Linköping Studies in Science and Technology Dissertation No. 1036

Multiband LNA Design and RF-Sampling Front-Ends for Flexible Wireless Receivers

Stefan Andersson



INSTITUTE OF TECHNOLOGY

Electronic Devices Department of Electrical Engineering Linköping University, SE-581 83 Linköping, Sweden

Linköping 2006

ISBN 91-85523-22-4 ISSN 0345-7524

Multiband LNA Design and RF-Sampling Front-Ends for Flexible Wireless Receivers

Stefan Andersson

ISBN 91-85523-22-4

Copyright © Stefan Andersson, 2006

Linköping Studies in Science and Technology Dissertation No. 1036 ISSN 0345-7524

Electronic Devices Department of Electrical Engineering Linköping University SE-581 83 Linköping Sweden

Author e-mail: stefan.h.andersson@gmail.com

Cover Image

Picture by the author illustrating an RF-sampling receiver on block level. The chip microphotograph represents a multiband direct RF-sampling receiver front-end for WLAN fabricated in 0.13 μ m CMOS.

Printed by LiU-Tryck, Linköping University Linköping, Sweden, 2006

I nådens år 2006! As my grandfather would have said.

Abstract

The wireless market is developing very fast today with a steadily increasing number of users all around the world. An increasing number of users and the constant need for higher and higher data rates have led to an increasing number of emerging wireless communication standards. As a result there is a huge demand for flexible and low-cost radio architectures for portable applications. Moving towards multistandard radio, a high level of integration becomes a necessity and can only be accomplished by new improved radio architectures and full utilization of technology scaling. Modern nanometer CMOS technologies have the required performance for making high-performance RF circuits together with advanced digital signal processing. This is necessary for the development of low-cost highly integrated multistandard radios. The ultimate solution for the future is a software-defined radio, where a single hardware is used that can be reconfigured by software to handle any standard. Direct analog-to-digital conversion could be used for that purpose, but is not yet feasible due to the extremely tough requirements that put on the analog-to-digital converter (ADC). Meanwhile, the goal is to create radios that are as flexible as possible with today's technology. The key to success is to have an RF front-end architecture that is flexible enough without putting too tough requirements on the ADC.

One of the key components in such a radio front-end is a multiband multistandard low-noise amplifier (LNA). The LNA must be capable of handling several carrier frequencies within a large bandwidth. Therefore it is not possible to optimize the circuit performance for just one frequency band as can be done for a single application LNA. Two different circuit topologies that are suitable for multiband multistandard LNAs have been investigated, implemented, and measured. Those two LNA topologies are: (*i*) wideband LNAs that cover all the frequency bands of interest (*ii*) tunable narrowband LNAs that are tunable over a wide range of frequency bands.

Before analog-to-digital conversion the RF signal has to be downconverted to a frequency manageable by the analog-to-digital converter. Recently the concept of direct sampling of the RF signal and discrete-time signal processing before analog-to-digital conversion has drawn a lot of attention. Today's CMOS technologies demonstrate very high speeds, making the RF-sampling technique appealing in a context of multistandard operation at GHz frequencies. In this thesis the concept of RF sampling and decimation is used to implement a flexible RF front-end, where the RF signal is sampled and downconverted to baseband frequency. A discrete-time switched-capacitor filter is used for filtering and decimation in order to decrease the sample rate from a value close to the carrier frequency to a value suitable for analog-to-digital conversion. To demonstrate the feasibility of this approach an RF-sampling front-end primarily intended for WLAN has been implemented in a 0.13 μ m CMOS process.

Preface

This Ph.D. thesis presents the results of my research during the period from February 2001 to September 2006 at the Electronic Devices group, Department of Electrical Engineering, Linköping University, Sweden. I started working on low-noise amplifiers and later continued with RF-sampling radio-receiver design. The following papers are included in the thesis:

- Paper 1: Stefan Andersson, Peter Caputa, and Christer Svensson, "A Tuned, Inductorless, Recursive Filter LNA in CMOS", in *Proceedings of the European Solid-State Circuit Conference (ESSCIRC)*, pp. 351-354, Florens, Italy, September 2002.
- Paper 2: Stefan Andersson and Christer Svensson, "An Active Recursive RF Filter in 0.35 μm BiCMOS", *Journal of Analog Integrated Circuits and Signal Processing* by Springer, pp. 213-218, vol. 44, no. 3, September 2005.
- Paper 3: Stefan Andersson and Christer Svensson "A 750 MHz to 3 GHz Tunable Narrowband Low-Noise Amplifier", in *Proceedings of the Norchip 2005 Conference*, pp. 8-11, Oulu, Finland, November 2005.
- Paper 4: Stefan Andersson, Christer Svensson, and Oskar Drugge, "Wideband LNA for a Multistandard Wireless Receiver in 0.18 μm CMOS", in *Proceedings of the European Solid-State Circuit Conference (ESSCIRC)*, pp. 655-658, Estoril, Portugal, September 2003.
- Paper 5: Rashad Ramzan, Stefan Andersson, Jerzy Dabrowski, and Christer Svensson, "Wideband LNA for a Multistandard RF-Sampling Front-End in 0.13 μm CMOS", *manuscript*.
- Paper 6: Stefan Andersson, Christer Svensson, "Channel Length as a Design Parameter for Low Noise Wideband LNAs in Deep Submicron CMOS Technologies", in *Proceedings of the Norchip 2004 Conference*, pp. 123-126, Oslo, Norway, November 2004.

- Paper 7: Stefan Andersson, Jacek Konopacki, Jerzy Dabrowski, and Christer Svensson, "SC Filter for RF Downconversion with Wideband Image Rejection", in *Proceedings of the ISCAS 2006 conference*, pp. 3542-3545, Kos, Greece, May 2006.
- Paper 8: Stefan Andersson, Jacek Konopacki, Jerzy Dabrowski, and Christer Svensson, "SC Filter for RF Sampling and Downconversion with Wideband Image Rejection", *Journal of Analog Integrated Circuits and Signal Processing* by Springer, special issue: MIXDES 2005, published online June 2006.
- Paper 9: Stefan Andersson, Jacek Konopacki, Jerzy Dabrowski, and Christer Svensson, "Noise Analysis and Noise Estimation of an RF-Sampling Front-End using an SC Decimation Filter", in *Proceedings of the MIXDES 2006 Conference*, pp. 343-348, Gdynia, Poland, June 2006.
- Paper 10: Stefan Andersson, Rashad Ramzan, Jerzy Dabrowski, and Christer Svensson, "Multiband Direct RF-Sampling Receiver Front-End for WLAN in 0.13 μm CMOS", *manuscript*.

The following publications related to this research project are not included in the thesis:

- Stefan Andersson, Jacek Konopacki, Jerzy Dabrowski, and Christer Svensson, "RF-Sampling Mixer for Zero-IF Receiver with High Image-Rejection", in *Proceedings of the MIXDES 2005 Conference*, pp. 185-188, Kraków, Poland, June 2006.
- Jerzy Dabrowski, **Stefan Andersson**, Jacek Konopacki, and Christer Svensson, "SC Filter Design for RF Applications", in *Proceedings of the ICSES 2006 Conference*, Lodz, Poland, September 2006.

During the period July to November 2004, I was on an internship at the Intel Communication Circuit Lab, Hillsboro, Oregon, USA. There I was working on RF sampling and decimation filters for RF-sampling front-ends. Parts of this work are described in the following publication and in two pending patent applications:

• Hasnain Lakdawala, Jing-Hong C. Zhan, Ashoke Ravi, **Stefan Andersson**, Brent R. Carlton, Richard B. Nicholls, Navid Yaghini, Ralph E. Bishop, Stewart S. Taylor, Krishnamurthy Soumyanath, "Multi-band (1-6 GHz) Sampled, Sliding-IF Receiver With Discrete-Time-Filtering in 90 nm Digital CMOS Process", in *Proceedings of the VLSI Symposium Conference*, Hawaii, USA, June 2006.

- Hasnain Lakdawala, Krishnamurthy Soumyanath, Stewart S. Taylor, and **Stefan Andersson**, "Discrete Time Filter having Gain for Digital Sampling Receivers", pending patent.
- Hasnain Lakdawala, Ashoke Ravi, Yorgos Palaskas, **Stefan Andersson**, and Krishnamurthy Soumyanath, "Filter with Gain", pending patent.

The following papers, falling outside the scope of this thesis, present other research topics I have been involved in during my Ph.D. studies:

- Peter Caputa, Henrik Fredriksson, Martin Hansson, **Stefan Andersson**, Atila Alvandpour, and Christer Svensson, "An Extended Transition Energy Cost Model for Buses in Deep Submicron Technologies", in *Proceedings* of the Fourteenth International Workshop on Power and Timing Modeling, Optimization and Simulation Conference, pp. 849-858, Santorini, Greece, November 2004.
- Ingvar Carlsson, **Stefan Andersson**, Sreedhar Natarajan, Aditya Sankar Me-dury, and Atila Alvandpour, "A High Density, Low Leakage, 5T SRAM for Embedded Caches", in *Proceedings of the European Solid-State Circuit Conference (ESSCIRC)*, pp. 215-218, Leuven, Belgium, September 2004.
- Stefan Andersson and Christer Svensson, "Direct Experimental Verification of Shot Noise in Short Channel MOS Transistors", *IEE Electronics Letters*, pp. 869-871, vol. 41, no. 15, July 2005.
- Christer Svensson, **Stefan Andersson**, and Peter Bogner, "On the Power Consumption of Analog-to-Digital Converters", *manuscript submitted to Norchip 2006*.
- Anton Blad, Christer Svensson, Håkan Johansson, and Stefan Andersson, "An RF-Sampling Radio Front-end Based on ΣΔ-Conversion", *manuscript* submitted to Norchip 2006.

I have also co-authored one book chapter:

• Christer Svensson and **Stefan Andersson**, "Software Defined Radio — Visions, Challenges and Solutions", chapter 3 in "Radio Design in Nanometer Technologies", Mohammed Ismail and Delia Rodríguez de Llera González, Eds., Springer, October 2006, ISBN 1402048238.

Contributions

The main contributions of this dissertation are as follows:

- Analysis and design of key circuits for multistandard receivers in CMOS.
- Implementation of wideband low-noise amplifiers in CMOS for wireless receivers.
- Implementation of widely tunable narrowband low-noise amplifiers. Tuned, inductorless LNAs are implemented using the concept of recursive filters which are electrically tuned over a large frequency range.
- A comprehensive study of switched-capacitor filters suitable for RF-sampling and decimation.
- A careful noise analysis and noise estimation of switched-capacitor decimation filters for RF-sampling front-ends.
- Design of an RF-sampling front-end consisting of a wideband low-noise amplifier and a switched-capacitor decimation filter with wideband image rejection in CMOS.

Abbreviations

AC	Alternating Current
ADC	Analog-to-Digital Converter
ASIC	Application-Specific Integrated Circuit
BER	Bit Error Rate
BiCMOS	Bipolar Complementary Metal Oxide Semiconductor
BJT	Bipolar Junction Transistor
BP	Band Pass
BPSK	Binary Phase-Shift Keying
CDMA	Code-Division Multiple Access
CMOS	Complementary Metal Oxide Semiconductor
СР	Compression Point
DAC	Digital-to-Analog Converter
DC	Direct Current
DAB	Digital Audio Broadcasting
DMB	Digital Multimedia Broadcasting
DSP	Digital Signal Processing
DSSS	Direct Sequence Spread Spectrum
DVB	Digital Video Broadcasting
EDGE	Enhanced GSM Evolution
EDR	Enhanced Data Rate
FIR	Finite-Impulse Response
FM	Frequency Modulation
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
HBT	Heterojunction Bipolar Transistor
HiperLAN	High Performance Radio Local Area Network
IC	Integrated Circuit
IF	Intermediate Frequency
IIR	Infinite-Impulse Response
IMD	Intermodulation Distortion
IO	Input Output

IIP3	Input-Referred Third-Order Intercept Point
IP3	Third-Order Intercept Point
I/Q	In-phase and Quadrature-phase
ISM	Industrial, Scientific, Medical
LNA	Low-Noise Amplifier
LO	Local Oscillator
LP	Low Pass
MIM	Metal-Insulator-Metal
MIMO	Multiple-Input Multiple-Output
MOSFET	Metal-Oxide-Semiconductor Field-Effect Transistor
NF	Noise Figure
NMOS	N-channel Metal-Oxide-Semiconductor
NMT	Nordic Mobile Telephone System
OFDM	Orthogonal Frequency-Division Multiplexing
OSR	Oversampling Ratio
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
PMOS	P-channel Metal-Oxide-Semiconductor
RF	Radio Frequency
SAW	Surface Acoustic Wave
SC	Switched-Capacitor
SDR	Software-Defined Radio
SFDR	Spurious Free Dynamic Range
SNDR	Signal-to-Noise and Distortion Ratio
SNR	Signal-to-Noise Ratio
SoC	System on a Chip
SOI	Silicon-On-Insulator
TDD	Time-Division Duplex
TDMA	Time-Division Multiple Access
TD-SCDMA	Time-Division - Synchronous Code-Division Multiple Access
T/H	Track-and-Hold
UMTS	Universal Mobile Telecommunication System
UWB	Ultra Wideband
VCO	Voltage-Controlled Oscillator
VGA	Variable Gain Amplifier
VLSI	Very Large Scale Integrated Circuits
WCDMA	Wideband Code-Division Multiple Access
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network
WPAN	Wireless Personal Area Network

Acknowledgments

It is with a lot of joy, as well as a bit of sadness, I have reached the point where it is time to summarize a few fantastic years. The work presented here is not only the result of many working hours, but also the result of a very inspiring and competent working environment. Therefore, perhaps this part of the thesis is the most important one where I get the chance to thank all the people who I have had the pleasure to meet and work with during these years. The following people have supported and encouraged me and deserve my deepest gratitude and many thanks:

- My advisor and supervisor Prof. Christer Svensson for giving me this opportunity, for his guidance, encouragement, and very inspirational way of doing research.
- Prof. Atila Alvandpour for the hint about a job at Ericsson Mobile Platforms and thereby indirectly sending me off to Lund. Who except him would let an RF designer take part in a memory design, and what a success!
- Anna Folkeson for help with all administrative stuff.
- Dr. Kalle Folkesson for proofreading this thesis, I cannot thank you enough for that. Now when I am moving to the south as well I expect you to teach me the local language down there called *Skånska*.
- Dr. Peter Caputa "The flying Doctor". My colleague for many years and coauthor of many lab- and solution-manuals. Thanks for all and interesting discussions both on-topic and off-topic, particularly about Ice Hockey.
- Lic. Eng. Martin Hansson for being a really good friend and for keeping me company in the summer of 2004 when we both were on internship at Intel in Oregon, USA. United we stand!
- M.Sc. Henrik Fredriksson for being my colleague as an undergraduate, while working for Ericsson Microelectronics, and now as a PhD student. How about moving south after your Ph.D.?

- Our Research Engineer Arta Alvandpour for solving all practical problems and always designing PCBs for me with a smile on this face.
- M.Sc. Rashad Ramzan for cooperation in chip design and measurements.
- All past and present members of the Electronic Devices group, especially Dr. Henrik Eriksson, Dr. Daniel Eckerbert, Professor Per Larsson-Edefors (the man who introduced me to VLSI design), Lic. Eng. Mindaugas Drazdziulis, Dr. Ulf Nordquist, Dr. Tomas Henriksson, Dr. Daniel Wiklund, Lic. Eng. Mikael Olausson, Dr. Håkan Bengtsson, M.Sc. Joacim Olsson, Lic. Eng. Behzad Mesgarzadeh, M.Sc. Timmy Sundström, Dr. Darius Jakonis, Dr. Ingemar Söderqvist, Dr. Mattias Duppils, Dr. Annika Rantzer, Ass. Prof. Jerzy Dabrovski, Naveed Ahsan, Adj. Prof. Aziz Ouacha, Isabel Ferrer, Rahman Aljasmi, Lic. Eng. Sriram Vangal, and Sreedhar Natarajan.
- Prof. Dake Liu, Dr. Eric Tell, and Lic. Eng. Anders Nilsson (the Radio Pirate) from the division of Computer Engineering.
- All the people at the Intel Communications Circuit Lab in Hillsboro, Oregon, USA, especially Krishnamurthy Soumyanath, Hasnain Lakdawala, Ashoke Ravi, Jing-Hong C. Zhan, Yorgos Palaskas, Richard Nicholls, Stewart Taylor, Mostafa Elmala, Brent Carlton, Gaurab Banerjee, Hossein Alavi, Navid Yaghini, Robert Wang, Ralph Bishop. Thanks for a wonderful time and for making my internship such a great experience.
- My fellow Intel interns Jeyanandh Paramesh, Charles Peach, Nebojsa Stanic, and Sunghyun Park. Thanks for many memorable moments. Hiking, CPR, Teriyaki food, all spiced up with a little knowledge in "Swedish".
- ACREO AB for great cooperation during the SoCTRix demonstrator project and for giving me a helping hand with chip measurements.
- Thanks to all friends and people I have had the pleasure to meet and work with that I could not fit in here.
- Last but certainly not least I would like to thank my parents Gun and Leif Andersson for all support and encouragement without really knowing what I am doing. My brother Johan Andersson for being a great friend whenever needed.

xvi

Stefan Andersson Linköping, September 2006

Contents

	Abs	tract		v
	Pref	face		vii
	Con	tributio	ons	xi
	Abb	oreviatio	ons	xiii
	Ack	nowled	gments	XV
Ι	Ba	ckgrou	ınd	1
1	Intr	oductio	n	3
	1.1	The Te	elecommunication Era	. 3
	1.2	The H	istory of Transistors and Integrated Circuits	. 4
	1.3	Introd	uction to the Expansion of Wireless Systems	. 7
	110	131	Wireless Standards	8
		1.3.2	Global-Range Wireless Systems	. 8
		1.3.3	Short-Range Wireless Systems	10
	1.4	Future	Challenges and Possibilities	. 13
	1.5	Motiv	ation and Scope of Thesis	. 15
	1.6	Refere	inces	. 16
2	Silic	on Tech	nology Development from an RF Design Perspective	19
	2.1	CMOS	S Technology	. 20
		2.1.1	The MOSFET Device	. 20
		2.1.2	Impact of Scaling on the MOSFET Device	. 22
		2.1.3	Silicon-on-Insulator Technology	. 25
	2.2	BiCM	OS Technology	. 27
		2.2.1	The Bipolar Device	. 27
		2.2.2	Impact of Scaling on the Bipolar Device	. 28

	2.3	Integration of Passives	28
	2.4	A Brief Device Comparison	31
	2.5	Technology Trends	32
	2.6	References	33
3	Rad	io-Receiver Architectures	35
	3.1	Receiver Specifications	35
		3.1.1 Receiver Requirements	36
		3.1.2 The Image Problem	38
	3.2	Superheterodyne Receivers	41
	3.3	Homodyne Receivers	43
	3.4	Low-IF Receiver	44
	3.5	Subsampling Receivers	45
	3.6	Digital Receivers	45
	3.7	Receiver Architectures — Summary and Trends	47
	3.8	References	48
4	Desi	gn of Low-Noise Amplifiers for Multiband Wireless Receivers	51
	4.1	Introduction to Multiband LNAs	51
	4.2	Low-Noise Amplifier Requirements and Performance Metrics	52
		4.2.1 Small-Signal Parameters	53
		4.2.2 Large-Signal Parameters	53
	4.3	LNA Design	55
		4.3.1 Design of Wideband LNAs	56
		4.3.2 Design of Widely Tunable LNAs	63
	4.4	Aspects on LNA Design and Noise in Deep Submicron CMOS	
		Technology	67
	4.5	References	69
5	RF-	Sampling Receivers	73
	5.1	Architectural Trends for Wireless Receivers	74
	5.2	Radio Receiver Front-Ends Utilizing RF Sampling	75
		5.2.1 Issues Related to the Sampling Process	76
		5.2.2 Noise in Switched-Capacitor Circuits	77
		5.2.3 Basic Circuit Techniques to Implement High-Frequency	
		SC Decimation Filters	78
	5.3	An RF-Sampling Receiver Front-End for WLAN in 0.13 μ m CMOS	84
		5.3.1 Circuit Description	84
		5.3.2 Required ADC Performance	85
		5.3.3 Preliminary Measurement Results	87
	5.4	References	89

6	Con	clusions and Future Work	93
	6.1	Conclusions	93
	6.2	Future Work	94

Due copyright restrictions the articles are removed but links has been provided to each article if published electronically.

II Papers

97

7	Pape	er 1 99
	7.1	Introduction
	7.2	Active Recursive Filter Approach to the LNA 100
	7.3	Measured Performance
	7.4	Discussion
	7.5	Acknowledgments
	7.6	References
8	Раре	er 2 109
	8.1	Introduction
	8.2	Design Theory
	8.3	Circuit Design
	8.4	Simulation and Measurement Results
	8.5	Discussion and Conclusion
	8.6	References
9	Pape	er 3 119
	9.1	Introduction
	9.2	Design Theory
	9.3	Circuit Design
	9.4	Simulation Results
	9.5	Measurement Results
	9.6	Discussion and Conclusion
	9.7	Acknowledgments
	9.8	References
10	Pape	er 4 131
	10.1	Introduction
	10.2	Design of the LNA Input Stage
	10.3	Design of the LNA Output Stage
	10.4	Measured Performance and Discussion
	10.5	Conclusion
	10.6	Acknowledgments
	10.7	References

11	Pape	er 5	143
	11.1	Introduction	144
	11.2	LNA Input Stage	145
	11.3	LNA Output Stage	148
	11.4	Measured and Simulated Performance	149
	11.5	Conclusions	151
	11.6	References	151
12	Pape	er 6	153
	12.1	Introduction	154
	12.2	Noise Modeling	155
	12.3	Noise Measurements	155
	12.4	Design Considerations	159
	12.5	Design Example	160
	12.6	Conclusion	163
	12.7	Acknowledgments	163
	12.8	References	163
13	Pape	er 7	165
	13.1	Introduction	166
	13.2	Filter/Decimator Prototype	167
	13.3	Filter Implementation	169
		13.3.1 Filter Structure	169
		13.3.2 Buffer Design	172
		13.3.3 DC-Offset Cancellation Loop	173
		13.3.4 Noise Characterization	174
	13.4	Conclusions	174
	13.5	References	175
14	Pape	er 8	177
	14.1	Introduction	178
	14.2	Filter/Decimator Prototype	180
	14.3	Filter Implementation	182
		14.3.1 Filter Structure	183
		14.3.2 Amplifier Design	185
		14.3.3 DC-Offset Cancellation Loop	186
		14.3.4 Noise Characterization	187
	14.4	Summary and Conclusions	192
	14.5	References	192

15	Pape	er 9	195
	15.1	Introduction	196
	15.2	RF-Sampling Filter Topology	197
	15.3	Noise in Simple SC Circuits	198
	15.4	Noise in FIR Filter Section	201
	15.5	Amplifier Related Noise	203
	15.6	Noise Analysis of Complete RF FIR Filter	207
	15.7	Summary and Conclusions	209
	15.8	The Authors	210
	15.9	References	210
16	Pape	er 10	213
	16.1	Introduction	214
	16.2	RF-Sampling Front-End Implementation	215
		16.2.1 Wideband LNA	217
		16.2.2 Decimation Filter Topology	217
		16.2.3 Decimation Filter Noise	219
		16.2.4 Required ADC Performance	220
	16.3	Preliminary Measurement Results	222
	16.4	Summary and Conclusions	224
	16.5	References	225
III	[A	ppendix	227

A	Tran	sistor Properties and Equations		229
	A.1	The Bipolar Transistor	. 2	229
	A.2	The MOS Transistor	•	231
	A.3	References	•	234

Part I Background

Chapter 1

Introduction

"You see, wire telegraph is a kind of a very, very long cat. You pull his tail in New York and his head is meowing in Los Angeles. Do you understand this? And radio operates exactly the same way: you send signals here, they receive them there. The only difference is that there is no cat."

