Downloaded from orbit.dtu.dk on: Dec 18, 2017

Technical University of Denmark

Low Power RF Filtering for CMOS Transceivers

Christensen, Kåre Tais; Olesen, Ole; Bruun, Erik

Publication date:
2002

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Christensen, K. T., Olesen, O., & Bruun, E. (2002). Low Power RF Filtering for CMOS Transceivers.

DTU Library
Technical Information Center of Denmark

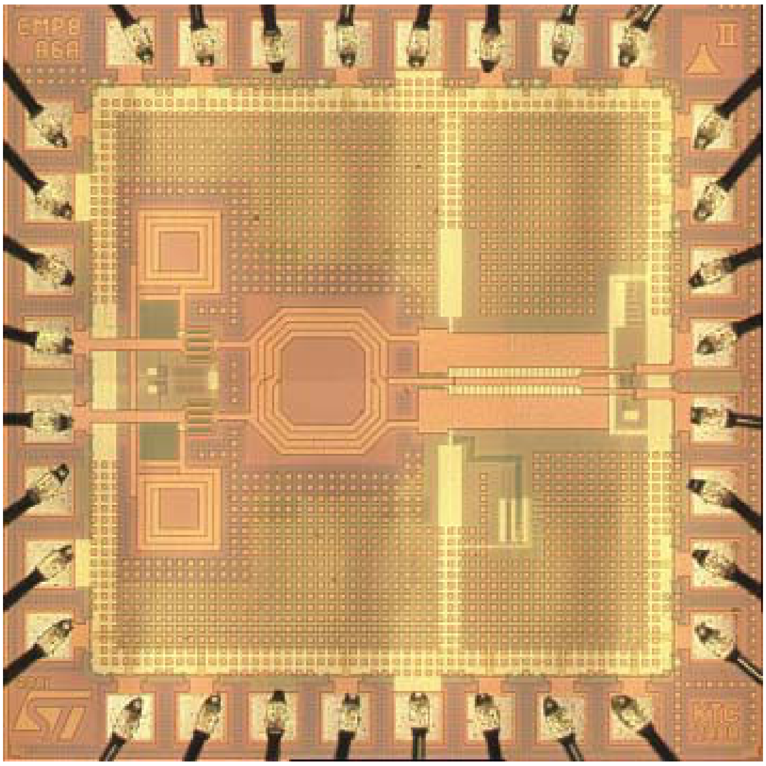
General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LOW POWER RF FILTERING FOR CMOS TRANSCEIVERS

Kåre Tais Christensen



Dissertation submitted to the faculty of the Technical University of Denmark
in partial fulfillment of the requirements for the degree of:

Doctor of Philosophy in Electrical Engineering

September 2001

Low Power RF Filtering for CMOS Transceivers

List of Symbols and Abbreviations iii

Preface v

1. Introduction to RF Filtering in Wireless Terminals 1

1.1 Trends in Wireless Terminals 1
1.2 Filtering Requirements in Wireless Terminals 5
1.3 Common Filtering Solutions 10
1.4 Future Challenges in Integrated RF Filtering 14
1.5 Summary 16
1.6 References 16

2. Input Power Constrained LNA Design 17

2.1 Classic LNA Topologies 17
2.2 Input Match of Inductively Degenerated LNA 19
2.3 ClassA - Operation in the presence of Strong Blocking Signals 20
2.4 Addition of External Gate-Source Capacitance 23
2.5 Mobility Degradation in Short Channel MOS Transistors 25
2.6 Noise Modelling of the LNA with External Gate-Source Capacitance 26
2.7 LNA Design Example 31
2.8 Other Considerations 34
2.9 Summary 37
2.10 References 39

3. Resonator Design 41

3.1 Properties of LC Resonators 41
3.2 Design of Metal-Metal Capacitors in Standard CMOS 47
3.3 Inductor Design in CMOS Processes with Highly Doped Substrates 55
3.4 Summary 72
3.5 References 73

4. Frequency Tuning 75

4.1 Introduction to Frequency Tuning 75
4.2 Analysis of Switched Capacitance Tuning of LC Resonators 81
4.3 Design Techniques for CMOS Switch Optimization 89
4.4 The Differential Waffle Switch 99
4.5 A Note on PMOS Switches and Multibit Tuning 104

4.6 Preliminary Measurements 105
4.7 Summary 107
4.8 References 108

5. Quality Factor Tuning 109

5.1 Feedback Techniques for Loss Compensation 110
5.2 Large Signal Linearity 114
5.3 Noise 116
5.4 Tuning the Negative Resistance 118
5.5 Systems for Control of the Q-tuning 120
5.6 Summary 122
5.7 References 122

6. Integrated Front-End Filter Design 123

6.1 Filter Topology 123
6.2 Component Values and Biasing 128
6.3 Filter Noise Analysis 132
6.4 Other Considerations 137
6.5 Example: A Fully Integrated Front-End Filter for GSM1800/1900 & WCDMA 138
6.6 Summary 144
6.7 References 144

7. Low Power LC Quadrature Generation for Image Rejection 145

7.1 Introduction 145
7.2 Allpass Filters 147
7.3 Quadrature Phases 148
7.4 Norton Transformation 149
7.5 Common Gate Amplifier Load 150
7.6 Parasitics 151
7.7 Example 152
7.8 Summary 153
7.9 References 154

8. Conclusion 155

8.1 Summary 155
8.2 Future Work 158

Appendix A: Publications 159

Appendix B: Contributions 161

List of Symbols and Abbreviations

A_i	virtual parallel plate capacitance area between two fingers in metal layer i
A_{vi}	virtual parallel plate cap. area between opposite vias on two fingers in metal layer i
A_{ij}	virtual parallel plate capacitance area between conducting metal layer i and metal layer j (i = 0 represents substrate or poly)
A/D	analog to digital converter
BAW	bulk acoustic wave filter
BB	baseband
BPF	bandpass filter
BW	band width
C_{fix}	fixed capacitance in a switched capacitor array
C_{gs}	gate-source capacitance
C_{gsm}	intrinsic gate-source capacitance
C_{gsx}	external/extra/extrinsic gate-source capacitance
C_i	lateral flux capacitance contribution from metal layer i on one side of one VMC
C_j	junction capacitance
C_{mesh}	capacitance of one vertical mesh (see Section 3.2 for more symbol explanations)
C_{ox}	total oxide intrinsic capacitance
CMOS	complementary metal oxide semiconductor
DECT	Digital European Cordless Telephone
DSP	digital signal processor
EPI	epitaxy process (used on an "EPI substrate")
f	frequency
f_0	center frequency, resonance frequency
f_b	frequency of blocking signal
f_T	transit frequency (unity current gain frequency)
FBAR	thin-film bulk wave acoustic resonator
G_m	transconductance
GPS	global positioning system
GSM	global system for mobile communication
I	in phase
I_D	total LNA drain current
i_{eddy}	eddy current - parasitic induced currents
IC	integrated circuit
ICP	1-dB input compression point
IF	intermediate frequency
IIP3	input referred 3rd order intermodulation point
ISM	industrial, scientific, medical
k	Boltzmann’s constant, attenuation konstant
k_{CD}	capacitive division factor
k_{QE}	Q-enhancement factor
L_{drawn}	drawn transistor channel length
L_{eff}	effective transistor channel length
L_g	inductor connected to the gate of an LNA input transistor

L_s	inductor connected to the source of an LNA input transistor
LC	passive network made out of inductors and capacitors
LNA	low noise amplifier
LO	local oscillator
LPF	lowpass filter
MM9	MOS model 9
MOS	metal oxide semiconductor
NF	noise figure
P_{in}	input power of the strongest out-of-band blocking signal
PA	power amplifier
PGS	patterned ground shield
PLL	phase locked loop
PWB	printed wire board (new name for printed circuit board)
Q	in quadrature, quality factor
Q_j	drain/source junction quality factor
Q_{tank}	uncompensated tank quality factor
QPSK	quadrature phase shift keying
R_C	series resistance in the capacitive branch of an LC polyphase network
R_L	load resistance
R_{neg}	negative resistance
R_{on}	MOS on-resistance
R_s	source impedance
RF	radio frequency
RX	receiver
S	metal spacing
S_{out}	output referred power spectral density
SAW	surface acoustic wave
SMD	surface mount device
SNR	signal-to-noise ratio
TDMA	time division muliple access
TX	transmitter
V_{peak}	maximum single-sided voltage amplitude
V_T	threshold voltage
VCO	voltage controlled oscillator
VGA	variable gain amplifier
VMC	vertical mesh capacitor
W	transistor width, wire width
W_{opt}	optimal switch transistor width
Z_{tank}	tank impedance
α	MOS transconductance degradation factor, relative frequency tuning range
δ	MOS gate noise constant
Δf	frequency bandwidth
γ	MOS drain noise constant (~ 2/3)
μ_n	effective surface mobility of an NMOS device
θ	MOS mobility degradation fitting parameter
ρ	fill factor (a measure of an inductors hollowness)
ω	frequency in radians
ω_0	center frequency, resonance frequency in radians
ω_T	transit frequency in radians
ξ	normalized resonator impedance

Preface

In 1997 a cooperation between the Technical University Denmark and Nokia Mobile Phones was initiated. Several M.Sc and Ph.D. projects were started in the area of RF integrated circuit design. The main research areas at that time were power amplifier design, linearization techniques and data converters. Then in the beginning of 1998 this project was launched. The topic was open so a study was initiated with the main purpose of identifying a research topic with significant business potential for Nokia. A Ph.D. project takes three to four years and implementation in a commercial chip set takes another two years. Therefore the time horizon was chosen to be year 2003. In 1998 most of the RF modem functionality was expected to be integrated on a single or a few RF integrated circuits by that time so to have business potential the research topic would have to be something with significant potential cost savings and which was not likely to be integrated during that five year time span.

RF filtering is a very fundamental need in all wireless communication, it accounts for several Euros of the total radio modem cost and it was not likely to be integrated in any commercial form before 2003 so it was decided to choose this area for the project.

It is obvious that RF filtering posed a significant challenge as most if not all of the published attempts had far from useful dynamic ranges and often much too high power consumption for deployment in battery powered wireless terminals like GSM phones.

The lack of dynamic range lead to research in the design of high dynamic range active circuits and on-chip passive components like capacitors and inductors in standard CMOS processes as well as filter design methodologies. It is the results of that work which is presented in this thesis.

The first chapter presents an introduction to RF filtering in wireless terminals. The second chapter presents a new LNA design methodology which enables sufficient dynamic range with an order of magnitude lower power consumption than with the common CMOS LNA topology. Chapter 3 describes in detail how state-of-the-art CMOS resonators can be designed. Chapter 4 describes how switched capacitor frequency tuning can be used to implement wide tuning ranges with low loss and noise. Especially the design and optimization of the MOS switches is given a rigorous treatment. Chapter 5 describes high dynamic range linear loss compensation circuits which are needed to achieve sufficient filter selectivity. Chapter 6 presents a complete design methodology for the design of fully integrated front-end filters in CMOS. The methodology provides closed form expressions for all component values and bias levels and it is believed that it can reach GSM performance with current bulk CMOS processes.

Chapter 7 presents a new LC quadrature generating circuit which can be highly useful for image reject mixers or for low power applications with narrowband operation. Finally Chapter 8 concludes the thesis.

This work has been made possible through the help of many individuals. Allan Jørgensen from the Centre for Integrated Electronics (CIE) is greatly acknowledged for providing good Cadence CAD support to a large number of students. Ole Olesen director of CIE is acknowledged for financial support for chip fabrication and for providing important contacts and encouragement. Many thanks go to Jacob Pihl and Rasmus Jensen for working with me on software defined radio while they were students at DTU and now for being good colleagues.

I would like to express my gratitude to Tom Lee from Stanford University in California for being a great source of inspiration and for letting me join his group for a year during 1999 and 2000. Many thanks go to the SMIrC group a Stanford University for treating me like a full member. I am especially grateful for the many fruitful conversations I have had with Rafael Betancourt, Hirad Samavati, Hamid Rategh and Sunderajan Mohan. At Aalborg University in Denmark Troels Kolding and Erik Pedersen are acknowledged for their help with microwave probe measurements and general discussions on passive devices.

At Nokia Mobile Phones in Denmark my supervisors Dan Rebild and Ivan Riis Nielsen deserve the most gratitude for always taking good care of me. Furthermore my supervisor at DTU, Erik Bruun is acknowledged for always being there when I needed him.

Finally The Danish Academy of Technical Sciences (ATV) and Nokia Mobile Phones are greatly acknowledged for making it all possible by funding this project.

Kåre Tais Christensen

Nokia Mobile Phones

Copenhagen, Denmark, September 2001

Chapter 1: Introduction to RF Filtering in Wireless Terminals

This chapter serves as an introduction to the field of RF IC filtering in wireless communication terminals. The first section describes the current trends in the market and highlights the key drivers and their impact on the hardware development which is going on at the moment in the industry. RF filtering is identified as one of the major obstacles in the quest to meet the market demands of cost, size and functionality. The second section presents the different types of RF filtering which are needed in modern wireless communication devices. Then in Section 3 it is described how these filtering requirements are met today and it is explained why some of these solutions are not practical for future wireless devices. This discussion leads to Section 4 which lists the future challenges of RF filtering in integrated circuits. It is these challenges that form the basis for the work which is presented in this thesis.

1.1 Trends in Wireless Terminals

The wireless revolution of the late 1990’s was enabled by the phenomenal improvements in semiconductor technology. After more than forty years of exponential performance improvements, semiconductor technology had reached a maturity level at which robust communication devices could be built at relatively low cost and with low power consumption. The impact was an astonishing boom in the deployment of mobile phones. The wireless boom lead to massive investments throughout the world by numerous players who wanted their share of a seemingly endless market. This lead to overcapacity, prohibitively expensive frequency auctions and together with a dot-com business which was built on dreams rather than reality it all lead to a high-tech recession. Sales and prices dropped to disturbing levels in less than one year. Overnight one of the most profitable businesses of the world had turned into one of the worst with most of the major players in the world loosing billions of dollars each quarter. At the moment many people fail to see anything positive about the future of the wireless business.

What these people forget is that the situation really has not changed. It is people’s expectations which have changed. Now the glass is half empty rather than half full. But most people today can not imagine a future without widespread use of wireless multimedia terminals and somebody will be making those terminals so there is no reason to believe that this business will not once again be very profitable. When all comes to an end the only thing that matter is what the customer wants. Therefore it is wise to start by considering what the market mechanisms are.

1.1.1 Market Mechanisms

Cost. The single most important driver in the wireless market is cost. It is true that some people are ready to pay anything for a high tech toy but those people are few and therefore they are not important in the overall picture. The truly large volumes are found at the basic level and at the intermediate levels and the target customers at these levels are very conscious about the price of a product. This means that it is vitally important for a hardware company to have the lowest possible manufacturing costs.

Features. The market prices at the moment drop at approximately 30% per year. This is faster than the manufacturing costs can be reduced so the prices must be kept at a higher level. This is done by continuously implementing new features like games and picture messaging. This development has made software the most important component in wireless terminals.

Fashion. The wireless consumer electronics business today is a fashion business. Any company which fails to realize this is doomed. Most people bring their mobile phones with them anywhere and any time because they want other people to be able to contact them. This makes the mobile phone an accessory which automatically becomes associated with our personality. At the moment individuality is one of the most important trends and therefore it is critically important that the end used can customize the product to fit his personal taste.

Usability. It is obvious that a piece of wireless electronics should be easy to operate, but this requirement translates into a number of perhaps less intuitive requirements. For instance lack of standardization, political protection of home markets, general development, free competition and compromises over patents have lead to a very fragmented deployment of wireless standards. The user expects a product to work wherever he is and therefore in many cases a wireless terminal needs to support several wireless standards. This is particularly true in some parts of the world as well as for people who travel frequently. Another issue is the power consumption. It is very important that the user can use his equipment whenever he feels like it and that he should not plan or limit his use. This means that the product with a normal combination of talk time and standby should be functional for several days on a single battery charging.

Size. The size of a mobile phone is not as important anymore as it has been but it is still important. Many think that the smaller the phone is the more advanced is the technology inside and therefore they feel that prestige is gained by having the smallest possible phone. For most people however the optimal size is a trade-off between ease of operation and convenience in portability i.e. it should be large enough to be easy to use and small enough to be easy to carry.

Timing. The wireless market is changing rapidly and therefore it is very difficult to make reliable forecasts only a few years into the future. Unfortunately the development of a wireless terminal takes several years and its lifetime is only approximately one year so the delay of a new product of just one quarter can have a devastating impact of the profitability of the product. Naturally the same applies to incorrect forecasts.

Reliability. Finally there is the issue of reliability. It may not be too obvious but wireless products are becoming personal trusted devices. People trust that when they expect a very important call then people will be able to reach them. Also in the near future electronic payments will be carried out with wireless terminals. Imagine that you are at a restaurant with some customers and you can not pay for the meal because the software on your mobile terminal crashed and subsequently you can not get into your car for the same reason. This is not a good situation. Therefore reliability is a very important parameter in the mobile information society.

1.1.2 Impacts on Wireless Hardware Development

In the previous some of the trends and requirements of the wireless market were described. Here it will be explained how these requirements and trends affect the development of hardware for wireless terminals. In other words the trends in hardware development - specifically for the radio modem - will be explained in the following.

The key driver at the moment in hardware development is increased integration level. This is because many of the market requirements can only be met by integrating more and more functionality into fewer integrated circuits. For instance if multi-mode (= multi standard) functionality is to be implemented in a wireless product then the component count and cost may double. On a short term basis this may be acceptable but in the long run this is unacceptable due to natural price erosion. The only solution is to develop flexible hardware solutions which can be configured to work with different standards and this is most conveniently done on an integrated circuit.

Another market requirement which pushes hardware development to higher integration levels is the terminal size. Implementing more functionality in a smaller area is only possible through a higher integration level.

One of the largest and most expensive components in a mobile phone today is the battery so if the power consumption of the hardware can be reduced then a smaller battery can be used and thus both size and cost can be lowered. On an integrated circuit the parasitic capacitances are much smaller than at the board level and therefore it is possible to use higher impedance levels. This enables lower power consumption in the RF circuits and therefore integration leads to lower power consumption, as well as cost and size reductions in more than one way.

One additional argument for going to higher integration levels is the dynamic market environment which makes forecasting difficult. In order to be able to follow the wireless market at the moment it is very important that wireless phone and terminal development is carried out over a short period of time. At the moment development times can be counted in years where as the future demands require development times shorter than one year. This is only possible through reuse of verified modules and integrated circuits are ideal in that scope.

It seems obvious that higher integration is the right path to follow. This understanding has caused very widespread debate about whether or not systems-on-chip and single chip radios were feasible. For many - particularly in Academia - the pursuit for single chip systems has turned into a religion. It is definitely true that single-chip is the optimal solution for some applications but many fail to see that single-chip is not the optimal solution for many other applications. One example is Bluetooth, a short distance wireless local area network, which many thought was ideal for a single-chip solution. Earlier this year Alcatel proudly announced that they had achieved this goal by implementing a fully functional single-chip Bluetooth modem with memory, DSP processor, microprocessor, radio modem and even an antenna integrated in the IC package. Then only six months later they had to publicly announce that they were abandoning the single-chip concept partly because they had problems with the RF circuitry and partly because they were not able to technology migrate the digital parts. Technology migration to a new (CMOS) process is a standard practice which lowers cost and power consumption while increasing performance - but that is not possible when there is analog and RF circuitry on the chip because these parts need to be completely redesigned and that can easily take one to two years.

Apart from not being able to perform technology migrations there are a number of other issues that can make a single-chip solution less attractive. One common argument against single-chip is the coupling of digital switching noise to sensitive parts. Another argument is lack of isolation between the RF transmitter and receiver in full duplex systems i.e. systems where information is transmitted and received at the same time. Yet another argument is that standard CMOS technology is not the optimal technology for all blocks. For instance memory can be made cheaper in a specialized memory technology and power amplifiers for the RF transmitter can be made more power efficient in specialized power processes. Finally large ICs have lower fabrication yield so just from a pure cost point of view it may make sense to split a single chip solution into two or three smaller ICs.

The situation at the moment in many wireless systems is far from this two or three chip solution. Currently wireless phones and terminals have several hundred electronic components and the functionality is going to be much higher in the near future. Therefore higher integration levels are needed. Most of the components in the RF transceiver (transmitter and receiver) part which are off chip today can be integrated on an RF system IC. This is the case for VCOs (voltage controlled oscillators), PLLs (phase locked loops), baseband filters, data converters, buffers, crystal oscillators (with an off-chip crystal), LNAs (low noise amplifiers) even PAs (power amplifiers) for some systems. There are also a few components which have not been successfully integrated. One example is the crystal which provides the very accurate reference frequency that is needed in most communication systems. This however is not a major problem because one reference frequency is sufficient also for multi-mode applications. A much worse problem is RF filtering.

RF filtering is one of the fundamental challenges in RF communication systems which has been hardest to integrate and the problem only becomes worse because the frequency spectrum is being used more with closer frequency allocations as a result. Actually there are RF filtering tasks which have never been successfully integrated for high performance systems like GSM. One example is the front-end filters i.e. the filters which are located close to the antenna - either in the RX (receiver) path or in the TX (transmitter) path or in both. These filters are unique for each frequency band and therefore it is necessary to use at least one filter for each frequency band. For a multi-mode phone this can easily add up to ten or more filters. What is worse, these filters are expensive and they prevent the high integration level which is so important. The clear conclusion is that RF filtering is a major bottleneck in the quest for highly integrated wireless transceiver solutions. This means that if the RF filtering functionality could be integrated onto an RF chip in a standard technology then there would be a tremendous business potential extending far beyond the cost of the eliminated filters. This is the main motivation behind the work presented in this thesis.

1.2 Filtering Requirements in Wireless Transceivers

This section provides a list of the different RF filtering requirements which must be met for the transceiver architectures which traditionally are used in wireless phones and terminals.

1.2.1 Out-of-Band Blocker Rejection

The antenna in a wireless device picks up all the RF signals that are present in the air. Usually the antenna has some selectivity i.e. it attenuates unwanted signals that are located far away from the wanted signal and thereby the antenna provides some RF filtering. Unfortunately the need for production tolerances and the uncertainty of the location of the hand makes antenna designers aim at wideband coverage and therefore even far away blockers (100MHz away) are only attenuated by one dB or less and as a consequence the antenna can not be used as a filter. It can at the most provide some additional design margin.

The unwanted signals can be categorized according to how and if they are potentially harmful to the reception of the wanted signal. First there are very strong out-of-band blocking signals (Figure 1). These signals are problematic because they can saturate the LNA and thereby desensitize the receiver. Alternatively out-of-band blockers can create intermodulation products that may corrupt the wanted signal. Therefore these out-of-band blocking signals are almost always attenuated with a passive front-end filter that attenuates all but the in-band signals and thereby ensures that the LNA is not saturated.

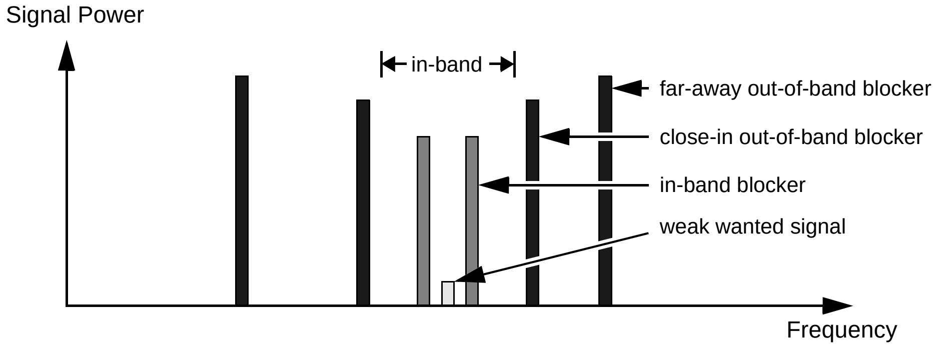


Figure 1. Unwanted blocking signals received from the antenna together with a weak wanted signal. The out-of-band blocking signals are the strongest and the in-band blocking signals are the closest to the wanted signal.

1.2.2 In-Band Blocker Rejection

The strong unwanted signals that are found in the allocated frequency band of a wireless system are called in-band blocking signals or just in-band blockers. These blockers are not as strong as the out-of-band blockers because their power levels are continuously monitored and controlled by the network. Instead they are located very close to the wanted signals and therefore it would be necessary to have an RF filter with extremely sharp edges if these in-band blocking signals were to be filtered out at the RF frequency. What is done instead is that this part of the frequency spectrum is moved to a lower frequency, an intermediate frequency (IF) or directly to baseband (around DC) before these blockers are filters out. The filtering at these frequencies is then called channel selection filtering.

1.2.3 Example - GSM1800 and GSM1900 Blocking Signal Requirements

The requirements for GSM1800 and GSM1900 are presented here to give an idea about what realistic blocking signal power levels are in a commercial wireless communication system. The requirements of GSM1800 (also called DCS1800) are shown in Figure 2 and the requirements of GSM1900 (also called PCS1900) are shown in Figure 3. At a first glance they look very similar and that is not a coincidence because by making the requirements as identical as possible it is easier to make multiband functionality i.e. to support different frequency bands at a reasonable cost.

GSM 1800 covers the frequency band from 1805-1880 MHz and GSM 1900 covers the frequency band from 1930-1990MHz. For the in-band blockers the requirements are the same: maximum -43dBm at 600kHz offset or more, maximum -33dBm at 1.6MHz offset or more and maximum -26dBm at 3MHz offset or more.

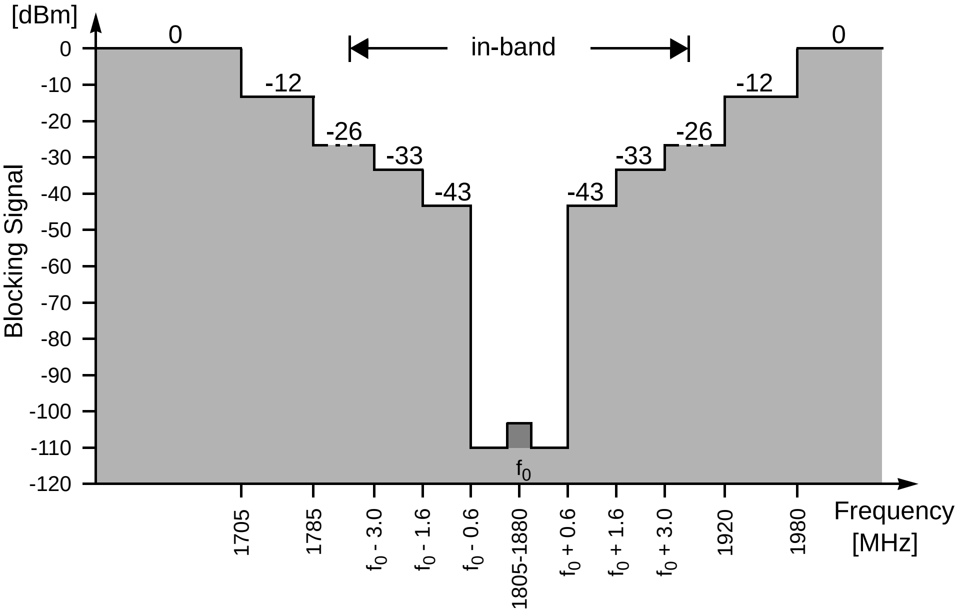


Figure 2. GSM 1800 Blocking Requirements.

For the strong out-of-band blockers the situation is a little different. For GSM 1800 the strongest close-in blockers are -12dBm below 1785MHz and above 1920MHz. The strongest far-away blockers are 0dBm and they are located below 1705MHz and above 1980MHz i.e. at least 100MHz from the allocated frequency band. For GSM 1900 the strongest close-in blockers are -12dBm below 1910MHz and above 2010MHz. The strongest far-away blockers are 0dBm and they are located below 1830MHz and above 2070MHz. It is here that the two frequency bands differ. The 0dBm blocker at 2070MHz is located at only 80MHz offset and this makes it the toughest front-end filtering requirement of these two GSM frequency bands.

Typically the out-of-band blocking signals are removed with passive front-end filters which have a worst-case in-band insertion loss of 3.5-4 dB. This corresponds to a reduction in sensitivity of the same amount and that is acceptable for GSM. If we try to implement the same functionality with an active filter it is necessary to consider the noise figure of such a filter. From [1] we have an expression which gives the maximum tolerable noise figure of the receiver for a given standard:

$$NF_{receiver,\,max}\big|_{dB} = P_{in,\,min}\big|_{dBm} - kT\big|_{dBm/Hz} - SNR_{min}\big|_{dBm} - 10\log B \tag{1}$$

In this equation $P_{in,min}$ is the weakest wanted signal, k is Boltzmann's constant, T is the absolute temperature, SNR_{min} is the minimum required signal-to-noise ratio and B is the signal bandwidth.

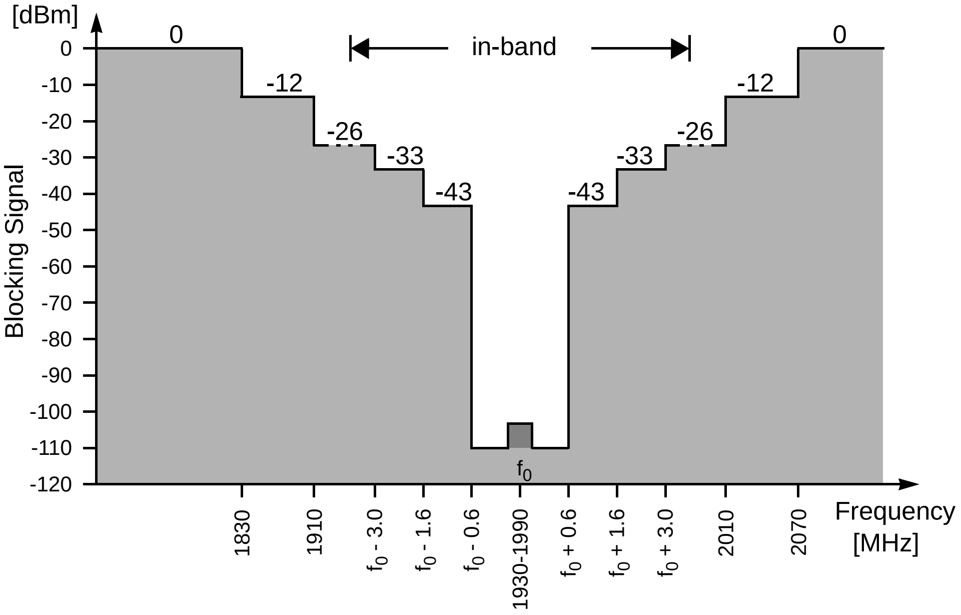


Figure 3. GSM 1900 Blocking Requirements

Then to reach an expression for the noise figure of the filter itself we need to take loss L before the filter into account. This loss mainly comes from the PWB, SMD matching components and perhaps from a balun. Finally we need to take the noise contribution of the subsequent stages N_{extra} into account. According to Friis equation this is not significant if the gain of the first stage is high but we still need to remember it. This leads to:

$$NF_{filter,\,max}\big|_{dB} = P_{in,\,min}\big|_{dBm} - kT\big|_{dBm/Hz} - SNR_{min}\big|_{dBm} - 10\log B - L - N_{extra} \qquad (2)$$

For GSM this is equal to:

$$NF_{filter,\,max} = -102dBm + 174dBm/Hz - 9dB - 10\log 200kHz - 1dB - 1dB = 8dB \qquad (3)$$

To fulfill the requirements of a GSM terminal in production it is not sufficient that the filter achieves this noise figure of 8dB at one temperature for a typical device. There must be a couple of dB's of margin to take production tolerances and temperature variation into account. Therefore to make a functional fully integrated front-end filter for GSM it must have a noise figure which is better than 6dB.

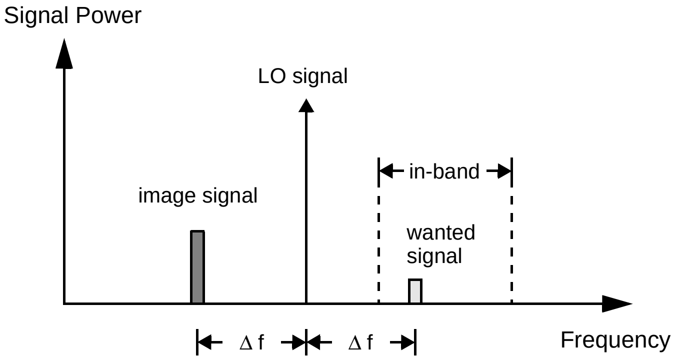


Figure 4. Problem of image signal in super-heterodyne receivers. The image signal is superimposed on the wanted signal after the first frequency conversion.

1.2.4 LO Attenuation and DC Offset Compensation

In a direct conversion receiver the local oscillator reference signal (LO) is located at the center frequency of the wanted signal. Some fraction of this LO will leak to the input of the receiver and be superimposed on the wanted signal. This LO component can not be removed at the RF and therefore it should be minimized through cautious layout and through the use of a balanced topology. The LO component will follow the wanted signal all the way to baseband where it takes the form of a DC offset. At this point the DC offset, if it is not too large, can be removed with DC offset compensation circuits which basically provide some form of highpass filtering.

1.2.5 Image Rejection

Weak signals at most other frequencies have negligible effect on the received signal and therefore they are irrelevant with respect to filtering. There is one exception and that is the case where a super-heterodyne receiver architecture is used. In this case there is one frequency, the image frequency, which is superimposed on the wanted signal after the first frequency conversion (see Figure 4). Once that has happened it is not possible to recover the wanted signal if the image is stronger than the wanted signal. Therefore it is important to ensure that the image is rejected by a sufficient amount before or during the first frequency conversion. This image rejection can be done with traditional (passive) bandpass filtering or with a special class of mixers (frequency converters) called image-reject mixers.

1.2.6 Noise Filtering

In some situations it is necessary to filter wideband RF noise. One of the most common examples is transmitter noise in the receive band. When a strong signal is transmitted in the upper end of the transmit band it may have wideband noise with sufficient energy to corrupt a weak wanted signal in the lower end of the receive band (Figure 5).

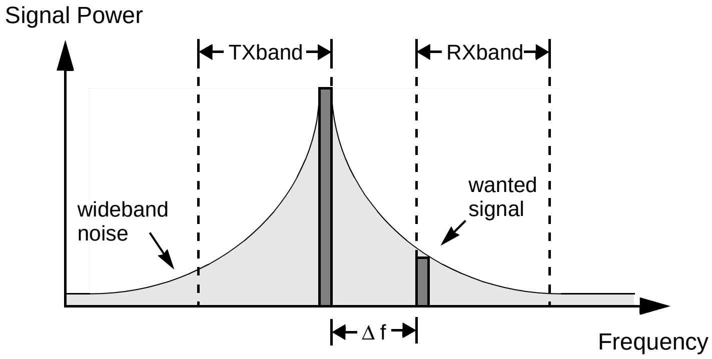


Figure 5. Transmitter wideband noise in the receive band. A transmitted signal can have wideband noise which contains more energy than a weak signal in the receive band. Therefore most systems have strict requirements for TX noise. RF filters are often needed to meet these requirements.

Therefore most systems have strict requirements for transmitter noise. For instance in GSM1800 and GSM1900 the requirement is that the noise in a one hertz bandwidth at a 20MHz offset (Δf) must be at least 151dB weaker than the carrier.

1.2.7 TX - RX Isolation

Finally there is the problem of the transmitted signal coupling to the receiver input. If the receiver is connected to the antenna during transmission then significant power is dissipated in the receiver input leading to poor PA efficiency and possible destruction of the receiver. Naturally therefore some means of isolation is always provided between the receiver and the transmitter. This can be antenna switches and/or passive filters traditionally called duplex filters. If the system is a full-duplex system then the requirements for isolation are tightened further to avoid saturation of the receiver during operation.

1.3 Common Filtering Solutions

It is obvious that all the filtering requirements that were presented in the previous can be met with current technology as there are countless products in the market that can pass a type approval (a test that documents conformity with a given standard). This section describes the filtering solutions which are commonly used today and in some cases it is also explained why the given technique is not attractive for future wireless terminals.

1.3.1 Passive Filters

Passive filters are by far the most common components used to provide RF filtering in wireless transceivers. In the following the most common passive filters are listed.

LC Type Filters. The most basic passive filter class is the LC type filters. These filters are constructed out of inductors and capacitors - typically coupled in a ladder structure. They have low cost but their edges are not very steep and they have relatively high insertion loss due to the moderate Q-value of the passive devices. Furthermore they have high temperature coefficients so this class is rarely used for demanding RF filtering situations like the front-end filter or the duplexer. As an additional comment it can be mentioned that these filters generally can not be integrated on a silicon chip due to very low Q-values and they are not tunable i.e. they can only be used for one frequency band. It is however possible to achieve both monolithic compatibility and tunability if micro-machining is used. Unfortunately micro-machining is not compatible with low cost, and mass production at the moment.

Edge Coupled Filters. Edge coupled filters can be made as striplines in a printed wire board (PWB) and therefore they can be made with the lowest cost. Unfortunately the wavelengths in the lower GHz range lead to very large structures which can not be tolerated in most mobile phones or wireless terminals. Insertion loss and out-of-band attenuation can be acceptable for some less demanding standards like for instance DECT.

Dielectric or Ceramic Filters. In order to reduce the physical volume needed for the edge coupled filters it is possible to use a material with a dielectric constant which is much higher than that of FR4 (standard PWB material) i.e. much larger than 4. Typically ceramic material with a dielectric constant in the order of 100 is used - hence the name ceramic filter. These filters can provide low insertion loss, steep skirts and very good power handling capabilities at acceptable component cost. For that reason these filters are commonly found in wireless products. The negative side is that they are still relatively large and they are not tunable so it is necessary to have one filter for each frequency band.

Electro-Acoustic or Piezoelectric Filters. Electro-acoustic or piezoelectric filters are filters that use the special properties of piezoelectric material namely that electric waves are coupled to mechanical waves. This is important because the mechanical properties of these materials can determine the filter transfer characteristics very accurately. Common variants are SAW (surface acoustic wave), BAW (bulk acoustic wave), FBAR (thin-film bulk wave acoustic resonator) and crystal filters. These filter types which are very common at the moment offer low insertion loss, steep skirts, small area and moderate power handling capabilities at reasonable cost [4]. Unfortunately these filters, like the others, are not tunable nor compatible with standard IC technology so the overall cost when used in a multi-band or multi-mode phone is still high. The FBAR/BAW structures can be fabricated on a silicon substrate but then the process is no longer a standard process and then cost becomes an issue again.

1.3.2 Active Filters

Active filters are rarely used for challenging RF filtering tasks because the requirements for dynamic range are extremely high. We do however have to consider active filters because they facilitate frequency tuning and can be integrated in standard IC technology. Properties which are much needed in future generations of wireless multi-mode and multi-band terminals. This sub-section presents the most common active filter techniques.

Gm-C and Active RC Filters. The Gm-C and Active-RC filter types are continuous time filter topologies which are very common in lower frequency applications (below 100 MHz). Unfortunately at higher frequencies these filter types have serious limitations in terms of linearity and noise which disqualifies them from being used as RF filters in the GHz range.

Switched Capacitor Filters. Switched capacitor filters are also very common in lower frequency applications and a few higher frequency applications. These structures do provide better dynamic range than the Gm-C and active-RC filters but they are discrete time filters which means that they rely on sampling in the time domain. When the sampling is performed then signal power which is located beyond the sampling frequency is folded down to a low frequency corrupting the desired signal. If there is not any significant power at higher frequencies then switched-cap could be a good solution but in these applications there can be signals at tens of GHz that are several orders of magnitude stronger than the desired signal. This means that switched-cap is ruled out for GHz RF filtering - at least without some continuous-time filtering used first.

Q-enhanced LC Filters. Filters which use loss compensation of LC resonators are called Q-enhanced LC filters. The philosophy behind these filters is that passive filters have higher dynamic range than active so the best RF performance is achieved by helping LC resonator type filters. First the passive LC filter is designed to have the highest possible Q-value (which is often not more than 2-4) and then the filter is helped by inserting a few highly linear negative resistance circuits. This type of filter has shown the best RF performance of all the active filter topologies [5], but it is still not enough for e.g. a GSM type product and the power consumption is too high. It is however believed (by the author) that this topology has potential to meet the toughest RF filtering requirements of GSM.

Gyrator-C or Active Inductor Filters. There is a circuit topology which can invert an impedance. This circuit is called a gyrator. The idea is that if a capacitor with a reasonably high Q-value is used in a gyrator coupling then it will emulate a high-Q inductor hence the name Active Inductor. This active inductor can then be used with other high-Q capacitors to form high-Q resonators at RF frequencies. The idea is sound but the problem is that the gyrator circuits - as most of the other active circuits - have serious problems with noise and linearity [6]. Some experts, however, believe that when transistor f_T's reach 200GHz then active inductor filters might yield satisfactory RF performance.

Actively Coupled LC Resonators. Another way of exploiting the inherent high dynamic range of passive LC resonators is to increase the filters order by simply cascading many low-Q passive LC filters. Each filter is separated by a low gain amplifier which provides isolation between the resonators and reduces the loss of the subsequent filter stages. The problem with this topology is that it requires very high dynamic range in the input stage (as the other active filter topologies) and has a high power consumption due to the many filter stages. It is however believed that the method is feasible perhaps in combination with the Q-enhancement techniques.

ADC - DSP Filters. An analog-to-digital converter followed by a digital filter perhaps implemented in a digital signal processor is more of an academic solution because it would require approximately 20 bits of resolution and a sampling rate of several GHz together with an extremely large DSP. This concept which is often referred to as software radio is considered impossible by most experts. Mostly a result of the very slow improvements in the data converter field which suffers from the low supply voltages of current IC technologies. As a final note it must be emphasized that the term software-defined radio is not in any way related to the software radio concept. It simply expresses that a chosen radio architecture is flexible enough to be reconfigured to support other similar standards.

1.3.3 Architectural Choices

In the previous most commonly used filtering techniques were mentioned. In this subsection architectural choices are presented which can eliminate the need for some of the filtering requirements.

Direct Conversion. When the direct conversion receiver architecture is used the wanted signal is converted directly to baseband. This means that there is no intermediate frequency and therefore no intermediate frequency filtering is needed. Also the wanted signal is its own image and therefore image rejection is not an issue - or more accurately it is much easier to deal with. Front-end filtering is still necessary and for this topology DC offset compensation is also important. One of the most attractive features of this architecture is that a part of the front-end filtering can be moved to baseband by increasing the linearity of the LNA and mixer [7] and by reducing the noise of the baseband filter [2]. Another advantage of direct conversion is that multi-mode operation is relatively simple once the front-end filtering problem is solved.

Image Reject Mixers and Topologies. For super heterodyne receiver architectures i.e. architectures that use intermediate frequencies it is very important that the image components are rejected. The traditional approach is to use passive off-chip filters. This is not compatible with the requirement of higher integration levels. Therefore a number of on-chip image reject architectures have been investigated. Most of these exploit that by using complex signal processing it is possible to process the wanted signal with the same sign and the image with different signs and then after addition the image is rejected. The quality of this image rejection is determined by matching properties in the mixer and by the quality of the quadrature signals (typically four phases of the same signal) which are needed for the complex signal processing. Amongst others the Hartley and Weaver architectures have been used for image rejection using complex signal processing. It is possible to use the LO or the RF or both in quadrature for image reject mixers. When the LO or the RF is used in quadrature then approximately 40 dB of image attenuation can be provided and with both the RF and LO in quadrature then approximately 60dB of image rejection can be achieved [3]. In many cases this is barely enough and there are also some serious power/noise trade-offs associated with the quadrature generation of the RF signal. Therefore image reject mixers and quadrature generation circuits have received significant attention in recent years.

There are a few other image reject techniques worth mentioning. If the LO frequency is chosen to be exactly half the frequency of the wanted signal then the image falls on 0Hz i.e. at DC and then the image can be rejected effectively by a simple AC coupling. Then we have the wanted signal at half the frequency and we have a new image problem but because the frequency is lower it might be easier to handle the image. Finally it is possible to implement a Q-enhanced LC resonator notch filter in the LNA rejecting the image by 10-20 dB.

Offset Loop TX Architectures. For the transmitter noise there are also a few architectural choices which can relax the filtering requirements. For instance the use of a power VCO together with an offset loop can produce a very pure output spectrum i.e. produce very low wideband noise and thus a passive filter may not be need.

1.4 Future Challenges in Integrated RF Filtering

In order to solve most of the RF filtering requirements in standard CMOS technology there are a number of IC design challenges which need to be met first. Most of these challenges deal with dynamic range i.e. the ability to handle very large signals while adding the least possible amount of noise. This section describes some of the most important challenges that we were facing in the beginning of 1998 when this work was initiated.

High-Linearity Low Power LNA Design. On-chip inductors have low Q-values and therefore passive filtering gives too much insertion loss. This means that the first block which the signal from the antenna meets on the IC is an LNA. This LNA must be able to process 0dBm blocking signals while amplifying -110dBm to -105dBm weak signals. This kind of linearity and noise performance has not been shown at reasonable power consumption levels (less than 10-20 mA) so this is a major challenge.

High-Q Inductor Design. At the output of the LNA the signal will usually meet an LC resonator of some kind. This resonator must have the lowest possible loss to provide frequency selectivity and to reduce the noise. Typically the loss of the resonator is dominated by the loss of the inductors. This means that research in high-Q inductor design is critically important.

High-Q Capacitor Design. Because the overall resonator loss is so important for the filter performance it is also highly desirable to improve the quality of capacitors. It is especially important to conduct research in capacitor design for standard processes because these usually do not include MIM (metal-insulator-metal) and for other capacitor structures Q-values can be rather poor. Also it would be highly beneficial if capacitors could be designed with small bottom plate capacitances because the parasitic capacitances tend to reduce RF performance. Finally small shielded High-Q capacitors would be very useful as unit capacitors in binary weighted switched capacitor banks.

Accurate Capacitance Modeling. Accurate capacitance modelling reduces development time because fewer design iterations are needed. Alternatively it improves RF performance because smaller design margins are needed. For RF filters in particular, accurate capacitance modelling reduces the needed tuning range and thus improves noise and linearity.

Wide Tuning Range. Because multi-band and multi-mode operation is the clear trend in RF development for wireless terminals it is important that on-chip RF filters can be tuned to cover several frequency bands over a wide frequency range. This is a major challenge because the linearity must still be good and the loss needs to be very low.

Low Loss Switch Design. Switched tuning is perhaps the only way to implement a wide tuning range without significantly degrading linearity or noise. Unfortunately switches are or were known to be very lossy. This means that investigation of optimal switch design is very important for the development of on-chip RF filters for multi-mode and multi-band transceivers.

High Resolution. Switched frequency tuning can be used for coarse tuning like frequency band switching but it can also be used for fine tuning. If a high resolution binary weighted switched capacitor bank can be realized at a given RF frequency then it may be possible to exclude varactor tuning. This represents a significant step towards a robust filter solution because frequency drift is greatly reduced and noise coupling is much less of a problem. With a switched capacitor tuning solution it may be possible to tune the filter with an adaptive algorithm implemented in DSP software and it may even be sufficient to use a lookup table after a single calibration has been performed.

High Dynamic Range Loss Compensation. It has been pointed out several times that on-chip inductors have very poor quality factors. This means that some kind of loss compensation is almost always needed to produce sufficient frequency selectivity. These loss compensation circuits also have extreme requirements for dynamic range when they are used for e.g. fully integrated front-end filters. In other words research in loss compensation with high linearity and low noise is extremely important for the feasibility of the most demanding RF filtering tasks.

Good Balance Without Tail Currents. It has been explained why direct conversion is attractive for multi-band and multi-mode operation. In direct conversion good signal balance is important to reduce DC offsets. Traditionally good balance is achieved through common-mode rejection implemented with a tail current. Unfortunately the tail current noise is modulated into the frequency band of interest. This is especially the case when strong blocking signals are present and that is the case for these filters. As a result it is important that the resonator structures are well balanced.

Front-End Filter Design Method. Tunable front-end filters may reduce the hardware cost of multi-band and multi-mode wireless terminals by as much as several Euros per terminal - this is a significant amount of money in mass production. Unfortunately, even if the needed circuit components can be designed, there is at present no design methodology for fully integrated front-end filters. Therefore, naturally, the development of a fully integrated front-end filter design methodology is of utmost importance for the integration of this filter.

Low Noise RF Quadrature Generation. In superheterodyne receivers image rejection is very important. Typical image reject mixers that use the LO in quadrature can only provide partial image rejection so a different solution is necessary. Double quadrature mixers that use both the LO and the RF in quadrature can provide sufficient image rejection in many cases. Unfortunately the RF can only be brought into a quadrature form by using RC polyphase networks which are lossy and noisy. Consequently a new low noise RF quadrature generation circuit would be very useful.

1.5 Summary

This chapter presented an introduction to RF filtering in wireless terminals. The first section covered trends in the consumer market and in hardware development and concluded that RF filtering is one of the major bottlenecks in the further development. Section 2 listed the most common RF filtering requirements and Section 3. presented the filtering solutions which are used at the moment. Then finally Section 4. presented a list of future challenges in integrated RF filtering.

1.6 References

[1] B. Razavi, "RF Microelectronics", *Prentice Hall*, 1998.
[2] R.G. Jensen, K.T. Christensen and E. Bruun, "Programmable Filter for Multi-Mode Phones", *submitted for the 2001 NorChip Conference*.
[3] J. Crols and M. Steyaert, "CMOS Wireless Transceiver Design", *Kluwer Academic Publishers*, 1997.
[4] R. Ruby et. al., "Ultra-Miniature High-Q Filters and Duplexers Using FBAR Technology", *In Proc. of 2001 IEEE - ISSCC*, pp.120-121, 438.
[5] W.B. Kuhn et. al., "Q-Enhanced LC Bandpass Filters for Integrated Wireless Applications", *IEEE Trans. on Microwave Theory and Techniques,* Vol. 46, No. 12, Dec. 1998, pp. 2577 - 2586.
[6] T.H. Lee, "The Design of CMOS Radio-Frequency Integrated Circuits", *Cambridge University Press*, 1998, pp. 268-269.
[7] J. Pihl, K.T. Christensen and E. Bruun, "Direct Downconversion With Switching CMOS Mixer", *In Proc. of 2001 IEEE - ISCAS*, Vol. 1, pp. 117-120.

Chapter 2: Input Power Constrained LNA Design

If an integrated RF frontend filter is to replace an off-chip SAW filter it must have an input stage that can process weak signals typically in the order of -100 dBm in the presence of very strong out-of-band blocking signals - often in the order of 0 dBm. Such extremely high dynamic range is rarely found in the literature. The few exceptions are LNAs intended for basestations and they typically use several hundred mW of power which can not be tolerated in battery powered devices.

This chapter describes a new LNA design methodology which gives closed form expressions for all the component values and bias levels which are needed to design an LNA with a specific real input impedance and a specific input power handling capability with a desired current consumption. By using this method it is possible to reduce the power consumption of such LNAs by up to an order of magnitude without degrading the noise figure by more than approximately 1 dB.

2.1 Classic LNA Topologies

The LNA is not only responsible for amplifying weak signals while adding the least possible amount of noise. Being the first active circuit in a receiver chain these amplifiers need to interface to passive structures which are located outside the integrated circuit. Typically these structures are ceramic or SAW filters that attenuate out-of-band blocking signals and unwanted signals at an image frequency. Alternatively the LNA can be connected directly to an antenna through an unknown length of transmission line or perhaps through a balun if the LNA is balanced. In either case it is important that the LNA has a well defined real input impedance typically in the order of 50 ohms because filters, microstrips and baluns almost always need to be terminated in a real impedance.

