UNIVERSITÉ CATHOLIQUE DE LOUVAIN
LABORATOIRE D'HYPERFRÉQUENCES
Louvain-la-Neuve

Nonlinear Devices Characterization and Micromachining Techniques for RF Integrated Circuits

Bertrand Parvais

Jury

Prof. A. Cerdeira
Prof. D. Flandre
Prof. J.-D. Legat (président)
Prof. J.-P. Raskin (promoteur)
Prof. D. Schreurs
Prof. D.Vanhoenacker-Janvier
Prof. P. Wambacq

Thèse présentée en vue de l'obtention du grade de docteur en Sciences Appliquées

December 2004

Acknowledgments

First I would like to thank my advisor, Prof. Jean-Pierre Raskin, for giving me the opportunity to work on such an exciting subject, and for his help throughout this work. I would also like to thank Prof. Denis Flandre and Prof. Danielle Vanhoenacker for their guidance as member of my training staff. Also, I would like to thank Prof. Dominique Schreurs and Prof. Antonio Cerdeira not only for being part of the jury of this thesis, but also for the collaboration we had. Finally I would like to thank Prof. Piet Wambacq for acting as an outstanding reader of this thesis and Prof. Jean-Didier Legat for chairing the examining board.

I am grateful to the whole technical staff of the microwave and microelectronics laboratories. My special thanks go to Pascal Simon for his enthusiasm and the quality of his measurements and to the clean room staff, among them André Crahay, Christian Renaux and David Spôte for providing me their experience in microelectronics processing. I would also like to thank Brigitte Dupont who manages brilliantly the computer park.

I must thank all my colleagues that I have interacted with in the past few years. They have made my experience rich and satisfying. Among them, I reserve a special thank to Morin Dehan, Antoine Pallandre, Jean Laconte, Rémy Charavel, François Iker, Laurent Vancaillie, and Alexandre Siligaris. Furthermore, I thank the staff of CISSOID company, Pierre Delatte, Laurent Demeûs, Vincent Dessart and Gonzalo Picun, for their guidance and interesting collaboration.

Thank you finally to all those who, at UCL and in my professional but also personal circle, encouraged me during these four years.

TABLE OF CONTENTS

Abstract

The huge demand for mobile communications leads to an increasing request of low-cost and low-power mixed-mode integrated circuits. Both the recent silicon technology improvements and the evolution of design techniques contributed to the development of systems-on-chip (SoC) that combines digital and high-speed communication circuits, providing new power-saving design opportunities.

The development of SoC involves multi-disciplinary skills at several levels, from system design to semiconductor physics. The present work focuses on three of these levels: technology, device, and circuit. The highlight is put on the design of radio-frequency (RF) integrated voltage-controlled oscillators (VCOs).

Because of its high-frequency and high-isolation properties, and of its low-power capabilities, Silicon-on-Insulator (SOI) CMOS is a promising technology for SoC and is used in the frame of this work. In telecommunication applications, distortion is responsible for the generation of spurious frequency bands. An accurate nonlinear characterization of SOI devices is therefore required. Most of the existing tools are today based on a low-frequency analysis. We discuss the frequency validity range of these methods and provide new simple techniques for the nonlinear characterization of systems at both low- and high-frequencies. Our analysis is based on the integral function method and on the Volterra series. These methods are applied to analyze the linearity behavior of different SOI transistors.

On the other hand, high performance RF circuits also rely on the quality of the passive components, such as inductors, capacitors, and transmission lines. New three-dimensional high quality passive devices can be realized in silicon thanks to micromachining techniques. The fabrication of micromachined capacitors is investigated in this work. Original dry etching techniques are proposed for the patterning of sub-micron devices. We show that cobalt silicided polysilicon may be used as a low-stress, highly conductive mechanical material. Being a critical process step, the release of micromachined devices requires a specific attention to avoid the stiction of the movable structure. We show that the coating of silanes in both liquid- and vapor-phases reduces the stiction issue, highly improving the reliability of micromachined thin-film devices.

The use of high quality devices is however not sufficient to develop high performance circuits. The last part of this work is dedicated to the design of VCOs. The design trade-off between phase noise and power consumption is discussed. The describing function formalism is used to evaluate the oscillation amplitude and is embedded in a design methodology. Since an accurate frequency tuning is another desired feature, SOI accumulation-mode varactors are analyzed in both small- and large-signal regimes. It is demonstrated that the intrinsic doping of the varactor provide high quality factor and symmetrical tuning curve, which are desirable features for their integration in VCOs.

Scientific Publications

Articles in periodicals

[1] B. Parvais, A. Cerdeira, D. Schreurs and J.-P. Raskin, "Nonlinear performance comparison for FD and PD SOI MOSFETs based on the Integral Function Method and Volterra modeling," *Wiley International Journal of Numerical Modeling*, to be published.

