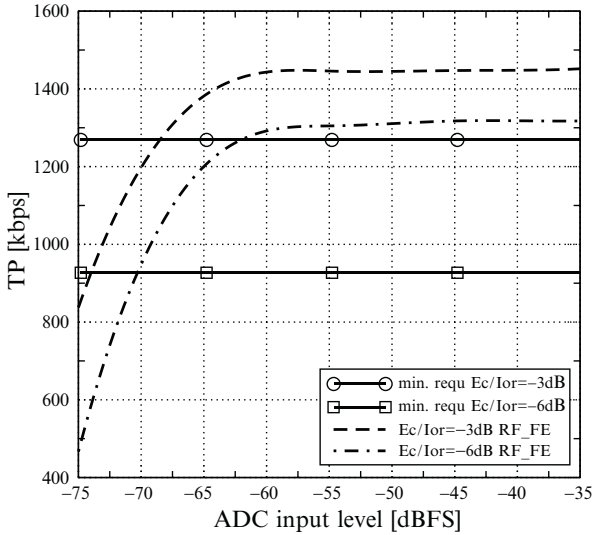


Fig. 11.10 Throughput performance for different signal power levels in front of the ADC (solid lines are minimum requirements from the HSDPA standard)

is too low at the ADC input are shown in Fig. 11.10, which displays the simulated High Speed Downlink Packet Access (HSDPA) throughput (TP) for a signal-to-noise level I_{or}/I_{oc} of 10 dB versus the signal power level in front of the ADC given in dB-Full Scale (dBFS). The solid lines show the minimum requirements according to [13]. It can be seen that a signal level of less than -65 dBFS in front of the ADC leads to a significant degradation of the TP. Thus a proper Automatic Gain Control (AGC) leveling mechanism must be applied.

The absolute input signal power level is important but so is the rate of change in high-speed scenarios. Figure 11.11 shows the measured power difference between consecutive slots for HSDPA data transmission over the ITU Vehicular A channel for a mobile speed of 120 km/h. The maximum gain step between two slots, which has to be leveled by the AGC, is in the range of about $+8$ dB and -8 dB.

Due to the use of the DFE, the gain calculation can be performed in a straightforward manner in the digital domain. Based on the calculated AGC word, the analog gain level is set as displayed in Fig. 11.12. The calculation of the signal power is carried out after the CIC filter during a predefined interval.

The result of a proper AGC strategy for HSDPA data transmission over the ITU Pedestrian A channel for a mobile speed of 3 km/h is shown in Fig. 11.13. The input power level at the antenna of the receiver and the ADC input power level are shown. The required amplification of 45 dB is carried out by the LNA, the mixer stage, and the analog filter, which are leveled between two fixed gain settings depending on the actual signal power. The fine leveling is carried out by the adjustable amplifier, and it can be seen that this strategy leads to an almost constant signal power level in front of the ADC.

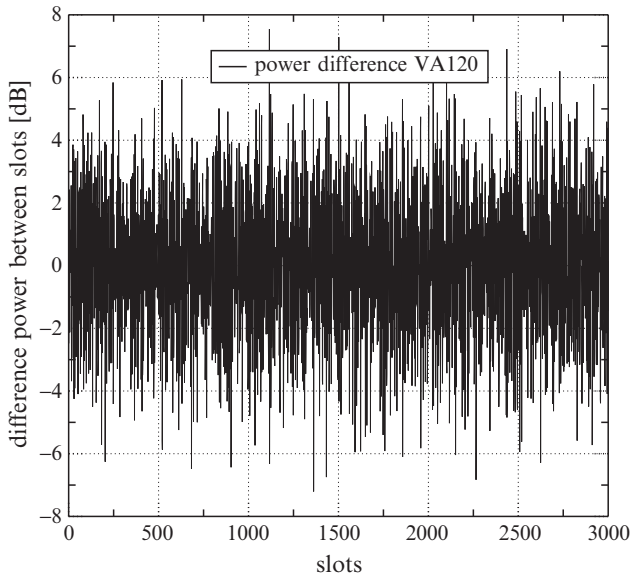


Fig. 11.11 Power difference between consecutive time slots for HSDPA data transmission over the ITU channel Vehicular A for a mobile speed of 120km/h

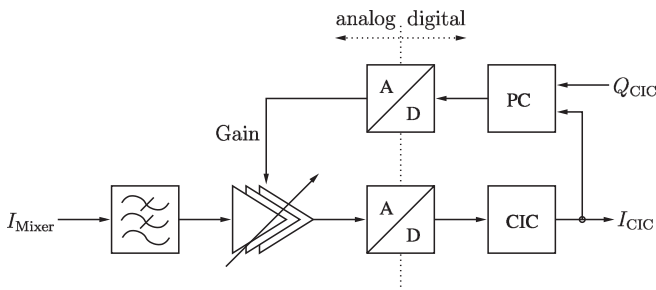


Fig. 11.12 DC offset and gain calculation in the DFE and compensation in the analog domain

11.4.9 Digital Interface

The digital interface to the BB-IC is a low-voltage differential serial digital connection for the transfer of the processed data samples. The interface allows a maximum data rate of 208 MBit/s with an arbitrary bitwidth of 4–16 bits per word. The digital interface is restricted by a maximum data rate to avoid cross-talk to sensitive analog parts of the chip, and the need to save cost by minimizing the number of pads of the chip package. This minimization is achieved by interleaving the serialized I and Q data into a single serial bitstream. These issues have been considered in the design,

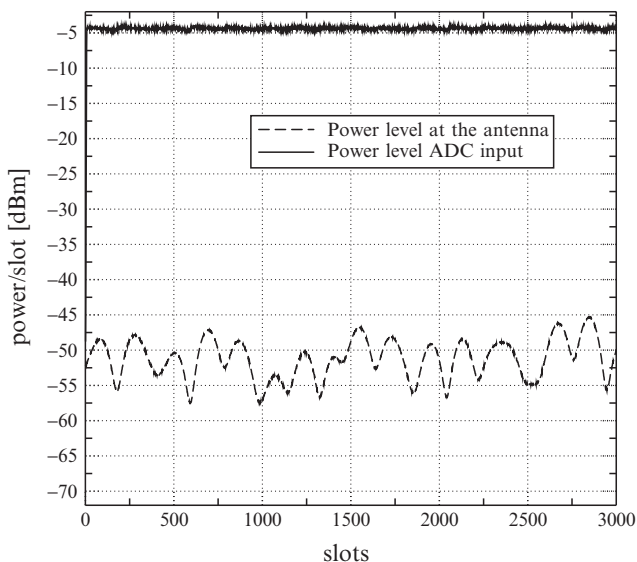


Fig. 11.13 Power level at the receiver antenna and AGC leveled ADC input level for HSDPA data transmission over the ITU Pedestrian A channel for a mobile speed of 3 km/h (power measured once per slot)

and measurements have not found any spurious or unwanted influence due to the high-rate data signal, while the bit-error rate (BER) of the interface exceeds 10^{-11} , a value far beyond the limits of the respective standards.

11.5 Simulation Results for the Digital Front-End

A simulation environment for the DFE architecture together with a first-order Butterworth analog filter and a $\Delta\Sigma$ ADC model has been implemented and used for evaluation of dedicated test cases for UMTS.

The first-order Butterworth anti-aliasing filter has a cut-off frequency of 1.5 MHz. The degradation of the wanted signal is compensated for in the DFE. The behavioral model of the ADC is a fourth-order continuous time $\Delta\Sigma$ ADC with a fourth-order feedback architecture and a sample rate of more than 100 MHz. This ADC has been adapted from [24] to improve attenuation in the adjacent channel frequency range.

For the time being, all modules of the DFE are implemented in floating point. The delivered signal fidelity after the DFE was checked by means of error vector magnitude (EVM). The simulations were performed for various DFE partitionings with or without the presence of a possible adjacent channel signal.

A summary of investigated DFE partitionings is listed in Table 11.4 with the filter orders for CIC2, IIR, AP, and FIR, the EVM in normal case and in ACS test case,

Table 11.4 Filter settings for different overall attenuation requirements for UMTS. N_{CIC} , N_{IIR} , N_{AP} , N_{FIR} are the filter orders for CIC, IIR, AP, and FIR, respectively. Furthermore, EVM with and without adjacent channel and the relative power consumption (P_{rel}) normalized to the most complex line-up are listed

N_{CIC}	N_{IIR}	N_{AP}	N_{FIR}	EVM	EVM _{ACS}	P_{rel}
2	-	-	15	4.8	13.6	0.38
2	-	-	17	4.5	13.6	0.41
2	-	-	21	3.5	13.3	0.46
4	-	-	15	4.7	7.2	0.44
4	-	-	17	4.7	7.9	0.47
4	-	-	21	3.5	5.3	0.52
4	3	2	21	4.4	4.9	0.67
4	3	2	25	4.2	4.7	0.72
4	3	2	29	4.0	4.6	0.77
4	5	2	25	3.3	4.2	0.78
4	5	2	29	3.4	4.0	0.84
4	5	2	33	3.2	3.9	0.89
4	5	2	37	3.1	3.7	0.95
4	5	2	41	2.9	3.6	1

and a relative power consumption normalized to the most complex filter line-up. The power consumption has been provided by power estimation on implemented and synthesized hardware. A reduction of the FIR and IIR filter order causes a degradation of EVM for cases with and without adjacent channel signal. In low-power modes, with the IIR and allpass filters bypassed, the EVM degrades abruptly in ACS test case. However, the published EVM of 9% for a current product [16] is by far exceeded for all configurations without an adjacent channel.

The receiver performance in terms of EVM versus the adjacent channel power ratio (ACPR), which is the ratio of the out-of-band power P_{ioc} over the wanted channel power P_{ior} , is depicted in Fig. 11.14 for some filter setups from Table 11.4. A very low EVM can be achieved with all given filter settings if no interferer is present. However, setups with low filter orders and bypassed IIR fail to achieve a sufficiently low EVM with increasing ACPR, especially in the ACS test case with an ACPR of 40 dB. Furthermore, an EVM limit of 5% has been defined, which is considered to give sufficient margin for a high modulation quality in a full receiver path. This requirement can be achieved by splitting the ACPR range into three segments and applying optimized DFE configurations for each of them. The switching is performed at an ACPR of 20 and 30 dB, respectively. Notably, the IIR filter does not need to be activated for ACPR less than 30 dB. The maximum modulation quality of 2.9% over almost the full range of ACPR is achieved with all filters active and using high filter orders; this would be necessary for high throughput HSDPA classes, for example. Assuming that the probability of very high interferer levels is low [12], it is very beneficial to obtain sufficient modulation quality with a simple and efficient setup of the DFE by a configuration of CIC2 order 2, bypassing IIR and allpass,

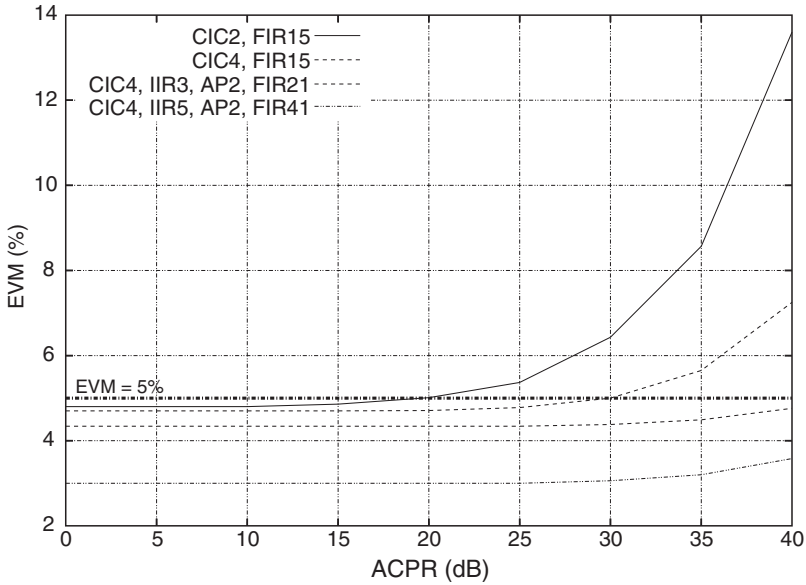


Fig. 11.14 Comparison of achievable EVM versus ACPR for several DFE setups. With a versatile filter-switching partitioning, the EVM stays below 5 %. The filter partition considered are CIC4_IIR5_AP2_FIR33 (fourth-order CIC, fifth-order IIR, second order allpass, and 41-tap FIR), CIC4_IIR3_AP2_FIR21 (fourth-order CIC, third-order IIR, second order allpass, and 21-tap FIR), CIC4_FIR21 (fourth-order CIC and 15-tap FIR), and CIC2_FIR15 (second-order CIC and 15-tap FIR)

and 15 taps in the FIR. Furthermore, increasing interferer levels are restrained by increasing the CIC2 order to 4. The highest adjacent channel power has to be attenuated by employing fifth-order IIR and second-order allpass stages. As a result, the power consumption is minimized in the scenario of little or no interferer while in the less probable cases of high interferer levels, the filter performance and thus power consumption are both increased.

Figure 11.15 depicts the DFE output spectra while varying the complexity of the DFE filter line-up. The simulations with bypassed IIR and AP filters exhibit a significantly larger spectral power in the adjacent channel due to the reduction of the adjacent channel attenuation. However it can be seen that without an adjacent channel signal, only a minor performance degradation in terms of EVM occurs but the power consumption of the setup can be significantly reduced by more than 62 % (see Table 11.4) compared to the line-up in [9], which draws 35 mW. The maximum inband amplitude ripple of the DFE together with ADC and lowpass filter varies between 0.09 and 0.52 dB for the given DFE configurations. At the same time the group delay is in the range of 2.74–31.21 ns. Notably, the group delay is worst when no allpass is active to mitigate contributions of the ADC and analog low-pass filter. Depending on the actual channel requirements, a significant power reduction is possible by adapting the DFE.

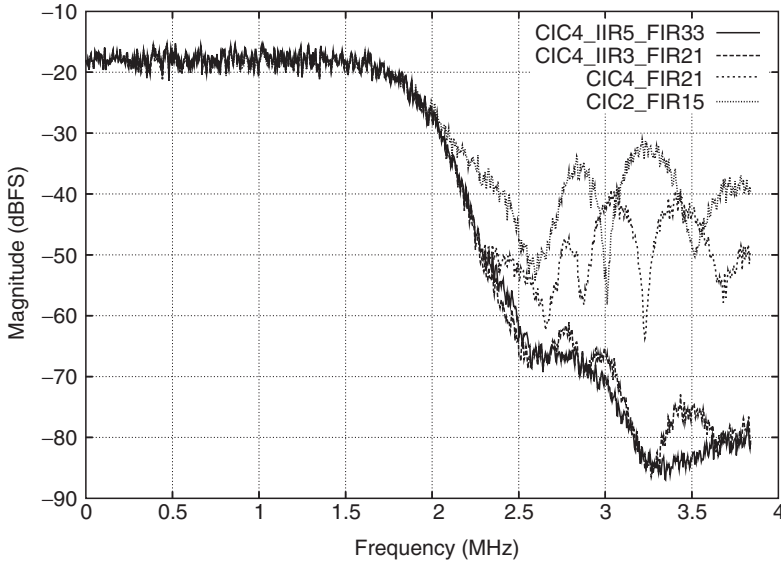


Fig. 11.15 Spectra of the DFE output signal for the ACS test case with adapted filter partitioning. The filter partition considered are CIC4_IIR5_FIR33 (fourth-order CIC, fifth-order IIR, and 33-tap FIR), CIC4_IIR3_FIR21 (fourth-order CIC, third-order IIR, and 29-tap FIR), CIC4_FIR21 (fourth-order CIC and 21-tap FIR), and CIC2_FIR15 (second-order CIC and 15-tap FIR)

In CDMA2000 mode, the demands for channel selection are increased due to the stringent blocker requirements. In order to deal with these scenarios, a seventh-order IIR and a third-order allpass stage are used. Consequently, adaptivity allows for the reduction of a total power of 19 mW in [9] by 58 % if the IIR and allpass stages are bypassed and CIC and FIR are reduced to order 2 and 15, respectively.