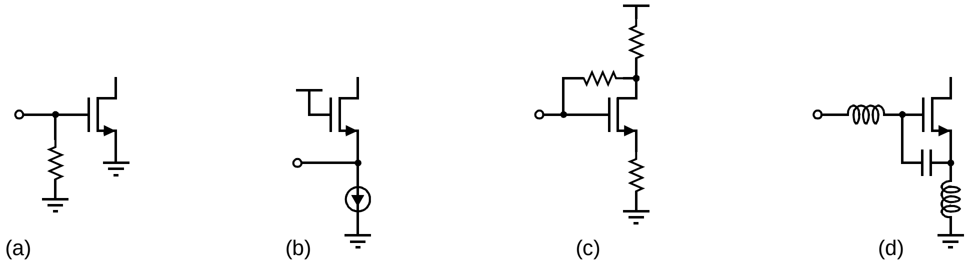


Figure 6. LNA topologies. Resistive termination (a), common gate (b), shunt-series feedback (c) and inductive source degeneration (d).

It has been argued that a good power match is not necessary for CMOS LNAs because it does not matter if the input power is reflected back to the antenna. What matters in CMOS is whether or not the input signal causes a sufficient voltage swing at the gate of the input transistor and that can easily be achieved without a good power match. This is all true but for most practical cases of interest the input must have a good power match to make the board level design issues manageable. Therefore this chapter only considers LNA topologies which can provide a well defined real input impedance.

The simplest way to provide a well defined real input impedance is to use a resistor at the input Figure 6 - (a). Even though this topology provides a good wideband power match it is largely useless for low noise amplification because the resistor contributes with excessive thermal noise and simultaneously reduces the input power before amplification. The next structure which is shown in Figure 6 - (b) is the common-gate amplifier stage. If the current source has an infinitely high output impedance then the input impedance is simply the inverse of the transconductance so provided a reasonably low impedance level at the drain it is a simple matter to achieve the desired input impedance and over a wide frequency band. A simplified analysis shows that the lower limit for the noise factor of this LNA is given by (4), [8] where γ is the drain noise constant which is approximately 2/3 in the long channel regime when the transistor is operated in saturation. α is the mobility degradation factor which is 1 at low gate-source voltages and decreases with the gate source voltage.

$$F_{1/gm} = 1 + \frac{\gamma}{\alpha} \tag{4}$$

This means that the theoretical lower limit for the noise figure of this LNA topology is 10Log(5/3) ~ 2.2 dB but at useful bias levels α is lower than 1 and γ may be higher than 2/3 due to hot-electron effects etc. so the more practical lower limit is a NF in the order of 3 dB. For some applications this is quite acceptable and the topology is also attractive because it enables good linearity and uses no passive on-chip components so it can be implemented in a very small silicon area.

The third approach (Figure 6 - c) uses resistive shunt and series feedback to provide the desired input and output impedances. This leads to a good wideband power match but unfortunately it also implies high power consumption to provide the desired gain. Furthermore good, well defined resistors are usually not available in silicon processes and even though the topology does not suffer from the same problems as the first topology the resistors still contribute with substantial thermal noise. Together these issues make the shunt-series feedback topology less attractive for LNAs in portable narrowband applications.

Then finally we have the inductive source degeneration LNA topology (Figure 6 - d) in which the input signal is fed to the gate terminal and to the source terminal through the parasitic gate-source capacitance. The topology uses a combination of phase shift and capacitive coupling to provide a real LNA input impedance [9]. The real part of the input impedance does not change with frequency but the imaginary part does so this topology must be considered a narrow band topology even though a reasonably good power match can be provided over a fairly wide frequency band. The topology’s main advantage is that it enables good simultaneous power and noise match i.e. it allows high gain and a well defined real input impedance while adding a minimum of noise. Sub-1dB noise figures have been reported for inductively source degenerated CMOS LNAs operating in the lower GHz range [10]. These features have made this the most commonly used topology for CMOS LNAs and it is also the one that is pursued in this work.

2.2 Input Match of the Inductively Degenerated LNA

In order to be able to design the LNA input stage in an efficient manner it is very important that the design constraints are expressed in a clear formalized manner. This allows identification of the degrees of freedom in the topology and thus leads to design insight which can be used to enhance circuit performance.

The first design constraint is the demand for a good power match at the input. Figure 7 shows a source degenerated LNA with a cascode transistor as load. C_{gs} is the parasitic gate-source capacitance of the input transistor which has a transconductance of g_m. L_s is an inductance that lowers the gain through negative feedback. This improves linearity but the main reason why it is there is that it helps creating the desired input impedance. L_g represents the bondwire inductance and perhaps some PWB trace. It is shown because it can be used to null the imaginary part of the input impedance.

$$Z_{in} = \frac{v_{in}}{i_{in}} = \frac{g_m L_s}{C_{gs}} + j\left(\omega(L_g + L_s) - \frac{1}{\omega C_{gs}}\right) \tag{5}$$

It is straightforward to show that the input impedance of this circuit is given by (5). From this equation it is evident that the real part of the input impedance is independent of frequency.

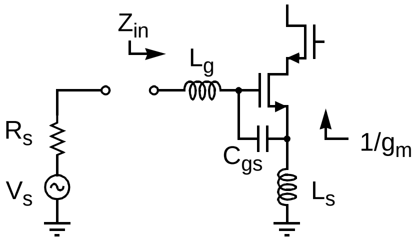


Figure 7. Source degenerated low noise amplifier input stage. L_s can be an on-chip inductor or a bondwire, C_{gs} is the parasitic gate-source capacitance of the input transistor and L_g is the bondwire and perhaps some PWB trace.

The imaginary part on the contrary is not. It assumes the form of a series LC resonator which only equals zero at one frequency. This means that the input impedance can only provide a good power match over a certain bandwidth.

A good power match means that the input impedance Z_{in} should equal the source impedance R_s. As we are dealing with complex numbers this corresponds to the two design equations (6) and (7) that both need to be fulfilled.

$$Re(Z_{in}) = \frac{g_m L_s}{C_{gs}} = R_s \tag{6}$$

$$Im(Z_{in}) = \omega(L_g + L_s) - \frac{1}{\omega C_{gs}} = 0 \tag{7}$$

2.3 Class A - Operation in the Presence of Strong Blocking Signals

Usually LNAs are preceded by passive front-end filters that attenuate out-of-band blocking signals by 20-30 dB. This means that the needed power handling capability of LNAs is usually quite manageable so reducing the noise figure tends to be the main concern in the design phase. For LNAs that are used as input stages for fully integrated RF front-end filters this is not the case. The whole point of integrating RF front-end filters is to eliminate the often quite considerable cost of passive front-end filters. This means that such LNAs must be able to tolerate very strong out-of-band blocking signals. For instance according to the GSM standard a receiver must be able to tolerate out-of-band blocking signals that are as strong as 0dBm and without passive front-end filters this translates to LNA power handling capabilities in the same order. These are extremely tough requirements so it is important that the LNA design procedure which is developed here takes the maximum input power level into account.

Instead of focusing on IIP3 or ICP it is simpler and perhaps more relevant to design the circuit such that it operates in class A also when a strong out-of-band blocker is present.

After all the circuit must be able to process weak signals in the same manner with or without the blocking signals.

Class A operation implies that the peak output current amplitude should not exceed the bias current (I_D) at the maximum input power level (P_{in}). To achieve the best possible power efficiency it is also important that the current swing is maximized i.e. that the maximum output current amplitude is invoked when the maximum input power is fed to the LNA. This means that the power-to-current gain of the input stage is fixed for a given bias current and maximum input power level.

Before formulating the class-A operation requirement let us review a few properties of input power, voltage and current under matched conditions. The input power can be expressed as (8) which is equivalent to peak input voltage and currents as given by (9).

$$P_{in} = \frac{v_{in,\,peak}^2}{2R_s} = \frac{i_{in,\,peak}^2 R_s}{2} \tag{8}$$

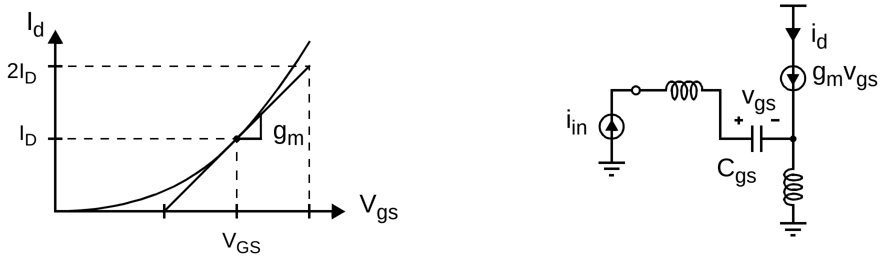
$$\begin{aligned} v_{in,\,peak} &= \sqrt{2P_{in}R_s} \\ i_{in,\,peak} &= \sqrt{2P_{in}/R_s} \end{aligned} \tag{9}$$

Now let us make a linearized approximation for the transistor I-V curve. Figure 8 - (a) shows such an approximation. When the transistor is biased at a certain gate source voltage (V_{GS}) with a corresponding bias current (I_D) and a transconductance (g_m) at this point then it is assumed that the transistor has a perfectly linear I-V curve with a slope of g_m for gate source voltages from V_{GS} - I_D/g_m to V_{GS} + I_D/g_m.

This might seem like a very crude approximation but it is actually not that bad because short channel MOS transistors experience significant mobility degradation when biased at the high gate-source voltages which are needed to accommodate the high input power levels. This means that they have a much more linear behavior than long-channel MOS transistors which are quadratic in nature. Even at the maximum peak-to-peak gate voltage swing of $2I_D/g_m$ the current swing is still very close to $2I_D$ because the LNA enters class AB operation instead of simply saturating. Therefore the approximation is good enough for derivation of design equations and for gaining design insight.

Next let us look at the current gain of the source degenerated LNA stage. Figure 8 - (b) shows a simplified circuit diagram of the LNA input circuit. The drain current I_d is given by the gate-source voltage multiplied by the transconductance g_m and the gate-source voltage is equal to the input current i_{in} multiplied by the impedance of the gate-source capacitance Z_{Cgs}. This leads to an expression for the current gain of the input stage (10).

$$i_d = g_m v_{gs} = g_m Z_{Cgs} i_{in} = g_m \frac{-j}{\omega_0 C_{gs}} i_{in} \tag{10}$$



(a) (b)

Figure 8. Linearized MOS transistor I-V curve (a) and simplified schematic for calculation of the current gain of a source degenerated input stage (b).

The linearized circuit approximation allows us to assume that equation (10) holds for the full output current swing from 0 to $2I_D$. As argued before the peak drain current $i_{d,peak}$ should equal I_D when the maximum input power is delivered to the LNA. Equation (10) can now be used to express I_D in terms of the peak input current which can be substituted with expression (9). Thereby we have derived an expression that links the maximum input power to the current consumption of the LNA (11).

$$I_D = i_{d,peak} = \frac{g_m}{\omega C_{gs}} i_{in,peak} = \frac{g_m}{\omega_0 C_{gs}} \sqrt{2P_{in}/R_s} \qquad (11)$$

Rearranging (11) yields a design equation (12) that must be fulfilled to guarantee class-A operation at the maximum LNA input power. This expression can then be used to derive expressions for the other passive components i.e. the inductors. The value of the source degenerating inductor L_s can be found by rearranging expression (6) and substituting Cgs with expression (12). This is done in (13) and as it can be seen the expression for L_s is very simple and only involves external parameters i.e. L_s is fixed when source impedance (R_s), maximum input power (P_{in}), drain bias current (I_D) and center frequency (ω_0) are given.

$$C_{gs} = \frac{g_m}{\omega_0 I_D} \sqrt{\frac{2P_{in}}{R_s}} \qquad (12)$$

$$L_s = \frac{R_s C_{gs}}{g_m} = \frac{R_s}{\omega_0 I_D} \sqrt{\frac{2P_{in}}{R_s}} = \frac{\sqrt{2R_s P_{in}}}{\omega_0 I_D} \qquad (13)$$

With expressions for C_{gs} and L_s it is now possible to derive an expression for the gate inductance L_g. The expression (14) is found by simply isolating L_g in (7) and substituting C_{gs} and L_s with (12) and (13).

$$L_g = \frac{1}{\omega_0^2 C_{gs}} - L_s = \frac{\sqrt{R_s}}{\omega_0} \left(\frac{I_D}{g_m \sqrt{2P_{in}}} - \frac{\sqrt{2P_{in}}}{I_D} \right) \qquad (14)$$

2.4 Addition of External Gate-Source Capacitance

Now let us consider the implications of (12). First recall that the unity current gain frequency measure ω_T of a MOS transistor is approximately equal to g_m/C_{gs} [9]. Rearranging (12) yields (15) which can be cast into the form (16) which can tell us how much bias current is needed to guarantee class-A operation.

$$\omega_T = \frac{g_m}{C_{gs}} = \frac{\omega_0 I_D}{\sqrt{2P_{in}/R_s}} \tag{15}$$

$$I_D = \frac{\omega_T \sqrt{2P_{in}}}{\omega_0 \sqrt{R_s}} = \frac{f_T \sqrt{2P_{in}}}{f_0 \sqrt{R_s}} \tag{16}$$

If for example f_0 = 2.0 GHz, R_s = 50 Ohm, P_{in} = 0dBm and f_T = 30 GHz (which is common at these bias levels in standard 0.25 micron CMOS) then I_D = 95 mA! This is clearly too much for most portable applications. Usually the highest power consumption that can be tolerated in an LNA for portable applications is at least an order of magnitude lower. R_s could be increased a little through off-chip matching networks but because R_s is found under a square-root sign it is far from enough to reduce the current consumption to an acceptable level.

The conclusion is that f_T must be lowered in order to reach an acceptable bias current level. This can only be done if another design parameter is introduced that decouples the relation between g_m and C_{gs} i.e. an extra degree of freedom is needed.

The most straightforward way of reducing f_T is to increase the transistor length L. This reduces g_m and increases C_{gs}. Increasing L also increases the output impedance because channel length modulation becomes less pronounced and it lowers the lateral fields in the channel so it might also reduce the excess thermal noise which is believed to come from hot carrier effects [8]. However as the latter effects are not well understood they are difficult to take into account. Increasing L also has some undesirable effects. First it increases the transistors threshold voltage and second it reduced the mobility degradation effects so the transistor gets a more quadratic I-V behavior i.e. the input stage becomes less linear.

Alternatively PMOS transistors which have an approximately 3 times lower f_T can be used in the input stage. This could yield quite good results as PMOS transistors are believed to have less excess thermal noise than their NMOS counterparts and they are usually also made with lower threshold voltages because leakage is less of a problem.

A third method which does not alter the input transistor is to add an external linear capacitor C_{gsx} between the gate and source terminals of the input transistor.

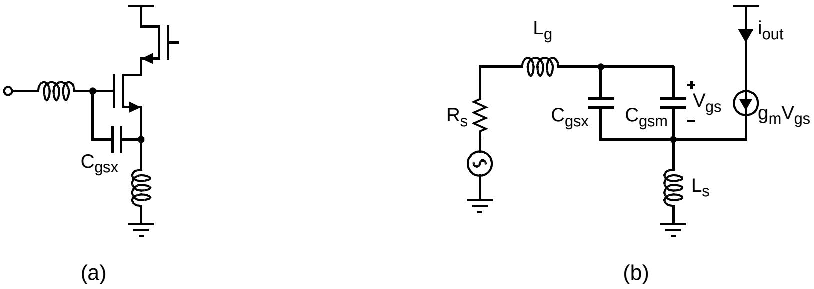


Figure 9. LNA input circuit with external gate source capacitance C_{gsx}. Schematic (a) and small signal diagram (b). C_{gsm} represents the intrinsic gate-source capacitance.

This way the effective f_T ($f_{T,eff}$) is reduced without reducing the intrinsic transistor f_T. As we will see later this reduces the effect of induced gate noise which is a very welcome noise reduction. Furthermore the use of an external linear gate source capacitor may make the input matching network more linear because otherwise the intrinsic gate-source capacitance which is essentially a bias dependent varactor may lead to significant distortion at high input power levels.

Summing up there are three ways f_T can be reduced without changing the architecture: By increase the transistor length L, by using PMOS transistors in the input stage or by adding an external linear capacitor C_{gsx} between the gate and source terminals of the input transistor. Of these three the two latter appear most appealing though more knowledge on excess noise might change this picture. In any case using PMOS transistors alone will not be sufficient in most cases as this only reduces the current consumption by a factor of 3-4. Therefore an external gate-source capacitor seems to be needed if the transistor length is 0.25 micron or shorter regardless of the chosen transistor type (NMOS or PMOS).

Now let us see what value this external capacitor C_{gsx} should have. The intrinsic gate-source capacitance C_{gsm} is approximately given by (17) which simply expresses the parallel plate capacitance under the gate. C_{gsx} equals C_{gs} minus C_{gsm} so substituting C_{gs} according to (12) and C_{gsm} according to (17) yields an expression for the external gate source capacitance (18). Notice that in this equation I_D is a *chosen* bias current - not a necessary bias current. This is possible because the introduction of an external capacitor C_{gsx} decouples the binding between g_m and C_{gs}.

$$C_{gsm} = \frac{\varepsilon_r \varepsilon_0 L W}{t_{ox}} \tag{17}$$

$$C_{gsx} = C_{gs} - C_{gsm} = \frac{g_m}{\omega_0 I_D}\sqrt{2P_{in}/R_s} - \frac{\varepsilon_r \varepsilon_0 L W}{t_{ox}} \tag{18}$$

2.5 Mobility Degradation in Short Channel MOS Transistors

The design of low power LNAs that can tolerate input power levels in the 0dBm range automatically implies overdrive voltages of several hundred millivolts. At such bias levels deep submicron CMOS transistors develop a strong electric field between the gate and the channel. This makes the charge carriers flow in a narrower region below the silicon-oxide interface and therefore lowers the mobility of the device [11]. One way to model this effect is (19) where μ_0 is the “low-field” mobility and θ is a mobility degradation fitting parameter which is in the order of 1.0 - 1.5 V^{-1} for a 0.25 micron CMOS process.

$$\mu_{eff} = \frac{\mu_0}{1 + \theta(V_{GS} - V_T)} \tag{19}$$

The impact of reduced mobility at high gate bias levels is a departure from the simple square-law behavior. The resulting drain current is given by (20) which shows that the I-V relationship approaches a linear function at high bias levels. This is good news because it means that highly linear RF amplifiers can be built by simply biasing the transistors at high overdrive voltages.

$$I_D = \frac{\mu_n C_{ox} W (V_{GS} - V_T)^2}{2L} \cdot \frac{1}{1 + \theta(V_{GS} - V_T)} \tag{20}$$

$$V_{gs} = \frac{I_D \theta L}{\mu_n C_{ox} W}\left(1 + \sqrt{1 + \frac{2\mu_n C_{ox} W}{I_D \theta^2 L}}\right) + V_T \tag{21}$$

If the transistor width W and the bias current I_D have been chosen for a design then what is needed is an expression that gives the bias voltage. Such an expression is found by simply isolating V_{GS} in (20). This gives equation (21). As the drain current deviates from the square-law behavior so does its derivative - the transconductance. Equation (22) shows what g_m amounts to when mobility degradation is taken into account.

$$g_m = \frac{\partial I_D}{\partial V_{GS}} = \frac{\mu_n C_{ox} W (V_{gs} - V_T)}{L} \cdot \frac{1 + \theta(V_{gs} - V_T)/2}{(1 + \theta(V_{gs} - V_T))^2} = g_{d0} \cdot \alpha \tag{22}$$

$$\alpha \equiv \frac{g_m}{g_{d0}} = \frac{1 + \theta(V_{gs} - V_T)/2}{(1 + \theta(V_{gs} - V_T))^2} \tag{23}$$

The first term in (22) is the transconductance that corresponds to the square-law behavior. To avoid confusion this term can be called g_{d0} for the zero-bias drain conductance which is given by the same expression [8]. The second term then expresses the transconductance deviation which is caused by mobility degradation. It turns out that this is a useful parameter in noise analysis and therefore it has been given a name α as in [8] i.e. the transconductance deviation α is defined as (23).

2.6 Noise Modelling of the LNA with External Gate-Source Capacitance

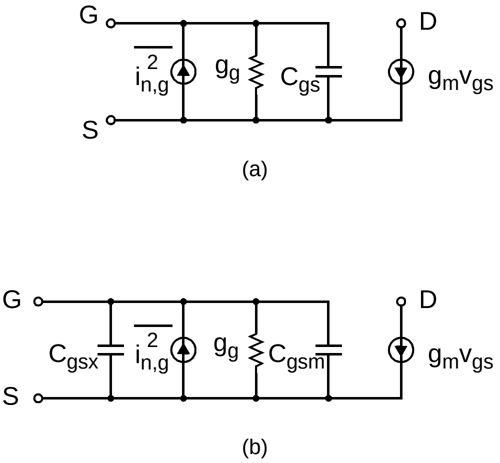
Noise in deep submicron CMOS transistors seems to be an eternal source of confusion. In the last couple of years a lot of research has been carried out in the field of noise characterization of short channel CMOS transistors but there is still no consensus as to how noise should be modelled in such devices and there does not even seem to be an understanding about which noise mechanisms are the dominating ones. Most recent treatments start out with a description of the long channel behaviour as described by Aldert Van der Ziel [17] including induced gate noise and then they diverge into different more or less clear “new” models. Typically such attempts try to explain excess noise observed in devices or circuit measurements. One popular explanation is that the excess noise is caused by hot-electron effects near the drain [8], but recent studies claim that this effect is less dominant [16]. This reference instead points to shot noise in the channel as a dominant noise source. Other references believe a lossy parasitic gate-bulk capacitance is to blame for the excess noise [14]. Still others just try to fit a model to measurements.

Though many of these references state that the excess noise will be worse with scaling due to stronger short channel effects the reported noise figures seem to suggest otherwise. Only a couple of years ago a 3 dB NF, CMOS LNA at 1 - 2 GHz was considered good but now several 1 dB NF LNAs in that range have been reported. This shows a remarkable improvement which can not be attributed to the increased f_T of the involved CMOS transistors alone. One explanation could be that the excess noise really comes from the distributed substrate effects i.e. from lossy drain-source junctions etc. This would explain why the noise is dropping so rapidly as the shrinking physical dimensions greatly improve the Q-values of junction capacitances. Whatever the explanation is it seems like the excess noise is lower than anticipated by many. For these reasons this work will only consider the basic noise model with gate induced noise as presented in [17] and [8] rather than resorting to doubtful guess works.

2.6.1 High Frequency Gate Model With External Gate-Source Capacitance

Let us take a look at the non-quasistatic gate model as presented in [8]. This model is shown in Figure 10 - (a). The conductance g_g models the real part of the gate impedance that originates in the distributed nature of the gate capacitance. According to the equation given this parameter is apparently frequency dependent. The more intuitive Thévenin equivalent network however reveals that this is only apparently because the real part in this form is independent of frequency. Likewise the gate noise current that would otherwise appear frequency dependent is also frequency independent i.e. white in nature when cast into the Thévenin equivalent form. When this form is chosen despite its less intuitive appearance it is because it leads to simpler calculation of the overall circuit noise performance.

$$\left(\frac{C_{gs}}{C_{gsm}}\right)^2 = \left(\frac{C_{gs}}{g_m} \cdot \frac{g_m}{C_{gsm}}\right)^2 = \left(\frac{f_T}{f_{T,\,eff}}\right)^2 \tag{24}$$



$$g_g = \omega^2 C_{gs}^2 / 5 g_{d0}$$

$$\begin{aligned}\overline{i_{n,g}^2} &= 4kT\delta g_g \Delta f \\ &= 4kT\delta\omega^2 C_{gs}^2 \Delta f / 5 g_{d0}\end{aligned}$$

$$g_g = \omega^2 C_{gsm}^2 / 5 g_{d0}$$

$$\overline{i_{n,g}^2} = 4kT\delta_{eff}\omega^2 C_{gs}^2 \Delta f / 5 g_{d0}$$

$$\delta_{eff} = \delta(C_{gsm}^2 / C_{gs}^2)$$

Figure 10. Gate circuit model including induced effects. Standard model as presented in [8] (a) and model with external gate-source capacitance added (b).

Before going into a detailed noise analysis of the LNA with its partially correlated noise sources [8] it is interesting to see how the external gate source capacitance affects the non-quasistatic gate model. This is shown in Figure 10 - (b) where $C_{gs} = C_{gsm} + C_{gsx}$ and where g_m is assumed identical to the g_m of Figure 10 - (a). The first thing we notice is that g_g is now $(C_{gs}/C_{gsm})^2$ times smaller. We can rewrite this expression as (24). From the discussion in Section 2.4 recall that in order to achieve high input power capability at reasonable power consumption is was necessary to reduce f_T by a factor of 5-10. This means that C_{gs}/C_{gsm} is approximately 5-10 and therefore g_g is reduced by 25-100 i.e. the real part of the gate input impedance which comes from distributed effects is also reduced by a factor of 25-100. In the section on input matching we ignored the distributed gate effect but actually the real part can be a quite significant part of source impedance e.g. 10-12 ohm for a 50 µm transistor [19] so usually this effect should be taken into account. In our case the external gate-source capacitor reduces this real part to a few hundred milli-ohms so we do not have to change the equations developed so far. The next thing we notice is that the induced gate noise is reduced by an equal amount i.e. by a factor of 25-100. This is good news for two reasons. First the noise figure is greatly improved by the external gate-source capacitor and second the gate noise contribution is now so low that we do not need to take it into account for the noise figure expression and this means a greatly simplified analysis which could result in more design insight.

2.6.2 Noise Figure Calculation

Figure 11 shows a small signal diagram which can be used for calculation of the LNAs noise figure. The capacitors C_{gsm} and C_{gsx} are shown as one capacitor C_{gs} to simplify the noise calculation. The first noise generator in the circuit is $v_{n,Rs}$ which models the thermal noise of the source.

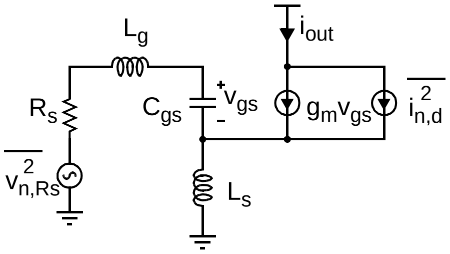


Figure 11. LNA small signal diagram with noise sources for calculation of the noise figure. The gate noise source is omitted because the external capacitor makes this contribution negligible.

It is equivalent to the noise of a resistor with a value of R_s i.e. it is given by (25). The next noise generator is the drain noise current of the input transistor $i_{n,d}$ which is given by (26), [8] where γ is the drain noise constant and g_{d0} is the zero bias drain conductance. See Section 2.5 for more detail. Notice that the gate noise generator has been omitted because the external gate-source capacitor makes the induced gate noise negligible as shown in Section 2.6.1.

$$\overline{v_{n,Rs}^2} = 4kTR_s\Delta f \tag{25}$$

$$\overline{i_{n,d}^2} = 4kT\gamma g_{d0}\Delta f \tag{26}$$

To be able to calculate the noise figure of this stage we can start by referring these two noise sources to the output. The output noise current which is caused by thermal noise in the source $i_{out,Rs}$ can be calculated as shown in (27) where it is assumed that the input is power matched. The output noise current which is caused by thermal noise in the channel of the input transistor can be calculated as shown in (28) where the expression in the parenthesis becomes 1/2 when the input is power matched.

$$i_{out,Rs} = g_m v_{gs,Rs} = g_m \frac{1}{sC_{gs}} i_{g,Rs} = g_m \frac{1}{sC_{gs}} \frac{v_{n,Rs}}{2R_s} = -j\frac{g_m}{\omega 2 R_s C_{gs}} v_{n,Rs} \tag{27}$$

$$i_{out,d} = \left(1 - \frac{1}{1 + \frac{R_s C_{gs}}{g_m L_s} + \frac{1 - \omega^2 (L_g + L_s) C_{gs}}{g_m L_s}}\right) i_{n,d} = \left(1 - \frac{1}{1 + 1 + \frac{0}{g_m L_s}}\right) i_{n,d} = \frac{i_{n,d}}{2} \tag{28}$$

Having referred the noise generators to the output it is now a simple matter to derive the corresponding output noise power spectral densities. The output noise power spectral density which arises due to the noise from the source is calculated as shown in (29). Likewise the output noise power spectral density which is caused by the channel noise in the input transistor is calculated as shown in (30) where g_{d0} is replaced by g_m/α as specified in Section 2.5.

$$S_{out,\,Rs}(\omega_0) = |H(\omega_0)|^2 S_{n,\,Rs}(\omega_0) = \frac{g_m^2}{\omega^2 4R_s^2 C_{gs}^2}\frac{\overline{v_{n,\,Rs}^2}}{\Delta f} = \frac{g_m^2}{\omega^2 4R_s^2 C_{gs}^2}4kTR_s = \frac{g_m^2 kT}{\omega^2 R_s C_{gs}^2} \tag{29}$$

$$S_{out,\,d}(\omega_0) = |H(\omega_0)|^2 S_{n,\,d}(\omega_0) = \frac{1}{4}\frac{\overline{i_{n,\,d}^2}}{\Delta f} = \frac{1}{4}4kT\gamma\frac{g_m}{\alpha} = \frac{kT\gamma g_m}{\alpha} \tag{30}$$

The noise figure measure (NF) has been defined in a number of different ways. One definition is the signal-to-noise power ratio at the input ($\mathrm{SNR_{in}}$) divided by the signal-to-noise power ratio at the output ($\mathrm{SNR_{out}}$), [12]. This emphasizes that the NF is a measure of the SNR degradation inflicted by the circuit. Another definition is given in [8] where the NF is defined as the total output noise divided by the output noise which is due to the source. Equation (31) shows that these definitions are identical. Reference [8] calls this definition the noise factor (F) when it is not given in dB. Other references define NF and F opposite so to avoid confusion this presentation uses the term noise figure (NF) whether (31) is given in dB or not.

$$\begin{aligned} NF \equiv \frac{SNR_{in}}{SNR_{out}} &= \frac{\text{Signal}/N_{in}}{G\cdot\text{Signal}/N_{out}} = \frac{N_{out}}{GN_{in}} = \frac{\text{Total output noise}}{\text{Output noise due to the source}} \\ &= \frac{GN_{in}+N_{circuit}}{GN_{in}} = 1+\frac{N_{circuit}}{GN_{in}} = 1+\frac{S_{out,\,circuit}(\omega)}{S_{out,\,source}(\omega)} \end{aligned} \tag{31}$$

With this definition it is straightforward to calculate the noise figure of the inductive source-degeneration LNA with external gate source capacitor. Equations (29) and (30) are simply substituted into equation (31). Then after simplification C_{gs} is substituted by expression (12) which was found in Section 2.3. This leads to a very simple expression for the noise figure of this circuit which is highly convenient for extracting design insight.

$$\begin{aligned} NF &= 1+\frac{S_{out,\,d}(\omega_0)}{S_{out,\,Rs}(\omega_0)} = 1+\frac{\dfrac{kT\gamma g_m}{\alpha}}{\dfrac{g_m^2 kT}{\omega_0^2 R_s C_{gs}^2}} = 1+\frac{\gamma\omega_0^2 R_s C_{gs}^2}{\alpha g_m} = 1+\frac{\gamma\omega_0^2 R_s}{\alpha g_m}\left(\frac{g_m}{\omega_0 I_D}\sqrt{\frac{2P_{in}}{R_s}}\right)^2 \\ &= 1+\frac{\gamma g_m 2P_{in}}{\alpha I_D^2} \end{aligned} \tag{32}$$

2.6.3 Noise Figure Optimization

Now that we have an expression for the noise figure of the LNA it is possible to consider what should be done to minimize the noise figure. If we consider L_s, L_g, C_{gsx}, V_{gs} and W as being the available design parameters then there is one degree of freedom left in the design. One could say that L_s is chosen to provide the desired real part of the input impedance (6) and that L_g is chosen to null the imaginary part of the input impedance (7). Likewise one could say that C_{gsx} is chosen such that the circuit can maintain class-A operation at a desired maximum input power level (11), (18) and that V_{gs} is chosen such that the desired bias current is achieved (21). This leaves only one parameter - the transistor width W - which can be used for noise optimization.

If we look at (32) then γ, I_D and P_{in} can be viewed as constants so it is actually only the g_m/α ratio which can be used for noise optimization. When the input transistor is made narrower i.e. when W is reduced then g_m drops. In order to maintain the same bias current it is necessary to increase V_{gs} and this lowers α. However α decreases slower than g_m so overall the g_m/α ratio is reduced by lowering the transistor width W. Therefore W should be minimized to achieve the best noise figure. This is a direct result of using an external gate-source capacitor because it makes the induced gate noise negligible.

According to (32) it should then be possible to achieve a 0 dB noise figure by making W infinitely narrow. Clearly this is not the case because this would demand an infinitely high gate-source voltage an no practical circuit can tolerate an infinitely high gate source voltage. What this boils down to is that the gate-source voltage should be maximized to minimize the noise figure.

There can be several issues that limit the gate source voltage. The process itself has some limitations on how large a gate source voltage the thin oxide can withstand. For a 0.25 micron CMOS process this is typically 2.5 volts. So in order not to exceed this voltage the gate-source bias voltage is limited to approximately 1.5 volts because at this level the gate voltage swing is in the order of 1 volt. In a slightly more formalized manner we can say that the maximum gate-source bias voltage $V_{GS,max} = (V_{dd,max}+V_T)/2$. Then there are circuit level trade-offs which can limit the gate-source bias voltage. For instance if a cascode configuration is used then it can be difficult to support such a high gate-source voltage and still keep all transistors operating in saturation while providing a reasonable power gain. Typically this will give a practical maximum V_{GS} that is several hundred millivolt lower than the constraint set by the process. Then there is the not-well-understood excess noise which is believed to increase with gate-source voltage due to hot-electron effects. Until a better understanding of this phenomenon is gained it is wise to be a little conservative with regards to the chosen V_{GS}. Therefore at the moment practical maximum V_{GS} values are in the 1.0 -1.2 volt range for a 0.25 micron CMOS technology.

Once a V_{GS} value has been chosen it is possible to derive an expression (33) for the transistor width. This is done by simply isolating W in (20). Thereby the last design parameter has been fixed and we can cast the design equations into their final form.

$$W = \frac{2I_D L_{eff}}{\mu_n C_{ox}} \cdot \frac{1 + \theta(V_{GS} - V_T)}{(V_{GS} - V_T)^2} \tag{33}$$

Both C_{gsx} and L_g depend on g_m so let us first see what this parameter amounts to with the transistor width given by (33). By substituting (33) into (22) it is possible to show that g_m can be expressed as (34).

$$g_m = \frac{I_D}{(V_{GS}-V_T)}\left(1+\frac{1}{1+\theta(V_{GS}-V_T)}\right) \tag{34}$$

This equation can then be substituted into the expression for C_{gsx} (18) and into the expression for L_g (14) and the following design equations are found:

$$C_{gsx} = \frac{\sqrt{2P_{in}/R_s}}{\omega_0(V_{GS}-V_T)}\left(1+\frac{1}{1+\theta(V_{GS}-V_T)}\right)-\frac{2I_D L^2(1+\theta(V_{GS}-V_T))}{\mu_n(V_{GS}-V_T)^2} \tag{35}$$

$$L_g = \frac{\sqrt{R_s}}{\omega_0}\left(\frac{(V_{GS}-V_T)}{\sqrt{2P_{in}}\left(1+\frac{1}{1+\theta(V_{GS}-V_T)}\right)}-\frac{\sqrt{2P_{in}}}{I_D}\right) \tag{36}$$

The noise figure equation (32) which incorporates g_m/α can be expressed such that it becomes a function of W:

$$\begin{aligned} NF &= 1+\frac{\gamma g_m 2P_{in}}{\alpha I_D^2} = 1+\frac{\gamma g_{d0} 2P_{in}}{I_D^2} = 1+\frac{\gamma \mu_n C_{ox} W(V_{GS}-V_T)2P_{in}}{L I_D^2} \\ &= 1+\gamma\theta\left(1+\sqrt{1+\frac{2\mu_n C_{ox}W}{I_D\theta^2 L}}\right)\frac{2P_{in}}{I_D} \end{aligned} \tag{37}$$

And then equation (33) can be substituted into this expression which after simplification becomes (38). This last equation expresses the minimum noise figure which is achieved when the transistor width W is chosen according to (33). This expression reveals plenty of insight: For a given maximum input power level (P_{in}) it is only possible to lower the noise figure by increasing the bias current (I_D) or by increasing the gate-source bias voltage (V_{GS}). It is perhaps also surprising that the NF of the input power constrained LNA is independent of frequency (ω) and independent of the source impedance (R_s).

$$NF_{min} = 1+\frac{4\gamma P_{in}}{I_D}\left(\frac{\theta}{2}+\sqrt{\frac{\theta^2}{4}+\frac{1+\theta(V_{GS}-V_T)}{(V_{GS}-V_T)^2}}\right) \tag{38}$$

2.7 LNA Design Example

A small example is given here to illustrate the ease with which the design equations can be used. First assume that the system requirements are: f_0 = 2.0 GHz, R_s = 50 ohms, P_{in} = -3 dBm, I_D = 8 mA and that the used IC process is a 0.25 micron CMOS process with the following parameters: V_T = 0.52 V, θ = 1.3 V^{-1}, γ = 2/3, $\mu_n C_{ox}$ = 250 $\mu A/V^2$ and μ_n = 0.0353 m^2/Vs. The maximum gate-source and drain-source bias voltages are assumed to be 1.0 volt to maintain compatibility with the cascode topology. Therefore the input transistor is biased at V_{GS} = V_{DS} = 1.0 V and it is further assumed that the cascode transistor is chosen wide enough to have a sufficiently low input impedance.

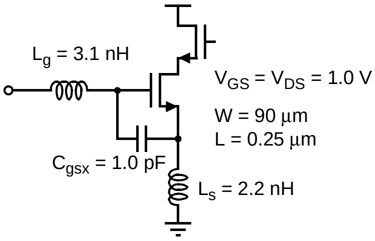


Figure 12. LNA Design Example. Input stage with low power consumption (8 mA), 50 Ω input impedance, very high linearity (-3 dBm input power in class A) and low noise (2 dB NF).

It is now possible to calculate all the necessary device parameters. First we can use (13) to calculate the source inductance (39). Then (36) can be used to calculate the gate inductance (40) and (35) can be used to calculate the external gate-source capacitance (41). Finally (33) can be used to calculate the input transistor width (42).

$$L_s = \frac{\sqrt{2 \cdot 50 \cdot 0.5\times10^{-3}}}{2\pi 2\times10^{9} \cdot 8\times10^{-3}} = 2.2 \text{ nH} \tag{39}$$

$$L_g = \frac{\sqrt{50}}{2\pi 2\times10^{9}}\left(\frac{1.0-0.52}{\sqrt{2 \cdot 0.5\times10^{-3}}\left(1+\frac{1}{1+1.3(1.0-0.52)}\right)} - \frac{\sqrt{2 \cdot 0.5\times10^{-3}}}{8\times10^{-3}}\right) = 3.1 \text{ nH} \tag{40}$$

$$\begin{aligned} C_{gsx} &= \frac{\sqrt{2 \cdot 0.0005}\left(1+\frac{1}{1+1.3 \cdot 0.48}\right)}{\sqrt{50}\,2\pi 2\times10^{9}\,0.48} - \frac{2 \cdot 8\times10^{-3}\,0.25^{2}\times10^{-12}(1+1.3 \cdot 0.48)}{0.0353(0.48)^{2}} \\ &= 1.0 \text{ pF} \end{aligned} \tag{41}$$

$$W = \frac{2 \cdot 8\times10^{-3}\,0.2\times10^{-6}(1+1.3(1.0-0.52))}{250\times10^{-6}(1.0-0.52)^{2}} = 90\ \mu\text{m} \tag{42}$$

Figure 12 shows the resulting circuit with component values. Details regarding power gain and output matching are not given as they can be designed separately by adjusting the supply voltage and the load impedance. Now let us see what noise figure we can achieve with this circuit. This parameter can be calculated with equation (38) i.e. the circuits NF amounts to approximately 2 dB (43). Naturally in a real circuit there will be other noise sources and the input transistor may experience some excess noise. If the effective γ turns out to be twice the long channel value 2/3 then the noise figure becomes 3.3 dB which is still very good considering the high linearity and low bias current.

$$NF = 1 + \frac{4\frac{2}{3}0.5\times10^{-3}}{8\times10^{-3}}\left(\frac{1.3}{2} + \sqrt{\frac{1.3^{2}}{4} + \frac{1+1.3(1.0-0.52)}{(1.0-0.52)^{2}}}\right) = 1.56 = 1.9 \text{ dB} \tag{43}$$

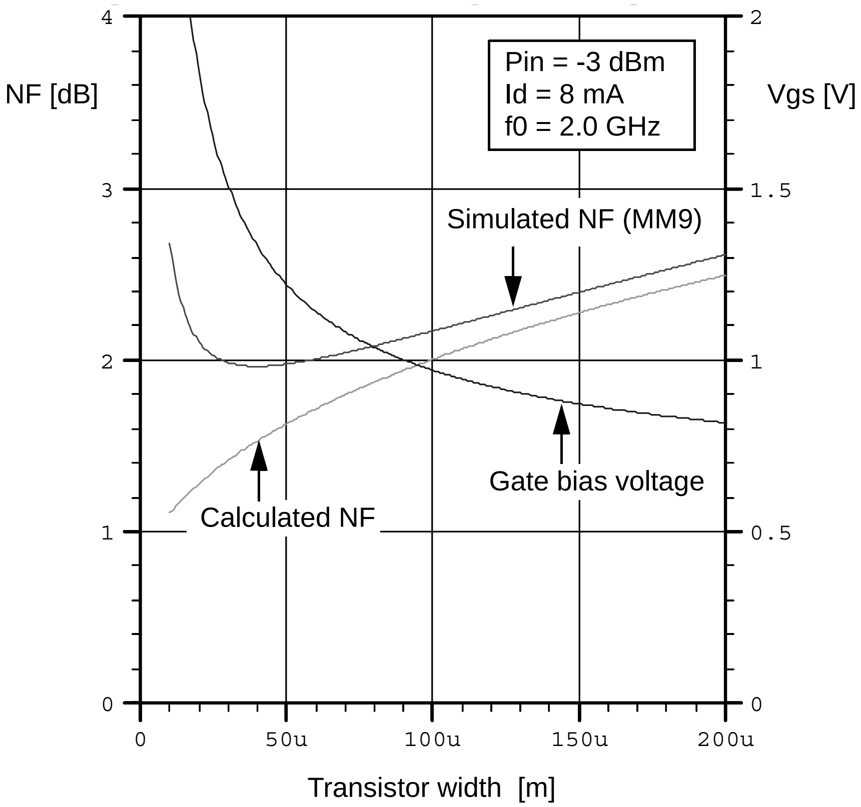


Figure 13. Simulated and calculated LNA performance. Noise figure and gate bias voltage versus transistor width W. All component values are recalculated for each data point.

A number of simulations have been made to evaluate the quality of the circuit model developed here. The simulations were made with the APLAC circuit simulator and the used transistor model is MOS Model 9 which contrary to many other transistor models does take induced gate noise into account. Harmonic balance methods were used for the simulations of large signal linearity i.e. input compression point.

Figure 13 shows the calculated and simulated noise figure of the circuit not only with the chosen gate-source voltage i.e. with the calculated transistor width but also for other transistor widths ranging from 10 µm to 200 µm. As expected the noise figure drops when the transistor width is reduced. However at very short transistor widths and high gate-source voltages the simulated noise figure starts to increase again. This may be due to inaccuracies in the model or perhaps due to incomplete parameter fitting. On the other hand for more practical transistor widths (V_{GS} < 1.5 volt) the calculated noise model shows a surprisingly good correspondence with the simulated values.

Table 1: Calculated and simulated performance of the LNA from Figure 12

Parameter	Calculated	Simulated (MM9)
NF	1.9 dB	2.1 dB
ICP	---	+ 3.0 dBm
Gain Compression (P_{in} = -3 dBm)	---	0.25 dB
Input Match (S_{11})	Ideal match	- 18 dB
Drain Current	8.0 mA	8.3 mA

The simulated linearity, input match and current consumption also match closely the design targets. The specific performance for the LNA design example (Figure 12) is summarized in Table 1. The good correspondence confirms that the developed LNA design methodology has a sufficient accuracy to be useful for finding initial component values in an input power constrained LNA design. Second order effects like inductor loss and channel length modulation can then be compensated for by adjusting these calculated component values.

2.8 Other Considerations

In the previous sections only the necessary design considerations have been treated. This does not mean that other considerations are not important but it means that they were not necessary for the derivation of the design equations. Having completed this task it is now time to look at some of the remaining design issues.

2.8.1 Source Impedance

So far the source impedance R_s has been considered a fixed parameter that could not be used for circuit optimization. Typically off-chip filters or baluns need to be terminated in a specific impedance but off-chip matching networks can transform this impedance up or down at the expense of a narrower operating bandwidth. This means that the source impedance seen from the IC (R_s) to some extent can be used to improve the circuit performance.

According to (32) the noise figure of an input power constrained LNA is independent of the source impedance. APLAC simulations have been made to check the validity of this statement when more complete transistor models (MOS Model 9) are used. It turns out that at the center frequency this is the case. Simulations also showed that the noise figure some distance away from the center frequency does depend on R_s. Typically the noise figure remains low at a wider bandwidth if R_s is maximized. Maximizing R_s also makes it easier to design PWB, bondwire and on-chip interconnect such that the series loss is negligible. Otherwise R_s can be used to adjust the values of the on-chip matching components such that they become more practical for integration. Fortunately these component values are almost always practical at the lower GHz frequencies so significant adjustment of R_s is rarely needed. However adjustment of R_s is a degree of freedom in the design that is worth keeping in mind.

2.8.2 Ohmic Gate Resistance

The polysilicon material which is used for the gate of MOS transistors usually has a relatively high resistivity. This means that any MOS transistor will have a distributed gate resistance that comes from the loss of the polysilicon material. The input transistor of an LNA is especially sensitive to this effect because the gate resistance leads to extra noise and loss of signal power before amplification. Therefore it is important to model the ohmic gate resistance. The gate resistance R_g can be expressed as (44), [20] where $R_{\square\text{poly}}$ is the sheet resistance of the polysilicon gate material, n is the number of transistor fingers, L is the length of the transistor and W is the width of the transistor.

$$R_g = \frac{R_{\square\text{poly}} W}{3n^2 L} \tag{44}$$

From this equation it is apparent that the ohmic gate resistance can be reduced by simply increasing the number of transistor fingers. The noise figure of an LNA typically becomes independent of the ohmic gate resistance when it is less than one percent of the source impedance so we can make the gate resistance negligible by laying the input transistor out with the number of transistor fingers that makes R_g equal to 1% of the source impedance. An expression for this number is found by isolating n in (44) and substituting R_g with $R_s/100$. Doing so yields (45). In some cases this leads to slightly more fingers than what is practical in a layout and if this is the case then the number of fingers can be cut in half by connecting both ends of the polysilicon fingers.

$$n = 10\sqrt{\frac{R_{\square\text{poly}} W}{3R_s L}} \tag{45}$$

2.8.3 Cascoding

For the design methodology which was developed in this chapter it has been assumed that the LNA should use a cascode topology. This is by no means a coincidence. First of all most recent wireless products are required to have very low radiation from the receiver through the antenna. This translates to fairly strict requirements on the reverse isolation of the LNA because otherwise too much LO signal may leak to the antenna. The cascode structure is a very effective way of improving the reverse isolation of the LNA. At the same time it mitigates the Miller capacitance and reduces the effect of channel length modulation by presenting a low impedance to the drain of the input transistor.

There are no exact design guidelines as to what the width of the cascode transistor should be but it should be wide enough to reduce the voltage swing at the drain of the input transistor. On the other hand it should not be too wide either because this increases the noise contribution of the cascode transistor. It can be shown that if there is no parasitic capacitance at the node between the two transistors then the cascode transistor can not contribute with noise. This means that the noise of the cascode transistor can be removed by resonating this capacitance out with an inductor. This is rarely done because an on-chip inductor consumes area and contributes with thermal noise of its own. It is however important to keep this option in mind if a very wide cascode transistor is used.

2.8.4 Power Gain, Supply Voltage and Load Impedance

The issues of power gain, supply voltage and load impedance were not covered in the developed input power constrained LNA design methodology. This is because when a cascode structure is used it is possible to achieve a very high output impedance and great reverse isolation. This means that the LNA noise figure, input match and power to current gain behavior becomes largely independent of the supply voltage and load impedance.

As a result it is possible to design the LNA as specified in this chapter and then consider its output as an ideal current source. Naturally the load impedance should be chosen small enough to make sure the cascode transistor stays in the saturation region but that is straightforward when the peak current swing is known (I_D) i.e. the load impedance should not exceed $V_{out,peak}$ / I_D at the frequency where the maximum current swing is invoked.

2.8.5 Differential Operation

Everything in this chapter has been presented in a single ended form. This is because it is easier to model the input power constrained LNA design in a single ended form. Sometimes a differential LNA topology is necessary for instance to make sure it can coexist with a larger RF system. If a differential LNA is needed then it should be designed as follows:

Consider one half circuit of the LNA and calculate the component values as specified in this chapter but this time use $R_s/2$, $P_{in}/2$ and $I_D/2$ instead of R_s, P_{in} and I_D. This will give the component values that are needed in one half of the LNA and for symmetry reasons the components which are used in the other half are chosen to be identical.

Notice that the differential LNA according to equation (38) will have exactly the same noise figure as the single ended version because P_{in}/I_D is not changed.

2.9 Summary

This chapter started with a short presentation of classic LNA topologies and the inductive source degeneration topology was chosen as a basis for this work. Then the input matching constraint was formulated. This lead to the identification of two design equations - one ensuring that the real part of the input impedance equals the source impedance (6) and one that nulls the imaginary part of the input impedance (7). Next the ability to handle very strong out-of-band blocking signals was translated to a requirement of Class-A operation. This requirement and the desire to have high power efficiency lead to another design equation (11).

Based on these equations and knowledge of the f_T of the IC process it was possible to derive an expression for the needed bias current (16). This expression revealed the necessity to lower the effective f_T which in turn lead to the addition of an external gate-source capacitance in order to get an extra degree of freedom in the design. Next some basic MOS transistor equations were formulated that take mobility degradation into account. These transistor equations together with the power dissipation constraint i.e. the maximum drain current constraint lead to one more design equation (20).

The following section modeled the noise of the LNA. It was explained that the external gate source capacitance makes the induced gate noise negligible and therefore can be excluded from the noise model. Then the noise figure of the LNA was calculated and the conclusion was that the transconductance of the input transistor should be minimized. This means that the transistor width should be minimized. The lower limit is set by the maximum gate-source bias voltage which can be implemented in the design taking practical considerations into account. Essentially this means that another design equation ($V_{GS} = V_{GS,max}$) has been identified.

Thereby we have formulated the five design equations that are necessary to derive closed form expressions for the five degrees of freedom there are in the design i.e. the source inductance, the gate inductance, the external gate-source capacitance, the input transistor width and the gate-source bias voltage. Closed form expressions for all design parameters were then derived and an example was given of how the design expressions can be used. The resulting circuit was then simulated and the results reveal good performance and good correlation with simulations.

Then some remaining design considerations were covered. Amongst others a design expression was derived that identifies the number of transistor fingers which guarantees that the distributed ohmic gate resistance is negligible.