[2] B. Parvais, A. Pallandre, A.M. Jonas and J.-P. Raskin, "Liquid and Vapor Phase Silanes Coating for the Release of Thin Films MEMS," *IEEE Transactions on Device and Materials Reliability*, to be published.

[3] C. Craeye, B. Parvais, and X. Dardenne, "MoM simulation of Signal-to-Noise patterns in infinite and finite receiving antenna arrays," *IEEE Transactions on Antennas and Propagation*, December 2004.

[4] B. Parvais and J.-P. Raskin, "Integral Function Method for the Analysis of Harmonic Distortion of SOI MOSFETs," *Revue HF*, no. 2, 2004.

[5] B. Parvais, G. Torrese, A. Cerdeira, D. Schreurs, J.-P. Raskin, "Comparison of Harmonic Distortion Characterization Techniques for SOI MOSFETs," *Revue HF*, no. 4, pp. 43-56, December 2002.

[6] E. Quévy, B. Parvais, J.P. Raskin, L. Buchaillot, D. Flandre, D. Collard, "A modified Bosch-type process for precise surface micro-machining of polysilicon," *Journal of Micromechanics and Microengineering*, vol. 12, pp. 328-333, May 2002.

[7] J.-P. Raskin, J. Laconte, A. Akheyar, C. Dupont, S. Adriaensen, A. Nève, M. Dehan, B. Parvais, D. Vanhoenacker, L. Demeûs, P. Delatte, V. Dessart, D. Flandre, "Fully-depleted SOI CMOS technology for heterogeneous micropower, high-temperature or RF Microsystems," *Revue HF*, no. 2, pp. 53-68, July 2001.

[8] D. Flandre, S. Adriaensen, A. Akheyar, A. Crahay, L. Demeûs, P. Delatte, V. Dessard, B. Iniguez, A. Nève, B. Katschmarskyj, P. Loumaye, J. Laconte, I. Martinez, G. Picun, E. Rauly, C. Renaux, D. Spôte, M. Zitout, M. Dehan, B. Parvais, P. Simon, D. Vanhoenacker and J.-P. Raskin, "Fully-depleted SOI CMOS technology for heterogeneous micropower, high-temperature or RF microsystems," *Solid-State Electronics*, vol. 45, no. 4, pp. 541-549, April 2001.

Conference Proceedings

[1] B. Parvais, J.-P. Raskin, A. Cerdeira, M. Estrada, "Application of Integral Function Method for Distortion Analysis of Microwave Transistors," *Asia Pacific Microwave Conference (APMC'04)*, New Delhi, India, December 15-18, 2004.

[2] B. Parvais, P. Delatte, H. Matsuhashi, F. Ichikawa, P. Simon, D. Schreurs, D. Flandre and J.-P. Raskin, "Small- and Large-Signal RF Characterization of Fully-Depleted Accumulation-mode Varactors for Low-Voltage LC-VCO SOI Design," *IEEE Int. SOI conference*, Charleston, South Carolina - USA, October 4 - 7, 2004.

[3] B. Parvais and J.-P. Raskin, "Analytical Expressions for Distortion of SOI MOSFETs Using the Volterra Series," *European Gallium Arsenide and other Compound Semiconductors Application Symposium 2004 (GAAS'04)*, Amsterdam, the Netherlands, October 11-15, 2004.

[4] B. Parvais, A. Cerdeira, D. Schreurs, J.-P. Raskin, "Harmonic Distortion Characterization of SOI MOSFETs," *European Gallium Arsenide and other Compound Semiconductors Application Symposium 2003 (GAAS'03)*, Munich, Germany, October 6-7, 2003.

[5] A. Cerdeira, M. A. Aleman, M. Estrada, D. Flandre, B. Parvais, G. Picun, "The Integral Function Method: A New Method to Determine the Non-linear Harmonic Distortion," *SBMicro 2003*, Sao Paulo, Brazil, September 8-11, 2003.

[6] B. Parvais, A. Siligaris, A. Cerdeira, J.-P. Raskin, "Oscillateurs micro-ondes en technologie CMOS/SOI: étude grand signal des dispositifs et performances," *Journées Nationales Micro-ondes (JNM'03)*, Lille, France, paper 5B-5, May 21-23, 2003.

[7] B. Parvais, A. Pallandre, A. M. Jonas, J.-P. Raskin, "A Fluoro-ethoxysilane-Based Stiction-Free Release Process for Submicron Gap MEMS," *Proc. Nanotech2003*, Vol. 1, pp. 522-525, San Francisco CA, USA, February 23-27, 2003.

[8] M. Dehan, B. Parvais, G. Dambrine, J.-P. Raskin, "Intérêts de la technologie CMOS SOI pour les applications micro-ondes faible tension faible Consommation," *3ème Journées Francophones d'Etudes Faible Tension Faible Consommation (FTFC'01)*, pp. 63-72, May 30-31 - June 1, 2001.