Furthermore, the satellite navigation system GNSS, with a bandwidth similar to UMTS, requires the partitioning derived from UMTS. Due to the very relaxed blocking requirements, a minimum configuration of a second-order CIC2, no IIR and allpass, and 15 taps in the FIR stage draws only 38 % of the power compared to the UMTS configuration.

It should be mentioned, that the detection of the adjacent channel power level (or any other relevant interference power level) is a key enabler for saving any power by means of adaptivity in the DFE [25, 26]. However, these concepts need further investigations with respect to suitability for chip-level integration.

11.6 Conclusion

This chapter has described the design of an adaptive digital front-end for multi-mode wireless receivers. Our approach is to use a generic direct-conversion analog front-end enhanced with a highly reconfigurable DFE to account for the major part of

the reconfigurability of the receiver. The main building blocks in the DFE are (fractional) sample-rate conversion and filtering. In advantageous spectral conditions (no adjacent channel present), the DFE's power consumption can be reduced by more than 60% due to adaptivity of filtering. As a result, our research work demonstrates the feasibility of multi-mode receivers in current semiconductor technologies for future cellular terminals.

Acknowledgment Part of this work was supported by the Linz Center of Competence in Mechatronics (LCM) within the framework of the Kplus program of the Austrian government and by the "Austrian Center of Competence in Mechatronics (ACCM)" in the context of the COMET Program of the Austrian federal government and the strategic program "Innovative Upper Austria 2010" of the Province of Upper Austria. The COMET Program is funded by the Austrian federal government, the Province of Upper Austria and the Scientific Partners of ACCM.

References

1. R. Bagheri, A. Mirzaei, S. Chehrizi, M. Heidari, M. Lee, M. Mikhemar, W. Tang, and A. Abidi, "An 800-MHz-6-GHz Software-Defined Wireless Receiver in 90-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 41, no. 12, pp. 2860–2876, Dec. 2006.
2. K. Lim, S. Lee, S. Min, S. Ock, M. Hwang, C. Lee, K. Kim, and S. Han, "A Fully Integrated Direct-Conversion Receiver for CDMA and GPS Applications," *IEEE J. Solid-State Cir.*, vol. 41, no. 11, pp. 2408–2416, Dec. 2006.
3. A. Tasić, W. Serdijn, and J. Long, "Adaptive Multi-Standard Circuits and Systems for Wireless Communications," *IEEE Circ. Syst. Mag.*, vol. 6, no. 1, pp. 29–37, 2006.
4. J. Craninckx, M. Liu, D. Hauspie, V. Giannini, T. Kim, J. Lee, M. Libois, B. Debaillie, C. Soens, M. Ingels, A. Baschiroto, J. V. Driessche, L. V. der Perre, and P. Vanbekbergen, "A Fully Reconfigurable Software-Defined Radio Transceiver in 0.13 μm CMOS," in *IEEE Proceedings of International Solid-State Circuits Conference (ISSCC 2007)*, Feb. 2007, pp. 346–348.
5. G. Hueber, L. Maurer, G. Strasser, R. Stuhlberger, K. Chabrak, and R. Hagelauer, "The Design of a Multi-Mode/Multi-System Capable Software Radio Receiver," in *Proceedings of IEEE Symposium on Circuits and Systems (ISCAS 2006)*, May 2006, pp. 3958–3961.
6. K. S. Yeung and S. C. Chan, "The Design and Multiplier-Less Realization of Software Radio Receivers With Reduced System Delay," *IEEE Trans. Circuits Syst.*, vol. 51, no. 12, pp. 2444–2459, Dec. 2004.
7. C. Martelli, R. Reutemann, C. Benkeser, and Q. Huang, "A 50mW HSDPA Baseband Receiver ASIC with Multimode Digital Front-End," in *IEEE Proceedings of International Solid-State Circuits Conference (ISSCC 2007)*, Feb. 2007, pp. 260–162.
8. A. Tasić, S. Lim, W. Serdijn, and J. Long, "Design of Adaptive Multimode RF Front-End Circuits," *IEEE J. Solid-State Circuits*, vol. 42, no. 2, pp. 313–322, Feb. 2007.
9. G. Hueber, L. Maurer, G. Strasser, K. Chabrak, R. Stuhlberger, and R. Hagelauer, "A GSM-EDGE/CDMA2000/UMTS Receiver IC for Cellular Terminals in 0.13 μm CMOS," in *Proceedings of European Conference on Wireless Technology 2006 (ECWT 2006)*, Sep. 2006, pp. 23–26.
10. R. Stuhlberger, L. Maurer, C. Wicpalek, E. Goehler, G. Heinrichs, J. Winkel, C. Drewes, G. Hueber, and A. Springer, "System Design of a Configurable Highly Digital UMTS/NAVSAT RF-Receiver," in *Proceedings of IEEE Vehicular Technology Conference (VTC 2006-Spring)*, May 2006, pp. 1787–1791.

11. R. Stuhlberger, L. Maurer, G. Hueber, and A. Springer, "The Impact of RF-Impairments and Automatic Gain Control on UMTS-HSDPA-Throughput Performance," in *Proceedings of IEEE Vehicular Technology Conference (VTC 2006-Fall)*, Sep. 2006, pp. 1–5.
12. 3GPP, *Technical Specification Group Radio Access Network: Radio Frequency (RF) system scenarios (Release 7)*, Mar. 2007.
13. 3GPP, *Technical Specification Group Radio Access Network: User Equipment (UE) radio transmission and reception (FDD) (3GPP TS 25.101 Release 6.9.0)*.
14. 3GPP2, *Recommended Minimum Performance Standards for cdma2000 Spread Spectrum Mobile Stations*, Jan. 2005, no. 3GPP2 C.S0011-C.
15. A. Springer, L. Maurer, and R. Weigel, "RF System Concepts for Highly Integrated RFICs for W-CDMA Mobile Radio Terminals (invited)," *IEEE Trans. on Microwave Theory and Tech.*, vol. 50, no. 1/II, pp. 254–267, Jan. 2002.
16. R. Koller, T. Rühlicke, D. Pimingsdorfer, and B. Adler, "A Single-Chip 0.13 μm CMOS UMTS W-CDMA Multi-band Transceiver," in *IEEE Radio Frequency Integrated Circuits Symposium (RFIC 2006)*, June 2006.
17. M. Valkama, M. Renfors, and V. Koivunen, "Blind signal estimation in conjugate signal models with applications to I/Q imbalance compensation," *IEEE Signal Processing Lett.*, vol. 12, pp. 733–736, Nov. 2005.
18. T. Hentschel, *Sample Rate Conversion in Software Configurable Radios*. Boston, MA/London: Artech House Publishers, 2002.
19. H. Aboushady, Y. Dumonteix, M.-M. Louerat, and H. Mehrez, "Efficient polyphase decomposition of comb decimation filters in $\Sigma\Delta$ analog-to-digital converters," *IEEE Trans. Circuits Syst.*, vol. 48, no. 10, pp. 898–903, Oct. 2001.
20. J. Volder, "The CORDIC trigonometric computing technique," *IRE Trans. Electron. Comput.*, vol. EC-8, pp. 330–334, 1959.
21. B. Razavi, *RF Microelectronics*. Prentice Hall, 1998.
22. W.-S. Lu, "Design of nonlinear-phase FIR digital filters: a semidefinite programming approach," in *Proceeding of IEEE Circuits and Systems (ISCAS '99)*, Aug. 1999.
23. A. Dempster and M. Macleod, "User of Minimum-Adder Multiplier," *IEEE Trans. Circuits Syst. II*, vol. 42, no. 9, pp. 569–577, Sep. 1995.
24. L. Dörrer, F. Kutter, P. Greco, P. Torta, and T. Hartig, "A 3-mW 74-dB SNR 2-MHz continuous-time delta-sigma ADC with a tracking ADC quantizer in 0.13- μm CMOS," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2416–2427, Dec. 2005.
25. A. Mayer, L. Maurer, G. Hueber, T. Dellsperger, T. Christen, T. Burger, and Z. Chen, "RF Front-End Architecture for Cognitive Radios," in *IEEE Symposium on Personal Indoor Mobile Radio Communications (PIMRC 2007)*, Oct. 2007.
26. J. Groe, "Adaptive receiver system that adjusts to the level of interfering signals, patent no. us 6980786b)," Patent, Jan. 2002.

Chapter 12

FEC Decoders for Future Wireless Devices: Scalability Issues and Multi-standard Capabilities

John Dielissen, Nur Engin, Sergei Sawitzki, and Kees van Berkel

12.1 Introduction

During the last decade we have witnessed a proliferation of transmission standards for wireless communication. This holds for cellular communication, but also for broadcast and connectivity standards. All these transmission standards employ an FEC coding scheme: by adding some redundancy and appropriate coding of a signal, high transmission quality can be achieved even over noisy channels. Here transmission quality is typically measured in terms of error rates per bit or per packet. For many of these standards there also has been a rapid succession of different generations, typically driven by a quest for ever higher bit rates, a more efficient spectrum utilization, and features. These improved spectrum utilization and dramatically higher bit rates have led to an explosion in required GOPS (Giga Operations per Second), adequately supported by Moore's Law (see Fig. 12.1). As a result, the overall silicon footprint of a state-of-the-art FEC decoder for consumer products measured in square millimeter remained more or less constant over time: from 1 mm² for a simple decoder up to 10 mm² for demanding standards. Fortunately, FEC codes most relevant for wireless communication standards are based on a few code families only: Reed-Solomon, convolution codes (including Turbo codes), and low-density parity codes (LDPC). Some standards apply multiple codes, either in a

J. Dielissen and N. Engin
NXP Semiconductors, Eindhoven, The Netherlands
e-mail: {John.Dielissen; Nur.Engin}@nxp.com

S. Sawitzki (✉)
FH Wedel, University of Applied Sciences, Wedel, Germany
e-mail: Sergei.Sawitzki@gmail.com

K. van Berkel
ST-Ericsson, Eindhoven, The Netherlands
e-mail: kees.van.berkel@nxp.com

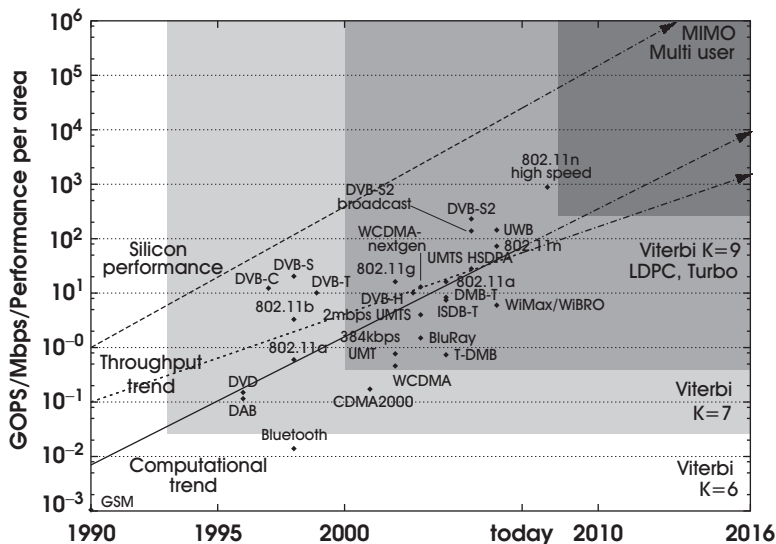


Fig. 12.1 Growing Complexity of the FEC Coding. Dots representing different standards are referring to computational requirements in GOPS. The three lines represent trends in silicon performance (gates/mm^2 , dashed line), throughput (Mbps, dotted line) and computational complexity (GOPS, solid line)

cascaded/concatenated manner or by specifying a particular code family for different logical channels.

Another important trend is set by an ever increasing number of different transmission standards a single user terminal is capable to receive and process. This triggers the need for so-called multi-standard FEC decoders. When the standards belong to the same code family, a solution can be based on parameterization of the decoder hardware. Parameters may include block size, coding rates, constraint lengths, polynomials, and so on. When the standards belong to different code families, it becomes tempting to look for area savings by sharing hardware resources, as will be explored in this chapter. Although channel decoding algorithms can be implemented quite efficiently on programmable DSPs, such software implementations become prohibitively power inefficient for data rates of tens of Mbps and beyond. For dedicated hardware implementations, the required power consumption remains challenging. Until recent years, straightforward CMOS scaling took care of increasing bit rates and increasing algorithmic complexity. This trend, however, has significantly slowed down. As supply voltages and parasitic capacitances will hardly decrease in the coming years, algorithmic and architectural innovations are now required to keep power consumption levels within tight budgets.

Multi-standard FEC decoders will become key components for a variety of product categories, including mobile phones, portable entertainment, digital television, set-top boxes, and PCs. In this overview chapter we discuss multi-standard FEC

decoding from a VLSI implementation perspective. We assume that the reader has some basic knowledge of the corresponding FEC decoder structure, thus, traditional implementation of the decoders will not be discussed herein. An interested reader may refer to [1–4] for reference designs. In this chapter all area figures are given for standard cell 90 nm CMOS technology.

12.2 Decoders Supporting a Single Code Family

12.2.1 Reed-Solomon

12.2.1.1 General Description

RS codes are based on polynomial division and Galois-field (GF) arithmetic. The code word is generated by appending the parity check symbols to the message to be encoded. The parity check symbols are the coefficients of the remainder from dividing the message by the generator polynomial $g(x)$ which is constructed based on the primitive element α of $GF(2^m)$. The decoding process consists of four main steps (Fig. 12.2). Upon receiving the encoded message, the parity symbols are checked through syndrome computation.

If all syndromes are zero, the code word is correct and the processing is finished. Otherwise the processing is passed to the key equation solver and Chien search and Forney error computation steps. The most computationally intensive step of the decoding algorithm is the key equation solver, for which there exist a number of implementations. For codes with 8 bit and larger symbols, the dual-line equation solver is the state of the art offering a good trade-off in terms of area vs. throughput [1], especially for high throughput decoders. The basic operations involved are finite field addition and (constant) finite field multiplication.