As a final note one should remember that this design methodology is developed for the situation where the LNA is required to have high input power handling capability and low power consumption. This is the case if the desired current consumption is at least 3-4 time lower than the value given by equation (16).

Table 2: Input Power Constrained LNA Design Parameters and Achievable Noise Figure.

Parameter	Expression or Trade-Off	Described
V_{GS}	The NF is reduced by maximizing this parameter	Section 2.6.3
I_D	The NF is reduced by maximizing this parameter	Section 2.6.3
L_s	$\frac{\sqrt{2R_sP_{in}}}{\omega_0 I_D}$	Section 2.3
L_g	$\frac{\sqrt{R_s}}{\omega_0}\left(\frac{(V_{GS}-V_T)}{\sqrt{2P_{in}}\left(1+\frac{1}{1+\theta(V_{GS}-V_T)}\right)}-\frac{\sqrt{2P_{in}}}{I_D}\right)$	Section 2.6.3
C_{gsx}	$\frac{\sqrt{2P_{in}/R_s}}{\omega_0(V_{GS}-V_T)}\left(1+\frac{1}{1+\theta(V_{GS}-V_T)}\right)-\frac{2I_DL^2(1+\theta(V_{GS}-V_T))}{\mu_n(V_{GS}-V_T)^2}$	Section 2.6.3
W	$\frac{2I_DL(1+\theta(V_{GS}-V_T))}{\mu_nC_{ox}(V_{GS}-V_T)^2}$	Section 2.6.3
n	$10\sqrt{\frac{R_{\square\mathrm{poly}}W}{3R_sL}}$	Section 2.8.2
R_s	Can be used to adjust bandwidth or to achieve practical passive component values.	Section 2.8.1
NF	$1+\frac{4\gamma P_{in}}{I_D}\left(\frac{\theta}{2}+\sqrt{\frac{\theta^2}{4}+\frac{1+\theta(V_{GS}-V_T)}{(V_{GS}-V_T)^2}}\right)$	Section 2.6.3

If this is not the case then the noise figure estimate will be less accurate and the design may not give the lowest possible noise figure. If the value given by equation (16) is in the same order or lower than the amount of bias current that can be tolerated in the design then the methodology presented in [8] should be used instead of the one presented here.

This completes the developed input power constrained LNA design methodology which is the first LNA design methodology to simultaneously take power dissipation, input power handling capability and noise figure into account. The developed closed form expressions and the most essential design trade-offs are summarized in Table 2.

2.10 References

[8] D. Shaeffer and T. Lee, “A 1.5-V, 1.5 GHz CMOS Low Noise Amplifier,” *IEEE J. Solid-State Circuits*, vol. 32, May 1997.

[9] T. Lee, “The design of CMOS Radio-Frequency Integrated Circuits”, *Cambridge University Press*, 1998.

[10] P. Leroux et. al., “A 0.8dB NF ESD-protected 9mW CMOS LNA”, *In Proceedings of the Int. Solid-State Circuits Conference 2001*, pp. 410-411, 2001.

[11] B. Razavi, “Design of Analog CMOS Integrated Circuits”, *McGraw-Hill*, 2001.

[12] B. Razavi, “RF Microelectronics”, *Prentice-Hall*, 1998.

[13] R.R.J. Vanoppen et. al., “RF Noise Modelling of 0.25 μm CMOS and Low Power LNAs”, *Tech. Dig. IEDM 1997*, pp. 317-320.

[14] S.V. Kishore et. al., “Substrate-Induced High-Frequency Noise in Deep Sub-Micron MOSFETs for RF Applications”, *Custom Integrated Circuits Conference*, 1999, pp. 365-368.

[15] T. Manku, “The Impact of Induced Gate Noise When Simultaneously Power and Conjugate Noise-Matching MOS Transistors”, *IEEE Transactions on Circuits and Systems-II*, June 1999, Vol. 46, No. 6, pp. 842-844.

[16] T. Manku et. al., “RF Simulation and Physics of the Channel Noise Parameters within MOS Transistors”, *Custom Integrated Circuits Conference*, 1999, pp. 369-372.

[17] A. Van der Ziel, “Noise in Solid State Devices and Circuits”, *Wiley*, 1986.

[18] G. Knoblinger, et. al., “A new Model for Thermal Channel Noise of Deep Submicron MOSFETS and its Application in RF-CMOS Design”, *In Symp. VLSI Circuits Digest*, 2000, pp. 150-153.

[19] H-Y. Tsui and J. Lau, “SPICE Simulation and Trade-offs of CMOS LNA Performance with Source-Degeneration Inductor”, *IEEE Trans. on Circuits and Systems-II*, January 2000, Vol. 47, No. 1, pp. 62-65.

[20] B. Razavi et. al., “Impact of Distributed Gate Resistance on the Performance of MOS Devices”, *IEEE Transactions on Circuits and Systems-I*, November 1994, Vol. 41, No. 11, pp. 750-754.

Chapter 3: Resonator Design

Resonators are some of the most commonly used components in RF circuits. Almost every receiver, transmitter or frequency synthesizer uses at least one resonator. Resonators are used in oscillators to generate a reference frequency, in filters they are used to pass the frequency interval of interest from the source to the load while attenuating all other frequencies. Between consecutive blocks in a signal processing chain they are used to provide impedance matching to ensure maximum power transfer and in amplifiers or buffers they are used to increase narrowband gain or they are used to reduce power consumption by cancelling out parasitic loads. These many uses illustrate just how useful resonators are in radio frequency circuits. In this presentation the focus is on the frequency selective nature of resonators because it is the most relevant for a treatment on RF filtering. The chapter starts with a description of the most important properties of resonators used for filtering in integrated RF circuits. More specifically transfer functions, selectivity, Q-values, coupled resonators and Q-enhancement techniques are reviewed. Passive components with low loss i.e. high-Q inductors and capacitors are of utmost importance for the performance of integrated filters, especially because these components are relatively poor compared to their discrete counterparts. Therefore the rest of this chapter is dedicated to the design of good inductors and capacitors in standard CMOS processes. First it is explained in great detail how high-Q capacitors with low process variation can be designed in a process without special capacitor options and then it is explained how useful inductors can be made in a CMOS process with low substrate resistivity and no thick top metal layers.

3.1 Properties of LC Resonators

In this first section it will be explained how the parallel LC resonator behaves over frequency, what Q-value is needed to achieve a certain attenuation at a specified offset from the center frequency, how single resonators can be coupled passively and actively and then it is explained how Q-enhancement techniques can be used to increase the effective Q-value of a resonator. Finally it is explained how tapped inductors and capacitive division can be used to reduce the loading of the resonator Q from the source and the load while simultaneously easing the matching conditions.

3.1.1 Parallel LC Tank Impedance

A parallel LC tank is simply an inductor (L) and a capacitor (C) coupled in parallel. The tank behavior is strongly dependent on how lossy these components are so it is necessary to model these losses. Over a relatively narrow bandwidth the loss of the inductor and the capacitor can be modelled by a single parallel resistor (R) which leads to the model shown in figure 14.

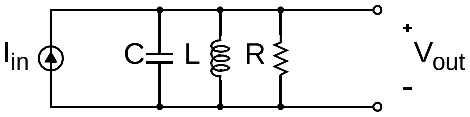


Figure 14. A parallel LC resonator driven by a current source.

The parallel LC resonator behaves in the following manner. At low frequencies the inductor has a very low impedance and at high frequencies the capacitor has a very low impedance. This means that the impedance of the LC resonator approaches zero at low and high frequencies. The impedance of L and C in parallel is 1/j(ωC - 1/ωL). It can be seen from this expression that at one particular frequency the denominator becomes zero i.e. the effects of the inductor and the capacitor cancel and leave the resonator impedance real and equal to R. This frequency (ω_0) which is called the resonance frequency is very important because it marks the frequency where the parallel LC resonator has the highest impedance and therefore it sets the center frequency in a bandpass filter or the oscillation frequency in an oscillator.

$$f_0 = \frac{\omega_0}{2\pi} = \frac{1}{2\pi\sqrt{LC}} = \text{resonance frequency} \tag{46}$$

$$Q = \frac{R}{\omega_0 L} = \omega_0 RC = \frac{R}{\sqrt{L/C}} = \frac{f_0}{BW_{3\text{dB}}} \qquad \text{(parallel RLC tank at resonance)} \tag{47}$$

Another important descriptive parameter for the parallel LC tank is the quality factor or simply the Q-value (47). The Q-value is an inverse measure of how lossy the resonator is i.e. how much power a resonator can store compared to the amount of power that it dissipates. It turns out that the Q-value is equal to the ratio of the center frequency and the -3 dB bandwidth so 1/Q directly gives the relative bandwidth of a bandpass filter made with a single resonator. Finally the characteristic impedance ($=\sqrt{L/C}$) is a measure of the magnitude of the capacitive and inductive reactances at the resonance frequency. With these three measures f_0, Q and $\sqrt{L/C}$ it is possible to completely characterize a parallel LC resonator in terms of center frequency (f_0), bandwidth (f_0/Q) and impedance at resonance ($Q \cdot \sqrt{L/C}$).

For filtering purposes it is often desired to know the attenuation at frequencies beyond the -3 dB frequencies i.e. it is useful to have a measure of the attenuation of the tank impedance over a wider frequency range. Such a measure can be a normalized tank impedance $|Z_{\text{tank}}|/\sqrt{L/C}$ which removes the dependence on the characteristic impedance. In this presentation it was chosen to label this measure ξ and as it can be seen from (48) it is possible to express ξ as a function of Q and normalized frequency f/f_0.

$$\xi = \frac{|Z_{\text{tank}}|}{\sqrt{L/C}} = \frac{Q}{\sqrt{1 + Q^2\left(\frac{f}{f_0} - \frac{f_0}{f}\right)^2}} = \text{normalized tank impedance} \tag{48}$$

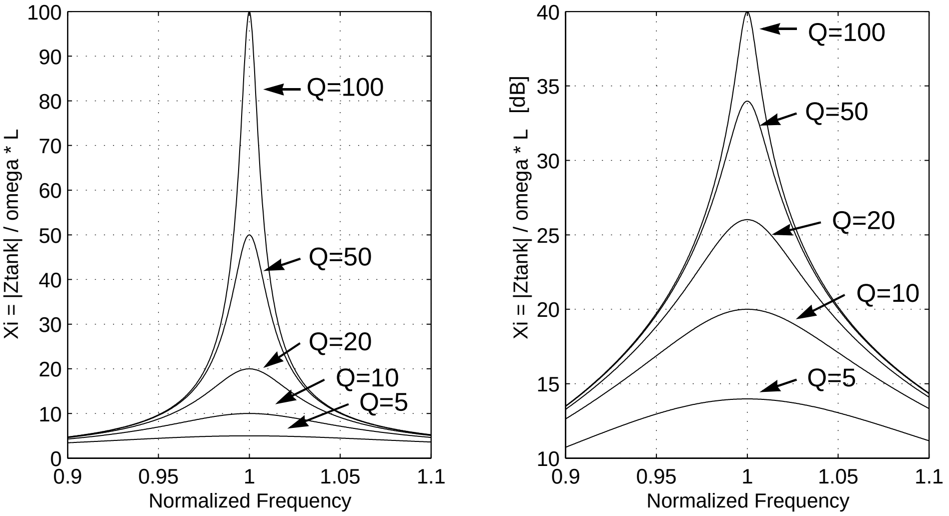


Figure 15. Normalized Resonator Impedance $\xi = |Z_{\text{tank}}|/\sqrt{L/C}$ vs. Normalized Frequency f/ f_0. Shows ξ for several resonator Q-values (left) and ξ in decibel for the same Q-values (right).

The value of ξ for different Q-values is shown in Figure 15. In a bandpass filter using a single resonator the relative attenuation at *f* is simply $\xi(f_0)/\xi(f)$. By solving for Q we get (49) which expresses the minimum resonator Q-value that gives the desired attenuation at a given frequency *f* located some distance away from the passband center frequency f_0.

$$Q_{min} = \frac{\sqrt{1-k^2}}{k\left|\frac{f}{f_0} - \frac{f_0}{f}\right|} \qquad k = 10^{-(\text{attenuation in dB}/20)} \tag{49}$$

3.1.2 Coupled Resonators

When (49) is used it quickly becomes apparent that in most radio frequency applications of interest Q-values of 50 - 100 or more are needed to provide useful filtering. Unfortunately today LC resonators that can be integrated in RF ICs have Q-values that are an order of magnitude lower and therefore it seems like we have a problem. However the problem of insufficient resonator Q-value caused by lossy passive components is by no means new. In the early days of radio i.e. in the 1920's people faced exactly the same problems. One classic approach to increasing the attenuation of a filter has been to increase the order of the resonator or the frequency selective network by coupling several resonators and thereby increasing the number of poles in the passive network. Figure 16 shows three different ways of coupling two resonators with passive components a) with a capacitor, b) with an inductor and c) with a transformer coupling.

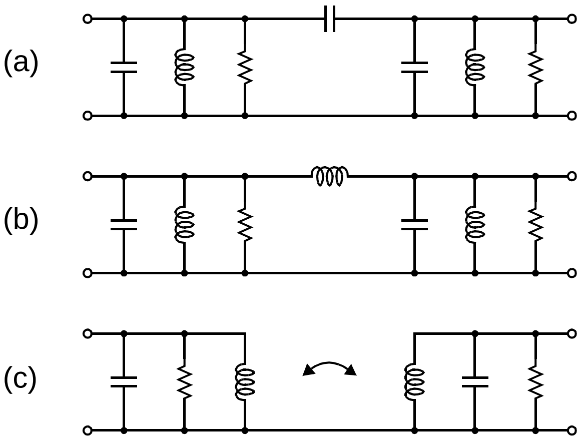


Figure 16. Passive Resonator Coupling. Capacitive (a), inductive (b) and transformer (c).

The first two coupled resonator topologies provide asymmetric filtering as a capacitor will provide additional attenuation at low frequencies while the inductor provides additional rejection at high frequencies. Common to all three coupled resonator topologies is the fact that their transfer functions are highly dependent on the exact value of the coupling component. If the coupling which by the way should be weak is too weak the passband is attenuated and if it is too high, then the transfer function will have a tendency to get twin peaking or a significant ripple in the pass band. Furthermore the close-in bandwidth actually increases when coupled resonators are used while the far-away attenuation is increased. This might be a desirable option when high-Q passive components are available but in integrated circuits this is not the case so it seems that passively coupled resonators are not very attractive in integrated circuits for narrow band filtering. However coupled resonators could be useful for image rejection or when better inductors become available. As a final remark on passively coupled resonators it can be said that two resonators can also be coupled by a third resonator - this is simply a bandpass version of a three element ladder filter and in this case it is also true that for narrow bandwidths it is really important to have passive components with low loss.

An alternative to the passive resonator coupling is the active resonator coupling (figure 17) which exploits the fact that two resonators can be coupled without affecting each others impedances. Expressed in different words individual resonators can be cascaded and if they are separated by transconductance stages their frequency selective nature will not be affected. This means that at the expense of higher power consumption the needed attenuation can be handled in e.g. three steps of 7 dB instead of one step of 21 dB. This strongly reduces the need for high-Q resonators but the needed Q-values are typically still much higher than what current silicon processes can provide, but it is a step in the right direction.

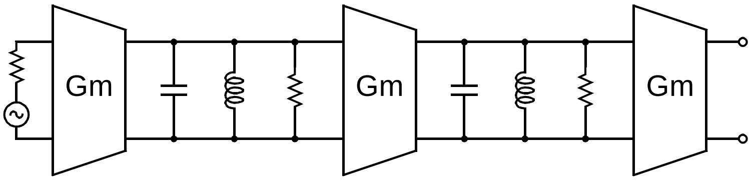


Figure 17. Active Resonator Coupling.

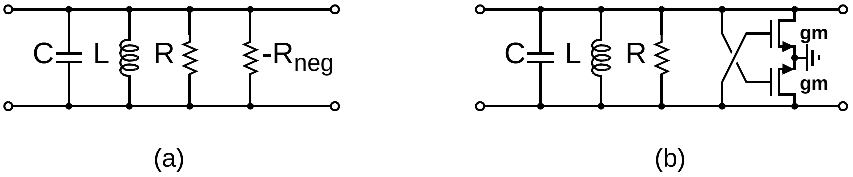


Figure 18. Resonator Q-enhancement with negative resistance circuit. Topology (a) and simplified implementation in a differential configuration (b).

3.1.3 Loss Compensation

Another option for increasing the effective resonator Q-value is to use active circuits to provide positive feedback. This technique is called Q-enhancement and the circuit element that presents a negative impedance to the resonator is called a negative resistance circuit (figure 18 - a). It is certainly possible to achieve very high effective Q-values with this technique. In fact when $R_{neg} = R$ the effective tank Q is infinite and if $R_{neg} < R$ the network is an oscillator. This highlights some of the problems using this technique. If the exact value of R_{neg} is not accurately controlled the filter is likely to become unstable. This effect becomes more and more pronounced as R_{neg} approaches R i.e. as the effective Q approaches infinity and therefore only a moderate amount of Q-enhancement can be handled without jeopardizing stability. Another concern is that the noise in the resonator rises quickly with the Q-enhancement factor (k_{QE}) which simply reflects the ratio of the enhanced Q-value and the uncompensated tank Q-value. Again moderate values of k_{QE} may lead to acceptable noise performance.

$$Q_{\text{enhanced}} = k_{\text{QE}} \cdot Q_{\text{tank}} \tag{50}$$

$$gm = \frac{2(1 - 1/k_{\text{QE}})}{Q_{\text{tank}}\sqrt{L/C}} \tag{51}$$

The simplest method for generating a negative resistance in a balanced configuration is to use a cross coupled transistor stage (figure 18 - b). Even though this is a simplified diagram it enables a calculation of the necessary transistor transconductance (gm) to provide a desired amount of Q-enhancement (EQ 51).

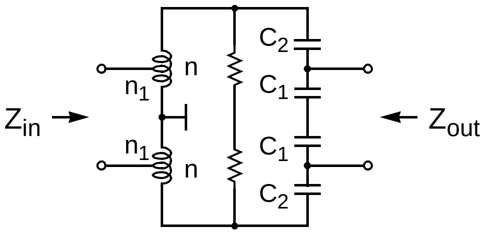


Figure 19. Two convenient ways of feeding or tapping a resonator. Tapped inductor (left) and capacitive division (right).

From this equation it is clear that gm must be increased if the uncompensated tank Q-value is decreased or if the characteristic impedance of the resonator is reduced. This means that in order to minimize power consumption, Q_{tank} and $\sqrt{L/C}$ must be maximized. As a final remark it should be emphasized that moderate Q-enhancement (e.g. $k_{QE} < 5\text{-}10$) may very well be what is needed together with state-of-the-art resonator design and active resonator coupling to make fully integrated front-end filters feasible in high-performance radio systems like GSM and WCDMA.

3.1.4 Impedance Matching

We have seen that $|Z_{tank}| = Q \cdot \sqrt{L/C}$ at the resonance frequency. Q is set by the desired amount of attenuation (EQ 49) and the characteristic impedance has a lower limit depending on the amount of power dissipation that can be tolerated. This means that the tank impedance has a lower limit which is typically still very high. Minimum tank impedances at resonance can easily be in the 500Ω - 1kΩ range. This may lead to very large voltage swings and therefore may be problematic for both the source and the load at the output. In order to circumvent the impedance matching problems it may be necessary to transform this high impedance level down to a more useful level at both the source and at the load. Figure 19 shows two convenient ways of doing this. One is the use of tapped inductors and the other is the use of capacitive division. Both can be used for the source and both can be used for load but the configuration shown is often the most practical. A tapped inductor is convenient at the input because it allows biasing of the input transconductance stage through the center point. Capacitive division in turn is convenient at the output because it allows accurate control of the output voltage swing and DC bias level. Furthermore the load is often capacitive in nature so it can simply be incorporated as a part of the resonator. To get the desired characteristic impedance and center frequency of the resonator in figure 19 one must make sure that the series connection of the inductors equals L and the series connection of the capacitors equals C. An added benefit of tapped inductance and capacitive division in a resonator is that the losses associated with the source and the load are reduced because they are connected to nodes with lower impedance.

3.2 Design of Metal-Metal Capacitors in Standard CMOS

Only a decade ago it was considered impossible to make good radio frequency integrated circuits in digital CMOS processes due to the lack of good passive components and slow transistors. Since then the massive investments in digital CMOS have resulted in exceedingly fast devices and many metal layers as standard. These many metal layers have made on-chip inductors feasible by reducing ohmic loss. Furthermore on-chip inductors have one very nice property. Because the inductance value is dependent on large physical dimensions in processes where dimensions are controlled to less than 0.1 micron the inductance value is highly repeatable. Typical lot-to-lot variations are better than 1%. Meanwhile RF IC designers have grown accustomed to using huge design margins in for instance VCO tuning range to make up for at least ±10% capacitance variation in the thin film capacitors (poly-poly or metal-metal) they use in the specialized RF IC processes. These design margins lower the achievable RF performance and therefore high-Q capacitors with better tolerances are very desirable.

Another consequence of the rapid CMOS process development is that now CMOS switches have very good RF performance. This has created an alternative to the use of varactors as frequency selective devices. By switching between a number of binary weighted capacitors it is now possible to achieve the same tuning range and as-good-as or better Q-value with a linear metal - metal capacitance.

This section explains how RF capacitors with superior performance can be made in a standard CMOS processes without thin film capacitor options. The density is lower than for the thin film capacitors but at RF frequencies this is often not a problem because the needed capacitance values are typically small.

3.2.1 Capacitor Construction

Today oxide layers and metal layers in digital CMOS processes are typically kept larger than 0.5 micron to minimize RC delays. At the same time minimum metal spacing is dropping to significantly below 0.5 micron to improve the density. This means that now lateral capacitance contributions can be significantly larger than the vertical contributions between different metal layers. Therefore using interleaved metal fingers results in higher capacitance per area (see Figure 20). Higher capacitance per area can be achieved by shifting the fingers in each layer [24] but the improvement is minor compared to the solution that is shown in Figure 20.

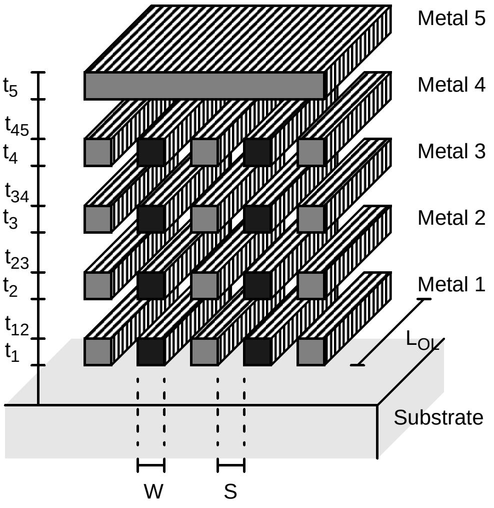


Figure 20. Four Stacked Lateral Flux Capacitors. Fingers with dark cross-sections are connected to one port. The remaining fingers are connected to the other port.

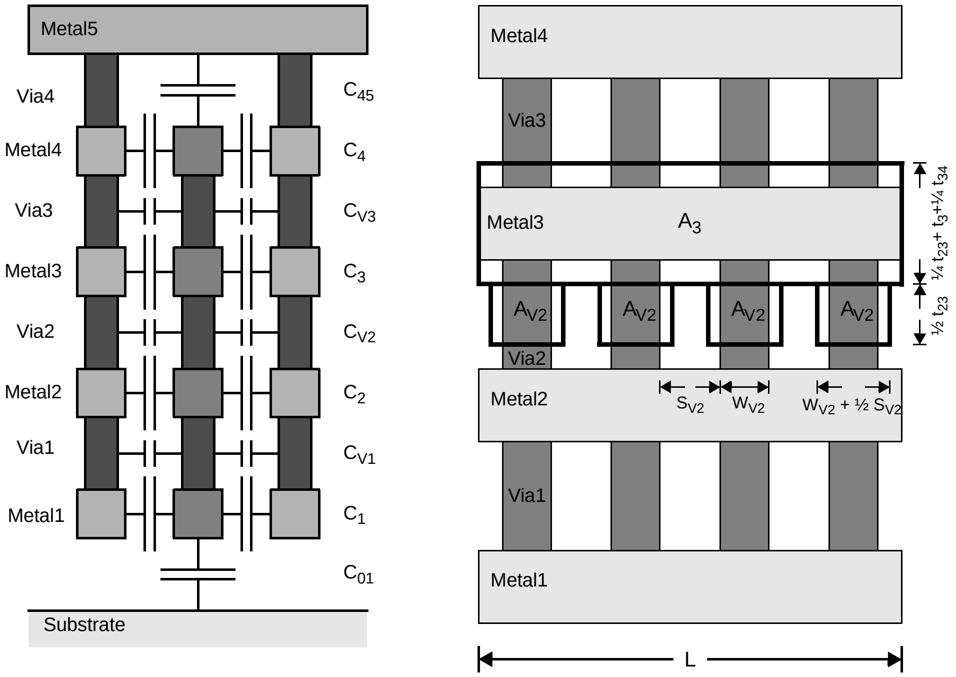
Furthermore this solution can be improved by connecting all fingers with vias and thereby creating extra lateral capacitance contribution (Figure 21). By doing so we have formed vertical metal meshes that are completely shielded from each other. This and the fact that each vertical capacitor unit is only roughly two microns wide makes them ideal as unit capacitors used in binary weighted switched capacitor banks that need to be well matched even at RF frequencies. In Figure 21 we can also see that there is some parasitic bottom plate capacitance. For this structure the bottom plate capacitance is very small i.e. in the order of 5% of the mesh capacitance but if the bottom plate capacitance is a problem e.g. due to substrate losses or capacitive division metal1 can be used as a ground plane or as a bottom shield just like metal5 is used as a top shield in figure 21. This reduces loss in the first case and removes C_{par} in the second case at the expense of lower capacitance per area.

3.2.2 Capacitance Calculation

For a vertical mesh capacitor using four layers of lateral capacitance contributions like the one in Figure 21 it is clear that the capacitance of each vertical mesh (C_{mesh}) is a combination of seven lateral capacitance contributions and one vertical (52) while the parasitic capacitance to ground is equal to C_{01} (53).

$$C_{mesh} = 2(C_1 + C_{V1} + C_2 + C_{V2} + C_3 + C_{V3} + C_4) + C_{45} \tag{52}$$

$$C_{par} = C_{01} \tag{53}$$


Metal5
Via4
Metal4
Via3
Metal3
Via2
Metal2
Via1
Metal1
Substrate
C_{45}
C_4
C_{V3}
C_3
C_{V2}
C_2
C_{V1}
C_1
C_{01}
Metal4
Via3
Metal3
A_3
A_{V2}
A_{V2}
A_{V2}
A_{V2}
Via2
Metal2
S_{V2}
W_{V2}
W_{V2} + ½ S_{V2}
Via1
Metal1
L
¼ t_{23}+ t_3+¼ t_{34}
½ t_{23}

Cross-Section Side View

Figure 21. Cross-Section of a Vertical Mesh Capacitor (left) and side view (right). Stacked fingers are connected by vias and thereby forming vertical meshes. The capacitance is the combination of several individual lateral flux capacitances. Parallel plate capacitance areas (A_3 and A_{V2}) are increased to take the fringing field capacitance contributions into account.

In order to be able to estimate what C_{mesh} and C_{par} amount to it is necessary to characterize each of these nine contributions separately. Fortunately there are only three types of capacitance contributions that are different in nature so it is sufficient to derive three capacitance formulas.

Before doing so a number of definitions are made to ease the presentation (see next page). Each of these definitions refers to a parameter that is shown in Figure 20 or Figure 21. Note that the virtual parallel plate capacitance areas are parallel plate capacitance areas that are increased enough to take the fringing capacitance contributions into account as well. I.e. if the fringing capacitance constitutes 25% of the parallel plate capacitance then the virtual parallel plate capacitance area is simply made 25% larger than the actual parallel plate capacitance area. This way the individual capacitance contributions including the fringing effects can be calculated by simply using the parallel plate capacitance formula (54) with A_{plate} being the virtual parallel plate capacitance area.

A_i	virtual parallel plate capacitance area between two fingers in metal layer i
A_{vi}	virtual parallel plate capacitance area between opposite vias on two fingers in metal layer i
A_{ij}	virtual parallel plate capacitance area between conducting metal layer i and metal layer j (i = 0 represents substrate or poly)
C_i	lateral flux capacitance contribution from metal layer i on one side of one vertical mesh
C_{ij}	capacitance between metal layer i and metal layer j from one vertical mesh
C_{main}	main capacitance
C_{mesh}	capacitance of one vertical mesh
C_{par}	parasitic bottom plate capacitance of one mesh
C_{vi}	capacitance between opposite vias on two fingers in metal layer i capacitance between via i and metal layer j from one vertical mesh
ε_0	permittivity of vacuum (= 8.854e-12AsV^{-1}m^{-1})
ε_{rij}	relative permittivity of the dielectric between metal layer i and metal layer j
k	number of topmost metal layer in the capacitor
L_{ol}	length where two vertical meshes overlap
L_{eff}	effective mesh length taking end contributions into account as well ($L_{eff} = L_{ol} + W + \pi S/2$)
n	number of individual meshes connected to one port of the capacitor.
S	vertical mesh spacing
S_{vi}	spacing between vias in layer i
t_i	thickness of metal layer i
t_{ij}	thickness of dielectric between metal i and metal j
W	width of the metal fingers
W_{vi}	width of vias in layer i
W_{eni}	metal enclosure of via i

$$C_{\text{ParallelPlate}} = \varepsilon_{\text{relative}} \cdot \varepsilon_0 \cdot \frac{A_{\text{Plate}}}{\text{distance}} \tag{54}$$

Naturally estimating these virtual parallel plate capacitance areas is not trivial and usually it is necessary to carry out time consuming 3D electromagnetic simulations but in this case it is possible to make some simplifying assumptions that allow a very rapid estimation of the virtual parallel plate capacitance area.

Figure 21 shows a side view of a vertical mesh. An estimate of the virtual parallel plate capacitance area of C_3 (i.e. A_3) is indicated by the upper rectangle. The width of the rectangle is L_{eff} i.e. the length of the vertical mesh. The height is more involved because this is the parameter that takes the fringing effects into account. The rectangle height is some fraction of t_{23} representing the lower fringing contribution plus the thickness of metal3 (t_3) plus some fraction of t_{34} representing the upper fringing contribution.

Now guessing the fraction of t_{23} (and t_{34}) is the challenging part. The lower bound must be zero because this corresponds to no fringing capacitance. The upper bound must be 1/2 because this corresponds to a solid plate i.e. if the holes were filled with metal.

The lower limit is reached when S/t_{23} approaches zero because the fringing contribution becomes negligible compared to t_{23} and the upper bound is reached when S/t_{23} approaches infinity. For reasonable choices of mesh spacing S/t_{23} (and S/t_{34}) are in the order of one and therefore a justified guess is that the fraction is halfway between the two bounds i.e. 1/4 of t_{23} and likewise 1/4 of t_{34}. This is the simplifying assumption that allows hand calculation of the vertical mesh capacitance. Therefore the lateral capacitance contribution of metal layer i including the fringing capacitance contributions can be expressed as (55).

$$C_i = \varepsilon_{\mathrm{ri}}\varepsilon_0\frac{A_i}{S} = \varepsilon_{\mathrm{ri}}\varepsilon_0\frac{L_{eff}\left(\frac{t_{(i-1)i}}{4} + t_i + \frac{t_{i(i+1)}}{4}\right)}{S} \tag{55}$$

The lateral capacitance contributions from the vias as for instance C_{V2} also need to be characterized. In the same manner as before these capacitance contributions are calculated using virtual parallel plate capacitance areas. Figure 21 shows the virtual parallel plate capacitance area of C_{V2} i.e. A_{V2} as a number of small rectangles. Each rectangle corresponds to the contribution from one via.

The height of the rectangle is limited to $t_{23} - 2 \times (t_{23}/4) = ½\, t_{23}$ because the same area can not be counted twice. The width of each rectangle is estimated in the same way as before i.e. halfway between the upper and the lower bound. The lower bound is W_{V2} and the upper bound is $W_{V2} + 2 \times (S_{V2}/2)$ and thus the width is estimated to be $W_{V2} + S_{V2}/2$. In order to find the total amount of capacitance from one layer of vias this rectangle area should be multiplied by the number of vias. The number of vias is approximately $L_{eff}/(W_{V2} + S_{V2})$ and therefore the capacitance contribution from one side of a mesh from vias in layer i (C_{Vi}) can be expressed as (56).

$$C_{vi} = \varepsilon_{\mathrm{rvi}}\varepsilon_0\frac{A_{\mathrm{vi}}}{S + 2W_{\mathrm{eni}}} = \varepsilon_{\mathrm{rvi}}\varepsilon_0\frac{L_{eff}\cdot\frac{W_{\mathrm{vi}} + S_{\mathrm{vi}}/2}{W_{\mathrm{vi}} + S_{\mathrm{vi}}}\cdot\frac{t_{\mathrm{i(i+1)}}}{2}}{S + 2W_{\mathrm{eni}}} \tag{56}$$

The remaining capacitance contributions are the vertical capacitance contributions i.e. the top and the bottom capacitances. These contributions are also estimated using a virtual parallel plate capacitance area. Both virtual capacitance areas are estimated by the mesh overlap length plus half t_{ij} (fringing) multiplied by the mesh width W plus half the mesh spacing (fringing). This leads to the expression for the top and bottom capacitances (57).

$$C_{ij} = \varepsilon_{\mathrm{rij}}\varepsilon_0\frac{(L_{ol} + t_{ij}/2)\cdot(W + S/2)}{t_{ij}} \tag{57}$$

Now we have expressions for all the individual capacitance contributions of one vertical mesh capacitor. Thus all we need to do to get the total capacitance C_{main} is to add the contributions and multiply by the number of vertical meshes (n) i.e. (58). Note that the bottom plate capacitance C_{01} is not included because it is considered a parasitic capacitance. If the other terminal is connected to ground this capacitance should be added to C_{main}.

$$C_{main} = n \cdot \left(\sum_{i=1}^{k} 2C_i + \sum_{i=1}^{k-1} 2C_{vi} + C_{k(k+1)} \right) \tag{58}$$

3.2.3 Process Variation

Perhaps the most important feature of the vertical mesh capacitor is that it uses a combination of many capacitance contributions that are formed in different processing steps. This means that the process variations of these capacitances are *not correlated* and therefore the combined mesh capacitor experiences significant averaging. For instance if a vertical mesh capacitor is made out of four equal sized contributions with ±10% variation each then the combined capacitor has $\pm 10\% / \sqrt{4} = \pm 5\%$ variation and if the capacitor is made of nine equal sized contributions then the variation is $\pm 10\% / \sqrt{9} = \pm 3.3\%$. The individual capacitances depend on three main parameters; layer thickness, spacing and relative dielectric constant. If the capacitor is designed with sufficiently large mesh spacing then the capacitance variation is dominated by the layer thickness variation which is typically in the order of ±10%. Therefore well designed vertical mesh capacitors have as little as ±3-5% process variation.

3.2.4 Measurement Results

A vertical mesh capacitor was designed for an LC resonator with switched capacitor tuning (figure 23) in a standard 0.25 micron CMOS process with 6 metal layers. Also a separate test structure with 42 capacitor meshes was fabricated to be able to characterize the capacitor through probe measurements. Metal 2, 3 and 4 were used for the meshes, metal 5 was used as top shield and metal 1 was used as a (relatively) low loss ground plane. The mesh spacing was chosen to be 0.5 micron and the meshes were made 51 micron long to give the desired unit capacitance value. The measured capacitance (figure 22) shows excellent agreement with the calculated values. The model overestimates the capacitance by only 2.4% which must be considered excellent and the average capacitances of the three measured capacitors varied by as little as ±0.2%. Naturally this is a very small data set but still the result is satisfying because it hints that the process variation is better than 3-5% as expected.

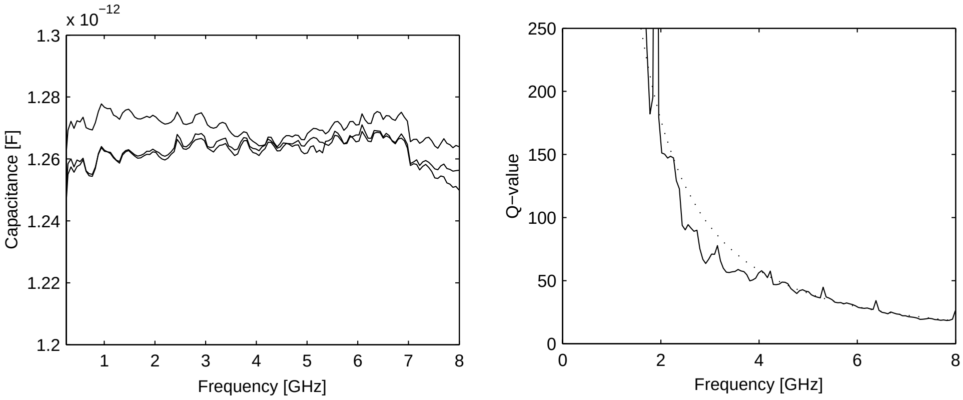


Figure 22. Capacitance of 42 meshes (M2, M3 and M4) each 51 micron long. Measurements from three dice (left). Quality factor of the fabricated capacitor (right). Measured Q = -Im(Z) / Re(Z) (solid), measured with second order polynomial fit of Re(Z) from 4 to 8 GHz (dotted).

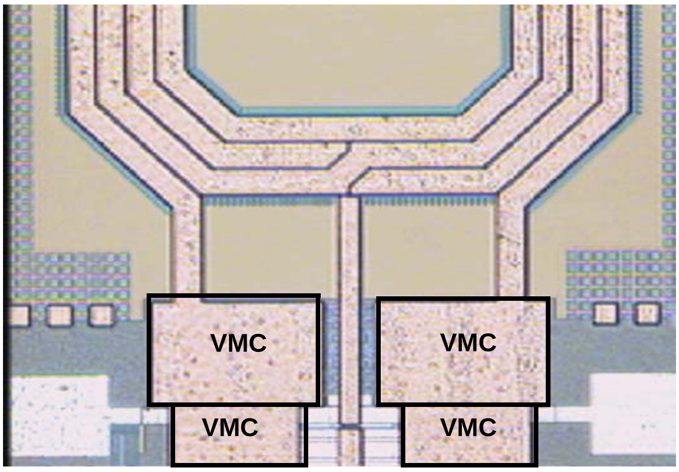


Figure 23. Chip photo of Vertical Mesh Capacitors used in an LC resonator with programmable centre frequency.

Although not optimized for maximum Q the measurements also show very good Q-values (Figure 22). Measuring high-Q devices is not trivial because the S-parameters are located at the edge of the Smith chart. Therefore for this device the Q-values are not very accurate below 3-4 GHz and the loss is most likely dominated by the test fixture and especially by the probe contact resistance so the actual device Q-value is probably better than what the measurements indicate.

Table 3: Summary of Capacitor Characteristics

Parameter	Calculated	Measured
Capacitance	1.30 pF	1.27 pF
Bottom plate capacitance	74fF ~ 6%	---
Capacitance per Area (M2-M4)	0.27 fF/μm^2	0.26 fF/μm^2
Capacitance per Mesh (51 μm)	30.6 fF	29.9 fF
Q-value at 4.0 GHz	> 100	> 57
Chip-to-Chip Capacitance Variation (average 250MHz - 8.0GHz)	<± 3-5%	± 0.2%

This section has explained how good RF capacitors can be made in standard CMOS processes. By stacking a number of interleaved metal fingers and connecting them with vias it is possible to form vertical metal meshes that provide good capacitance density and high Q-value. Each unit capacitor is completely shielded yet it is only two microns wide which makes it ideal for binary weighted switched capacitor banks. The capacitors show low process variation due to averaging of capacitance tolerances adhering from processing steps that are not correlated. Also the measured capacitance is almost constant up to 8GHz so the series inductance is negligible over that frequency span. Finally the section presents a simple yet accurate method for calculating the capacitance of these structures. By using this method it was possible to predict the measured capacitance with an error of only 2.4%.

3.3 Inductor Design in CMOS Processes with Highly Doped Substrates.

For on-chip RF filters it is absolutely crucial that the best possible inductors are used as the filter performance scales with the square of the resonator quality factor which is almost always dominated by the on-chip inductor quality factor. Therefore a substantial amount of work has been put into the design of good on-chip inductors. This section will try to present the results of this work. First the parasitics that affect the inductor quality are explained. Then modelling and layout aspects are treated. After this a number of measurement results are presented and finally a design guide is provided that explains how close-to-optimal inductors can be designed in CMOS processes with low substrate resistivity. Inductors made in silicon processes with relatively high substrate resistivity are generally superior compared to inductors made in processes with low substrate resistivity. When this project was initiated, fabrication in CMOS processes with low substrate resistivity was the only option. Furthermore at that time this kind of process (EPI) was considered standard and standard CMOS is the target of this research so the choice of a low resistivity substrate was natural. Fortunately CMOS processes with higher substrate resistivity (Bulk CMOS) appear to be gaining in popularity after some breakthroughs in process development. This raises the hope that very soon significantly better RF CMOS filters can be made at low cost thanks to process development.

3.3.1 Inductor Parasitics

On-chip inductors in CMOS processes with highly doped substrates (EPI substrates) are affected by the proximity of the low resistivity substrate. Typically the substrate resistivity is in the order of 0.01 Ωcm which is three orders of magnitude lower than what is normal for bipolar, BiCMOS and Bulk CMOS processes. This means that eddy currents in the substrate have a much greater impact on the inductor quality in highly doped CMOS processes than in other silicon processes. A cross-section of a circular inductor is shown in figure 24. It shows how the magnetic flux generated by the inductor current, penetrates the substrate and the innermost inductor turns and thereby induces eddy currents. There are three dominating loss mechanisms in on-chip inductors in highly doped CMOS processes. The first is the series resistance of the inductor. At DC the series resistance is determined mainly by the metal sheet resistance and by the inductor length. At higher frequencies eddy currents, skin effect, Hall effect and proximity effects make the current-flow in the inductor turns non-uniform and thereby increase the effective series resistance. The second loss mechanism is the power that is dissipated due to eddy currents in the substrate. The third loss mechanism is the capacitive coupling to the conducting substrate. This coupling makes the substrate potential immediately under the inductor turns vary which makes replacement currents flow in the substrate and thereby power is dissipated. At low GHz frequencies this component is often negligible but at higher frequencies or with higher substrate resistivities this component could be significant.

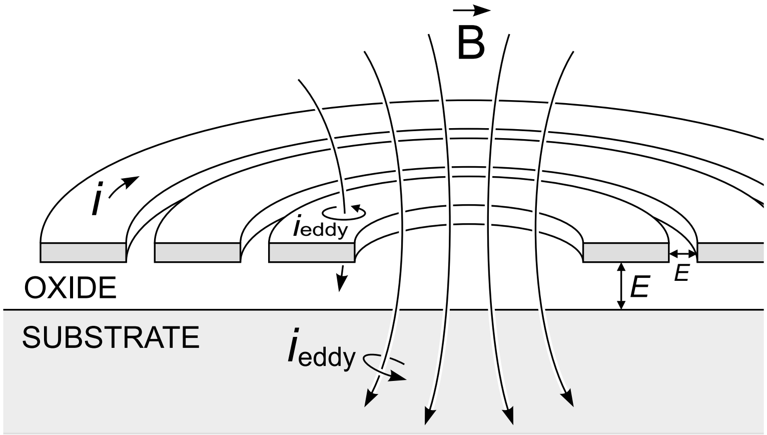


Figure 24. Low Resistivity (EPI) CMOS inductor physics. Shows magnetic flux (B) causing induced eddy currents and electric fields (E) causing non-uniform current distribution in the coil and replacement currents in the substrate.

The capacitive couplings between windings and to the substrate do not directly reduce the quality of the inductor but they do affect the inductor impedance and above a certain frequency (the self-resonance frequency) they make the inductor useless as an inductor. Furthermore they may affect the current distribution in the windings and thereby increase the series resistance.

3.3.2 Layout of On-Chip Inductors

On-chip inductors can be laid out in a number of different ways with many degrees of freedom. Generally the design parameters can be divided into two groups - those that deal with lateral dimensions and those that deal with vertical dimensions. Starting with the lateral design parameters there are the metal width of one winding in the coil (W) and the spacing between windings (S). Then there is the number of inductor turns (n). Next there is average inner radius (R_I) and the average outer radius (R_O) and finally there is the inductor shape.

The most typical inductor shapes are the square shape and the octagonal shape but circular inductors are sometimes also allowed by the silicon foundry.

Figure 25 shows two octagonally shaped inductors with two turns. The left one is a traditional inductor layout with turns wound from the inside and out. The right one is a balanced inductor layout that gives the inductor almost symmetrical electrical characteristics i.e. $Z_{11} = Z_{22}$. This kind of inductor is especially useful for resonators that are driven differentially. Figure 25 also shows how W, S, n, R_I and R_O should be interpreted.

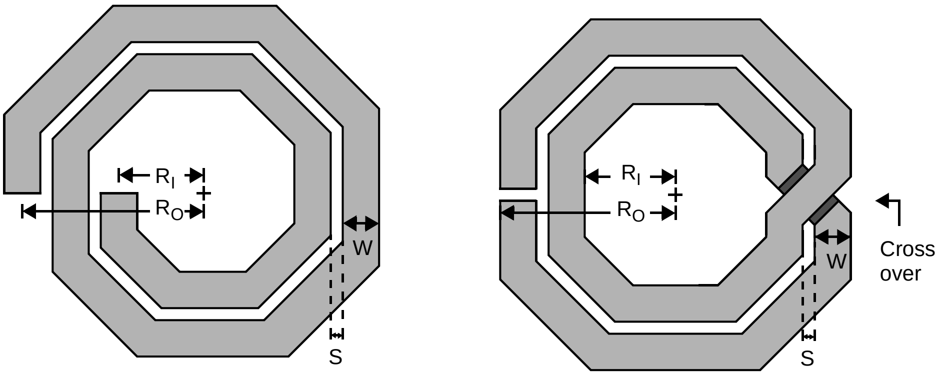


Figure 25. A two turn octagonal inductor with layout parameters shown. Traditional (left) and symmetrical (right).

Today most digital CMOS processes have many metal layers. This can be exploited in several ways to increase the inductance value or to reduce the effective series resistance. These methods can be regarded as vertical design parameters. The first technique is known as stacking and it exploits that identical coils in e.g. the top three metal layers can be stacked and connected with vias to reduce the effective series resistance thereby increasing the inductors quality factor.

A second vertical design technique is known as multi-layer inductor design. Again lets assume that we have three inductors in the three top layers. Now if they are connected in series instead of being shunted the resulting inductor will have three times as many turns and as the inductance value is proportional to the square of the number of turns the inductance can ideally be nine times higher than the inductance of a single inductor. Unfortunately the series resistance and the substrate losses are also at least nine times higher so usually multilayer inductors have slightly worse quality factor than stacked inductors. Nevertheless multi layer inductors can be very useful for situations where large inductances are needed or when silicon area is more important than the quality factor.

For all kinds of inductors it is necessary to cross windings to be able to connect to both ports from the outside. For the traditional layout style (Figure 25 - left) the connection is made with a connection in a lower laying metal layer directly to the innermost turn. This connection is usually referred to as the underpass. For the symmetrical or balanced layout style (Figure 25 - right) the connection is made by crossing windings once every half turn thereby gradually making its way to the center and out again. These winding crossings can be called cross-overs.

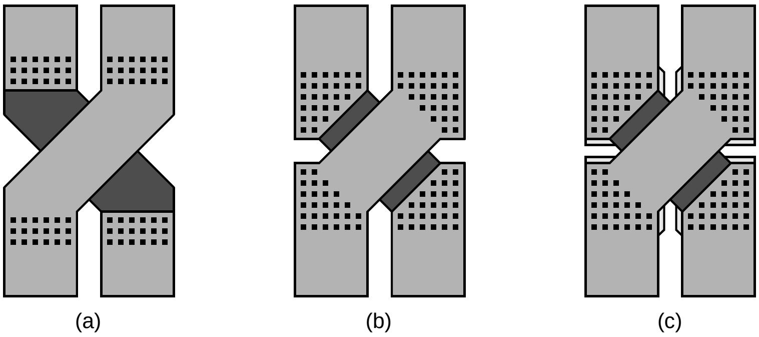


Figure 26. Cross-over layout for balanced inductors, for balanced transformers or for differential interconnect. Traditional cross-over (a), cross-over with reduced series resistance (b) and cross-over with reduced series resistance and capacitive shield for enhanced symmetry (c).

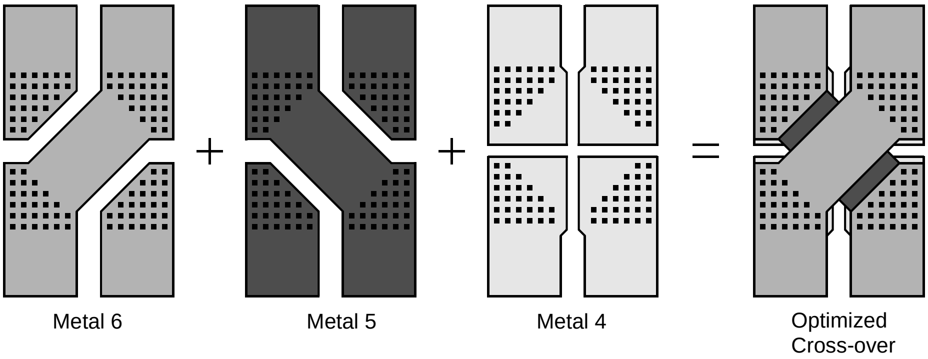
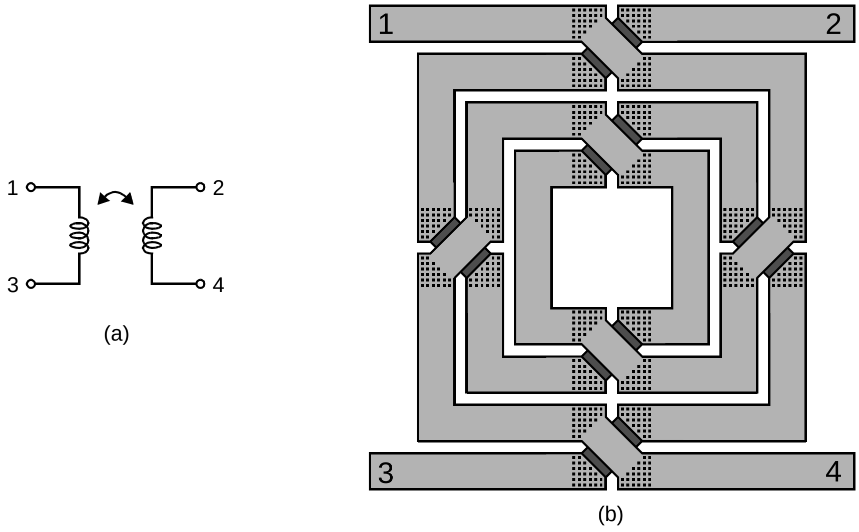


Figure 27. Layout of a cross-over with reduced series resistance and capacitive shield (metal 4) for enhanced symmetry.

Common to both underpasses and cross-overs is that unless careful layout is exercised the series resistance of the inductor can easily be increased by as much as 10-20%. If extreme caution is paid in the layout of the underpass or the cross-overs the series resistance increase can usually be kept in the order of 5-10%.