Albert Einstein, when asked to describe radio.

1.1 The Telecommunication Era

Different techniques for long-distance communication have always been of large interest to humans, no matter if it is a way to achieve important information or just getting the latest gossip. It was not that long ago people were forced to rely on techniques like smoke signals, mirrors, jungle drums, and carrier pigeons. With the knowledge about electricity, electrical signals, and electromagnetic wave propagation new technical solutions for long-distance communication were invented. It all started with the early telegraph¹. Almost 40 years later the first telephone² was invented. I think it is fair to say that telephony has had a great impact on peoples lives since then. In parallel with the telegraph and the telephone, who both require a wire, another technique was developed. The use of electromagnetic waves transmitting data invisibly through the air. Guglielmo Marconi was the first to prove the feasibility of radio communication in 1895. In 1902 the first successful transatlantic radiotelegraph message was sent.

More than 100 years after the invention of the telephone it was time for the next revolution of the telecommunication area when the mobile phones were introduced. The first analog standard in Scandinavia, NMT450 (Nordic Mobile Tele-

¹First electric telegraph 1837.

²Patented 1876 by G. Bell.

phone System) was soon followed up by the worldwide GSM standard³ widely used today. In less than 10 years more or less every citizen got a GSM phone and right now the third generation (3G) is breaking through the ice. One should of course not forget all wireless short-range standards like Bluetooth, WLAN, DECT, etc. that also are commonly used today.

However, this tremendous expansion within the wireless area would never have taken place if it was not for the great invention of the component called *transistor*. One can of course argue about the importance of different inventions, but one thing is for sure: Transistors do have a great impact on our lives whether we know they exist or not. The invention of the transistor was also the beginning of one of the largest and most successful business areas of today, i.e. *the semiconductor industry*.

1.2 The History of Transistors and Integrated Circuits

Before the invention of the transistor electromagnetic switches and vacuum tubes were used instead. The electromagnetic switch worked excellently as a switch, but since it was mechanical it was slow. It took about a thousand of a second to open and close. Its competitor, the vacuum tube could work both as a switch and amplifier, and since electrons can travel at high speed in vacuum it could operate at high frequencies. However, the vacuum tubes used considerable standby power and their lifetime was limited. Trying to replace the vacuum tubes finally lead to the invention of the transistor.

The birth of solid-state electronics was already in 1874 when the scientist Ferdinand Braun discovered the first metal-semiconductor contact. The first transistor was described already in 1925 in a patent by Lilienfeld. It was a field-effect transistor where the conductivity could be altered by applying a voltage to a poorly conducting material, in other words a semiconductor. Interestingly, the field-effect transistor was discovered more than 20 years earlier than the bipolar transistor. But, due to the lack of good semiconductor materials and manufacturing difficulties the development of field-effect transistors was delayed for many years until the 60s. The first working transistor (Fig. 1.1) was invented in 1947 by Bardeen, Brattain, and Shockley at Bell telephone laboratories. It was a point-contact transistor, i.e. a primitive type of a bipolar transistor. In fact, it was discovered by accident while trying to build a field-effect transistor. The name *transistor* was suggested about a half year later and the story behind the naming can be read in

³In 1990 the first GSM specification is born with over 6000 pages of text. Commercial operation started in 1991.



Figure 1.1: The first working transistor from 1947.

[1]. The three inventors of the bipolar transistor later received the Nobel Prize for physics⁴ in 1956.

In 1958 Jack Kilby, working for Texas Instruments, built the first integrated circuit (IC). He used germanium with etched mesa structures to separate the components and connected them electrically using bond wires of gold. The technique was patented in 1959 [2], and in 2000 he was awarded the Nobel Prize in physics⁵ for his breakthrough discovery. A couple of years earlier Shockley had started his own company called Shockley Semiconductors in what later became Silicon Valley⁶ in California. In 1957 eight of his researchers⁷ resigned and started Fairchild Semiconductors. Among those eight were Robert Novce and Gordon. E. Moore who later founded Intel. The year after Kilby demonstrated the first IC Noyce fabricated the first IC with planar interconnects using photolitography and etching techniques. This way of manufacturing ICs was the same as is used today. It is also worth noting that Intel competitor AMD was also founded by engineers from Fairchild Semiconductors. Among them Eugene Kleiner, one of the eight who once left Shockley Semiconductors. The first ICs used bipolar transistors and at the time the bipolar device was considered as THE transistor. The first Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET) was fabricated

⁴Unfortunately there is no Nobel Prize for electronics.

⁵At the same time he gave a speech at the Royal Institute of Technology in Stockholm and I was there listening to him.

⁶"Shockley is the man who brought silicon to Silicon Valley", [3].

⁷Shockley later named them "the Traitorous eight".



Figure 1.2: Intel duo-core processor containing 1.72 billion transistors.

by Khang at Bell Labs in 1960. However, it would take another 10 years before the manufacturing problems had been resolved and the era of field-effect transistors really started. The Complementary-MOS (CMOS) process was invented in 1963 by Wanlass at Fairchild Semiconductor. He found that CMOS shrank standby power by six orders of magnitude compared to similar gates using bipolars or PMOS transistors [4]. In 1965 Gordon Moore, when still working for Fairchild Semiconductor, wrote a paper Cramming more components onto Integrated circuits [5]. There he made the prediction that the number of devices on an IC will double every 12 months, which was later revised to a doubling every 24 months instead. This is the famous *Moore's law* which still have a great impact on the semiconductor business today. The way I see it *Moore's law* is not related to pure technical limitations and improvements, it's rather a driving force for the whole semiconductor business to improve at a regular pace. From the first working single transistor in 1947 to the 1.72 billion transistors on one of Intels duo-core processors presented in 2006 [6] (Fig. 1.2) in less than 60 years. Today even single ICs containing analog-, RF-, and digital-circuits all on the same die are quite common. It is most likely better to talk about whole systems on chip nowadays rather than single circuits. Many, more clever persons than I, have speculated about how long this development will continue. I just think and hope it will continue for many many years more.

More details about the history of the transistor and integrated circuits can be found in several published papers and books as well as on the home pages of companies like Intel [7] and Fairchild semiconductor [8]. Vol. 86, issue 1 of *Proceedings of the IEEE* contains papers all related to the history of transistors and ICs. Among these paper are *The Invention of the Transistor* [9], *The Naming*

of the Transistor [1], and Cramming More Components onto Integrated Circuits [5].

1.3 Introduction to the Expansion of Wireless Systems

Wireless communication is far from a new phenomena. As discussed earlier the knowledge how to transfer voice and data through the air using radio has been known for more than a hundred years. In the beginning the wireless communication technology was mostly used for broadcasting to a large audience rather than point-to-point communication like for example the telephone. Even though radar and radio communication was used very early for military purpose, the real break through for wireless communication started when mobile communication became available to the civil market. Wireless communication is today a very large and important market affecting our lives in many ways. Today in our part of the world almost everyone can be reached by mobile phone whenever and wherever he or she is. Moreover, due to the huge expansion of Internet, wireless transfer to and from handheld devices and computers has become another extremely important area. Wireless systems are today developing very fast. This can also be seen from the huge number of new companies created within this area.

The main driving forces behind the evolution of mobile terminals from a customers point of view is:

- New applications
- Increased and improved functionality
- Lower cost
- Miniaturization

but how is this going to be accomplished? Of course the market is the main driver for the industry. To fulfill the market needs both architectural and technology improvements are necessary.

The number of wireless standards are increasing for every year and higher data rates are requested by even more users than before. To make this possible a numerous number of different standards and technologies are used. A few examples are GSM, WCDMA, Bluetooth, and WLAN (802.11a,b,g...) etc. With the increased number of wireless standards used today it is no longer reasonable that customers should have one handset for each standard. From a customer's perspective one handset capable of switching between all standards is the optimal

solution. The simplest, and most straight forward, solution would of course be to include one chip-set for each standard into for example a mobile phone or a laptop. This would however also lead to a highly increased cost and a very bulky terminal with short standby time due to high power consumption. The smallest, cheapest, and most elegant solution would be to use a multiband RF front-end capable of covering all wanted standards, i.e. a software-defined radio.

1.3.1 Wireless Standards

The number of systems that use radio links is increasing quickly. At the same time, the number of standards for such systems is increasing very quickly as well [10]. To make this possible the number of frequency bands dedicated for wireless communication (voice and data) has also been increased. This applies for both licensed and unlicensed frequency bands. However, the frequency spectra is limited and has to be used as efficiently as possible. One, and maybe the most important, thing is to transfer as much data per unit bandwidth as possible. Another important thing to address is the difference in standards and frequency bands used in Europe, North America, and Japan. Making a single product for all these markets handling all the differences is not an easy task, particularly when different frequency bands are used.

To distinguish between different standards they are here divided into globalrange and short-range communication standards. Global range is for example cellular systems while short range is wireless up to about 100 meters as is the case for DECT, WLAN, and Bluetooth etc. Fig. 1.3 illustrates the range, coverage, and mobility versus bandwidth for various wireless systems. The solid line basically represents the "Shannon bound" of communication theory [11]. This limit is thus set by the minimum amount of energy needed for reliable transfer of every bit of information (assuming constant transmit power).

1.3.2 Global-Range Wireless Systems

Cellular systems have seen three generations of evolution and the next generation is already being considered. The first generation (1G) cellular systems used analog frequency modulation (FM) and were operating in frequency bands around 450 MHz and 900 MHz. Different standards were used in Europe, USA, and Japan making phones incompatible and prohibiting efficient roaming [12]. The main problem with 1G was the lack of available spectrum needed when the number of users increased. This first generation is being phased out and here in Sweden it is already closed down.

The second generation (2G) was introduced in the early 1990s. This was also a transfer from traditional analog modulation schemes to digital ones. Technology



Figure 1.3: Range/Coverage/Mobility - Bandwidth relationship [11].

scaling and low-cost compact digital processing techniques made this transfer possible. New methods like time-division multiple access (TDMA) and code-division multiple access (CDMA) could be used for more efficient multiple access of many users [13]. TDMA means the same band is available to many users but at different time slots (the channel is divided up in time). CDMA on the other hand is a form of multiplexing. It encodes the data with a special code associated with each channel and uses the constructive interference properties of the special codes to perform the multiplexing. Both TDMA and CDMA have a number of advantages [14]. Among them:

- Better sharing of the available frequency spectra among multiple users.
- The power consumption can be reduced in radio transmission.
- Increased flexibility for both voice and data.
- Better security since the transmission is encrypted.

GSM is in many ways an international standard and is today used in almost 200 countries. Originally GSM used the 900 MHz spectrum, but Europe, Africa, and Asia later added additional capacity at 1800 MHz. In North America the frequency bands 800 MHz and 1900 MHz are used instead. Luckily, most phone manufacturers are today able to offer products that can be used anywhere GSM

systems are found, i.e. tri-band (900, 1800, 1900 MHz) or quad-band (800, 900, 1800, 1900 MHz) phones.

While moving from 2G to 3G (the third generation) the so called 2.5G was introduced. This is essentially an upgraded version of 2G. EDGE (Enhanced Data GSM Evolution) and GPRS (General Packet Radio Service) both allow higher data rates while using the GSM network.

The third generation (3G) has higher data rate (2 Mbps [15]) than GSM (14.4 kbps [15]). Therefore, it is mainly intended for applications like video telephony, Internet, email, and instant messaging rather than traditional voice calls even if that of course is possible. 3G is defined by a worldwide standard, International Mobile Telecommunications Standard IMT-2000 [16]. In this standard a number of different systems are defined. Among them the WCDMA (Wideband Code-Division Multiple Access) technology used in Europe and Japan. In Europe the 3G standard is also known as UMTS (Universal Mobile Telecommunication System). In China another technology called Time Division - Synchronous Code Division Multiple Access (TD-SCDMA) is used instead, while North and South America use CDMA2000. What the fourth generation (4G) will be is still not clear. However, higher bandwidths and higher data rates will clearly be required as well as handover between different wireless systems.

Then there is the whole entertainment business. Receiving digital radio, TV, and video is a must for future multi-media terminals (including mobile phones). Also here a number of standards exist. Among them:

- Digital Multimedia Broadcasting (DMB)
- Digital Audio Broadcasting (DAB)
- Digital Video Broadcasting (DVB)
- Integrated Services Digital Broadcasting (ISDB)
- Satellite radio

It is easy to realize the difficulties in order to combine all these different standards into one single phone. Including other functionalities as FM radio and Bluetooth for short-range communication do not make it any easier. There is obviously a huge need for efficient multistandard receivers to meet market requirements.

1.3.3 Short-Range Wireless Systems

Typical for short-range wireless systems, also referred to as wireless data systems, is that they offer higher data rates and shorter range of wireless coverage. Typical applications today are handsfrees and wireless connections for personal computers. A number of different standards and flavors of standards exist:

- Bluetooth
- IEEE 802.11 (Wi-Fi)
- HiperLAN
- Ultra WideBand (UWB)
- WiMAX, IEEE 802.16
- HIPERMAN

Bluetooth⁸ [17] is a low-cost low-power technology for wireless personal area networks (WPANs), originally intended for cable replacement and is commonly used in handsfrees. It is geared towards voice and data applications and operates in the unlicensed 2.4 GHz spectrum (the 2.4 GHz ISM band). Several classes exist that can operate over a distance of 10-100 m. The peak data rate with EDR (Enhanced Data Rate) is 3 Mbps.

The IEEE 802.11 [18] is a WLAN standard targeted for a number of different data rates and many flavors of the standard exists. The most commonly mentioned are:

- 802.11a that operates in the unlicensed 5 GHz band. It uses OFDM and has a maximum data rate of 54 Mbps.
- 802.11b that operates in the unlicensed 2.4 GHz band with a maximum data rate of 11 Mbps using DSSS. 802.11b is the original Wi-Fi standard.
- 802.11g that operates in the unlicensed 2.4 GHz band. It uses OFDM and has a maximum data rate of 54 Mbps. It is also backwards compatible with 802.11b.
- 802.11n is expected to work in the unlicensed 5 GHz band. The use of higher bandwidths than the others and MIMO (Multiple Input Multiple Output) techniques will offer a maximum data rate of over 100 Mbps.

A part of the technical receiver requirements for IEEE 802.11a are summarized in Tab. 1.1 and the frequency channel plan for the USA is shown in Fig. 1.4.

The European Telecommunications Standards Institute (ETSI) has adopted the High Performance Radio Local Area Network (HiperLAN) standard [19] for WLAN. HiperLAN2 use the 5 GHz band has has similar performance as 802.11a.

⁸The name Bluetooth comes from the Danish King Harald Bluetooth, who unified Denmark and Norway in the 10^{th} century.

Data Rate	Constellation	Minimum Sensitivity	Required SNR
[Mbps]		[dBm]	[dB]
6	BPSK	-82	4.2
9	BPSK	-81	5.5
12	QPSK	-79	7.5
18	QPSK	-77	8.8
24	16-QAM	-74	12.9
36	16-QAM	-70	14.8
48	64-QAM	-66	19.3
54	64-QAM	-65	20.1

Table 1.1: Technical specifications for IEEE 802.11a.



Figure 1.4: OFDM frequency channel plan for IEEE 802.11a in the USA.

It also seems like the IEEE 802.11 standards will drive HiperLAN out of the market.

Ultra Wideband (UWB) use a different approach. Transmission of digital data is made over a wide spectrum of frequency bands (3.1-10.6 GHz) with very low power [20]. To date, UWB only has regulatory approval in the USA and two

competing standards make the situation complicated. The UWB Forum [20] is promoting one standard based on Direct Sequence (DS-UWB) and the WiMedia Alliance [21] is promoting another standard based on OFDM. Each standard allows for data rates up to 500 Mbps at a range of 2 m and a data rate of approximately 100 Mbps at a range up to 10 m.

WiMAX stands for Worldwide Interoperability for Microwave Access and is a wireless metropolitan area network (MAN) technology [22]. WiMAX is actually a medium-range communication technology rather than a short-range standard. WiMAX has a maximum range of about 50 km with data rates of 70 Mbps. However, a typical cell has a much smaller range. WiMAX (IEEE 802.16a) operates in the 2-11 GHz frequency bands. There is also a fixed wireless access standard called HIPERMAN developed by ETSI in Europe. HIPERMAN also operates in the spectrum between 2-11 GHz and is compatible/interoperable with the IEEE 802.16a standard. However, there has been delays in regulatory approval in Europe due to issues regarding the use of spectrums in the 2.8 GHz and 3.4 GHz range. WiMAX is heavily supported by the computer industry and Intel has been one of the main drivers. It is created to compete with DSL and cable modem access. WiMAX technology is also considered ideal for rural, hard to wire areas.

Even though many different standards exist and many more are on the way, Bluetooth and the IEEE 802.11a,b,g standards are the commonly used for shortrange communication today. As the frequency spectra below 10 GHz is expected to get extremely crowded within the near future, the work on using new frequency spectra at several tenths of GHz is ongoing. One suggested frequency spectra is in the 60 GHz band [23]. The most difficult challenge for the future will be to combine a large set of both global-range and short-range standards into the mobile phone or the computer platform.

1.4 Future Challenges and Possibilities

The engine behind the market growth in the wireless communication area is the availability of cheap RF ICs. The only technologies available that are suitable for high-level integration low-cost chipsets are CMOS and BiCMOS. Fig. 1.5 shows an application spectrum and semiconductor devices likely to be used in that frequency range today. The boundaries between the kinds of RF technologies (Si, SiGe, GaAs, and InP) shown in Fig. 1.5 are diffuse and strongly related to the manufacturing cost and therefore change with time. The obvious reason why these borders change with time is the ongoing scaling. Therefore it is reasonable to assume that silicon technologies will take over much of the market shares held by GaAs technologies in the near future. Since Dr. Gordon E. Moore stated his famous law (Moore's law) [5] almost forty years ago the technology scaling has



Figure 1.5: Application spectrum and semiconductor devices likely to be used today [24].

been the main factor behind the semiconductor evolution. According to *ITRS* [24] this scaling will continue for quite many years. With the scaling, RF CMOS will offer great possibilities for integration of complete receivers utilizing carrier frequencies far into the mm-wave range. Using RF CMOS, the possibility to develop and mass produce low-cost chipsets for wireless applications seems very promising for the future. The new generation of radio tranceivers for WLAN is almost exclusively in CMOS and GSM solutions also exist [25, 26].

One of the most thrilling challenges for the future will be to find new receiver architectures suitable for highly integrated multiband receivers. Wireless multistandard full CMOS SoCs [27, 28] are already a reality [29], but to date complete radio solutions are in most cases developed as a chipset composed of several independent chips [30]. One chip each for radio frequency (RF), analog baseband (ABB), power management (PM), and digital baseband (DBB) together with a front-end module (FEM) consisting of duplexers or transmit/receive (T/R) switches and finally a power amplifier (PA). When moving towards multistandard applications a high level of integration of radio functions thus becomes a necessity. This can only be accomplished by improved radio architectures and utilization of the technology scaling. At the same time new applications require large signalprocessing capabilities together with a high level of memory integration. This can only be fulfilled using high-end CMOS technology. Even though a multistandard radio front-end is usually considered as the most difficult design task, there is also a tremendous need for area-efficient programmable multistandard baseband processors [31, 32, 33]. Single chip solutions (SoC) containing RF, analog-, and digital-baseband have potential to be low cost and with a small form factor.

Software-defined radio (SDR) would be the ideal solution to achieve a totally
reconfigurable radio. However, implementing multistandard or reconfigurable radio is not that easy. One of the key issues with SDR is early digitization, ideally at the antenna. This puts incredibly tough requirements on the analog-to-digital converter (ADC). Usually these requirements cannot be met and the ADC is the bottleneck in SDR. Therefore, a compromise between flexibility and the usage of parallel receivers has to be made. Never the less, by using new RF front-end architectures that take advantage of process scaling and by using DSP horsepower to ease analog and RF design (for example calibration and digital error correction) a lot more flexible radio chips can be made. A lot remains to be done in this area and new innovative solutions are most needed.

1.5 Motivation and Scope of Thesis

The wireless market is developing very fast today. An increasing number of users and the constant need for higher and higher data rates have led to an increasing number of emerging wireless communication standards as seen in Section 1.3. At the same time consumer electronics have become very cost sensitive. As a result there is a huge demand for flexible and low-cost radio architectures for portable applications. Moving towards multistandard radio, a high level of integration becomes a necessity and can only be accomplished by new improved radio architectures and full utilization of technology scaling. Modern nanometer CMOS technologies have the required performance for making high-performance RF circuits together with advanced digital signal processing. This is necessary for the development of low-cost highly integrated multistandard radios.

One of the key components in such a radio front-end is a multiband multistandard low-noise amplifier (LNA). The LNA must be capable of handling several carrier frequencies within a large bandwidth. Therefore it is not possible to optimize the circuit performance for just one frequency band as can be done for a single application LNA. It is also necessary to minimize the number of passive components like inductors to reduce area and cost. Two different circuit topologies that are suitable for multiband multistandard LNAs have been investigated, implemented, and measured. Those two LNA topologies are tunable narrowband LNAs and wideband LNAs. In **Paper 1-3** the concept of active recursive filters is used for implementing tunable LNAs and filters. The three different circuit implementations described in those papers also demonstrate the feasibility of implementing tuned circuits without the use of inductors. Paper 1 shows an implementation of a recursive filter LNA in 0.8 μ m CMOS and Paper 3 describes an implementation of a widely tunable narrowband LNA in 0.18 μ m. Both LNAs show excellent frequency tunability and could be tuned over a wide frequency range. In Paper 2 a recursive filter is implemented in a 0.35 μ m SiGe BiC-

MOS technology. This bipolar implementation also shows excellent frequency tunability but at higher frequencies, 6-10 GHz. In **Paper 4** and **Paper 5** two implementations of wideband LNAs for WLAN (802.11a,b,g,...) are demonstrated. Elementary wideband amplifiers show a severe trade-off between noise figure and input impedance matching. In **Paper 4** and **Paper 5** we show a way to decouple the requirement on input matching from the overall noise properties of a wideband LNA. **Paper 6** address the increased noise observed for short channel lengths in deep submicron technologies, and how the channel length can be utilized as a design parameter when optimizing the noise figure for wideband LNAs.