[9] B. Parvais, M. Goffioul, D. Vanhoenacker, and J.-P. Raskin, "Oscillateurs RF intégrés en technologie CMOS/SOI," *Journées Nationales Micro-ondes (JNM'01)*, pp. 63-72, May 30-31 - June 1, 2001.

Lectures and Workshops

[1] B. Parvais, A. Cerdeira, D. Schreurs and J.-P. Raskin, "Nonlinear performance comparison for FD and PD SOI MOSFETs based on the Integral Function Method and Volterra modeling," *MOS-AK/ESSCIRC Workshop*, Leuven, Belgium, September 20, 2004.

[2] B. Parvais and J.-P. Raskin, "Integral Function Method for the Analysis of Harmonic Distortion of SOI MOSFETs," *Proc. International Union of Radio Science (URSI)*, Brussels, Belgium, December 2003.

[3] B. Parvais, "Non-linearities in CMOS SOI technology: characterization, and impact on RF circuits," in *GSEC - AS10 -Integrated Circuits challenges in Broadband Telecommunication Systems: Process, modeling, design techniques and applications*, Louvain-la-Neuve, Belgium, September 22-24, 2003.

[4] B. Parvais, G. Torrese, A. Cerderia, D. Schreurs and J.-P. Raskin, "Harmonic distortion characterization techniques for SOI MOSFETs," *Proc. International Union of Radio Science (URSI)*, Brussels, Belgium, December 2002.

[5] B. Parvais and J.-P. Raskin, "Design of Integrated Voltage-Controlled-Oscillators in CMOS/SOI technology," *Proc. International Union of Radio Science (URSI)*, Louvain-la-Neuve, Belgium, December 2001.

[6] B. Iniguez, M. Goffioul, B. Parvais, J.-P. Raskin, "Deep-Submicron DC to RF SOI MOSFET Macro-Model for Designing Non-Linear RF Circuits," *Proc. 6th Workshop on Integrated Nonlinear Microwave and Millimeterwave Circuits (INMMC'01)*, Leuven, Belgium, September 20-21, 2001.

[7] B. Parvais, "Design of a microwave transceiver for wireless communication in SOI technology using new microfabrication techniques," *Proc. International Union of Radio Science (URSI)*, Brussels, Belgium, December 2000.

Glossary

List of Symbols

α	oscillator start-up gain [-]
α_T	thermal expansion coefficient [K^{-1}]
α_{FET}	coefficient that measures the efficiency of the coupling between the gate potential and the channel potential of a MOSFET at the Si-SiO_2 interface [-]
δ	induced gate noise factor [-]
ϵ	strain [-]
ε	permittivity [F/m]
γ	MOSFET channel noise factor referred to g_d at $V_{ds} = 0$ [-]
Γ	surface tension [N/m]
μ	mobility of the carriers [$m^2s^{-1}V^{-1}$]
ν	Poisson's ratio [-]
σ	stress [Pa]
θ	contact angle [^{o}]
ω_o	oscillation angular frequency [rad]
A	signal amplitude [V]
$\mathcal{A}$	anisotropy ratio [-]
A_S	surface area [m^2]
C_{ds}	MOSFET drain-to-source capacitance [F]
C_{gd}	MOSFET gate-to-drain capacitance [F]
C_{gs}	MOSFET gate-to-source capacitance [F]
d	distance between the plates of a MEMS capacitor [m]
E	Young's modulus [Pa]
F	oscillator noise factor [-]
$f_o = \frac{\omega_o}{2\pi}$	oscillation frequency [Hz]
f_{pHD_2}	pole of HD_2 [Hz]
f_{piHD_3} $(i = 1, 2)$	poles of HD_3 [Hz]
f_{zHD_2}	zero of HD_2 [Hz]
f_{ziHD_3} $(i = 1, 2)$	zeros of HD_3 [Hz]
g_d	MOSFET output conductance [S]
g_{d1}	MOSFET output conductance nonlinear coefficient of order 1 [S]
g_{d2}	MOSFET output conductance nonlinear coefficient