Figure 26 shows different ways of laying out cross-overs for inductors with one, two or three metal layers stacked. Figure 26 - (a) is a traditional layout style. Figure 26 - (b) shows a cross-over with reduced series resistance because the current is only flowing through one square in one layer and because the contacting is more effective. Figure 26 - (c) shows a cross-over with reduced series resistance and with a capacitive shield for enhanced symmetry. Figure 27 shows a break-down of the layout of this optimized cross-over.



(a)

(b)

Figure 28. Symmetrical square shaped transformer conveniently laid out with cross-overs on four sides. Schematic (a) and layout (b).

If the sheet resistances of the two top metal layers are different the series resistances can be made equal by making one strip narrower and the other wider in the crossing area. This enables the layout of a cross-over with almost perfect electrical symmetry and low series resistance. Such cross-overs are useful not only for balanced inductor but also for differential interconnect and for symmetrical transformers. Usually transformers are laid out with double crossings i.e. two connections crossing two others at the same point or with other involved layout schemes. This usually leads to significant extra series resistance and sometimes low coupling coefficients. Figure 28 shows a very convenient way of laying out on-chip transformers using cross-overs on four sides resulting in low series resistance, good coupling coefficients and almost perfect symmetry.

The capacitively coupled losses in the substrate can be significant in processes with higher substrate resistivity. In such cases these losses can be reduced by placing a conducting ground shield right over the substrate i.e. in poly or in metal 1. If this shield is made solid then currents are induced in the shield leading to much more loss. Instead Patterned Ground Shields [34] can be used. Figure 29 shows how patterned ground shields can be laid out. The basic principle is that a conducting shield in poly or metal 1 is placed under the inductor and sliced up in a number of fingers to prevent eddy currents from flowing in the shield. The fingers are then connected to a common point which is connected to ground. It is very important that the ring that connects the fingers is not closed because this will enable induced currents to flow in the ring. Usually the connecting ring is placed at the outer perimeter of the ground shield. This could give significant series resistance in the shield itself and the magnetic flux may penetrate the ring and induce some loss at this point (Figure 29 - left).

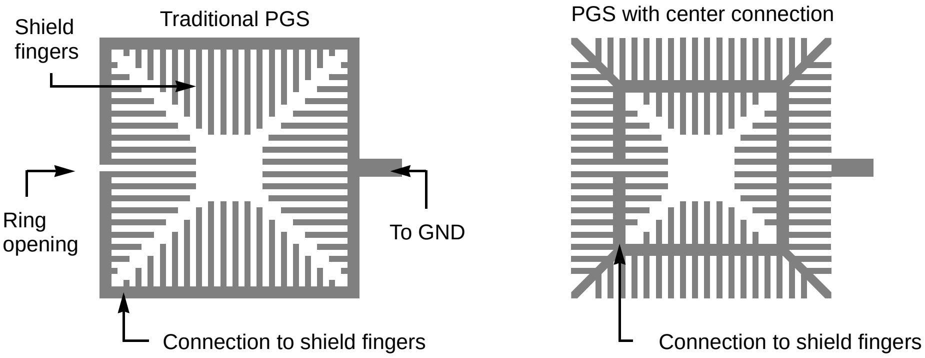

Traditional PGS
Shield fingers
Ring opening
To GND
Connection to shield fingers
PGS with center connection
Connection to shield fingers

Figure 29. Patterned Ground Shields in poly or metal1 for a square shaped inductor. Traditional shield with connections made from the outside (left) and shield with connections at the center of each finger to reduce eddy currents and series resistance (right).

If instead the ring is moved to the center of the fingers the series resistance of the shield is reduced and the magnetic flux is not penetrating the ring because the field lines are parallel to the shield layer at this point (Figure 29 - right). For inductors in silicon processes with low substrate resistivity the main loss mechanisms remain ohmic loss in the coil and induced loss in the substrate. The induced losses in the substrate can be reduced if options like deep trench isolation etc. are available. Usually they are not in CMOS processes but to some extent induced eddy current can be avoided in the surface by placing strips of N-well under the inductor. Unfortunately in EPI processes the major substrate loss contributer is not in the surface so the overall inductor quality can not be greatly improved this way.

3.3.3 Modeling

The behavior of CMOS inductors has been modelled in a number of different ways [26] - [30]. The dominating electrical effects are shown in Figure 30 - a. The coil is modelled as an ideal inductance L coupled in series with a resistor R that models the ohmic loss in the coil. The capacitive coupling to the conducting substrate is modelled with two capacitances C_1 and C_2 from each port in series with two resistors R_{SUB}. Note that R_{SUB} should be excluded from the model if a patterned ground shield is used. The distributed capacitive coupling between the inductor turns is modelled with a capacitor C_{12} placed between port1 and port2. Finally the induced substrate currents can be modelled with a coupling coefficient to a one turn inductor (with dimensions comparable to those of the coil) in series with a substrate resistance. Unfortunately the estimation of these parameters is not trivial and generally requires 3D finite element simulations. Furthermore the model is inconvenient for hand calculation. Therefore this loss contribution is often modelled by simply increasing R until the inductor Q-value is correct at a given desired frequency. The inductance value itself is reduced by this coupling but usually by less than 10% [37] so often this effect is simply neglected.

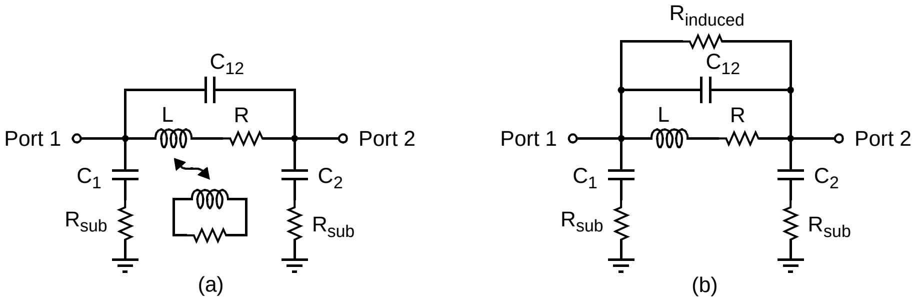


Figure 30. Lumped element inductor model. With inductively coupled substrate losses (a) and simple model with a parallel resistance modelling the substrate losses at a desired frequency (b).

An alternative is to model the induced substrate loss as a parallel resistance (Figure 30 - b). This gives a slightly better model over a wider bandwidth [32] compared to simply increasing R. Now lets see how the different component values corresponding to (Figure 30 - b) can be estimated. The inductance L can be estimated by using 3D field solvers such as Maxwell from Ansoft that solve Maxwell's equations numerically at the expense of time and memory usage. Other options include using custom tools like ASITIC and FastHenry. These tools use simplifying assumptions to reduce the computational burden and are good for verification but inconvenient for circuit design and synthesis. A very simple empirical formula (59) that is good for first cut designs is given in [31].

$$L = l \cdot (1.05 + 0.19 \cdot n) \qquad nH \tag{59}$$

$$l_{oct} = 1.027 \cdot \frac{\pi}{W+S} \cdot (R_O^2 - R_I^2) \tag{60}$$

$$L = \frac{\mu n^2 d_{avg} c_1}{2}\left[\ln\left(\frac{c_2}{\rho}\right) + c_3\rho + c_4\rho^2\right] \tag{61}$$

Where l is the inductor length in mm and n is the number of turns. (60) can be used to calculate the length of an octal inductor. S is the metal spacing and W is the wire width. R_O and R_I denote the distances from the center of the inductor to the middle of the outermost and innermost segments respectively. Recently a more accurate closed form expression (61) was presented [32]. In this expression μ is the permeability, d_{avg} is the arithmetic mean of the inner and the outer diameters, ρ is a measure of the inductors hollowness called the fill factor with $\rho = (d_{out} - d_{in}) / (d_{out} + d_{in})$ and the various c_n coefficients depend on the inductor shape. For a square C_1 - C_4 are 1.27, 2.07, 0.18, 0.13, for an octagon C_1 - C_4 are 1.07, 2.29, 0, 0.19 and for a circle C_1 - C_4 are 1, 2.46, 0, 0.2.

The inductor series resistance R can easily be calculated at DC as the metal sheet resistance multiplied by the inductor length and divided by the metal width W. At lower GHz frequencies the current distribution in the inductor turns is affected by a wealth of parasitics such as eddy currents, skin effect, Hall effect and proximity effects. These effects make the current distribution non-uniform and therefore increase the effective series resistance. This increase can easily be in the order of 50% - 100% [37]. At the present moment there is no accurate analytical expression for this increase so one needs to rely on finite element simulations and/or experiments. Note that this increase is strongly frequency dependent so R must be evaluated at the frequency of interest and can only be expected to be valid near this frequency. The parallel resistance $R_{induced}$ modeling the induced losses in the substrate is equally difficult to estimate. For a well designed inductor in a CMOS process with low substrate resistivity these losses are comparable to the losses in R so a very approximate expression is $R_{induced} = 2Q\omega_0L$. R_{sub} can usually be neglected in a silicon processes with low substrate resistivity and if not it can be eliminated with a patterned ground shield.

The capacitances can be calculated using specialized tools, empirical formulas [33] or by using the following simplified formulas.

$$C_{VA} \approx C_A(A_{IND} + A_{FRI} - A_{\text{underpass}}) \tag{62}$$

$$C_1 \approx C_{VA}/2 \tag{63}$$

$$C_2 \approx C_{VA}/2 + C_{\text{underpass}} \tag{64}$$

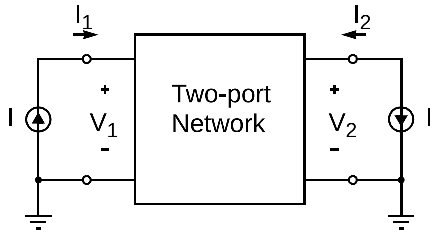
Where C_{AV} is the capacitance of a virtual inductor area seen from the substrate. C_A is the parallel plate-capacitance per area between the inductor layer and the substrate. A_{IND} is the area that the inductor covers (including the metal spacing). A_{FRI} is an area corresponding to the additional fringing contribution (approximately 3-4 micron multiplied by the total inductor perimeter). $A_{underpass}$ is the area of the underpass connecting the center of the inductor to the outside (port2). Finally $C_{underpass}$ is the capacitance between the underpass and the substrate. Note that if the inductor is a balanced inductor then $A_{underpass}$ and $C_{underpass}$ can be ignored.

C_{12} can be estimated using (65) where C_{IW} (Inter-Winding Capacitance) denotes the lateral capacitance between inductor turns per length. C_{IW} can be calculated using the techniques explained in the section on capacitor design or by using the formulas presented in [33]. Note that l_S in (65) is the total length of inductor spacing and C_{UP2IND} is the capacitance between the underpass and the inductor. The reason why C_{IW} is divided by (n-1), where n is the number of turns, is that the average voltage swing between inductor turns is (n-1) times lower than the voltage swing between port1 and port 2. A similar argument can be used to explain why C_{UP2IND} must be divided by 2.

$$C_{12} \approx l_S \cdot \frac{C_{IW}}{n-1} + \frac{C_{UP2IND}}{2} \tag{65}$$

$$C_{12\_\text{balanced}} \approx (l_S \cdot C_{IW} + C_{\text{crossover}})/k \tag{66}$$

Equation (65) only holds for the simple inductor type. If a balanced inductor is used instead, C_{12} can be estimated with (66) where $C_{crossover}$ is the total crossover capacitance.



$$V_1 = Z_{11}I_1 + Z_{12}I_2 = Z_{11}I - Z_{12}I$$
$$V_2 = Z_{21}I_1 + Z_{22}I_2 = Z_{21}I - Z_{22}I$$
$$V_{diff} = V_1 - V_2 = (Z_{11} - Z_{12} - Z_{21} + Z_{22})I$$

$$Z_{diff} = V_{diff} / I = Z_{11} - Z_{12} - Z_{21} + Z_{22}$$
$$Y_{diff} = 1 / Z_{diff}$$

Figure 31. Derivation of the differential impedance of a 2-port driven by a differential current source.

k is a measure of the average voltage swing between the inductor turns compared to between port1 and port2. Notice that in a balanced inductor the turns are interleaved and therefore k is approximately 2 for inductors with many turns and approximately 3 for inductor with few turns. This means that C_{12} can be much larger in a balanced inductor than in a simple inductor.

3.3.4 Quality Factor Definitions

When talking about the quality factor of an inductor there seems to be some confusion because the relevant measure can be different from one situation to another. The traditional inductor Q-value expression (67) is a good measure at frequencies where the inductor reactance is still primarily inductive. At higher frequencies i.e. when approaching the inductors self resonance frequency this Q measure goes to zero even though the inductor is not infinitely lossy. If the inductor is intended for use in a resonator a more appropriate measure is (68) because it expresses what the resonator Q would be at a given frequency assuming that the capacitive part of the resonator is lossless. This means that the Q-value does not go to zero when approaching the inductors self resonance frequency and thus gives a better picture of what can be expected at these frequencies. The inductance value L can be extracted from measurements at lower frequencies.

$$Q_{\text{single\_ind}} = \frac{-Im(Y_{11})}{Re(Y_{11})} \tag{67}$$

$$Q_{\text{single\_res}} = \frac{R_{\text{parallel}}}{\omega_0 L} = \frac{1}{\omega_0 L \cdot Re(Y_{11})} \tag{68}$$

$$Q_{\text{diff\_ind}} = \frac{-Im(Y_{\text{diff}})}{Re(Y_{\text{diff}})} \tag{69}$$

$$Q_{\text{diff\_res}} = \frac{R_{\text{parallel}}}{\omega_0 L} = \frac{1}{\omega_0 L \cdot Re(Y_{\text{diff}})} \tag{70}$$

Expressions (67) and (68) assume that the inductor is operated in a single ended configuration i.e. that one port is grounded. Here it was assumed that port 2 was the grounded port. Naturally (67) and (68) could just as well be defined with port 1 grounded by using Y_{22} instead of Y_{11}. If instead the inductor is intended for differential operation it would be more appropriate to express (67) and (68) in terms of the differential admittance Y_{diff} because especially the capacitive effects are different. This leads to (69) and (70) where Y_{diff} is defined as shown in Figure 31.

3.3.5 Measurement Results

A large number of different test inductors intended for single ended operation have been fabricated in a three metal layer 0.5 micron standard CMOS process with 3 metal layers on a highly doped substrate (10mΩcm). Furthermore a number of test inductors designed for a differential resonator have been fabricated in a 0.25 micron standard CMOS process 6 metal layers and 10mΩcm substrate resistivity. The purpose has been to investigate the impact of spacing S, inner radius R_I and patterned ground shields on the achievable Q-value plus characterization of a balanced inductor intended for use in a Q-enhanced LC filter.

A large spacing gives higher series resistance because W must be reduced - all other things being equal. On the other hand a too small spacing gives a high capacitive coupling and thus a lower frequency of operation plus perhaps non-uniform current distribution in the turns. Five inductors with the spacing swept from 1.2µm to 4.0µm (1.2µm, 1.5µm, 2.0µm, 2.8µm & 4.0µm) were integrated in order to investigate what the optimal spacing is. The metal pitch (S+W) was kept constant at 8 µm and all other parameters were left unchanged. S-parameter measurements were performed and the data was converted into Y-parameters. Then the pad parasitics were de-embedded and the Q-values were calculated using the classic Q-value definition (67).

The resulting Q-curves were smoothed using a Loess smoothing function and placed in a matrix. Finally a Q-surface was generated by applying curve fitting in the spacing-domain. The resulting contour plot is shown in Figure 32. We see an optimum Q-value for a spacing that is close to two micron. The peak (Q = 4.3) is found at a spacing of 1.9µm. It can also be observed that the capacitive coupling reduces the peak-Q frequency considerably below 1.5µm. Above 2.2 µm we can see a steady reduction of the quality factor corresponding to the increased series resistance.

The wire length, the number of turns and the size of the center hole are the parameters that have the greatest effect on the self-inductance of an inductor. In Figure 33 the L/l ratios of 21 different inductors (large, small, square, octal, large center hole and small center hole) are shown against the number of inductor turns. The line is a second order polynomial regression that shows the average relation. It is interesting to note that the measurements below the line generally are inductors with small center holes and the ones above the line are inductors with large center holes. This confirms the theory that energy is lost due to negative coupling and eddy currents in the innermost turns when the center hole is too small. Equation (59) is based on these measurements.

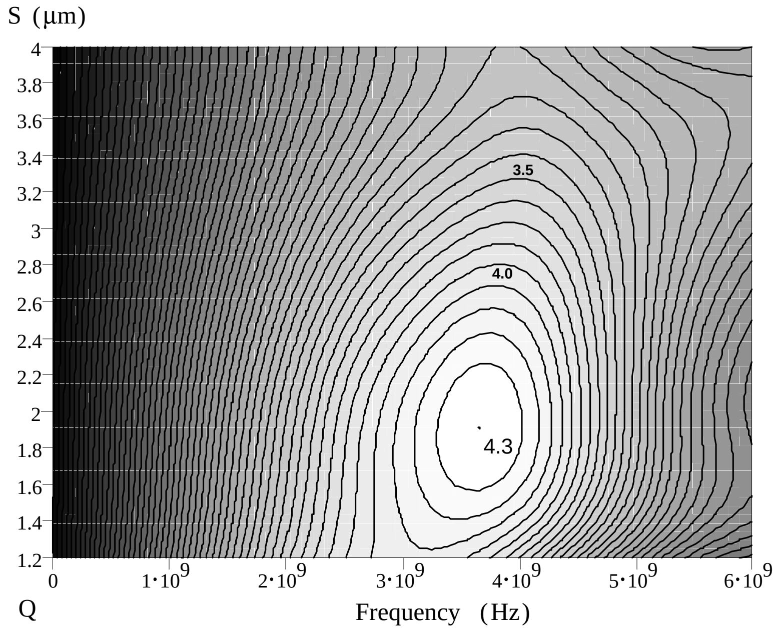


Figure 32. Measured Q-surface. Metal spacing versus frequency.

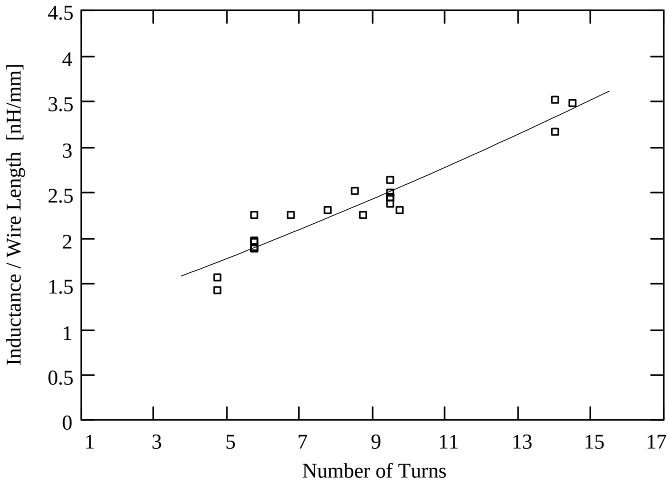


Figure 33. Inductance per wire length compared to the number of inductor terns.

The peak-Q frequency is dependent on induced losses (which are strongly frequency dependent) and capacitive couplings. Both are correlated to the inductor area and therefore it is relevant to see how the peak-Q frequency and inductor area correspond for the measured inductors. This is shown in Figure 34.

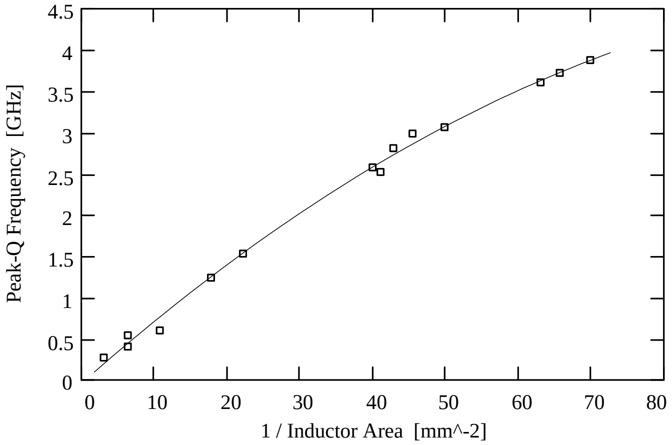


Figure 34. Measured Peak-Q Frequency versus inverse area. The measurements show surprisingly strong correlation considering the fact that the data points correspond to very different inductors.

It is interesting to note the strong correlation. Furthermore measurements from the 0.25 micron CMOS process with similar substrate resistivity agree well with Figure 34. Therefore it seems that the frequency dependent losses are dominating in setting the peak Q-frequency and therefore Figure 34 seems to be useful for first cut estimation of the optimal inductor area which can be used to fix the wire width W.

The innermost inductor turns in an inductor which is filled up to the center do not contribute positively to the quality of the inductor. Therefore the inner diameter should be large. On the other hand a too large inner diameter will reduce the inductance due to a smaller number of inductor turns. Therefore an optimum inner diameter exists.

Five inductors were fabricated in order to investigate what the optimum inner diameter is. They are based on an inductor with, S=2µm, W=6µm and 10 turns filled up to the center. Each inductor has one turn more taken away than the next. This gives inner diameters of 8, 24, 40, 56 and 72µm. As with Figure 32, a contour plot was generated (Figure 35).

It shows an optimum Q-value (4.6) somewhere near D_I = 64µm but as the peak-Q frequency is increasing along with the inner diameter (due to the smaller area) the Q-value could be improved by slightly increasing the wire width. This means that the optimum is probably found at a slightly higher diameter. It seems like the optimum is near a 80-100µm inner diameter for a 5 nH inductor in highly doped CMOS. This result matches well with the results obtained in [37] through 3D electromagnetic simulations.

A microphotograph of the chip with the test inductor used for these experiments is shown in Figure 36.

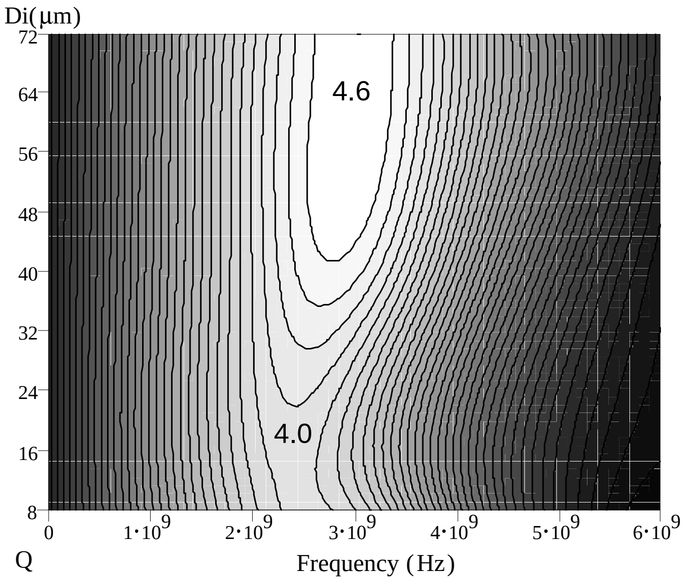


Figure 35. Measured Q-surface. Shows the effect on inductor Q when the inductor turns are gradually removed from the inside and out - thereby increasing the inner diameter.

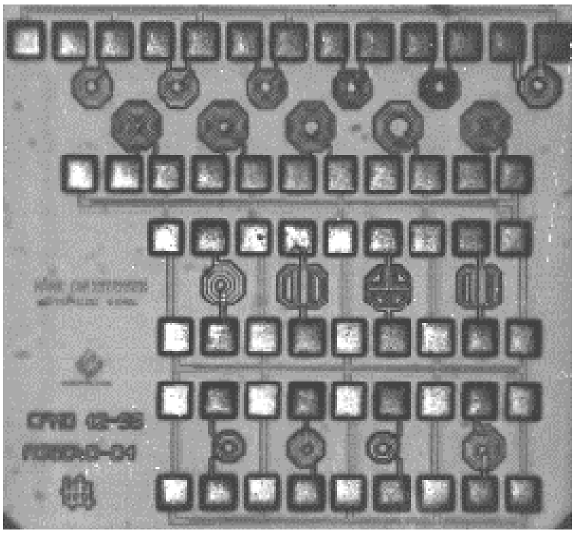


Figure 36. Microphotograph of the IC with test structures for investigation of optimal inductor spacing (top row) and inner radius (next row). The remaining inductors are experiments with alternative layout styles. Process: 0.5 μm, 3 metal, CMOS with 0.01Ωcm substrate resistivity.

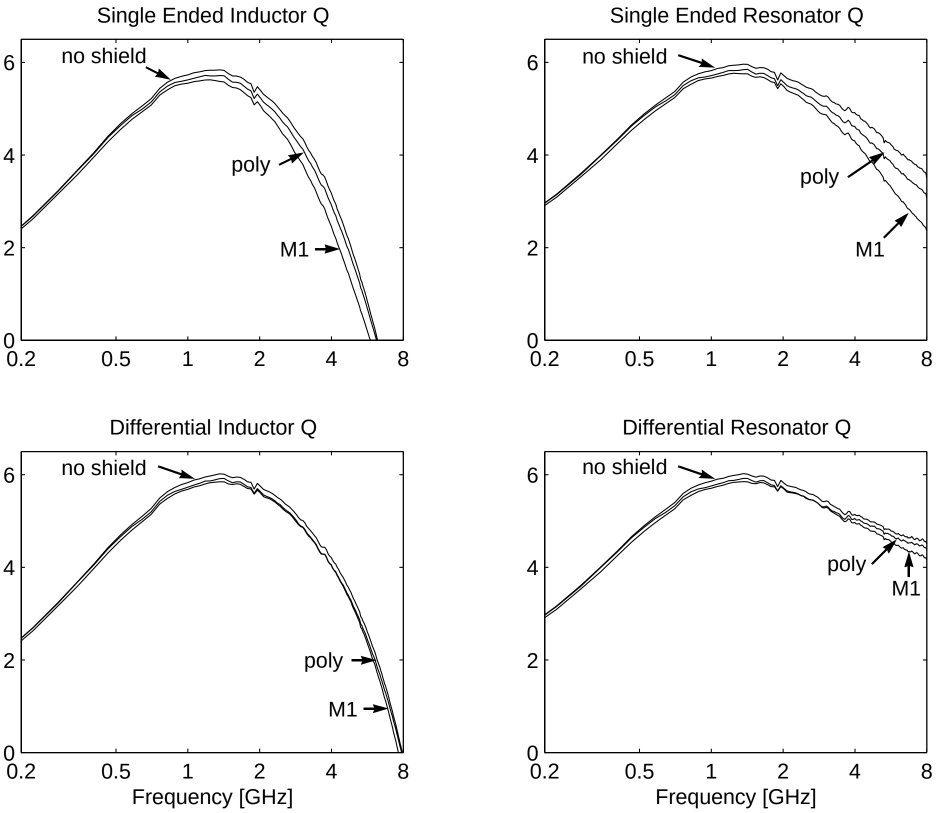


Figure 37. Measured Quality Factors of a 4.5 nH balanced inductor without patterned ground shield (no shield), with poly silicon ground shield (poly) and with metal1 ground shield (M1).

Later a balanced 4.5 nH inductor was designed for a Q-enhanced LC filter operating at 1750 MHz - 2250 MHz. It has an outer diameter of 240 µm, 4 turns, W=11µm, S=1.8µm and uses cross-overs optimized for best possible symmetry. The inductor shape is octagonal but squeezed to make it closer to a square shape in order to minimize layout overhead when integrated with the final circuit (see Figure 40). A stacked topology was used with metal 6, metal 5 and metal 4 connected for the inductor turns. This corresponds to a total thickness of 2.6µm and a DC sheet resistance of 13mΩ per square. In order to investigate the effect of patterned ground shields two other test structures were fabricated - one with a center connected ground shield (Figure 29) in poly and one with a center connected metal 1 shield. The measured Q-values for these three test structures are shown in Figure 37. Because the inductor is designed for a differential resonator it is the Q_{diff_res} definition (70) that is the most appropriate. The peak Q-value of the inductor without patterned ground shield is 6.0 at 1.4 GHz. This is lower than the intended frequency range 1750 MHz - 2250 MHz but it is not a major issue as the Q-value stays above 5.7 over a very wide frequency band (820MHz - 2370 MHZ).

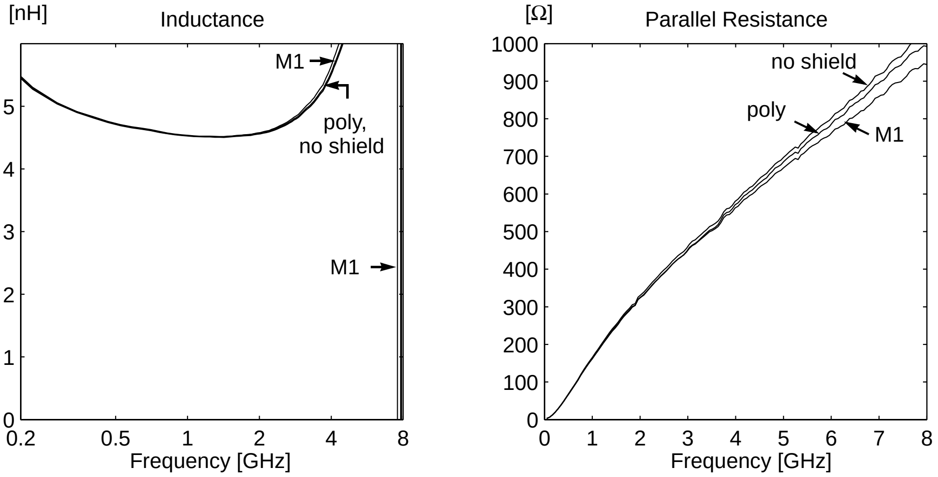


Figure 38. Parallel inductance and parallel resistance of a 4.5 nH balanced inductor without patterned ground shield (no shield), with poly silicon ground shield (poly) and with metal1 ground shield (M1).

It does however indicate that the inductor could be a little better in the desired frequency range if an outer diameter of 210 - 220 µm was used and perhaps if W was chosen to be 10 µm to further reduce the inductor area. Still a Q-value of 6.0 for a 4.5 nH inductor is, to the best of the authors knowledge, the best measured Q-value reported to date in a standard digital CMOS process with low substrate resistivity.

The measurements also show that a poly shield has no significant effect on the Q-value. Actually it lowers the Q-value by a couple of percent and this is not due to capacitive coupling but due to loss in the shield as can be seen from the parallel resistance (Figure 38 - right). This means that the capacitively coupled substrate loss is negligible and therefore a ground shield should only be used in a low resistivity CMOS process if the purpose is reduction of cross-talk. Often it is said that metal 1 can not be used for patterned ground shields because of induced losses. These measurements show that a metal 1 patterned ground shield can be made without significant loss below 4GHz when using sufficiently narrow fingers that are connected at the center point.

Figure 38 shows the measurements of the parallel inductance $L_P = 1/-2\pi f \cdot Im(Y_{diff})$ and the parallel resistance $R_P = 1/Re(Y_{diff})$. It can be seen that the patterned ground shields have very little effect on the inductance. The parallel resistance in turn is lowered above 4 GHz corresponding to extra loss caused by the shield.

The inductor was specifically designed for good balance i.e. $Z_{11} = Z_{22}$. The magnitude and phase of Z_{11}/Z_{22} is shown in Figure 39. It can be seen that the balance is extremely good up to 2-3 GHz i.e less that 0.5 degrees phase difference and 0.5% magnitude difference. In fact it is so good that it is likely that these numbers reflect accuracy of the measurement equipment rather than the imbalance of the inductor.

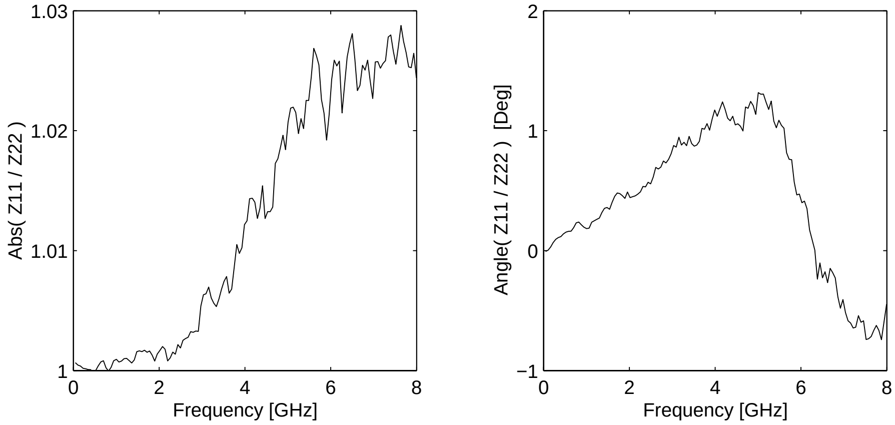


Figure 39. Balance of an octagonal inductor with optimized cross-overs. Ratio of the magnitudes of Z_{11} and Z_{22} (left) and the difference in phase of Z_{11} and Z_{22} (right).

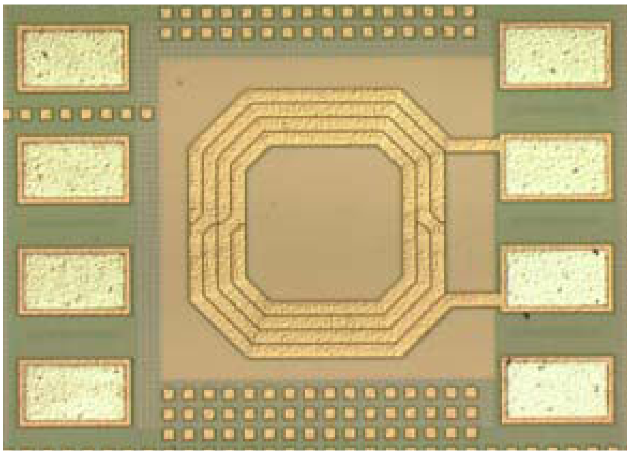


Figure 40. Microphotograph of the balanced test inductor with optimized cross-overs. Process: 0.25 µm, 6 metal, digital CMOS with 0.01Ωcm substrate resistivity.

3.3.6 Design Guide to On-Chip CMOS Inductors.

This design guide is based on the results found in the analysis of the test inductors and the guidelines given in [37]. It covers most issues involved with the design of good on-chip inductors in CMOS processes with low substrate resistivity.

1) Process. If it is possible, a process should be chosen with A) Low metal resistivity in the upper metal layers. B) As many metal layers as possible. C) Large oxide thickness (between substrate and metal layers). D) Highest possible substrate resistivity.

2) Layer topology. Metal1 should not be used for inductor turns - it is simply too close to the substrate. In a two metal-layer process Metal2 must be used for the inductor turns. In a three metal-layer process Metal3 is used for the inductor turns. Metal2 may also be used in a stacked configuration if the relative improvement in conductivity is significant. The same scheme applies for processes with more metal layers.

3) Inductor area. There is a close correlation between inductor area and the Peak-Q frequency so an area should be estimated that gives approximately the desired peak-Q frequency. Figure 34 may be a help in choosing an inductor area. Note that this area includes wire spacing and fringing fields.

4) Inductor shape. The shape that has the shortest perimeter for a given area is the circle. Therefore an inductor with circular turns has the highest number of turns for a given wire length and center-area. This means that the circular inductor gives the highest inductance value and hence the highest Q-value. An octagonal shape has a 2.7% longer perimeter and a square has a 12.8% longer perimeter than the circle. This means that octal inductor can be used with a minimal penalty. A square inductor can be used if the Q-value is of minor importance.

5) Average inner radius. Inductor turns close to the center have a negative effect on the quality of the inductor. Therefore a large hole should be made in the center of the inductor. The measurements indicate that for a 5.0 nH inductor a center hole with an average inner radius of 40-50µm give optimal conditions. It is reasonable to assume that a larger inductor should have a slightly bigger hole due to the stronger field.

6) Average outer radius. The inductor area and average inner radius have been chosen. This means that the average outer radius can be calculated from these numbers. Otherwise a good rule of thumb is that the outer radius should be 2-3 times the inner radius [35], [37]. The factor of 2 is probably the best for processes with low substrate resistivity. This corresponds to an outer radius of 80 - 100µm.

7) Number of turns. With the inner and outer radius fixed it is a simple matter to choose the number of turns. Use (EQ 61) for different values of n (number of turns) and choose the number that gives an inductance value that is closest to the desired.

8) Wire width and spacing. The wire width plus spacing is roughly the difference between the outer and inner radius divided by the number of turns so W+S is easily found. The measurements indicated that a spacing of 2 microns gives the best Q-value so this is a reasonable choice. Fixing S gives the wire width W because their sum is known. Now the inner radius should be adjusted to give the desired inductance more accurately.

9) Inductor layout. As low series resistance is of utmost importance be sure to connect the inside of the inductor to the outside with a bridge with sufficiently low resistance.

Use lower laying metal layer(s) with a width that give a sheet resistance per length not too far from the one used for the turns. 10-15µm is acceptable. Further be sure to use enough vias. The resulting via resistance should be in the order of the sheet resistance. Usually 40-50 vias is enough. At the outside of the inductor bring the connecting bridge back up into the same layers as used for the turns. Slightly wider metal may be used for the connecting strips to the inductor. Finally make sure an area surrounding the inductor is cleared to ensure that the inductor does not interact with neighboring circuitry. The inner radius can be used as a measure of a reasonable distance to surrounding circuitry/ metal. If a balanced inductor design is chosen use the guidelines for cross-over design presented earlier in this chapter.

10) Test layout considerations. For test purposes use the pad configuration specified by the available RF probes. Use a metal 1 ground shield under the pads. Furthermore place the pads far from the inductor (70-100µm or more) and use plenty of substrate contacts near the probe pads. Remember to make an identical layout with substrate contacts, identical proximities and everything but without the inductor itself. This way the parasitics of the pads can be measured and de-embedded from the inductor measurements.

3.4 Summary

This chapter explained how state-of-the-art on-chip resonators can be designed in a standard CMOS process.

The chapter started by reviewing different LC resonator topologies and then comments and design equations were given which are relevant for the design of fully integrated RF bandpass filters.

The most critical part of on-chip resonator design is the design of the used capacitors and inductors and therefore these components have received special attention in this chapter. First a new capacitor structure, the Vertical Mesh Capacitor, was presented. It has great RF performance and it is compatible with digital CMOS processes. Then it was explained in detail how state-of-the-art on-chip inductors can be designed without the use of expensive and slow electromagnetic simulators.

The chapter also presented a large number of measurement results from several test chips. Especially the Vertical Mesh Capacitors and an inductor that was designed for a tunable bandpass filter for 1750 - 2250 MHz showed excellent performance. Both the capacitor and the inductor appear to be state-of-the-art for standard CMOS processes with highly doped substrates.

3.5 References

[21] K.T. Christensen, "Design and Characterization of Vertical Mesh Capacitors in Standard CMOS," *in Digest of the 2001 VLSI Circuits Symposium*, pp. 201-204, 2001.

[22] D. Pederson and K. Mayaram, "Analog Integrated Circuits for Communication - Principles, Simulation and Design," *Kluwer Academic Publishers*, 1991.

[23] C. Bowick, "RF Circuit Design," *Newnes*, 1982.

[24] T. Lee, “The Design of CMOS Radio-Frequency Integrated Circuits,” *Cambridge University Press*, 1998, Chapter 2.

[25] H.Samavati et. al., “Fractal Capacitors,” *IEEE Journal of Solid-State Circuits*, Vol. 33 - 12, Dec. 1998, pp. 2035-2041.

[26] H.M. Greenhouse, " Design of Planar Rectangular Micro-electronic Inductors", *IEEE Trans. on Parts, Hybrids and Packaging*, Vol. PHP-10, no.2, pp. 101-109, June 1974.

[27] J.R. Long and M. A. Copeland, "The Modeling, Characterization, and Design of Monolithic Inductors for Silicon RF IC’s," *IEEE J. Solid-State Circuits*, Vol. 32. No. 3, pp. 357-369, March 1997.

[28] J. Craninckx and M. Steyaert, "A 1.8-GHz Low-Phase-Noise CMOS VCO Using Optimized Hollow Spiral Inductors," *IEEE J. Solid-State Circuits*, Vol. 32 No.5, May 1997.

[29] B. K. Ashby. et al., "High Q Inductors for Wireless Applications in a Complementary Silicon Bipolar Process," *IEEE J. Solid–State Circuits*, Vol. 31, no.1, pp. 4-9, Jan 1996.

[30] H. Ronkainen et al., "IC compatible planar inductors on silicon," *IEE Proc.-Circuits Devices Syst.*, Vol.144, No.1, pp.29-35, February 1997.

[31] K.T. Christensen and A. Jørgensen, "Easy Simulation and Design of On-Chip Inductors in Standard CMOS Processes", *Proceedings of the 1998 IEEE International Symposium on Circuit and Systems*, Vol. 4, pp. 360-364.

[32] S.S. Mohan, "The design, modeling and optimization of on-chip inductors and transformer circuits," *Ph.D. dissertation, Stanford University, Stanford, CA*, 1999.

[33] N. Weste and K. Eshraghian, "Principles of CMOS VLSI Design – A Systems Perspective," *Addison Wesley*, 1993.

[34] C.P. Yue and S. S. Wong, "On-chip spiral inductors with patterned ground shields for Si-based RF ICs," *In Digest of Tech. Papers 1997 VLSI Circuits Symposium*, June 1997.

[35] J. Craninckx and M. Steyaert, "Wireless CMOS Frequency Synthesizer Design," *Kluwer Academic Publishers*, 1998.

[36] T.H. Lee and S.S. Wong, "CMOS RF Integrated Circuits at 5 GHz and Beyond," *Proceedings of the IEEE*, Vol. 88, No. 10, October 2000.

[37] K.T. Christensen, "Design and Optimization of CMOS Switches for Switched Tuning of LC Resonators," *in Proceedings of the 2001 IEEE ISCAS*, Vol. 4, pp. 770-773, May 2001.

Chapter 4: Frequency Tuning

Frequency tuning of on-chip RF filters is very important because with current IC technologies there are both temperature and process variations that need to be taken into account. The frequency tuning problem can be divided into two parts - one is how to design the circuit component that causes the frequency shift and the other is how to design the system that controls the tuning. The first of these problems involves the basic RF performance of the filter i.e. the noise and linearity of the LC network that implements the filter function and the second deals with the more practical aspects of how the tuning is controlled.

Currently the feasibility of on-chip RF filters for higher performance systems is questioned by many because of their limited dynamic range. This is understandable because providing sufficient dynamic range is truly a very challenging problem. For this reason it was chosen to focus on the fundamental dynamic range issues rather than on the more practical aspects of implementing the tuning system itself i.e. it was chosen to focus on the design of tuning elements with very good linearity and very low noise. It is believed that switched metal-metal capacitor banks is the best option for such tuning elements. The design of the metal-metal capacitors themselves was treated in Chapter 3 so the majority of this chapter is dedicated to the design of optimal switches for switched frequency tuning.

The chapter starts with a short introduction to the field of frequency tuning. Then a rigorous treatment on the design of RF switches for switched tuning is given. Switched tuning is modelled and equations are derived for the optimum MOS switch width and Q-value. These expressions are then used to evaluate layout techniques and finally an enhanced differential waffle switch is developed.

4.1 Introduction to Frequency Tuning

This section gives a short introduction to frequency tuning. Continuous and discrete tuning elements are reviewed and the classic tuning methods are presented along with some suggestions for tuning systems that are well suited for fully integrated front-end RF filters.

4.1.1 Tuning Elements

In order to be able to vary the resonance frequency of an LC resonator it is required that the effective reactance of either the capacitor or the inductor can be changed (see the previous chapter). This can be done in a number of different ways which can be divided into two groups - tuning elements that enable continuous tuning and tuning elements that enable discrete frequency tuning.

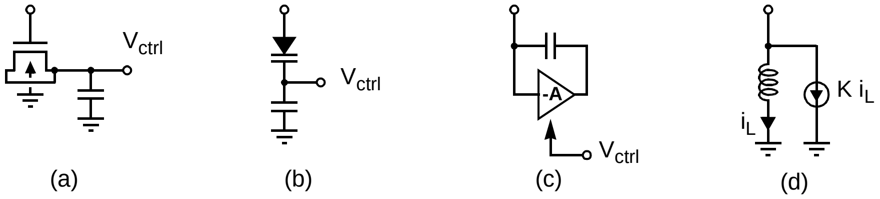


Figure 41. Components for continuous frequency tuning. MOS varactor (a), diode varactor (b), Miller feedback capacitor (c) and inductor with current feedback (d).

The most common frequency tuning elements are varactors because they are used in most LC oscillators. Varactors are typically implemented as either reverse biased diodes or MOS gate capacitors (Figure 41 - a & b). When the potential across these elements is varied the extension of the depletion layer in the semiconducting material is changed which effectively changes the parallel plate capacitance of the device. With current silicon technologies these components can be implemented with very low loss and therefore they would be the obvious choice for LC filters if it were not for the fact that these structures are inherently nonlinear devices. For some applications this might not be the worst problem but for the RX front-end filter it is important to use linear elements as the requirements for dynamic range are extremely high. Another issue is the sensitivity to noise coupled through the tuning node but this can usually be alleviated through the use of an AC shorting capacitor (Figure 41 - a & b). An alternative to varactor tuning that implements linear capacitors is the use of active circuitry to provide feedback. One well known example is the Miller capacitance (Figure 41 - c) which uses negative feedback to alter the effective value of a linear capacitor. At a first glance this method may appear attractive but the design of the inverting amplifier involves some serious noise, linearity, bandwidth and power trade-offs which almost inevitably lead to a reduced dynamic range for the filter. Feedback can also be used to alter the effective inductance. One example is shown in (Figure 41 - d) but this structure involves the same trade-offs as the Miller capacitance.

Strictly speaking the above mentioned tuning elements could be considered discrete if their control signals were applied through a D/A converter. This is not a very attractive solution because it does not solve the fundamental limitations of noise and linearity. Switched tuning on the other hand does. Switched tuning is simply the technique of switching passive elements like capacitors or inductor in or out of an LC resonator circuit. Especially switching binary weighted elements is useful because it makes it possible to cover a desired frequency range with a resolution set by the smallest unit element. Figure 42 - (a) shows a binary weighted switched capacitor bank and Figure 42 - (b) shows a binary weighted switched inductor bank. The binary weighted switched capacitor bank is the most practical because the capacitors are easier to lay out in a small area and thus easier to match.

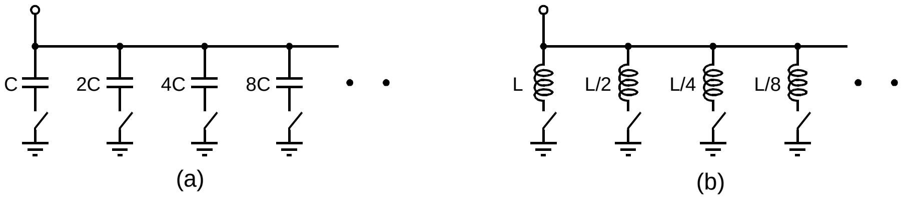


Figure 42. Components for discrete frequency tuning. Switched capacitor bank (a) and switched inductor bank (b).

Using both switched capacitance and switched inductance opens for the possibility of digitally programming or calibrating the characteristic impedance of the resonator at a given frequency or for keeping a constant characteristic impedance over frequency. These are some very nice degrees of freedom which are rarely if ever mentioned. Switched capacitor banks, however still appear to be the most practical solution for frequency tuning over a narrow bandwidth.

Switched tuning can, if the switched elements are implemented with linear passive components, result in a very linear resonator behavior. Furthermore because switched tuning does not involve active circuitry it does not suffer from the same fundamental power/noise trade-offs that the feedback techniques (Figure 41c & d) do. At the same time switched tuning opens for a wealth of new tuning options. For instance if the temperature drift is not too high calibration can be performed during fabrication or the resonance frequency corresponding to each switch setting can be stored in a lookup table. Also the used tuning control methods can be implemented in the digital domain or even handled by the DSP processor i.e. be handled in software. This at very low power consumption, great flexibility and very robust RF behavior. Naturally switched tuning is not without challenges of its own. Perhaps most pronounced is the need for RF switches with very low loss. Another challenge is the layout of switched capacitor banks with sufficient resolution i.e. with sufficiently good matching to give the desired resolution. This is challenging because capacitive as well as resistive and inductive parasitics must be well matched. Fortunately often the needed resolution is not terribly high. As an example a bandpass filter with a Q-value of 50 at 2GHz and 10% tuning range only needs a resolution of 5-6 bits to be well within the -3 dB bandwidth of 40MHz. Generally speaking it is believed that these challenges are manageable so in order to optimize RF performance and to increase the level of programmability it was chosen to pursue switched tuning as a means for frequency tuning in this research project.

4.1.2 Tuning Systems

Having discussed the different elements that allow frequency tuning it is now time to turn to the tuning systems that control the tuning elements. The tuning system does not involve fundamental limitations in terms of RF performance for the filter so in theory the tuning control problem can be solved.

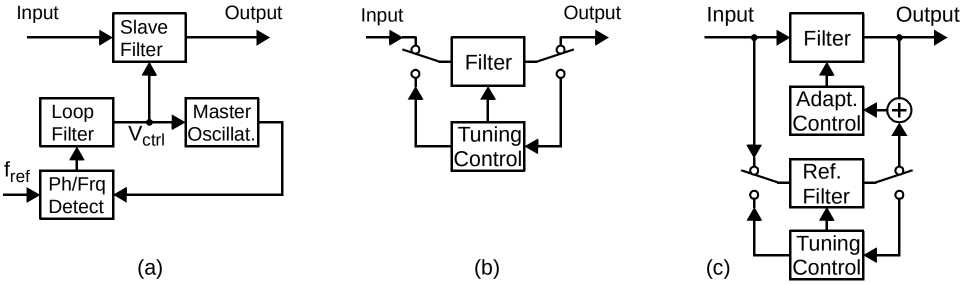


Figure 43. Classic Tuning Approaches. Master-Slave (a), Self-Tuning (b) and Adaptive filter tuning (c).

In practice however there are many issues such as element matching, continuous operation and power consumption that complicate the design of a tuning system. This section presents the classic tuning systems along with some ideas for tuning systems that exploit the circuitry that is already present in an RF receiver.

The first and most common technique for tuning continuous-time filters is called the Master-Slave technique (Figure 43 - a). It exploits the fact that an almost identical copy of the filters resonator can be turned into an oscillator with the same resonance frequency. This oscillator can then be locked to a reference frequency through a phase locked loop. If the filters frequency is controlled by the same signal as the oscillator then the filter will have the desired center frequency once the PLL is in lock. The Master-Slave technique relies on matching between two different structures. Matching can be made with 1-2% accuracy so the maximum filter Q-value is limited to 50 or lower [39]. Furthermore the technique involves substantial extra chip area and power consumption.

Another technique that enables much higher Q-values is the Self-Tuning technique (Figure 43 - b), [43]. In this tuning system the filter is periodically taken off-line and tuned directly by passing a known reference frequency through it or through some other suitable means. Because this technique does not depend on matching it can be used for filters with very high Q-values. The main problem is that the filter can not be used while it is being tuned. For cellular standards using time division multiple access (TDMA) or frequency hopping this is not a problem because the signal transmission is not continuous.

A third approach that allows higher Q-values and continuous operation is Adaptive Filter Tuning (Figure 43 - c), [44]. The idea is that a reference filter can be tuned with the self-tuning technique. The input signal is passed through both the main filter and the reference filter and then the two outputs are compared. The differences in the outputs is used to adjust the main filter until the outputs are identical. This scheme also implies substantially increased circuit complexity and chip area.