Recently the concept of direct sampling of the RF signal and discrete-time signal processing before analog-to-digital conversion has drawn a lot of attention. Today's CMOS technologies demonstrate very high speeds, making the RF-sampling technique appealing in a context of multistandard operation at GHz frequencies. Once the signal is sampled and downconverted it has to be decimated before analog-to-digital conversion. A discrete-time switched-capacitor filter is used for filtering and decimation in order to decrease the sample rate from a value close to the carrier frequency to a value suitable for analog-to-digital conversion. There are two essential design aspects to consider. First of all images introduced during decimation have to be suppressed and secondly the noise has to be low. In **Paper 7** and **Paper 8** a decimation filter are discussed in **Paper 8** and **Paper 9**. Finally, to demonstrate the feasibility of this approach an RF-sampling front-end primarily intended for WLAN has been implemented in a 0.13 μ m CMOS process. This RF-sampling front-end is described in **Paper 10**.

1.6 References

- [1] J. Pierce, "The Naming of the Transistor," *Proc. of the IEEE*, vol. 86, January 1998.
- [2] J. Kilby, "Minituarized Electronic Circuits," U.S. Patent 3 138 743, 1959.
- [3] Silicon Valley, http://www.stanford.edu/dept/news/pr/02/shockley1023.html, 2006.
- [4] F. Wanlass, "Low Stand-By Power Complementary Field Effect Circuitry," U.S. Patent 3 356 858, 1967.
- [5] G. E. Moore, "Cramming More Components onto Integrated Circuits," *Electronics*, vol. 38, no. 8, 1965.

- [6] S. Naffziger, B. Stackhouse, and T. Grutkowski, "The Implementation of a 2-core Multi-Threaded Itanium-Family Processor," in *Proc. of the ISSCC Conf.*, pp. 182–183, 2005.
- [7] Intel, http://www.intel.com, 2006.
- [8] Fairchild Semiconductor, http://www.fairchildsemi.com, 2006.
- [9] I. Ross, "The Invention of the Transistor," vol. 86, January 1998.
- [10] M. Vidojkovic, J. van der Tang, P. Baltus, and A. van Roermund, "Design of Flexible RF Building Blocks — A Method for Implementing Configurable RF Tranceiver Architectures," in *Proc. of Asia-Pacific Conference on Communications*, pp. 445–449, 2005.
- [11] M. Ismail and D. Rodríguez de Llera González, eds., *Radio Design in Nanometer Technologies*. Springer, 2006.
- [12] H. Elwan, H. Alzaher, and M. Ismail, "A New Generation of Global Wireless Compatibility," *IEEE Circuits and Devices Magazine*, vol. 17, pp. 7–19, 2001.
- [13] B. Razavi, RF Microelectronics. Prentice Hall, 1998.
- [14] J.E Padgett, C.G. Günther, and T. Hattori, "Overview of Wireless Personal Communications," *IEEE Communications Magazine*, vol. 33, no. 1, pp. 28– 41, 1995.
- [15] L. Larson, "Silicon Technology Tradeoffs for Radio-Frequency/Mixed-Signal "Systems-on-a-Chip"," *IEEE Tran. on Electron Devices*, vol. 50, pp. 683–699, March 2003.
- [16] IMTS, http://www.itu.int/home/imt.html, 2006.
- [17] Bluetooth website, http://www.bluetooth.com, 2006.
- [18] IEEE website, http://www.ieee.org, 2006.
- [19] ETSI website, http://www.etsi.org, 2006.
- [20] UWB Forum, http://www.uwbforum.org, 2006.
- [21] WiMedia Alliance, http://www.wimedia.org, 2006.
- [22] WiMAX Forum, http://www.wimaxforum.org, 2006.

- [23] S. Moore, "Cheap Chips for Next Wireless Frontier," *IEEE Spectrum*, pp. 8–9, June 2006.
- [24] ITRS, http://public.itrs.net, 2006.
- [25] A.A. Abidi, "RF CMOS Comes of Age," *IEEE Journal of Solid State Circuits*, vol. 39, no. 4, pp. 549–561, April 2004.
- [26] K. Muhammad, Y-C. Ho, T. Mayhugh, C.M. Hung, T. Jung, I. Elahi, C. Lin, I. Deng, C. Fernando, J. Wallberg, S. Vemulapalli, S. Larson, T. Murphy, D. Leipold, P. Cruise, J. Jaehnig, M.C. Lee, R.B. Staszewski, R. Staszewski, and K. Maggio, "Discrete Time Quad-band GSM/GPRS Receiver in a 90nm Digital CMOS Process," in *Proc. of the CICC 2005 Conf.*, pp. 809–812, 2005.
- [27] Z. Zu, S. Jiang, Y. Wu, H. Jian, G. Chu, K. Ku, P. Wang, N. Tran, Q. Gu, M.-Z. Lai, C. Chien, M.F. Chang, and P.D. Chow, "A Compact Dual-Band Direct-Conversion CMOS Tranceiver for 802.11a/b/g WLAN," in *Proc. of the ISSCC Conf.*, pp. 98–99, 2005.
- [28] T. Maeda, T. Yamase, T. Tokairin, S. Hori, R. Walkington, K. Numata, N. Matsuno, K. Yanagisawa, N. Yoshida, H. Yano, Y. Takahasni, and H. Hida, "A Low-power Dual-Band Triple-Mode WLAN CMOS Tranceiver," in *Proc. of the ISSCC Conf.*, pp. 100–101, 2005.
- [29] J. van der Tang, H. van Rumpt, D. Kasperkovitz, and A. van Roermund, "RF Building Blocks and Entertainment SoCs for Mobile Telecommunication Platforms," in *Proc. of Asia-Pacific Conference on Communications*, pp. 440–444, 2005.
- [30] K. Muhammad, R.B. Bogdan, and D. Leipold, "Digital RF Processing: Toward Low-Cost Reconfigurable Radios," *IEEE Communications Magazine*, pp. 105–113, August 2005.
- [31] A. Nilsson, Design of Multi-Standard Baseband Processors. Lic. thesis, Linköpings universitet, Department of Electrical Engineering, SE-581 83 Linköping, Sweden, 2005.
- [32] E. Tell, Design of Programmable Baseband Processors. PhD thesis, Linköpings universitet, Department of Electrical Engineering, SE-581 83 Linköping, Sweden, 2005.
- [33] Coresonic AB, http://www.coresonic.com, 2006.

Chapter 2

Silicon Technology Development from an RF Design Perspective

Ever since Dr. Gordon E. Moore stated his famous law (Moore's law) [1] almost forty years ago, the technology scaling has been the main factor behind the semiconductor evolution. According to *ITRS* [2] this scaling will continue for quite many years more. It can of course not continue forever due to the laws of physics. Many attempts have been made this far to predict where it all ends, but where exactly is the limit [3, 4]? It might not at all be the technical limit that ends the scaling. Increased manufacturing cost is a problem that comes along with scaling and is a possible show-stopper for the process scaling to continue.

Essentially there are three silicon technologies for integration of RF circuits. CMOS (bulk CMOS), BiCMOS where both MOS- and bipolar-transistors are available, and the third one is Silicon on Insulator (SOI) where the MOS transistors are built on an insulator (usually sapphire or an oxide layer). CMOS and Si/SiGe BiCMOS are the two dominant process technologies used for RF transceivers today and will remain to be so in the nearest future [2]. Today, BiC-MOS has the biggest share in terms of volume compared to CMOS in cellular transceivers. That may however change in the near future. Using a logic CMOS process is not always possible for RF applications. Adding an RF option (inductors and MIM capacitors) to a logic CMOS process typically adds both technology delay, about one year compared to logic, and cost. However, RF CMOS is still considered a cheaper technology than BiCMOS when both analog and digital parts are implemented on the same die. The main drawback with BiCMOS is that the MOS-part lags pure CMOS technology by at least one to two generations. Therefore it is not suitable to use BiCMOS except for the RF part. Some of the main things to address to reduce the cost of the RF parts in the future is to use low-cost RF-compatible technologies (CMOS is a good example), robust designs that are tolerant against process variations, and minimize the number of inductors since

they occupy large area. This also removes one of the benefits of using BiCMOS, since one of the reasons for using BiCMOS is the possibility to integrate better passive components like inductors. In this chapter the CMOS, BiCMOS, and SOI technologies will be described and their benefits and drawbacks will be discussed.

2.1 CMOS Technology

Due to the attractive scaling properties of CMOS technology, low standby power, low cost, and fast development, CMOS has for a long time been considered as THE technology for advanced digital circuits. Along with the technology scaling and improved performance of the transistors CMOS has become very popular for both analog and RF circuits as well. The consequence of this is the great opportunity to integrate analog-, RF-, and digital-circuits on the same die, which makes CMOS an excellent technology for future System-on-Chip implementations.

2.1.1 The MOSFET Device

A schematic symbol and a cross section view of an NMOS transistor is illustrated in Fig. 2.1. The basic function of the transistor is to control the current flowing between the drain and source terminals by applying a voltage on the gate terminal. The gate is isolated from the positively (p) doped substrate by a thin layer of an insulating material, usually SiO₂. When the gate voltage is increased above a certain threshold voltage, V_{TH} , a conducting channel of electrons is formed in the p-doped area under the gate. This allows a current to flow between the highly negatively doped drain and source. This on-off model is however an oversimplified view of the threshold voltage. The channel is built up gradually and a subthreshold current is flowing even below V_{TH} . For proper operation, a voltage also has to be applied to the substrate (bulk) in order to have a well defined potential. Appendix A contains the equations describing the current through the transistor as a function of the terminal voltages for the different operating regions of a MOSFET. The MOSFET can be used as a switch as is the case in digital circuits, a voltage controlled resistor or an amplifying device in analog circuits.

The most frequently used high-frequency performance metric for MOSFETs is the maximum cutoff frequency, f_T . f_T is defined as the frequency where the AC-signal short-circuit current gain is unity. It is therefore proportional to the ratio between the transconductance (g_m) and the input capacitance (C_{in}) . The cutoff frequency for long channel NMOS transistors and for short channel NMOS transistors operating at low gate over-drive voltages is expressed as:

$$f_T \approx \frac{3\mu_n (V_{GS} - V_{TH})}{4\pi L^2}$$
 (2.1)



Figure 2.1: (a) NMOS schematic symbol. (b) NMOS cross section.

 μ_n is the electron mobility, $(V_{GS} - V_{TH})$ is the gate over-drive voltage, and L is the channel length. As seen from Eq. 2.1 f_T depends on $(V_{GS} - V_{TH})$ and is inversely proportional to the square of the channel length. It is thus obvious why transistor performance improve by scaling down the channel length. In deep submicron technologies with very short channel lengths the carriers get velocity saturated, particularly at high $(V_{GS} - V_{TH})$. Velocity saturation, v_{sat} decreases the electron mobility [5]. Considering this, f_T for these devices can be rewritten as:

$$f_T = \frac{v_{sat}}{2\pi L} \tag{2.2}$$

As seen from Eq. 2.2 f_T is in this case inversely proportional to the channel length at the first order and the benefit of scaling is therefore lowered. To date, transistors used for analog and RF are rarely biased in a way that the carriers get velocity saturated. This since reaching velocity saturation still requires a relatively high gate over-drive voltage for technologies used today and that would cause an unacceptably high power consumption for most analog an RF circuits. However, due to scaling, lower and lower gate over-drive voltage will be required to reach velocity saturation. The most important thing for analog and RF is to have high f_T for low bias current together with low noise. Biasing at maximum f_T is thus not really practical for low power.

Another often used high-frequency measure is f_{max} . f_{max} is defined in the same way as f_T , but instead of looking at the current gain it is when the power gain has dropped to unity. f_{max} is a function of f_T , gate resistance and gate-drain capacitance.

2.1.2 Impact of Scaling on the MOSFET Device

The device scaling will continue within the near future. From the roadmap over technology scaling from *ITRS* [2] the scaling will continue and year 2019 the drawn channel length is predicted to be 16 nm. Table 2.1 shows the predicted scaling of RF and analog mixed-signal CMOS technology. It can thus be expected that scaling will continue to give huge performance improvements.

Year of production	2007	2010	2013	2016	2019
Technology node [nm]	65	45	32	22	16
Nominal V_{DD} [V]	1.2	1.1	1.0	1.0	1.0
Saturation V_{TH} [V]	0.2	0.18	0.15	0.11	0.1
Peak f_T (NMOS) [GHz]	170	280	400	550	730
Peak f_{max} (NMOS) [GHz]	270	420	590	790	1020
NF _{min} @5 GHz (NMOS) [dB]	0.25	<0.2	<0.2	<0.2	<0.2
σV_{TH} matching [mV· μ m]	6	5	5	4	4
I_{DS} for f_T =50 GHz [$\mu A/\mu m$]	13	8	6	4	3
g_m/g_{ds} at $5{\cdot}L_{min-digital}$	32	30	30	30	30

Table 2.1: Predicted scaling of RF and analog mixed-signal CMOS technology according to *ITRS* 2005.

Increased f_T and f_{max} will certainly be most welcome. Faster devices due to scaling will make CMOS usable for RF frequencies up to many tenths of GHz. However, there are also problems arising from the scaling. The most severe one for RF and analog is the lower supply voltage. As a direct consequence of the reduced supply voltage, the available dynamic range is also reduced. To compensate for this reduction in dynamic range, the noise floor has to be lowered. As an example, the noise in switched-capacitor (SC) circuits can only be reduced by increasing the capacitor size. Inevitably, this means increased chip area and higher current dissipation required to drive the larger capacitance. For some analog applications, where the reduced signal swing due to the lower supply voltage cannot be accepted, one option is to use high-voltage transistors (IO transistors) instead. The IO transistors have thicker gate oxide and longer channel lengths and are therfore slower than the thin-gate devices, but in some cases this can be accepted and give an extra degree of freedom for circuit designers. For fixed-power applications, like power amplifiers, the lower power supply voltage also requires increased current. This make designs more sensitive to parasitic resistance in the supply lines causing voltage drops that further reduce the usable supply voltage.

In digital design the subthreshold leakage and gate leakage are two of the main concerns since they affect the reliability and cause an excessive amount of standby power. In fact, scenarios where the leakage power and the active power are of the same order are quite common for large digital circuits like for example microprocessors. The subthreshold leakage is a result of the scaling of the threshold voltage. Lower threshold voltage gives a faster device at the expense of more leakage. For analog and RF the subthreshold leakage becomes a problem in for example sampling circuits. Charge stored at the capacitor will leak away even though the transistor is "turned off". Since the leakage strongly depends on the voltage across the transistor, the leakage will give arise to distorsion. The gate leakage on the other hand is caused by tunneling through the extremely thin gate oxide. Further decreasing the thickness of the oxide will increase the gate leakage to unacceptable levels. One solution to the gate-leakage problem that has been suggested is the use of high- κ dielectrics instead of SiO₂ for the gate-oxide [2, 6]. Using the same example as above, a sampling circuit will be affected by the gate leakage in roughly the same way as by the subthreshold leakage. However, the use of MOS-capacitors will be very difficult in the future considering the increased gate leakage.

The noise properties of scaled transistors are also very important. Regarding the 1/f noise, the 1/f-noise corner will move up in frequency due to the scaling. 1/f-noise originates from the trapping and de-trapping of carriers in the gate oxide. The introduction of high- κ dielectric materials in the gate of future CMOS technology nodes will tend to increase the 1/f-noise levels [7, 8]. Due to the nature of carrier transport in MOS transistors taking place at the interface between SiO₂ and Si, the 1/f-noise corner is much higher than for bipolars. Here CMOS has a physically-based drawback compared to bipolar transistors. Regaring the thermal noise in a MOSFET the two main contributors are the drain current noise:

$$i_{nd}^2 = 4kT\gamma g_{d0}\Delta f \tag{2.3}$$

and the thermal noise contributed by the extrinsic gate resistance:

$$v_{nR_a}^2 = 4kT\delta R_q \Delta f \tag{2.4}$$

where δ depends on the contacting of the gate (one end or both ends). Along with the scaling the gate resistance is starting to become a problem. The gate resistance is becoming a significant noise source when designing for low noise, and short transistor fingers have to be used in order to reduce the gate resistance. γ in Eq. 2.3 is the channel noise factor. For long-channel devices $\gamma = 2/3$. For shortchannel devices larger values of γ has been reported. In [9] a $\gamma = 2$ is given for short-channel devices, but many different values of γ can be found in literature. One of the main concerns is that the noise level can become higher with scaling in such a way that the increased gain due to the scaling cannot compensate for the increased noise. In this case scaling would start producing higher noise figures at some point.

Figure 2.2 shows the typical characteristics for drain-current and transconductance (g_m) for a long-channel and a short-channel transistor. For the long-channel device the drain current is a quadratic function of the gate-source voltage (V_{GS})



Figure 2.2: Typical transistor characteristics for: (a) I_{DS} versus V_{GS} for longchannel MOSFET, (b) g_m versus V_{GS} for long-channel MOSFET, (c) I_{DS} versus V_{GS} for short-channel MOSFET, (d) g_m versus V_{GS} for short-channel MOSFET.

resulting in a transconductance which has a linear dependence on (V_{GS}) . For the short-channel device the current becomes linear to (V_{GS}) as the device becomes velocity saturated. This causes g_m to be more or less constant for higher V_{GS} . Since f_T is related to g_m , the f_T curve looks more or less the same as the one for g_m , compare with Eq. 2.1 and Eq. 2.2.

Regarding the linearity, for example the linearity of an LNA is strongly coupled to the transconductance of the device. Nonlinearities arise from the fact that g_m is not constant when a signal is applied to the gate. From Fig. 2.2 it is seen that a velocity-saturated device has a nearly constant transconductance yielding very high linearity. Thus, as a result of scaling transistors will be velocity saturated even for low gate-source voltages resulting in better linearity. However, for example reduced output impedance, g_{ds} , will worsen the linearity instead. Improved transistor speed can also be traded for improved linearity by the use of feedback even at high frequencies.

Increased variation in threshold voltages due to random dopant fluctuations in the channel is another important phenomena that comes along with the scaling [10]. This will increase the mismatch between devices even though they are placed close together. Particularly when reaching the 32-to-16 nm generation there will only be tenths of dopant atoms left in the channel causing large statistical fluctuations in the threshold voltage.

2.1.3 Silicon-on-Insulator Technology

In SOI CMOS a thin silicon layer where the devices are laid out is placed on top of an insulating layer (usually oxide for SOI CMOS). Fig. 2.3 shows the cross section of an SOI-NMOS transistor.



Figure 2.3: NMOS cross section when implemented in SOI technology.

There are two commonly mentioned types of SOI-CMOS devices, partially depleted (PD) and fully depleted (FD). In the case of PD SOI an undepleted region is present in the body region of the device if the thickness of the silicon layer is larger than the depletion depth. The body is therefore only partially depleted. The undepleted body region can be charged during operation, giving rise to several unwanted effects like the kink and history effect [11]. These effects can be solved by adding body contacts. PD SOI technology is now relatively mature [12]. Due to process scaling many of the advantages using PD SOI over bulk CMOS diminishes [11]. For FD SOI, where the silicon layer is made thinner than the depletion depth, the whole body-region is depleted. Consequently, the body region cannot be charged and the unwanted effects appearing for PD SOI are thereby avoided. FD SOI looks more promising for the future but is not yet commercially available.

The SOI-CMOS technology has several advantages, but also a number of disadvantages over bulk-CMOS technologies. The most important advantages usually mentioned when motivating the use of SOI CMOS are:

- Lower parasitic capacitance, leading to higher f_T .
- Reduced substrate coupling.
- Higher gain.
- Due to the dielectric isolation high quality passive components can be integrated.

but there are also a number of drawbacks:

- Thermal effects, self heating.
- Partly depleted SOI suffer from kink and history effects.
- · Fabrication cost.

The self-heating problem is due to the insulator layer in SOI CMOS. The thermal conductivity is about 100 times lower for the devices compared to bulk-CMOS devices [13]. Most of the heat generated in a bulk-CMOS device is transferred to the substrate below and only little heat is transferred to neighboring devices. For an SOI-CMOS device the situation is different. Due to the poor thermal conductivity of the insulating layer, more of the heat generated by the device will remain in the device, which increases the temperature. In addition, more heat is transferred to neighboring devices. Increased temperature also leads to decreased electron and hole mobility, which in turn decrease the drain current and thereby the transistor speed. The variation in drain current due to self-heating can be as much as 25%

[14]. The thermal effects should thus be included when simulating circuits based on SOI CMOS.

SOI-CMOS technology might become a mainstream technology in the future. It has several advantages over bulk CMOS, better isolation and better passive components. However, issues like self-heating, history effects, and silicon cost are major obstacles for SOI CMOS. The near future belongs to standard CMOS and BiCMOS technologies. The question is if there ever will be a market for SOI technology for RF applications?

2.2 BiCMOS Technology

In a BiCMOS process bipolar transistors have been added to a CMOS technology. By adding bipolars, high-frequency circuits, where CMOS simply are not fast enough, can be integrated together with digital parts where CMOS is the only reasonable choice. Modern BiCMOS processes use SiGe heterojunction bipolar junction transistors (HBTs). A graded Germanium profile within the base reduce the base transit time, thus improving f_T . BiCMOS technologies usually use a substrate with higher resistivity, compared to high-performance CMOS technologies, better suited for integration of high-Q passive components.

2.2.1 The Bipolar Device

A schematic symbol and a cross-section view of an NPN transistor is illustrated in Fig. 2.4. The operation of a bipolar transistor is physically much more complicated than for the MOSFET. A simplified view of the operation for an NPN bipolar junction transistor (BJT) is as follows [15]. When the base-emitter junction is forward biased, it starts conducting just as any forward-biased junction. The current consists of majority carriers from the base (holes since the base is



Figure 2.4: (a) NPN schematic symbol. (b) NPN cross section.

positively (p) doped) and majority carriers from the emitter (electrons since the emitter is negatively (n) doped) diffusing across the junction. Since the emitter is more heavily doped than the base, there are many more electrons injected from the emitter than holes injected from the base. Further assuming the collector voltage is large enough to ensure that the collector-base junction is reversed biased, no holes from the base will go to the collector. However, the electrons that travel from the emitter to the base, where they are now minority carriers, diffuse away from the base-emitter junction because of the large gradient of minority-carrier concentration in the base region. Any of these electrons that get close to the collector-base junction will immediately be swept across the junction by the drift mechanism due to the large positive voltage on the collector. Accordingly, the bipolar transistor is a current-controlled device in opposite to the voltage-controlled MOS transistor. The total emitter current is the sum of the base current and the collector current. Since the base current is small compared to the collector current, the emitter current is approximately equal to the collector current. Appendix A contains the equations describing the current through the bipolar transistor as a function of the terminal voltages.