	of order 2 [S/V]
g_{d3}	MOSFET output conductance nonlinear coefficient of order 3 [S/V^2]
g_m	MOSFET gate transconductance [S]
g_{m1}	MOSFET gate transconductance nonlinear coefficient of order 1 [S]
g_{m2}	MOSFET gate transconductance nonlinear coefficient of order 2 [S/V]
g_{m3}	MOSFET gate transconductance nonlinear coefficient of order 3 [S/V^2]
$H_{nk}(s_1, s_2, \ldots, s_n)$	Volterra kernel of order n at circuit node k (in the Laplace domain)
HD_n	harmonic distortion of order n [dB]
I	moment of inertia [kg m^2]
I_B	oscillator bias current [A]
I_D	MOSFET drain current [A]
IMD_n	intermodulation distortion of order n [dB]
k	spring constant [N/m]
$k_B = 1.380658 \cdot 10^{-23}$	Boltzmann's constant [JK^{-1}]
k_v	varactor sensitivity [F/V]
$L(\Delta\omega)$	phase noise at $\Delta\omega$ of the carrier [dBc/Hz]
N	noise factor [-]
N_C	elasto-capillary number [-]
P	MOSFET channel noise factor referred to g_m [-]
P_d	dissipated power [W]
Q	quality factor [-]
q_i	ion charge [C]
q	capillary pressure [Pa]
R_g	MOSFET gate resistance [Ω]
T	temperature [K]
U	energy [J]
V_{dd}	supply voltage [V]
V_D	MOSFET drain voltage referred to the source [V]
V_G	MOSFET gate voltage referred to the source [V]
V_{GS}	control voltage of a MOSFET varactor [V]
VIP_{3h}	harmonic voltage intercept point of order 3 [V]
VIP_{3i}	intermodulation voltage intercept point of order 3 [V]

V_{PI}	pull-in voltage [V]
V_{th}	threshold voltage [V]
Y_L	load admittance [S]
$Z_L = Y_L^{-1}$	load impedance [Ω]

List of Abreviations

AC	alternative current
AM	amplitude modulation
BC	body-contacted
CAD	computer-aided design
CMOS	complementary metal-oxide semiconductor
CS	common-source
CVD	chemical vapor deposition
DC	direct current
DF	describing function
DR	dynamic range
DSP	digital signal processing
DTMOS	dynamic-threshold MOSFET
FD	fully depleted
FET	field effect transistor
FOM	factor of merit
GVO	gate voltage overdrive
HD	harmonic distortion
HF	high frequency
IC	integrated circuit
ICP	inductively-coupled plasma
IFM	integral function method
IMD	intermodulation distortion
ISF	impulse sensitive function
LF	low frequency
LNA	low noise amplifier
LO	local oscillator
LPCVD	low pressure chemical vapor deposition
LSNA	large signal network analyzer
LSTF	large signal transfer function
LTI	linear time invariant

LTV	linear time variant
MEMS	micro-electromechanical system
MOSFET	metal-oxide-semiconductor field-effect transistor
PA	power amplifier
PD	partially depleted
PECVD	plasma enhances chemical vapor deposition
PLL	phase-locked loop
PM	phase modulation
RF	radio frequency
RBW	resolution bandwidth
RIE	reactive ion etching
RMS	root-mean square
SAW	surface acoustic wave
SNR	signal-to-noise ratio
SoC	system-on-chip
SOI	silicon-on-insulator
THD	total harmonic distortion
TMAH	tetramethylammonium hydroxide
UCL	université catholoque de Louvain
VBW	video bandwidth
VCO	voltage-controlled oscillator
VIP	voltage intercept point
XRR	X-ray reflectometry

CHAPTER 1
INTRODUCTION: RADIO FREQUENCY INTEGRATED CIRCUITS

Wireless technology experienced a booming evolution since in 1901, Guglielmo Marconi successfully transmitted radio signals across the Atlantic Ocean. Today, few of us are surprised when voice, images, and data are transmitted from a cellphone to another one placed at the opposite corner of the planet. The domain of applications of wireless technologies increases from day to day: cellphones, global positioning systems, satellite communications, short-range data links such as home wireless networks or intelligent sports equipments, radio-frequency (RF) identification, wireless monitors, remote sensors and other uses in the industrial and automotive sectors, medical and implantable sensors, etc. In all these applications, the information is transmitted from a point to another by an electromagnetic wave at RF. In the view of the increasing number of applications, the sharing of the frequency spectrum becomes more difficult, resulting today in a dense occupation of the spectrum and the use of higher frequency bands (several GHz).

This explosive evolution has been possible thanks to the evolution of technologies, the emergence of digital communication systems, and the ability to deal with the ever growing complexity of circuits. A transmission of high quality, a very high degree of integration, a low power consumption and the use of a low supply voltage are the goals set for new developments in wireless transceiver design. Because transceivers have to be mass fabricated in order to satisfy the needs of the present market, low-cost is another concern. Therefore, as the submicron complementary-metal-oxide-semiconductor (CMOS) technology permits to achieve good RF performances and high level of integration, this technology has been pointed out for some ten years as a good candidate for low-cost RF integrated circuits (ICs).