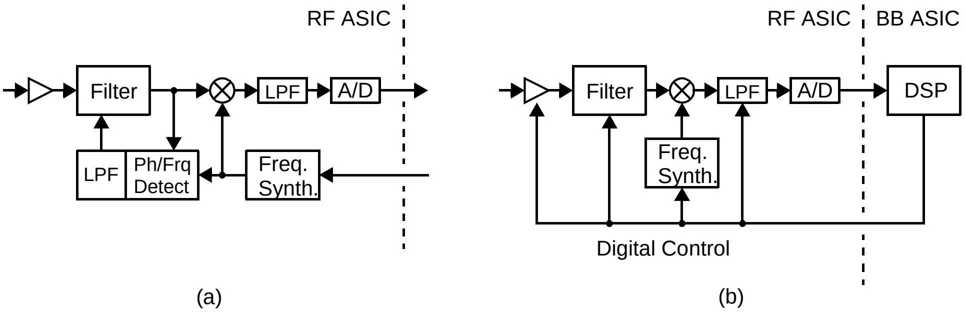


Figure 44. Frequency Tuning Schemes Using Existing Receiver Circuits. Analog or Digital PLL tuning exploiting reference frequency a in Direct Conversion Receiver (a) and DSP controlled tuning (b).

The two first tuning systems can be used for RF front-end filters while the third might be problematic because the input can not be fed to two filters without splitting the potentially very weak signal power. Note that even though it might not seem intuitive the frequency control signals to the resonators do not have to be analog signals - they could just as well be digital.

Because the RF front end filter is part of a receiver there is a lot of other circuitry implemented that could be used to simplify the tuning system. Especially the frequency synthesizer and the receiver chain are useful components. Figure 44 shows two examples of tuning systems exploiting existing circuitry. The first (Figure 44 - a) is a tuning scheme that can be used in a direct conversion receiver. It works like a combination of the master-slave and the self-tuning techniques. First the LNA is turned off and then the bandpass filter is turned into an oscillator by over-compensating the losses. This oscillator is then locked to the receiver system reference frequency in a phase locked loop. Next the control signal is sampled either in the analog or in the digital domain and then the over-compensation is removed - turning the oscillator back into a bandpass filter. Finally the LNA is turned on again. If the PLL is implemented in the digital domain then the added cost of such a tuning system can be negligible. Furthermore if the tuning element is implemented as a switched capacitor bank the frequency drift can be very low and thus it might be sufficient to perform this frequency tuning only once after which the resulting control value can be stored in a lookup table.

The second tuning scheme (Figure 44 - b) involves the receiver chain and the DSP and therefore it represents a very flexible solution with virtually no added circuit complexity - the tuning algorithm can simply be implemented in software. One way to implement the tuning is to turn the bandpass filter into an oscillator like before. Then the frequency of this oscillator or the received frequency is swept over its entire range. The setting that yields a signal at the output is stored in a lookup table. This can be repeated until the entire frequency range of the bandpass filter is recorded. An alternative approach which can be used during operation is to change the frequency control signal a little bit and note if the receiver gain drops or increases. If the gain increases then the control signal can be changed more the same way until no change is observed. If the gain decreases the control signal can be changed the opposite way until no gain change is observed. Then the filter is tuned and the frequency control setting can be recorded in a lookup table.

4.2 Analysis of Switched Capacitance Tuning of LC Resonators

When switched capacitance tuning is used a binary array of capacitors can be switched in or out of an LC tank. If a very accurate centre frequency is needed then switching can be combined with a small varactor and a simple control loop. The advantages of switched tuning, with or without a small varactor, are greatly reduced sensitivity to noise coupling through the tuning node, potentially very fast tuning perhaps through a look-up table, possibly reduced complexity and power consumption of the control loop and increased frequency tuning range. This sounds like very compelling reasons for using switched tuning but surprisingly switched tuning has seen very limited deployment [49], [52], [53].

The explanation may be partly historical because in the discrete RF realm and in RF ICs dating back only a decade the complexity level was very low and matching many capacitors and switches was considered excessively difficult or impossible if not too expensive. In modern IC processes transistors can be considered free and good matching is readily available. Furthermore digital signal processing is often used together with the RF circuits enabling programmable tuning. Therefore it seems that modern IC processes should be very well suited for switched tuning.

Another explanation why switched tuning has not gained more grounds might be that CMOS switches have had a reputation of being very lossy and therefore seriously degrade the quality of the LC tank. Driven by the enormous investments that have been made in digital CMOS processes, switches are now much better than their reputation. Today well designed switches lead to lower loss than that of varactors. Surprisingly the design and optimization of switches for tuning of LC tanks has received very limited attention in the literature.

This section presents a complete treatment of the topic. After a thorough analysis an expression is derived for the optimal switch size and the conditions leading to the highest Q-value are identified. This insight is then used to generate a comprehensive collection of design techniques. The impact of each of these techniques is modelled and it is shown that superior performance can be achieved by using these techniques together.

The common procedure for design of RF CMOS switches is to choose NMOS switches because their mobility is 3-4 times higher than that of PMOS transistors and make them as big as possible without their parasitic capacitance becoming too dominating. While this reasoning sounds intuitively satisfying, it turns out that this procedure leads to sub-optimal designs for several reasons. In this section the design of NMOS switches will be analyzed and the conditions for optimum transistor width are identified. Finally this result is used to express the best possible Q-value of the capacitive part of an LC tank.

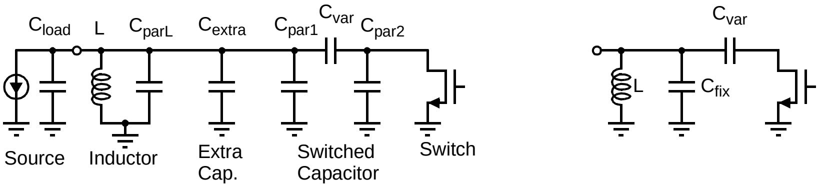


Figure 45. LC tank with parasitic capacitances (left) and equivalent model (right).

4.2.1 A Model of Switched Capacitance Tuning

This subsection gives a short introduction to LC resonators in CMOS and frequency tuning with switched capacitors. At this point only the case with one-bit of programmability is covered because once the optimal C_{var} and the optimal switch size and type are found for the one-bit case, then the solution can easily be scaled to several bits of tuning.

When an LC tank with switched capacitor tuning is connected to an oscillator, an LNA or a filter it sees a number of parasitic capacitances (Figure 45). The source, an output and any interconnect wiring will have parasitic capacitance that needs to be modelled (C_{load}). On-chip inductors, when optimized for highest Q, have a substantial amount of capacitance (C_{parL}) and therefore this contribution should also be taken into account. When off-chip inductors or bond-wire inductors are used C_{parL} also represents the parasitic capacitance of bondpads and ESD protection. To adjust the tuning range it is often desirable to add an extra high-Q capacitor (C_{extra}). When maximum tuning range is desired C_{extra} will typically be small or omitted and when smaller tuning range is needed C_{extra} can be the dominating capacitance. The capacitor that is switched in and out (C_{var}) should, as all other reactances in contact with the tank, be high-Q. There are several options in standard silicon processes. Parallel plate metal-metal or poly-poly capacitors are obvious choices. They typically have a parasitic bottom plate capacitance to ground (C_{par1}) of 5-10% of C_{var} and negligible top-plate capacitance (C_{par2}). Many different mutations of metal-metal capacitors can be designed to exploit a combination of parallel-plate, lateral flux (contributions from sidewalls) and fringing field capacitance. In some cases it is desirable for density and Q reasons to design C_{var} such that C_{par2} is not negligible. One example is when the lowest used metal layer constitutes two inter digitized comb structures. In such cases C_{par2} as well as C_{par1} generally both are in the order of 5-10% of C_{var}. In the following description C_{par2} is ignored for the sake of simplicity. If C_{var} is implemented as the later kind this will introduce a small error in the results, but in most cases this will be quite acceptable because both in the on and off states the swing across the switch is only marginally affected by C_{par2}.

By ignoring C_{par2} and by representing all the fixed capacitance contributions to ground with a single capacitor $C_{fix} = C_{load} + C_{parL} + C_{extra} + C_{par1}$ the LC tank can be modelled as in Figure 45- right.

Typically varactors when used as tuning elements are characterized by minimum Q and the ratio of the maximum capacitance and the minimum capacitance. For switched capacitor tuning an equivalent characterization would yield very high C_{max}/C_{min} ratios. In most cases these ratios lead to much higher tuning ranges than what is desired. Therefore usually a significant amount of fixed capacitance is used together with the switched capacitance. For this reason it is more convenient to focus on the achievable Q for a given frequency tuning range. Doing so also makes it much easier to compare the quality of different solutions.

If the relative desired frequency tuning range (plus/minus) is denoted α then ω_{min} and ω_{max} can be expressed as in (71).

$$\begin{aligned} \omega_{min} &= \omega_{on} = \omega_0(1-\alpha) = 2\pi f_0(1-\alpha) \\ \omega_{max} &= \omega_{off} = \omega_0(1+\alpha) = 2\pi f_0(1+\alpha) \end{aligned} \tag{71}$$

The resonance frequency of an LC tank is given by $\omega^2 = 1/LC$. Therefore the equivalent capacitance values can be expressed as (72) and (73).

$$C_{max} = C_{on} = \frac{1}{L\omega_0^2(1-\alpha)^2} \tag{72}$$

$$C_{min} = C_{off} = \frac{1}{L\omega_0^2(1+\alpha)^2} \tag{73}$$

When the switch is turned on the capacitance seen from the tank is $C_{on} \approx C_{fix} + C_{var}$ (Figure 45 - right) because the switch can be approximated with a short to ground. When the switch is turned off the capacitance seen from the tank is C_{fix} plus the series combination of C_{var} and C_j. C_j is usually much smaller than C_{var} and therefore the capacitance of the combination is approximately C_j. Further C_j is also usually much smaller than C_{fix} so $C_{off} \approx C_{fix}$. These approximations and (72) yield:

$$\frac{1}{L\omega_0^2} = (C_{fix} + C_{var})(1-\alpha)^2 = C_{fix}(1+\alpha)^2 \tag{74}$$

Which leads to the expressions for C_{fix} (75) and C_{var} (76) as functions of tuning range, frequency and inductance.

$$C_{fix} = \frac{1}{(1+\alpha)^2\omega_0^2 L} \tag{75}$$

$$C_{var} = \frac{4\alpha}{(1-\alpha)^2}C_{fix} = \frac{4\alpha}{(1-\alpha^2)^2\omega_0^2 L} \approx \frac{4\alpha}{\omega_0^2 L} \tag{76}$$

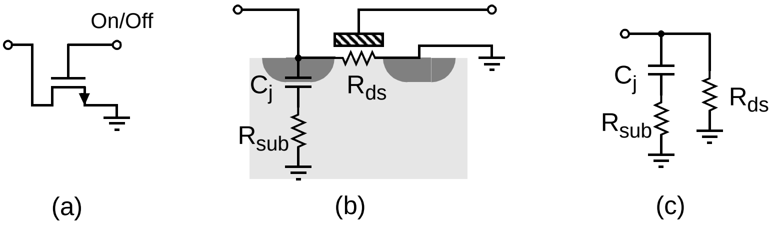


Figure 46. The MOS transistor operated as switch. Symbol (a), simplified physical model (b) and equivalent circuit diagram (c).

4.2.2 The MOS Transistor as a Switch

When an NMOS transistor is used as a switch in an LC tank a very simple transistor model is sufficient because it is never operated in saturation. The transistor can be modelled as shown in Figure 46. When the switch is turned on R_{ds} equals the drain-source resistance of the transistor operated in the triode region (= R_{on}) and when the switch is turned off R_{ds} can be considered infinite. The junction capacitance C_j of the transistor drain/source areas is lossy so even though C_j is small compared to C_{fix}, the junction capacitance needs to be modelled to be able to take this loss component into account. The loss of the junction capacitance can be modelled with a series resistance R_{sub}.

4.2.3 Q-value with Switch Turned On and Off

It is necessary to establish what is meant by optimal design. There could be several different design targets but generally what is desired is to maximize the Q of an LC tank, perhaps for a given L. If the inductor is well designed its contribution to the overall tank Q is roughly uniform over the frequency tuning range and thus the switch optimization goal reduces to the achievement of the highest possible uniform Q-value of the capacitive part of the tank.

Expressed in other terms the goal is to get the highest possible Q both at the high frequency when the switch is turned off and at the low frequency when the switch is turned on.

In the on-state the MOS transistor is conducting and therefore R_{ds} can be considered very low impedance (= R_{on}). The junction capacitance in turn is high impedance so the parallel combination is clearly dominated by R_{on} and thus the junction capacitance can be ignored in the on state (Figure 47 - left). When the switch is off, R_{ds} can be considered infinite and thus R_{ds} can be ignored. Furthermore the series combination of C_{var} and C_j is dominated by the smaller C_j (higher impedance) so C_{var} can be replaced by a short (Figure 47 - right).

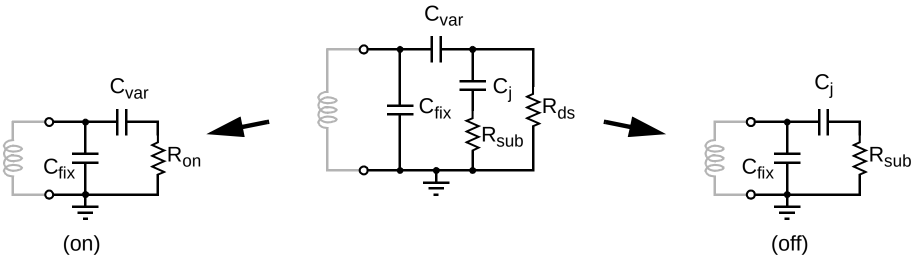


Figure 47. Capacitive part of the LC tank in the switch-on and switch-off states.

From this it is evident that both the on and off states can be modelled by the same generic structure. One capacitor C_1 in parallel with the series connection of a capacitor C_2 and a resistor R_2. Further if Q_2 denotes the Q of the series connection of C_2 and R_2 ($Q_2 = 1/(\omega R_2 C_2)$) it can be shown that the overall Q of the generic structure is given by:

$$Q = Q_2 \frac{C_1}{C_2}\left(1 + \frac{1}{Q_2^2} + \frac{C_2}{C_1}\right) \tag{77}$$

When (77) is used for the switch on-state, C_1 represents C_{fix}, C_2 represents C_{var} and R_2 represents R_{on}. With these parameters Q_2 is usually substantially larger than 5 so the $1/(Q_2)^2$ term in (77) can be excluded. This leads to the following expression for the Q of the capacitive part of the LC tank when the switch is turned on.

$$\begin{aligned} Q_{on} &\approx \frac{1}{\omega_{on} R_{on} C_{var}} \cdot \frac{C_{fix}}{C_{var}} \cdot \left(1 + \frac{C_{var}}{C_{fix}}\right) = \frac{1}{\omega_{on} R_{on} C_{var}} \cdot \left(\frac{C_{fix}}{C_{var}} + 1\right) \\ &\approx \frac{1}{\omega_0 (1-\alpha) R_{on} C_{var}} \cdot \left(\frac{(1-\alpha)^2}{4\alpha} + 1\right) = \frac{1}{\omega_0 R_{on} C_{var}} \cdot \frac{(1+\alpha)^2}{4\alpha(1-\alpha)} \end{aligned} \tag{78}$$

When the MOS transistor is used as a switch it is operated in the deep triode region and thus the drain current is given by the following expression. Note that the drain-source voltage is very small compared to the gate overdrive voltage so V_{ds}^2 can be excluded.

$$I_d \approx \mu_0 C_{ox} \frac{W}{L}\left[(V_{gs} - V_t)V_{ds} - \frac{V_{ds}^2}{2}\right] \approx \mu_0 C_{ox} \frac{W}{L}(V_{gs} - V_t)V_{ds} \tag{79}$$

$$R_{on} = \frac{V_{ds}}{I_d} = \frac{L}{\mu_0 C_{ox}(V_{gs} - V_t)} \cdot \frac{1}{W} = R'_{on} \cdot \frac{1}{W} \tag{80}$$

$$R'_{on} = \frac{L}{\mu_0 C_{ox}(V_{gs} - V_t)} = \text{normalized on-resistance} \tag{81}$$

Which leads to the final expression for the Q-value in the on-state.

$$Q_{on} = \frac{W}{\omega_0 R'_{on} C_{var}} \cdot \frac{(1+\alpha)^2}{4\alpha(1-\alpha)} \tag{82}$$

In the switch-off state (Figure 47 - left), (77) can be used with $C_{fix} = C_1$, $C_j = C_2$, $R_{sub} = R_2$ and $Q_2 = 1/(\omega_{off} R_{sub} C_j) = 1/((1+\alpha)\omega_0 R_{sub} C_j) = Q_j/(1+\alpha)$. Where Q_j is the Q of the drain/source junction capacitance at the centre frequency ω_0. The Q_j of NMOS transistors can be very low. Typical values are in the order of 2-3 at 2 GHz for a 0.5 µm transistor and therefore the $1/(Q_2)^2$ term can not be excluded. Generally C_j is much smaller than C_{fix} so instead the C_2/C_1 term can be excluded. This leads to the expression for the Q of the capacitive part of the LC tank in the off-state. Note that $C'_j = C_j/W$ is the normalized junction capacitance i.e. the junction capacitance per transistor width.

$$Q_{off} \approx \frac{Q_j C_{fix}}{W(1+\alpha)C'_j} \cdot \left(1 + \frac{(1+\alpha)^2}{Q_j^2}\right) \tag{83}$$

4.2.4 Optimal Switch Size

Because Q_{on} is proportional to the transistor width W and Q_{off} is inversely proportional to W one optimal transistor width W_{opt} exists where $Q_{on} = Q_{off}$. Any other transistor width will produce a lower Q-value either in the on- or in the off-state. This optimum can be found by simply equating the two Q-expressions and solve for W.

$$\begin{gathered} Q_{on} = Q_{off} \\ \frac{W}{\omega_0 R'_{on} C_{var}} \cdot \frac{(1+\alpha)^2}{4\alpha(1-\alpha)} = \frac{Q_j C_{fix}}{W(1+\alpha)C'_j} \cdot \left(1 + \frac{(1+\alpha)^2}{Q_j^2}\right) \end{gathered} \tag{84}$$

$$W^2 = \frac{Q_j C_{fix}}{(1+\alpha)C'_j} \cdot \left(1 + \frac{(1+\alpha)^2}{Q_j^2}\right) \cdot \frac{\omega_0 R'_{on} C_{var}}{\dfrac{(1+\alpha)^2}{4\alpha(1-\alpha)}} \tag{85}$$

To further simplify this expression there are a number of terms that can be substituted. First C_{var} can be substituted by $4\alpha C_{fix}/(1-\alpha)^2$ according to (76). Second C_{fix} can be substituted by $1/L\omega_0^2(1+\alpha)^2$ according to (75). The substitutions enable direct computation of W_{opt} from a previously determined inductance value L.

$$\begin{aligned} W^2 &= \frac{Q_j \omega_0 R'_{on} C_{fix} C_{var} 4\alpha}{C'_j} \cdot \left(\frac{1}{1+\alpha} + \frac{1+\alpha}{Q_j^2}\right) \cdot \frac{1-\alpha}{(1+\alpha)^2} \\ &= \frac{Q_j C_{fix}^2 R'_{on} \omega_0 16\alpha^2}{C'_j (1-\alpha)^2} \cdot \left(\frac{1}{1+\alpha} + \frac{1+\alpha}{Q_j^2}\right) \cdot \frac{1-\alpha}{(1+\alpha)^2} \\ &= \frac{Q_j R'_{on} 16\alpha^2}{C'_j (1+\alpha)^6 L^2 \omega_0^3} \cdot \left(\frac{1}{1-\alpha^2} + \frac{1+\alpha}{Q_j^2(1-\alpha)}\right) \end{aligned} \tag{86}$$

$$W_{opt} = \frac{4\alpha}{\omega_0 L(1+\alpha)^3}\sqrt{\frac{Q_j R'_{on}}{\omega_0 C'_j}\left(\frac{1}{1-\alpha^2}+\frac{1+\alpha}{Q_j^2(1-\alpha)}\right)} \tag{87}$$

If Q_j is expressed as:

$$Q_j = \frac{1}{\omega_0 R_{sub} C_j} = \frac{1}{\omega_0 (R'_{sub}/W)(C'_j \cdot W)} = \frac{1}{\omega_0 R'_{sub} C'_j} \tag{88}$$

where R'_{sub} is the normalized substrate resistance i.e. the substrate resistance multiplied by the transistor width, (87) can also be expressed as:

$$W_{opt} = \frac{\sqrt{R'_{on}/R'_{sub}}}{\omega_0^2 L C'_j} \cdot \frac{4\alpha}{(1+\alpha)^3} \cdot \sqrt{\frac{1}{1-\alpha^2}+\frac{(1+\alpha)\omega_0^2 R'^2_{sub} C'^2_j}{1-\alpha}} \tag{89}$$

The last term is for common values of α, R'_{sub} and C'_j in the lower GHz range between 1 and 1.1. Therefore (89) can be approximated by:

$$W_{opt} \approx \frac{4\alpha\sqrt{R'_{on}/R'_{sub}}}{(1+\alpha)^3 \omega_0^2 L C'_j} \tag{90}$$

Finally W_{opt} can be expressed in terms of C_{var} by using that, according to (76), $C_{var} = 4\alpha/((1-\alpha^2)^2\omega_0^2 L)$.

$$W_{opt} = \frac{\omega_0 C_{var}(1-\alpha)^2}{(1+\alpha)}\sqrt{\frac{Q_j R'_{on}}{\omega_0 C'_j}\left(\frac{1}{1-\alpha^2}+\frac{1+\alpha}{Q_j^2(1-\alpha)}\right)} \tag{91}$$

4.2.5 Q-value with Optimally Sized Switch

Now the optimal Q-value of the capacitive part of the LC tank can be found by simply substituting W_{opt} into one of the Q expressions.

$$\begin{aligned} Q_{opt} &= \frac{W_{opt}}{\omega_0 R'_{on} C_{var}} \cdot \frac{(1+\alpha)^2}{4\alpha(1-\alpha)} \\ &= \frac{(1+\alpha)^2 \omega_0 C_{var}(1-\alpha)^2}{\omega_0 R'_{on} C_{var} 4\alpha(1-\alpha)(1+\alpha)}\sqrt{\frac{Q_j R'_{on}}{\omega_0 C'_j} \cdot \left(\frac{1}{1-\alpha^2}+\frac{1+\alpha}{Q_j^2(1-\alpha)}\right)} \\ &= \frac{1}{4\alpha}\sqrt{\frac{Q_j(1-\alpha^2)}{\omega_0 R'_{on} C'_j} \cdot \left(1+\frac{(1+\alpha)^2}{Q_j^2}\right)} \end{aligned} \tag{92}$$

The two correction factors $(1-\alpha^2)$ and $(1+(1+\alpha)^2/Q_j^2)$ are both close to unity. Excluding these factors will for common values of α and Q_j in the lower GHz range produce an error of no more than 5-10%. If this is acceptable or if insight is the goal then the following approximation can be used.

$$Q_{opt} \approx \frac{1}{4\alpha}\sqrt{\frac{Q_j}{\omega_0 R'_{on} C'_j}} \tag{93}$$

For more insight (93) can be expressed as:

$$Q_{opt} \approx \frac{1}{4\alpha\omega_0 C'_j \sqrt{R'_{sub} \cdot R'_{on}}} \tag{94}$$

Together (93) and (94) show exactly how the basic parameters need to be optimized in order to achieve the best possible switch design.

4.3 Design Techniques for CMOS Switch Optimization

The result in (93) explains that there are five ways Q can be improved. The first way is to minimize the centre frequency of operation ω_0. The second is to minimize the tuning range α. The third is to maximize Q_j, the quality factor of the junction capacitance. The fourth way of optimizing overall Q is to minimize the normalized transistor on-resistance R'_{on} and finally the fifth is to minimize the normalized junction capacitance C'_j. At a first glance these five parameters seem to be either predetermined at the system level (ω_0 and α) or device constants (Q_j, R'_{on} and C'_j). Therefore it may seem like there is not much the designer can do to improve the Q of the capacitive part an LC tank with switched capacitance frequency tuning. In the following it will be explained why this is not the case. It is shown that for every single parameter there are techniques that can improve the Q-value.

4.3.1 Reduction of Centre Frequency

Usually the centre frequency ω_0 is predetermined at the system level and thus out of the hands of the circuit designer once the system level has been fixed. This emphasizes the importance of realizing during the system optimization that the Q of the capacitive part of the LC tank drops with ω_0 (94). Specifically this should be kept in mind when choosing intermediate frequencies or for instance when deciding whether or not to operate a VCO at twice the frequency to generate quadrature signals.

4.3.2 Minimization of Tuning Range

Also it seems that the relative tuning range 2α is predetermined at the system level. While the frequency tuning range is established at this level the actual needed tuning range might very well be substantially larger due to high tolerances and poor modelling of the inductance and capacitance values. This implies that to maximize Q the inductance value and the capacitance value must be modelled as accurately as possible. Also interconnect parasitics as well as parasitics associated with sources and loads need to be modelled accurately. Therefore if accurate modelling is used the tolerance on the centre frequency is minimized and thus 2α is minimized.

4.3.3 Maximizing the Quality of the Junction Capacitance

The quality of the junction capacitance Q_j is clearly process dependent but there are several options the circuit designer can exploit to increase Q_j. First recall that $Q_j = 1/\omega_0 R_{sub} C_j$. This shows that Q_j can be increased by reducing either R_{sub} or C_j. R_{sub} is not one clearly identifiable resistor rather it is a model of the losses in the substrate caused by the replacement currents that are coupled through C_j. Figure 48 shows the distributed nature of the substrate resistance from the bottom plate of the junction capacitance of the drain/source terminal of an NMOS transistor to the substrate contact and to the other source/drain terminal of the device. To reduce R_{sub}, as many substrate contacts as possible should be placed as close to the transistor as possible.

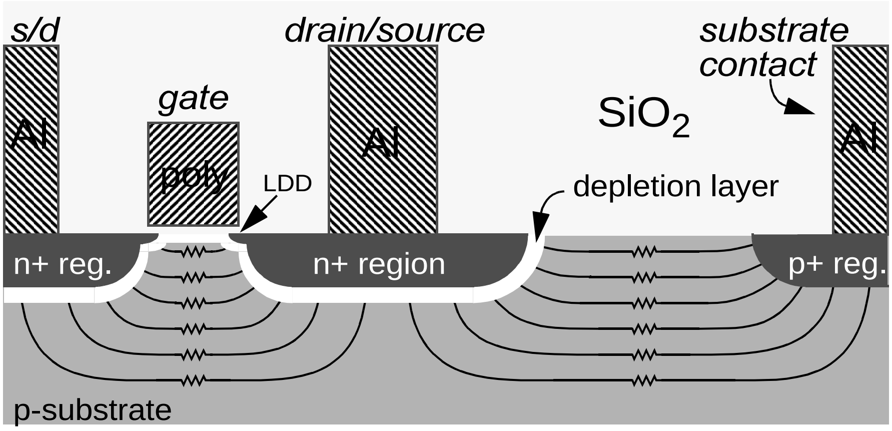


Figure 48. Cross-section of an NMOS transistor. Shows the distributed nature of the substrate resistivity R_{sub}.

In Figure 48 the relative dimensions are representative for a standard deep submicron NMOS transistor laid out with minimum sizes and minimum distances (except from the gate oxide thickness that is roughly 5 times thinner than shown). This reveals that the distances to the other drain/source terminal (left) are significantly shorter than the distances to the substrate contact (right) even when it is placed as close as possible to the transistor. Therefore resistance contributions from the first (left in Figure 48) should be the smallest and therefore the dominating ones. In other words if these contributions can be reduced, a significant reduction of R_{sub} should be expected. When the transistor is laid out with minimum dimensions it seems impossible to reduce R_{sub} further. While this is certainly true at the physical level it is possible to use topological techniques to reduce the effective R_{sub}. For instance when the LC tank is used in a differential configuration the switch can be operated in a differential mode. If the other source/drain region (left-most in Figure 48) is driven 180 degrees out of phase and with the same amplitude instead of being held at AC ground a virtual ground will be formed right under the gate (Figure 49). This will effectively reduce the common mode replacement currents and thus increase Q_j. That these effects are not pure speculation was verified in [46] where the same principle was used to increase the Q of varactor diodes made in standard CMOS. They measured close to a doubling in Q, when the differential mode was used, in good agreement with the above reasoning. Therefore differential operation should be used when it is applicable. It is difficult to estimate the actual improvement in Q-value for the drain/source junction capacitance but [46] indicates that a doubling is possible.

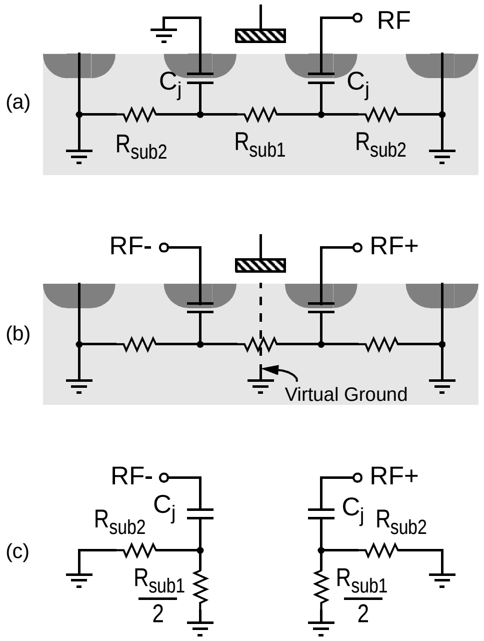
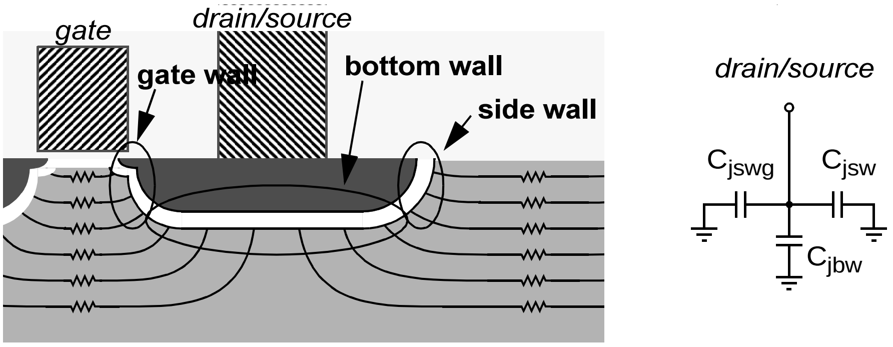


Figure 49. Simple model of an NMOS switch in the off state. (a) single-ended operation, (b) differential operation and (c) equivalent circuit diagram of differential operation.

Finally the depletion region can be expanded by applying a larger reverse bias at the drain/source terminal when the transistor is off. This leads to shorter substrate distances and thus a lower R_{sub}. The effect is a second order effect that improves with device scaling so even though it might not be very significant with present processes it might give a significant boost in coming process generations. At the present moment a more important consequence of applying a larger reverse bias is found in the Q_j boost provided by the reduction in junction capacitance. The reduction in junction capacitance can be in the order of 35-40% (see next section) and the reduction in substrate resistivity can be perhaps in the order of 5-10%. This means that Q_j improvements from applying a larger reverse bias can be up to 70-80% assuming that the bias circuit that applies the reverse bias potential contributes with negligible loss of its own.

4.3.4 Minimization of the Junction Capacitance

The junction capacitance appears in (93) both as part the normalized junction capacitance C_j' and as part of Q_j. Therefore it is very important that both the capacitance per transistor width and the capacitance per unit drain/source area is minimized. The first is strongly layout and bias dependent while the second is mainly bias dependent.



(a) (b)

Figure 50. MOS drain/source junction capacitance model. Cross section of the MOS transistor (a) and equivalent circuit diagram (b).

The junction capacitance of the drain/source region can be expressed [47] as the sum of three different capacitance contributions C_{jswg}, C_{jbw} and C_{jsw} (Figure 50). C_{jbw} represents the contribution from the bottom wall, C_{jsw} represents the contribution from the outer side wall and C_{jswg} represents the contribution from the gate side wall i.e. the wall facing the gate. The reason why the contributions from the outer side wall and from the gate side wall are modelled separately is that they may differ considerably. Often this difference is explained as a result of differences in doping level and profile. Supposedly these differences should be particularly pronounced when lightly doped drain (LDD) is used. A different explanation is given later in this chapter but for now it is sufficient to note that they are modelled as two separate parameters.

For a given technology the bottom wall capacitance is typically specified as a capacitance per unit area C''_{jbw}. For the gate wall and side wall capacitances it is more convenient to specify a capacitance per length because the capacitance per unit area is not constant with depth and further the wall is not flat but closer to cylindrical. If C'_{jswg} denotes the capacitance per length of gate wall (= transistor width W) and C'_{jsw} denotes the capacitance per length of outer side wall (= drain/source perimeter P), C_j can be expressed as (95) where A denotes the drain/source area.

$$\begin{aligned} C_j &= C_{jswg} + C_{jbw} + C_{jsw} \\ &= C'_{jswg}W + C''_{jbw}A + C'_{jsw}P \end{aligned} \tag{95}$$

$$C_j' = \frac{C_j}{W} = C'_{jswg} + C''_{jbw}\frac{A}{W} + C'_{jsw}\frac{P}{W} \tag{96}$$

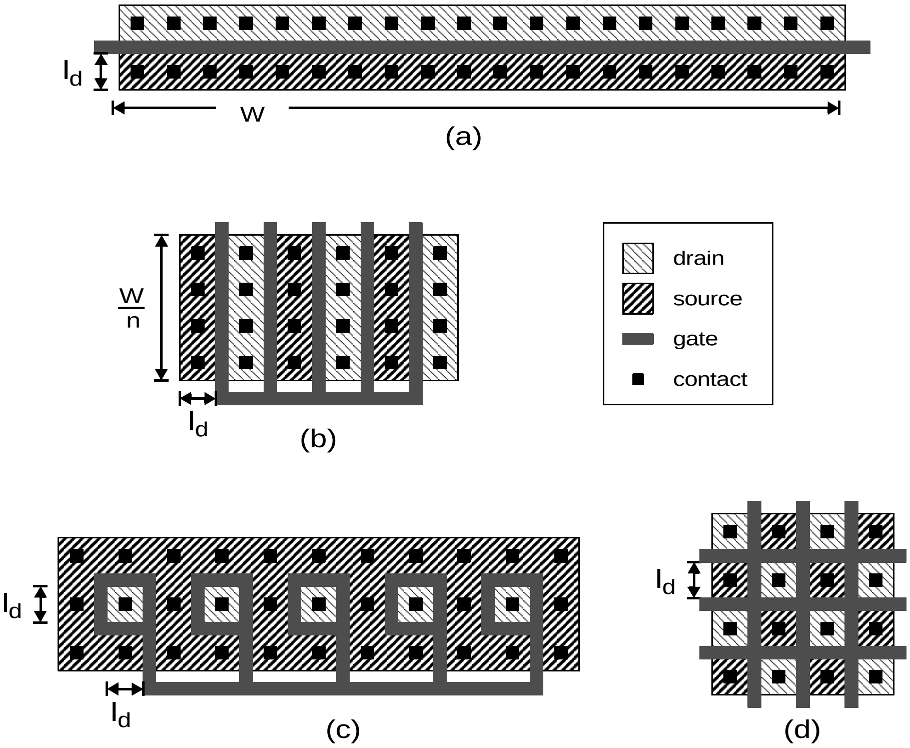


Figure 51. Transistor layout techniques. (a) simple, (b) finger/folded, (c) ring/doughnut, (d) waffle.

Which leads to the expression for normalized junction capacitance (96). Notice that C''_{jbw}, C'_{jswg} and C'_{jsw} are process and bias dependent. Thus for a given process and bias level C'_j can only be reduced by minimizing the A/W and P/W ratios.

The most basic transistor layout is shown in Figure 51 - (a). The transistor is W wide, and the length of the drain and source areas are l_d, where l_d should be chosen to be the minimum allowed distance in the used CMOS process. With these definitions C'_j for the basic structure is given by (97). Notice that the bottom wall capacitance contribution is given by $C'_{jbw}l_d$.

The most common transistor layout style is the finger transistor style (Figure 51 - b). Basically with this technique one transistor is separated into n parts that are connected in parallel. It is said that the transistor has n fingers if it has n strips of gate. The advantage of using this layout style is that the drain/source areas can be shared such that most strips of drain/source region have gate on both sides rather than on one side.

$$C'_{j\,basic} = C'_{jswg} + C''_{jbw} l_d + C'_{jsw}\left(1 + \frac{2l_d}{W}\right) \tag{97}$$

$$C'_{j\,finger} = C'_{jswg} + C''_{jbw} \frac{l_d}{2}\left(1 + \frac{1}{n}\right) + C'_{jsw}\left(\frac{1}{n} + \frac{(n+1)l_d}{W}\right) \tag{98}$$

As can be seen from the expression for C'_j for the finger transistor (98), the result is that the bottom wall contribution approaches half that of the basic structure and also the contribution from the outer side wall is reduced considerably.

While this layout technique by far is the most common it is not the one that provides the smallest drain junction capacitance per gate width. If instead of having gate on two sides of every drain contact, gate is placed on all four sides, then the effective gate width has almost doubled for roughly the same junction capacitance. This corresponds to placing rings around every drain contact (Figure 51 - c). Transistors made in this layout style are usually called ring transistors or doughnut transistors. As seen from (99) ring transistors have a bottom wall contribution of $C'_{jbw} l_d/4$, less than half the contribution provided by finger transistors and four times smaller than the basic layout style. Also the contribution from the outer side wall is completely eliminated giving ring transistors the lowest C'_j of the drain region. At the same time the ring layout style produces a huge source area and thus a large source junction capacitance. This is not a problem if there is no AC signal component on the source terminal. Therefore this layout style is probably the best when switching to AC ground, but the large source capacitance makes it unsuitable for differential operation and other operation that implies AC swing on the source terminal.

$$C'_{j\,ring} = C'_{jswg} + C''_{jbw} \frac{l_d}{4} \tag{99}$$

$$C'_{j\,waffle} = C'_{jswg} + C''_{jbw} \frac{l_d}{4}\left(1 - \frac{1}{2m_c} - \frac{1}{2m_r}\right)^{-1} + C'_{jsw}\left(2\left(\frac{1}{m_r} + \frac{1}{m_c}\right)^{-1} - 1\right)^{-1} \tag{100}$$

To circumvent this problem the rings can be placed in a waffle-like shape (Figure 51 - d). If m_r denotes the number of rows and m_c denotes the number of columns the C'_j of the waffle transistor can be expressed as (100). For higher number of rows and columns they have close to minimum C'_j and they are symmetrical devices so they are well suited for differential operation. Unfortunately the physical dimensions in modern processes are so small that the inner contacts can not be reached with diagonal metal strips as they could be in older processes. The structure is simply so dense that there is no room for interconnect between the contacts. This explains why, to the best of the authors knowledge, the waffle structure has not been used for the last 5-10 years. Even though the interconnect becomes fairly involved it is possible to lay out waffles structures with few rows and columns e.g. four rows and four columns by using several metal layers. Finally for the sake of completeness it should be mentioned that the C'_j of finger transistors can be improved somewhat by shaping the gate strips as sinusoids/square waves and skipping every other contact.

Table 4: Comparison of C'_j for different layout styles in 0.25 micron CMOS

Layout style	**Parameters**	**Equation**	**C'_j [fF/µm]**
Basic	l_d = 0.8µm, W = 19.2µm	(97)	1.47
Finger	l_d = 0.8µm, W = 19.2µm, n = 5	(98)	0.89
Ring	l_d = 0.8µm	(99)	0.42
Waffle	l_d = 0.8µm, m_r = 4, m_c = 4	(100)	0.60

The different layout styles can now be compared. Using typical numbers for a 0.25 micron process at zero reverse bias ($C''_{jbw} = 1.2*10^{-3}$F/m^2, C'_{jsw} = 0.3 fF/µm, C'_{jswg} = 0.18 fF/µm) yields the numbers given in Table 4. Note that for large waffle structures C'_j approaches that of ring transistors because the contribution from the edge of the structure becomes insignificant.

The results in Table 4 reveal that C'_j can be reduced by a factor of 2.12 if a ring style layout is used instead of the standard finger layout and by 1.48 if a waffle structure is used instead of the a finger layout. This corresponds to an overall improvement in Q according to (93) of 45% for the ring and 20% for the waffle.

As previously mentioned the junction capacitance can also be reduced by applying a larger reverse bias to the drain/source area. The capacitance is lowered because the increased voltage difference across the junction extends the depletion layer and thus minimizes the parallel plate capacitance (Figure 48). In the switch-on case this is not a good approach because it reduces the gate source voltage and thus increases the on-resistance. Therefore to take advantage of this capacitance lowering effect without worsening the switch behavior a circuit must be applied that biases the drain/source node to the highest tolerable voltage when the switch is off and to ground when the switch is on. In designing such a bias circuit it is important to make sure the bias circuit itself does not contribute with significant loss or capacitive loading.

The different parts of the junction capacitance C'_{jswg}, C''_{jbw} and C'_{jsw} are all bias dependent. They posses different bias voltage dependencies because of their different doping profiles. Generally though they are modelled with the same equation (101) - [47] where C'_{jswg0}, C''_{jbw0} and C'_{jsw0} represent the zero reverse bias capacitance, V_{biswg}, V_{bibw} and V_{bisw} represent the built-in junction potential and M_{jswg}, M_{jbw} and M_{jsw} model the behavior of the specific doping profile. Further V is the reverse bias voltage i.e. the drain/source voltage when the bulk terminal is connected to ground.

$$
\begin{aligned}
C'_{jswg}(V) &= \frac{C'_{jswg0}}{(1 + V/V_{biswg})^{Mjswg}} \\
C''_{jbw}(V) &= \frac{C''_{jbw0}}{(1 + V/V_{bibw})^{Mjbw}} \\
C'_{jsw}(V) &= \frac{C'_{jsw0}}{(1 + V/V_{bisw})^{Mjsw}}
\end{aligned}
\tag{101}
$$

For a typical 0.25 micron process (Table 6) applying a voltage of 2.5V to the drain/source region will reduce the capacitance by a factor of roughly 1.55 which leads to an increase in Q_{opt} according to (93) of 25%. At the same time Q_j is improved by an equal amount and thus contributes another 25% to the overall Q. In the 2V range the capacitance is not very sensitive so if a lower voltage turns out to be more practical it can be used without too much penalty. For instance applying a voltage of 1.8V will reduce the capacitance by a factor of roughly 1.45 which corresponds to a Q_{opt} improvement of 21% from reduced C'_j and another 21% from improved Q_j.

4.3.5 Reduction of Transistor On-Resistance

As shown earlier (80) the normalized on-resistance R'_{on} is given by (102) - repeated here for convenience. Ron is inversely proportional with $\mu_0 C_{ox}$ so a process with a high $\mu_0 C_{ox}$ should be chosen in order to improve the switch. Likewise the overdrive voltage V_{gs}-V_t should be maximized. This means that the gate voltage in the switch-on stage must be maximized. It also means that the threshold voltage should be minimized. If different transistors are available then one with a low V_t can be chosen. Also in current deep submicron CMOS processes V_t is dependent on the transistor length. Fortunately V_t is minimized by minimizing the transistor length L which is already desirable due to its presence in the nominator. Further notice that L represents not the drawn transistor length but the effective transistor length. For 0.25 micron CMOS processes the minimum effective transistor length is usually in the order of 0.20 micron.

$$
R'_{on} = \frac{L_{eff}}{\mu_0 C_{ox}(V_{gs} - V_t)} \tag{102}
$$

It may seem like maximizing V_{gs} and minimizing L is all that can be done to reduce the normalized on-resistance once a process has been chosen. This is true if the LC tank is operated in a single ended configuration. If instead the LC tank is operated in a differential configuration a virtual ground can be exploited to reduce the effective on-resistance by a factor of two. Therefore the effective on-resistance in the differential mode is given by (EQ 103).

$$
R'_{ondiff} = \frac{L_{eff}}{2\mu_0 C_{ox}(V_{gs} - V_t)} \tag{103}
$$

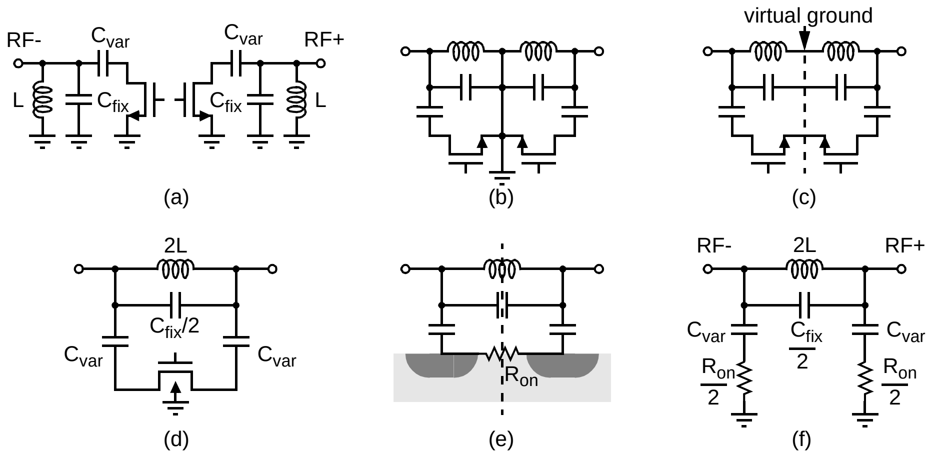


Figure 52. Translation from single-ended to differential operation. (a) two single-ended switch structures, (b) equivalent schematic, (c) virtual ground - electrically equivalent in differential operation, (d) two minimum length transistors replaced by one minimum length, (e) effective transistor length and on-resistance thereby cut in half, (f) electrically equivalent schematic.

Figure 52 explains how the effective on-resistance can be halved by using a virtual ground in a differential configuration. Figure 52 - (a) shows the basic LC tank structure with a switched capacitance - only now it is duplicated to allow for differential operation. Each side behaves just like the single-ended case that has been used as a basis for the analysis. Figure 52 - (b) is the same schematic rewritten so it shows the symmetry of the structure. In Figure 52 - (c) the connection to ground has been removed. Because of the symmetry of the structure the AC component of the nodes along the dashed line is zero. In other words removing the ground connection does not change the electrical behavior provided that the two transistors are biased the same way. In Figure 52 - (d) the inductors and C_{fix} capacitors are collapsed. The electrical behavior seen from the RF terminals is still not changed. When the transistors are collapsed to one, the transistor length should be $2*L_{min}$ to give the same electrical behavior but if L_{min} is maintained the on resistance will appear as if it was cut in half (Figure 52 - (e)). Seen from each of the RF terminals (Figure 52 - (f)) the only electrical change appears to be the R_{on} cut in half. Therefore the analysis developed applies equally well in the differential LC tank case provided that (103) is used instead of (102). Note that this reasoning assumes that C_j is not increased by differential operation. For the simple model (Figure 50, [47]) this is certainly the case but for an actual transistor the distributed effects may come into play. If this is the case the transistor width must be reduced and the improvement becomes less than a factor of two.

4.3.6 Summary

In Section 4.2.5 on page 87 the expressions for optimum Q were found ((92) - (94)). They show that five parameters ω_0, α, Q_j, C'_j and R'_{on} determine the optimum Q of the

Table 5: Q optimization techniques for switched capacitance tuning of LC tanks

Section	**Technique**	**Effect**	**Q - improvement according to (93)**
4.3.1	Topological choices	ω_0 ▼	0 - 100%
4.3.2	Accurate inductance modelling	α ▼	~ 25%
	Accurate capacitance modelling	α ▼	~ 25%
	Choose capacitors with low process tolerance	α ▼	~ 25%
4.3.3	Differential operation	Q_j ▲	~ 40%
	Layout - many substrate contacts	Q_j ▲	---
	Increase reverse bias in the switch-off state	Q_j ▲	25% - 30%
4.3.4	Layout - minimize drain/source dimensions	C_j' ▼	---
	Layout - waffle/ring style	C_j' ▼	20 - 45%
	Increase reverse bias in the switch-off state	C_j' ▼	20 - 25%
4.3.5	Layout - minimum gate length	R'_{on} ▼	---
	Differential operation	R'_{on} ▼	~ 40%
	Maximize overdrive voltage in switch-on state	R'_{on} ▼	---

capacitive part of LC tanks with switched capacitance tuning. In this section each of these parameters has been thoroughly studied and several techniques and circuit topologies were identified that can be used to optimize the overall Q. Table 5 summarizes these techniques. The right most column shows estimates of how much Q_{opt} can be improved over a reference switch by applying each of these techniques. The reference switch is a traditional finger transistor with minimum gate length in a single ended configuration. Further drain/source regions are biased at ground, maximum gate voltage is applied and many substrate contacts are used. These techniques are considered to be standard techniques so the right most column indicate these with three dashes. If the standard techniques are not used it may result in a significant penalty in Q_{opt}. The remaining techniques are considered to be non-standard. The actual numbers were to the extent possible elaborated on in the individual sections. Note that the estimates for improvement by using differential operation are very approximate.

4.4 The Differential Waffle Switch

This section presents the efforts made to combine all the loss reduction techniques identified in the previous section. The result is a Differential Waffle Switch as well as an Enhanced Differential Waffle Switch. After the description of these switches they are compared to the switch types which are normally used.

4.4.1 Functionality of the Differential Waffle Switch

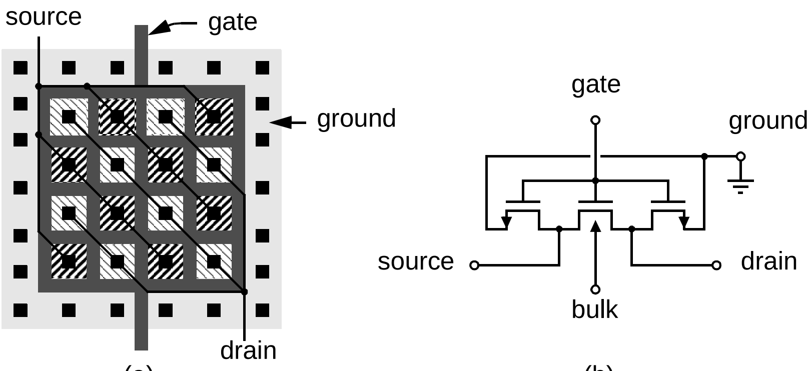
The previous section explains in detail how Q_{opt} can be improved. Specifically it is noted that the differential structure can be used to reduce the switch on-resistance (R'_{on}) and to improve the Q of the junction capacitance (Q_j). Further a ring shaped layout style can be used to minimize the amount of junction capacitance per transistor width (C'_j). To the best of the authors knowledge the combination of these two techniques has never been used for frequency tuning of LC resonators. The basic ring shape is as mentioned earlier not suited for differential operation because of its excessive source area. Instead a waffle shaped structure, a different way of laying out transistor rings, must be used (Figure 51 - d). Therefore the combination of these two techniques leads to a *Differential Waffle Switch*.