The dual-line architecture is a modification of the original Berlekamp-Massey algorithm [5] for the computation of the error location and evaluation polynomials. The same hardware is reused to compute first the error location polynomial and then the error evaluation polynomial, achieving both comparatively low area cost at low latency. Chien search and Forney algorithms are computing the exact location of every error and its value based on location and evaluation polynomials from the previous step. These algorithms are by far less complex than the key equation

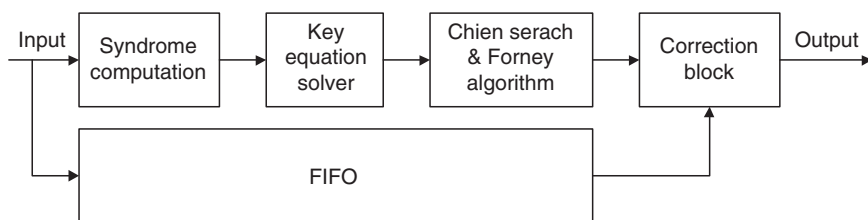


Fig. 12.2 Reed-Solomon decoder architecture view

solver (at least for the symbol sizes of 8 bits and above). Finally, having the error information in place, the symbols can be corrected by the correction block (finite field addition).

12.2.1.2 Scalability

From the scalability point of view, RS-codes do not introduce a major burden, since the decoding hardware is working not on single bits but on symbols. For symbol sizes of 8 bit a CMOS 90 nm solution clocked at more than 200 MHz is easily achievable, resulting in more than 1.6 Gbps effective throughput [6]. Such a decoder can be realized using less than 0.12 mm^2 and dissipating less than 10 mW when operating at 1 Gbps data rate ($\approx 0.01 \text{ nJ/Bit}$) as shown in [7]. Due to their specific properties, RS codes are frequently used as outer codes in combination with convolutional or other code families (for example in UWB, DVB-T or DMB-T standards). Note, that the maximum achievable clock speed (and throughput) of an RS decoder decreases with the growing symbol size. For wireless communication and connectivity standards it is not an issue, since only RS codes with symbols of 8 or less bits are used. Some optical storage formats, however, work with larger symbol sizes.

12.2.1.3 Multi-Standard

Decoding different RS codes on the same hardware (as long as the symbol sizes of these codes remain the same) is possible. It implies, however, that the set of constants for the finite field multipliers increases (each additional standard = new set of constants). The key equation solver block is not affected by this change, since it uses variable-to-variable multipliers anyway, but both syndrome computation and Chien search blocks will suffer from increasing multiplication complexity (variable–variable multipliers instead of fixed–variable multipliers). Still, the overall area overhead for multi-standard solution is reasonable, since syndrome computation and Forney algorithm implementation are much smaller than key equation solver and FIFO blocks. In addition, if the throughput requirements are low to mid range ($\leq 1 \text{ Gbps}$), semi-serial solution can be used for these blocks, so the overall area cost stays manageable.

Concerning codes with different symbol sizes, it is also possible to reuse the hardware used to decode RS-codes with larger symbol size for decoding the codes with smaller symbol size (but not vice versa).

12.2.2 Viterbi

12.2.2.1 General Description

The traditional implementation of the Viterbi decoding algorithm consists of three stages: branch metrics (BM) computation, path metrics (PM) computation and trace

back (TB), see Fig. 12.3. In the branch metric computation step, the Manhattan distance or any other appropriate difference measure between the ideal symbols (whose number depends on the coding rate) and the received symbol is computed. In the path metrics computation step several add–compare–select (ACS) operations have to be performed. For the convolutional code with constraint length K , 2^{K-1} ACS operations need to be computed per user bit. Especially for larger values of K , ACS operations are responsible for up to 90% of the computational load.

The result of the ACS operation is fed back as input to other ACS units. The decision bit as the result of the comparison is stored to be used in the last step of the algorithm called trace-back. During the trace back step, 2^{K-1} decision bits are produced by the ACS units as MSB of subtraction operation (which is the equivalent of comparison) and used to decode the original values of the user bits. The easiest way to implement the trace back is to map the state-transition diagram of the code directly to hardware as registers connected via multiplexers. This implementation technique is called “register exchange” trace back and is characterized for quite high area and power consumption especially for larger values of K . Due to this fact, trace back is quite often implemented with embedded memory blocks instead of registers (multiplexer logic is replaced by address generation unit in this case).

The estimated area cost of different Viterbi decoders is shown in Table 12.1, area/throughput dependency is summarized in Fig. 12.4. Note, that although both

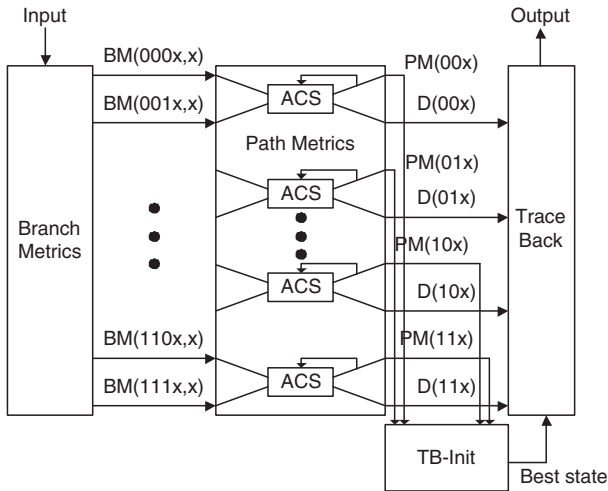


Fig. 12.3 Viterbi decoder architecture view

Table 12.1 Estimated area of Viterbi decoder

Standard	Branch metrics, mm ²	Path metrics, mm ²	Trace back, mm ²	Total
UWB	0.001	0.040	0.092	0.133
GSM	Neglectable	0.004	0.02	0.024
802.11a/g	0.001	0.015	0.092	0.108

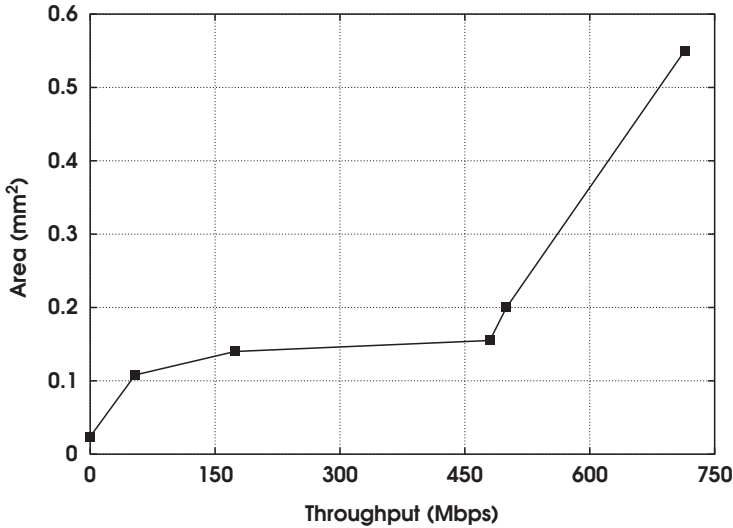


Fig. 12.4 Estimated area/throughput relation for Viterbi decoders at $K = 7$

UWB and WLAN Viterbi decoders have the same number of ACS units, the area of the path metric units differs by a factor of almost 2.5. This is partially due to the fact, that WLAN requires a much lower throughput (decoder is clocked at lower frequency). In addition we can think of a scenario, where 32 or 16 ACS units suitable for UWB decoding clocked at high speed are shared in time to process 64 states for WLAN, so the area will be $\frac{1}{2}$ or $\frac{1}{4}$ plus some control overhead to allow time sharing. In this case two or four cycles are consumed respectively to process one output sample.

Depending on the throughput requirements, an average energy consumption for a $K = 7$ Viterbi decoder is estimated at 0.4–0.8 nJ/Bit (decreasing for smaller values of K). Figure 12.5 shows the estimated power consumption for some Viterbi decoders.

12.2.2.2 Scalability

The major challenge in design of a Viterbi decoder is to provide the required throughput. Since it is possible to execute all 2^{K-1} operations per bit in parallel the main efforts are spent on optimization of the ACS operation itself [2]. Using CMOS 90 nm standard cell technology decoders for $K = 7$ can provide up to 700 Mbps output data rate, which is good enough for the current and next generation standards [8]. For the standards with independent code blocks and/or relaxed latency constraints, i.e. all broadcast standards, it is possible to achieve even higher data rates through coarse grain parallelization (several decoders running in parallel each one processing its own block of data) although at linearly growing area cost [9].

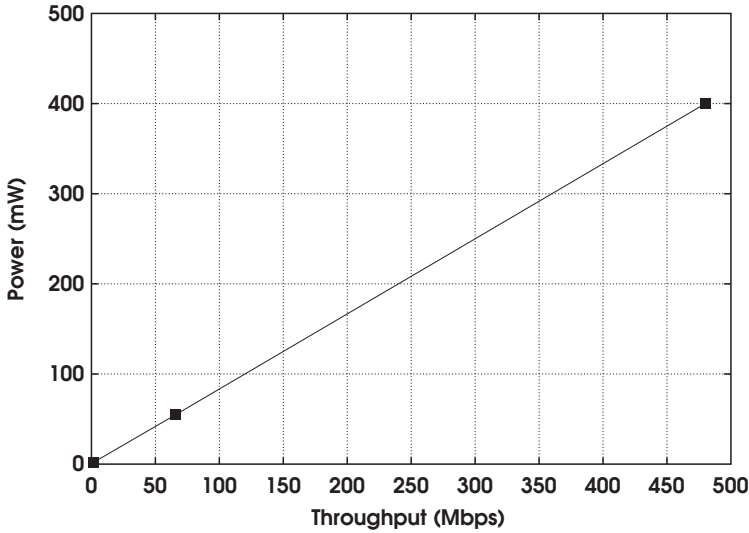


Fig. 12.5 Estimated power/throughput relation for Viterbi decoders at $K = 7$

12.2.2.3 Multi-Standard

For decoding of the codes with the same constraint length but different sets of generator polynomials only the interconnect structure between the branch metric unit and path metric unit needs to be changed. A code with smaller K can be decoded on the hardware for the code with larger K if the feedback interconnect between the ACS units is made flexible. Indeed, if this interconnect introduces enough flexibility (e.g. partially populated cross bar), a decoder for constraint length K can be used to decode in parallel 2^m streams encoded with $K - m$ code and having roughly the same bit rate. Alternatively, the same decoder can be used to decode a stream encoded with $K + m$ code and having roughly the bit rate of $\frac{1}{2^m} \times \text{Bit_rate_at_}K$. Introducing reconfigurable interconnect reduces the overall clock frequency the decoder can be run at, but if the number of standards/streams is limited to 2–4, this reduction should be less than 10–15%.

In general Viterbi decoders used to decode convolutional codes can easily meet multi-standard and multi-channel requirements at relatively little additional cost. The scalability is given for the most current real-world applications.

12.2.3 Turbo

12.2.3.1 General Description

The main parameters of a Turbo decoder are the code block size (determined by the size of the interleaver), code rate and the constraint size corresponding to the

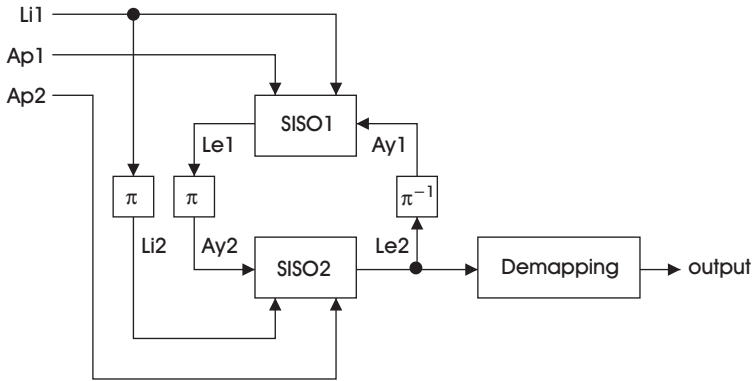


Fig. 12.6 Turbo decoder functional view

convolutional code decoded by the SISO blocks). A Turbo decoder typically consists of two main building blocks, a Soft-Input-Soft-Output (SISO) decoder and an interleaver. Figure 12.6 shows a functional view of a Turbo decoder. The decoding process includes several iterations; during each iteration SISO decoding is applied once to linearly ordered information words and once to interleaved information words. During each of these half iterations, systematic and parity information (Li and Ap) as well as a-priori information from previous iteration (Ay) are used to produce the extrinsic information word Le . When the predefined number of iterations is reached or convergence is detected, the iterations are stopped and the soft outputs are demapped into output bits. A more thorough treatment of Turbo coding can be found in [27] and [10] and more on Turbo decoder implementation aspects can be found in [3] and [11–13].

Turbo decoding is very computationally intensive. For one decoded bit up to around 1,500 operations are performed. As the bit rate supported by the wireless standards increases, higher throughput Turbo decoders are needed. For example, the UMTS HSDPA (High-Speed Downlink Packet Access) channels require 14 Mbps user bits throughput and in the UMTS LTE (Long-Term Evolution) data rates as high as 100 Mbps are envisioned. For such mobile applications the cost and power consumption must be kept as low as possible even at these rates. For this reason, the design of new Turbo decoder architecture concepts reaching high data rates at low cost and power consumption emerges as a major challenge.

In general, increasing the throughput of a digital signal processing design is possible through increasing the clock rate or adding parallelism. In case of Turbo codes, increasing the clock rates alone will not be sufficient: for a conventional Turbo decoder architecture, 100 Mbps throughput requires a clock rate of around 1.8 GHz. As Turbo codes are block-based, increasing the decoding throughput with fixed clock rate translates to either employing multiple decoders or speeding up the inner loop of the Turbo decoder. The first solution, although a fairly scalable one, means that the decoder area increases linearly with increasing throughput. The remaining option is speeding up the inner Turbo decoding loop without increasing clock speed,

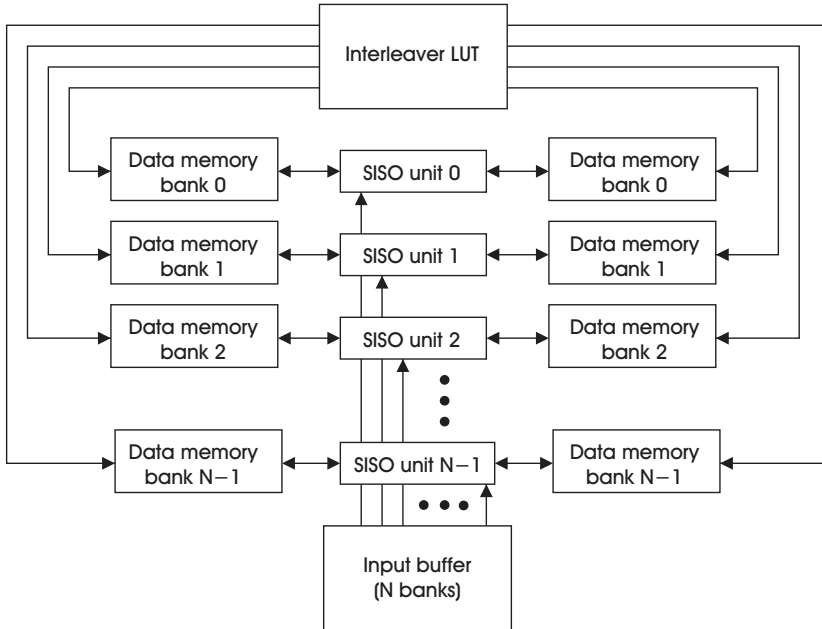


Fig. 12.7 Architecture view of a parallel-SISO Turbo decoder ($N =$ parallelism factor)

thus by adding parallelism. A parallel Turbo decoder block diagram is given in Fig. 12.7. In the coming sections, the scalability and multi-standard decoding for both the conventional and parallel architectures will be discussed.