Today's pocket phones contain more than one million transistors, with only a small fraction operating in the radio-frequency range and the rest performing low-frequency "baseband" analog and digital signal processing. The largest part of new wireless applications uses indeed digital data communication. Digital wireless has many advantages such as high quality communication due to signal regeneration and error correction systems, computer control of the communication, and integration

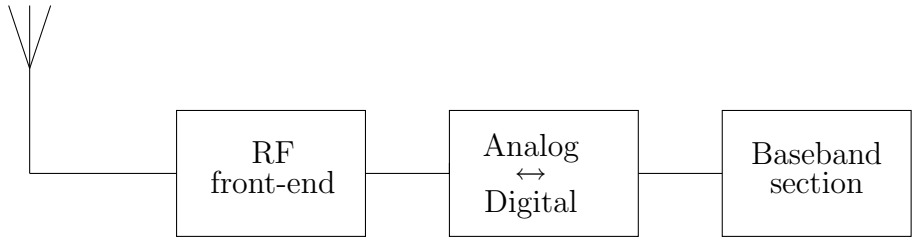


Figure 1.1: Principle of a transceiver.

of different services in one application. Because of the high frequency involved and because the signals we experience in the real life are analog, a part of the signal processing has however to be performed analog.

A typical transceiver is constituted of four main parts (Fig. 1.1): the antenna that picks up and sends data from and to the atmosphere, the RF front-end that converts the information between low- and high-frequencies (LF and HF, respectively), the analog-to-digital and digital-to-analog converters, and a baseband digital signal processing unit. Integrability and power consumption reduction of the digital part will further improve with the continued downscaling of technologies, but this is completely different for the analog front-end, for which integrability and power consumption are closely related to the physical limitations imposed by the transceiver topology and not by the technology. The present trend is thus to push more radio signal processing into baseband digital signal processors[1] (DSPs). Due to the complexity of present-day transceivers and the technological limitations, mobile systems are often constituted of a three or four chips-set combined with external components. A further reduction of the number of components is essential to obtain lower cost, power consumption and weight. For that reasons, a single-chip CMOS or more flexible multi-chip-modules (MCM) solutions are very attractive. The analog RF front-end is the only part of communication systems that will be investigated in the present work. We further restrict the analysis to integrated silicon technologies.

The RF section remains the bottleneck of the entire system. This is so for three reasons [1]. First, in contrast to other types of analog and mixed-signal circuits, RF systems demand a good understanding of many areas that are not directly related to integrated circuits. Indeed, skills in signal propagation, wireless standards, microwave and communication theories are needed in addition to the knowledge of transceiver architectures, computer-aided-design (CAD) tools, and IC technologies and design techniques. Second, RF circuit must process analog signals with a wide

[1]If analog becomes more and more digital, it is however worth noting that even in computer chips design, the treatment of digital signals relies more and more on analog techniques due to the very high frequencies involved.

dynamic range at high frequencies. The designer has thus to face the trade-offs between power, noise, high-frequency, gain, supply voltage and linearity. Third, computer-aided analysis and synthesis tools for RF ICs are still in their infancy, forcing the designer to rely on experience, intuition, or inefficient simulation techniques to predict the performances.

The present work is intended to push this difficulties a little bit further away. Due to the multi-disciplinary aspects of RF IC, three different levels are investigated:

- *Circuit* design techniques, filling the gap between by-hand design and numerical analysis using SPICE-like simulating tools. Oscillator circuits are detailed.

- *Technology* and processing techniques, as a determinant factor for the circuit performances. Highlight is put on recent CMOS silicon-on-insulator (SOI) technology and micromachining techniques.

- *Devices* analysis and characterization, for a better understanding of the trade-offs in analog RF design. Details on the devices linearity are given.

Before entering into the details of our work, let us describe in this introducing chapter some useful items on RF front-ends and on the technologies we used.

1.1 RF Front-ends

The RF front-end is the circuitry directly connected to the antenna. In the receive path, it has to deal with the very small signals coming from the antenna. These signals are amplified and down-converted to a given frequency, called the *intermediate frequency* (IF). In the transmit path, the role of the front-end is to up-convert the digital data to the antenna by providing enough amplification of the signals.

Fig. 1.2 is a general picture of the different constitutive elements of a front-end. In the receive path, after a band-selective filter, the low-noise amplifier (LNA) improves the sensitivity of the system. The signals are then translated down to IF in a mixer. The *mixer* is a circuit that functions as a mathematical multiplier of two input signals. In the present case, these two signals are the RF signal to be demodulated at frequency f_{RF} and a sine wave at frequency f_{LO} provided by a local oscillator (LO). The mixer output is then constituted of two tones, respectively at $f_{IF} = f_{RF} - f_{LO}$ and at $f_{RF} + f_{LO}$. A subsequent filter is needed to eliminate the high frequency component. In the transmit path, the signals are up-converted in the same manner and are later amplified in a power amplifier (PA).