4.4.2 Enhanced Differential Waffle Switch

For large waffle structures i.e. many rows and columns, the junction capacitance per transistor width approaches that of isolated rings. Unfortunately the physical dimensions in modern CMOS processes only allow for a very limited number of rows and columns and thus C'_j is significantly higher than that of isolated rings (Table 4). Part of the reason is that each drain/source contact does not have gate on all four sides (Figure 51 - d). If gate is placed on all sides of each contact what terminal should the outside region be? It can not carry any signal as it would add significant parasitic capacitance and loss. Instead it can be connected to ground. Doing so leads to a greatly *Enhanced Differential Waffle Switch* because it:

- Reduces the on-resistance of the switch by providing an extra path to ground.
- Ensures that the drain/source regions are biased at ground when the switch is on.
- Minimizes the junction capacitance because C'_{jswg} is generally smaller than C'_{jsw}.
- Improves the Q_j because R_{sub} is reduced by the proximity of a AC ground terminal.

The layout of such a switch is shown in Figure 53 - (a). Effectively this is a five terminal device where the fifth terminal is the ground ring that surrounds the waffle. The electrical equivalent is shown in Figure 53 - (b) where it becomes clear that the switch is composed of three parts. First a transistor part that is operated differentially - this is the original waffle switch - and then two transistor parts that work as switches to ground and therefore are operated in a single-ended mode.



(a) (b)

Figure 53. Enhanced Differential Waffle Switch. Ground ring as 5th terminal. (a) layout, (b) circuit diagram.

The fact that this device works in a combined differential and single-ended manner implies that neither (EQ 102) nor (EQ 103) can be used to describe the on-resistance of the switch without proper modification of the model.

Because of the targeted differential operation (EQ 103) is chosen as the most natural choice to describe the on-resistance. The implication is that the on-resistance of the part of the transistor that goes to ground is artificially doubled. The correct on-resistance can be achieved if only half of the width of this part is accounted for. Thus the effective transistor width seen from one side is $W_{effective} = W_{differential} + W_{single\text{-}end}/2$. $W_{differential}$ is the width of the middle transistor in Figure 53 - (b) and $W_{single\text{-}end}$ is the width of each of the two remaining transistors. Now there is a difference between the actual transistor width seen from one side $W_{actual} = W_{differential} + W_{single\text{-}end}$ and the effective transistor width $W_{effective}$. Therefore the generic expression for normalized junction capacitance (EQ 96) must be modified to (EQ 104).

$$C_j' = \frac{C_j}{W_{effective}} = C'_{jsw}\frac{W_{actual}}{W_{effective}} + C''_{jbw}\frac{A}{W_{effective}} + C'_{jsw}\frac{P}{W_{effective}} \qquad (104)$$

For the Enhanced Differential Waffle Switch, $W_{actual} = 2*l_d*m_r*m_c$, $W_{effective} = 2*l_d*m_r*m_c - ld*(m_r+m_c)/2$, $A = l_d^2*m_r*m_c/2$ and $P = 0$. Where l_d continues to denote the minimum drain/source width and where m_r and m_c continue to represent the number of rows and columns.

$$\begin{aligned} C'_{j\,edw} &= C'_{jswg}\left(1-\frac{1}{4m_c}-\frac{1}{4m_r}\right)^{-1} + C''_{jbw}\frac{l_d}{4}\left(1-\frac{1}{4m_c}-\frac{1}{4m_r}\right)^{-1} \\ &= \left(C'_{jswg}+C''_{jbw}\frac{l_d}{4}\right)\left(1-\frac{1}{4m_c}-\frac{1}{4m_r}\right)^{-1} = C'_{j\,ring}\left(1-\frac{1}{4m_c}-\frac{1}{4m_r}\right)^{-1} \end{aligned} \qquad (105)$$

Table 6: Typical 0.25 micron NMOS parameters

Parameter	BSIM3	typical value
C''_{jbw0}	CJ	$1.2*10^{-3}$ F/m^2
C'_{jsw0}	CJSW	$3.0*10^{-10}$ F/m
C'_{jswg0}	CJSWG	$1.8*10^{-10}$ F/m
M_{jbw}	MJ	0.4
M_{jsw}	MJSW	0.24
M_{jswg}	MJSWG	0.24
V_{bibw}	PB	0.85V
V_{bisw}	PBSW	0.85V
V_{biswg}	PBSWG	0.85V
$\mu_0 C_{ox}$	---	250 µA/V^2
Q_{j0} @ 2.0GHz	---	6
l_d	---	0.8 µm

This leads to the expression for the normalized junction capacitance of the enhanced differential waffle switch $C_j'_{edw}$ (105). Notice that it is proportional to the capacitance of the individual ring structure $C_j'_{ring}$. For a waffle with four rows and four columns for instance the difference is 8/7 i.e. it is only 14% larger than the capacitance of individual rings. For zero reverse bias this gives $C_j'_{edw}$ = 0.48 fF/µm which is substantially better than the 0.60 fF/µm of the basic waffle structure (Table 4).

4.4.3 Comparison of Switches

To be able to compare the different types of switches it is necessary to establish a set of numbers for a reference technology. A standard 0.25 micron CMOS process is chosen as reference. Usually CMOS processes from the same generation (same minimum transistor length) do not differ much in behavior but the modelling of especially the drain/source junction capacitances is often poorly executed or perhaps more correctly not a priority. Therefore the model parameters of several different 0.25 micron processes were compared and used to convey a set of “typical” parameters. These parameters are presented in Table 6. Some parameters, especially the Q-value, would benefit from a more comprehensive set of data but the numbers given are good enough for the objective of comparing the different switch topologies.

When comparing different topologies it is always difficult to ensure an actual apple-to-apple comparison. In this case the comparison is not particularly difficult, but it would be relatively comprehensive because of the many different combinations of loss-reduction techniques that can be created. Fundamentally most of these techniques are independent of each other and therefore the merits of a specific technique can be estimated by considering Table 5 together with Table 4.

Here a comparison is not made between the individual techniques. Instead a comparison is made between five different configurations. One configuration (A) represents the un-optimized case, two configurations (B and C) represent optimized cases that are considered typical and by many considered to be the state-of-the-art and finally two configurations (D and E) are presented that use all the techniques presented in this work.

All five configurations work at 4GHz. Furthermore all five use many substrate contacts, minimum drain/source dimensions and minimum gate length. The un-optimized configuration (A) is a basic transistor layout. It assumes that the inductor and the fixed capacitors are designed using standard inductance formulas and the capacitance numbers given in a process description. This leads to large tolerances in the centre frequency which is why the switch is configured with a relatively high α of 0.2 (40% tuning range). Furthermore a V_{dd} and V_{gs} of 1.8 volts is used because it is typical in recent RF CMOS designs.

Configuration B uses the isolated ring layout technique to reduce the amount of lossy drain/source junction capacitance. This is the type of switch that is used in [49]. Furthermore it is assumed that the inductors are designed either with the latest analytical expressions [50] or with inductance simulation tools like ASITIC [51]. Therefore the tolerances in centre frequency are lower and thus the switch is configured with a lower α of 0.15 (30% tuning range). Also a higher V_{dd} and V_{gs} of 2.5 volts is assumed either by accepting higher power consumption of other circuitry or by generating a higher local voltage for this particular use.

Configuration C uses the same setup as configuration B but it is used in a differential structure to take advantage of virtual grounds to lower the effective on-resistance and to increase Q_j by lowering the effective R_{sub}. Because the individual ring layout style is incompatible with differential operation configuration C uses a finger layout technique instead.

Configuration D uses an enhanced differential waffle switch layout to take advantage of both minimum junction capacitance and differential operation. Further it is assumed that, along with the techniques used for configuration B and C, all parasitics in the layout including interconnect capacitance and inductance are modelled and perhaps even simulated with EM field solvers. This brings down the tolerances a little more which is why an α of 0.125 (25% tuning range) is chosen.

Table 7: Comparison of different switches. Values calculated with (93).

	Switch type and configuration	**f_0** **[GHz]**	α	**Q_j**	**C'_j** **[fF/µm]**	**R'_{on}** **[Ω∗µm]**	**Q_{opt}**
A	Basic. $V_{dd} = V_{gs} = 1.8$ volt. $V_{rb} = 0$ volt.	4.0	0.2	3	1.47	620	15
B	Single-Ended Ring. $V_{dd} = V_{gs} = 2.5$ volt. $V_{rb} = 0$ volt.	4.0	0.15	3	0.42	400	47
C	Differential Finger. $V_{dd} = V_{gs} = 2.5$ volt. $V_{rb} = 0$ volt.	4.0	0.15	6	0.89	200	61
D	Enhanced Diff. Waffle Switch. $V_{dd} = V_{gs} = V_{rb} = 1.8$ volt.	4.0	0.125	9	0.33	310	118
E	Enhanced Diff. Waffle Switch. $V_{dd} = V_{gs} = V_{rb} = 2.5$ volt.	4.0	0.125	10	0.31	200	160

For this configuration a V_{dd} and V_{gs} of 1.8 volts is chosen to demonstrate its performance at the more typical voltage of 1.8 volts. Finally the switch is configured with a bias circuit that pulls the drain/source terminals up to V_{dd} only when the switch is in the off-state. Doing so without adding parasitic capacitance and loss is not completely trivial but it is possible. Curiously this does not seem to be a standard technique which is why it is not used in configuration B and C.

Configuration E uses all the bells and whistles presented in this text. The only difference from configuration D is that V_{gs} in the on state and $V_{drain/source}$ in the off-state are chosen to be 2.5 volts. Again these bias voltage could be generated on-chip using DC-DC converters. Higher than 2.5 volts should not be used because it is considered the maximum operating voltage the 0.25 micron transistors can tolerate.

The combined effects of the techniques used in these five configurations can be seen in Table 7. It shows that configuration B and C with Q-values in the 50's at 4.0 GHz with 30% tuning range provide roughly three times better Q than the perfectly sized but otherwise un-optimized case (A). These numbers are very good and therefore they show that switched frequency tuning is a serious alternative to varactor tuning.

Using the enhanced differential waffle switch along with the other techniques presented here (configuration E) brings the Q-values up to 160 which is outstanding. It is three times better than configuration B and C and an order of magnitude better than the un-optimized case. Even with a lowered supply of 1.8 volts (configuration D) the Q-value of 118 is still twice as good as configuration B and C.

4.5 A Note on PMOS Switches and Multi Bit Tuning

Generally it is assumed that PMOS transistors are not suitable for RF switches because of their high on-resistance. Typically the drain-source resistance of a PMOS transistor in the triode region (the on-resistance) is three to four times higher than the on-resistance of an equivalent NMOS transistor. This implies that a PMOS switch would be 3-4 times more lossy than a similar NMOS switch. Based on the analysis presented here it is obvious that this is not necessarily the case. According to (94) what matters is the geometric mean of the normalized on-resistance and the normalized substrate resistance $\sqrt{R'_{sub} \cdot R'_{on}}$.

The N-well, PMOS transistors are placed in, usually has a higher conductivity than the epi-layer or the bulk substrate the NMOS transistors are placed in. Therefore the R'_{sub} of the PMOS transistors can be lower than the R'_{sub} of NMOS transistors. This means that $\sqrt{R'_{sub} \cdot R'_{on}}$ for PMOS transistors could be lower than for NMOS transistors. The result is that PMOS switches, when sized correctly, theoretically could yield better performance than NMOS switches.

Naturally the PMOS transistors must be in the order of five times wider than the NMOS transistors leading to significantly higher parasitic capacitances. In some cases this may not be desirable. Also it means that the approximations used in the analysis presented in Section 4.2.3 on page 84 are not valid for the PMOS case, or more precisely, it means that even though the trends are the same, the results presented in (87) and (92) pose much greater inaccuracies and should therefore be used with caution. Except from this all the other issues that were treated and all the loss reduction techniques that were presented apply equally well for the PMOS case. I.e. the enhanced waffle switch is also believed to be the best solution when using PMOS transistors.

The analysis and design methodology presented in this work only deals with the design of switches for one-bit tuning. In other words only one capacitance C_{var} is switched in or out of the LC tank leading to only two resonance frequencies $f_{on} = f_{min}$ and $f_{off} = f_{max}$. In many applications it is highly desirable to have several bits of tuning i.e. to be able to set several resonance frequencies between f_{min} and f_{max}. The design of such an N-bit tuning system is straightforward once W_{opt} has been found using (87) or (91). C_{var} is simply split into 2^N parallel unit capacitors $C_{unit} = C_{var} / 2^N$ and each of these is connected with at unit transistor switch of size $W_{unit} = W_{opt} / 2^N$. If this is done then the unit capacitors can be switched in or out of the LC tank in a binary array. Because the switches usually get poorer when they get small the minimum tolerable switch size usually sets the maximum number of bits to somewhere in the 5-7 bit range. Beyond this range matching gets very tricky so even if more bits are desired it might be very difficult to achieve such resolution in the lower GHz range.

An alternative solution that takes care of this is to implement the frequency tuning as a combination of a low resolution, high-Q switch capacitor bank (e.g. 3-4 bits) for coarse tuning and a small high resolution switched capacitor bank for fine tuning.

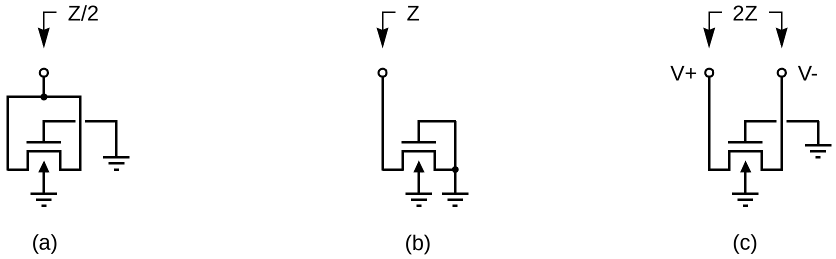


Figure 54. Different Setups for Measuring the Impedance of the Drain/Source Junction Capacitance (Z). Drain and Source in common mode configuration (a), One Drain/Source terminal grounded (b) and Drain and Source in differential mode configuration (c).

4.6 Preliminary Measurements

To be able to optimize an RF switch for frequency tuning of LC resonators it is necessary to know the quality factor of the drain/source junction capacitor Q_j at the desired frequency. The natural approach would be to measure the junction capacitance of a test transistor in the given process. Unfortunately these measurements are complicated by the presence of the low-loss gate-source overlap capacitance that has a capacitance which is comparable to the junction capacitance. The measurements are further complicated by the proximity effects under the gate and the distributed nature of the substrate resistance. These effects lead to significantly different results depending on measurement setup. Figure 54 shows three different methods for measuring the junction capacitance. In the first setup (a) the drain and source terminals are connected and measured together while the gate is grounded. In this common mode setup the substrate replacement currents are forced to flow relatively far and therefore the junction capacitance is expected to have a low Q-value. When the measurement is made in differential setup (c) the replacement currents only need to flow a minimal distance and thus the Q-value is expected to be high. A single-ended measurement (b) can be viewed as a combination of the two so the Q-value should be somewhere in between the Q-values observed for the configurations (a) and (c).

Erik Pedersen from Aalborg University has made some measurements of the drain/source junction capacitance of different finger layouts measured in the configuration (a). For a 0.5 micron CMOS process he found the Q-value to be in the order of 2 at 2.0 GHz and for a 0.25 micron CMOS process he found the Q-value to be in the order of 6 at 2.0 GHz. These results were used for the comparisons of the switches presented so far.

A number of test structures were fabricated for this project. First 20 minimum size NMOS ring transistors fabricated in a 0.25 micron CMOS process were measured in configuration (b). The preliminary results are that the capacitance was approximately 0.75 fF/µm gate and that the Q was approximately 40 at 2.0 GHz. Roughly half of this capacitance is expected to be overlap capacitance so this should correspond to a Q-value for the drain/source junction capacitance itself of approximately 20 at 2.0 GHz.

The impedance of the junction capacitance under differential operation i.e. configuration (c) can be extracted from a two-port measurement (see the treatment on balanced inductors). Finger transistors and enhanced differential waffle switches were fabricated in the same 0.25 micron CMOS process. The 2-port measurements of these structures were then converted into one differential impedance. The junction and overlap capacitances were as expected but the measurements were not accurate enough to give a reasonable prediction of the Q-value at lower GHz frequencies but the indications were that the losses were significantly lower than what was measured for the ring transistors in configuration (b).

These measured Q-values were significantly higher than expected. This emphasizes that switched tuning is very attractive also for applications that at present are considered impossible by many. These applications include combinations of low voltage operation (< 1.5V), high frequency (> 5GHz), large tuning range (> 50%) and cheap CMOS (> 0.5 μm). Based on these measurements it is possible to make some conclusions:

A) Q-values scale rapidly with technology development, B) Common mode replacement currents dominate losses and C) Common mode replacement currents can be almost eliminated with finger or waffle switches operated in differential mode.

It seems that as all dimensions are dropping with technology scaling so do the series resistances and thereby the drains/source junction capacitance Q-values increase. Curiously this might hold the answer to the mysterious excess noise observed in MOS transistors used for LNA designs. In early papers [54] the noise factor was estimated to be 2-3 (in a 0.7 μm process) and these values have been dropping ever since and are now approaching the theoretical value of 2/3 predicted by Van der Ziel. Usually this excess noise has been blamed on hot electron effects but if this were true then the noise factor should increase with smaller transistors - not decrease. If instead this excess noise was caused by the losses of the drain/source junctions the excess noise would decrease with smaller transistor sizes as observed.

Another curious observation is that the short gate length implies a low impedance connection between the gate wall capacitance contributions. This should lead to high Q-values for the gate wall capacitance contributions in good agreement with the differential measurements. This also implies that if the drain/source junction capacitance is characterized with a measurement in configuration (b) then what is seen at the gate wall is not the gate wall capacitance but the gate wall capacitance in series with the gate wall capacitance of the other drain/source junction i.e. roughly half the gate wall capacitance. Curiously the gate wall capacitance is often extracted to be roughly half the outer sidewall capacitance. Often this difference is explained as coming from differences in the doping profiles but it might be simpler than that - namely that what is observed is the series connection of the two gate wall capacitances. If this is true then the model parameters are often extracted with wrong values for the gate wall capacitance.

4.7 Summary

This chapter first gives an introduction the field of frequency tuning. Both tuning elements and tuning systems are covered. Then an analysis of switched capacitance tuning of LC resonators is presented. Based on this analysis, symbolic expression are derived for the optimal width of a switch transistor and for the optimal Q-value of the capacitive part of the resonator. It is shown that the Q value of an optimal sized switch is inversely proportional to the tuning range 2α, the centre frequency ω_0, the junction capacitance C'_j, the square root of the substrate resistance R'_{sub} and the square root of the transistor on-resistance R'_{on}. Therefore minimizing each of these parameters separately will lead to the best possible overall Q value.

The nature of each of these five parameters is explained and several suggestions are given on how to minimize each of the parameters. Based on these observations it is clear that it is important to have transistor gate on all sides of the drain/source contacts and important to use a differential structure. The traditional ring structure is not suited for differential operation. By using a waffle layout in a differential coupling (a differential waffle switch) both objectives can be met simultaneously leading to superior performance.

Due to the physical dimensions of modern processes only small waffle structures can be laid out. These have higher normalized junction capacitance because many contacts are not completely surrounded by transistor gate. This problem can be solved in part by adding a fifth terminal - a ground ring - to the waffle structure. This brings the normalized capacitance close to the value of large waffle structures and individual rings while retaining the symmetry and thus the applicability of differential operation. Furthermore this enhanced differential waffle switch provides a very convenient way of biasing the structure.

Using the techniques presented for frequency tuning of LC resonators leads to superior designs that can easily compete with the Q-values provided by the best varactors available [46]. For instance Q values in the order of 160 can be achieved for a 25% tuning range at 4GHz. At the same time the use of switches leads to a much more linear capacitance and a significant reduction in sensitivity to noise injection from tuning nodes. Furthermore tuning by switching is ideal for programmability of RF circuits.

Based on these results it is concluded that switched frequency tuning of LC resonators is likely to get a major impact on future RF systems-on-chip either alone or in combination with a smaller continuous tuning element like a varactor. Furthermore it is concluded that the Enhanced Differential Waffle Switch will be very useful for these applications.

4.8 References

[38] K. Christensen, “Design and Optimization of CMOS Switches for Switched Tuning of LC Resonators”, in *Proceedings of the IEEE Int. Symp. on Circuits and Systems*, vol. 4, pp. 770-773, May 2001.

[39] W. B. Kuhn, "Design of Integrated, Low Power, Radio Receivers in BiCMOS Technologies", *Ph.D. Dissertation, Virginia Polytechnic Institute and State University*, Blacksburg, Virginia, 1995.

[40] R. Duncan, "Active-RLC Filters in Silicon", *Ph.D. Dissertation, University of Toronto*, 1995.

[41] R. Castello, I. Bietti, F. Svelto, “High-Frequency Analog Filters in Deep-Submicron CMOS Technology” *in ISSCC Digest of Technical Papers*, Feb. 1999, pp. 74-75.

[42] K-S. Tan and P.R. Grey, "Fully Integrated Analog Filters Using Bipolar-JFET Technology," *IEEE J. Solid-State Circuits*, vol. SC-13, no. 6, pp. 814-821, December 1978.

[43] Y. Tsividis, "Self-Tuned Filters," *Electronic Letters*, pp. 406-407, June 1981.

[44] K.A. Kozman et al. "Automatic Tuning of Continuous-Time Integrated Filters Using an Adaptive Filter Technique," *IEEE Trans. Circuits Systems*, 11, pp. 1241-1248, 1991.

[45] E. Pedersen, “Performance Evaluation of CMOS Varactors for Wireless RF Applications”, *Proceedings of the 17th NORCHIP Conference*, pp. 73-78, November 1999.

[46] A.-S. Porret et al., “Design of High-Q Varactors for Low-Power Wireless Applications using a Standard CMOS Process”, *IEEE Journal of Solid-State Circuits*, Vol. 35, pp. 337-345, March 2000.

[47] Y. Tsividis, “Operation and Modelling of The MOS Transistor”, *McGraw-Hill*, 1999.

[48] C. Enz, “MOS Transistor Modelling for RF Integrated Circuit Design”, *Proceedings of the Custom Integrated Circuits Conference*, pp. 189-196, 2000.

[49] A. Kral et al., “RF-CMOS Oscillators with Switched Tuning”, *Proceedings of the Custom Integrated Circuits Conference*, pp. 555-558, 1998.

[50] S. S. Mohan et al., “Simple Accurate Expressions for Planar Spiral Inductors”, *IEEE Journal of Solid-State Circuits*, Vol. 34, pp. 1419-1424, October 1999.

[51] ASITIC homepage, http://formosa.EECS.Berkeley.EDU/~niknejad/asitic.html

[52] W. Kuhn, et al., “A 200 MHz CMOS Q-Enhanced LC Bandpass Filter”, *IEEE Journal of Solid-State Circuits*, Vol. 31, No. 8, pp. 1112-1122, August 1996.

[53] T. Cho, et al. “A Single-Chip CMOS Direct-Conversion Transceiver for 900MHz Spread-Spectrum Digital Cordless Phones”, *Proceedings of the IEEE International Solid-State Circuits Conference*, pp. 228-229 & 464, 1999.

[54] A.A. Abidi, "High-Frequency Noise Measurements on FET’s with Small Dimensions," *IEEE Transactions on Electron Devices*, vol. ED-33, no.11, pp. 1801-1805, Nov. 1986.

Chapter 5: Quality Factor Tuning

The design of low noise input stages, on-chip LC resonators and frequency tuning has been covered in the past chapters. This chapter explains how loss compensation or equivalently negative resistance circuits can be designed such that they can be used for Q-enhancement of LC resonators.

The first section presents different circuit topologies that provide a negative impedance which can be used for loss compensation. The most common options are feedback to the source terminal combined with impedance transformation and cross coupled feedback to the gate terminal.

The next section explains how a large resonator voltage swing can be enabled without jeopardizing sensitivity. More specifically a level shift realized with a source follower or an AC coupling in the feedback path enables a single sided peak tank voltage swing of one third of the supply voltage.

Then the noise of a differential Q-enhanced LC resonator with capacitive division in the feedback path is modelled and it is shown that in order to minimize the noise of a Q-enhanced resonator it is important that the uncompensated tank quality factor is maximized, that the Q-enhancement factor is minimized and that the capacitive division in the feedback path is minimized.

After this it is explained how the negative impedance of the loss compensation circuits can be tuned. The traditional method of varying the bias current is not well suited for high overdrive voltages in deep submicron CMOS. Instead varying the capacitance in the feedback path or switching unit transistors in or out of the loss compensation circuit may be used.

Finally Q-tuning systems are considered. Many of the traditional tuning topologies which were presented in Chapter 4 - Frequency Tuning can also be used for Q-tuning with little or no additional circuitry.

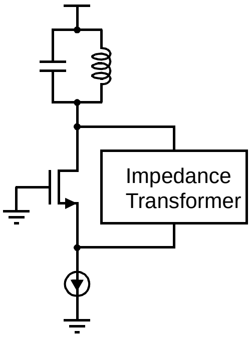


Figure 55. Loss compensation through feedback to the transistor source terminal. An impedance transformer in the feedback path reduces the 1/gm loading of the resonator.

5.1 Feedback Techniques for Loss Compensation

Simple Q-Enhanced LC filters are closely related to LC oscillators despite the fact that filters have both an input and an output and are not producing a reference tone. VCOs and RF filters share many of the same circuit level trade-offs. Both structures are frequency selective in nature and both need to deal with signals with very large voltage swing. At the same time for high performance systems it is imperative that an absolute minimum of noise is added. Furthermore it is desirable that the oscillator is linear to reduce up- or down-conversion of noise while it is necessary for the filter to avoid intermodulation. These simultaneous demands are very challenging so it is tempting to assume that the best performance is achieved through the use of simple circuit structures optimized to the limit. This is exactly the lessons learned from the intensive research carried out in integrated LC oscillator design. Therefore it is natural to start the search for good loss compensation circuits for the Q-enhanced LC filters amongst the LC oscillator loss compensation techniques.

5.1.1 Negative Resistance Circuits

One of the most basic ways of providing loss compensation for a resonator is through feedback around a single transistor. Typically the resonator signal is fed back to the source terminal of the transistor. This means that the tank is loaded with 1/gm which is unfortunate because the tank should have the highest possible uncompensated Q-value. Therefore it is common to provide some sort of impedance transformation between the tank and the source terminal (Figure 55) to reduce the loading of the resonator.

One of the most popular means of impedance transformation used in the feedback is capacitive division. Oscillators with capacitive division in the feedback path are generally known as Colpitts oscillators. Though originally referring to the single ended case it is possible to use the same approach for balanced oscillator structures as shown in Figure 56 - (a).

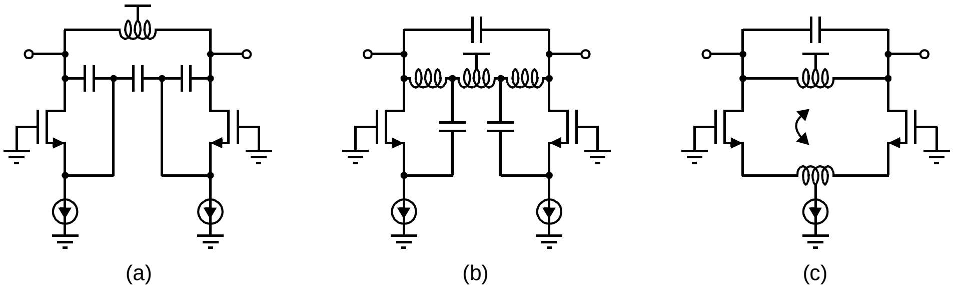


Figure 56. Loss compensation of differential resonators with passive impedance transformation in feedback path to the transistor source terminal. Capacitive Division (Colpitts) - (a), Tapped inductor (Hartley) - (b) and transformer feedback (c).

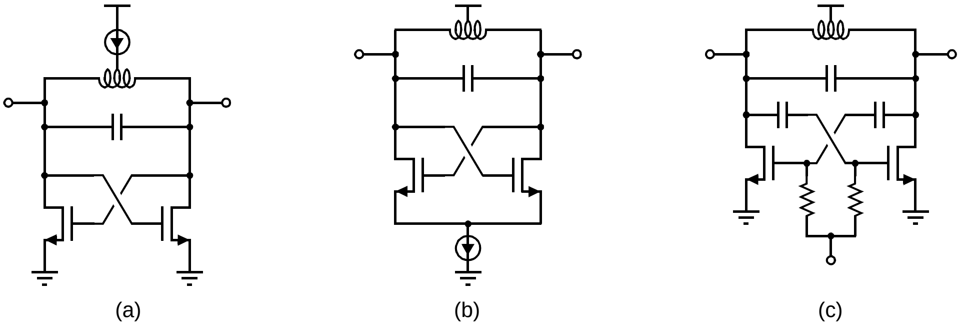
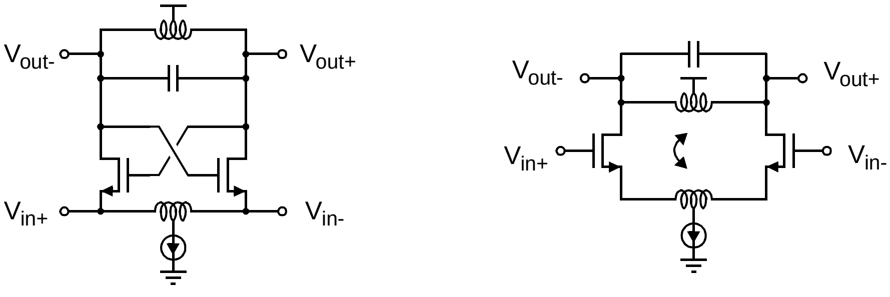


Figure 57. Loss compensation of differential resonators with feedback to the transistor gate terminal (cross-coupled pair). Biasing through center tapped inductor (a), biasing through tail current (b) and biasing at the gates of the buffers (c).

An alternative impedance transformation technique is inductor tapping. Oscillators with tapped inductors in the feedback are often referred to as Hartley oscillators. A balanced version is shown in Figure 56 - (b). Finally transformer feedback can be used either with loose coupling or with different numbers of windings. Again a balanced version is shown in Figure 56 - (c).

In balanced circuits it is very easy to provide 180 degrees of phase shift through simple cross coupling. This can be exploited to provide loss compensation through feedback to the gate terminal of the loss compensating transistors. This is desirable because the gate terminal is usually a node with a relatively high impedance and therefore it does not load the tank. Figure 57 shows different examples of how this can be done. The first (a) uses a top-biased structure using a center tapped inductor, the second (b) uses tail-current biasing and the third (c) uses biasing directly at the transistor gate terminals through large biasing resistors.



(a) (b)

Figure 58. Examples of Input Buffers with Negative Output Impedance. 1/gm input termination (a) and inductive source degeneration (b).

All of these oscillator structures and many others following the same feedback topologies can be used for loss compensation of LC resonators that are used in Q-enhanced LC filters provided that the gm of the transistors is chosen appropriately i.e. such that the desired effective Q-value is achieved without jeopardizing the stability of the filter.

5.1.2 Filter Input Stages with Negative Output Impedance

Negative resistance circuits that are intended for loss compensation of LC resonators are usually viewed as separate circuit units in the sense that they are designed separately and have their own bias current etc. This means that the LC filter will consume power in both the input stage and in the loss compensation circuit and this while both contribute with their own noise. This seems obvious but there is actually something that can be done to reduce power consumption and simultaneously reduce the noise level.

In the previous section it was explained how loss compensation could be provided either through feedback to the gate terminal or through feedback to the source terminal. This can be exploited because if the feedback is led to one terminal then the input signal can be injected through the other terminal. This way the stage will function as a low noise input buffer with a negative output impedance that can be used to loss compensate the LC resonator load.

Figure 58 shows two examples of how this could be done. The first (a) uses cross-coupled feedback to the gate terminal and the input signal is injected at the source terminal. Notice the similarity with the traditional 1/gm LNA topology. It is necessary to choose an appropriate transistor gm value to avoid over compensation of the resonator. If this transconductance is not compatible with the desired input impedance it is possible to split the input transistor into several unit transistors and only lead the feedback to some of these while the gates of the remaining unit transistors are simply AC grounded at the desired bias level.

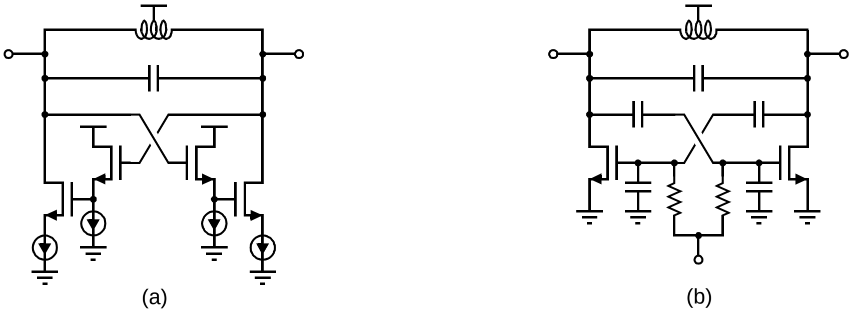


Figure 59. Increasing the maximum tank voltage swing. Source follower in feedback path (a) and capacitive division in feedback path (b).

An alternative solution (Figure 58 - b) is to inject the input signal at the gate terminal and implement the loss compensation through transformer feedback to the source terminal. Notice the similarity with the very popular inductive source degeneration LNA topology. Furthermore in this configuration there is an extra degree of freedom since the coupling coefficient can be used to adjust the loss compensation with a given input transistor and source inductance.

These input buffers do enable reduced power consumption and noise but they have other implications as well. One implication is the increased complexity of tuning the loss compensation without affecting the input impedance and gain. Another implication is that the gate-source capacitance reduces the achievable reverse isolation. For some applications this may be very problematic and for others it might not. For instance a front-end filter intended for the receiver in a high performance cellular system usually must have very good reverse isolation to guarantee that no signal e.g. from the reference oscillator leaks to the antenna. Another problem with reduced reverse isolation is that if the off-chip impedance changes considerably, perhaps due to a finger placed on the antenna, the resonance frequency of the resonator may change and thus detune the filter. If instead the filter is used to reduce noise in the transmitter the impedance levels at the input will be well defined and there is not risk of unintentional leakage to the antenna so in this case low noise buffers with negative output impedance might be interesting.

5.2 Large Signal Linearity

The requirements for an RF filter vary depending on the application but typically the requirements for dynamic range are high and especially for a front-end filter connected directly to the antenna the requirements for dynamic range are extremely high. This implies that the filter must be able to handle large voltage swings while processing very weak signals. The passive elements are highly linear when implemented as metal structures so they do not pose any difficulty in terms of linearity. Likewise the frequency tuning as well as the input and output circuits can be implemented such that they do not limit the voltage swing (see previous chapters).

The loss compensation circuit, on the contrary, does limit the achievable voltage swing. The cross coupled loss compensation technique (feedback to the gate terminal) is desirable because it does not load the resonator. Unfortunately the pure cross coupled structure has a peak voltage swing in one side of only a few hundred millivolts because the drain-gate DC voltage drops to below minus one threshold voltage and thus the transistors leave the saturation region. One common way to alleviate this problem is to place a source follower in the feedback path to provide a level shift and thus increase the drain-gate voltage (Figure 59 - a). While this certainly increases the possible tank voltage swing it has another less desirable implication. Namely that the source follower adds noise and therefore reduces the filters sensitivity.

An alternative method for increasing the possible voltage swing is to place a capacitor in the feedback path. This enables higher voltage swing in two different ways. First the bias level at the gate of the transistors can be set independently of the resonator voltage. This means that the drain-gate voltage can be chosen such that it does not limit the voltage swing. Instead the voltage swing is limited by the supply voltage.

It can be shown (Figure 60) that the maximum voltage swing that can be sustained with the transistors operating in saturation is one third of the supply voltage. The maximum voltage swing can be increased further by implementing capacitive division in the feedback path (Figure 59 - b). Actually there will always be some capacitive division due to parasitic capacitance at the gate terminal. Capacitive division reduces the voltage swing at the gate such that this terminal can be biased at a lower level and the tank swing can be augmented. The ratio of the voltage swings at the gate and in the tank can be expressed as a capacitive division constant (k_{CD}) as shown in (106). The fact that there is a difference between the voltage swings at these two terminals implies that the gm expression derived in Chapter 3 needs to be modified to take this effect into account as well. This is done in (107).

$$V_{\mathrm{gate}} = k_{\mathrm{CD}} \cdot V_{\mathrm{tank}} \tag{106}$$

$$gm = \frac{2(1 - 1/k_{\mathrm{QE}})}{k_{\mathrm{CD}} Q_{\mathrm{tank}} \sqrt{L/C}} \tag{107}$$

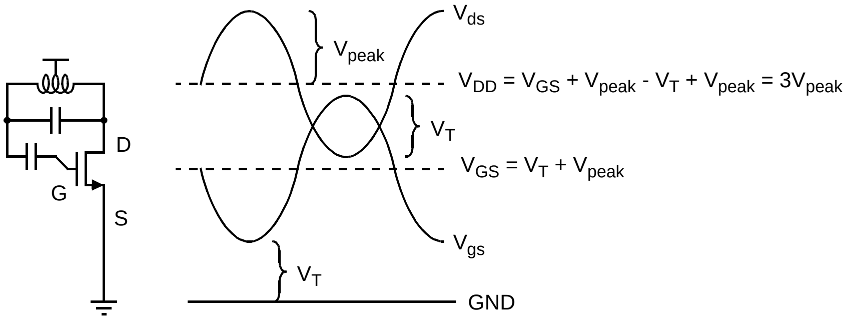

D
G
S
V_{ds}
V_{peak}
$V_{DD} = V_{GS} + V_{peak} - V_T + V_{peak} = 3V_{peak}$
V_T
$V_{GS} = V_T + V_{peak}$
V_{gs}
V_T
GND

Figure 60. Cross-Coupled Pair Biased for Maximum Voltage Swing. V_{DD} equals three times the peak voltage swing and V_{GS} equals the threshold voltage plus the peak voltage swing.

From this expression it is clear that a lower capacitive division constant leads to a higher gm value and therefore higher power consumption and increased noise. This means that substantial capacitive division is not beneficial but moderate amounts of capacitive division could be advantageous.

A high voltage swing that can be sustained while the transistors remain in saturation is not the only concern because other large signal effects can easily limit linearity of the filter. One effect that may be problematic is the non-linear CMOS output impedance. If the output impedance is low enough compared to the tank impedance to matter it is possible to increase the output impedance through the cascoding technique. This will reduce loading but also reduce the achievable voltage swing and add noise of its own so cascoding should only be in the loss compensation circuit if it is necessary.

Another effect that needs to be taken into account as well is the quadratic nature of the drain current - gate voltage relation. Fortunately deep submicron CMOS transistors biased in strong inversion exhibit substantial velocity saturation which results in an almost linear I-V curve. For instance a 0.25 micron NMOS transistor with the gate biased at 1.1 Volt has an almost linear I-V curve over +/- 500mV. As a result a loss compensation circuit with limited capacitive division can be designed with tank voltage swings approaching one third of the supply voltage.

5.3 Noise

In order to achieve high dynamic range in an RF filter it is necessary to enable high signal swing and low noise levels. The previous section dealt with the design of loss compensation circuits that can work with high voltage swings. This section deals with the design and modelling of loss compensation circuits with low noise.

5.3.1 Basic Noise Sources

A Q-enhanced LC filter has a number of noise sources. The input stage usually contributes with a significant amount of noise. The design and modelling of low noise input stages was explained in Chapter 2 so this is not covered here. Instead the focus of this section is the noise of the Q-enhanced resonator itself i.e. the noise from the tank and the noise from the compensation circuit.

The first noise source is thermal noise coming from loss in the resonator. As explained earlier these losses can be modelled at the frequency of interest as a parallel resistor. This is also the case for the thermal noise i.e. the thermal noise can be modelled as a parallel noise current source with the value (108) where R_P is expressed in terms of the uncompensated tank Q and the characteristic impedance of the resonator.

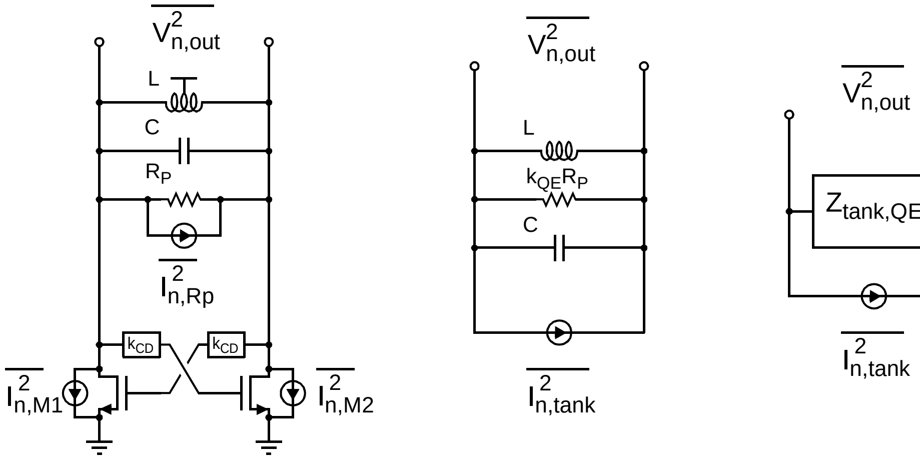
$$\overline{I_{n,\,Rp}^2} = \frac{4kT\Delta f}{R_P} = \frac{4kT\Delta f}{Q_{\text{tank}}\sqrt{L/C}} \tag{108}$$

$$\overline{I_{n,\,M1}^2} = \overline{I_{n,\,M2}^2} = 4kT\gamma g_{d0}\Delta f = \frac{4kT\gamma g_m\Delta f}{\alpha} = \frac{8kT\gamma(1-1/k_{\text{QE}})\Delta f}{\alpha k_{\text{CD}}Q_{\text{tank}}\sqrt{L/C}} \tag{109}$$

The other significant noise source is the thermal drain current noise of the loss compensating MOS transistors. This noise source is expressed as (109) where γ is the drain noise constant which has a value of 2/3 in saturation in the long channel regime [59]. g_{d0} - the zero voltage drain conductance is equal to the transconductance g_m divided by α where α represents the amount of velocity saturation which occurs at a given bias point (see Chapter 2 for more detail). Furthermore (109) substitutes g_m according to (107) such that all noise sources are expressed in terms of tank Q and characteristic impedance.

5.3.2 Q-Enhanced LC Resonator Noise

A simplified schematic of the cross coupled loss compensation circuit with capacitive division in the feedback path is shown in Figure 61 - (a) with noise sources (ignoring induced gate noise). The contributions from these noise sources can be computed by calculating the differential output noise voltage resulting from each noise source.



(a) (b) (c)

Figure 61. Q-Enhanced LC Resonator. With noise sources (a), equivalent schematic (b) and symbol (c).

The differential output noise voltage coming from transistor M1 can be calculated by nulling the other noise sources and using Kirchoffs current law on each of the output nodes. Doing so leads to the expression (110). The noise contribution from M2 is similar to that of M1 because the circuit is symmetrical so it is given by (111). Similarly the noise coming from loss in the resonator can be calculated to be (112).

These output noise powers can be added directly because they are not correlated. Therefore the total output noise power is given by (113) where the total noise power is expressed in terms of the differential Q-enhanced tank impedance ($Z_{\mathrm{tank,QE}}$), the uncompensated Q-value (Q_{tank}) and the characteristic impedance ($\sqrt{L/C}$).

Finally this enables us to model the Q-enhanced LC resonators noise as one parallel noise current with the value given by (114).

$$\overline{V^2_{n,\,out}}\Big|_{M1} = \left|\frac{1}{2(j\omega C + 1/j\omega L + 1/k_{\mathrm{QE}}R_P)}\right|^2 \cdot \overline{I^2_{n,\,M1}} = \frac{Z^2_{\mathrm{tank,QE}}}{4}\overline{I^2_{n,\,M1}} \tag{110}$$

$$\overline{V^2_{n,\,out}}\Big|_{M2} = \left|\frac{1}{2(j\omega C + 1/j\omega L + 1/k_{\mathrm{QE}}R_P)}\right|^2 \cdot \overline{I^2_{n,\,M2}} = \frac{Z^2_{\mathrm{tank,QE}}}{4}\overline{I^2_{n,\,M2}} \tag{111}$$

$$\overline{V^2_{n,\,out}}\Big|_{R_P} = \left|\frac{1}{j\omega C + 1/j\omega L + 1/k_{\mathrm{QE}}R_P}\right|^2 \cdot \overline{I^2_{Rp}} = Z^2_{\mathrm{tank,QE}}\overline{I^2_{Rp}} \tag{112}$$

$$\begin{aligned}\overline{V_{n,\,out}^2} &= \frac{Z_{\text{tank,QE}}^2}{4}\overline{I_{n,\,M1}^2} + \frac{Z_{\text{tank,QE}}^2}{4}\overline{I_{n,\,M2}^2} + Z_{\text{tank,QE}}^2\overline{I_{Rp}^2} \\ &= Z_{\text{tank,QE}}^2\left(\frac{2kT\gamma(1-1/k_{\text{QE}})\Delta f}{\alpha k_{\text{CD}}Q_{\text{tank}}\sqrt{L/C}} + \frac{2kT\gamma(1-1/k_{\text{QE}})\Delta f}{\alpha k_{\text{CD}}Q_{\text{tank}}\sqrt{L/C}} + \frac{4kT\Delta f}{Q_{\text{tank}}\sqrt{L/C}}\right) \\ &= Z_{\text{tank,QE}}^2\frac{4kT\Delta f}{Q_{\text{tank}}\sqrt{L/C}}\left(1+\frac{\gamma(1-1/k_{\text{QE}})}{\alpha k_{\text{CD}}}\right)\end{aligned} \tag{113}$$

$$\overline{I_{n,\,\text{tank}}^2} = \frac{\overline{V_{n,\,out}^2}}{Z_{\text{tank,QE}}^2} = \frac{4kT\Delta f}{Q_{\text{tank}}\sqrt{L/C}}\left(1+\frac{\gamma(1-1/k_{\text{QE}})}{\alpha k_{\text{CD}}}\right) \tag{114}$$

5.4 Tuning the Negative Resistance

It is not sufficient that an on-chip RF filter achieves a sufficient dynamic range and that it can be frequency tuned. Process and temperature variations make it absolutely necessary that the amount of loss compensation can also be tuned or calibrated. In this section different techniques for tuning the amount of loss compensation will be presented and some ideas for systems that can control the tuning are presented. As for the frequency tuning the basic RF performance is prioritized over the more practical tuning control issues so this last topic is only covered briefly.

5.4.1 Continuous tuning

There are several ways the negative resistance of a loss compensation circuit can be tuned. The first and perhaps the most common is to change the transconductance of the transistors by varying the bias tail current (Figure 62 - a). Unfortunately deep submicron transistors when biased at high overdrive voltages have transconductances that are almost independent of the bias current. Actually the gm drops above a certain gate-source voltage due to velocity saturation and other short channel effects so this approach is not well suited for tuning of Q-enhanced RF filters with very high dynamic range. Strictly speaking this effect is fortunate with regards to dynamic range but not with regards to tuning. However the first is clearly more important as there are alternatives for tuning the negative resistance. One alternative is to use varactors for capacitive division in the feedback path (Figure 62 - b). This way the voltage swing at the gates can be tuned with a control voltage. The problem with this tuning approach is mainly that the capacitive division leads to higher noise and power consumption. Also the non-linearity of the varactors might impose problems at maximum voltage swing. Furthermore the change in capacitance directly results in a change in resonance frequency which might complicate the tuning or calibration system substantially. A third approach for linear tuning of the negative resistance is to vary the bias voltage at the transistor gates (Figure 62 - c) but this method fundamentally suffers form the same problems the first did when deep submicron MOS transistors are used.

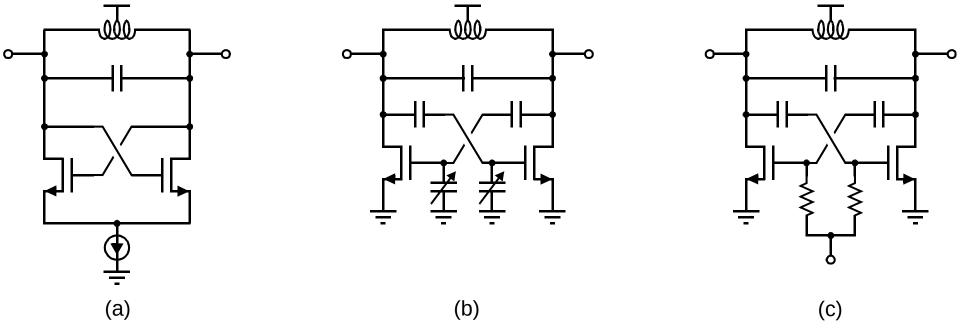


Figure 62. Continuous Q-tuning techniques. Bias current adjusted transconductance (a), varactor tuned capacitive division ratio (b) and gate-source bias voltage adjusted transconductance (c).

5.4.2 Switched Tuning

If the continuous tuning techniques are abandoned or if they are used in combination with discrete calibration there are several much more attractive solutions. For instance switched tuning enables optimum RF performance and great flexibility at the expense of more complicated design and verification. The simplest form of switched tuning is the use of NMOS switches either at the source or at the drain of a number of unit transistors that are all cross coupled. When the switch is placed at the source it can have a linearizing effect in the on-state which might be useful but it also leads to extra noise. If the switch is placed at the drain they are effectively used as switched cascodes so if a cascode structure is acceptable with regards to noise and voltage swing this might be a very useful solution.

Another slightly more complicated solution is to use a switched capacitor bank in the capacitive feedback (Figure 63 - a). This takes care of the varactor linearity concern but it still leads to increased power consumption and noise due to the capacitive division.

A more complicated solution that enables optimum RF performance is shown in Figure 63 - (b). This solution uses a number of separate binary weighted transistors that are each cross coupled with their own feedback capacitor. This enables switched bias voltage for each transistor such that they can be turned on or off individually. Notice that the switches that are used for the switched bias should be sized such that they have a high on resistance because otherwise the voltage swing will be strongly reduced and substantial noise will be added.

As a final note one should also be cautious of making this on-resistance too high because this leads to much increased noise at low frequencies which may be up-converted due to nonlinearities in the presence of strong blocking signals.

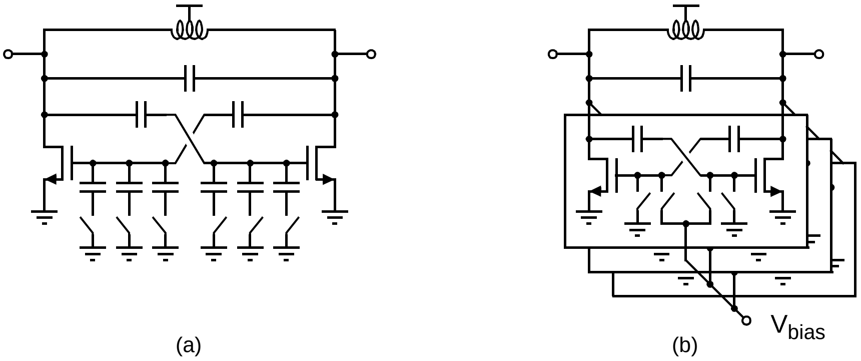


Figure 63. Switched tuning of the loss compensation. Switched capacitor tuning altering the capacitive division ratio (a) and switched bias of binary weighted transistor array (b).

5.5 Systems for Control of the Q-tuning

Controlling the Q-value is definitely not trivial but as it has been stated earlier basic RF performance is more important. Therefore this section will only give a brief introduction to tuning systems. Most of the tuning systems presented are the same as those presented in the chapter on frequency tuning. This is not a coincidence because these tuning systems are developed for simultaneous tuning of the frequency and the quality factor.