12.2.3.2 Scalability

The parallel architecture of Fig. 12.7 is enabled by the idea that it is possible to parallelize the block interleaving, and therefore it is only applicable for Turbo decoders where block interleaving is specified in the standard. For introducing parallelism, a data memory structure with multiple banks is needed. By making the number of banks equal to the number of interleaving rows, it is possible to prevent memory conflicts even when the memory is accessed in an interleaved fashion. The same idea can be expanded to smaller number of banks taking a divider of the number of interleaver rows. Because pruning is included in block interleaving schemes, however, it is not possible to prevent bank conflicts altogether and a small number of bank conflicts still occur. It has been shown with simulations using UMTS Turbo interleaving scheme that it is possible to limit the bank conflicts to 15% of the overall number of accesses (with the average case around 10%). Thus, a throughput gain very close to N times can be achieved using this type of parallel interleaving. Due to the parallelization of the interleaver, the throughput limit is reached just below $N \times f_{clk} / N_{iter}$ (in Mbps), where N is the parallelism factor, f_{clk} is the clock frequency and N_{iter}

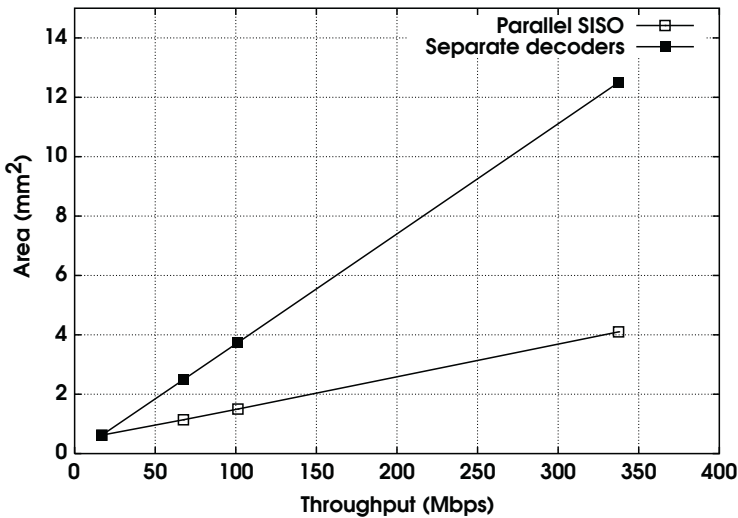


Fig. 12.8 Estimated area/throughput comparison of separate Turbo decoders with parallel-SISO Turbo decoder architectures

is the number of iterations. Other approaches to parallel interleaving have also been presented in [14–16].

With parallel architecture (compared to the conventional Turbo decoder) each memory unit is replaced by multiple memory banks. Although the total memory capacity remains the same, there is an area increase due to the inefficiency of using smaller memories. This area increase, however, is less than the factor N , which would be the case if separate decoders were used. Due to this difference the estimated area costs of parallel-SISO Turbo compared to the several separate Turbo decoders (providing the same throughput) is lower, as Fig. 12.8 shows.

The difference in the estimated power consumption for both architectures is represented in Fig. 12.9. Both area and power comparisons are based on estimated area/power figures obtained by scaling from a reference Turbo decoder design for UMTS [3]. However, the introduced technique is applicable in the same way to other decoders as well. As can be seen from the examples, throughput scales linearly with area and power given a certain code block size. The estimated area costs of Turbo decoders for various standards are shown in Table 12.2. In case of area figures, the parallel SISO has a clear area advantage compared to separate decoders. When power is concerned, the difference is much smaller, as the energy consumption per bit is around 2 nJ/bit for all cases.

12.2.3.3 Multi-Standard

Concerning multi-standard support for SISO, a reasoning similar to that of Viterbi applies: the support of different polynomials requires a more complex branch metric

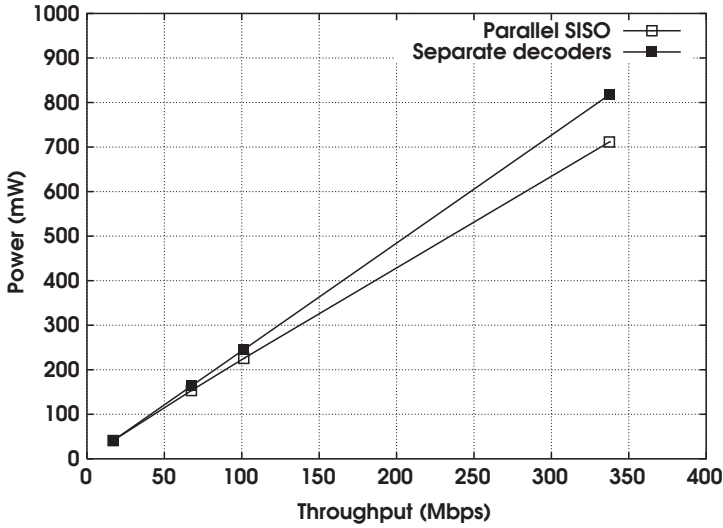


Fig. 12.9 Estimated power/throughput comparison of separate Turbo decoders with parallel-SISO Turbo decoder architectures

Table 12.2 Estimated area of Turbo decoder

Standard	SISO mm ²	Data mem mm ²	Input buffer mm ²	LUT mm ²	Total incl. 0.1 mm ² overhead
UMTS (R'99)	0.14	0.16	0.13	0.14	0.67
UMTS HSDPA	0.14	0.16	0.13	0.14	0.67
UMTS LTE	0.81	0.24	0.15	0.16	1.46
CDMA2000	0.14	0.64	0.52	0.56	1.95
802.16e	0.24	0.29	0.12	0.13	0.90

unit. However, this overhead will be relatively small, as the datapath accounts for less than 20% of the complete decoder area for a single-SISO Turbo decoder. The interleaver is less simple to adapt. In general, each standard has a unique Turbo interleaver, and a large part of the address generating hardware can not be reused. An extra address generation unit results in an increase of up to 20% of the decoder area, depending on the complexity of the interleaving scheme.

12.2.4 LDPC

12.2.4.1 General Description

With LDPC coding, the structure of the code is specified by the so-called *H*-matrix. The *H*-matrix describes the parity check equations which must be obeyed to obtain

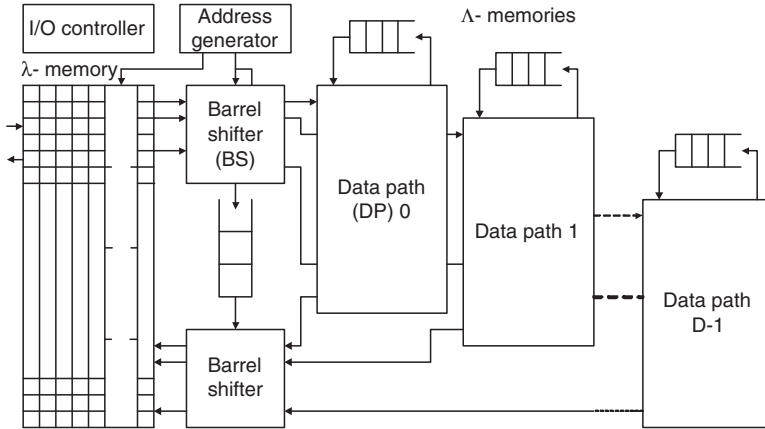


Fig. 12.10 LDPC decoder architecture view, showing the soft input memories (λ), checknode message memories (Λ), and the parallel data paths

a correct code word. Each row in the matrix is an equation, and each “1” denotes the participation of the corresponding bit in that equation. K stands for the number of participants in an equation. J is the number of equations in which a bit participates. All currently defined communication standards use quasi-cyclic LDPC codes. The H -matrix of a quasi-cyclic code consists of circulants of a predetermined size z , and each circulant contains either the identity matrix rotated over a certain angle or the all-zero matrix. Since all inputs related to a circulant are used only once in the related group of equations, it makes sense to process this group in parallel. It is then possible to organize the inputs such that they are stored in a single row in the memory. In [4] it is shown that it is possible to take sub-sets of this group, preserving the structure while scaling the performance.

LDPC decoding essentially means solving a large set of parity check equations, using soft bit information. By extracting information from one equation, another equation can be solved more accurately which, again, can be used to improve the solution of the first equation resulting in an iterative decoding algorithm. Detailed information on LDPC codes can be found in [17–19], on decoding in [20–22], and on architectures for decoding in [4] and [23–25].

This chapter uses the architecture template shown in Fig. 12.10 as reference, whose detailed explanation can be found in [4]. This architecture processes quasi-cyclic LDPC codes specified in DVB-S2, DMB-T, 802.11n, and 802.16e standards. Table 12.3 shows the post-synthesis area of the corresponding LDPC decoders.

12.2.4.2 Scalability

Scalability of the LDPC-decoder stretches over multiple dimensions. First of all, it regards the LDPC-code itself. Whereas in Turbo-codes the rate of the code is

Table 12.3 Estimated area of LDPC decoder. Architectural components can be found in Fig. 12.10

Standard	λ -mem mm ²	Λ -mem mm ²	DP mm ²	BS mm ²	Total incl. 0.1 mm ² overhead	
DVB-S2	2.0	0.40	0.70	0.50	0.10	3.8
2×DVB-S2	2.1	0.50	0.70	1.00	0.30	4.7
DMB-T	0.4	0.07	0.12	0.25	0.10	1.1
802.11n	0.3	0.04	0.09	0.70	0.10	1.3
802.16e	0.1	0.03	0.04	0.03	0.01	0.3

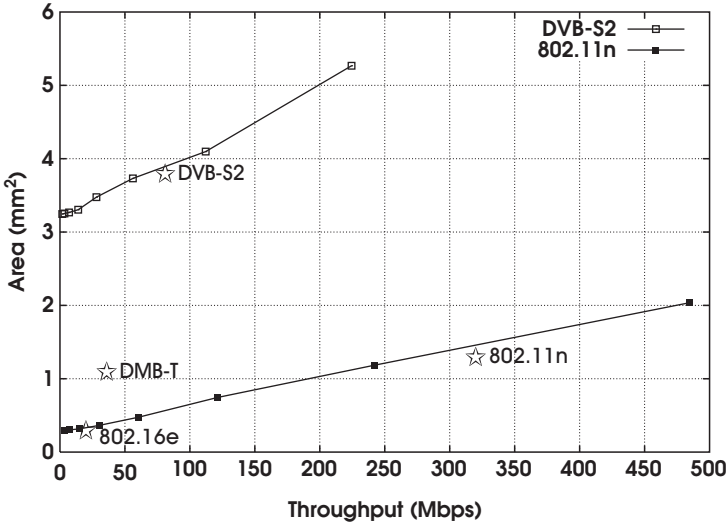


Fig. 12.11 Estimated area of LDPC decoder

determined by puncturing, and has no influence on decoding, in LDPC different rates lead to different codes. So, as a requirement the architecture must support multiple codes, even for a single standard. The second dimension of scaling is the code word length. It can be supported by changing the circulant size or by introducing new codes. In both cases the scalability is similar to multi-standard capability discussed below. The third and most important scaling dimension is the throughput. By adding more data paths, the throughput of the decoder scales linearly. However, the area of the decoder scales super linearly. The area scales with three main contributors: data paths, barrel shifters and memories. As the parallelism factor increases, the area of the data paths scales linearly (0.007 mm² per data path). The area cost of the memories is determined by the storage capacity, and thus by the targeted code word length. Note, however, that both narrow (=small word width) and shallow (=small address space) memories suffer from efficiency loss. The barrel shifter scales quadratically with the parallelism factor. Figure 12.11 shows the area of an LDPC decoder as a function of the maximum achievable throughput. Contrary to

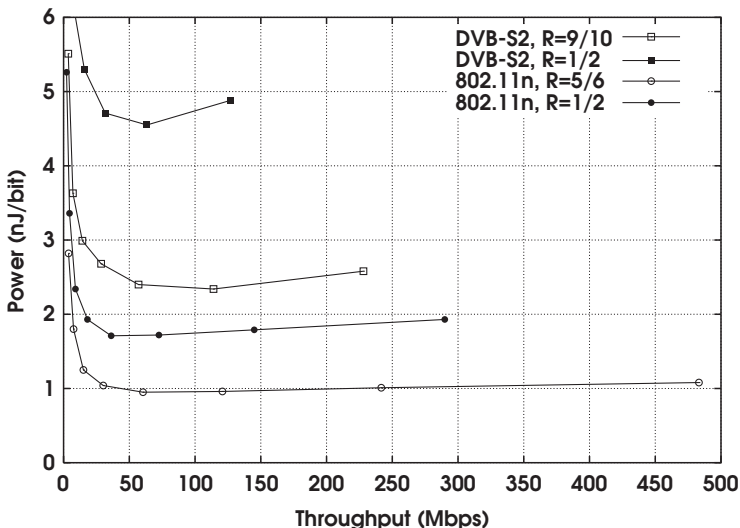


Fig. 12.12 Estimated power efficiency of LDPC decoding for different standards and different coding rates R . Throughput is in uncoded (user) bits

RS, Viterbi, and Turbo, the implementation cost of LDPC depends on the uncoded bit rate, while for the other decoders the cost depends on the coded bit rate. For the presented area consumption, the highest rate, and thus the highest throughput are taken. For DVB-S2 and 802.11n codes the decoder area is shown as continuous line drawn through a set of data points (parallelism factors $2^x, 0 \leq x \leq 7$). Note that these points do not represent valid designs, but are introduced purely for illustrative purposes. Valid design points for the existing standards are marked with stars.

The power efficiency of an LDPC decoder is shown as a function of the user bit rate in Fig. 12.12. The efficiency is determined by two factors: the code and the required throughput, which translates to the required parallelism factors. In general, larger code word lengths lead to more iterations. The workload per iteration scales linearly with the number of equations in which nodes participate. This results in varying number of computations and varying power consumption per standard. Furthermore, the number of computations per code word does not depend on the coding rate R . As a result, the power consumption per uncoded bit is approximately linearly proportional to the code rate. As mentioned, the power also scales with the parallelism factor. The power is a sum of a fixed control overhead, and a power consumption per data path. For low parallelism factors, this control overhead is dominant, and combined with narrow (inefficient) memories this leads to high power consumption. For high parallelism factors, both the quadratical scaling of barrel shifters' power consumption and the shallow (inefficient) memories become more significant.