2.2.2 Impact of Scaling on the Bipolar Device

Table 2.2 shows the predicted scaling of RF and analog mixed-signal bipolar technology according to *ITRS* 2005. The MOS part in a BiCMOS process will be facing the same problems as a pure CMOS process described in section 2.1.2. Technology requirements for bipolar transistors are driven by the need for lower power consumption, lower noise, and lower cost in the same way as for CMOS. One of the major challenges for BiCMOS technology is to scale the supply voltage of the bipolars to reduce the power consumption versus CMOS technology. Reduced power consumption and lower noise for bipolar devices are achieved through higher f_T and f_{max} , scaling of the vertical profile and scaling of the emitter width. Reducing the base resistance is essential to reduce the noise. For Si/SiGe bipolars f_T is determined by the electron transit times for the base and collector [9]. Therefore, scaling the dimensions will improve the high-frequency performance.

2.3 Integration of Passives

In many cases the integration of RF, analog, and mixed-signal circuits require integration of inductors, high-precision capacitors, resistors, and varactors alongside with the active devices. Table 2.3 shows the predicted scaling of RF passives according to *ITRS* 2005. The improved performance of the passives does not come

Year of production	2007	2010	2013	2016	2019		
Emitter width [μ m]	0.13	0.1	0.09	0.08	0.07		
σ current matching [%· μ m]	2	2	2	2	2		
High speed NPN							
Peak $f_T (V_{bc} = 1 \text{ V}) \text{ [GHz]}$	265	370	420	480	550		
Peak f_{max} [GHz]	300	410	460	520	590		
BV_{ceo} [V]	1.8	1.7	1.6	1.4	1.3		
J_c at peak f_T [mA/ μ m ²]	12	15	18	21	24		
RF NPN							
Peak $f_T (V_{bc} = 1 \text{ V}) \text{ [GHz]}$	90	100	120	130	150		
Peak f_{max} [GHz]	170	200	230	260	290		
BV_{ceo} [V]	3.1	2.9	2.6	2.5	2.4		
NF _{min} @5GHz (NMOS) [dB]	0.26	<0.2	<0.2	<0.2	<0.2		
I_c for $f_T = 50$ GHz $[\mu A/\mu m]$	28	15	12	9	6		

Table 2.2: Predicted scaling of RF and analog mixed-signal bipolar technology according to *ITRS* 2005.

from process scaling, rather from the evolution of device implementation and material choices [16]. The two most challenging passive components to integrate are high-precision high-density capacitors and inductors. Since these usually occupy much more silicon area than the active devices, scaling of the capacitors and inductors is highly desireable to reduce the overall die size.

Precision capacitors with high density, good matching properties, together with high linearity are usually implemented as metal-insulator-metal (MIM) capacitors. These capacitors are integrated near the top of the interconnect stack by adding a couple of extra masks during the manufacturing. Looking at the scaling properties of MIM capacitors, the density in terms of $fF/\mu m$ will improve. The resulting Q-value due to scaling will more or less remain the same. This since the Q-value is proportional to the product of the capacitance (C) and the series resistance (R). The total area is reduced while the resistance per unit area is increased and the product of R and C remains the same throughout scaling [9, 17].

Year of production	2007	2010	2013	2016	2019				
Technology node [nm]	65	45	32	22	16				
P+ polysilicon resistor									
Sheet resistance $[\Omega/sq]$	200-300	200-300	200-300	200-300	200-300				
σ Matching [%· μ m]	1.7	1.7	1.7	1.7	1				
Metal-Insulator-Metal capacitor									
Density [fF/µm ²]	2	5	7	10	12				
σ Matching [%· μ m]	0.5	0.4	0.3	0.2	0.2				
Q (5 GHz, 1 pF)	>50	>50	>50	>50	>50				
Inductor									
Q (5 GHz, 1 nH)	29	34	40	46	52				
MOS varactor									
Tuning range	5.5	5.5	5.5	5.5	5.5				
Q (5 GHz, 0 V)	35	40	50	55	60				

Table 2.3: Predicted performance improvements of passives according to *ITRS* 2005.

Inductors are implemented in the thicker top-level metal layers of the interconnect stack to minimize the resistive loss, as well as the capacitive and magnetic losses to the dissipative silicon substrate. The Q-factor is limited by the inductor series resistance at low frequencies, while at high frequencies the substrate resistivity is the limiting factor [16]. Traditionally CMOS technologies used a low-ohmic p+ substrate to avoid latch-up. This is usually not used for RF CMOS technologies because of the need for high-quality passives. Instead a p- substrate is used or a thin p- epi on a p+ substrate. One of the main issues with inductors is that they consume a lot of area and scaling their size is not as easy as for the MIM capacitors. The fundamental problem for integrated inductors is that they need to store much more energy than they dissipate per cycle, and naturally it is very difficult to store a lot of energy in the small volume of an integrated circuit die. The scaling properties of inductors are thus not very promising, since scaling the size of an inductor also means an unwanted scaling of the quality factor [9]. Therefore, inductors will not scale as transistors and silicon areas occupied by onchip inductors will not scale down eventhough CMOS technology advances. This will make inductors a really expensive component to use in terms of silicon area and cost in the future.

2.4 A Brief Device Comparison

A fair device comparison between a state-of-the-art CMOS technology and corresponding BiCMOS technology is a little difficult to make since the BiCMOS technoloy lags about two generations with respect to the pure CMOS technology. In [9] the silicon technology tradeoffs are dicussed for RF and mixed-signal system-on-chip implementations. Comparable performance for LNA implementations has been shown using 0.25 μ m and 0.18 μ m RF CMOS and SiGe BiCMOS [18]. In this section a few of the most important differences between MOS and bipolar transistors are discussed.

Some of the drawbacks with using MOS transistors compared to bipolar transistors are:

- Lower gain than bipolar transistors.
- Higher power consumption for noise matching.
- Higher parasitics associated with MOS and worse passives.
- Much higher 1/f-noise corner frequency.
- Lower power supply voltage.

but there are also a number of advantages, among them:

- Better linearity.
- Superior for switches, sample and hold circuits, and mixers.
- More aggressive process scaling rapidly improving f_T .
- More suitable for high-level integration (SoC), since the technology is compatible with the rest of the design (baseband).
- Allows for low-voltage operation, which is not very easy using bipolars.
- Lower cost in high-volume production.

The lower power supply voltage for MOS transistors compared to bipolar transitors is on one hand an advantage since it allows for low voltage operation, but on the other hand it becomes a real bottleneck since it is limiting the dynamic range. The lower power supply voltage is one of the main issues when using RF CMOS compared to BiCMOS technology. The two most important transistor parameters when considering power consumption are the transconductance (q_m) and the effectiveness, the ratio between transconductance and bias current (g_m/I) , of the transistor. g_m/I for a bipolar transistor is equal to $V_T^{-1} \approx 40 V^{-1}$, while for a MOS transistor it is approximately $2/(V_{GS} - V_{TH}) \approx 1 - 25 V^{-1}$. In this sense the bipolar transistor uses the current much more efficiently. However, the MOS-FET can be operated at a much lower bias and supply voltage than the bipolar transistor, leading to lower power dissipation in some cases. Also, as explained in Appendix A, the maximum possible gain that can be achieved with a singletransistor amplifier is higher for bipolars. Another very important parameter is the maximum transition frequency, f_T , since RF devices should operate at very high frequencies. Today f_T for both bipolar and MOS devices are similar.

2.5 Technology Trends

The first thing to consider is single-chip versus multi-chip solutions. Regarding the digital baseband a high-end CMOS technology is the only resonable choice. Considering a multi-chip solution a BiCMOS implementation for RF could be the best solution. Several reasons can be mentioned, for example better device modelling, better isolation between the analog and digital parts, and perhaps lower power consumption. Regarding the manufacturing cost of BiCMOS, most comparisons of CMOS and BiCMOS are based on comparing processes using the same generation of photolithography as in [18]. The result from this comparison is obviously that CMOS is cheaper. This comparison is however a little skewed. Comparisons must not be made within the same lithography generation without considering what technology generation is needed for the wanted application. For a particular RF-IC design, a more agressive CMOS generation must be chosen, while a less agressive BiCMOS process will be at least as capable. In many cases, a less aggressive BiCMOS process will be less expensive than the CMOS alternative [19]. One of the main reasons for using bipolars today is the lack of accurate RF transistor models for MOSFETs. This also explains why many new RF products are first realized in BiCMOS and later transferred to pure CMOS implementations. However, once CMOS solutions hit the market bipolar solutions quickly become obsolete. For single-chip solutions (power amplifier most likely not included) the requirements from digital signal-processing capability and memory integration directly imply the use of CMOS. For most low-cost high-volume

applications CMOS will be used, at least for frequencies up to ~ 10 GHz. SiGe bipolar technology is driven by products like automotive radar and high data rate short-range WLAN emerging in the 24-77 GHz range [20].

RF circuits implemented in SiGe BiCMOS may have better performance than corresponding CMOS circuits, but if CMOS is "good enough" it will be used. Considering the amount of research going on in the field of implementing CMOS circuits for RF and the number of new companies working on CMOS radio, the technology trend clearly points in the CMOS direction. *Hopefully, all these engineers all around the world cannot be wrong*.

2.6 References

- G. E. Moore, "Cramming More Components onto Integrated Circuits," *Electronics*, vol. 38, no. 8, 1965.
- [2] ITRS, http://public.itrs.net, 2006.
- [3] C. Svensson, "Forty Years of Feature-Size Predictions," in Proc. of the 50th IEEE International Solid-State Circuits Conf. (ISSCC), 2003.
- [4] D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur, and H-S. Wong, "Device Scaling Limits of Si MOSFETs and Their Application Dependencies," vol. 89, March 2001.
- [5] B.G. Streetman, Solid State Electronic Devices. Prentice-Hall, 1995.
- [6] Y-C. Yeo, T-J. King, and C. Hu, "MOSFET Gate Leakage Modeling and Selection Guide for Alternative Gate Dielectrics Based on Leakage Considerations," *IEEE Transactions on Electron Devices*, vol. 50, no. 4, April 2003.
- [7] L.M Franca-Neto, R. Eline, and B. Balvinder, "Fully Integrated CMOS Radios from RF to Millimeter Wave Frequencies," *Intel Technology Journal*, vol. 8, pp. 241–258, August 2004.
- [8] A. Mercha, W. Jeamsaksiri, J. Ramos, S. Jenei, S. Decoutere, D. Linten, and P. Wambacq, "Impact of Scaling on Analog/RF CMOS Performance," in *Proc. of the ICSICT Conf.*, pp. 147–152, 2004.
- [9] L. Larson, "Silicon Technology Tradeoffs for Radio-Frequency/Mixed-Signal "Systems-on-a-Chip"," *IEEE Tran. on Electron Devices*, vol. 50, pp. 683–699, March 2003.

- [10] S. Borkar, "Designing Reliable Systems from Unreliable Components: The Challenges of Transistor Variability and Degradation," *IEEE Micro*, pp. 10– 16, November 2005.
- [11] E. Säll, Implementation of Flash Analog-to-Digital Converters in Siliconon-Insulator Technology. Lic. thesis, Linköpings universitet, Department of Electrical Engineering, SE-581 83 Linköping, Sweden, 2005.
- [12] R. Huang, H. Liao, and G. Zhang, "SOI CMOS Technology for RF/MMIC Applications — Yes or No?," in *Proc. of the ESSDERC Conf.*, pp. 241–244, 2005.
- [13] B.M. Tenbroek, M.S.L. Lee, W. Redman-White, R.J.T. Bunyan, and M.J. Uren, "Impacy of Self-Heating and Thermal Coupling on Analog Circuits in SOI CMOS," *IEEE Journal of Solid-State Circuits*, vol. 33, pp. 1037–1046, July 1998.
- [14] J.B. Kuo and S-C. Lin, Low-Voltage SOI CMOS VLSI Devices and Circuits. John Wiley & Sons Inc., 2001.
- [15] D.A. Johns and K. Martin, Analog Integrated Circuit Design. Wiley, 1997.
- [16] C. Diaz, "CMOS Technology for MS/RF SoC," in Proc. of the IEEE Workshop on Microelectronics and Electron Devices, pp. 24–27, 2004.
- [17] K. Lee, I. Nam, I. Kwon, J. Gil, K. Han, S. Park, and B-I. Seo, "The Impact of Semiconductor Technology Scaling on CMOS RF and Digital Circuits for Wireless Application," *IEEE Tran. on Electron Devices*, vol. 52, pp. 1415– 1422, July 2005.
- [18] B.A. Floyd and D. Ozis, "Low-Noise Amplifier Comparison at 2 GHz in 0.25-μm and 0.18-μm RF-CMOS and SiGe BiCMOS," in *Proc. of the IEEE Radio Frequency Integrated Circuits Symposium*, pp. 185–188, 2004.
- [19] D. Foty, "New Directions in Technology for High-Speed Wireless Data Communications," in Proc. of the MIXDES Conf., pp. 19–27, 2006.
- [20] M. Racanelli and P. Kempf, "SiGe BiCMOS Technology for RF Circuit Applications," *IEEE Transactions on Electron Devices*, vol. 52, no. 7, July 2005.

Chapter 3 Radio-Receiver Architectures

The basic function of a radio front-end is to take the modulated carrier signal from the antenna, amplify and downconvert the signal, select the wanted channel, and finally extract the baseband information. From this point of view a radio receiver does not appear to be very complicated. However, when considering that the signal could be very week or very strong depending on how far from the transmitter the receiver is, or even worse when the wanted signal is very week but another strong interferer (blocker) is present, then the true challenges of radio receiver design are brought into the light. There exist a number of different ways to build a receiver which takes care of all those issues in one way or the other. In this chapter the most common traditional radio-receiver architectures are presented. More details can be found in [1, 2, 3].

3.1 Receiver Specifications

A traditional RF front-end typically consists of a low-noise amplifier (LNA), mixer(s) for downconversion, local oscillator(s) (LO), filters, and finally an analog-to-digital converter (ADC). The first stage after the antenna and RF filter in the receiver chain is typically an LNA. Since every stage in the receiver chain adds noise to the signal, very weak signals will be buried in this noise and be lost. The main function of the LNA is to provide high enough gain to overcome the noise of subsequent stages (mixer etc.), while adding as little noise as possible. At the same time it should be linear enough to handle strong interferers without introducing intermodulation distortion. The LNA is followed by a mixer that performs the downconversion to an intermediate frequency (IF) determined by the local oscillator. Due to the nature of the mixing process, the image problem occurs. Several techniques to handle the image problem exists and will be discussed later in this chapter. After the mixer filtering is required. Then, if necessary, another step of

downconversion and filtering is made before the signal is amplified and ready for analog-to-digital conversion.

The performance of the receiver depends on the system design, circuit design, and working environment [4]. The acceptable level of noise and distortion varies with the application. Noise and distortion specify a lower limit on the usable signal-level at the output. For the output-signal to be useful, the signal power has to be larger than the noise power by an amount specified by the required minimum signal-to-noise ratio (SNR). Including the distortion produced in the receiver, both noise and distortion should be at least the minimum specified signal-to-noise and distortion ratio (SNDR) below the signal power.

3.1.1 Receiver Requirements

Due to the very different operating conditions of a receiver depending on the surrounding environment a number of requirements can be specified.

- *Sensitivity*: The receiver sensitivity is the ability to receive a week signal. The requirement is the specified SNR for the analog front-end corresponding to the bit error rate (BER) in the digital domain. The sensitivity directly translates to the receiver noise figure (NF).
- *Selectivity*: The selectivity is the ability to suppress signals on adjacent channel frequencies. The selectivity sets the requirements on the channel select filter.
- *Blocker immunity, intermodulation rejection*: Interfering RF signals together with receiver nonlinearities can generate intermodulation products that fall into the channel of interest. The receiver linearity is usually specified through the IIP3 and compression point.
- *Power consumption*: Low power consumption is usually required. The previous three requirements should not be solved at the expense of severely increased power consumption.

As mentioned above, noise and linearity are two very important specifications for a radio receiver front-end. The total gain and how well interfering signals have to be suppressed, is at the end determined by the ADC. Too tough requirements on the ADC will result in unacceptably high power consumption. Therefore, the task of the receiver has to be distributed among the different stages in the receiver chain to find the best possible solution. A link budget is usually made to specify the required noise and linearity properties for each component in the front-end in order to fulfill the required specification for the whole receiver. In a typical frontend the LNA and the mixer are the main stages contributing to the overall noise and linearity [5]. The receiver noise figure specify how much noise is added to the signal while passing the receiver, i.e. the SNR degradation by the receiver. The noise figure of the receiver is defined as:

$$NF = 10\log \frac{SNR \ at \ input}{SNR \ at \ output} = \frac{S_i/N_i}{S_o/N_o} = \frac{N_o}{G \cdot N_i}$$
(3.1)

where S_i is the input signal, S_o is the output signal, N_i is the noise at input, N_i is the noise at output, and G is the total gain of the receiver. When designing a receiver front-end, each component has to be designed with a noise figure and gain so that the specification for the whole front-end is met. The noise figure for a system of cascaded circuits, and the impact of every circuit on the overall noise figure, can be calculated using *Friis formula* where the noise factor (F) of stage *i* is divided by the total gain from the input to stage *i*:

$$F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \dots + \frac{F_N - 1}{G_1 G_2 \dots G_{N-1}}$$
(3.2)

The noise figure is calculated from the noise factor as NF = 10log(F). The gain dependency of F in Eq. 3.2 results in the fact that once the signal is amplified, the noise of subsequent stages is less important. Therefore, the system noise figure is often dominated by the noise from the first stage(s) in the receiver chain. Trying to achieve a good noise figure, most of the design effort must be put on the earliest stages in the receiver chain, particularly on the LNA.

Nonlinearities in a receiver result in intermodulation distortion, desensitization (blocking), and cross-modulation. When two signals at frequencies f_1 and f_2 are applied to a nonlinear system they generate intermodulation (IM) products according to $mf_1 \pm nf_2$, m, n = 0, 1, 2, ... [4]. The linearity of a circuit or a whole receiver can be evaluated by using a two-tone test, i.e. applying two frequency tones to the input and look at the frequency spectra at the output. Second-order intermodulation is not that troublesome because the products can be filtered out [6]. However, this is not true for direct-conversion (homodyne) receivers. The two-tone third-order IM products are of primary interest since they can show up within the passband of the signal that is going to be received. Fig. 3.1(a) shows a typical output spectra from a two-tone test. The two fundamental tones are at f_1 and f_2 , while the third-order intermodulation distortion (IMD) shows up at frequencies $2f_1 - f_2$ and $2f_2 - f_1$. If the IMD is larger or equal to the noise floor, the spurious-free dynamic range (SFDR) is estimated as the difference between the fundamental tone and the IMD product [7]. There are many measures of linearity, but the most commonly used are third-order intercept point (IP3) and 1dB compression point (CP). At the 1-dB compression point the gain is 1 dB lower than the small-signal gain due to nonlinearities. This is because the linear smallsignal assumption is no longer valid for such high input-power. This will reduce



Figure 3.1: Linearity characteristics and dynamic range definitions: (a) Intermodulation spectrum. (b) Input power versus output power.

the gain of a weak signal in presence of a strong interferer (blocking). The most common measure to specify the linearity is the input-referred third-order intercept point (IIP3). Fig. 3.1(b) illustrates the result of a two-tone test. The output power of the fundamental and third-order tone are plotted as a function of input power. By extrapolating the output power curves of the fundamental and third-order tone, IIP3 is found as the input power where the two lines intersect. When designing a front-end the linearity (IIP3) of every component has to be considered. The overall linearity of cascaded nonlinear components can be calculated as:

$$\frac{1}{IIP3} = \frac{1}{IIP3_1} + \frac{G_1}{IIP3_2} + \dots \frac{G_1G_2\dots G_{N-1}}{IIP3_N}$$
(3.3)

When comparing Eq. 3.2 and Eq. 3.3 the trade-off between noise, gain, and linearity is obvious. As discussed before, large gain in the first stage (the LNA) implies a low overall noise figure. However, large gain in the first stage also means the overall linearity gets worse. Thus, a compromise between noise and linearity has to be made when determining the gain of each stage in the receiver chain.

3.1.2 The Image Problem

Since the mixer acts as a multiplier it performs both up and downconversion simultaneously. For a receiver only the downconverted signal is of interest why it is here assumed that the up-converted signal is filtered out. Thus, by mixing



Figure 3.2: Image problem in downconversion: (a) Heterodyne receivers. (b) Homodyne receivers. (c) Low-IF receivers.

an RF signal of frequency f_{RF} with an LO signal of frequency f_{LO} results in an intermediate-frequency (IF) signal at $f_{IF} = f_{RF} - f_{LO}$. Now, the problem is that a signal equally spaced from the LO on the opposite side with respect to the RF signal will also be downconverted to the same IF frequency. This signal is usually called the image frequency, f_{IM} , and is a result of the mixing process. A more mathematical explanation can be found in [7]. Fig. 3.2 illustrates the image-frequency problem for three different receiver architectures downconverting to different intermediate frequencies. Different techniques exist to eliminate the image problem. The most straight forward approach is to choose a high enough IF so that the image frequency can be removed by the RF filter. This technique is commonly used in superheterodyne receivers, as described in Section 3.2. A specific case is the homodyne or zero-IF receiver. Since the image frequency in this case is the signal itself, no image rejection filter is needed. However, for frequency-or phase-modulated signals this does not solve the problem. This is because the upper and lower sidebands does not contain the same information [3].

Another solution to the image problem is to use special image-rejection architectures. The idea is that instead of removing the image by using filters, the wanted signal and the image are separated by using quadrature downconversion. Using quadrature downconversion to separate the image from the wanted signal is more suitable for chip implementations since it is usually very difficult to realize the sharp high-quality filter needed to remove the image. Two commonly used image-rejection architectures are the Hartley architecture shown in Fig. 3.3, and



Figure 3.3: Hartley architecture.