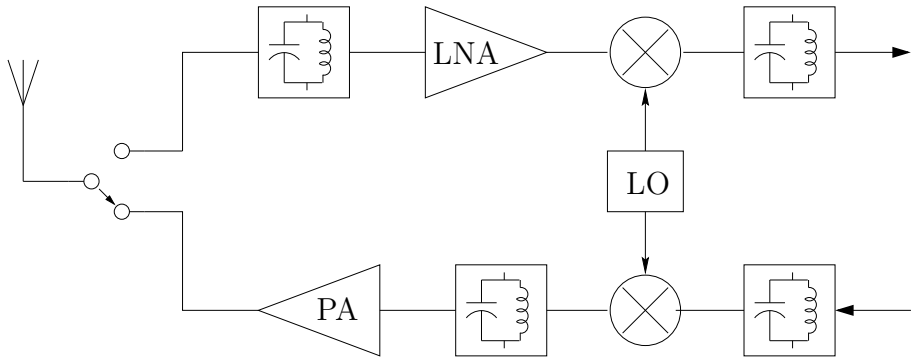


Figure 1.2: RF front-end principle.

Implementation options Several implementation styles of front-ends exist, depending on the IF value (f_{IF}). In *heterodyne* transceivers, f_{IF} is high and several conversion stages are sometimes needed. The relaxed specifications on the filters make this widely used implementation very attractive. The major drawback of heterodyne architectures is that the spectral component at $f_{LO} - f_{IF}$, called the *image* frequency, is also converted into the IF band of interest. To avoid this, costly filters are required. In the second option, the *homodyne* architecture, the problem of the image frequency is circumvented by the use of a zero-IF and a quadrature demodulation. It is worth noting that, as almost all DSPs perform the demodulation on a quadrature baseband signal, the production of quadrature signals at the output of the front-end is necessary in most cases. The disadvantage of the homodyne receiver is that extraneous offset voltages can corrupt the signal or saturate the following stages. The origin of these DC offsets is related to the mixing of signals at the same frequency, which are coming from the LO and RF paths, and from leakages at these frequencies inside the circuits. However, a high level of integration is possible with this architecture. The third implementation option, the *low-IF* topology, takes advantages of the two previous ones, at the price of a higher complexity. The IF is chosen low enough to avoid the frequency image problem, but it is not set to zero to avoid the presence of DC offsets.

Frequency synthesizer Let us now inspect one of the key-subsystem, the local oscillator. Its role is to synthesize a precise output frequency. The wireless applications may require multiple standards to be embedded in one transceiver. In that case, the synthesizer should provide several output frequencies. The tremendous growth of mobile communications placed stringent requirements on channel spacing and, as a consequence, on the spectral purity and stability of the LO. The exact frequency should be tracked digitally, precisely and quickly. In practice, the stabil-

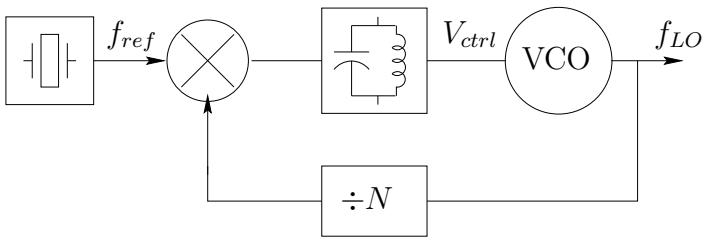


Figure 1.3: PLL principle.

ity achievable by electrical oscillators does not meet these needs. Crystal resonators allow on the other hand defining a frequency with high precision, but they are limited to low-frequency. A solution to obtain a stable HF oscillation is to embed the oscillator in a feedback loop, which is referenced to a crystal. A *phase-locked-loop* (PLL) is usually used for that purpose.

The principle of such a PLL is depicted in Fig. 1.3. The wanted signal is located at the output of the voltage-controlled oscillator (VCO). In VCOs, the output frequency f_{LO} changes as a function of a given input voltage (V_{ctrl}). The PLL provides this control voltage in such a way to ensure the stability of the output oscillation. A stable signal at f_{LO} is obtained through the phase comparison of this signal to a reference signal defined by a crystal resonator. The comparator may be implemented by a mixer, whose output voltage depends on the difference between the reference frequency f_{ref} and f_{LO}. Because f_{LO} is several times higher that f_{ref}, the frequency of the signal at the output of the VCO must be divided by a certain factor N in a frequency divider before entering in the comparator. The filtered output of the mixer is used as the control voltage of the VCO.

The bottleneck of PLLs is the design of the VCO. Indeed, even if the output wave of the VCO is stabilized in the loop, the performances of transceivers still depend highly on the purity of this signal. This is especially true in modern applications for which channels are closely spaced to each other. As it will be explained in chapter 3, the spectral purity may be improved at the price of power consumption. However, high power consumption is not desirable in mobile applications. To face those trade-offs and to improve the performances, we can act on two levels: technologies and design techniques. Technology should improve the quality of the devices. But the connection of high quality devices is not enough to make high performances circuits. Design techniques are needed. This two levels scheme asks implicitly for a third one: the behavior of the devices should be known with a good accuracy. Modeling and characterization techniques are thus of great importance.