5.5.1 Continuous Tuning

The more traditional tuning systems are continuous in nature. One of the most common is that master-slave tuning approach which was presented in the previous chapter. For controlling the Q-value it is necessary to use some extra circuitry. An envelope detector can be used in the feedback loop to control the negative resistance such that the oscillation amplitude is very low (Figure 64 - a). When the oscillation amplitude is very low the distortion is very low too and therefore the negative resistance matches well the losses of the resonator. If the negative resistance in the filter is implemented with transistors that are e.g. 20% smaller (transistor width W 20% lower) than those used in the oscillator then this loss compensation circuit will only compensate for 80% of the losses in the filter resonator i.e. the filter will have a Q-enhancement factor k_{QE} of approximately 5.

Another classic tuning approach which has also been presented before is the self-tuning approach (Figure 64 - b). The filter is taken off-line periodically and tuned with a separate tuning circuit. This circuit can be implemented in a number of different ways. For instance after frequency tuning is performed a signal can be passed through the filter and the gain will give the Q-value because the characteristic impedance does not vary much with process.

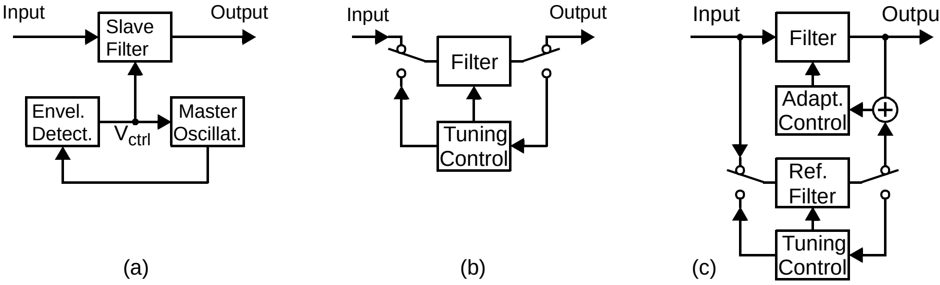


Figure 64. Q-tuning systems. Envelope detector in feedback to keep to oscillation amplitude low (a), Self-tuning (b) and adaptive tuning (c).

Therefore a feedback loop that sets the gain to a certain value will simultaneously fix the Q-value. After tuning the control signal can be sampled and held and the filter can be put back on-line.

Finally adaptive tuning which was also presented before can be used for tuning the Q-value as well as the frequency. Once the reference filter is tuned to the desired frequency and Q-value the adaptive control circuitry can compare both phase and magnitude of the outputs of the two filters and use this information to simultaneously tune both frequency and Q-value.

5.5.2 Discrete Tuning and DSP Assisted Tuning

Q-tuning does not have to be continuous. The above mentioned tuning methods can also be performed in an iterative digital manner with either discrete Q-tuning elements or through dedicated D/A converters.

Also it is possible to take advantage of the existing circuit blocks in an RF transceiver as mentioned earlier. Especially the presence of a frequency synthesizer and a receiver and/ or a transmitter makes this possible. One simple approach to Q-tuning is to turn the LNA off, over compensate the resonator such that it starts oscillating. This can be observed through the receiver chain. Then the loss compensation is slowly reduced until the oscillation ceases. At this point the negative resistance circuit matches the losses of the resonator so reducing the loss compensation by another 20% will imply that 80% of the losses are compensated for i.e. that a Q-enhancement factor of 5 is chosen. This control value can now be sampled either in the digital or in the analog domain and the LNA can be turned back on again.

5.6 Summary

This chapter has covered most of the fundamental issues involved with the design of loss compensation circuits for Q-enhancement of LC resonators that are used in RF filters. The material presented is based on classic LC oscillator techniques because they enable both large voltage swing and low noise. Put in other terms it is believed that the use of such techniques can lead to the highest possible dynamic range.

The first section presented different circuit topologies that provide a negative impedance which can be used for loss compensation. Specifically feedback to the source terminal with impedance transformation in the feedback path and cross coupled feedback to the gate terminal enables a negative output impedance which can be used for loss compensation. The cross coupled feedback to the gate terminal appears to be most attractive because it involves the least amount of loading of the resonator. The following section explained how large resonator voltage swings can be enabled without jeopardizing sensitivity. The main challenge is DC biasing of the drain and gate terminals. With a source follower or AC coupling in the feedback path it is possible to reach single sided peak tank voltage swings approaching one third of the supply voltage.

Then the noise of differential Q-enhanced LC resonators with capacitive division in the feedback path was modelled and it was shown that to minimize the noise it is important that the uncompensated tank quality factor is maximized, that the Q-enhancement factor is minimized and that the capacitive division is minimized. After this it was explained how the negative impedance of the loss compensation circuits can be tuned. Both continuous and discrete tuning methods were presented. It appears that the traditional bias current tuning is not well suited for this application because strong velocity saturation makes the gm almost independent of the bias current. Instead switched transistors or switched bias is believed to be the safest way to good RF performance. Finally a few control systems were presented that can perform the needed Q-tuning.

5.7 References

[55] B. Razavi, "RF Microelectronics," *Prentice Hall*, 1998.

[56] B. Razavi, "Design of Analog CMOS Integrated Circuits", *McGraw-Hill*, 2000.

[57] W. Kuhn, "Design of Integrated, Low Power, Radio Receivers in BiCMOS Technologies," *Ph.D. Dissertation Virginia Polytechnic Institute and State University*, 1995.

[58] S. Pipilos, et al., "A Si GHz RLC Filter with Tunable Center Frequency and Quality Factor," *IEEE J. Solid-State Circuits*, vol. 31, no. 10, pp. 1517-1524, October 1996.

[59] A. van der Ziel, "Thermal Noise in Field Effect Transistors," *Proc. IEEE*, August 1962, pp. 1801-1812.

[60] D. Shaeffer and T. Lee, "A 1.5V, 1.5GHz CMOS Low Noise Amplifier," *IEEE J. Solid-State Circuits*, May 1997.

Chapter 6: Integrated Front-End Filter Design

A fully integrated front-end filter for a high performance system like GSM or WCDMA in a standard silicon process poses a significant challenge due to its exceedingly high requirements for dynamic range. To the best of the authors knowledge this goal has never been achieved with an acceptable current consumption i.e. less than 30-40 mA.

Chapter 2 explained how a low noise input stage can be designed with an exceptionally high input power handling capability. Chapter 3 explained how state-of-the-art inductors and metal-metal capacitors can be designed in a standard digital CMOS process with low substrate resistivity - the worst case situation. Furthermore it has been explained in Chapter 4 how frequency tuning can be accommodated with very low loss through the use of switched capacitor banks with optimized switches and finally Chapter 5 explained how loss compensation can be achieved with maximum signal swing and the least possible noise added.

This chapter combines all the results of the previous chapters in an attempt to prove the feasibility of the fully integrated front-end filter for wireless communication. Furthermore a complete closed form design description is presented including all component parameters and all biasing details. This was made possible through the analytical work of the previous chapters. The chapter is concluded with an example of a fully integrated front-end filter for GSM1800,GSM1900 and WCDMA which uses these techniques.

6.1 Filter Topology

The filter topology which is used in this work was chosen based on the philosophy that as long as the basic RF performance is not achieved it does not matter if the far away attenuation is a little better or how practical the tuning mechanisms are i.e. the topology was chosen because it is believed to be the one that enables the best dynamic range and thus is the topology which is most likely to fulfill the very stringent requirements of a fully integrated front-end filter. The reasoning is as follows.

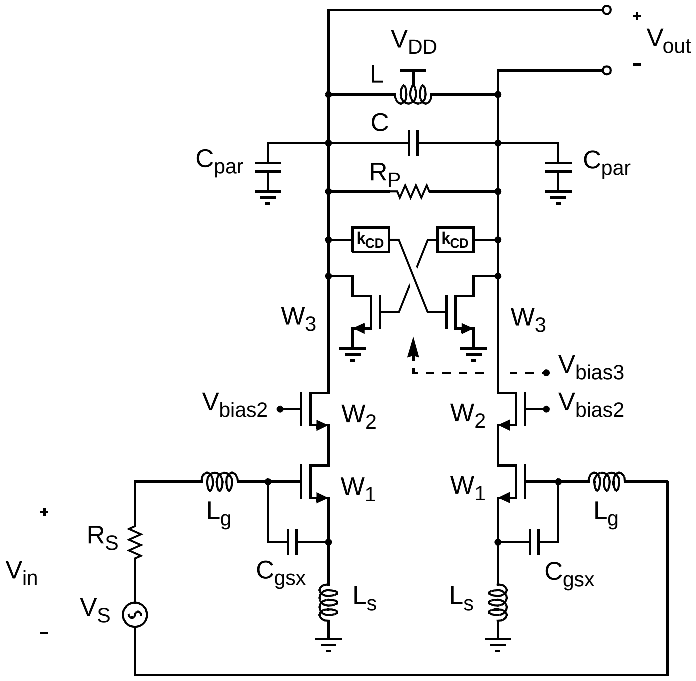


Figure 65. Overall filter structure used in the modelling and optimization of a fully integrated front-end filter.

Purely passive filters have been attempted for integration of fully-integrated filters [63], but their performance was relatively discouraging (10dB insertion loss and 1GHz BW at 6GHz) due to the high losses of on-chip inductors. Also they are not easily tunable so they offer a less flexible filtering solution. This means that some amplification must be provided before the filtering can take place in order to reduce the noise figure of the filter.

After amplification there are a few options: a single resonator, passively coupled resonators and actively coupled resonators. The passively coupled resonators provide additional attenuation far away from the passband but reduced attenuation close to the passband (see Chapter 3) and as this is the main issue in these filters the passively coupled resonator topologies are abandoned. A single resonator topology can be viewed as the first stage of an actively coupled resonator topology and because this is the most challenging stage it was chosen to focus on the single resonator topology - knowing that if one stage is possible then a cascaded solution can provide additional attenuation if desired. Finally moderate quality factor enhancement (< 10) is used to achieve a reasonable amount of selectivity. The reason why high quality factor enhancement in single resonators and moderate quality factor enhancement in passively coupled resonators are not considered is that they have serious stability issues and may have unreasonably high noise or power consumption (Chapter 3). As a conclusion of this line of argumentation it was chosen to base the analysis of a fully integrated front-end filter on a single Q-enhanced resonator placed after a highly linear low noise input amplifier stage. A common-source, open-drain output buffer structure was assumed because this structure provides very good linearity with a low-loss capacitive input impedance which can be incorporated as a part of the resonator. Finally the entire filter structure is chosen to be fully differential to make it robust to cross-talk when integrated on a larger RF IC.

The target filter topology is shown in Figure 65. Notice that the input stage is made out of two single ended input stages rather than out of one differential stage with a tail current. One reason for this is that the noise from a tail current source in a differential stage is modulated into the frequency band of interest at large input signal levels. Another reason is that when very large input signals are present then the input stages will raise their bias point leading to a much softer roll-off i.e. leading to a significantly higher 1dB ICP. This choice of input stage also implies that all the analytical results from Chapter 2 can be reused with only minor modifications.

We will start with the effective transconductance G_m of the input stage. In the single ended case it was given by (115). When the same topology is used in one side of a differential input stage then the source impedance is $R_s/2$ and the blocker input power is $P_{in}/2$ but the drain current is also only half ($I_D/2$). This results in exactly the same transconductance as the single ended case (116) and thus we can continue to use this equation in the analysis of the filter.

$$G_{m,\ \text{single-end}} = \frac{I_D}{v_{in,\ peak}} = \frac{I_D}{\sqrt{2P_{in}R_s}} \tag{115}$$

$$G_{m,\ \text{diff, one side}} = \frac{I_D/2}{v_{in,\ peak,\ \text{one side}}} = \frac{I_D/2}{\sqrt{2(P_{in}/2)(R_s/2)}} = \frac{I_D}{\sqrt{2P_{in}R_s}} = G_{m,\ \text{single-end}} \tag{116}$$

After the input transistors in Figure 65 there is a level of cascode transistors. These are chosen because they improve the reverse isolation and raise the output impedance and thereby reduce the loading of the resonator. They were not included in the noise modelling because their noise is usually negligible and therefore for now we will assume that all they do is to increase the output impedance enough to be considered infinite.

Next there is the loss compensation circuit. The intended loss compensation circuit is the one shown in Chapter 5 in Figure 63-b because it allows quality factor tuning while providing maximum voltage swing and minimum noise. For the analysis of the filter it is sufficient to use a simplified representation of this circuit as shown in Figure 65 where the factor k_{CD} simply illustrates a reduction in voltage swing from the tank.

The resonator itself is modelled as a parallel RLC tank with some parasitic capacitance to ground. The inductor L is expected to be designed as explained in Chapter 3 and the capacitor is expected to be implemented as Vertical Mesh Capacitors also as explained in Chapter 3. The frequency tuning is expected to be made with optimized switches as explained in Chapter 4 and thus achieving very low loss and therefore it is not necessary to incorporate these effects into the model shown in Figure 65. Finally R_p models all the losses in the resonator i.e. mainly the losses of the inductor. R_p should therefore not be considered a physical component in the design but rather a parasitic component.

With these assumptions in place it is possible to proceed with the analysis and turn to the choice of the impedance of the resonator. The choice of the resonator impedance or equivalently the choice of the inductor and capacitor values depends on several factors. First we need to realize that the filter can only sustain a certain voltage swing. In Chapter 5 it was shown that the peak single-ended voltage swing is one third of the supply voltage. It was also shown that the loss compensating transistor must be biased at $V_{gs} = V_T + V_{peak} = V_T + V_{DD}/3$ for this to be possible.

The peak voltage swing is usually invoked when the peak current injected into the tank is $I_D/2$ and this is a result of the strongest possible out-of-band blocker being present in the minimum distance from the center frequency. This hints that we need to know the supply voltage (V_{DD}), the drain current (I_D), the center frequency (f_0), the closest frequency of the strongest out-of-band blocker (f_b) and that we need to know the effective Q-value of the resonator i.e. the quality factor after Q-enhancement ($Q_E = k_{QE}Q_{tank}$) - in order to know the relative attenuation of the tank impedance at the blocker frequency.

The expression for the tank impedance at the center frequency ($k_{\mathrm{QE}}Q_{\mathrm{tank}}\sqrt{L/C}$) can be found in the following manner. The peak single-ended voltage swing should equal the peak single-ended current multiplied by half the magnitude of the tank impedance at the frequency of the blocker (117). The expression for the magnitude of the tank impedance as a function of frequency was found in Chapter 3 in (48).

$$V_{peak,\,SE} = I_{peak,\,SE}\frac{|Z_{\mathrm{tank}}(f_b)|}{2} = \frac{I_D}{2}\frac{Q_E\sqrt{L/C}}{2\sqrt{1+Q_E^2\left(\frac{f_b}{f_0}-\frac{f_0}{f_b}\right)^2}} = \frac{I_D}{4}\frac{k_{\mathrm{QE}}Q_{\mathrm{tank}}\sqrt{L/C}}{\sqrt{1+k_{\mathrm{QE}}^2Q_{\mathrm{tank}}^2\left(\frac{f_b}{f_0}-\frac{f_0}{f_b}\right)^2}} \quad (117)$$

It is now possible to isolate the expression for the tank impedance at the center frequency (118).

$$k_{\mathrm{QE}}Q_{\mathrm{tank}}\sqrt{L/C} = V_{peak,\,SE}\frac{4}{I_D}\sqrt{1+k_{\mathrm{QE}}^2Q_{\mathrm{tank}}^2\left(\frac{f_b}{f_0}-\frac{f_0}{f_b}\right)^2} = \frac{4}{3}\frac{V_{DD}}{I_D}\sqrt{1+k_{\mathrm{QE}}^2Q_{\mathrm{tank}}^2\left(\frac{f_b}{f_0}-\frac{f_0}{f_b}\right)^2} \quad (118)$$

With the knowledge of the effective transconductance Gm and the knowledge of the magnitude of the resonator at the center frequency it is now possible to derive an expression for the voltage gain of the fully integrated front-end filter or in this context perhaps more appropriately the fully integrated bandpass amplifier $A_V(f)$.

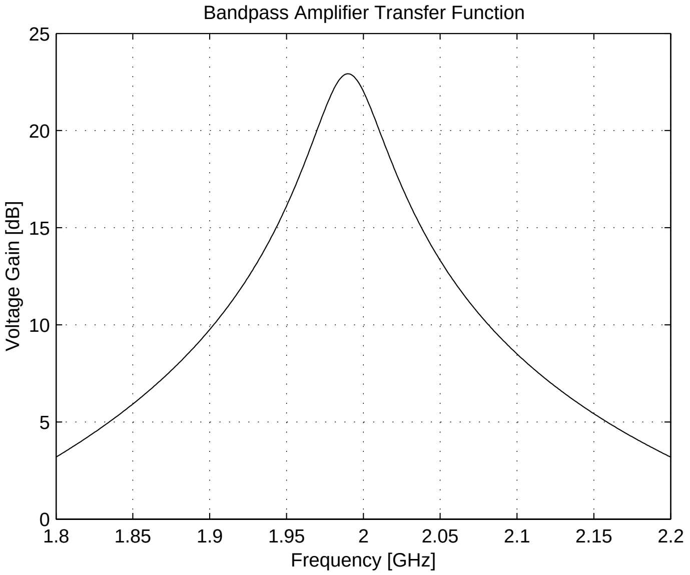


Figure 66. Voltage transfer function of a bandpass amplifier with 1.99GHz center frequency, 2.07GHz blocker frequency, 0dBm blocker power, a resonator Q-value of 48, 100Ω input impedance and 2.4V supply voltage.

This is done in (119) where (48) is used again. It is perhaps surprising that the voltage gain is independent of the bias current. This is due to the fact that the voltage swing is limited by the supply voltage and therefore it is necessary to reduce the tank impedance every time I_D is increased.

$$\begin{aligned} A_V(f) &= G_m \cdot \frac{\left|Z_{\text{tank}}(f)\right|}{2} = \frac{I_D}{\sqrt{2P_{in}R_s}} \cdot \frac{k_{\text{QE}} Q_{\text{tank}} \sqrt{L/C}}{2\sqrt{1 + k_{\text{QE}}^2 Q_{\text{tank}}^2 \left(\frac{f}{f_0} - \frac{f_0}{f}\right)^2}} \\ &= \frac{I_D}{\sqrt{2P_{in}R_s}} \frac{\frac{4V_{DD}}{3I_D}\sqrt{1 + k_{\text{QE}}^2 Q_{\text{tank}}^2 \left(\frac{f_b}{f_0} - \frac{f_0}{f_b}\right)^2}}{2\sqrt{1 + k_{\text{QE}}^2 Q_{\text{tank}}^2 \left(\frac{f}{f_0} - \frac{f_0}{f}\right)^2}} = \frac{2V_{DD}}{3\sqrt{2P_{in}R_s}} \frac{\sqrt{1 + k_{\text{QE}}^2 Q_{\text{tank}}^2 \left(\frac{f_b}{f_0} - \frac{f_0}{f_b}\right)^2}}{\sqrt{1 + k_{\text{QE}}^2 Q_{\text{tank}}^2 \left(\frac{f}{f_0} - \frac{f_0}{f}\right)^2}} \end{aligned} \qquad (119)$$

Finally an example is provided in Figure 66 to give a feeling of the numbers involved in a fully integrated front-end filter designed for the 2GHz range. The center frequency is 1.99 GHz and the blocker power is 0dBm at 2.07GHz. Furthermore V_{DD} is assumed to be 2.4V, the differential input impedance is assumed to be 100Ω and the enhanced resonator Q-value is assumed to be 48. Then the in-band voltage gain will be 23dB and the relative attenuation of the blocker will be 11.9dB.

6.2 Component Values and Biasing

The purpose of this work is not only to prove the feasibility of fully integrated front-end filters. The purpose is also to gain enough insight to identify a filter circuit topology that guarantees performance which is close to the fundamental limitations of a semiconductor process. Finally it is the purpose of this work to develop a design methodology which efficiently and reliably identifies the component values needed to implement this filter topology. The purest form of a design methodology is a closed form description of the design. This is because only this form contains the full design insights as well as the easiest and fastest circuit synthesis. This section explains how all component values and bias levels in the fully integrated front-end filter are calculated.

The first component in the signal path is the series inductor at the gate of the input transistor L_g. Because this input stage uses the topology developed in Chapter 2 we can use (36) to find the inductance value. Only in this case R_s, P_{in} and I_D should be replaced by $R_s/2$, $P_{in}/2$ and $I_D/2$. Doing so yields (120). Using the same procedure with (35) leads to an expression for the external gate-source capacitance (121). Repeating this procedure with the inductor at the source of the input transistor leads to the same expression as for the single ended input stage i.e. (122). Finally the width of the input transistor W_1 can be found using (33) with I_D replaced by $I_D/2$. The choice of the width of the cascode transistor W_2 has not been explained in detail so far. A common choice is to use the same transistor width as input transistor because this gives some minor layout advantages. However in this case we choose the width of the cascode to be three times the width of the input transistor i.e. $W_2 = 3W_1$. This is because a wider cascode transistor raises the the DC level at the drain of the input transistor and thus enables a larger V_{GS1}. The cascode should be biased a little lower than the supply voltage minus the peak swing plus a threshold voltage to guarantee operation in saturation. This means that V_{bias2} as shown in Figure 65 should be chosen according to (124).

$$L_g = \frac{\sqrt{R_s/2}}{\omega_0}\left(\frac{V_{GS1}-V_T}{\sqrt{P_{in}}\left(1+\frac{1}{1+\theta(V_{GS1}-V_T)}\right)} - \frac{2\sqrt{P_{in}}}{I_D}\right) \tag{120}$$

$$C_{gsx} = \frac{\sqrt{2P_{in}/R_s}}{\omega_0(V_{GS1}-V_T)}\left(1+\frac{1}{1+\theta(V_{GS1}-V_T)}\right) - \frac{I_D L_{drawn}^2(1+\theta(V_{GS1}-V_T))}{\mu_n(V_{GS1}-V_T)^2} \tag{121}$$

$$L_s = \frac{\sqrt{2(R_s/2)(P_{in}/2)}}{\omega_0(I_D/2)} = \frac{\sqrt{2R_sP_{in}}}{\omega_0 I_D} \tag{122}$$

$$W_1 = \frac{I_D L_{eff}(1+\theta(V_{GS1}-V_T))}{\mu_n C_{ox}(V_{GS1}-V_T)^2} \tag{123}$$

$$V_{bias2} \approx V_{DD} - V_{peak,SE} + V_T - 100mV = \frac{2V_{DD}}{3} + V_T - 100mV \tag{124}$$

Now we turn to the resonator. From (118) and (47) we can identify an expression for the equivalent parallel resistance R_p (125). Using (125) we can isolate the resonator inductance L (126). Now it is a simple matter to find the resonator capacitance C. Isolating C in (46) and substituting L with the expression given in (126) leads to (127) where the parasitic capacitances to ground C_{par} are subtracted.

$$R_p = Q_{\mathrm{tank}}\sqrt{L/C} = Q_{\mathrm{tank}}2\pi f_0 L = \frac{4}{3}\frac{V_{DD}}{I_D}\sqrt{\frac{1}{k_{\mathrm{QE}}^2} + Q_{\mathrm{tank}}^2\left(\frac{f_b}{f_0} - \frac{f_0}{f_b}\right)^2} \tag{125}$$

$$L = \frac{2V_{DD}}{3\pi f_0 I_D}\sqrt{\frac{1}{k_{\mathrm{QE}}^2 Q_{\mathrm{tank}}^2} + \left(\frac{f_b}{f_0} - \frac{f_0}{f_b}\right)^2} \tag{126}$$

$$C = \frac{1}{\omega_0^2 L} - \frac{C_{par}}{2} = \frac{3I_D}{8\pi f_0 V_{DD}\sqrt{\frac{1}{k_{\mathrm{QE}}^2 Q_{\mathrm{tank}}^2} + \left(\frac{f_b}{f_0} - \frac{f_0}{f_b}\right)^2}} - \frac{C_{par}}{2} \tag{127}$$

The only remaining part of the circuit is the transistor which provides loss compensation. To find the width of this transistor (W_3) we must first find an expression for the needed transconductance. Such an expression was presented in Chapter 5, (107). By using this equation and inserting the result from (125) we arrive at expression (128). This is the transconductance value which is needed. Now let us revise the expression for transconductance of a MOS transistor under the influence of mobility degradation - (22), (129).

$$gm = \frac{2(1 - 1/k_{\mathrm{QE}})}{k_{\mathrm{CD}}Q_{\mathrm{tank}}\sqrt{L/C}} = \frac{3I_D(k_{\mathrm{QE}} - 1)}{2k_{\mathrm{CD}}V_{DD}\sqrt{1 + k_{\mathrm{QE}}^2 Q_{\mathrm{tank}}^2\left(\frac{f_b}{f_0} - \frac{f_0}{f_b}\right)^2}} \tag{128}$$

$$gm = \frac{\mu_n C_{ox} W(V_{gs} - V_T)}{L_{eff}} \cdot \frac{1 + \theta(V_{gs} - V_T)/2}{(1 + \theta(V_{gs} - V_T))^2} \tag{129}$$

It is now possible to equate (128) and (129) and then isolate the transistor width W. This is done in (131). However this expression does not give the final form because the bias level of the transistor is still in the equation. What is needed to simplify the expression is the expression of the bias level of this transistor. Recall from Chapter 5 that this bias level is given by (130).

$$V_{bias3} = V_{GS3} = V_T + V_{peak} = V_T + V_{DD}/3 \tag{130}$$

When this expression is inserted in (131) we reach a useful expression for W_3 ((131) - second line).

$$\begin{aligned} W_3 &= \frac{L_{eff}}{\mu_n C_{ox}(V_{gs}-V_T)} \cdot \frac{(1+\theta(V_{gs}-V_T))^2}{1+\theta(V_{gs}-V_T)/2} \cdot \frac{3I_D(k_{\mathrm{QE}}-1)}{2k_{\mathrm{CD}}V_{DD}\sqrt{1+k_{\mathrm{QE}}^2 Q_{\mathrm{tank}}^2\left(\frac{f_b}{f_0}-\frac{f_0}{f_b}\right)^2}} \\ &= \frac{L_{eff}I_D(k_{\mathrm{QE}}-1)}{\mu_n C_{ox}k_{\mathrm{CD}}\sqrt{1+k_{\mathrm{QE}}^2 Q_{\mathrm{tank}}^2\left(\frac{f_b}{f_0}-\frac{f_0}{f_b}\right)^2}} \cdot \frac{(3/V_{DD}+\theta)^2}{2+\frac{\theta V_{DD}}{3}} \end{aligned} \tag{131}$$

Now we have all component values and all bias levels expressed in closed form. This allows us to calculate the total current consumption I_{total} of the fully integrated front-end filter as a function of the LNA drain current I_D. This is done in (132) where (131) is inserted in the expression for the drain current of a MOS transistor (20). After simplification it becomes evident that there is a linear relationship between I_D and the total current consumption I_{total}. This means that it is easy to find the drain current I_D if the total current budget is known.

$$\begin{aligned} I_{total} &= I_D + 2I_{M3} = I_D + 2\frac{\mu_n C_{ox}W_3(V_{GS}-V_T)^2}{2L_{eff}(1+\theta(V_{GS}-V_T))} = I_D + \frac{\mu_n C_{ox}(V_{DD}/3)^2}{L_{eff}(1+\theta V_{DD}/3)} \cdot W_3 \\ &= I_D + \frac{\mu_n C_{ox}(V_{DD}/3)^2}{L_{eff}(1+\theta V_{DD}/3)} \cdot \frac{L_{eff}I_D(k_{\mathrm{QE}}-1)}{\mu_n C_{ox}k_{\mathrm{CD}}\sqrt{1+k_{\mathrm{QE}}^2 Q_{\mathrm{tank}}^2\left(\frac{f_b}{f_0}-\frac{f_0}{f_b}\right)^2}} \cdot \frac{(3/V_{DD}+\theta)^2}{2+\frac{\theta V_{DD}}{3}} \\ &= I_D\left(1 + \frac{k_{\mathrm{QE}}-1}{k_{\mathrm{CD}}\sqrt{1+k_{\mathrm{QE}}^2 Q_{\mathrm{tank}}^2\left(\frac{f_b}{f_0}-\frac{f_0}{f_b}\right)^2}} \cdot \frac{3+\theta V_{DD}}{6+\theta V_{DD}}\right) \end{aligned} \tag{132}$$

This concludes the derivation of the different component values and bias levels. Thereby the filter design is modelled completely in closed form - making it well suited for semi automated design. Finally all the results from this section are collected in Table 8 for easy reference.

Table 8: Fully Integrated Filter Design Parameters and Total Current Consumption.

Parameter	**Expression or Trade-Off**	**Described**
$\mathrm{V_{bias1}}$	The NF is reduced by maximizing this parameter	Section 2.6.3
$\mathrm{I_D}$	The NF is reduced by maximizing this parameter	Section 2.6.3
$\mathrm{L_g}$	$\frac{\sqrt{R_s/2}}{\omega_0}\left(\frac{V_{GS1}-V_T}{\sqrt{P_{in}}\left(1+\frac{1}{1+\theta(V_{GS1}-V_T)}\right)}-\frac{2\sqrt{P_{in}}}{I_D}\right)$	Section 2.6.3
$\mathrm{C_{gsx}}$	$\frac{\sqrt{2P_{in}/R_s}}{\omega_0(V_{GS1}-V_T)}\left(1+\frac{1}{1+\theta(V_{GS1}-V_T)}\right)-\frac{I_D L_{drawn}^2(1+\theta(V_{GS1}-V_T))}{\mu_n(V_{GS1}-V_T)^2}$	Section 2.6.3
$\mathrm{L_s}$	$\frac{\sqrt{2R_sP_{in}}}{\omega_0 I_D}$	Section 2.3
$\mathrm{W_1}$	$\frac{I_D L_{eff}(1+\theta(V_{GS1}-V_T))}{\mu_n C_{ox}(V_{GS1}-V_T)^2}$	Section 2.6.3
$\mathrm{W_2}$	$3W_1$	Section 6.2
$\mathrm{W_3}$	$\frac{L_{eff}I_D(k_{\mathrm{QE}}-1)}{\mu_n C_{ox}k_{\mathrm{CD}}\sqrt{1+k_{\mathrm{QE}}^2Q_{\mathrm{tank}}^2\left(\frac{f_b}{f_0}-\frac{f_0}{f_b}\right)^2}}\cdot\frac{(3/V_{DD}+\theta)^2}{2+\frac{\theta V_{DD}}{3}}$	Section 6.2
$\mathrm{V_{bias2}}$	$\approx\frac{2V_{DD}}{3}+V_T-100mV$	Section 6.2
$\mathrm{V_{bias3}}$	$V_T+V_{DD}/3$	Section 5.2
L	$\frac{2V_{DD}}{3\pi f_0 I_D}\sqrt{\frac{1}{k_{\mathrm{QE}}^2Q_{\mathrm{tank}}^2}+\left(\frac{f_b}{f_0}-\frac{f_0}{f_b}\right)^2}$	Section 6.2
C	$\frac{3I_D}{8\pi f_0 V_{DD}\sqrt{\frac{1}{k_{\mathrm{QE}}^2Q_{\mathrm{tank}}^2}+\left(\frac{f_b}{f_0}-\frac{f_0}{f_b}\right)^2}}-\frac{C_{par}}{2}$	Section 6.2
$\mathrm{I_{total}}$	$I_D\left(1+\frac{k_{\mathrm{QE}}-1}{k_{\mathrm{CD}}\sqrt{1+k_{\mathrm{QE}}^2Q_{\mathrm{tank}}^2\left(\frac{f_b}{f_0}-\frac{f_0}{f_b}\right)^2}}\cdot\frac{3+\theta V_{DD}}{6+\theta V_{DD}}\right)$	Section 6.2

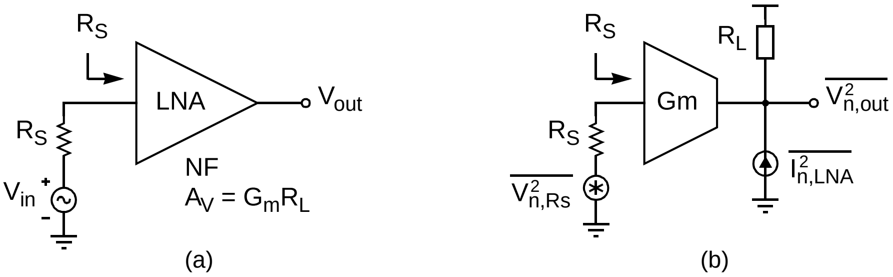


Figure 67. Single Ended LNA with input power match. General structure (a) and with noise sources (b). In the latter the noise of the LNA is referred to the output.

6.3 Filter Noise Analysis

In Chapter 2 the input power constrained LNA design methodology was developed for the single-ended case. The fully integrated front-end filter described here uses a differential version so in order to make a noise analysis of the filter we must first verify that the results from the single ended case can also be used in the differential case.

Lets start by considering one single ended LNA designed using the input power constrained design methodology developed in Chapter 2. If at first we assume that this stage is loaded by a noiseless resistor R_L then the situation is as shown in Figure 67 - (a). The noise from this LNA stage can be modeled as an output referred noise current as shown Figure 67 - (b).

The noise figure of Figure 67 - (b) is given by (133). By isolation the output referred noise current in this equation we have an expression for this noise source (134).

$$NF = 1 + \frac{\text{output noise from LNA}}{\text{output noise from source}} = 1 + \frac{\overline{I^2_{n,LNA}}R_L^2}{\overline{V^2_{n,Rs}}\left(\frac{1}{2}G_mR_L\right)^2} = 1 + \frac{\overline{I^2_{n,LNA}}}{kTR_SG_m^2} \tag{133}$$

$$\overline{I^2_{n,LNA}} = (NF-1)\cdot kTR_SG_m^2 \tag{134}$$

We can then proceed and insert two of these stages in a differential arrangement with a differential source. Remember that if the differential input stage is to have the same input impedance (R_s), input power handling capability (P_{in}) and current consumption (I_D) as the single ended version then each half circuit must have an input impedance of $R_s/2$, input power handling capability of $P_{in}/2$ and a current consumption of $I_D/2$. From (116) we know that the effective transconductance Gm is the same in both cases so the situation is as shown in Figure 68 where the mean noise voltage from the source is shown as $V_{n,RS}$.

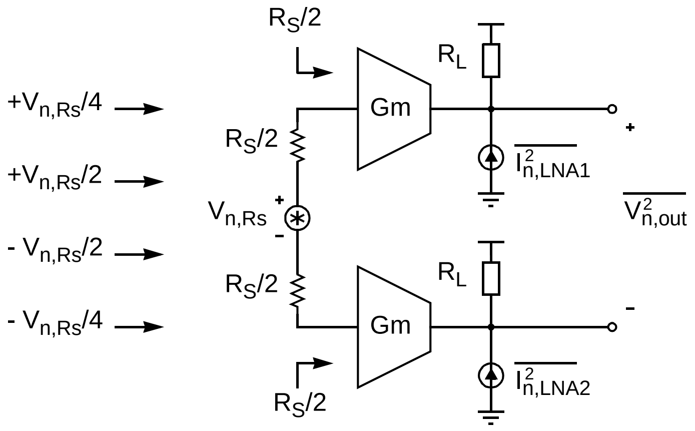


Figure 68. Noise sources in a differential LNA input stage made out of two single-ended input power constrained LNAs. The load resistors R_L are considered noiseless.

This figure allows us to calculate the noise figure of a differential input stage made out of two single-ended input power constrained LNAs. The thermal noise from the load resistors is not included at this point (this will be done later) because it is not necessary for the comparison of the single-ended and the differential versions. The noise figure is given by (135) and the result is - as expected - that the noise figure is identical to the noise figure of the single-ended case. This is interesting because it is common belief that differential LNA input stages always have higher noise figures than single ended.

$$\begin{aligned} NF_{\text{Diff}} &= 1 + \frac{\text{output noise from LNA}}{\text{output noise from source}} = 1 + \frac{\overline{I^2_{n,LNA1}}R_L^2 + \overline{I^2_{n,LNA2}}R_L^2}{\overline{\left(V_{n,Rs}\frac{1}{4}G_mR_L - \left(-V_{n,Rs}\frac{1}{4}G_mR_L\right)\right)^2}} \\ &= 1 + \frac{2 \cdot (NF_{\text{Single-End}} - 1)kT\frac{R_S}{2}G_m^2 \cdot R_L^2}{4kTR_S \cdot \left(2\frac{1}{4}G_mR_L\right)^2} = NF_{\text{Single-End}} \end{aligned} \tag{135}$$

We are now ready to find an expression for the overall noise figure of the fully integrated front-end filter. What is left is to incorporate the noise from the load. The load of the filter is not two resistors but instead a differential LC resonator which is loss compensated by a cross coupled transistor stage with capacitive division in the feedback path. This means that the impedance seen at each output is $Z_{\text{tank}}/2$ where Z_{tank} is the frequency dependent impedance of the Q-enhanced LC resonator.

The noise of this Q-enhanced resonator comes from thermal noise caused by loss in the resonator and it comes from drain noise in the loss compensating transistors. The total

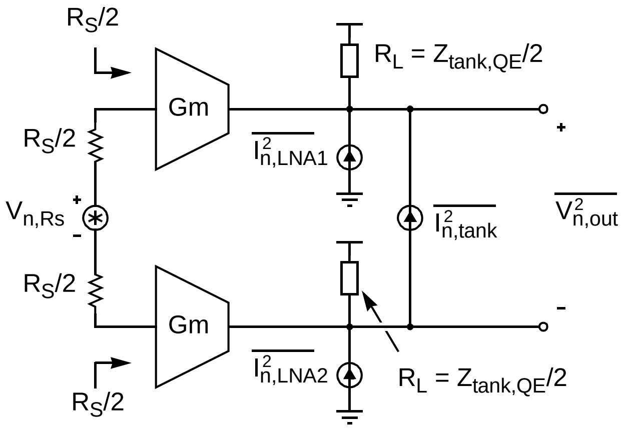


Figure 69. Noise sources in a fully integrated front-end filter with a differential low noise input stage made out of two single ended input stages and loaded with a single differential resonator which is Q-enhanced.

noise of the resonator can be modelled as a single parallel noise source with the value given by (114) as shown in Chapter 5. The noise figure of the filter takes the following form when these noise sources are all included (136).

$$\begin{aligned} NF_{\text{Filter}} &= 1 + \frac{\text{output noise from LNA}}{\text{output noise from source}} = 1 + \frac{(\overline{I^2_{n,LNA1}} + \overline{I^2_{n,LNA2}})R_L^2 + \overline{I^2_{n,\text{tank}}}(2R_L)^2}{kTR_S G_m^2 R_L^2} \\ &= NF_{\text{SE}} + \frac{4\overline{I^2_{n,\text{tank}}}}{kTR_S G_m^2} = NF_{\text{SE}} + \frac{4\overline{I^2_{n,\text{tank}}}}{kTR_S(I_D/\sqrt{2P_{in}R_s})^2} = NF_{\text{SE}} + \frac{8P_{in}\overline{I^2_{n,\text{tank}}}}{kTI_D^2} \end{aligned} \tag{136}$$

We can then insert the result from Chapter 5 i.e. (114) into this expression.

$$NF_{\text{Filter}} = NF_{\text{SE}} + \frac{8P_{in}}{kTI_D^2} \cdot \frac{4kT\left(1 + \dfrac{\gamma(1 - 1/k_{\text{QE}})}{\alpha k_{\text{CD}}}\right)}{Q_{\text{tank}}\sqrt{L/C}} = NF_{\text{SE}} + \frac{32P_{in}\left(1 + \dfrac{\gamma(1 - 1/k_{\text{QE}})}{\alpha k_{\text{CD}}}\right)}{I_D^2 Q_{\text{tank}}\sqrt{L/C}} \tag{137}$$

In order to arrive at an expression which contains the most insight it is necessary to eliminate the component values from the expression. It is possible for instance to substitute $Q_{\text{tank}}\sqrt{L/C}$ with the expression from (125). This leads to:

$$NF_{\text{Filter}} = NF_{\text{SE}} + \frac{24P_{in}\left(1 + \dfrac{\gamma(1 - 1/k_{\text{QE}})}{\alpha k_{\text{CD}}}\right)}{I_D V_{DD}\sqrt{\dfrac{1}{k_{\text{QE}}^2} + Q_{\text{tank}}^2\left(\dfrac{f_b}{f_0} - \dfrac{f_0}{f_b}\right)^2}} \tag{138}$$

Next it is possible to insert the expression for the noise figure of the single ended input (38) in this equation and then we arrive at (139).

$$NF_{\text{Filter}} = 1 + \frac{4\gamma P_{in}}{I_D} \cdot \left(\frac{\theta}{2} + \sqrt{\frac{\theta^2}{4} + \frac{1 + \theta(V_{GS} - V_T)}{(V_{GS} - V_T)^2}} + \frac{\dfrac{1}{\gamma} + \dfrac{1 - 1/k_{\text{QE}}}{\alpha k_{\text{CD}}}}{\dfrac{V_{DD}}{6}\sqrt{\dfrac{1}{k_{\text{QE}}^2} + Q_{\text{tank}}^2\left(\dfrac{f_b}{f_0} - \dfrac{f_0}{f_b}\right)^2}} \right) \tag{139}$$

$$\alpha = \frac{1 + \theta(V_{gs} - V_T)/2}{(1 + \theta(V_{gs} - V_T))^2} = \frac{1 + \theta(V_T + V_{DD}/3 - V_T)/2}{(1 + \theta(V_T + V_{DD}/3 - V_T))^2} = \frac{1 + \theta V_{DD}/6}{(1 + \theta V_{DD}/3)^2} \tag{140}$$

Finally we can eliminate α from this expression by using (23) and (130). According to these equations α can be expressed as (140) and this expression can then be inserted in (139) leading to the final expression for the noise figure of a fully integrated front-end filter (141).

$$NF_{\text{Filter}} = 1 + \frac{4\gamma P_{in}}{I_D} \left(\frac{\theta}{2} + \sqrt{\frac{\theta^2}{4} + \frac{1 + \theta(V_{\text{GS1}} - V_T)}{(V_{\text{GS1}} - V_T)^2}} + \frac{\dfrac{1}{\gamma} + \left(\dfrac{1 - 1/k_{\text{QE}}}{k_{\text{CD}}}\right)\dfrac{(1 + \theta V_{DD}/3)^2}{1 + \theta V_{DD}/6}}{\dfrac{V_{DD}}{6}\sqrt{\dfrac{1}{k_{\text{QE}}^2} + Q_{\text{tank}}^2\left(\dfrac{f_b}{f_0} - \dfrac{f_0}{f_b}\right)^2}} \right) \tag{141}$$

This expression contains device constants (γ, θ, V_T and Q_{tank}), filter requirements (P_{in}, k_{QE}, f_0, f_b) and a few parameters which can be used for noise optimization (V_{GS1}, k_{CD}, V_{DD} and I_D). We know from Chapter 2 that V_{GS1} should be maximized to minimize noise. The upper limit is set by the V_{DD} and the width of the cascode device. Practical upper limits are in the 0.9-1.1V range. The capacitive division factor k_{CD} should also be maximized to minimize noise. This factor is always smaller than one due to parasitic capacitances but even if it were possible to come very close to one it would not be desirable. The reason is that the loss compensating transistors can not handle the full resonator voltage swing at their gates due to linearity constraints. With short-channel effects (which improve linearity) the upper limit is in the order of 0.8 - 0.9. Usually the supply voltage V_{DD} is fixed but if it is not then it should be maximized to reduce the noise. The upper limit is set by reliability issues in the process. For instance the maximum supply voltage in a 0.25 micron CMOS process is typically 2.5V. Because the resonator swings above the supply voltage it may be necessary to chose a slightly lower V_{DD} value e.g. 2.4V. Finally the noise figure is reduced by increasing the current consumption I_D. Usually there is a power budget which limits this parameter.

If the maximum current consumption does not give a sufficiently low noise figure then there is only one thing to do and that is to use a different IC process which has passive components (typically inductors) with lower loss such that a higher tank quality factor Q_{tank} can be achieved.

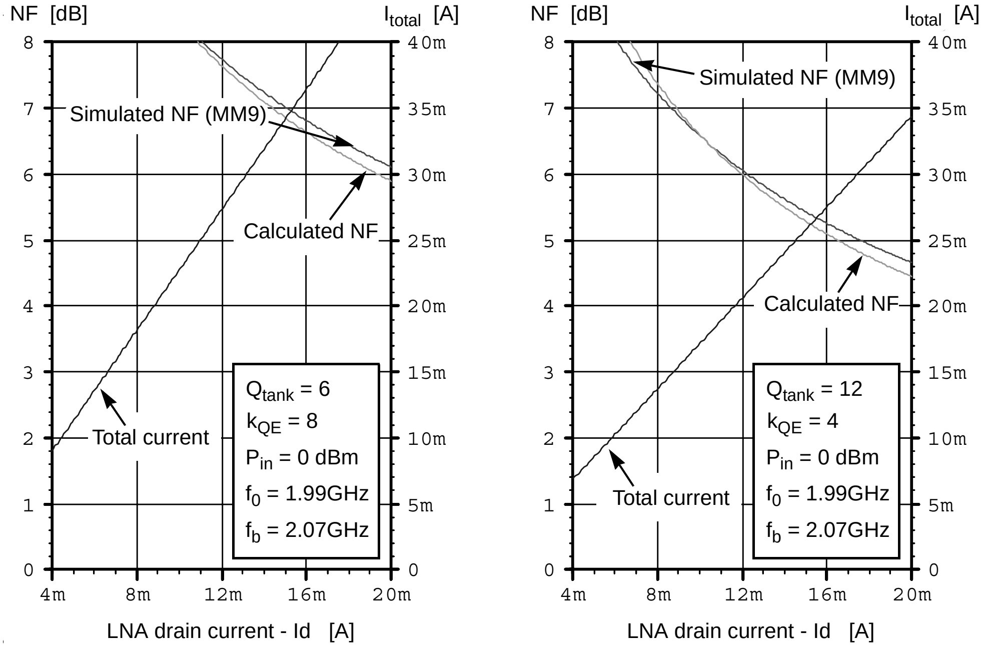


Figure 70. Noise figure and current consumption of a bandpass amplifier with a resonator Q-value of 6 (left) and with a Q-value of 12 (right). The simulated noise figure includes all noise sources except from parasitic losses in the input matching network.

Example. We want to see if GSM1900 performance can be achieved in a standard 0.25 micron digital CMOS process on an EPI substrate. According to Chapter 1 a NF of less than 8 dB is sufficient to meet the GSM requirements, but a NF of 6 dB must be achieved to meet production and temperature variations. We will further assume that an attenuation of 12 dB at the blocker frequency is required. According to (49) this corresponds to an effective tank Q-value of ~ 48. In the given process the maximum resonator Q-value is 6 as the maximum inductor Q-value is ~ 6.5 due to significant eddy current losses (see Chapter 3 for more detail). This means that the Q-enhancement factor must be 48/6 = 8 a value which is quite comfortable. The parameters for the fully integrated front-end filter are: $\gamma = 2/3$, $P_{in} = 0\text{dBm}$, $\theta = 1.3\text{V}^{-1}$, $V_T = 0.52\text{V}$, $V_{DD} = 2.4\text{V}$, $V_{GS1} = 1.0\text{V}$, $k_{CD} = 0.85$, $f_0 = 1.99\text{GHz}$, $f_b = 2.07\text{GHz}$, $R_s = 100\Omega$. Most of these are given by the GSM standard, others by the CMOS process and finally the rest are chosen according to the previous recommendations. The used circuit schematic is that of Figure 65.

A sweep is made with drain currents ranging from 4mA to 20mA. For each bias point the corresponding component values are calculated according to Table 8 and then a noise simulation is performed in the APLAC simulator using the MOS Model 9 (MM9) transistor model. The use of MM9 ensures that drain noise as well as induced gate noise is included. In the simulations all noise sources were included except from those coming from parasitic losses in the input matching network. Figure 70 - (left) shows the simulated NF along with the calculated noise figure and the total current consumption. The calculated noise figure shows excellent agreement with the simulated noise figure despite the fact that the calculated noise figure does not include the noise from the cascode transistor nor induced gate noise of any of the devices. This is satisfying as these noise sources were not expected to be significant.

The noise figure results themselves are also very satisfying as they show that GSM1900 performance can be met in a standard digital 0.25 micron CMOS process. An 8dB NF is achieved with 25mA of current consumption and the 6dB noise figure which is needed to meet production tolerances is achieved with 45mA of current consumption. This is absolutely the upper limit of what can be tolerated but nevertheless it is a very encouraging result.

In CMOS processes which are fabricated on bulk substrates or in the slightly more specialized RF CMOS processes the substrate resistivity is three orders of magnitude higher than that of EPI substrates. This means that inductors with Q-values in the order of 15 can be integrated and this translates to resonator Q-values of approximately 12. If this kind of process was used then the Q-enhancement factor k_{QE} would be 48/12 = 4. We can now repeat the simulations with these new parameters. The results are shown in Figure 70 - (right). Again there is an excellent agreement between the simulated and the calculated noise figures. This time however the current consumption that gives an 8dB noise figure is only 12mA and the current consumption which gives a 6dB noise figure is only 21mA. These values are much more tolerable for a wireless handset. Based on these numbers the conclusion is the there is strong evidence that a fully integrated GSM front filter *can* be integrated today with reasonable performance in a standard CMOS process with high substrate resistivity.

6.4 Other Considerations

In these simulations certain issues were not included. For instance the issues of frequency tuning and control of the quality factor tuning. Likewise the noise and linearity of subsequent stages has not been modelled.

Regarding the frequency tuning then a switched capacitor bank with 5-7 bits of resolution is assumed to be able to do this without adding significant losses as explained in Chapter 3. The control of the tuning is expected to be handled by the DSP, lookup tables and adaptive algorithms. That is also the case for the Q-factor tuning which is assumed to be implemented by a switched g_m configuration with 5-7 bits of resolution (see Figure 63 for more detail). The output stage is expected to be implemented as an open drain (common source) stage or as a source follower. Due to the relatively high voltage gain (Figure 66) noise contributions from these stages or from subsequent filter stages is not expected to be significant. If linearity is an issue then moderate capacitive division can be used at the output (see Figure 19 for more detail).

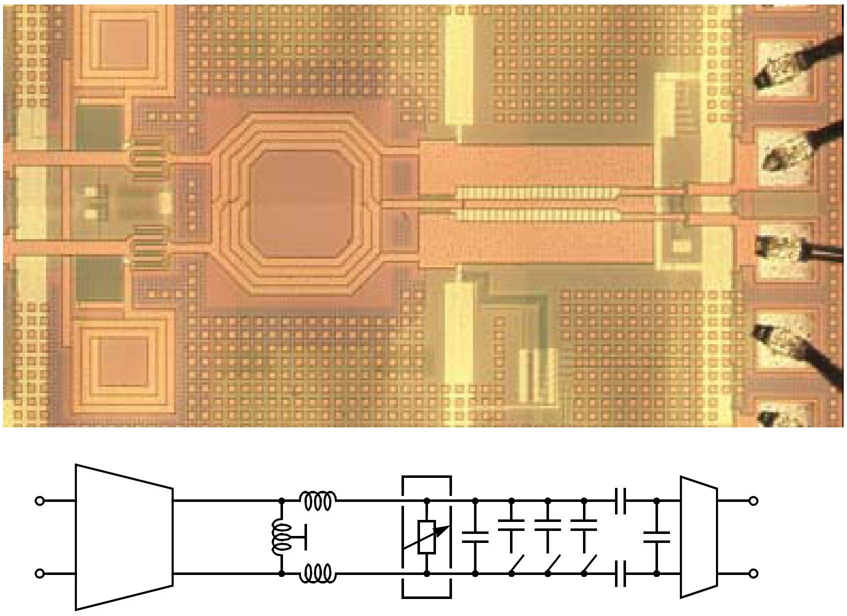


Figure 71. A fully integrated front-end filter for GSM1800/1900 and WCDMA. Chip microphotograph (top) and toplevel circuit schematic (bottom).