The maximal parallelism factor is limited to the circulant size inside the code. When this number is exceeded, data cannot be read from one memory line anymore,

thus sophisticated scheduling techniques need to be introduced. For lower throughputs, solutions are presented in [4], where the parallelism factor is limited to dividers of the circulant size. It has been shown in [25] that these techniques can be generalized for any parallelism factor at the cost of some efficiency loss.

12.2.4.3 Multi-Standard

Different standards have different codes and parameters like the number of participants in parity check equations, and different number of equations in which a symbol participates. In fact, the support for this feature is already implemented inside one decoder due to the different specification of codes for different rates, as explained previously.

For the architecture to support multiple codes with the same properties, only a different address generation unit and different rotation angles must be used. This is realized by a ROM-based address generation. The architecture therefore supports scaling in rate, basically running multiple codes on the same hardware. The control of the ROM depends on the parameters of the code, which can be configured easily. The flexible addressing enables the architecture presented in Fig. 12.10 to support multiple standards/codes with the same parallelism factor.

To take into account the different parallelism factors in the codes, only the barrel shifter must be modified. It must be able to rotate over any angle for any parallelism factor. The number of data paths must be chosen in such a way, that it satisfies the most computationally demanding standard. Consequently, the height must be adapted to the standard with the largest code word length taking the possible parallelism factor for the code into account.

The consequence of using an enhanced barrel shifter for less demanding standards is that only a limited number of columns are used within the memory. Whereas the area of the multi-standard decoder is (far) less than the sum of the individual decoders, the power consumption, even for the simplest standard, is close to that of the most demanding one. Filling the non-used columns with zeros might reduce this overhead.

12.3 Decoders Supporting Multiple Code Families

In the previous section every decoder family was discussed separately. Some preliminary observations on resource sharing for combined decoder designs were made based on the analysis of the data-path structure and complexity of the corresponding decoder. For example, Reed-Solomon requires very different operations compared to other code families, so it is very difficult to find any sensible combination incorporating hardware resource sharing for this code family. On the other side, an LDPC decoder is so much larger than almost any Viterbi decoder, that effort to combine them (although in principle possible) would be much higher than the potential gain

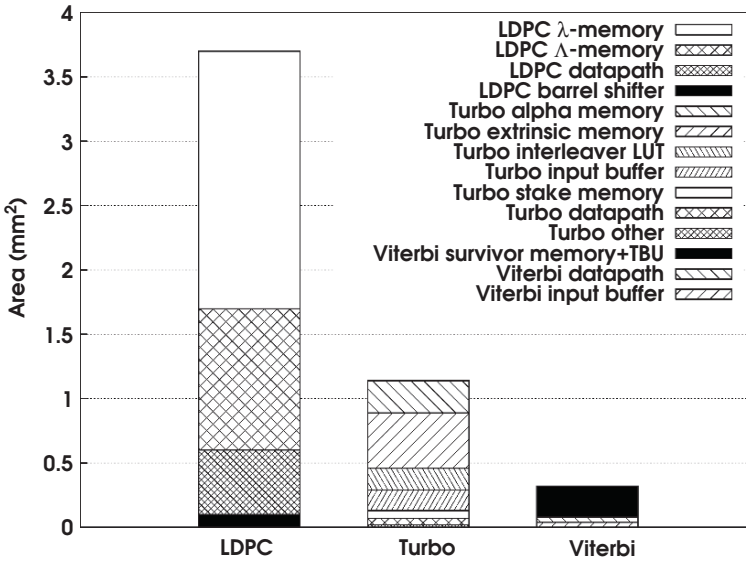


Fig. 12.13 Area distribution of LDPC (DVB-S2), Turbo (UMTS) and Viterbi (UMTS) decoders

in area or power. Taking these observations into account, we limit the following discussion to the two cases of combined Viterbi/Turbo and combined Turbo/LDPC decoders.

Before analyzing the resource sharing possibilities for combined decoders, we will support the above statement with a reference case: a design where three decoders are completely separate. For this example, area numbers are estimated for Viterbi, Turbo, and LDPC decoder. Figure 12.13 shows the result of this estimation. The distribution of area between main data processing and memory units within each decoder can be observed in the figure. For Turbo block size and LDPC codeword length, the UMTS and DVB-S2 standards are assumed respectively. Sufficient resources have been included in the Viterbi decoder figures for supporting different polynomials and constraint sizes up to $K = 9$. Looking at Fig. 12.13, it is already clear that any resource sharing in case of combined Turbo/Viterbi or Turbo/LDPC decoder will not be larger than 20–25%, as this is roughly the area relationship of complete Viterbi to Turbo decoder or complete Turbo to LDPC decoder respectively. For the same reason the potential for hardware sharing in case of combined Viterbi/LDPC decoder is almost neglectable (less than 6–8%).

As the given example corresponds to only one case, we can further use Fig. 12.13 to reason about alternative scenarios. For Turbo and LDPC decoders, the most important parameter affecting the decoder area is the Turbo code block size. For example, CDMA2000 specifies four times larger block size compared to UMTS, which increases the Turbo decoder area roughly by the same factor. In this case, the combination with Viterbi decoder will deliver even less area advantage than the case in Fig. 12.13.

12.3.1 Combining Viterbi and Turbo

12.3.1.1 Algorithms

Up to now, few approaches combining the Turbo and Viterbi decoding have been reported [26,28]. The main emphasis is on the multi-channel aspect, and the flexibility in coding schemes has not been studied so far. At the same time, the throughput requirements of the considered decoders are quite different.

The Turbo and Viterbi datapath hardware is determined by the choice of decoding algorithms used. The most commonly used algorithms for decoding of Turbo and convolutional codes are Max-log-MAP algorithm and hard-output Viterbi respectively. It is in principle possible to use other algorithms, for example a soft-output Viterbi algorithm (SOVA) within the Turbo soft-input soft-output (SISO) unit or Max-log-MAP algorithm for decoding convolutional codes. A more detailed study on the algorithm options for a combined Turbo-Viterbi data path was conducted previously [29]. In this study, both algorithms have been compared with respect to complexity, error correction performance and scalability. With respect to these criteria, the conventional hard-output Viterbi for convolutional decoding and Max-log-MAP algorithm for Turbo decoding emerge as the best choices.

12.3.1.2 Data Path Reuse

The fundamental fact to be understood about the datapath sharing between Turbo and Viterbi decoding is that both hard-output Viterbi and Max-log-MAP algorithms work with Add-Compare-Select (ACS) operations. Furthermore, it is possible to combine datapaths corresponding to different values of constraint length K . In some cases, Turbo decoding can exploit a wide datapath by working on parallel windows, while Viterbi can work on one sliding window and even time-share the datapath for supporting larger constraint lengths. In this way, it is possible to scale a combined Turbo-Viterbi data path with respect to the throughput and number of states required by different standards. An example of such a datapath is discussed in detail in [29]. A good measure of the processing power required for Viterbi/Turbo decoding in a certain standard is the number of butterflies per second. This parameter compares the processing requirements regardless of the different code types (Turbo or Viterbi) and coding parameters (e.g. constraint length $K = 4$ for Turbo and $K = 7...9$ for Viterbi). For different wireless standards there is a spread from $<10^1$ (CDMA Viterbi) to $>10^3$ (WiMax WiBRO Turbo) in terms of processing requirements. Because of this large spread, it is not realistic to make one generic datapath for all these standards together. Ad hoc combinations of Turbo and Viterbi standards make sense in some cases. The size and sharing potential of the datapath is, however, different in every case. When it is required to combine UMTS R'99 and DAB on one decoder, for instance, one datapath of eight states will be sufficient, but when the same decoder has to support UWB as well, a wider datapath will be needed to support the high Viterbi throughput required. Similarly, on a data path that has to support

future generations of the UMTS standard (up to 100 Mbps), support for a 802.11a/g Viterbi decoding can be added at relatively small cost. In general, when multiple standards are combined, the standard with the higher throughput requirement will determine the datapath size and the clock rate.

12.3.1.3 Memory Reuse

As can be seen in Fig. 12.13, the largest memory in the Viterbi decoder is the survivor memory. The survivor memory size depends on the constraint size: the width of the memory is the number of states (2^{K-1}) and the depth is often at least 20 times the constraint length for wireless fading channels (a smaller traceback depth is sufficient when AWGN channels are considered). For a $K = 9$ Viterbi decoder, this results in a memory as large as 256 bits wide and 256 lines deep. This size is comparable to the size of Turbo extrinsic memory, which is normally 12 bits wide for a $R = 1/3$ Turbo decoder, and has depth equal to the block size, e.g. 5120 in case of UMTS. In case of the reconfigurable 64-state datapath described above, the required bandwidth to the survivor path is 64 bits, as a maximum of 64 ACS operations can be done per clock cycle. In case of Turbo, eight windows are processed in parallel datapath slices. Each window produces a soft word of 7 bits or larger. By reserving 8 bits per soft word, a 64-bit bandwidth is required in this case as well. It results in Viterbi survivor memory and Turbo extrinsic memory having similar size and bandwidth when both data paths are running at maximum speed. The input buffers of Turbo and Viterbi can also be combined, although this would not save much space as the input buffer of the Viterbi decoder is in general one tenth of that of Turbo decoder or even smaller.

12.3.1.4 Area Comparison

In [29], a combined Turbo/Viterbi architecture is proposed. Here we will summarize the findings of this study. For more details the reader is referred to the full paper. The architecture suggested in this work is shown in Fig. 12.14. The datapath can be reconfigured as eight Max-log-MAP decoding datapaths for Turbo or one 64-state ($K = 7$) datapath for Viterbi. Interconnect for supporting larger Viterbi constraint sizes $K = 8$ and $K = 9$ is also included. In case of Turbo decoding, the datapath is configured to support forward (α) or backward (β) Max-log-MAP processing. An extension is possible by adding a parallel datapath for backward (β) computations for Turbo decoding, thereby increasing the Turbo decoding throughput by two. In this case, the memory area does not change but the datapath becomes larger due to the added β units.

The described architecture enables not only the sharing of the 64 ACS units (organized as 32 butterfly operations) in the datapath but also the sharing of the survivor/extrinsic memory because of the matching bandwidth requirements. The main purpose of this example is to show this point. Different instantiations are possible

for other ratios. According to the architecture example, area figures have been estimated and compared for the three following architecture cases:

- Architecture I is as explained above and shown in Fig. 12.14. In the case of Turbo, the datapath is time-shared between α and β computations. For Viterbi, parallel state decoding is supported up to $K = 7$ by the 64-state datapath. For higher K , folding is applied.
- Architecture II is the same as architecture I, except that a parallel β datapath is included, doubling the Turbo data rate and increasing the datapath area.
- Separate decoders are the reference case, consisting of a reconfigurable Viterbi decoder and a reconfigurable Turbo decoder. Turbo has parallel α and β computation, and Viterbi has a $K = 7$ 64-state datapath with folding-mode support for $K = 8$ and $K = 9$. In this case Turbo and Viterbi decoders can work completely in parallel.

A clock speed of 300 MHz is assumed for throughput calculations.

As shown in Table 12.4, the area advantage of combining Viterbi and Turbo decoding in one architecture is around 20%. This advantage comes at the cost of scheduling overhead arising from sharing the same datapath for both algorithms.

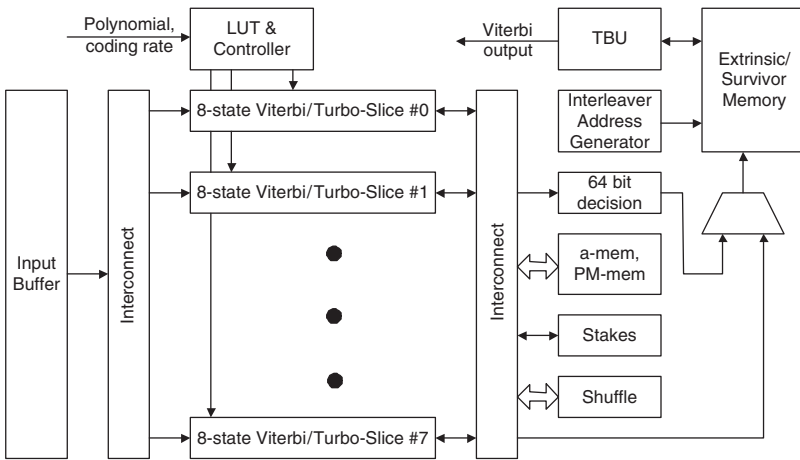


Fig. 12.14 Reconfigurable sliced architecture for multi-standard Viterbi and Turbo FEC decoding

Table 12.4 Reconfigurable Turbo–Viterbi architectures

	Total area (mm ²)	Viterbi throughput (K = 7)	Viterbi throughput (K = 9)	Turbo throughput (K = 4)
Architecture I	1.69	300 Mbps	75 Mbps	60 Mbps
Architecture II	2.01	300 Mbps	75 Mbps	120 Mbps
Separate decoders	2.35	300 Mbps	75 Mbps	120 Mbps

Especially when designing for standards that have strict latency requirements such as IEEE 802.11a/g, sharing the same datapath can cause several context switches within the decoding of one Turbo code block.

12.3.1.5 Results and Conclusion

When combining Viterbi and Turbo decoders, a modest area saving is possible by sharing resources. The main resources that can be shared are the datapath consisting of butterfly units and the survivor memory in Viterbi shared with the extrinsic memory in the Turbo decoder. In case of data path sharing, the area gain will be higher if the processing requirements for Turbo and Viterbi are close, $(M_{butterflies}/s)_{Turbo} \approx (M_{butterflies}/s)_{Viterbi}$. For sharing of memories, both memory capacity and bandwidth required need to match. The capacity condition requires that the Turbo extrinsic data memory and Viterbi survivor memory have similar sizes:

$$2 \times P \times B_{Turbo} \approx 2^{(K_{Viterbi}-1)} \times TB_{Viterbi} \quad (12.1)$$

where B_{Turbo} is the Turbo block size, $K_{Viterbi}$ is Viterbi constraint length and $TB_{Viterbi}$ is Viterbi traceback depth. The factor P in the equation is the precision of the Turbo extrinsic information and usually varies from 6 to 10 depending on the standard and implementation. At the same time the memory bandwidth for the two memories should also be close:

$$\frac{2 \times Throughput_{Turbo} \times Iterations}{P} \approx \frac{Throughput_{Viterbi}}{2^{K_{Viterbi}-1}} \quad (12.2)$$

Another fundamental point is that, when it is possible to share resources, the area gain is dependent mainly on how the sizes of two decoders relate. In cases where Turbo decoder has a very large block size (such as CDMA2000), the percentage gain will be meager as the area occupied by the Viterbi resources will be relatively small. So for a high percentage gain we have the condition $Area_{Turbo} \approx Area_{Viterbi}$. However, in some cases it is desirable to have an area advantage in absolute numbers even if the ratio to total decoder area is not very large. So, as a rule of thumb we require $B_{Turbo} \gg 2000$ for the absolute area gain to become significant.