CMOS for RFIC The CMOS technology is today widely used in RFICs. The evolution of high speed and low-power DSPs has been possible thanks to the downscaling of CMOS. High cut-off frequencies of several tens of GHz have in turn permitted the use of CMOS in RF applications. This evolution makes of CMOS a strong competitor to silicon bipolar and III-V based technologies in the RF field. It further presents a strong interest as its mass fabrication helps to reduce the production costs. The integration of both the RF and the digital parts in a single chip renders CMOS technology even more attractive.

RF building blocks and integrated transceivers have been demonstrated for about ten years [2]. The MOS technology is now mature for RF applications and (almost) fully integrated transceivers are available [3]. CMOS has been shown to be particularly adequate for low-costs, short-range and low-power communication systems in the low GHz range. The MOS field-effect transistors (FETs) have been studied deeply, both at low and high frequencies [4]. Deterioration of the performances due to short-channels effects has also been widely studied [4]. Noise behavior has been characterized [5, 6]. If the linear behavior of the MOSFETs is now well understood, some lacks persist in their nonlinear characterization at HF, as today, only few papers are dedicated to this subject [7, 8, 9]. A large part of this dissertation is thus dedicated to the distortion characterization of MOSFETs.

1.2 Technology

Technology fixes the limitation of the performances achievable for a given application. The quality of the devices is directly related to the technology in which they are implemented. As stated before, the high cut-off frequencies of submicron MOSFETs permits their use in today's RFICs. Because digital applications takes more than 90 % of the electronic market, the silicon analog circuits has to follow the trends of the digital world. The well-known Moore's law, that predicts a doubling of the performances every 18 months, has thus to be fulfilled. The downscaling of bulk MOSFETs has permitted to follow this law for many years, but the physical limitations of MOSFETs principally due to short-channels effects in deep-submicron technologies, prevent standard bulk MOS technology to follow this law. Silicon-on-Insulator (SOI) technology helps however to keep the Moore's law valid for some years [10]. This explains the present growing interest for that technology.

High performances RF circuits also require high quality passive components, such as inductors, capacitors, or transmission lines. Unfortunately, the fabrication of high quality passive elements is quite difficult in CMOS due to the relatively high

losses linked to the silicon substrate and the relatively thin metal layers used for interconnections. To face this, some solutions were proposed. On one hand, the conductor losses are reduced by thicker metallization and stacking of multiple metal layers. On the other hand, a decrease of the substrate losses is obtained by the use of high resistivity substrates, very thick isolating oxide layers between passive components and the substrate, or pattern ground shields that avoid the electric fields to penetrate into the substrate. Finally, high-quality three-dimensional devices may be fabricated by micromachining technologies, as detailed in the following.

The present work is intended to fabricate some building blocks of RF silicon integrated transceivers. As both micromachining and SOI technologies present some interests in that field, because of they allow to build CMOS compatible, high-quality RF components, they are introduced hereafter and used throughout our work.

1.2.1 Silicon-on-insulator technology

SOI technology has now been widely demonstrated and recognized to be a mature and viable alternative to mainstream bulk Si for the realization of high-speed, low-power digital and analog CMOS circuits, as well as niche applications under extreme high temperature or radiation operating conditions [11, 12].

Compared to bulk, the basic characteristic of SOI technology is the separation of the top active region from the underlying mechanical substrate by a thick insulator layer (Fig. 1.4). This difference leads to some advantages over bulk silicon [13]:

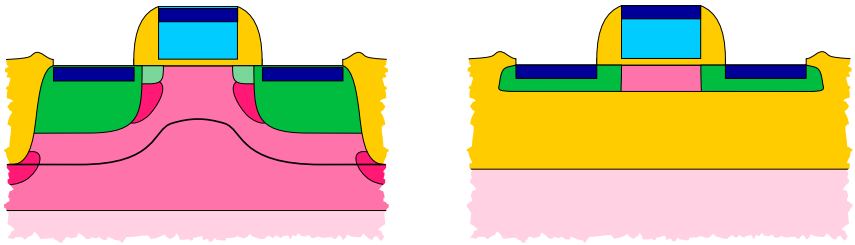
- *Reduction of the substrate parasitic capacitances.* The reduction of the total capacitance of a MOSFET will increase the cut-off frequency of the transistor, allowing to use circuits at higher frequency.

- *Small devices without latch-up.* The buried oxide prevents any current to flow between devices, eliminating the latch-up effect.

- *Reduction of short-channel effects.* In fully-depleted devices, the space charge in the thin film is well controlled by the gate, reducing the loss of gate control experienced in short-channel MOSFETs.

- *Improved subthreshold slope.* The lower inverse subthreshold slope, compared to bulk, allows better performances at low supply and the FET shows higher efficiency due to the lower dependence on the body potential.