Figure 3.4: Weaver architecture.

the Weaver architecture shown in Fig. 3.4. Both architectures need a 90° phase shift and the main difference is that instead of the passive phase shift typically used in the Hartley architecture a second mixing is used in the Weaver architecture. To demonstrate how the image rejection is accomplished using the Hartley architecture refer to Fig. 3.3 and assume an input signal composed of the wanted signal and its image:

$$v(t) = A_{RF}\cos(\omega_{RF}t) + A_{IM}\cos(\omega_{IM}t)$$
(3.4)

After quadrature mixing of the input signal using the signals $\cos(\omega_{LO})$ and $\sin(\omega_{LO})$ from the LO and after LP filtering, the frequency relations $\omega_{IF} = \omega_{RF} - \omega_{LO} =$

 $\omega_{LO} - \omega_{IM}$ give signal and image parts in the branches as:

$$v_{RF,I} = \frac{A_{RF}}{2} \sin((\omega_{RF} - \omega_{LO})t) = \frac{A_{RF}}{2} \sin(\omega_{IF}t)$$

$$v_{RF,Q} = \frac{A_{RF}}{2} \cos((\omega_{RF} - \omega_{LO})t) = \frac{A_{RF}}{2} \cos(\omega_{IF}t)$$

$$v_{IM,I} = \frac{A_{IM}}{2} \sin((\omega_{LO} - \omega_{RF})t) = \frac{A_{IM}}{2} \sin(-\omega_{IF}t) = -\frac{A_{IM}}{2} \sin(\omega_{IF}t)$$

$$v_{IM,Q} = \frac{A_{IM}}{2} \cos((\omega_{LO} - \omega_{RF})t) = \frac{A_{IM}}{2} \cos(-\omega_{IF}t) = \frac{A_{IM}}{2} \cos(\omega_{IF}t)$$
(3.5)

Performing the phase shift of 90° on the Q branch result in:

$$v_{RF,Q} = -\frac{A_{RF}}{2}\sin(\omega_{IF}t)$$

$$v_{IM,Q} = -\frac{A_{IM}}{2}\sin(\omega_{IF}t)$$
(3.6)

Subtracting the signals from the I and Q branches gives the IF signal:

$$v_{IF} = A_{RF} \sin(\omega_{IF} t) \tag{3.7}$$

As seen from the equations above the desired part of the I and Q components are in phase and adds up while the image parts are in opposite phase thus cancelling each other out. Gain and phase mismatch between the I and Q components will result in degraded image rejection. This is because the image parts no longer cancel when they have different amplitude and phase. In modern receivers the phase shift and addition of the signals are done in the digital domain (in the baseband processor). Static gain and phase errors can be digitally corrected and compensated for [8, 9, 10]. This ease the design of the RF front-end considerably. It is also a good example of how digital compensation can be utilized to ease the design requirements of RF and analog circuits.

3.2 Superheterodyne Receivers

The superheterodyne receiver topology was invented by Edwin H. Armstrong in 1918¹. Most radio receivers used today use more or less a similar approach for frequency downconversion etc. as the classical superheterodyne, even though some refinements and changes have been made compared to the original architecture. The superheterodyne receiver is still commonly used today due to its high sensitivity and selectivity combined with good image rejection. Fig. 3.5 illustrates a classical dual superheterodyne receiver, while Fig. 3.6 shows a more modern dual superheterodyne receiver architecture for digital baseband. The first band-

¹A Frenchman named Lévy later claimed he was first and in 1928 Armstrong lost his superheterodyne patent.



Figure 3.5: Dual superheterodyne receiver.



Figure 3.6: Dual-superheterodyne receiver for digital-baseband demodulation.

pass filter between the antenna and LNA selects the desired RF band and also acts as an image rejection filter for the first mixer. Then the RF-signal is amplified by a low-noise amplifier (LNA) and downconverted to an intermediate frequency (IF) in the first mixer. The first IF is determined by the first local-oscillator (LO) frequency and is chosen in such a way that the image is removed by the RF filter. Another bandpass filter is used before the second mixer to remove the image frequencies associated with the second mixing. Consider the receiver shown in Fig. 3.5, a third bandpass filter is used after the second mixer to select the desired channel and suppress the neighboring channels. The desired channel is chosen by tuning the second LO to a frequency so that the IF of the desired channel coincides with the passband of the channel filter. Thus the channel filter has a fixed center frequency and the IF before the demodulator is the same for all channels. One way to demodulate the signal is to do it in digital domain. Fig. 3.6 shows how quadrature downconversion followed by low-pass filtering, amplification (VGA), and analog-to-digital conversion is performed before the digital demodulation in a digital signal processor (DSP).

The choice of IF is a trade-off between good selectivity and good image rejection. A high IF means the signal and image bands are further apart, thus relaxing the filter requirements and still achieve very good image rejection. A low IF on the other hand gives the advantage of relaxed channel-filter (IF filter) requirements. If the carrier frequency is very high compared to the signal bandwidth, several intermediate frequencies can be used to downconvert the RF signal. This improves the image rejection and suppression of adjacent channels [11]. Therefore, by using several intermediate frequencies, both good selectivity and sensitivity can be accomplished.

The main drawback with the superheterodyne receiver architecture is the need for high-Q filters used for image rejection. These filters are usually placed off-chip and tend to be bulky and expensive. This also means that single-chip solutions, where everything is integrated on the same die, are difficult to build using the superheterodyne architecture. Besides that, the power needed to drive off-chip components like the external filters tend to give relatively high power consumption [12].

3.3 Homodyne Receivers

As described in previous section, the major obstacle for high-level integration of superheterodyne receivers is the need for high-Q bandpass filters. A lot can be gained by using receiver architectures where these filters can be simplified or removed. In a homodyne (also called direct conversion or zero-IF) receiver the signal is downconverted directly to baseband without the use of any intermediate frequencies. Since the center of the desired channel is translated to DC (0 Hz), the portion of the channel on the negative frequency axis becomes the image to the portion of the signal on the positive frequency axis and vice versa. Therefore, the downconverted signal has to be reconstructed using complex signal processing. Otherwise the negative-frequency half-channel will be folded on the positive-frequency half-channel. This means that two downconversions are made, one for the in-phase (I) component and one for the quadrature-phase (Q) component. Fig. 3.7 shows a block diagram of a homodyne receiver using I/Q downconversion. As seen in Fig. 3.7 the band-pass channel-selection filter used in the superheterodyne



Figure 3.7: Homodyne receiver.

is now replaced by a lowpass (LP) filter. The LP filter can be implemented onchip using for example active filters, thus eliminating the need for external filters and thereby allowing for a high level of integration. If the filter in a homodyne receiver can be made programmable, then multiple frequency bands could be used (for example different signal bandwidths for different standards) [11]. After the LP filter the signal is amplified using a variable-gain amplifier (VGA) to adjust the signal level to the input-range of the ADC and thereby decrease the required dynamic range. The final demodulation is performed in the DSP.

Of course there are also a number of disadvantages with homodyne receivers. The most severe one is the DC-offset problem. This since the DC offset is superimposed on the baseband signals that are downconverted around DC. The DC offset is introduced in the mixer by the parasitic coupling of the LO signal causing the LO signal to be mixed with itself. This effect is commonly called *LO leakage* [13]. Except corrupting the baseband signal at DC, the DC offset might also cause saturation in the following stages after the mixer. Mismatch between I and Q signals might also give a DC offset. The DC offset is usually removed by adding a DC-offset cancellation circuit. For example, the DC offset is measured by the DSP and a compensating voltage is fed back using a digital-to-analog converter (DAC) to compensate for the offset.

Another common issue when using the homodyne architecture is the problem with 1/f noise. Since the 1/f-noise essentially is a problem at low frequencies, the baseband signal now downconverted around DC can be expected to suffer from the 1/f noise. Techniques to solve the 1/f-noise problem for mixers have been suggested in [14, 15]. Other techniques like chopper amplifiers [16] can also be used in for example switched-capacitor (SC) circuits to reduce the influence of 1/f noise.

The homodyne receiver is also sensitive to I/Q mismatch arising from gain and phase errors between the I and Q paths, which severely degrades the image rejection. However, this can be corrected for in the digital domain and is usually not considered that critical for homodyne receivers of the type shown in Fig. 3.7.

3.4 Low-IF Receiver

The low-IF receiver architecture is a compromise between the superheterodyne and the homodyne trying to combine the advantages of these two. The IF is typically chosen as one or two times the channel bandwidth [17]. Because of the non-zero IF, DC offset, 1/f-noise, and LO self-mixing problems are not as severe as for the homodyne receiver. The low-IF receiver architecture combines high performance and high degree of integration at the same time [18]. The analog part of the low-IF receiver is similar to the homodyne receiver (Fig. 3.8). The image



Figure 3.8: Low-IF receiver.

problem is solved by complex signal processing in the same manner as for the homodyne receiver. A low IF thus requires a higher sampling rate for the ADC compared to the homodyne. The needed-channel select filter of band-pass type is implemented in digital domain.

3.5 Subsampling Receivers

Almost all wireless receivers are narrowband. Since only a small portion of the entire RF-band is of interest subsampling can be used. The wanted signal-band is downconverted without aliasing as long as the Nyquist criterion is fulfilled for the channel bandwidth. The main difference between subsampling receivers and the previously described receivers is that no mixer is used to downconvert the RF signal to baseband. Instead the sample-and-hold circuit in the ADC is used to downconvert the RF signal to baseband. The main drawback with subsampling receivers is the aliasing of noise. Wideband noise is folded into the baseband unless the RF signal is properly band-pass filtered. Besides that, subsampling worsens the effect of noise in the sampling clock [1].

3.6 Digital Receivers

Signal processing can be made much more efficient in digital domain and increased flexibility can be achieved by doing most of the filtering there as well. This together with the recent development of high performance ADCs makes the idea of moving the ADC closer to the antenna very appealing. The most extreme solution is to place the ADC immediately after the antenna and perform both downconversion and demodulation in digital domain. Thus eliminating the nonidealities of the analog signal processing, since the implementation of high selectivity and high dynamic range is easier done in digital domain [19]. Having the whole signal bandwidth in digital form, multiple channels could be demodulated



Figure 3.9: Direct A/D-conversion receiver.

simultaneously using parallel digital blocks. This type of receivers could be successfully used for wireless communication systems employing different standards and utilizing receiver flexibility [20]. This reasoning is often used for software-defined radio (SDR), where the idea is to have a single radio that can receive and transmit multiple frequency bands and handle different standards by reconfiguring the software. Note that this does not say anything about what the receiver front-end should look like. However, the *old fashioned* view is to place the ADC directly after the antenna or possibly after an LNA.

The direct A/D-conversion receiver shown in Fig. 3.9, where the ADC is placed after the LNA, is here representing the traditional view of SDR. The reason for putting the ADC after the LNA is that real ADCs usually do not have the required resolution to pick up sub-microvolt, nanowatt radio signals. The main bottleneck in SDR is the ADC since it requires very high dynamic range at sampling frequencies comparable to the RF. So far such ADCs have shown to have very high power consumption, which makes this whole receiver architecture unpractical. The initial RF filter removes out-of-band blockers, which can generate in-band intermodulation products. However, this violates the whole point with SDR since the filter reduces the radio's flexibility.

When it is not possible to directly use a Nyquist ADC at RF frequency, the digital-IF receiver shown in Fig. 3.10 can be an alternative. In principal this is a dual-IF receiver where the second mixer stage has been replaced by an ADC and the second downconversion is performed in digital domain. The digital-IF receiver has a number of advantages [7], but the most important from a SDR point of view is that the programmable digital part simplifies channel filtering and makes the downconversion of multiple channels possible for wideband applications [21]. The main drawback is high dynamic range and relatively high sampling frequency needed in the ADC, which increases the power consumption.

RF-sampling receiver topologies have recently received much attention [7, 22, 23, 24]. Since both mixing and sampling performs a frequency translation, the RF-sampling receiver can be seen as a flavor of the single-conversion receiver architecture. The direct RF-sampling techniques allow for great flexibility in reconfigurable radio design [23]. Digital signal processing concepts are used to



Figure 3.10: Digital-IF receiver.



Figure 3.11: RF-sampling receiver.

help relieve analog design complexity. One version of an RF-sampling receiver is shown on block level in Fig. 3.11. The RF signal is first sampled after the LNA. Because of the sampling the signal is downconverted to baseband frequency, although highly oversampled. To relax the requirements on the ADC, the sampled discrete-time signal is decimated and filtered using a switched-capacitor (SC) filter before the A/D-conversion. The SC filter can be implemented in many different ways [7, 22, 23, 24]. Since the baseband signal is highly oversampled, the ADC can be implemented as a $\Delta\Sigma$ -ADC and thereby utilizing the the high oversampling ratio [23].

3.7 Receiver Architectures — Summary and Trends

As mentioned in the previous sections there are many requirements to fulfill when designing a radio receiver front-end. Sensitivity, selectivity, blocker immunity, power consumption, and at the same time minimize cost in terms of silicon area and external (off-chip) components. Despite the fact that the superheterodyne receiver is a well proven architecture with high sensitivity and selectivity, the cost penalty due to its complexity and need for external filters makes it less suitable for cost-driven applications.

The direct A/D-conversion receiver and the digital-IF receiver are not likely to

be used due to the extremely tough requirements on the ADC usually resulting in unacceptably high power consumption. However, according to *ITRS* long-term predictions [25] it is expected that full-digital implementations of receivers in CMOS will replace most analog functions except for the ADC.

Presently, the homodyne and the low-IF receivers appear to be the most popular architectures. The homodyne and low-IF result in the most simple filter implementations and have the lowest power consumptions. With some extra resolution in the ADC, some of the automatic gain control (AGC) can be achieved with low penalty further simplifying the design.

The RF-sampling receiver is a recent newcomer. Instead of traditional analog baseband processing it uses discrete-time baseband processing which appears to be in line with the use of fast deep-submicron CMOS processes with low supply voltages.

No matter which one of the homodyne, low-IF, or RF-sampling receiver architecture that is used, the most important and difficult design task is to achieve a good balance between gain and noise of individual blocks, baseband filtering, ADC requirements, and digital processing when partitioning a receiver. The receiver topology that best supports the implementation of flexible reconfigurable radio front-ends at the lowest cost will be used. With continued improvement in ADC performance the direct A/D-conversion receiver might become an option in the near future. However, it is required that the power consumption of those high performance ADCs is reduced. From a fundamental point of view, the power consumption of such ADCs could reach acceptable levels [26].

3.8 References

- [1] B. Razavi, RF Microelectronics. Prentice Hall, 1998.
- [2] T. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*. Cambridge University Press, 1998.
- [3] J. Crols and M. Steyaert, CMOS Wireless Tranceiver Design. Kluwer Academic Publishers, 1997.
- [4] K. Chang, *RF and Microwave Wireless Systems*. John Wiley & Sons, Inc., 2000.
- [5] B. Razavi, "Architectures and Circuits for RF CMOS Receivers," in Proc. of the IEEE Custom Integrated Circuits Conf., pp. 393–400, 1998.
- [6] J. B. Hagen, *Radio-Frequency Electronics, Circuits and Applications*. Cambridge University Press, 1996.

- [7] D. Jakonis, Direct RF Sampling Receivers for Wireless Systems in CMOS Technology. PhD thesis, Linköpings universitet, Department of Electrical Engineering, SE-581 83 Linköping, Sweden, 2004.
- [8] M. Windisch and G. Fettweis, "On the Performance of Standard-Independent I/Q Imbalance Compensation in OFDM Direct-Conversion Receivers," in *Proc. of the 13th European Signal Processing Conf.*, 2005.
- [9] J.P.F. Glas, "Digital I/Q Imbalance Compensation in a Low-IF Receiver," in *Proc. of the Global Telecommunications Conf.*, pp. 1461–1466, 1998.
- [10] G. Xing, M. Shen, and H. Liu, "Frequency Offset and I/Q Imbalance Compensation for Direct-Conversion Receivers," *IEEE Tran. on Wireless Communications*, vol. 4, no. 2, pp. 673–680, March 2005.
- [11] X. Li and M. Ismail, *Multi-Standard CMOS Wireless Receivers: Analysis* and Design. Kluwer Academic Publishers, 2002.
- [12] A. Abidi, "Direct-Conversion Radio Tranceivers for Digital Communications," *IEEE Journal of Solid-State Circuits*, vol. 30, no. 2, pp. 1399–1409, February 1995.
- [13] B. Razavi, "Challenges in Portable RF Tranceiver Design," *IEEE Circuits and Devices Magazine*, vol. 12, pp. 12–25, September 1996.
- [14] H. Darabi and J. Chiu, "A Noise Cancellation Technique in Active RF-CMOS Mixers," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 12, pp. 2628–2632, December 2005.
- [15] R. S. Pullela, T. Sowlati, and D. Rozenblit, "Low Flicker-Noise Quadrature Mixer Topology," in *Proc. of the ISSCC 2006 Conf.*, pp. 466–467, 2006.
- [16] R. Baker, CMOS Mixed-Signal Circuit Design. IEEE Press, Wiley Interscience, 2002.
- [17] S. Mirabbasi and K. Martin, "Classical and Modern Receiver Architectures," *IEEE Communications Magazine*, vol. 38, no. 11, pp. 132–139, November 2000.
- [18] J. Crols and M. Steyaert, "A Single-Chip 900 MHz CMOS Receiver Front-End with a High Performance Low-IF Topology," *IEEE Journal of Solid-State Circuits*, vol. 30, no. 12, pp. 1483–1492, December 1995.
- [19] S. Lindfors, CMOS Baseband Integrated Circuit Techniques for Radio Receivers. PhD thesis, Helsinki University of Technology, 2000.

- [20] F. Maloberti, "High-Speed Data Converters for Communication Systems," *IEEE Circuits and Systems Magazine*, vol. 1, no. 1, pp. 26–36, 2001.
- [21] B. Brannon, "Using Wide Dynamic Range Converters for Wideband Radios," *RF Design*, pp. 50–65, May 1995.
- [22] K. Folkesson, ADC Modeling from a System Perspective and Design of RF-Sampling Radio Receivers. PhD thesis, Linköpings universitet, Department of Electrical Engineering, SE-581 83 Linköping, Sweden, 2004.
- [23] K. Muhammad, R.B. Bogdan, and D. Leipold, "Digital RF Processing: Toward Low-Cost Reconfigurable Radios," *IEEE Communications Magazine*, pp. 105–113, August 2005.
- [24] H. Lakdawala, J. Zhan, A. Ravi, S. Andersson, B.R. Carlton, R.B. Nicholls, N. Yaghini, R.E. Bishop, S.S. Taylor, and K. Soumyanath, "Multi-band (1-6GHz) Sampled, Sliding-IF Receiver With Discrete-Time-Filtering in 90nm Digital CMOS Process," in *Proc. of the VLSI Symposium Conf.*, 2006.
- [25] ITRS, http://public.itrs.net, 2006.
- [26] M. Ismail and D. Rodríguez de Llera González, eds., Radio Design in Nanometer Technologies. Springer, 2006.
Chapter 4

Design of Low-Noise Amplifiers for Multiband Wireless Receivers

4.1 Introduction to Multiband LNAs

The simplest multiband wireless receivers are built using one front-end for every frequency band and the multiband functionality is achieved by having a number of stacked front-ends [1, 2]. Considering the increased number of standards to be covered by a single terminal and that every narrowband front-end uses 2-6 inductors [3], the silicon area and cost will be huge. Such obsolete solutions are not cost-effective [2]. The increased number of wireless systems calls for more flexible radio front-ends, for example the one shown in Fig. 4.1 [4, 5]. The design of more flexible front-ends will be further discussed in Chapter 5.

One of the most critical components in a multiband multistandard receiver is the LNA. This LNA must be capable of handling several carrier frequencies within



Figure 4.1: Sampling receiver front-end architecture [5].

a large bandwidth with the same performance requirements as if an LNA optimized for just one carrier frequency was used. Therefore, an LNA with sufficient gain, a low noise figure, and impedance-matching over a wide frequency range is needed. LNAs exploiting inductors normally fulfill these requirements relatively easy, but only in a very narrow frequency band around resonance. Hence, they are not very useful in these types of applications. However, there do exist dual band versions of this type of LNA such as in [6], covering two different frequency bands. Using many frequency selective LNAs in parallel is a typical method to achieve multiband functionality. Due to the increased number of standards this approach requires many inductors and thereby a large chip area [7], resulting in too high silicon cost. Therefore it is important to minimize the number of inductors used in the LNA. In this chapter focus will be on LNAs suitable for wideband multistandard wireless receivers. Such LNAs are important building blocks for low cost and highly integrated multistandard receivers. It is also important to find LNA topologies suitable for a high level of integration in CMOS. Other important aspects are to find solutions that are more tolerant against process variations, since this is expected to be an increased problem in the future [8] and the lack of good models, which is a very common Design Kit issue these days. Both wideband LNAs and tunable narrowband LNAs have been designed and are presented in this chapter.

4.2 Low-Noise Amplifier Requirements and Performance Metrics

A traditional RF front-end typically consists of a low-noise amplifier (LNA), mixers, a local oscillator (LO), filters, and finally an analog-to-digital converter (ADC). As an example, a typical single-conversion receiver architecture can be seen in Fig. 4.2. The first stage after the antenna and RF filter in the receiver chain is typically an LNA. The main function of the LNA is to provide high enough



Figure 4.2: Single conversion receiver architecture (homodyne).

gain to overcome the noise of subsequent stages (mixer etc.), while adding as little noise as possible. At the same time it should be linear enough to handle strong interferers without introducing intermodulation distortion. The dynamic range is therefore determined by the added noise at the low end and by non-linearities at the high end. All these LNA performance metrics aim for a sufficient dynamic range. An LNA should also in most cases present a specific impedance level to the input source and sometimes to the output load as well. Another important design parameter for many applications is the power consumption. It is particularly important for battery driven applications, where the battery life time is crucial.

4.2.1 Small-Signal Parameters

The typical small-signal design parameters for an LNA are gain and noise figure. To understand why these two parameters are important the overall noise figure of the receiver has to be studied. The noise figure of cascaded systems depends both on the individual noise figures and the gains. This is illustrated in *Friis' formula* [9]:

$$F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \dots + \frac{F_N - 1}{G_1 G_2 \dots G_{N-1}}$$
(4.1)

The total noise factor is the sum of the contribution from all stages, where the noise factor of stage i is divided by the total gain from the input to stage i. The gain dependency results in the fact that once the signal is amplified, the noise of subsequent stages is less important. Therefore, the system noise figure is often dominated by the noise from the first stage(s) in the receiver chain. Trying to achieve a good noise figure, most of the design effort must be put on the earliest stages in the receiver chain, particularly on the LNA.

The input impedance (and perhaps also output impedance) are also important when designing an LNA. The impedance levels are normally given by the specification. The main reason for having a certain input impedance is due to the passive band-selection filter between the antenna and the LNA. The transfer characteristics of many filters are sensitive to the quality of the termination [10]. Mismatch between filter and LNA could therefore destroy the filter function. The most common input impedance level is 50 Ω for a single-ended input and 100 Ω for differential inputs.

4.2.2 Large-Signal Parameters

Besides gain, noise, and input match, linearity is another important design consideration. It is true that the main task for an LNA is to provide gain without adding too much noise itself, but it must be able to do this even when strong signals are



Figure 4.3: Definition of 1-dB compression point (CP).

being received. If the wanted signal is strong it can just be passed by the LNA directly to the subsequent component and in this case the linearity of the LNA does not matter. The worst case occur when the LNA is receiving a weak signal in the presence of a strong interferer and the LNA must maintain its linearity. Nonlinearities will then result in intermodulation distortion, desensitization (blocking), and cross-modulation. Cross-modulation results when nonlinear interaction transfers the modulation of one signal to the carrier of another.

There are many measures of linearity, but the most commonly used are the 1-dB compression point (CP) and the third-order intercept point (IP3). For a visual definition of the 1-dB compression point (CP) see Fig. 4.3 and for IP3 see Fig. 4.4.