12.3.2 Combining LDPC and Turbo

12.3.2.1 Architecture

Both Turbo and LDPC have a memory dominated architecture, where typically multiple words are read from a memory per clock cycle, and are written back after a certain time when the equations are solved.

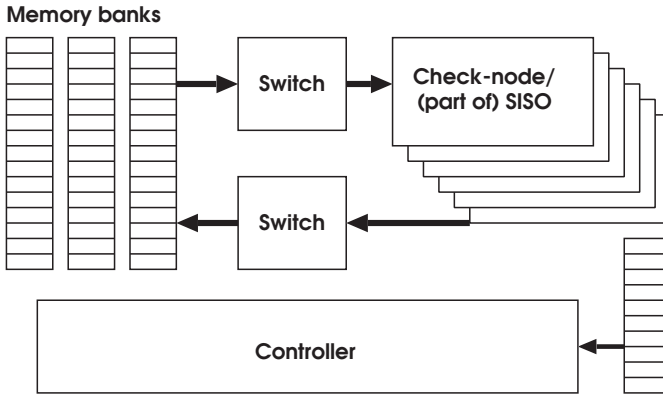


Fig. 12.15 Combined Turbo/LDPC architecture

In case of a high-speed parallel Turbo architecture the data exists in multiple banks since different addresses must be read. After reordering the data dedicated data paths solve the corresponding equations. The data are then typically written back at the same locations where they were read. Both Turbo and LDPC have a controller where interleaving addresses (or the H -matrix) are read from a pre-initialized memory. Figure 12.15 shows the targeted combined LDPC/Turbo architecture. In the next paragraphs we will evaluate the most important components of this architecture in terms of resource sharing.

12.3.2.2 Data Path Reuse

The LDPC data path for the min-sum variant of the algorithm basically consists of finding the minimum and one-but-minimum from a set of inputs. Such a data path mainly has a couple of comparators to administrate the minimum and one-but-minimum and a few adders, resulting in approximately 0.007 mm^2 silicon area. The number of data paths varies from 4 (802.16e) to 127 (DMB-T), and is typically around 60.

The SISO data path mainly consists of log-butterfly operators: 4 adders, and 2 maximum units. For constraint size $K = 4$, the α , β , and Λ data paths consist of respectively 4, 4, and 8 log-butterfly operators and the complexity grows exponentially with K (e.g. for $K = 5$ these numbers double). For high throughput Turbo decoding multiple SISO units are instantiated the same way as was shown in an example in Section 12.3.1.

The best match for a combined LDPC/Turbo data path is achieved when both have the same granularity, e.g. at the check-node and log-butterfly operator level. However the size of these data paths is so small that the additional configuration logic will have comparable area. We therefore exclude the option of efficient reuse of the data path logic.

12.3.2.3 Interconnect Reuse

The switching logic for all standardized LDPC-codes consist of barrel shifting, which has the task of rotating the input vector over a certain (programmable) degree. The standard cell complexity of this operation is $O(n \log(n))$, where n matches the number of data paths.

The switch in Turbo is a fully connected crossbar with complexity $O(n^2)$ at granularity of the number of SISOs (typically 2–8). In theory switching over α , β , and Λ or even log-butterfly operators is possible, but it results in significant area overhead, which is higher than potential gain due to the quadratic complexity (over the large number of parallel datapaths in LDPC). Thus, the option of efficient reuse of the switching logic is excluded from further consideration as well.

12.3.2.4 Memory Reuse

The storage requirements of LDPC decoding are mainly determined by λ and Λ values, typically around 18 bits per input bit, e.g. $18 \times B_{ldpc}$. For DVB-S2 this leads to 1.2 Mbits of storage, for DMB-T 0.14 Mbits of storage are required. Approximately half of it is a single port SRAM, the other half is two-port SRAM.

The storage capacity of Turbo decoding is mainly determined by the input values and internal data on which the iterations take place. For example, for a UMTS Turbo decoder this leads to 18 bits per input bit e.g. $18 \times B_{turbo}$ bits storage, up to 0.1 Mbit in total. Approximately one third of it is dual port, which can be implemented as two single port memories, leading to $24 \times B_{Turbo}$ bits storage with $B_{Turbo} \leq 6144$, so $Mem_{Turbo} \leq 0.11$ Mbit.

If the LDPC memory is 10 times larger compared to the Turbo memory, reuse will lead to only 10% area gain. Only in case of $B_{ldpc} \approx B_{turbo}$ effective reuse can take place. Next to the storage capacity also the bandwidth needs to be considered. The LDPC memory is dominated by the (dual port) λ memory where each bit is accessed between 60 and 150 times. This leads to high bandwidth requirements where in each clock cycle hundreds of bits are accessed, depending on the exact code and the required throughput. For e.g. DVB-S2 this results in a bandwidth of more than 600 bits read and written every clock cycle. In case of Turbo each bit is accessed around 6 times, thus, around 20 bits that are accessed per clock cycle per SISO. This again depends on the required throughput. For UMTS long-term evolution (LTE), with similar throughputs as DVB-S2 this results in 128 bits read and 48 bits written every clock cycle.

Combining the memories with huge bandwidth differences lead to multiplexing at the memory ports, which has limited area overhead, however, potentially results in massive power consumption overhead. If each clock cycle 600 bits are read from a memory, of which only 128 are used, the power consumption is almost five times too high compared to the optimal memory usage. Only in case of similar bandwidths this overhead is minimal, leading to the requirement $Throughput_{ldpc} \approx \frac{1}{4} \times Throughput_{turbo}$.

The last challenge in effective reuse of the memories is the memory organization. In case of LDPC this is one very wide memory, physically separated over banks of around 128 bits due to limitations of SRAMs. In case of Turbo it consists of memory banks of around 10 bits width each. In case of combined UMTS LTE Turbo ($B = 6144$) and StiMi ($B = 9216$), having throughput difference around factor 3, LDPC needs 1 dual-port memory of 160 bits width with 600 addresses and one single port memory of 112 bits width with 900 addresses. Turbo requires 8 memory banks, each consisting of 3 physical memories of 12, 6, and 6 bits width, and 800 addresses depth. In this case 112 out of $8 \times (12 + 6 + 6) = 192$ bits can be reused with the LDPC, the remaining 80 bit can be reused with the dual port RAM, requiring this dual port RAM to be partitioned into many banks resulting in high memory efficiency loss. Furthermore the layout and routing overhead for interconnecting these separate memory instances will further decrease the gain, making reuse less attractive. We therefore conclude that memory reuse is possible under certain conditions but the area advantage is small in most cases.

12.3.2.5 Results and Conclusion

The outcome that only memory can be reused for a small number of special cases, leads us to the conclusion that setting up a generic LDPC/Turbo processor architecture has little benefits. Furthermore, the overhead for being able to process two decoder families on the same hardware may further decrease the gain. Similarly, shared resources can be a disadvantage when multiple channels must operate simultaneously, requiring twice the bandwidth. In general, the reuse of memories can be implemented without designing the two IPs as one processor. To achieve this the memories must be accessible from other IP blocks preferably at low latency.

12.4 Simultaneous Multi-Standard FEC Decoding

Our analysis so far has focused on sharing same hardware resources to decode streams of different FEC families. Although making a decoder of one family reconfigurable for covering codes with different polynomials, block sizes, etc has already been shown in literature [27,28], resource sharing between families is more complex to design and delivers meager area advantage in most cases.

An additional dimension of complexity to consider is that of decoding multiple streams of different standards simultaneously. Whereas configuring the hardware for one standard at a time has been covered in our analysis, the additional costs and complexity of time-multiplexing between differently encoded streams has not been included. To give an example, let us assume that a combined Turbo/Viterbi decoder is required to time-switch between a convolutionally encoded 802.11(a/g) stream and a Turbo encoded UMTS HSDPA stream. As 802.11(a/g) latency requirements are fairly strict, the Turbo decoding will need to be interrupted before a block is

decoded. For this reason, state-saving and switching must be implemented for both standards. For Turbo, the state of the decoder is defined by the input buffer, state memory and the extrinsic memory. For Viterbi the input, survivor and path metric memories define the state of the decoder. This restricts the moments of context switch of the Turbo process to the end of every half-iteration, whereas a Viterbi process can be switched at any stage. Thus to support parallel streams we require an extra 125 Kbits memory for every Turbo stream and 64 Kbits memory for every Viterbi stream. This translates to around 0.8 mm^2 for each extra Turbo stream and 0.4 mm^2 for each extra Viterbi stream. For LDPC decoding, the reasoning is similar to Turbo, as it is possible to stop between iterations. All intermediate data has to be saved, which includes the check-node and symbol-node memories.

We concluded earlier that the most interesting candidate for reuse in a multi-standard decoder is the memory. However for the simultaneous multi-standard decoding we just showed that this memory must be instantiated multiple times since it contains the decoder's state. This reduces the potential area gain in multi-standard decoding significantly. Algorithm modifications for reducing the state size to solve this problem appear as an interesting topic of future research.

12.5 Conclusions

In order to reduce the number of different hardware implementations, it makes a lot of sense to parameterize the decoder hardware architecture for a particular decoder family (Reed Solomon, Viterbi, Turbo, and LDPC). With a relatively small overhead in energy/bit (or in die area), such a decoder can handle multiple data rates, block sizes, coding rates, and more specific parameters such as polynomials and constraint sizes. Based on this study, Figs. 12.16 and 12.17 offer estimates of the die area and energy consumption for single-standard decoders. Multiple bit streams can generally share the same decoder resources, time-multiplexed at block level. In specific cases preemption can be supported in order to accommodate tough real-time constraints encountered e.g. in 802.11x standards.

Although it looks tempting to share hardware among decoders of different decoder families, our study suggests the evidence that the area saving will be relatively low. The most promising combinations appear to be Viterbi/Turbo and Turbo/LDPC. Even in these cases, however, the overhead of sharing arithmetic hardware turns out to be larger than the area saved. Furthermore, the sizes of memories and required memory bandwidths for the different standards make memory sharing unattractive in general. Nevertheless, for specific combinations of demanding decoders, memory sharing may offer some cost savings as has been shown in this chapter. These savings can justify efforts on development and implementation of combined decoders in specific cases but do not provide a clear evidence for the development of generic Viterbi/Turbo or Turbo/LDPC processors.

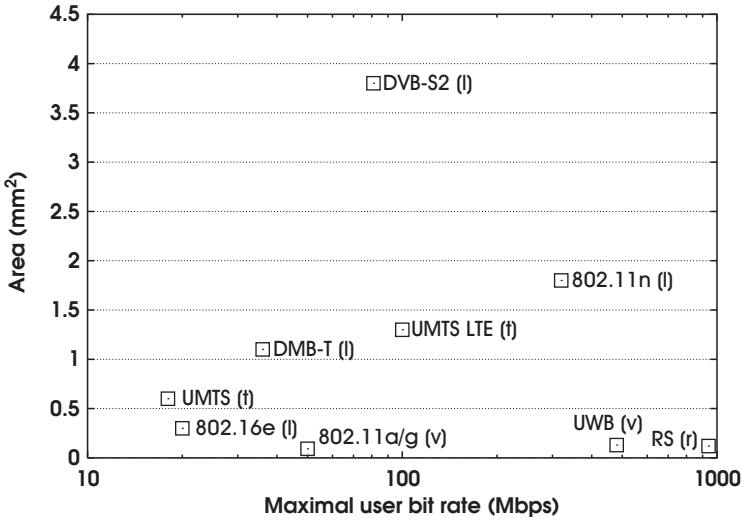


Fig. 12.16 Size estimations for channel decoders for a variety of transmission standards

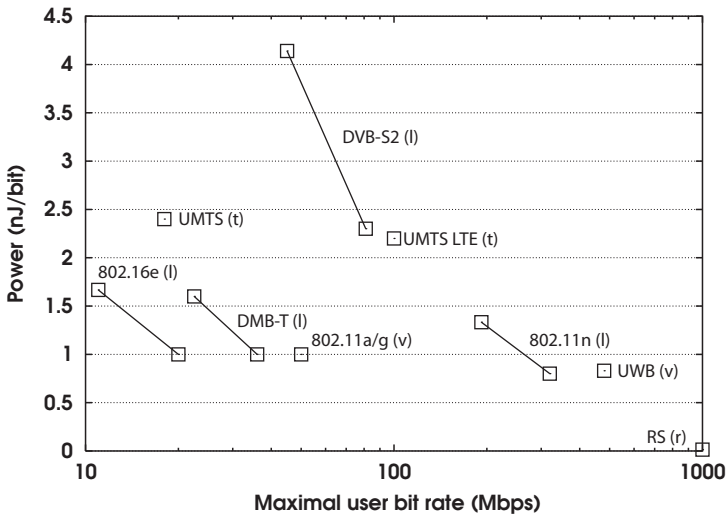


Fig. 12.17 Energy consumption estimations for channel decoders for a variety of transmission standards

References

1. Kang, H. J., Park, I. C.: A High-Speed and Low-Latency Reed-Solomon decoder based on a Dual-Line Structure, IEEE International Conference on Acoustics, Speech, and Signal Processing, May 2002
2. Fettweis, G., Meyr, H.: High-Speed Parallel Viterbi Decoding: Algorithm and VLSI-Architecture, IEEE Commun. Mag., 29(5), May 1991, pp. 46–55