6.5 Example - A Fully Integrated Front-End Filter for GSM1800/1900 and WCDMA.

A demonstrator chip was designed to verify that a fully integrated front-end filter for GSM1800/GSM1900 can be implemented in a standard 0.25 micron digital CMOS process. To further underline the usefulness of an on-chip filter which is tunable it was decided to include the 3G - WCDMA receive band frequencies 2110-2170MHz.

GSM1800 is located in the 1805-1880MHz band and GSM1900 is located in the 1930-1990MHz band. Beyond these frequencies it is necessary have some production margins. Because on-chip inductors and vertical mesh capacitors have very low process variation it is sufficient to use only 2-3% design margin i.e. roughly 50-60 MHz of margin is needed. This translates to the requirement that the filter should be tunable from approximately 1750MHz to 2250MHz.

It was also decided that the filter should attenuate the 0dBm blockers by at least 10 dB relative to the wanted signal. This corresponds to a resonator Q-value of 40-50 and this again corresponds to a Q-enhancement factor of 7-8 in this process. Finally it was chosen to implement both frequency and quality factor tunability by means of switching to achieve good RF performance and a high degree of flexibility through adaptive algorithms run on a DSP. The designed filter is shown in Figure 71.

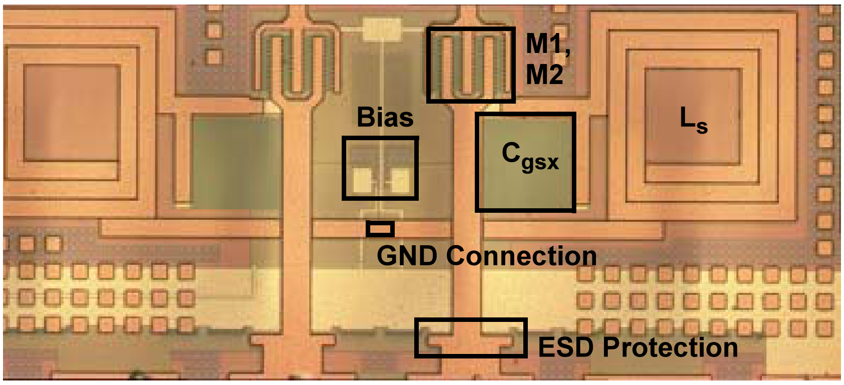


Figure 72. Microphotograph of the differential LNA input stage with ESD protection, 50 ohm differential input impedance, 10mA current consumption, 0dBm ICP and 2.5 dB (simulated) noise figure.

6.5.1 Low Noise Input Stage

The low noise input stage of the filter was designed before the theoretical work of Chapter 2 was completed so the design was synthesized in a different manner. The optimization features in the circuit simulator APLAC were used extensively. A goal of high IIP3 (simulated with harmonic balance), low noise figure, input power match and a fixed power consumption was used together with a simulated annealing algorithm. After several iterations this resulted in the parameters that were used for the above design. The components values used correspond very well with the findings of Chapter 2 despite the fact that they were found in a completely different manner.

The external gate-source capacitor C_{gsx} was implemented using a stacked inter-digitized metal-metal structure with a capacitance value of 1.2pF. The choice of a lateral flux capacitor was easy as no capacitors were provided in the given digital CMOS process. The inductors that are used for source degeneration L_s were implemented as a three-turn stacked inductor using the top-three metal layers yielding a total metal thickness of 2.6 μm and a Q-value of approximately 5. The inductance value itself is 1.4nH calculated with the equations of S. S. Mohan [64]. The input transistors as well as the cascode transistors were chosen to be 80 fingers each 1.6μm long corresponding to a total transistor width of 128μm. This is very short transistor fingers but it was necessary to mitigate the ohmic gate resistance. The bias resistors were chosen to be 50kohm to reduce the loss and noise introduced at the LNA gates to a negligible level.

The LNA was not measured stand-alone so the actual noise figure is not known but the power consumption and the gain were as expected. Also the input impedance was close to the desired 50 ohms (S_{11} ~ -9dB) and most importantly the 0dBm ICP which is very difficult to reach was achieved. The designed LNA is shown in Figure 72.

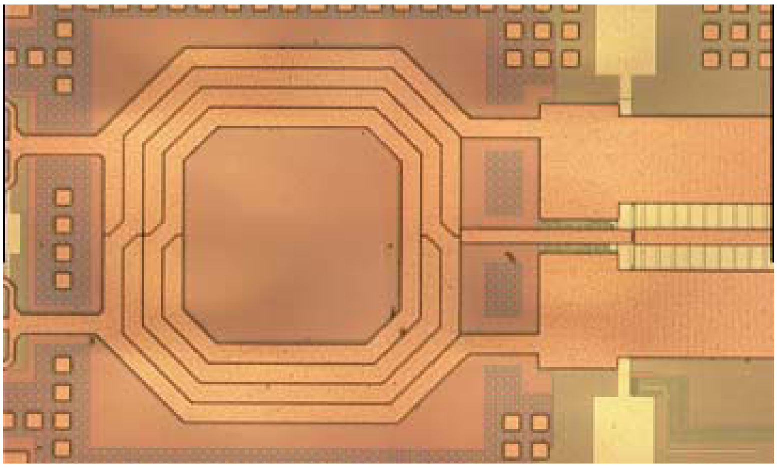


Figure 73. Microphotograph of the 1.75 - 2.25 GHz LC resonator with tapped inductor input and 7-bit switched capacitor tuning.

6.5.2 On-Chip LC Resonator

The on-chip LC resonator is perhaps the most important part of the filter because it dominates the noise of the filter or bandpass amplifier. Therefore extreme care was exercised in the design of these components. The 4.4nH inductor was described in detail in Chapter 3. It achieves a Q-value of 6.0 at 1.5 GHz and 5.8 at 2GHz i.e. the Q-value peaked a little before the desired frequency but it is still a very good result - perhaps the best ever achieved in a standard CMOS process with 10mΩcm substrate resistivity. The low inductor Q-value implies that it is simply not tolerable that the capacitor or any of the other circuitry connected to the resonator contributes with significant loss. This means that the capacitor with everything connected to it (LNA, output buffer, switched capacitor bank and loss compensation) should have an equivalent Q-value above 50. This is almost impossible but that was the goal of the resonator design. First high-Q vertical mesh capacitors were developed. They are described in Chapter 3 and [65]. Then it was investigated how MOS switches can be optimized to yield the best performance in a switched capacitor bank. This was also described in Chapter 3 and in [66]. And then finally moderate inductor tapping was used at the input and capacitive division was used at the output (see Figure 19).

It is possible to measure the uncompensated resonator quality factor by simply observing the bandwidth of the bandpass amplifier with the loss compensation turned off. The measured resonator Q-value was 5.5 and therefore the equivalent Q-value of the capacitor with everything else connected to it must be in excess of 50 - a very satisfying result. It was almost equally satisfying that the measured resonance frequency was within 50 MHz of the desired frequency. This is exceptionally good - or lucky - as all inductances and capacitances (including parasitics) were calculated by hand.

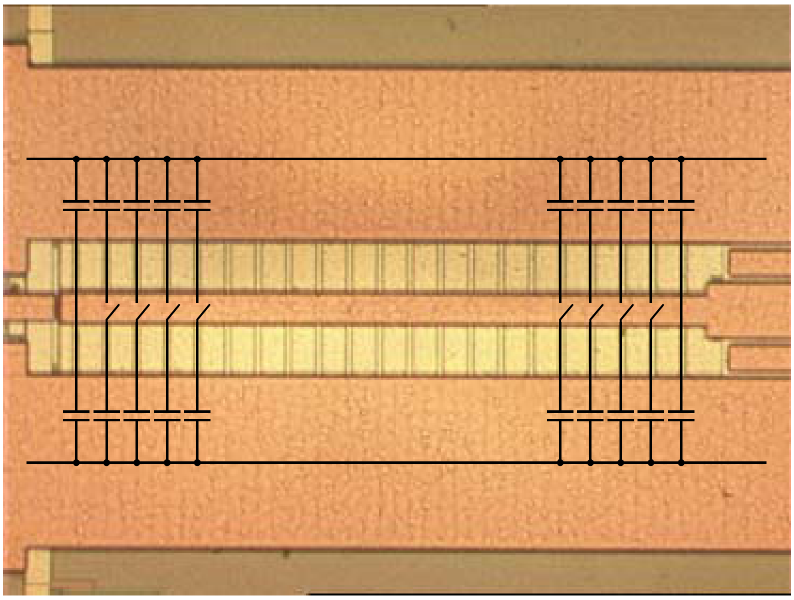


Figure 74. Microphotograph of the 7-bit switched capacitor bank which implements frequency tuning from 1.75-2.25 GHz with a resolution of 4MHz.

6.5.3 Switched Capacitor Frequency Tuning

The frequency tuning of the filter was implemented with a switched metal-metal capacitor bank to achieve low loss and high linearity. The bandwidth of the filter is approximately 2GHz/50 = 40MHz. This means that the frequency tuning system should have a resolution of no more than 4MHz to avoid attenuation of the wanted signal. The total tuning range is 500MHz so this requirement corresponds to a 7-bit resolution.

Achieving a 7-bit resolution at 2GHz is not trivial. Some would even question that it is possible. The reason is that at lower frequencies it is mainly unit capacitances that need to be matched but for this application it is also important that the parasitic inductances and resistances are well matched. Furthermore it is important that the switches are matched because they introduce significant parasitic components.

The switched capacitor bank (Figure 74) was implemented using 63 unit capacitors and one half ($32C_{unit}$, $16C_{unit}$, .. , $1C_{unit}$, $0.5C_{unit}$). The unit capacitor was designed as a 2.2 µm wide shielded vertical mesh capacitor with a value of 30fF (Chapter 3, [65]). This corresponds to 0.96pF for the most significant bit and 15fF for the least significant bit. The existence of such a narrow shielded unit capacitor is probably what makes 7-bit resolution possible because all the capacitors can be laid out in parallel in a common centroid style in only 140µm. More unit capacitors are laid out as dummy structures which at the same time provide the desired amount of fixed capacitance. Finally the switches were implemented as matching differential waffle switches Figure 53.

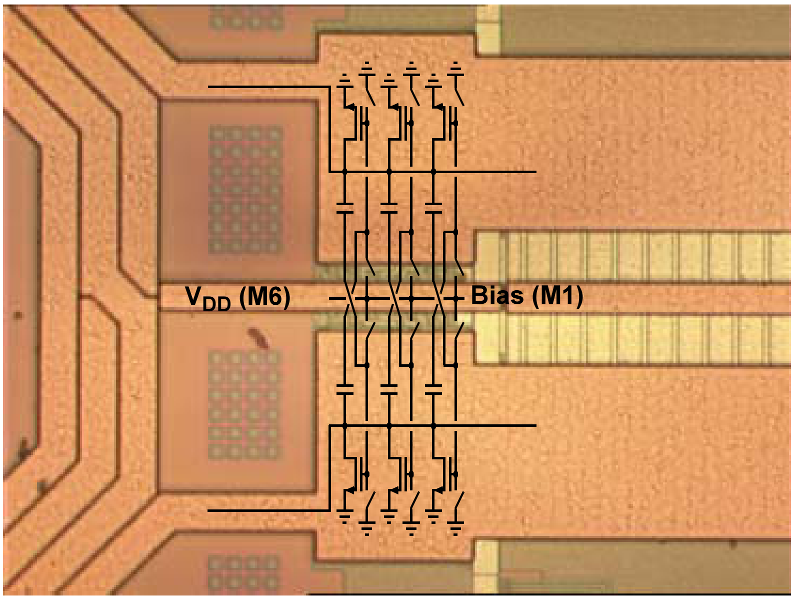


Figure 75. Microphotograph of the 7-bit switched-gm loss compensation circuit.

6.5.4 Switched Quality Factor Tuning

The purpose of the loss compensation circuit is to raise the Q-value of the resonator from approximately 6 to approximately 48 i.e. to implement a Q-enhancement factor of approximately 8. If we chose to accept that the filter can be tuned to a resonator Q-value between 45 and 51 then a resolution of 6 bits is needed. It was decided to increase this resolution to 7 bits by doubling the maximum loss compensation conductance. This was done to make sure that the filter would still be functional in case the uncompensated resonator Q-value turned out to be overestimated by as much as 100% i.e. to make sure the filter would function with a resonator Q-value down to 3. Furthermore it was chosen to use the loss compensation topology which is shown in Figure 63 - (b). This choice was made to secure the best possible dynamic range for the filter.

The layout of this loss compensation circuit is relatively involved because there are many cross couplings and parasitic effects that need to be well matched. For instance the capacitive division in each cross coupling branch must be the same and in the order of 0.85 - 0.9. This was done by sizing the capacitors such that they were ~ 10 times larger than all the parasitic capacitance on the gate nodes in each of the 7 binary weighted loss compensation networks. The shielded vertical mesh capacitors were also very useful for this purpose. The loss compensation transistors were implemented as combinations of unit transistors with W=1.2µm and the two least significant bits were implemented with transistors with W=0.6µm and W=0.35µm. Finally the MOS switches to the biasing node were chosen small enough to function as biasing resistors when turned on.

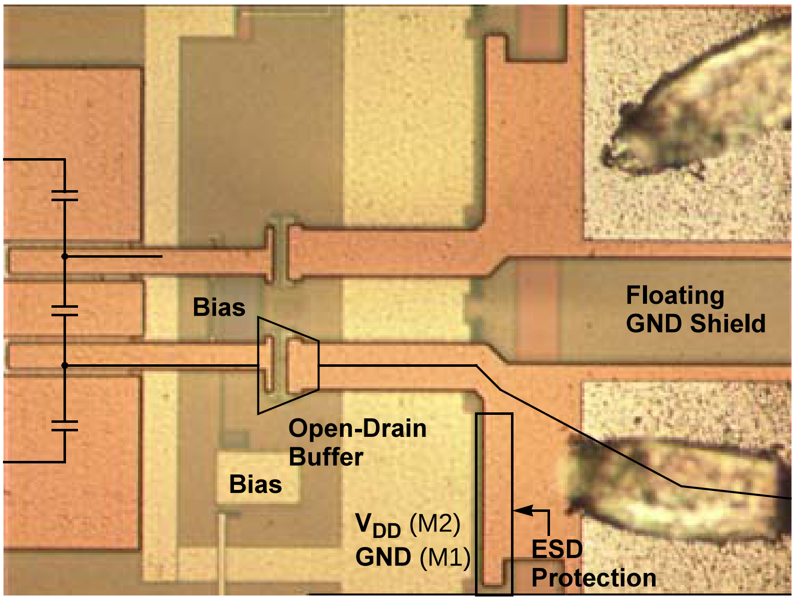


Figure 76. Microphotograph of the open-drain output buffer with ESD protection and a floating ground shield under the balanced output pads.

6.5.5 Output Buffer and ESD Protection

The output buffer of the circuit (Figure 76) is not limited by noise due to the relatively high voltage gain of the filter. The linearity on the other hand is a limitation due to the large voltage swings in the resonator. For this reason it was chosen to use a capacitive divider at the input of the output buffer leading to a moderate reduction of the input voltage swing without affecting the overall noise figure of the filter. The buffer itself was chosen to be an open-drain structure because it enables biasing for optimal linearity. Also the load of the output buffer is not a natural part of the filter because if the filter were to be used in a highly integrated transceiver IC then the load would differ significantly from the 50 ohms used in the off-chip environment.

Underneath the differential output pads there is a floating ground shield. The reason why it is there is that it gives a well defined differential impedance, low loss and reduced coupling to the substrate. This scheme was also used for the LNA input where these issues are particularly important. Finally some limited ESD protection was implemented on all pads to ease the testing of the chip and to accommodate a more realistic situation. The ESD protection was implemented using 20 NMOS and 20 PMOS ring transistors with gate and source connected to GND / V_{DD} and the drain connected to the pad. Also V_{DD} and GND were connected with a chain of diode connected NMOS transistors. The measured capacitance of the RF pads was 160 fF with a very high quality factor (>50 at 2 GHz) which is sufficient to not degrade the performance of the filter.

6.5.6 Measurement Results

The first version of the fully integrated front-end filter was fabricated and tested. Unfortunately it was not possible to turn on the loss compensation circuit in this version and therefore it was not possible to measure the IC as a filter. The reason remains unknown but it is most likely the bias circuit that was faulty because the loss compensating transistors were not drawing any current. One possible explanation is that antenna rule violations may have caused increased leakage currents pulling the bias level low. The antenna rule violations were identified at a much later stage because they were not included in the design rule checker (DRC). In the second version of this filter the antenna rule violations were resolved and the bias circuit was moved off-chip to minimize risk. In the switched capacitor bank it was not possible to turn the switches off. Attempting to do so resulted in a slightly higher resonance frequency and a lower resonator Q-value. This corresponds to the switch being turned off partially. The reason is believed to be that a diode connected pull-up circuits is not turned off completely - again due to much higher than anticipated leakage currents. In the second version of the chip this pull-up circuit is removed all together. Apart from these two problems all other aspects seemed to be very close to the simulated and calculated behavior. The second version of the chip has been fabricated but the measurements have not yet been made.

6.6 Summary

This chapter presented a complete analysis of a fully integrated front-end filter. The analysis which uses the main results from all the previous chapters provides an expression for the transfer function and closed form expressions for all the component values and bias levels. Furthermore it presents a complete noise analysis which directly gives the current consumption that is needed to achieve a given noise performance in a given CMOS process. This expression indicates that GSM performance is achievable in a standard CMOS process on a bulk substrate. Finally the chapter presents the details of a fully integrated front-end filter demonstrator IC for GSM1800/1900 and WCDMA.

6.7 References

[61] B. Razavi, "RF Microelectronics", *Prentice Hall*, 1998.

[62] H.T.Friis, "Noise Figure of Radio Receivers," *Proceedings IRE*, Vol. 32, pp. 419-422, July 1944.

[63] J. Burghartz, et. al., "RF Circuit Design Aspects of Spiral Inductors on Silicon", *in Proceedings of the 1998 ISSCC*, pp. 246-247, 1998.

[64] S.S. Mohan, "The Design, Modelling and Optimization of On-Chip Inductors and Transformer Circuits," *Ph.D. dissertation, Stanford University*, California, 1999.

[65] K.T. Christensen, "Design and Characterization of Vertical Mesh Capacitors in Standard CMOS," *Dig. of the 2001 VLSI Circuits Symposium*, pp.201-204, 2001.

[66] K.T. Christensen, "Design and Optimization of CMOS Switches for Switched Tuning of LC Resonators," *in Proceedings of the 2001 IEEE ISCAS*, Vol. 4, pp. 770-773, May 2001.

Chapter 7: Low Power LC Quadrature Generation for Image Rejection

When a superheterodyne receiver architecture is used instead of the direct conversion architecture then another RF filtering problem arises. This is the problem of attenuating the signal at the image frequency which will otherwise be folded on top of the desired signal and perhaps corrupt it. Traditionally image-rejection is achieved through complex demodulation using so-called image-reject mixers. These mixers need accurate quadrature signals in the RF path or in the LO path or in both. The generation of quadrature signals typically involves serious noise-power trade-offs but image-reject mixing is never-the-less quite common. This chapter presents a new method for generating quadrature signals in integrated circuits using only inductors and capacitors. This LC quadrature generation method enables significantly reduced noise and power consumption due to the absence of resistors in the signal path.

7.1 Introduction

A number of different topologies exist that can be used for complex modulation and demodulation in radio frequency communication systems. Some of these topologies are shown in Figure 77 and Figure 78. The topologies shown in Figure 77 use either the RF or the LO signal in quadrature [70] and the double quadrature topology [71] shown in Figure 78 uses both the RF and the LO in quadrature either to loosen the requirements of the quadrature generating circuits or to improve the accuracy of the down converted signals.

Common to these topologies is the need to have one or more of the high frequency signals in quadrature i.e. one version *In-Phase* and one version in *Quadrature* i.e. shifted 90 degrees and if the circuit uses differential signalling which is typically the case this translates to the need for four identical signals each shifted by 90 degrees.

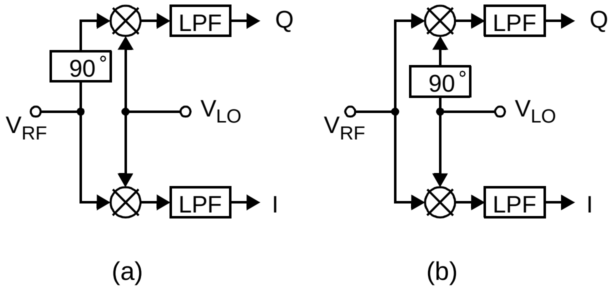


Figure 77. Quadrature mixing with quadrature generation in (a) RF path, (b) LO path.

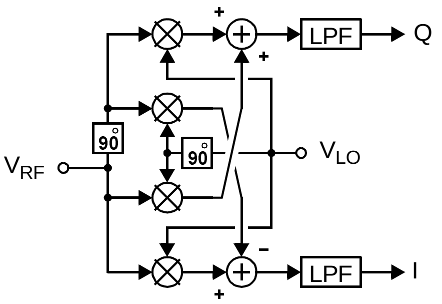


Figure 78. Double quadrature downconversion.

Such four quadrature signals can be generated in a number of different manners. The local oscillator signal (LO) can be generated in quadrature with a quadrature VCO or by dividing a double frequency VCO. Alternatively it can be generated with an RC polyphase filter. For the RF signal the only method that is used at present to generate quadrature signals is the RC polyphase filter.

All of these methods have their drawbacks. The Quadrature VCO and the double frequency VCO with dividers are fairly power hungry solutions and the resistors in RC polyphase filters introduce significant thermal noise and loss which results in a need for more high frequency amplification and thus added power consumption. Due to the noise-power implications of the RC polyphase filter the topology in Figure 77 - b is by far the most common today [70].

In order to reduce the power consumption needed for the topology in Figure 77 - b and in order to make the other two topologies more feasible what is needed is a quadrature generating circuit that introduces negligible noise and power consumption of its own. Such a circuit is presented in this chapter. It eliminates the thermal noise and loss of the RC polyphase filter by using only capacitors and inductors and it may be considered very low power because it can be integrated as part of a high frequency buffer that may be needed in any case.

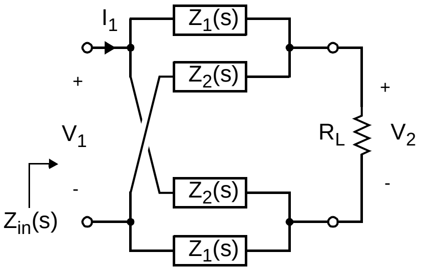


Figure 79. Symmetrical resistance reciprocal network.

7.2 Allpass Filters

When designing classic filter structures like Butterworth, Chebyshev or elliptical filters, synthesis is made without consideration of the phase response given by the network. In some cases a nonlinear phase response is of minor importance but in other cases such as digital communication systems that use phase modulation the signal may be seriously distorted by a non-linear phase response. In order to get a linear phase or equivalently a constant group delay it is often necessary to use phase linearizers. Typically such circuits should not change the filters magnitude response but only affect the phase response. This means that they must have a unity gain independent of frequency i.e. that $|H(j\omega)| = 1$. Because phase linearizers pass signals at all frequencies they are usually called allpass filters and because the sole purpose of these networks is to change the phase of a signal they are a natural place to start our search for a suitable quadrature generating circuit.

Allpass filters can be realized in a number of different manners. These include single ended LC networks using ideal transformers or center tapped inductors, active circuits exploiting negative feedback and symmetrical cross coupled resistance reciprocal networks. The latter requires a well balanced differential input signal, good matching between circuit elements and more reactive components than what would be needed for some of the other methods. In a discrete realization all of these three requirements may be problematic but on an integrated circuit well balanced differential signals and good component matching is readily available and the differential structure does not double the cost of the IC. The fact that the cross coupled resistance reciprocal network (Figure 79) exploits the nature of the differential signal makes the circuit especially attractive for an integrated circuit where absolute component tolerances are usually fairly high. For these reasons we will take a closer look at the usability of this structure as a quadrature generating network.

It can be shown [69] that for the circuit in Figure 79 to behave as an allpass filter it must be resistance reciprocal i.e. satisfy (142) and $Z_1(s)$ and $Z_2(s)$ must both be LC reactance networks.

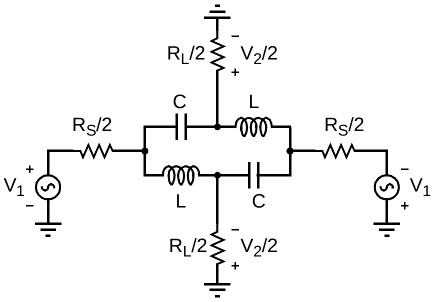


Figure 80. Four phase network made from a first order LC allpass filter.

$$\frac{Z_1(s)}{R_L} = \frac{R_L}{Z_2(s)} \tag{142}$$

If this is the case then the circuit will have unity gain and an input impedance $Z_{in}(s)$ that is equal to the load resistance R_L. Therefore the source impedance R_S must be equal to the load impedance to ensure maximum power transfer (143).

$$R_S = Z_{in}(s) = \frac{V_1}{I_1} = R_L \tag{143}$$

If $Z_1(s)$ and $Z_2(s)$ are made out of one reactive element each the network is called a first order passive allpass filter. If instead they are each made out of a series or a parallel combination of an inductor and a capacitor the network is called a second order allpass filter. Because these allpass networks have unity gain and an input impedance that is equal to the load impedance they can be cascaded without loosing their allpass nature. Therefore allpass filters with an arbitrary order can be made by cascading a number of first and second order allpass filter stages.

7.3 Quadrature Phases

The whole point of considering phase shifting networks is the target of generating quadrature phases i.e. four signals with equal amplitude and phases separated by 90 degrees. In order to minimize complexity we will start by considering first order allpass structures i.e. $Z_1(s) = sL$ and $Z_2(s) = 1/sC$ or vice versa (Figure 80).

For a cross coupled first order allpass network it can be shown [68] that if $\omega_0 L = 1/\omega_0 C$ i.e. at the resonance of an LC tank (144) the output will have a phase that is exactly the average of the two driving phases so if the driving signal is a perfect differential signal (180 degrees) the output is rotated 90 degrees compared to the input.

$$\omega_0^2 = \frac{1}{LC} \tag{144}$$

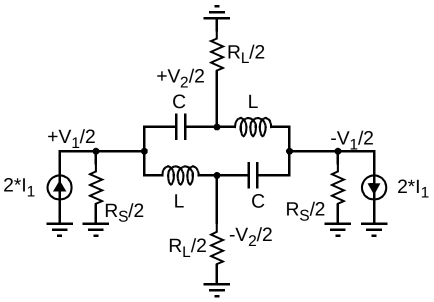


Figure 81. Current mode four phase network based on a first order LC allpass filter with matching resistive loads.

To get from the structure shown in Figure 79 to the circuit shown in Figure 80 a few extra considerations need to be made. Because the driving signal is expected to be a perfect differential signal the source including the source resistance can be split in two.

Likewise the load resistance can be split in two for symmetry reasons. It is now more intuitive how four different phases with equal amplitude are generated.

In order to get unity gain, maximum power transfer and quadrature phases (142), (143) and (144) must be fulfilled. Further because we are considering a first order allpass structure $Z_1(s) = sL$ and $Z_2(s) = 1/sC$ or vice versa. This enables us to formulate a set of simple design equations that guarantee quadrature signals.

$$C = \frac{1}{2\pi f_0 R_L} \tag{145}$$

$$L = \frac{R_L}{2\pi f_0} \tag{146}$$

7.4 Norton Transformation

We are not yet completely satisfied with the quadrature generating network shown in Figure 80. It does generate quadrature phases but if subsequent signal processing stages are to benefit from the quadrature signals they need to present an infinite impedance to the input of the allpass network because otherwise the quadrature relation is not maintained. Furthermore matching the source impedance R_S to the load impedance R_L is not trivial when these impedances arise from different physical effects.

By transforming the source to its Norton equivalent (Figure 81) it becomes apparent how these problems can be solved. If the signal is injected as a current from a source with an infinitely high impedance $R_S/2$ and $R_L/2$ can *both* be implemented as well matching resistors or as identical resistive loads from subsequent stages.

The differential current mode source can be implemented with a classic integrated circuit component - the basic differential pair with or without a tail current.

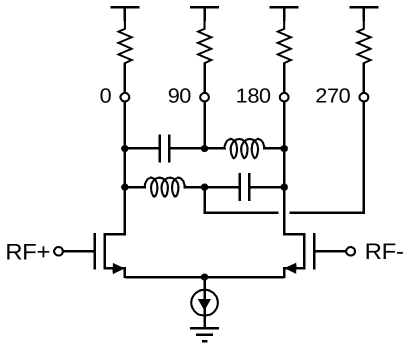


Figure 82. Quadrature generating RLC network with a basic differential pair and matching resistive loads.

Figure 82 shows the resulting circuit which is much more satisfying and practical than the circuit shown in Figure 80 because it uses standard integrated circuit components and ensures good matching between the loads.

7.5 Common Gate Amplifier Load

Despite the fact that the circuit in Figure 82 is satisfying as it is there are still a number of issues that may be resolved or improved. If each resistive load is replaced by a common gate amplifier (Figure 83) with $gm = 2/R_L$ several benefits can be achieved.

Most important is the buffering effect of common gate amplifiers. With a common gate amplifier the quadrature generation becomes independent of the subsequent stages input impedance. Almost equally important is the common gate amplifiers impedance transforming capabilities. At the lower GHz range it may be necessary to reduce the load impedance seen by the allpass network to 50Ω or less to avoid getting too large inductances.

Similarly in many cases it is beneficial to have a higher output impedance e.g. 100Ω or more in each branch which can easily be accommodated by a common gate amplifier. In short the common gate amplifier makes the impedance matching much easier to handle. An additional benefit of this structure is that the load impedance seen by the allpass structure $R_L/2 = 1/gm$ can now be tuned by simply adjusting the bias current.

As a final note it can be pointed out that center tapped inductors (top of Figure 83) can be used for biasing and thereby completely eliminate the need for lossy and noisy resistors in the quadrature generating network. The next stage can then use a shunt and/ or a series capacitor to achieve the appropriate tuning.

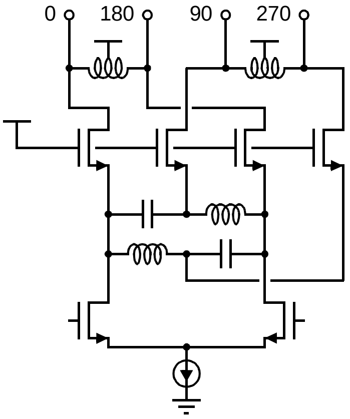


Figure 83. Quadrature generating circuit with common gate amplifier load.

7.6 Parasitics

Naturally passive components on an integrated circuit are not ideal components therefore we must take a look at how these non-idealities affect the quadrature generating network and possibly how they can be circumvented.

Perhaps the most obvious non-ideality is the poor quality of on-chip inductors. At present inductors with Q-values of 6-7 can be achieved on EPI substrates and Q-values of 14-16 can be achieved on bulk substrates in the lower GHz range, but much better Q-values (up to 35) are expected with upcoming semiconductor technologies. In the meantime we need to address this issue. A low Q-value implies that the allpass circuit can not be considered lossless and therefore it does not behave as an ideal allpass filter. The inductor loss introduces small but significant phase and amplitude errors. It can be shown [68] that for the first order allpass structure the amplitude and phase errors can be completely out compensated by introducing an equal amount of loss in the capacitance branch e.g. as a series resistance R_C.

Also the design equations need to be modified as specified in (147), (148) and (149).

$$C = \frac{(Q^2 + 1)}{2\pi f_0 R_L (Q^2 - Q)} \tag{147}$$

$$L = \frac{R_L (Q^2 - Q)}{2\pi f_0 (Q^2 + 1)} \tag{148}$$

$$R_C = \frac{2\pi f_0 L}{Q} \tag{149}$$

Another well know parasitic element of on-chip inductors is the parasitic capacitance across the terminals of the inductor. All on-chip inductors with reasonable Q-value will have a small but noticeable amount of parasitic capacitance (e.g. 10% of C). This capacitance can be considered an expansion of the first order allpass network to a second order allpass network. Therefore the effect of the parasitic inductor capacitance can be compensated for by simply adding a small series inductance to the capacitor C. Strictly speaking this is fortunate because a capacitor can not be realized without some parasitic series inductance of its own.

These allpass quadrature generating circuits provide a good quadrature output over a relatively narrow frequency band. Therefore it is important that the absolute values of L, C and R_L are well controlled. Fortunately on-chip inductors have very low process variation (<1%) and capacitors when implemented as lateral flux capacitors [72] also have low process variation (<3-5%).

This means that the center frequency can be made with only 2-3% process variation. R_L will have higher process variation but this parameter may be calibrated either by using a switched resistor array or by adjusting the bias current when the common gate amplifier load is used.

Therefore it should be possible to achieve sufficient accuracy for several narrow band and low cost applications like for instance GPS, ISM and DECT - especially when used with a double quadrature topology (Figure 78).

7.7 Example

A short example is given here to demonstrate how these design equations can be used. We want to design a quadrature generating circuit like the one in Figure 82 with a perfect quadrature output at 2.0 GHz. It is assumed that $R_L/2 = 100\Omega$, that the transistors are ideal and that the inductors have a Q-value of 8. Now (147) - (149) are used to find that $C = 0.462\text{pF}$, $L = 13.7\text{nH}$ and $R_C = 21.5\Omega$ where the latter is the series resistance that is added to the capacitive branches in the first order allpass network.

Figure 84 shows the amplitude and phase response of the resulting circuit. Notice that the circuit as expected has a relatively narrow band in which it provides 90 degrees phase shift and an excellent amplitude match over a very wide frequency range. This is opposite to the behavior of RC polyphase filters that only have a relatively narrow frequency band in which the amplitudes are well matched.

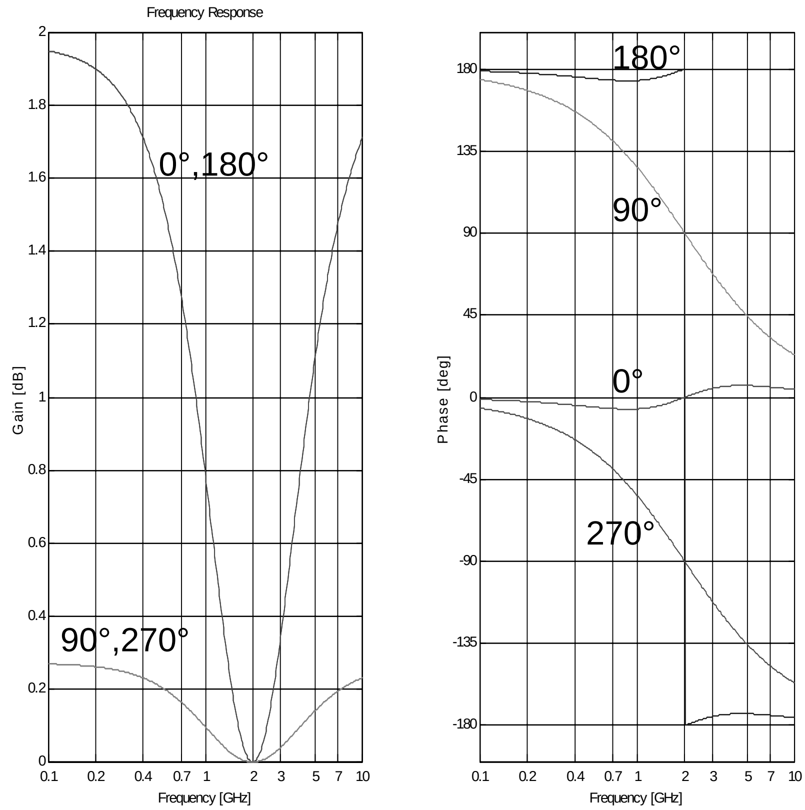


Figure 84. Simulated RLC quadrature generating circuit.

7.8 Summary

This chapter presented a new quadrature generating circuit based on LC allpass networks. Because the circuit does not use resistors to achieve the phase shifts it does not have the same thermal noise, loss and power consumption problems that RC polyphase filters do. Therefore this circuit enables considerable reduction in power consumption. Also it makes the double quadrature down-conversion scheme [71] much more attractive due to reduced noise in the RF path. The chapter discusses the functionality of the circuit and finally a loss compensation method is presented that enables compensation of the on-chip inductor losses.

7.9 References

[67] K.T. Christensen, “LC Quadrature Generation in Integrated Circuits”, in *Proceedings of IEEE Int. Symp. on Circuits and Systems*, vol. 1, pp. 41-44, May 2001.

[68] K.T. Christensen, “RLC Polyphase Filters in Silicon Integrated Circuit Technology”, *Nokia Mobile Phones Internal Report*, June 1999.

[69] H. Gaunholt, “Analoge og Digitale Filtre”, *48205 Class Notes - Technical Univ. of Denmark*, pp. 76-85, August 1998.

[70] B. Razavi, “RF Microelectronics”, *Prentice Hall*, 1998.

[71] J. Crols and M. Steyaert, “CMOS Wireless Transceiver Design”, *Kluwer Academic Publishers*, 1997.

[72] H. Samavati et al., “Fractal Capacitors”, *Journal of Solid-State Circuits*, Vol. 33, No. 12, December 1998.

[73] F. Behbahani and A. Abidi, “Analog RC Polyphase Filter and Mixer Design for Large Image Rejection”, *AACD Workshop Proceedings*, 1999.

[74] M. Gingell, “Single Sideband Modulation using Sequence Asymmetric Polyphase Networks”, *Electronic Communication*, vol. 48, no. 1-2, pp. 21-25, 1973.

[75] S. Maas, “Baluns and Hybrids” in *RF and Microwave Circuit Design for Wireless Communication, Artech House*, pp. 248-255, 1996.

Chapter 8: Conclusions

Today most radio modem functionally can be integrated on a few silicon dice. Passive RF filters are the most important exceptions because their number and overall cost scale linearly with the number of supported frequency bands and radio standards. The current trend of multi-band and multi-mode functionality will make passive off-chip filtering responsible for 20-30% of the total RF engine cost. It is the elimination of this cost that has been the main motivation behind the work presented in this thesis.

The reason why only very limited RF filter functionality is integrated today is that it has not been possible to achieve sufficient dynamic range with acceptable power consumption. As a result it has been the main focus of this work to identify design techniques which can enable sufficient dynamic range at GHz frequencies with low power consumption. This work has lead to significant advancements in the design of LNAs, passive on-chip components and switched tuning. Together these advancements should enable full integration of most RF filtering functions in most wireless terminals.

This final chapter summarizes the results of the work presented in this thesis and gives suggestions for future work.

8.1 Summary

The first chapter was an introduction to RF filtering in wireless terminals. It started with an overview of the trends in the consumer market which will dictate hardware development in the near future. It was pointed out that RF filtering is one of the major bottlenecks in the further development. The next section listed the most common RF filtering requirements from a system point of view and then the filtering solutions which are used at the moment were summarized. The chapter was concluded by listing the future RF filtering challenges in integrated circuit design.

The first circuit component in an RF receiver for a wireless communication system is typically an LNA. If the RF front-end filter is to be fully integrated on the receiver IC then the LNA needs to have extremely high dynamic range due to the existence of 0dBm blocking signals. Currently such dynamic range is only seen in base station applications and only with very high power dissipation. This problem was addressed in Chapter 2 where a new input power constrained LNA design methodology was presented. This methodology provides closed form expressions for all component values and bias levels for a given input impedance, input blocker power and current consumption. The noise optimization which is embedded in these design equations enables an order of magnitude lower power consumption while only degrading the noise figure by approximately one dB. This makes the input power constrained LNA design methodology the first LNA design methodology to simultaneously take power dissipation, input power handling capability and noise figure into account.

The most important parameter in the design of RF filters is the quality factor of the inductors and capacitors which are used. This is because high Q-values enable resonators with low noise, low power dissipation and good selectivity. Chapter 3 dealt with these issues. First different LC resonator options were reviewed and comments and design equations were presented that are relevant for the design of fully integrated RF bandpass filters. Then it was explained how good RF capacitors and inductors can be designed in a standard CMOS process with low substrate resistivity. Most important is perhaps the development of the Vertical Mesh Capacitor, a new metal-insulator-metal capacitor, which is compatible with standard digital CMOS processes and has excellent RF performance. It has high Q-value, very low process variation, low bottom plate capacitance and very narrow shielded unit capacitors which are ideal for switched capacitor banks with high resolution at high frequencies. Finally measurement results from several test chips were presented and especially the measurement results from a capacitor and an inductor that were designed for a tunable bandpass filter for 1750 - 2250 MHz were encouraging. Both the capacitor and the inductor appear to be state-of-the-art for standard CMOS processes with low substrate resistivity.

A narrowband filter is useless without frequency tuning because of process variations and because most systems apart from GPS applications need to cover a frequency band which is significantly wider than the signal bandwidth. Also the multi-band and multi-mode trend increases the need for a wide frequency tuning range. The tuning must be implemented with very low loss and high linearity to maintain the required dynamic range. This is not trivial because the varactor tuning principle commonly used in VCOs may give too much distortion and loss. Also a switched capacitor solution was by many not believed to be feasible due to loss in the MOS switches and insufficient resolution of the switched capacitor bank. Chapter 4 presented solutions to the problems associated with switched capacitor tuning at RF frequencies. First a rigorous study of switched capacitance tuning of LC resonators was presented and based on this analysis, symbolic expressions were derived for the optimal size of the MOS switch as well as for the achievable resonator Q-value. Then these results were used for an analysis of the optimal switch design and layout. This work resulted in the development of an Enhanced Differential Waffle Switch which enables wide tuning range with a very high resonator Q-value.

Even with state-of-the-art on-chip inductors and capacitors the achievable quality factors are too low to give useful selectivity in passive networks. This almost invariably leads to the use of loss compensation techniques which was the topic of Chapter 5. The requirement for extremely high dynamic range translates to a requirement for low noise and good large signal behavior for the loss compensation circuit. Such low noise and high-swing loss compensation techniques are well known from the oscillator world so it was natural to learn from their example. Two additional requirement which are needed for filters are the needs for linear behavior and accurate control of the amount of loss compensation. Traditionally this is handled by adjusting a bias current but in deep-submicron devices biased at high overdrive voltages this is not possible due to mobility degradation. Chapter 5 presented most standard loss compensation techniques along with a few new ones. Especially one based on switched bias tuning in a binary weighted transistor array was pointed out because it is believed to enable the best possible dynamic range at all gain settings. Finally a few control systems were presented that can perform the needed Q-tuning - some with the use of a digital signal processor and some without.

There is at present no published design methodology which can be used for the design of a fully integrated front-end filter. This problem was addressed in Chapter 6 where a complete analysis of a fully integrated front-end filter was presented. The analysis which uses the main results from all the previous chapters provides an expression for the transfer function and closed form expressions for all the component values and bias levels. Furthermore the chapter presented a complete noise analysis which directly gives the current consumption that is needed to achieve a given noise performance in a given CMOS process. This expression indicates that GSM performance is achievable in a standard CMOS process on a standard bulk substrate. Finally the chapter presented the details of a fully integrated front-end filter demonstrator IC for GSM1800/1900 and WCDMA.

For receivers which use the superheterodyne architecture i.e. that convert the wanted signal to an intermediate frequency before it is converted to baseband it is not always sufficient to have a front-end filter. Often the image frequency needs to be rejected by a significant amount. This can be done with an image reject mixer of which especially the double quadrature mixer, which uses both the RF and the LO in quadrature, is efficient. Unfortunately quadrature generation in the RF path involves serious noise-power trade-offs with the only known topology; the RC polyphase network topology. Chapter 7 presented a new quadrature generating circuit based on LC allpass networks. Because this circuit does not use resistors to achieve the phase shifts it does not have the same thermal noise, loss and power consumption problems that RC polyphase filters do. Therefore this circuit should be able to improve image rejection and power consumption in many applications which use the superheterodyne receiver architecture.

As a final conclusion it seems reasonable to claim that this thesis represents significant advancements in RF CMOS circuit design with special contributions in the fields of input power constrained LNA design, RF capacitor design, inductor design, MOS switch design, fully integrated front-end filter design and quadrature signal generation.

8.2 Future Work

As in most other areas of integrated circuit design there are plenty of issues to be looked at and improvements to be made in RF filtering. This is partially due to the constant silicon process developments and partially due to the fact that RF IC design for mass production is a relatively young field.

The most critical parameter for the integration of RF filters is the low quality factor of on-chip inductors. Fortunately bulk substrates with approximately 10 Ωcm resistivity are becoming common in both digital and analog CMOS processes as well as bipolar processes. This significantly reduces the magnetically coupled losses and thereby enables two to three times higher quality factors than those of inductors fabricated on low resistivity EPI substrates. This means that quality factors in the order of 10 - 15 can be achieved. This is a great improvement but the parameter remains critical. Therefore the on-chip inductor modelling and quality factor optimization is still very important. It could be argued that on-chip inductor modelling is now so complicated that it is not a circuit design task anymore but instead a device physics task. This may be true but it does not change the fact that it is a very important issue.

Another circuit block which is important for the RF performance of integrated RF filters is the loss compensation circuit. It is important that more work is done in this field to better understand the trade-offs between linearity, noise and power dissipation. It could for instance be interesting to consider what the optimal MOS transistor parameters are because technology scaling gives better linearity due to mobility degradation but also more noise due to a higher needed zero-bias drain conductance.

When switched frequency tuning is used it is only temperature variations and bias level variations which can cause a change in the resonance frequency. Because on-chip inductors and vertical mesh capacitors have low temperature gradients it is reasonable to assume that on-chip filters can be made with very low frequency drift. It is also reasonable to assume that the overall frequency temperature gradient does not change much from one wafer to another. This means that with some temperature compensation implemented it is possible that integrated RF filters could be implemented with open loop tuning only. For instance a lookup table and perhaps a one-time calibration could be sufficient for the frequency tuning of an on-chip filter. This would greatly simplify the tuning issues and lower the production costs. The whole concept of digital frequency tuning and quality factor tuning is ideal for digital signal processor control. This combination offers a number of benefits when the tuning is implemented in software. It would for instance be easy to support several wireless standards because the DSP could simply run different adaptive algorithms. Therefore research in adaptive DSP control of digitally tuned RF filters would be relevant as it goes well in line with the future requirement of multi-mode operation.

Finally it could be interesting to make a study of the optimal number of stages in a cascaded filter taking noise, linearity, power dissipation and die area into account.

Appendix A: Publications

Papers Published as First Author

(1) K.T. Christensen and A. Jørgensen, "Easy Simulation and Design of On-Chip Inductors in Standard CMOS Processes", *in Proceedings of the 1998 IEEE International Symposium on Circuit and Systems*, Vol. 4, pp. 360-364.

(2) K.T. Christensen, "Design and Optimization of CMOS Switches for Switched Tuning of LC Resonators," *in Proceedings of the 2001 IEEE International Symposium on Circuit and Systems*, Vol. 4, pp. 770-773, May 2001.

(3) K.T. Christensen, “LC Quadrature Generation in Integrated Circuits”, in *Proceedings of IEEE International Symposium on Circuits and Systems*, vol. 1, pp. 41-44, May 2001.

(4) K.T. Christensen, "Design and Characterization of Vertical Mesh Capacitors in Standard CMOS," *in Digest of the 2001 IEEE VLSI Circuits Symposium*, pp. 201-204, June 2001.

Papers Published as Second Author

(5) J. Pihl, K.T. Christensen and E. Bruun, "Switching CMOS Mixer for Direct Downconversion", *In Proceedings of the 2000 NorChip Conference*, pp. 282-287, November 2000.

(6) J. Pihl, K.T. Christensen and E. Bruun, "Direct Downconversion With Switching CMOS Mixer", *In Proceedings of 2001 IEEE International Symposium on Circuits and Systems*, Vol. 1, pp. 117-120 May 2001.

Papers Submitted

(7) R.G. Jensen, K.T. Christensen and E. Bruun, "Programmable Filter for Multi-Mode Phones", *submitted for the 2001 NorChip Conference*.

Patents Applied For

(8) K.T. Christensen, "Integrated Phase Shifting Using Coils", *Nokia Mobile Phones Invention Report NC29435*, April 1999.

(9) K.T. Christensen, "Frequency Tuning Scheme for Bandpass LC Filters", *Nokia Mobile Phones Invention Report NC33334,* March1999.

(10) K.T. Christensen, "Switched Tuning of Differential LC Resonators", *Nokia Mobile Phones Invention Report NC29580,* December 2000.

(11) K.T. Christensen, "Integrated Circuit Transformer", *Nokia Mobile Phones Invention Report NC 33377,* August 2001.

Reports Written

(12) K.T. Christensen, "CMOS RF Receiver Perspectives and Research Ideas", *Nokia Mobile Phones Internal Report*, July 1998.

(13) K.T. Christensen, "RLC Polyphase Filters in Silicon Integrated Circuit Technology", *Nokia Mobile Phones Internal Report*, June 1999.

(14) K.T. Christensen, "Design and Optimization of RF CMOS Switches for LC Resonators", *Nokia Mobile Phones Internal Report*, August 2000.

(15) K.T. Christensen, "Knowledge Creation and Innovation in a Consumer Electronics R&D Centre", *Business Report for the Danish Academy of Technical Sciences (ATV)*, January 2001.

Appendix B: Contributions

Table 9: Contributions to the field of RF filtering in CMOS Transceivers

Challenge	Contribution	Potential Benefit	Described
Achieve sufficient LNA linearity to process unfiltered signals from the antenna.	Input Power Constrained LNA Design	Elimination of passive front-end filter	Chapter 2
Design of High-Q capacitors in standard CMOS	Vertical Mesh Capacitor	High-Q resonator, low noise, low power.	Section 3.2
Design of accurate capacitors in standard CMOS	Vertical Mesh Capacitor	Accurate frequency, wide tuning range, low noise.	Section 3.2
Accurate closed form capacitance modelling	Virtual Capacitance Area Estimation	Short development time	Section 3.2
High-Q inductor design in Standard CMOS on low resistivity substrates	Optimized cross-over design	High-Q resonator, low noise, low power.	Section 3.3
Good Balance in Resonators	Optimized cross-over design	Low DC offset in direct conversion	Section 3.3
Wideband impedance transformation with good RF performance	Symmetrical Transformer Layout	Robust design, higher gain, lower noise	Section 3.3
Center Frequency Tuning	DSP Tuning With Oscillating Filter	Elimination of front-end filter	Section 4.1.2
Linear High-Q Resonator Tuning	Optimal MOS Switch Size Identification	high selectivity, low noise, good linearity	Section 4.2
Low Loss Switched RF Tuning	Enhanced Differential Waffle Switch	high selectivity, low noise, good linearity	Section 4.4
High Dynamic Range Loss Compensation	Switched Bias Binary Weighted Loss Comp.	Elimination of front-end filter	Section 5.4
Integrated Front-End Filter Design	Integrated Front-End Filter Design Method	Elimination of front-end filter	Chapter 6
Low Noise, Low Power RF Quadrature Generation for Enhanced Image Rejection and Complex Modulation	Low Power LC Quadrature Generation	Elimination of IF filters, low noise, low power.	Chapter 7