The straight forward specification of the upper power limit of an amplifier is the 1-dB compression point. At this point the gain is 1 dB lower than the small-signal gain due to nonlinearities. This is because the linear small-signal assumption is not valid for higher input powers. This will reduce the gain of a weak signal in presence of a strong interferer (blocker). In receivers the most troublesome of the intermodulation products is of the third order. The special problem with the third-order intermodulation terms is that they can fall within the wanted frequency band and interfere with the wanted signal. Second-order intermodulation is not so troublesome because the products can be filtered away [11]. However, this is not true in direct-conversion (homodyne) receivers, where the second-order intermodulation is equally important. The use of differential LNA architectures removes



Figure 4.4: Definition of third-order intercept point (IP3).

much of the second-order intermodulation distortion. In an ideal differential LNA there will not exist any second-order intermodulation. For weak nonlinearities the third-order intercept point and the 1-dB compression point are strongly correlated. The relationship between the 1-dB compression point and IIP3 is [3]:

$$IIP3 - CP_{-1dB} = 9.64 \, dB \tag{4.2}$$

The standard technique to measure or simulate IIP3 is a two-tone test, which uses two closely spaced signals of equal amplitude. The third-order intermodulation products of the output spectrum are compared with the fundamental terms as the input amplitude varies and the intercept point is then calculated, as illustrated in Fig. 4.4.

4.3 LNA Design

Two different approaches for implementing multiband multistandard LNAs have been investigated:

- Wideband LNAs that cover all frequency bands of interest.
- Tunable narrowband LNAs, tunable over the frequency bands of interest.



Figure 4.5: Transfer characteristics for wideband and tunable narrowband LNAs.

Wideband LNAs have gain over a large bandwidth and tunable narrowband LNAs have a smaller bandwidth that can be tuned within a wide frequency range, see Fig. 4.5. Both types of LNAs have been implemented and verified by measurements.

4.3.1 Design of Wideband LNAs

Elementary wideband amplifiers such as the shunt-series feedback amplifier and the common-gate amplifier in Fig. 4.6 show a severe trade-off between their noise figure and the input impedance matching requirement [12]. The reason behind this is that for both stages in Fig. 4.6 the transconductance, g_m , is more or less predefined due to the required input impedance for matching. The transconductance is also the parameter that sets the noise level in these circuits. A large g_m is required for low noise, but since g_m already is set by the input impedance requirement there is no freedom left to optimize for low noise. Obviously, to find a solution to this we need to find a way to decouple the requirement on input matching from the overall noise properties of the LNA. One way to do this is the "noise cancelling technique" proposed in [12] and shown in Fig. 4.7. It is not really as good as it sounds. All noise is of course not cancelled. However, the signals at node X and Y in Fig. 4.7 have opposite signs, while the noise at the nodes X and Y have equal



Figure 4.6: Common wideband LNAs. To the left a shunt-series feedback amplifier and to the right a common-gate amplifier.



Figure 4.7: "Noise-cancelling" wideband LNA [12].

signs. This means that for the case when:

$$\frac{g_{m2}}{g_{m3}} = 1 + \frac{R}{R_s} \tag{4.3}$$

the noise from the first stage cancels and the noise performance can be set by the design of the second stage. Note that this is only true if Eq. 4.3 is fulfilled.

Fig. 4.8 shows a wideband version of the G_m -boosted common-gate LNA topology described in [13]. Even this is an attempt to separate the input matching requirement from the noise performance. The benefit is best shown by comparing the noise factor of the traditional common-gate LNA with the G_m -boosted



Figure 4.8: G_m -boosted common-gate LNA.

common-gate LNA. The noise factor of a traditional common-gate LNA is [13] (neglecting the load resistor):

$$F = 1 + \frac{\gamma}{\alpha} \cdot \frac{1}{g_{m,M1}R_s} \tag{4.4}$$

For the G_m -boosted common-gate LNA the noise factor is described as [13] (neglecting the load resistor):

$$F = 1 + \frac{\gamma}{\alpha} \cdot \frac{1}{(1+A)^2 g_{m,M1} R_s}$$
(4.5)

Input matching requires $(1 + A)g_{m,M1} = 1/R_s$. Using this in Eq. 4.5 the noise factor becomes:

$$F = 1 + \frac{\gamma}{\alpha} \cdot \frac{1}{(1+A)} \tag{4.6}$$

since $g_{m,M1}R_s = 1$ for the traditional common-gate LNA due to the input matching requirement, the noise factor is reduced by 1/(1+A) when using the G_m boosting technique. Of course another amplifier with gain A will also introduce noise. However, A = 1 is easily achieved in a differential topology by just crosscoupling the inputs.

Another common method to implement wideband LNAs is to use resistive feedback [14, 15]. A large feedback resistor is required for low noise. However, a



Figure 4.9: The first proposed wideband LNA.

large feedback resistor requires high gain which affects the linearity in a negative way [14]. On the other hand, to some extent feedback also helps linearizing the LNA. One of the main drawbacks with resistive-feedback LNAs is that the circuit might become unstable. Such LNAs have to be designed carefully to avoid instability problems.

Our first proposed wideband LNA is described in **Paper 4** ([16]) and is shown in Fig. 4.9. A differential topology is chosen for its robustness and lower sensitivity to supply noise. The input stage is a common-source amplifier with a common-drain feedback stage, Fig. 4.9. The gain of the input stage is:

$$A_v = -g_{m,M1}R_1 (4.7)$$

Bandwidth enhancement with inductive shunt-peaking is used by connecting an inductor (L₁) in series with the resistive load (R₁), [17]. The cascode transistor M_4 is used to reduce the Miller effect and to improve the reverse isolation. By using a common-drain feedback stage a wideband impedance-matching to 50 Ω is achieved. The input impedance is given by (reactive components excluded):

$$Z_{in} = \frac{1}{g_{m,M2}(1+|A_v|)} = \frac{1}{g_{m,M2}(1+g_{m,M1}R_1)}$$
(4.8)

The inductor on the input side, L_2 , is used for optimizing the frequency behavior of the input impedance. Transistor M_3 together with resistor R_2 act as a current source. This topology has similar benefits as the noise cancelling technique proposed in [12]. The common-drain feedback stage used here devises another way to achieve a similar result. The chosen circuit topology gives the designer an extra degree of freedom when designing for a low noise figure and at the same time meet the required input impedance. Here, the input impedance is controlled by the feedback transistor (M_2), Eq. 4.8, whereas the noise figure and gain is controlled by the input transistor (M_1). Besides that, our solution partly cancels input transistor noise due to the feedback.

A theoretical expression for the noise figure of a single-ended version of our amplifier (reactive components excluded) is derived from [18]:

$$F = \frac{\sum v_n^2}{v_{n,R_S}^2} \tag{4.9}$$

where $\sum v_n^2$ is the total input-referred noise due to all noisy components (transistors and resistors) and v_{n,R_s}^2 is the noise due to the input source alone. Ideally, R_2 does not add any noise since it is in series with a current source with infinite impedance. Using Eq. 4.7 and Eq. 4.8 together with the assumption $Z_{in} = R_s$ (R_s is the source impedance) gives the following expression for the noise figure:

$$F = 1 + \left(1 + \frac{1}{1 + |A_v|}\right)^2 \cdot \frac{\gamma_1}{4 \cdot g_{m,M1} \cdot R_s} + \frac{\gamma_2}{4 \cdot (1 + |A_v|)} + \frac{g_{m,M3} \cdot R_s \cdot \gamma_3}{4} + \frac{R_1}{4 \cdot R_s \cdot (1 + |A_v|)^2}$$
(4.10)

where γ_1 , γ_2 , and γ_3 are the noise factors of M_1 , M_2 , and M_3 respectively [18]. The second term (including $g_{m,M1}$) represents the noise contribution from M_1 . The equivalent input noise voltage is partly cancelled (reduced by a factor of 4) through the feedback loop (half the M_1 equivalent noise voltage appear with opposite sign at transistor input via M_1 and M_2). The third term represents the noise contribution from M_2 . It is quite small for a reasonable amount of gain, A_v , and it is also reduced by a factor of 4. The two last terms represent noise from M_3 and R_1 respectively. $g_{m,M3}$ can be made small and does not contribute much to the overall noise figure.

This proposed circuit topology was implemented using a 0.18 μ m technology. A gain of 13.1 dB was measured together with an almost 7 GHz 3-dB bandwidth. The noise figure was 3.6 dB at 2 GHz and increased to about 5.5 at 6 GHz. More details on this implementation is found in **Paper 4**. The inductors used in Fig. 4.9 are not really necessary and had very little impact on the circuit performance. The inductors (L₂ in Fig. 4.9) could successfully be replaced by resistors instead leading to a more wideband input matching and without degrading the noise figure much. According to the measurement results in **Paper 4** the inductors used for shunt-peaking had very little effect on the bandwidth, or none at all. The cause for that might be that more parasitics are added. Another reason for the lower bandwidth might be the inductors' frequency dependency, where both Q value



Figure 4.10: The second proposed wideband LNA.

and inductance tend to decrease at higher frequencies [19]. About two years after the measurements, the LNA was re-simulated with new updated RF-models for both transistors and inductors from the foundry. The new simulations matched the measured results very well.

Our second implementation of a wideband LNA uses the same approach as the first one with a few exceptions. The schematic of the LNA is shown in Fig. 4.10. The output buffers are not shown here but are implemented as two common-drain stages. First of all no inductors are used and thereby the total area is very small, only 0.019 mm² (buffers included). From the first design (Fig. 4.9) it is known that a large transconductance is needed for the input transistor. This implies a large transistor and a certain amount of bias current to maintain a reasonable linearity. Pulling this bias current through the resistive wideband load, R_D , causes a voltage drop, which reduces the voltage headroom and thereby drives the cascode transistor into the linear region. This DC voltage drop is the problem with using resistive wideband loads [7]. Reducing the resistance of R_D is not an option since that would reduce the gain. For the first implementation in 0.18 μ m CMOS with a supply voltage of 1.8 V the voltage drop across the load resistor was not a limiting factor. However, for the second implementation in 0.13 μ m CMOS with a supply of 1.4 V the voltage drop would have been too large. As seen in Fig. 4.10 a PMOS transistor (M_5) is introduced. This transistor is primarily intended to feed a bias current to the main input transistor (M_1) so that not all of the current has to flow through the load resistor and the cascode transistor (M_4). By connecting the gate of M_5 to the input it will also contribute to the overall gain, even though the transconductance of the PMOS transistor is much smaller than for M₁. But, more important is that the noise contribution from M5 is reduced when connected to the input. The LNA shown in Fig. 4.10 is well suited even for low supply voltages. The total gain of the LNA shown in Fig. 4.10 without including reactive components can be expressed as:

$$A_v = -\frac{(g_{m,M1} + g_{m,M5})}{\frac{1}{R_D} + \frac{(1/R_D + g_{ds,M4})(g_{ds,M1} + g_{ds,M5})}{(g_{m,M4} + g_{ds,M4})}}$$
(4.11)

By assuming $g_{m,M4} \gg g_{ds,M4}$ Eq. 4.11 can be simplified to:

$$A_v = -(g_{m,M1} + g_{m,M5})R_D \cdot \frac{g_{m,M4}}{(g_{m,M4} + g_{ds,M1} + g_{ds,M5})} = -G_m R_D \quad (4.12)$$

where:

$$G_m = (g_{m,M1} + g_{m,M5}) \cdot \frac{g_{m,M4}}{(g_{m,M4} + g_{ds,M1} + g_{ds,M5})}$$
(4.13)

The common-drain feedback stage, consisting of M_2 and R_1 , provides wideband impedance matching to the source impedance. The feedback transistor M_2 is AC-coupled to be biased independently of the DC-level at the output. The output of the LNA drives the gate of M_2 with a signal of large amplitude. This large signal-swing at the output leads to a non-linear response from the feedback part resulting in poor linearity of the entire LNA. To circumvent this undesired behavior, the resistor (R_1) is added in series with M_2 to improve the linearity of the feedback part. The cross-coupled capacitors (C_X) between the inputs and the drain-side of M_1 improve the input matching at high frequencies. In this way some of the input capacitance is neutralized [10]. Ignoring the reactances, the input impedance can be expressed as:

$$Z_{in} = \frac{1 + g_{m,M2}R_1}{g_{m,M2}(1 + G_m R_D)} = \frac{1 + g_{m,M2}R_1}{g_{m,M2}(1 + |A_v|)}$$
(4.14)

From Eq. 4.13-4.14 it is seen that the gain is set by the common-source amplifier and the input impedance by the common-drain feedback stage.

To calculate the noise factor and highlight the noise cancelling mechanism for the LNA shown in Fig. 4.10, every transistor and resistor is assumed to produce a noise voltage appearing at the output. A scaled and bandlimited version of the same noise is fed back to the input by the common-drain feedback stage, where it gets amplified and inverted before appearing at the output. Therefore, at the output some of the correlated noise is cancelled and the overall output noise is reduced. All resistors and transistors except M_3 are within the feedback loop and their noise is partly cancelled due to the feedback. The noise from M_3 appears directly at the input and since this transistor is outside the feedback loop this noise is just amplified and summed at the output. Expressing the noise factor of the LNA shown in Fig. 4.10 a similar expression as Eq. 4.10 can be derived. Assuming perfect input matching and ignoring all reactive components, the noise factor can be expressed as:

$$F = 1 + \left(1 + \frac{1}{1 + |A_v|}\right)^2 \cdot \left(\frac{\gamma_1}{4 \cdot g_{m,M1} \cdot R_s} + \frac{\gamma_5}{4 \cdot g_{m,M5} \cdot R_s}\right) + \frac{\gamma_2}{4 \cdot (1 + g_{m,M2} \cdot R_1) \cdot (1 + |A_v|)} + \frac{\gamma_3 \cdot g_{m,M3} \cdot R_s}{4} + \frac{R_1}{4 \cdot R_s \cdot (1 + |A_v|)^2} + \frac{R_D}{4 \cdot R_s \cdot (1 + |A_v|)^2}$$
(4.15)

where γ_1 , γ_2 , γ_3 , and γ_5 are the noise factors of M₁, M₂, M₃, and M₅ respectively [18]. The same reasoning as for the previously described LNA regarding the different sources of noise can be applied here as well. The main difference between Eq. 4.10 and Eq. 4.15 is the addition of the noise added by R₁. However, the impact of the noise from this resistor is small as long as the gain is reasonably high. Far more important is to keep $g_{m,M3}$ as small as possible. This is possible since $g_{m,M3}$ has nothing to do with neither gain or input impedance. More about the implementation of this LNA can be found in **Paper 5**.

4.3.2 Design of Widely Tunable LNAs

The front-ends of most wireless receivers need a band-select filter and an LNA. The band-select filter is very often an off-chip component, i.e. a surface-acousticwave (SAW) filter or an LC-filter. For some cases an on-chip LC-filters is good enough together with the LNA. External LC-components have the benefit of high Q value but they are on the other hand extremely expensive to use. Integrated LC-components are preferred, but especially the inductors suffer from poor Q values due to parasitics and only small inductances (< 10 nH) should even be considered to integrate [10]. It would be even better if the inductors could be completely avoided as they require a special RF process with extra thick top metal layer in order to get a reasonably good Q value. From a cost perspective it would therefore be nice if a purely digital process could be used without any RF option (for CMOS implementations). Another drawback with using inductors is that the size of the inductors normally occupies much more area than the active circuitry itself. Furthermore, it would also be very attractive if the combined bandpass filter and LNA could be electronically tunable in both frequency and bandwidth. Even if an off-chip SAW filter cannot be avoided it is in some cases still desirable to have a bandpass transfer function of the LNA. For example to avoid that wideband noise from the LNA is folded in an RF-sampling receiver front-end.

Active recursive bandpass filters have recently become interesting for microwave applications [20]. A traditional recursive bandpass filter is accomplished



Figure 4.11: Traditional microwave model of an active recursive filter.



Figure 4.12: CMOS/BiCMOS block level design of Fig. 4.11.

according to the model in Fig. 4.11. This model is based on positive feedback, where some of the output power is fed back to the input through a delay (τ). α is the input coupling factor and β is the feedback coupling factor. By varying the delay in the feedback loop a phase shift is accomplished and the center frequency (resonance frequency) is changed. In microwave applications, bandpass filters are aimed for replacing the combination of LNAs and resonator-based filters, leading to smaller area and considerably increased tunability. Let us now see how the principle of active recursive filters can be applied to a standard silicon process (CMOS or BiCMOS) without using inductors, resonators, passive combiners, or transmission lines. To do this all passive elements in the traditional microwave model in Fig. 4.11 are replaced with active circuitry. We have chosen a scheme according to Fig. 4.12, where the input stage power combiner is replaced by an amplifier with two differential inputs. The necessary delay is replaced by the delays through the amplifier stages. Frequency tunability is achieved through variation of the amplifier delay, which can be studied through the amplifier time constant, τ . The delay through the amplifier chain corresponds to a phase shift ϕ_{delay} according to:

$$\phi_{delay} = \pi + 2\pi n = \omega_c \tau = 2\pi f_c \tau, \qquad (4.16)$$

where ω_c is the angular center frequency. By connecting the positive/negative

output to the negative/positive input, a 180° low-frequency phase shift is provided. Together with the delay through the amplifier chain, the necessary phase shift of 360° (and its multiples) is obtained. The multiples mentioned are strongly suppressed due to the limited bandwidth of the amplifiers. This implies:

$$f_c = k/\tau \tag{4.17}$$

where k is a positive integer. The proposed scheme in Fig. 4.12 can be viewed as a "non-oscillating ring oscillator" utilizing the same frequency tuning mechanism as used in VCOs based on ring oscillators [18, 21, 22]. The delay of each amplifier is proportional to its unity-gain frequency:

$$\tau = C_L/g_m \tag{4.18}$$

The time constant τ can thus be controlled by the bias current through each amplifier. This is due to the bias dependency of g_m . The Q value can be tunable through variation of the loop gain. Low gain broadens the amplification peak, which results in a lower Q value. Both gain and Q value are strongly correlated.

Three versions have been implemented, two in CMOS and one in BiCMOS, to demonstrate the feasibility and concept of this new approach. The first one is a widely tunable LNA implemented in 0.8 μ m CMOS and is described in detail in **Paper 1** ([23]). The second implementation, described in **Paper 2** ([24]), is a tunable RF filter implemented in 0.35 μ m BiCMOS. The reason for not calling this circuit an LNA as well is the relatively high noise figure of this particular implementation. While the two first circuit implementations are rather similar from a circuit perspective the third one is improved to solve some of the issues associated with the first two. The third one is implemented in 0.18 μ m CMOS and is further described in Paper 3 ([25]). The first two circuits have a common-gate and common-base input stage, respectively. This configuration is good for wideband input matching, but has the disadvantage of relatively high noise figure as discussed earlier for wideband LNA implementations. This is one of the reasons for the relatively high noise figure, particularly for the second implementation (Paper 2). The first circuit (Paper 1) also has the disadvantage that the frequency tuning affects the input impedance. This since the current through the commongate input-stage is changed while tuning the frequency. The third implementation (Paper 3) uses an input-stage similar to the wideband amplifier implementations previously described in this chapter. This resulted in a more consistent inputmatching and lower noise figure over the whole frequency tuning range.

Due to the chosen way to control the gain in the two first circuits, **Paper 1** and **Paper 2**, by using a cross-coupled transistor between the outputs of the amplifiers (Fig. 7.3 and 8.3), the linearity was affected. These cross-coupled transistors are nonlinear resistors that generates (and amplifies) harmonics. As a result relatively



Figure 4.13: Schematic of the amplifier/delay stages used in Paper 3.

large third-order harmonics are created that has a slope of two instead of three as expected when measuring IIP3. Another problem was that when tuning the center frequency by adjusting the current levels in the amplifier the DC-operating point for the cross-coupled transistors controlling the gain was changed. This resulted in difficulties since the frequency tuning affected the gain tuning and vice versa. In the third implementation a completely different strategy for controlling the gain was used to solve these problems. Fig. 4.13 shows the schematic of the amplifier/delay stages used in **Paper 3**. Frequency tunability is done by changing $g_{m.in}$ and is thus achieved by changing the bias current through the transistor marked *biasn* in Fig. 4.13. The gain should also be tunable. Therefore, the load impedance of the amplifier in Fig. 4.13 consists of an NMOS transistor in parallel with a PMOS transistor. The NMOS device is connected in a diode-coupled way, except that the gate is connected to a voltage Vb. The NMOS loads therefore have an output impedance approximately equal to $1/g_m$, assuming $g_m \gg g_{ds}$. The output impedance of the NMOS loads are therefore proportional to the amount of current flowing through them. This current is on the other hand adjustable by the PMOS transistor in parallel acting like a high-impedance current source. This because the total current through the load is determined by the tail-current of the NMOS current source at the bottom. The more of the current that is provided by the PMOS current source the higher the total output impedance is. By building like this, a circuit where the center-frequency tuning is independent from the gain tuning can be made. This makes frequency and gain tuning much easier to handle.

All three designs show excellent frequency tuning range and both gain and Q value can be tuned in a broad range. When designed with a low enough noise

figure, acceptable linearity, and high enough Q value such an LNA could be used without an external filter between antenna and LNA. This would be an excellent LNA for multistandard receivers.

4.4 Aspects on LNA Design and Noise in Deep Submicron CMOS Technology

The ongoing scaling of CMOS devices with shorter and shorter channel lengths has opened the opportunity to successfully use them in RF parts, for example in wireless systems. Unfortunately, the development of accurate noise models for deep submicron devices has not kept the same pace as the scaling, leading to difficulties for RF designers. For RF designs accurate noise models is a must to succeed without too many iterations. This is important from a time perspective as well as from a cost perspective. The most common way to deal with this problem nowadays is to use measured data for fixed device sizes and layouts of transistors. The main problem with this method is that the designer is tied to these nonscalable devices, limiting the degree of freedom when designing. In **Paper 6** we performed our own noise measurements on NMOS devices with different channel lengths to be able to compare with simulated data. This was done in order to understand how to utilize transistor length as a design parameter to achieve noise figures as low as possible.

It is well known that the thermal noise from MOSFETs normally is described with the formula:

$$\overline{i_n^2} = 4k_B T \gamma g_{d0}, \tag{4.19}$$

where k_B is the Boltzman constant, T is the temperature in Kelvin, g_{d0} is the conductance at zero V_{DS} , and γ is a constant. For long channel devices γ equals 2/3 when operating in saturation and $g_{d0} = g_m$, therefore g_m is sometimes used in Eq. 4.19. From papers published recent years large values of γ have been observed for devices using minimum channel length for a certain technology. However, when moving to a technology with smaller feature size the value of γ seems to decrease for a given channel length, while still remain very high (or increase somewhat) for the minimum channel length. From our measurements, large discrepancies between measured and simulated values of γ were observed, Fig. 4.14-4.15. The low-frequency 1/f noise (flicker noise) can usually be neglected when designing high-frequency LNAs and is not discussed here.

For low-noise RF designs the most important thing is to optimize the AC performance with the respect to the noise performance. Therefore, good knowledge about how the noise changes with channel length is of great importance. The optimal choice of channel length is not necessarily the minimum since the noise



Figure 4.14: Measured γ vs. V_{GS} , $V_{DS} = V_{DD}$.