3. Bekooij, M., Dielissen, J., Harmsze, F., Sawitzki, S., Huisken, J., van de Werf, A.: Power-Efficient Application-Specific VLIW Processor for Turbo Decoding, Proceedings of ISSCC'01, 4–8 February 2001
4. Dielissen, J., Hekstra, A., Berg, V.: Low cost LDPC Decoder for DVB-S2, Proceedings of DATE 2006, 6–10 March 2006, pp. 130–135
5. Sarvate, D. V., Shanbhag, N. R.: High-Speed Architectures for Reed-Solomon Decoders, IEEE Trans. VLSI Syst., 9(5), October 2001, pp. 641–655
6. Kumar, A.: High-Throughput Reed-Solomon Decoder for Ultra Wide Band, M.Sc. thesis, National University of Singapore, November 2004
7. Kumar, A., Sawitzki, S.: High-Throughput and Low-Power Architectures for Reed-Solomon Decoder, Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, 2005
8. Habib, I.: Automated Generation of Viterbi Decoders, M.Sc. thesis, Indian Institute of Technology, Delhi, July 2006
9. Lin, H. D., Messerschmitt, D. G.: Algorithms and Architectures for Concurrent Viterbi Decoding, IEEE International Conference on Communications, Boston MA., June 1989
10. Berrou, C. and Glavieux, A., Thitimajshima, P.: Near Shannon Limit Error-Correcting Coding and Decoding: Turbo Codes, IEEE Proceedings of ICC'93, May 1993, pp. 1064–1070
11. Berrou, C.: The Ten-Year-Old Turbo Codes are Entering into Service IEEE Commun. Mag., 41(8), August 2003, pp. 110–116
12. Bougard, B., Giulietti, A., Derudder, V., Weijers, J.-W., Dupont, S. and Holl, L.: A scalable 8.7nJ/bit 75.6Mb/s Parallel Concatenated Convolutional (Turbo-) CODEC, Proceedings of ISSCC'03, 9–13 February 2003, pp. 152–484
13. Gilbert, F., Thul, M. J., Wehn, N.: Communication Centric Architectures for Turbo-Decoding on Embedded Multiprocessors, Proceedings of DATE 2003, 3–7 March 2003, pp. 356–361
14. Schurgers, C., Carrhoor, F., Engels, M.: Optimized MAP Turbo Decoder, IEEE Workshop on Signal Processing Systems, October 2000, pp. 245–254
15. Giulietti, A., van der Perre, L., Strum, M.: Parallel Turbo Coding Interleavers: Avoiding Collisions in Accesses to Storage Elements, Electron Letts., 38(5), February 2002, pp. 232–234
16. Thul, M. J., Gilbert, F., Wehn, N.: Concurrent Interleaving Architectures for High-Throughput Channel Coding, Proceedings of International Conference on Acoustics, Speech, and Signal Processing IC, pp. II-613–616
17. Tarable, A., Benedetto, S., Montorsi, G.: Mapping Interleaving Laws to Parallel Turbo and LDPC Decoder Architectures, IEEE Trans. Inf. Theory, 50(9), September 2004, pp. 2002–2009
18. Kschischang, F. R., Frey, B. J., Loeliger, H. A.: Factor Graphs and the Sum-Product Algorithm, IEEE Trans. Inf. Theory, 47(2), February 2001, pp. 498–519
19. Richardson, T., Shokrollahi, A., Urbanke, R.: Design of Capacity-Approaching Irregular Low-Density Parity Check Codes, IEEE Trans. Inf. Theory, 47(2), February 2001, pp. 619–637
20. MacKay, D.: Good Error Correcting Codes Based on Very Sparse Matrices, IEEE Trans. Inf. Theory, 45(2), February 1999, pp. 399–431
21. Fossorier, M., Mihaljevic, M., Imai, H.: Reduced Complexity Iterative Decoding of Low-Density Parity Codes Based on Belief Propagation, IEEE Tran. Comm., 47(5), May 1999, pp. 673–680
22. Chen, J., Dholakia, A., Eleftheriou, E., Fossorier, M., Hu, X.: Reduced-Complexity Decoding of LDPC Codes, IEEE Tran. Comm., 53(8), August 2005, pp. 1288–1299
23. Zhao, J., Zarkeshvari, F., Banihashemi, A.: On Implementation on Min-Sum Algorithm and its Modifications for Decoding Low-Density Parity-Check (LDPC) Codes, IEEE Tran. Comm., 53(4), April 2005, pp. 549–554
24. Blanksby, A. J., Howland, C. J.: A 690-mW 1-Gb/s 1024-b, Rate-1/2 Low-Density Parity-Check Code Decoder, IEEE J. Solid-State Circ. 37(3), March 2002, pp. 404–412
25. Kienle, F., Brack, T., Wehn, N.: A Synthesizable IP Core for DVB-S2 LDPC Code Decoding, IEEE Proceedings of DATE 2005, 7–11 March 2005, pp. 1530–1535
26. Dielissen, J., Hekstra, A.: Non-fractional Parallelism in LDPC Decoder Implementations, Proceedings of DATE 2007, 16–20 April 2007, pp. 337–342

27. Thomas, C., Bickerstaff, M. A., Davis, L. M., Prokop, T., Widdup, B., Zhou, G., Garrett, D., Nichol, C.: Integrated Circuits for Channel Coding in 3G Cellular Mobile Wireless Systems, *IEEE Commun. Mag.*, 41(8), August 2003, pp. 150–159
28. Vogt T., Wehn N., Alves, P.: A Multi-Standard Channel Decoder for Base-Station Applications, *Proceedings of SBCCI*, February 2004, pp. 192–197.
29. Krishnaiah, G., Engin, N., Sawitzki, S.: Scalable Reconfigurable Channel Decoder Architecture for Future Wireless Handsets, *Proceedings of DATE 2007*, 16–20 April 2007, pp. 1563–1568

Index

A

ACPR. *See* Adjacent channel power ratio
ACS. *See* Adjacent channel selectivity
Adaptive multi-mode low-power DFE, concept and design, 254
 ADC leveling, 262–264
 CIC filters, 255–256
 CORDIC algorithm, 256–257
 DC notch filter, 259–260
 digital interface, 264–265
 finite impulse response filter and fractional sample-rate converter, 260–262
 IIR and allpass filter, 258–259
 simulation outcomes for, 265–268
ADC leveling, DFE, 262–264. *See also*
 Multi-mode wireless receivers, adaptive digital front-end
A/D Converter, 147
 $\Delta\Sigma$ A/D converter, noise shaping, 11
Add-compare-select (ACS) operations, 275, 287. *See also* FEC decoders, for future wireless devices
Adjacent channel power ratio, 123, 266
Adjacent channel selectivity, 250
AFC. *See* Automatic frequency control
AFE. *See* Analog front-end
AGC. *See* Automatic gain control
Allpass filter, DFE, 258–259. *See also*
 Multi-mode wireless receivers, adaptive digital front-end
Amplitude modulation, polar transmitter, 18–20. *See also* Multimode radio transceiver, cellular applications
Analog front-end, 250
Analog-to-digital converter, 90, 203
Automatic frequency control, 15
Automatic gain control, 9, 263

Average power efficiency, 125–126. *See also*
 Multi-mode power amplifiers, wireless handset applications

B

Bandpass filter, 90
Baseband (BB) signal, 152–153
Baseband processing and control unit, function of, 146
Baseband receiver, 9–13. *See also* Multi-mode radio transceiver, cellular applications
Binary phase shift keying, 90
BPF. *See* Bandpass filter
BPSK. *See* Binary phase shift keying
Bridged-T filter, 7

C

Cascaded-integrator-comb, 255–256. *See also*
 Multi-mode wireless receivers, adaptive digital front-end
CDMA. *See* Code division multiple access
Cellular applications, multimode radio transceiver, 5
 polar transmitter, 15–17
 amplitude modulation, 18–20
 phase/frequency modulation, 17–18
 Rx band noise, 20–23
 receiver, 5–6
 baseband receiver, 9–13
 PLL, 13–15
 RF front-end, 6–9
Cellular phone applications, IIP2 performance requirement, 170
CIC filters. *See* Cascaded-integrator-comb

- CML. *See* Current mode logic
- CMOS downconversion mixers, second order intermodulation distortion mechanism, 173
- CMOS IR-UWB transceiver system design, 85–86
- IR-UWB
 transceiver systems, architecture of, 92–93
 transmitter designs, 93–101
- UWB
 LNA design, 106–114
 transceiver, circuit implementations in, 114–117
 wireless and wireless links, 86–91
 wireless interconnect, magnetic coupling transformer, 102–106
- CMOS. *See* Complementary MOSFET
- CMOS technology, advantages, 123
- Code division multiple access, 89
- Cognitive radio system, 145–147
- Complementary MOSFET, 88
- Contact-less chip testing applications, CMOS
 IR-UWB transceiver system in, 85–86
- IR-UWB
 transceiver systems, architecture, 92–93
 transmitter designs, 93–101
- UWB
 LNA design, 106–114
 transceiver, circuit implementations, 114–117
 wireless and wireless links, 86–91
 wireless interconnect, magnetic coupling transformer, 102–106
- Coordinate-Rotation-Digital-Computer, 253, 256–257. *See also* Multi-mode wireless receivers, adaptive digital front-end
- CORDIC. *See* Coordinate-Rotation-Digital-Computer
- Current mode logic, 160
- D**
- D/A Converter, 147
- DAC. *See* D/A Converter; Digital to analog converter
- DA. *See* Distributed amplifiers
- DCB. *See* Dynamic current biasing
- DC notch filter, DFE, 259–260. *See also* Multi-mode wireless receivers, adaptive digital front-end for
- DCS. *See* Digital Cellular Communication System
- DDB. *See* Dual dynamic bias
- Decoders, for multiple code families, 285–286
 LDPC and turbo, combination, 290–293
 viterbi and turbo, combination, 287–290
- De-embedding techniques, UWB receivers, 44–46. *See also* UWB receivers
- Delay locked loop (DLL), 78–79, 146, 156
- Device-under-test, 45, 85
- DFE. *See* Digital front-end
- DFF. *See* D flip-flop
- D flip-flop, 158
- Digital Cellular Communication System, 171
- Digital front-end, 249. *See also* Multi-mode wireless receivers, adaptive digital front-end
 ADC leveling, 262–264
 allpass filter, 258–259
 DC notch filter, 259–260
 digital interface, 264–265
 FIR filter, 260
 in GNSS mode, functions, 253
 IIR, 258
 simulation outcomes, 265–268
- Digital interface, DFE, 264–265. *See also* Multi-mode wireless receivers, adaptive digital front-end
- Digital predistortion (DP) techniques, 131. *See also* Multi-mode power amplifiers, wireless handset applications
- Digital signal processing, 131, 254
- Digital to analog converter, 55
- Digital Video Broadcasting–Handheld, 204
- Direct conversion architecture, 169
- Direct-sequence (DS) ultra wideband, 55–56. *See also* Ultra wideband
- Distributed amplifiers, 57
- DLL MPCG Jitter, 157–158
 simulation outcomes for, 166
 and SR Jitter, comparison of
 jitter generated due to thermal noise, 160–163
 jitter transferred from reference clock, 159–160
 mismatch jitter analysis, 163–165
- DSP intensive radio transmitter, digital signal processing techniques, 152
- DSP. *See* Digital signal processing
- DS-UWB systems, pulse based transmitters, 71–73
 full band 3.1–10.6 GHz, 73–76
 low sidelobe sub-band 3.1–10.6 GHz, 76–78

Dual dynamic bias, 126–128
 DUT. *See* Device-under-test
 DVB. *See* Dynamic voltage biasing
 DVB-H. *See* Digital Video Broadcasting–
 Handheld
 Dynamic current biasing, 122, 140
 Dynamic voltage biasing, 122

E

ECMA 368 standard, 27
 EDGE band, features, 5
 E-GSM. *See* Enhanced-GSM
 EIRP. *See* Equivalent Isotropically Radiated
 Power
 Electrostatic discharge, 106, 107
 LNA design and, 107–110
 Enhanced-GSM, 171
 Equivalent Isotropically Radiated
 Power, 54
 Error vector magnitude, 12, 265
 ESD. *See* Electrostatic discharge
 EVM. *See* Error vector magnitude

F

FCC. *See* Federal Communications
 Commission
 FDD. *See* Frequency division duplexing
 FEC decoders, future wireless devices,
 271–273
 single code family, decoders
 LDPC, 281–285
 reed-solomon, 273–274
 turbo, 277–281
 viterbi, 274–277
 Federal Communications Commission, 53, 88
 emissions mask, for UWB communica-
 tions, 54
 Feedback (FB) and feedforward (FF) topology,
 208–210
 Figure of merit, 160
 Filter-less power up-converter, 153–156
 Finite impulse response (FIR) filter, DFE,
 260. *See also* Multi-mode wireless
 receivers, adaptive digital front-end
 Flexible radio hardware platform, 145
 Flexible RX/RFS, 146–147
 Flexible TX, 148–150
 Flicker and thermal noises, calibration circuit,
 188–189
 Flip logic (FL) circuit, 158
 FOM. *See* Figure of merit

Fractional sample-rate conversion, 255,
 260–262. *See also* Multi-mode
 wireless receivers, adaptive digital
 front-end
 Frequency division duplexing, 172
 Frequency down-conversion, sampling clock
 jitter, 147–148
 FSRC. *See* Fractional sample-rate conversion

G

GaAs HBT technology, advantages, 123
 Gain-bandwidth, 111
 Gain bandwidth product, 175
 Gaussian minimum shift keying, 121
 Gaussian monocycle pulses, spectral density,
 89
 GBW. *See* Gain-bandwidth; Gain bandwidth
 product
 3.1–10.6 GHz band, UWB LNAs, 57–62. *See
 also* Low noise amplifier
 Giga operations per second, 271
 Global navigation satellite system, 250
 g_m -C filters. *See also* Multimode radios, power
 efficient reconfigurable baseband
 filters
 design strategy for biquadratic, 224
 g_m -C biquad architecture, 225–226
 multimode reconfigurable, 234–240
 Nauta's transistor, 226–228
 GMSK. *See* Gaussian minimum shift keying
 GNSS. *See* Global navigation satellite system
 GOPS. *See* Giga operations per second
 4G radio systems, 203. *See also* Wireless
 communication systems
 ADCs specifications, 204–205
 power considerations, topology exploration
 of circuit level, 210–214
 power calculation, 214–217
 of system level, 205–210
 reconfigurable DSM, 217–220
 GSM band, features, 5
 GSM of wireless telecommunications market,
 171–172

H

High speed downlink packet access (HSDPA),
 263

I

Ideal DACs and hard-switched mixers,
 modeling, 148–149

- IIP2 degeradation
 - input stage, 175
 - mixer input stage, reason of, 174
 - IIP2 improvement techniques, multi-standard
 - mobile radio, 169–170
 - calibration technique, 183–185
 - circuit implementation, 187–188
 - flicker and thermal noises in calibrated mixer, 188–189
 - proposed calibration technique, 185–187
 - proposed mixer, 189–190
 - simulation outcomes, 190–191
 - of GSM, 171–172
 - second order nonlinearity of input stage
 - circuit implementation, 179–180
 - IM2 cancellation mechanism, 175–179
 - proposed mixer, 180–181
 - simulation outcomes of, 181–183
 - second order nonlinearity of switching
 - pairs
 - experimental outcomes, 198–200
 - implementation of, 196–197
 - programmable inductor, 192–195
 - second order nonlinearity sources, 172–175
 - UMTS system, 172
 - IIP2 performance of mixer, gigahertz
 - frequencies, 173
 - IIP2. *See* Second order input intercept-point
 - IIV2. *See* Input intercept second-order voltage point
 - Image rejection ratio, 253
 - IM2 cancellation technique, 175–179
 - IMD. *See* Inter-modulation distortion
 - Impulse-based ultra wideband, 89–90
 - Impulse-radio ultra-wideband, 86
 - transceiver systems, architecture, 92–93
 - transmitter designs, 93
 - impulse generation, 94–97
 - impulse shaping using LC BPF, 97–101
 - IM2. *See* Second order intermodulation
 - Infinite impulse response (IIR), 258. *See also*
 - Multi-mode wireless receivers, adaptive digital front-end
 - Input intercept second-order voltage point, 174
 - Integrated noise level, 241
 - Inter-modulation distortion, 123
 - Inter-stage filtering, in UMTS RX front-end, 173
 - I/Q down converter, effect of I/Q imbalance, 12
 - I/Q imbalance, reduction of, 13
 - IRN. *See* Integrated noise level
 - IRR. *See* Image Rejection Ratio
 - IR-UWB. *See* Impulse-radio ultra-wideband
 - I-UWB. *See* Impulse-based ultra wideband
- J**
- Joint test action group (JTAG), 86
- L**
- LDMOS technology, advantages of, 123
 - LDPC decoder, 281–285. *See also* FEC
 - decoders, for future wireless devices
 - LDPC. *See* Low-density parity codes
 - Linearity performance, expression, 229
 - LNA. *See* Low noise amplifier
 - LO. *See* Local oscillator
 - Local oscillator, 8, 152–153
 - Long-term evolution, 292
 - Look-up table, 20
 - Low-density parity codes, 271
 - Low noise amplifier, 6, 56–57, 86, 172
 - design, in UWB receivers, 31–32 (*see also* UWB receivers)
 - noise and power matching in, 32–36
 - receiver linearity and LNA selectivity, 36–37
 - gain settings, 7
 - linearity, in UWB receivers, 40–42
 - primary-secondary inductive ESD protected, 107–110
 - for UWB systems (*see also* Ultra wideband)
 - 3.1–10.6 GHz band and, 57–62
 - 0–960 MHz band and, 62–71
 - Low voltage active mixer, preparation, 172
 - LTE. *See* Long-term evolution
 - LUT. *See* Look-up table
- M**
- Magnetic coupling transformer, wireless interconnect, 102–106
 - Max-log-MAP algorithm, for turbo decoding, 287. *See also* FEC decoders, for future wireless devices
 - MB-OFDM. *See* Multi-band orthogonal frequency division modulation
 - MC-UWB. *See* Multicarrier-based ultra wideband
 - MEMs. *See* Micro-electrical-mechanical system
 - Micro-electrical-mechanical system, 145
 - Micro lead frame, 138
 - MLF. *See* Micro lead frame

- $\Delta\Sigma$ Modulator, 15, 203
 - Monte Carlo simulation, of mixer, 182, 190
 - MOS transistor, noise parameters, 34
 - MPCG. *See* Multi-phase Clock Generator
 - Multi-band orthogonal frequency division modulation, 27
 - Multicarrier-based ultra wideband, 89–91
 - Multi-functional terminal, radio of, 169
 - Multi-mode power amplifiers, wireless handset applications, 121–122
 - efficiency enhancement, 125–126
 - circuit design, 129–131
 - DDB, principle of, 126–128
 - power control and gain variation, 128–129
 - linearity improvement, 131–132
 - complex-gain digital predistortion, 134
 - practical issues, 135–138
 - predistortion algorithm, 133–134
 - system topology, 132
 - measurement outcomes, 138–141
 - power amplifier technologies, 122–125
 - Multimode radios, power efficient reconfigurable baseband filters. *See also* Wireless communication systems
 - circuit level implementation and validation, 231–234
 - design optimization and strategy, 228–231
 - g_m -C filter architecture and circuit design, 225–228
 - multimode reconfigurable g_m -C filters, 234–240
 - outcomes of, 240–247
 - Multimode radio transceiver, cellular applications, 5
 - polar transmitter, 15–17
 - amplitude modulation, 18–20
 - phase/frequency modulation, 17–18
 - Rx band noise, 20–23
 - receiver, 5–6
 - baseband receiver, 9–13
 - PLL, 13–15
 - RF front-end, 6–9
 - Multimode terminals, usage of, 223
 - Multi-mode wireless receivers, adaptive digital front-end, 249–250
 - adaptive multi-mode low-power DFE, concept and design, 254
 - ADC leveling, 262–264
 - CIC filters, 255–256
 - CORDIC algorithm, 256–257
 - DC notch filter, 259–260
 - digital interface, 264–265
 - finite impulse response filter and fractional sample-rate converter, 260–262
 - IIR and allpass filter, 258–259
 - simulation outcomes, 265–268
 - architecture of adaptive, 251–254
 - Multi-phase clock generator, 156
 - Multi-phase clock requirements, 156
 - Multiple code families, decoders, 285–286
 - LDPC and turbo, combination of, 290–293
 - viterbi and turbo, combination of, 287–290
 - Multi-standard FEC decoders, definition of, 272
 - Multi-standard FEC decoding, 293–294
 - Multi-standard mobile radio, IIP2 improvement techniques, 169–170. *See also* Wireless communication systems
 - calibration technique, 183–185
 - circuit implementation, 187–188
 - flicker and thermal noises in calibrated mixer, 188–189
 - proposed calibration technique, 185–187
 - proposed mixer, 189–190
 - simulation outcomes, 190–191
 - of GSM, 171–172
 - second order nonlinearity of input stage
 - circuit implementation, 179–180
 - IM2 cancellation mechanism, 175–179
 - proposed mixer, 180–181
 - simulation outcomes of, 181–183
 - second order nonlinearity of switching pairs, 191
 - experimental outcomes, 198–200
 - implementation of, 196–197
 - programmable inductor, 192–195
 - second order nonlinearity sources, 172–175
 - UMTS system, 172
- N**
- Nauta g_m -C biquadratic low-pass filter, 224–225
 - Nauta's transistor, 224–228. *See also* g_m -C filters
 - NF. *See* Noise figure
 - Noise and power matching, UWB LNA design, 32–36. *See also* UWB receivers
 - Noise cancellation technique, LNA wideband implementation, 38–39. *See also* UWB receivers
 - Noise factor equation, 65
 - Noise figure, 28, 56, 90
 - Noise transfer function (NTF), 207

O

- OFDM. *See* Orthogonal frequency-division multiplexing
- OFDM UWB, frequency synthesis, 78–80
- On-off keying, 90
- OOK. *See* On-off keying
- Operational transconductance amplifier, 213, 227
- Orthogonal frequency-division multiplexing, 54–55, 89
- OSR. *See* Oversampling rate; Oversampling ratio
- OTA. *See* Operational transconductance amplifier
- Oversampling rate, 11
- Oversampling ratio, 254

P

- Packet error rate, 28
- PAE. *See* Power-added efficiency
- PAM. *See* Pulse amplitude modulation
- PAs. *See* Power amplifiers
- PCS. *See* Personal communication services
- PDF. *See* Probability distribution function
- PER. *See* Packet error rate
- Personal communication services, 171
- Phase/frequency modulation, polar transmitter, 17–18. *See also* Multimode radio transceiver, for cellular applications
- Phase-locked loop (PLL), 13–15, 37, 54
- 2-Point modulation, definition, 17
- Polar modulation, 15–16
- Polar transmitter, 15–17
 - amplitude modulation, 18–20
 - phase/frequency modulation, 17–18
 - Rx band noise, 20–23
- Polyphase multipath circuits, spectral purity enhancement, 150–151
- Polyphase n-path transmitter system, 152
- Power-added efficiency, 121
- Power amplifier technologies, 121–125. *See also* Multi-mode power amplifiers, wireless handset applications
- Power spectral density, 73, 87, 211
- PPM. *See* Pulse position modulation
- PRF. *See* Pulse repetition frequency
- Probability distribution function, 125
- Programmable inductor (Q_e), equivalent quality factor, 194
- PSD. *See* Power spectral density
- Pulse amplitude modulation, 90
- Pulse position modulation, 90
- Pulse repetition frequency, 72

R

- Radio bands, radio transceiver role in, 13–14
- Radio frequency identifier, 87
- Radio frequency scanner, 146
- Radio receiver, 5–6, 146, 148
 - baseband receiver, 9–13
 - PLL, 13–15
 - RF front-end, 6–9
- Radio transmitter, 146
- Reconfigurable ADC, design challenges, 204–205. *See also* 4G radio systems
- Reconfigurable DSM, 4G radios, 217–220. *See also* 4G radio systems
- Reconfigurable multi-band OFDM UWB receivers, 27, 46–49
 - UWB receivers, 27
 - architecture, 39–42
 - broadband measurements and de-embedding techniques, 44–46
 - design challenges, 31–36
 - design simulation outcomes, 49–51
 - evolution and state-of-the-art in, 37–39
 - receiver specifications, 28–31
- Reed-solomon decoder, 273–274. *See also* FEC decoders, for future wireless devices
- RF front-end, in radio receiver, 6–9. *See also* Multimode radio transceiver, cellular applications
- RFID. *See* Radio frequency identifier
- RF power amplifiers, technologies, 123
- RFS. *See* Radio frequency scanner
- Root raised-cosine, 21
- RRC. *See* Root raised-cosine
- Rx band noise, polar transmitter, 20–23. *See also* Multimode radio transceiver, cellular applications
- RX. *See* Radio receiver

S

- Sample rate conversion, 255
- Sampling clock jitter, frequency down-conversion, 147–148
- SDDB. *See* Switched dual dynamic biasing
- SDR. *See* Software defined radio
- Second order input intercept-point, 169, 170
- Second order intermodulation, 169
- Second order nonlinearity of input stage. *See also* Multi-standard mobile radio, IIP2 improvement techniques
 - circuit implementation, 179–180
 - IM2 cancellation mechanism, 175–179

proposed mixer, 180–181
 simulation outcomes of, 181–183
 Second order nonlinearity sources, IIP2
 techniques, 172–175. *See also*
 Multi-standard mobile radio, IIP2
 improvement techniques
 SFDR. *See* Spurious free dynamic range
 Shift Register, 146
 SiGe HBT technology, 124
 Signal to noise ratio, 53, 87
 Signal-to-quantization noise ratio, 206
 Single-sideband, 78
 SISO. *See* Soft-input-soft-output
 SNR. *See* Signal to noise ratio
 SOA. *See* Switchable Op-Amp
 SOC. *See* System-On-Chip
 Soft-input-soft-output, 278, 287
 Soft-output Viterbi algorithm (SOVA), 287
 Software defined radio, 203, 224
 Spectral purity enhancement, polyphase
 multipath circuits, 150–151
 Spurious free dynamic range, 245
 SQNR. *See* Signal-to-quantization noise ratio
 SRC. *See* Sample rate conversion
 SR MPCG Jitter, 158–159
 simulation outcomes for, 166
 and SR Jitter, comparison
 jitter generated due to thermal noise,
 160–163
 jitter transferred from reference clock,
 159–160
 mismatch jitter analysis, 163–165
 SR. *See* Shift Register
 SSB. *See* Single-sideband
 Switchable Op-amp, 219
 Switched dual dynamic biasing, 122
 Switching pairs, second order nonlinearity. *See*
 also Multi-standard mobile radio,
 IIP2 improvement techniques
 experimental outcomes, 198–200
 implementation, 196–197
 programmable inductor, 192–195
SynthesizeNTF function, 207–208
 System-On-Chip, 88

T

TFI. *See* Time frequency interleaving
 Time frequency interleaving, 28
 Turbo decoder, 277–281. *See also* FEC
 decoders, future wireless devices
 TX. *See* Radio transmitter

U

Ultra wideband (UWB), 27, 53
 LNA, 106
 3.1–10.6 GHz band, 57–62
 G_m-boosted, 110–114
 0–960 MHz band, 62–71
 primary-secondary inductive ESD
 protected, 107–110
 technology, 53–54
 DS ultra wideband, 55–56
 OFDM proposal, 54–55
 transceiver, circuit implementations,
 114–117
 wireless and wireless links, 86–88
 design approaches for, 89–91
 UMTS system, 172
 UWB receivers, 27, 46–49
 architecture, 39–40
 highly linear mixer, 42–44
 selective and reconfigurable LNA,
 40–42
 broadband measurements and
 de-embedding techniques, 44–46
 design challenges, 31–32
 receiver linearity and LNA selectivity,
 36–37
 wideband LNA design, noise and power
 matching, 32–36
 design simulation outcomes of, 49–51
 evolution and state-of-the-art in, 37–39
 receiver specifications
 interference analysis, 28–30
 interferers scenario, 30
 receiver linearity specifications, 30–31

V

Variable gain amplifier, 9
 Variable-gain control amplifier, 90
 VCDL. *See* Voltage controlled delay line
 VCO. *See* Voltage-controlled oscillator
 Very-large-scale-integration, 85
 VGA. *See* Variable gain amplifier; Variable-
 gain control amplifier
 V–I converter transistors, 153–154
 Viterbi decoder, 274–277. *See also* FEC
 decoders, future wireless devices
 VLSI. *See* Very-large-scale-integration
 Voltage controlled delay line, 157
 Voltage-controlled oscillator, 13, 253

W

WBET. *See* Wide bandwidth envelope tracking
 WCDMA band, features of, 6

- Wideband phase shifters, applications, 152
- Wideband radio receiver, usage of, 146–147
- Wide bandwidth envelope tracking, 19
- Wireless chip testing, IR-UWB transceiver systems, 92–93
- Wireless communication systems
 - FEC decoders, future wireless devices, 271–273
 - LDPC, 281–285
 - reed solomon, 273–274
 - turbo, 277–281
 - viterbi, 274–277
- 4G radio systems
 - ADCs specifications, 204–205
 - power considerations, topology exploration on, 205–217
 - reconfigurable DSM, 217–220
- IR-UWB
 - transceiver systems, architecture of, 92–93
 - transmitter designs, 93–101
- multimode radios, power efficient
 - reconfigurable baseband filters for circuit level implementation and validation, 231–234
 - design optimization and strategy, 228–231
 - g_m -C filter architecture and circuit design, 225–228
 - multimode reconfigurable g_m -C filters, 234–240
 - outcomes, 240–247
- multimode radio transceiver, 5 (*see also* Multimode radio transceiver, cellular applications)
 - polar transmitter, 15–23
 - receiver, 5–15
- multi-mode wireless receivers, adaptive
 - digital front-end, 249–250
 - adaptive multi-mode low-power DFE, concept and design of, 254–265
 - architecture of adaptive, 251–254
 - simulation outcomes for, 265–268
- multiple code families, decoders for, 285–286
 - LDPC and turbo, association of, 290–293
 - viterbi and turbo, combination of, 287–290
- multi-standard FEC decoding, 293–294
- multi-standard mobile radio, IIP2
 - improvement techniques, 169–170
 - calibration technique, 183–191
 - of GSM, 171–172
 - input stage, second order nonlinearity, 179–183
 - second order nonlinearity sources, 172–175
 - switching pairs, second order nonlinearity, 191
 - UMTS system, 172
- UWB
 - LNA design, 106–114
 - wireless and wireless links, 86–91
- UWB receivers
 - architecture of, 39–44
 - broadband measurements and de-embedding techniques, 44–46
 - design challenges, 31–37
 - design simulation outcomes, 49–51
 - evolution and state-of-the-art in, 37–39
 - receiver specifications in, 27–31
- UWB technology (*see* Ultra wideband)
 - DS-UWB systems, 71–78
 - LNA, 56–71
 - OFDM, frequency synthesis, 78–80
 - technology, 53–56
 - wireless interconnect, magnetic coupling transformer, 102–106
- Wireless handset applications, multi-mode power amplifiers, 121–122. *See also* Wireless communication systems
 - efficiency enhancement, 125–126
 - circuit design, 129–131
 - DDB, principle of, 126–128
 - power control and gain variation, 128–129
 - linearity improvement, 131–132
 - complex-gain digital predistortion, 134
 - practical issues, 135–138
 - predistortion algorithm, 133–134
 - system topology, 132
 - measurement outcomes, 138–141
 - power amplifier technologies, 122–125
- Wireless interconnect, magnetic coupling transformer, 102–106
- Wireless local area network, 54, 204
- Wireless personal area network, 88
- Wireless telecommunications market, 171–172
- WLAN. *See* Wireless local area network
- WPAN. *See* Wireless personal area network