Figure 4.15: Simulated γ vs. V_{GS} , $V_{DS} = V_{DD}$.

properties of the transistor changes rapidly as the channel length is decreased. The actual noise performance on circuit level might be improved by not using the minimal channel length, despite the fact that g_m is decreased. From our own noise measurements, **Paper 6**, we draw the conclusion that today's noise models are not good enough to model noise from deep submicron devices. It also appears that the noise figure can be improved by not using the minimum channel length. Uncertainties in our noise measurements can be the reason for this, however it is clear that transistors with shorter channel lengths tend to be noisier. This theory is also supported by the recent work presented in [26]. In [26] NF_{min} and R_n are mea-

sured versus channel length at 5 GHz for a 90 nm technology, thus all devices are manufactured in the same lithographic node as in our case. Fig. 1 in [26] shows that R_n has a minimum for about three times the minimum drawn channel length. Of course there are also drawbacks with increasing the channel length such as increased loading, lower gain, and lower bandwidths. This may still mean that the minimum channel length results in the best performance.

4.5 References

- S. Mattisson, "Architecture and Technology for Multistandard Tranceivers," in *Proc. of the IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 82–85, 2001.
- [2] M. Vidojkovic, J. van der Tang, P. Baltus, and A. van Roermund, "Design of Flexible RF Building Blocks — A Method for Implementing Configurable RF Tranceiver Architectures," in *Proc. of Asia-Pacific Conference on Communications*, pp. 445–449, 2005.
- [3] Adiseno, Design Aspects of Fully Integrated Multiband Multistandard Front-End Receivers. PhD thesis, Kungliga Tekniska Högskolan, Stockholm, 2003.
- [4] K. Folkesson, D. Jakonis, J. Dabrowski, and C. Svensson, "Design of RF Sampling Receiver Front-End," in *Proc. of the International Conference Mixed Design of Integrated Circuits and Systems*, pp. 538–543, 2004.
- [5] D. Jakonis, Direct RF Sampling Receivers for Wireless Systems in CMOS Technology. PhD thesis, Linköpings universitet, Department of Electrical Engineering, SE-581 83 Linköping, Sweden, 2004.
- [6] H. Hashemi and A. Hajimiri, "Concurrent Multiband Low-Noise Amplifiers - Theory, Design, and Applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 1, pp. 237–244, January 2002.
- [7] J. Ryynänen, S. Lindfors, K. Stadius, and K.A.I. Halonen, "Integrated Circuits for Multi-Band Multi-Mode Receivers," *IEEE Circuits and Systems Magazine*, vol. 6, issue 2, pp. 5–16, 2006.
- [8] X. Wang, J. Pekarik, X. Yuan, R. Anna, and S. Parker, "Meeting the Design Challenges in Modern RFCMOS Technology," in 7th International Conference on Solid-State and Integrated Circuits Technology, pp. 1291–1294, 2004.

- [9] R. Ludwig and P. Bretchko, *RF Circuit Design, Theory and Applications*. Prentice-Hall inc., 2000.
- [10] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*. Cambridge University Press, 1998.
- [11] J. B. Hagen, *Radio-Frequency Electronics, Circuits and Applications*. Cambridge University Press, 1996.
- [12] F. Bruccoleri, E.A.M. Klumperink, and B. Nauta, "Noise Cancelling in Wideband CMOS LNAs," in *Proc. of the IEEE International Solid-State Circuits Conf.*, pp. 406–407, 2002.
- [13] D.J. Allstot, X. Li, and S. Shekhar, "Design Considerations for CMOS Low-Noise Amplifiers," in *Proc. of the Radio Frequency Integrated Circuits Symposium*, pp. 97–100, 2004.
- [14] J.-H.C. Zhan and S.S. Taylor, "A 5GHz Resistive-Feedback CMOS LNA for Low-Cost Multi-Standard Applications," in *Proc. of the IEEE International Solid-State Circuits Conf.*, pp. 200–201, 2006.
- [15] Adiseno, H. Magnusson, and H. Olsson, "A 1.8-V Wideband CMOS LNA for Multiband Multistandard Front-End Receiver," in *Proc. of the European Solid-State Circuits Conf.*, pp. 141–144, 2003.
- [16] S. Andersson, O. Drugge, and C. Svensson, "Wideband LNA for a Multistandard Wireless Receiver in 0.18 μm CMOS," in *Proc. of the* 29th European Solid-State Circuits Conf., pp. 655–658, 2003.
- [17] S.S. Mohan, M. del Mar Hershenson, S.P. Boyd, and T.H. Lee, "Bandwidth Extension in CMOS with Optimized On-Chip Inductors," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 3, March 2000.
- [18] B. Razavi, *Design of Analog CMOS Integrated Circuits*. McGraw-Hill, 2001.
- [19] S-M. Yim, T. Chen, and K.K. O, "The Effects of a Ground Shield on Spiral Inductors Fabricated in a Silicon Bipolar Technology," *IEEE Journal of Solid-State Circuits*, vol. 37, pp. 237–244, February 2002.
- [20] R. Malmqvist, *Tunable Recursive Active Monolithic Microwave Integrated Circuit Filters*. PhD thesis, Linköpings universitet, 2001.

- [21] A. Hajimiri, S. Limotyrakis, and T. Lee, "Jitter and Phase Noise in Ring Oscillators," *IEEE Journal of Solid-State Circuits*, vol. 34, pp. 790–804, June 1999.
- [22] D.A. Johns and K. Martin, Analog Integrated Circuit Design. Wiley, 1997.
- [23] S. Andersson, P. Caputa, and C. Svensson, "A Tuned, Inductorless, Recursive Filter LNA in CMOS," in *Proc. of the* 28th European Solid-State Circuits Conf., pp. 351–354, 2002.
- [24] S. Andersson and C. Svensson, "An Active Recursive RF Filter in 0.35 μm BiCMOS," *Journal of Analog Integrated Circuits and Signal Processing by Springer*, vol. 44, no. 3, pp. 213–218, September 2005.
- [25] S. Andersson and C. Svensson, "A 750 MHz to 3 GHz Tunable Narrowband Low-Noise Amplifier," in *Proc. of the Norchip 2005 Conf.*, pp. 8–11, 2005.
- [26] G. Banerjee, K. Soumyanath, and D.J. Allstot, "Measurement and Modeling Errors in Noise Parameters of Scaled-CMOS Devices," *IEEE Tran. on Microwave Theory and Techniques*, vol. 54, no. 6, pp. 2336–2345, June 2006.

Chapter 5

RF-Sampling Receivers

As discussed in Chapter 3 there are many things to consider when choosing a suitable RF front-end architecture. One key challenge is to find an architecture that can be used for multistandard radio receiver design. A high level of integration is also necessary to reduce the cost as well as having a manageable power consumption. The later requirement is particularly important for battery driven applications. Further assuming that the high-level of integration due to cost reasons implies a single-chip solution and that such an implementation is feasible. Then, a high-end deep-submicron CMOS technology is the only reasonable choice since the digital baseband requires a lot of signal processing capability together with a large amount of integrated memory [1].

Designing RF and analog circuits that are compatible with digital deep submicron CMOS technology is not easy. However, modern CMOS technology shows excellent RF performance for applications below 10 GHz [2]. Some of the excess performance in terms of transistor speed can be traded for analog and RF performance. By using the huge processing capability in a DSP it is possible to further ease analog and RF design by simplifying and relaxing the performance requirements of those components. Digital techniques can thus be used to alleviate the need for complex analog design. A few thousand extra gates come almost for free in modern deep submicron technologies [1]. Due to the continuous scaling of CMOS technology, other practical aspects on integration of CMOS radio front-ends are to take advantage of the increased transistor performance and ease technology node migration [3]. RF and analog circuits must also be designed early on immature processes. Most likely this means a lack of good models that are subject to frequent updates. Besides that, a digital CMOS process is likely designed for low cost and not optimized for RF. Due to cost reasons it will not be changed in order to improve performance of RF circuits. This means that conventional design techniques cannot be used. Analog and RF circuit area also has to shrink with process scaling to be cost competitive. However, analog and RF circuits do not scale very well [3]. For example, one inductor takes up an area corresponding to 10-100 kgates in a deep submicron technology of today [4]. Therefore, quite a lot of digital signal processing can be made per single passive RF component.

With this in mind, a lot of effort should be put on finding new RF front-end architectures. As stated in [3], digital radio solutions offer performance, low power consumption, high manufacturing yield, small area, and flexibility. Receiver frontends based on RF-sampling techniques have recently become popular to address some of the problems described above [1, 5, 6]. RF-sampling is of course not the only suggested option to implement flexible, or reconfigurable, radio receivers. Another suggested solution is to keep the LNA and mixer and make use of oversampling ADCs (e.g. $\Delta\Sigma$ -ADCs), where the digital decimation filter determines the channel bandwidth in combination with a switchable analog filter [4]. However, this solution still have the problem of requiring precise analog active filters using operational amplifiers at a low supply voltage. The rest of this chapter will be about RF-sampling and front-ends using this technique. In Section 5.3 our implementation of a wideband RF-sampling front-end is presented. The frequency synthesis is not included and is not further discussed except when looking at the effect of sampling-jitter-induced noise.

5.1 Architectural Trends for Wireless Receivers

As the technology of choice appears to be deep-submicron digital CMOS when ever that is applicable, this must be considered when deciding on front-end architecture. The low supply voltage in modern CMOS processes is extra troublesome for the analog circuits and has to be accounted for. For zero-IF receivers, circuit techniques that can reduce the effect of increased 1/f-noise due to process scaling are also needed [7, 8]. However, one of the more crucial questions on architectural level is whether to use analog channel-select filtering or not. Not performing the channel filtering in analog domain means letting the blocker reach the ADC and thus require a higher dynamic range in the ADC. The channel filtering is then performed in the digital domain instead.

The analog channel-select filter is there to help reduce the required dynamic range in the ADC. It also implements the required antialiasing filtering so that the ADC does not have to run that fast, i.e. a lower sampling frequency can be used. Since the power consumption of ADCs is proportional to the number of bits and the sampling frequency as $f_s 2^{2n}$ it clearly makes sense trying to reduce the number of bits and sampling frequency. However, due to process scaling traditional analog filters get worse due to the limited headroom available because of the low supply voltage [3]. Low-voltage techniques for analog components like filters have been suggested in [9] and a channel-filter for WLAN is presented in

[10]. One option is to design those type of circuits using thick-oxide MOSFETs commonly available for IOs that can tolerate higher supply voltages. However, these thick-oxide devices are in fact belonging to an older process generation. This is a very conservative approach and circuits built like that do not take benefit from process scaling. Besides that, with ongoing scaling those devices will themselves scale down until they cannot be used anymore [11]. For CMOS technologies at channel lengths of 90 nm and below, analog circuits like this filter will no longer be able to handle waveforms with a large dynamic range without a disproportionate rise in power consumption [11]. As a result, a point will be reached where it is wiser to spend the available power budget on an ADC with adequate dynamic range and bandwidth and instead perform the filtering in digital domain. The main problem with this is that GSM-like standards require a huge dynamic range. Removing the channel-select filter before the ADC thus increase the required number of bits in the ADC. On the flip side, process scaling also improves the speed of ADCs and necessary digital filtering needs less power. Utilizing highly oversampled $\Delta\Sigma$ -ADCs a very high resolution can be accomplished, particularly for narrowband signals (like GSM). The use of oversampled ADCs are also preferred since the antialiasing filtering is relaxed and maybe the RF filtering can be enough. According to ITRS [12] it is expected that full-digital implementations will replace most analog functions in the RF front-end. Therefore, sooner or later the channel-select filter might not be useful anymore.

5.2 Radio Receiver Front-Ends Utilizing RF Sampling

Designing multistandard radio is a major challenge in CMOS due to the low quality factor of tuned circuits [6]. Instead high frequency sampling can be used as an alternative to tuned circuits. This and the fact that passive SC filters can be made digitally reconfigurable makes RF-sampling very appealing in terms of building multistandard radio. Since the maximum sampling frequency increases with process technology scaling, the concept of RF-sampling can utilize the digital process scaling. The concept is also well suited for implementations with low supply voltages as is the case in modern CMOS processes. Furthermore, in sampling circuits the transistors are used as switches and accurate transistor models become less important. The yield can also be assumed to improve since the switches are being limited by defect yield (like digital circuits) rather than parametric yield. With process scaling it is expected that process variations increase and thus directly affect the parametric yield and reliability of analog circuits [13].

It is beneficial to use as high sampling frequency as possible, directly at RF

or at a high subsampling frequency, to minimize the risk of noise folding due to wideband noise at the output of the LNA. Using a tuned LNA also helps but prevents multiband operation. In an RF-sampling front-end the frequency translation is done by the sampling. Cascaded SC-filtering and decimation reduces the sample rate and prevents noise folding during the decimation. The image bands occurring in decimation are located at f_s/R between 0 and f_s , where R is the decimation rate. It is therefore important that the SC-filter provides high enough image rejection to suppress those frequencies. The discrete-time filter also implements the required antialiasing filtering for the ADC and can also be designed to provide some channel-select filtering. Due to the frequency translation and decimation the discrete-time signal at the output of the decimation filter is well suited for A/D-conversion. Since the RF signal is translated to baseband it is possible to digitize the signal at Nyquist rate for the baseband signal. Thus, high resolution ADC techniques can be used. A suitable candidate is therefore the $\Delta\Sigma$ -ADC, which can make use of the high oversampling ratio.

5.2.1 Issues Related to the Sampling Process

As mentioned before the sampling frequency should be as high as possible to minimize noise folding. A tuned LNA is required if subsampling is used, otherwise the impact of noise folding can be severe. When designing high-speed trackand-hold (T/H) circuits there are a number of switch nonidealities that have to be considered. Those nonidealities will affect the circuit performance and limit the maximum sampling frequency that can be used. The most important nonidealities limiting the performance of a simple T/H circuit are listed in Tab. 5.1.

Most of the nonidealities result in harmonic distortion, as seen in Tab. 5.1. The harmonic distortion due to the input-dependent sampling is analyzed in detail in [14]. The distortion depends on the amplitude of the input signal, input frequency, rise- or fall-time of the sampling clock, and the amplitude of the clock signal. In the case of sampling RF signals, the voltage-swing of the signal is relatively small at the input where the frequency is the highest. The linearity of a simple switch with proper sizing is usually enough. If this is not the case, fixed aperture time techniques like bootstrapping (Abo switch [15]) and bottom-plate sampling can be used if the linearity needs to be improved. Ultimately, the maximum sampling frequency is determined by the RC time constant. The capacitance is determined by the upper noise limit (kT/C noise) and can therefore not be reduced to improve the speed. The remaining option is to lower the switch resistance by sizing the MOS switch. Unfortunately, then the nonlinear parasitic capacitance associated with the transistor increases and adds to the sampling capacitance. The nonlinear parasitic capacitance also results in distortion of the signal. Obviously, there is a trade-off between on-resistance and parasitic capacitance when sizing the switches.

Nonideality	Cause	Consequences
Sampling pedestal	Clock feedthrough	Nonlinearity
	Channel charge injection	Harmonic distortion
Nonlinear channel	Input-dependent	Harmonic distortion
resistance	gate-source voltage	Input frequency limitation
Input-dependent	Finite clock transition time	Harmonic distortion
sampling instant	Input-dependent	Input frequency limitation
	gate-source voltage	
Finite acquisition	Finite bandwidth (RC)	Limited tracking accuracy
time	Finite slew rate	Limited sampling freq.
Aperture jitter	Clock jitter	Limited dynamic range
		Input frequency limitation
Charge leakage	Subthreshold leakage	Harmonic distortion
	Gate leakage	A lower bound on
		sampling frequency

Table 5.1: Nonidealities affecting the performance of high-speed T/H circuits.

Clock jitter increases the noise floor (Gaussian noise) in the sampled signalspectrum and if large enough it limits the signal-to-noise ratio (SNR). The maximum dynamic range achievable due to clock jitter can be expressed as [16]:

$$SNR_{max} = 10\log\left(\frac{1}{(2\pi\sigma f_{RF})^2}\right) \times \left(\frac{f_s}{f_{RF}}\right)^2 \times OSR$$
(5.1)

where σ is the sampling-clock jitter, OSR is the oversampling ratio, f_s is the sampling frequency, and f_{RF} is the carrier frequency. The effect of jitter (σ) is higher for lower sampling frequencies if the jitter is dominated by the synthesizer. Even though quite obvious, it is also worth mentioning that the jitter requirements on the local oscillator is comparable for sampler- and mixer-based downconversion [17]. Thus, the clock jitter is not a more severe problem for RF-sampling receivers (or fast ADCs for that matter) than for traditional mixer-based receivers.

5.2.2 Noise in Switched-Capacitor Circuits

Noise in a sampling circuit must be characterized in terms of the folding effect [18]. The noise at the output of a simple switched capacitor, i.e. a T/H circuit, can be expressed by the power spectral density S(f), which is shaped with respect to the sampling frequency f_s and hold time τ_h :

$$S(f) = (1 - \tau_h f_s) 4kTR + (\tau_h f_s)^2 sinc^2 (\tau_h f) \frac{2kT}{Cf_s}$$
(5.2)

The first term of S(f) can be referred to as the track noise and the other as the hold noise. Apparently, the track noise is proportional to the on-resistance of the switch and the duty factor $(1-\tau_h f_s)$, but its contribution to the total power spectral density (PSD) can be neglected when the *RC* time-constant is much smaller than the sampling period $(1/f_s)$ and the hold factor $\tau_h f_s$ does not approach zero. In practice, those conditions are often met and the hold noise tends to prevail over its track counterpart. Most of the power of the hold noise is located in the first lobe of the sinc²($\tau_h f$) function, being limited by a "zero" at $f = 1/\tau_h$. Thus, the noise spectral density is approximately equal to:

$$S(f) = (\tau_h f_s)^2 \frac{2kT}{Cf_s}$$
(5.3)

The equivalent noise bandwidth of S(f) can be estimated as $1/2\tau_h$, such that the PSD integrated over the whole frequency is:

$$I = (\tau_h f_s)^2 \frac{2kT}{Cf_s} \cdot \frac{1}{2\tau_h} = (\tau_h f_s) \frac{kT}{C}$$
(5.4)

Even though the theory of noise in switched-capacitor circuits is widely described in literature it becomes very intricate when dealing with more complicated circuits than a simple T/H circuit [19]. The theory of noise in SC circuits is nicely described in [19, 20]. For RF-sampling front-ends using SC decimation filters the noise properties are essential for the performance. The noise of such circuits are analyzed in Chapters 13-15 and in [5, 21].

5.2.3 Basic Circuit Techniques to Implement High-Frequency SC Decimation Filters

Discrete-time analog filtering is commonly used in the IF sections of analog frontends to implement the channel filter. Those filter implementations usually rely on operational amplifiers or transconductors in the case of G_m -C filters. Due to the high frequency, when sampling directly at RF, such solutions are not practical and operational amplifiers should be avoided at least until the signal has been decimated. SC circuits using operational amplifiers implemented in today's CMOS processes can perform well up to a few hundred MHz [22]. Therefore, other techniques must be used for higher sampling rates. In this section basic circuit techniques for SC-filters operating in the GHz range are addressed. Passive SC circuits are used for implementing FIR- and IIR-like filters and in the case of cascading filters simple buffers or low-gain amplifiers without feedback are used.

A passive SC circuit implementing an N-tap FIR filter with decimation by N is shown in Fig. 5.1. N succeeding samples are stored on C_1 - C_N . Averaging the



Figure 5.1: SC implementation of a passive N-tap FIR filter with decimation by N.

charge stored on all N capacitors, by simply shorting them all together, implements an FIR filter with decimation. From Fig. 5.1 it can also be noted that the output signal can only be updated every N-th clock cycle, which is equivalent to the decimation rate of the filter. If a filter with a length larger than the decimation rate is needed, then time-interleaved filter banks are required. If no decimation is desired, then as many as N time-interleaved filters are required. Fig. 5.2 visualizes the difference between the filter response of an 8-tap moving-average FIR filter and the one from a 32-tap moving-average FIR filter consisting of four timeinterleaved sections to maintain the same decimation rate as the first one. Furthermore, it is here assumed that the parasitic capacitance $C_p << C_1 + \ldots + C_N$ so that the memory effect of C_p can be neglected. The resulting filter function can be expressed as:

$$H(z) = \sum_{n=0}^{N-1} a_n z^{-n}$$
(5.5)

where a_n is the filter coefficients and N is the number of taps. The coefficient a_n is given as:

$$a_n = \frac{C_{n-1}}{\sum_{i=1}^N C_i}$$
(5.6)

Thus, by using weighted capacitor sizes all different types of FIR filter responses can be implemented. In the case where differential circuits are used both polarities of the input voltage is available. Negative coefficients can be achieved by crosscoupling the differential input. This is equivalent with a negative coefficient since the capacitor will be charged with a voltage of the opposite polarity. From a prac-



Figure 5.2: (a) Filter response of an 8-tap moving-average FIR filter. (b) Filter response of 4 time-interleaved 32-tap moving-average FIR filters

tical point of view one of the most important parameters, describing the quality of SC circuits, is the ratio of the biggest capacitor to the smallest one. Circuits with large such ratios can be expected to suffer from coefficient mismatch. In [5], two time-interleaved filters with a similar topology as the one in Fig. 5.1 were used to implement a 22-tap long FIR filter with a capacitor ratio between the largest and smallest of 17. Such FIR filters with high ratios are not very practical and a better way is to use cascaded FIR filters instead as described in **Paper 7-9**. One of the main drawbacks with the circuit shown in Fig. 5.1 is the need for multi-phase clocking.

Simple IIR filters can also be implemented. Consider the SC circuit shown in Fig. 5.3. The input voltage is sampled on the capacitor C_{sa} . During the opposite clock phase the two capacitors C_{sa} and C_{sum} are shorted and the new value stored on C_{sa} is averaged with the charge from previous samples. Thus, C_{sum} implements a "memory" function, i.e. an IIR filter. The transfer function for the circuit



Figure 5.3: SC circuit used for implementing an IIR filter.



Figure 5.4: (a) Filter implementation of an N-tap moving-average FIR filter (decimation by N) combined with a first-order IIR filter at the output. (b) Filter response of the circuit shown to the left (N = 8).

in Fig. 5.3 is:

$$H(z) = \frac{C_{sa}}{C_{sa} + C_{sum}} \cdot \frac{1}{1 - \frac{C_{sum}}{C_{sa} + C_{sum}} z^{-1}}$$
(5.7)

The FIR filter shown in Fig. 5.1 can be combined with an IIR filter to improve the filter selectivity. Such a circuit is shown in Fig. 5.4a where a summing capacitor, without reset, is added at the output and thus implementing the IIR filter functionality. The filter response of the filter is shown in Fig. 5.4b and corresponds to a transfer function equal to:

$$H(z) = \frac{(1-p_1)(1-z^{-N})}{(1-z^{-1})(1-p_1z^{-N})}$$
(5.8)

where N = 8 and $p_1 = 0.8$. As seen in Fig. 5.4b the filter bandwidth is much more narrow compared to the one without IIR filtering. To get a filter with even smaller bandwidth, to be used for channel filtering for example, another IIR filter can be implemented at the input of the FIR filter. Such an IIR-FIR-IIR filter is shown in Fig. 5.5a. The corresponding filter response is shown in Fig. 5.5b and corresponds to a filter transfer equal to:

$$H(z) = \frac{(1-p_1)(1-p_2)(1-z^{-N})}{(1-z^{-1})(1-p_1z^{-N})(1-p_2z^{-N})}$$
(5.9)

where N = 8 and $p_1 = p_2 = 0.8$. By cascading filter section of the type described here many different filters can be realized in order to achieve the wanted filter functionality.



Figure 5.5: (a) Filter implementation of an N-tap moving-average FIR filter (decimation by N) combined with a first-order IIR filter at both the input and at the output. (b) Filter response of the circuit shown to the left (N = 8).



Figure 5.6: Integrator implementing an FIR filter with all-one coefficients.

Another technique to implement FIR filters was presented in [1, 21]. This technique is based on a transconductor and a switched capacitor used as an integrator, see Fig. 5.6. In this case a current is integrated on one large single capacitor. The length of the FIR filter and also the decimation rate is determined by the reset rate of the capacitor. Thus, in opposite to the previously described circuits this one need a reset to be functional (when a voltage is sampled on a capacitor as in Fig. 5.1 it is automatically reset). This far only FIR filters with all-one coefficients, also known as moving average, has been demonstrated. However, different coefficients could be realized by adjusting the transconductor gain for different samples. The concept is shown in Fig. 5.7. By scaling g_m individually for every sample, different coefficients can be realized. A way to implement such a circuit is to use a similar concept as current-steering DACs. By dividing the total transconductance into a number of unit sized transconductors each coefficient is determined by how many of these that are switched in to charge the capacitor during that clock cycle. The method requires individual clock-gating for every



Figure 5.7: Integrator implementing an FIR filter with adjustable coefficients

transconductor element according to a predefined pattern determining the filter coefficients. New filters functions could thereby be implemented by simply reconfiguring this predefined pattern, which determines how many transconductor elements should be used during a certain clock cycle.

IIR filtering can also be made using the circuit in Fig. 5.6 as described in [1, 21]. By splitting the integrating capacitor into two on a periodic basis a first order IIR filter is created, see Fig. 5.8. A brief explanation of how the circuit works is as follows. The two rotating capacitors, C_R , are switched every N-th cycle for read out and reset. These two are implementing the FIR filter function. The history capacitor, C_H , implements the IIR filtering. This capacitor is never reset and only a portion of its charge is removed every N-th cycle when C_R rotates. This discrete-time IIR filter operates at f_s/N sampling rate and creates a single pole frequency response. The resulting pole, in the continuous-time domain for $f_c \ll f_s/N$, is:

$$f_c = \frac{f_s}{2\pi \cdot N} \cdot \frac{C_R}{C_H + C_R} \tag{5.10}$$

A more detailed explanation is given in [1, 21]. This type of filtering has successfully been used to implement both Bluetooth and GSM single-chip radios [1, 23, 24].



Figure 5.8: Simplified circuit showing the IIR operation implemented in [1, 21].

5.3 An RF-Sampling Receiver Front-End for WLAN in 0.13 μm CMOS

To demonstrate the feasibility of RF sampling a receiver front-end primarily intended for WLAN has been implemented in a 0.13 μ m CMOS process (**Paper 10**). The zero-IF front-end consists of a wideband LNA, described in **Paper 5**, and an I/Q SC decimation filter based on the topology discussed in **Papers 7-9**.

5.3.1 Circuit Description

At the output of the wideband LNA the RF signal is directly sampled. A zero-IF is achieved using a sampling frequency equal to:

$$f_s = f_c \cdot 2^{-n} \tag{5.11}$$

Where f_c is the carrier frequency and $n = 0, 1, 2, \ldots$ For the 2.4 GHz band the sampling frequency is chosen to be equal to the carrier frequency. The 5-6 GHz frequency band is subsampled by $2\times$, resulting in sampling frequencies between 2.5-3 GHz. The reason for using subsampling for the higher frequency band was that the multi-phase clocking needed in the decimation filter could not be realized for such frequencies in 0.13 μ m CMOS. The sampling frequency was thus limited by the digital circuitry rather than the analog. Otherwise it is beneficial to use as high sampling frequency as possible to minimize the risk of noise folding due to wideband noise at the output of the LNA. Using a tuned LNA helps but prevents multiband operation unless a tunable narrowband LNA can be implemented.

Each of the two (I and Q) SC decimation filters used consists of 4 cascaded filter sections. The front-end was implemented is such a way that the output can be taken both after 3 and 4 stages, respectively. Thus a decimation rate of 8 or 16 can be chosen. For a 2.4-2.5 GHz carrier frequency and decimation by 16, the output rate is 150-156.25 MHz, which is acceptable in terms of A/D conversion rate. The 5-6 GHz band is subsampled with half the carrier frequency giving a corresponding sample rate of 156.25-187.5 MHz at the input of the ADC. When scaling the decimation rate, the relative bandwidth of the filter will scale accordingly. Using a decimation of 8 instead of 16 will give twice the bandwidth at the price of higher conversion rate for the following ADC. The bandwidth also scale with the sampling frequency. For example, at half the sampling frequency the image rejection will remain the same if only half the bandwidth is used. Larger bandwidth can of course be used as well, but at the price of lower image rejection. Basically, the filter can be used for any carrier frequency up to 6 GHz with a maximum sampling frequency of 3 GHz. A special technique was also used to reduce the 1/f-noise and DC-offset related to the amplifiers used within the SC filters. By



Figure 5.9: Frequency response of the decimation filter, $f_s = 2.5$ GHz.

switching the amplifiers both 1/f-noise and DC-offset will have very little effect on the circuit performance.

In this particular RF front-end implementation there will be no or at least very limited blocker suppression as seen from the simulated filter transfer function (Fig. 5.9). Therefore, the total gain in the receiver is set by the maximum allowed blocker and the input range of the ADC. Maximum signal strength at the antenna is -30 dBm. Since the LNA has a differential input, this signal is transformed to a differential signal using a balun. This means that each input has a maximum signal swing of 28.3 mV_{pp}, which translates to a total differential input signal swing of 56.6 mV_{pp} at the input of the LNA. At the ADC input a full-scale differential signal swing of 1.41 V_{pp} is desired. The required voltage gain is then 25 corresponding to 28 dB power gain. This gain will be provided by the LNA and the SC decimation filter. It is further assumed that a noise factor of 10 for the whole front-end would be sufficient.

5.3.2 Required ADC Performance

Since the used decimation filter provides no or very limited blocker suppression it can be assumed that full blocker power can reach the ADC input. The dynamic range requirement is therefore given by the maximum blocker power, P_B , and a sensitivity determined by the noise figure of the front-end. A noise factor, F, leads to a needed signal-to-noise equal to:

$$SNR = \frac{P_B}{FkTf_B} \tag{5.12}$$

where k is Boltzmann's constant, T is the temperature in Kelvin, and f_B is the baseband bandwidth corresponding to a channel bandwidth of $2f_B$. An ADC performing Nyquist sampling of the baseband needs a sampling frequency $f_s = 2f_B$. Such an ADC have a signal-to-noise ratio equal to:

$$SNR = \frac{3}{2}2^{2n}$$
 (5.13)

where n is the ADC resolution in terms of number of bits. Using Eq. 5.12-5.13 and that $f_B = f_s/2$ gives the required performance of an ADC that can handle a week signal as well as the blocker [25]:

$$f_s 2^{2n} = \frac{4}{3} \cdot \frac{P_B}{FkT} \tag{5.14}$$

When using a sampling frequency higher than the required Nyquist frequency, the oversampling ratio $OSR = f_s/2f_B$ can be used to reduce the required ADC resolution. An ADC with plain oversampling will have a signal-to-noise ratio of [26]:

$$SNR = \frac{3}{2} \cdot OSR \cdot 2^{2n} \tag{5.15}$$

Another option is to use oversampled $\Sigma\Delta$ -ADCs with noise shaping. A first-order $\Sigma\Delta$ -ADC have a signal-to-noise ratio of [26]:

$$SNR = \frac{9}{2\pi^2} \cdot OSR^3 \cdot 2^{2n} \tag{5.16}$$

while a second-order $\Sigma\Delta$ -ADC have a signal-to-noise ratio of [26]:

$$SNR = \frac{15}{2\pi^4} \cdot OSR^5 \cdot 2^{2n}$$
(5.17)

For the RF-sampling front-end for WLAN we assume $P_B = -30$ dBm and a noise factor of 10 for the whole front-end. The channel bandwidth is 20 MHz giving $f_B = 10$ MHz. Further, assuming a sampling frequency of 2.4 GHz and decimation by 16 before the ADC leads to an OSR of 7.5. Using this and Eq. 5.15 gives a resolution of 9.4 bits. Thus a 10 bit ADC with a conversion rate of 150 MS/s would be enough. In reality 1-2 bits more would be used to get some design margin. Never the less, such ADCs are fully feasible and exists [27, 28, 29, 30]. However, a more efficient way is to make use of oversampled $\Sigma\Delta$ -ADCs.


Figure 5.10: Frequency response corresponding to a 1.6 GHz channel with: (a) A decimation rate of 16. (b) A decimation rate of 8.

In such a case it would also make sense to increase the sampling frequency by performing less decimation. Using a decimation of 8, i.e. an OSR of 15 together with Eq. 5.16 and Eq. 5.17, a 6 bit first-order $\Sigma\Delta$ -ADC or a 4 bit second-order $\Sigma\Delta$ -ADC with a sampling frequency of 300 MS/s would be enough to give the needed resolution.

5.3.3 Preliminary Measurement Results

To date some initial measurements have been made. The frequency response and the linearity of the front-end has been measured. The frequency response for a channel at 1.6 GHz carrier frequency sampled with a sampling frequency of 1.6 GHz is shown in Fig. 5.10 for a decimation rate of 16 and 8, respectively. In Fig. 5.11 a similar frequency response is shown but for a channel at 2.4 GHz carrier frequency sampled with a sampling frequency of 2.4 GHz.

Also the linearity of the front-end has been measured. Fig. 5.12 shows both the 1-dB compression measurement and the third-order intercept point measurement. The 1-dB compression point was measured to be -25.5 dBm and an IIP3 of -15.45 dBm was recorded. This should be enough to fulfill the requirements for WLAN [31]. Unfortunately the noise figure has not yet been measured. Even though not yet fully evaluated, the initial measurements indicate full functionality of the entire front-end. The digital part, i.e. the multiphase clocking has been tested up to 3 GHz.

The whole front-end (LNA plus two decimation filters for I and Q) consumes about 176 mW (126 mA from a 1.4 V supply). The power consumption can be divided into three parts originating from LNA, analog circuitry, and digital



Figure 5.11: Frequency response corresponding to a 2.4 GHz channel with: (a) A decimation rate of 16. (b) A decimation rate of 8.



Figure 5.12: (a) Measured 1-dB compression point (1dB-CP). (b) Measured third-order intercept point (IIP3).

circuitry. The wideband LNA dissipates 36.4 mW (22.4 mW from the LNA core and 14 mW from the output buffers). The power consumption of the rest of the analog circuitry is 112 mW and includes the 50 Ω output buffers for off-chip measurements. The digital power consumption is approximately 28 mW. This figure is an approximation since the digital power cannot be separated from the digital IO.

5.4 References

- K. Muhammad, R.B. Bogdan, and D. Leipold, "Digital RF Processing: Toward Low-Cost Reconfigurable Radios," *IEEE Communications Magazine*, pp. 105–113, August 2005.
- [2] J.-H.C. Zhan and S.S. Taylor, "A 5GHz Resistive-Feedback CMOS LNA for Low-Cost Multi-Standard Applications," in *Proc. of the IEEE International Solid-State Circuits Conf.*, pp. 200–201, 2006.
- [3] K. Maggio, "Practical Aspects of CMOS Integrated Radio," *IMS Workshop Slides*, 2004.
- [4] M. Ismail and D. Rodríguez de Llera González, eds., Radio Design in Nanometer Technologies. Springer, 2006.
- [5] D. Jakonis, K. Folkesson, J. Dabrowski, P. Eriksson, and C. Svensson, "A 2.4-GHz RF Sampling Receiver Front-End in 0.18μm CMOS," *IEEE Journal of Solid State Circuits*, vol. 40, pp. 1265–1277, June 2005.
- [6] H. Lakdawala, J. Zhan, A. Ravi, S. Andersson, B. Carlton, R.B. Nicholls, N. Yaghini, R.E. Bishop, S.S. Taylor, and K. Soumyanath, "Multi-band (1-6GHz) Sampled, Sliding-IF Receiver With Discrete-Time-Filtering in 90nm Digital CMOS Process," in *Proc. of the VLSI Symposium Conf.*, 2006.
- [7] H. Darabi and J. Chiu, "A Noise Cancellation Technique in Active RF-CMOS Mixers," *IEEE Journal of Solid State Circuits*, vol. 40, pp. 2628– 2632, December 2005.
- [8] R.S. Pullela, T. Sowlati, and D. Rozenblit, "Low Flicker-Noise Quadrature Mixer Topology," in *Proc. of the ISSCC 2006 Conf.*, pp. 466–467, 2006.
- [9] K. Soumyanath, "Low Voltage Techniques for sub 100 nm CMOS, RF Tranceivers," in *Proc. of the IEEE Radio Frequency Integrated Circuits Symposium*, pp. 501–504, 2005.
- [10] M. Elmala and B. Carlton and R. Bishop and K. Soumyanath, "A Highly Linear Filter and VGA Chain with Novel DC-Offset Correction in 90nm Digital CMOS Process," in *Proc. of the VLSI Symposium Conf.*, pp. 302– 303, 2005.
- [11] A.A. Abidi, "RF CMOS Comes of Age," *IEEE Journal of Solid State Circuits*, vol. 39, no. 4, pp. 549–561, April 2004.
- [12] ITRS, http://public.itrs.net, 2006.

- [13] X. Wang, J. Pekarik, X. Yuan, R. Anna, and S. Parker, "Meeting the Design Challenges in Modern RFCMOS Technology," in *Proc. of the 7th International Conference on Solid-State and Integrated Circuits Technology*, pp. 1291–1294, 2004.
- [14] D. Yu, S. Subhajit, and H. Leung, "Distortion Analysis of MOS Track-and-Hold Sampling Mixers Using Time-Varying Volterra Series," *IEEE Tran. on Circuits and Systems*, vol. 46, no. 2, pp. 101–113, February 1999.
- [15] A.M. Abo and P.R. Gray, "A 1.5-V, 10-bit, 14.3-MS/s CMOS Pipeline Analog-to-Digital Converter," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 5, pp. 599–606, May 1999.
- [16] S.Lindfors, A. Pärssinen, and K.A.I. Halonen, "A 3-V 230-MHz CMOS Decimation Subsampler," *IEEE Tran. on Circuits and Systems*, vol. 50, no. 3, pp. 105–117, March 2003.
- [17] V.J. Arkesteijn, E.A.M. Klumperink, and B. Nauta, "Jitter Requirements of the Sampling Clock in Software Radio Receivers," *IEEE Tran. on Circuits and Systems*, vol. 53, no. 2, pp. 90–94, February 2006.
- [18] R. Gregorian and G. Temes, Analog MOS Integrated Circuits. Wiley, 1986.
- [19] R. Schreier, J. Silva, J. Steensgaard, and G.C. Temes, "Design-Oriented Estimation of Thermal Noise in Switched-Capacitor Circuits," *IEEE Tran. on Circuits and Systems*, vol. 52, no. 11, pp. 2358–2368, November 2005.
- [20] K. Kundert, "Simulating Switched-Capacitor Filters with SpectreRF," http://www.designers-guide.com/documents.html, 2006.
- [21] K. Muhammad and R.B. Staszewski, "Direct RF Sampling Mixer with Recursive Filtering in Charge Domain," in *Proc. of the ISCAS 2004 Conf.*, pp. 557–580, 2004.
- [22] J. Dabrowski, S. Andersson, J. Konopacki, and C. Svensson, "SC Filter Design for RF Applications," in *Proc. of the ICSES 2006 Conf.; Accepted for publication*, 2006.
- [23] K. Muhammad, D. Leipold, B. Staszewski, Y.-C. Ho, C.M. Hung, K. Maggio, C. Fernando, T. Jung, J. Wallberg, J.-S. Koh, S. John, I. Deng, O. Moreira, R. Staszewski, R. Katz, and O. Friedman, "A Discrete-Time Bluetooth Receiver in a 0.13μm Digital CMOS Process," in *Proc. of the ISSCC 2004 Conf.*, pp. 268–269, 2004.

- [24] K. Muhammad, Y.-C. Ho, T. Mayhugh, C.M. Hung, T. Jung, I. Elahi, C. Lin, I. Deng, C. Fernando, J. Wallberg, S. Vemulapalli, S. Larson, T. Murphy, D. Leipold, P. Cruise, J. Jaehnig, M.C. Lee, R.B. Staszewski, R. Staszewski, and K. Maggio, "Discrete Time Quad-band GSM/GPRS Receiver in a 90nm Digital CMOS Process," in *Proc. of the CICC 2005 Conf.*, pp. 809–812, 2005.
- [25] S. Andersson and C. Svensson, Software Defined Radio Visions, Challenges and Solutions, in Radio Design in Nanometer Technologies. M. Ismail and D. Rodríguez de Llera González, Eds., Springer, 2006.
- [26] S.R. Norsworthy, R. Schreier, and G.C. Temes, eds., *Delta-Sigma Data Con*verters: Theory, Design, and Simulation. Wiley-IEEE Press, 1996.
- [27] Linear Technology, http://www.linear.com, 2006.
- [28] Maxim, http://www.maxim-ic.com, 2006.
- [29] P. Bogner, "A 28mW 10b 80MS/s Pipelined ADC in 0.13 μm CMOS," in Proc. of the ISCAS 2004 Conf., pp. 17–20, 2004.
- [30] B. Hernes, A. Briskemyr, T.N. Andersen, F. Telstø, T.E. Bonnerud, and Ø. Moldsvor, "A 1.2V 220MS/s Pipeline ADC Implemented in 0.13 μm Digital CMOS," in *Proc. of the ISSCC 2004 Conf.*, pp. 256–257, 2004.
- [31] M. Brandolini, P. Rossi, D. Manstretta, and F. Svelto, "Toward Multistandard Mobile Terminals — Fully Integrated Receivers Requirements and Architectures," *IEEE Tran. on Microwave Theory and Techniques*, vol. 53, no. 3, pp. 1026–1038, March 2005.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

Modern radio transceivers have to support several different standards. Due to cost reasons a high level of integration of radio functions has become a necessity. Since the baseband requires a lot of signal processing capability and many applications require an extensive amount of on-chip memory, the only option is to use a deep submicron CMOS technology. Designing RF and analog circuits that are compatible with digital CMOS technology is not an easy task. Adding the need for multistandard operation and this really becomes a major challenge. RF-sampling techniques have recently become popular in order to implement more flexible radio front-ends. Today's CMOS technologies demonstrate very high speeds, making the RF-sampling technique appealing in a context of multistandard operation at GHz frequencies. The goal is to move in the direction of software-defined radio to get more flexible radio receivers. In this thesis a few key building blocks for flexible multistandard radio front-ends have been suggested.

Two different LNA topologies suitable for multiband multistandard wireless receivers have been demonstrated. Extremely wideband LNAs using 0.18 μ m and 0.13 μ m CMOS technology. Both implementations are showing the great possibility of designing wideband LNAs in deep submicron CMOS technologies. A new form of bandpass LNA, based on the principle of active recursive filters has also been developed. Designed circuits in both CMOS and BiCMOS show that this approach is feasible, at least for applications with not too high performance requirements. The noise and linearity need to be improved. However, these designs show excellent frequency tunability and the possibility to tune both gain and Q value in a broad range. The goal with all these LNAs has been to cover a large bandwidth, either by using wideband LNAs or by using tunable narrow-band LNAs, while still having reasonable noise figures, linearity, and impedance

matching. Except for one of the wideband LNAs, inductors have been avoided to decrease the cost in terms of area and make the RF circuits compatible with digital CMOS technology. Such LNAs are essential for the implementation of low-cost multistandard radio.

Switched-capacitor circuits are used to implement the needed filtering and decimation. These filters initially have to operate at very high frequencies similar to the carrier frequency. The sampled signal is filtered and decimated to a frequency suitable for analog-to-digital conversion. During decimation new image bands are introduced. These have to be removed by the switched-capacitor filter. In this thesis we have presented switched-capacitor filters for RF sampling and decimation with wideband image rejection.

To demonstrate the concept and feasibility of receivers based on RF sampling, an RF-sampling front-end primarily intended for WLAN has been implemented in a 0.13 μ m CMOS process. The front-end consists of a wideband LNA and switched-capacitor filters with wideband image rejection. Initial measurements indicate full functionality of the entire front-end. The frequency response as well as the linearity of the whole front-end has been measured. A 1-dB compression point of -25.5 dBm and an IIP3 of -15.5 dBm have been measured. Unfortunately the noise figure has not yet been measured. The switched-capacitor decimation filters can be used for any carrier frequency up to 6 GHz with a maximum sampling frequency of 3 GHz. Depending on the wanted bandwidth and sample rate at the output, the decimation rate can be chosen as 8 or 16. This frequency scalability has the potential of being utilized for multistandard applications. The only limitation is the ADCs resolution and conversion rate. Even though not yet fully evaluated, the presented work shows the possibilities of designing flexible RF front-ends for wireless receivers using RF-sampling and decimation techniques.

6.2 Future Work

The presented LNAs built on the principle of active recursive filters show excellent frequency tunability. However, as a consequence the linearity and noise performance have to some extent been sacrificed. This, since the tuning made with the bias current also affects both noise and linearity. Instead of just tuning the frequency by adjusting the bias current, a combination of fine tuning using the current and a coarse tuning by changing the capacitive load on the amplifiers can be used. The capacitive load on the amplifiers can be implemented using an array of switched capacitors. Both noise and linearity could be improved in this way and made more consistent over the frequency tuning range.

To avoid unnecessary noise folding from the wideband LNA in an RF-sampling front-end, it would be advantageous if the bandwidth of the wideband LNAs could

be adjusted along with the sampling frequency. This can be accomplished by changing the load or the output impedance of the output stage in the LNA. This works as long as the sampling frequency and carrier frequency are roughly the same, i.e. subsampling is not used.

The concept of RF-sampling is still a quite new and unexplored field. Many interesting options still remains to be investigated regarding how to implement the decimation filters more efficient and how to reconfigure the filter function for different applications. It would also be interesting to look into if it is possible to use quite simple parallel filters operating at different sampling frequencies to cancel out certain frequency bands and interferers. Finally, finding a way to integrate the decimation filter, or at least part of it, with the $\Sigma\Delta$ -ADC would further improve the RF-sampling receiver architecture.