- *Reduction of the substrate noise.* The insulator prevents from high crosstalk through the substrate in mixed-mode HF circuits [14].



(a) (b)

Figure 1.4: Cross-section of a (a) bulk and a (b) SOI MOSFET. The SOI device is separated from the mechanical substrate by an insulating layer.

Depending on the thickness of the top silicon layer, the silicon thin-film may be fully- or partially depleted (FD or PD, respectively). The different properties of the devices will be studied in chapter 2.

1.2.2 Micromachining techniques

The term *micromachining* refers to the micro-fabrication of mechanical structures using processes from the mechanical as well as from the microelectronics fields. Micromachining techniques are used to realize *microsystems*, also called *micro-electromechanical systems* (MEMS). As the name suggests, a MEMS is a real system at the micro-scale (i.e., typical length is 1 μm= 10^{-6}m) that has to perform a certain functionality, involving both electronics and moving parts. The microelectronic ICs can be thought of as the "brains" of a system and the mechanical part augments this decision-making capability with "eyes" and "arms", to allow microsystems to sense and control the environment. MEMS promises to revolutionize nearly every product category by bringing together silicon-based microelectronics with micromachining technology, making possible the realization of complete systems-on-a-chip (SoC).

The interest for MEMS is thus multiple. First, its high integration level permits to diminish the number of chips for a given application. This is advantageous not only for the lower volume and weight, but also for the decreased number of required interfaces between chips, provided that each interface implies a loss of energy. The production costs are moreover drastically reduced. Moreover, as the signal is processed on the same chip it is sensed, the speed and the quality of the processing is improved.

Even if micromachining techniques are almost as old as micro-electronics [15], it remained a lab curiosity for a long time until the interest of silicon as mechanical material was pointed out in the eighties [16]. The area of applications primary

interested by three-dimensional devices and MEMS was that of intelligent sensors and actuators. The development of MEMS for RF began in the early nineties. Interest for RF MEMS has not stopped to grow until now.

Indeed, the majority of heterodyning communication transceivers relies heavily upon the high quality factor Q of surface-acoustic-wave (SAW) resonators to achieve adequate frequency selection in their RF and IF filtering stages and to realize the required purity and stability in their local oscillator. In addition, discrete inductors and variable capacitors are used to properly tune and couple the front-end sense and power amplifiers. Because these elements are placed off-chip, they must interface with integrated electronics at the board level, consuming a sizable portion of the total subsystem area, and of the power consumption as they are usually referenced to a low impedance level (50 Ω). The potentiality of integration offered by micromachined RF devices should allow a drastic diminution of these drawbacks. The high Q of mechanical devices should furthermore improve the performances of the transceivers of tomorrow.

1.3 This work

As previously explained, SoCs that combine digital and high-speed communication circuits present new opportunities for power-saving designs. These opportunities have been possible through improvements in both silicon technologies and design techniques. SOI CMOS is a promising technology for SoCs for several reasons: transistor scaling leads to active power reduction, standard interconnect supports the high-quality passive devices essential to communications circuitry, the crosstalk is reduced, and high-speed analog circuits on SOI are state of the art in terms of both performance and power dissipation [12]. On the other hand, new three-dimensional high quality passive devices can be realized in silicon thanks to micromachining fabrication techniques. Lots of RF MEMS has been demonstrated; however, few of them are today in the market. The reason is that some reliability problems remain, due to the young age of the technology. Through this work, SOI and thin-film MEMS technologies are investigated to improve the performances of today transceivers. As we are not focused on a specific application, the generality of the analysis is preserved and applies to a large range of wireless systems. The technological pros and cons of will nevertheless be discussed in some specific cases.

The technology advances and circuits performances are only measurable if valuable characterization methods exist. Techniques for the characterization of the linear

behavior of devices have been developed for a long time. The interest for nonlinear RF characterization has however grown recently as a consequence of the increase in applications requiring wireless communications. Indeed, as distortion causes spurious in telecommunication systems, more restrictive specifications are today put on the linearity. Simple and valuable tools for the nonlinear characterization are therefore desirable.

The development of SoCs is a team work that requires more than ever the conjunction of skills from the different engineering branches. Several levels of abstraction are involved, from the system design to the semiconductor physics. For high performances designs, optimization should take place at each of these (inter-dependent) levels.

The analysis provided in this work is focused on three levels: circuit, device and technology. The oscillator circuit is given as an example. Performances of VCOs highly rely on the design techniques as well as on the technology. SOI technology is used for its HF and low-power abilities. On the other hand, micromachining technologies are expected to improve the phase noise. To face the reliability lacks of micromachining, new fabrication techniques are developed. On the other hand, because oscillators are intrinsically nonlinear machines, there is a strong interest to study the nonlinearity of the devices they use. Therefore, valuable characterization techniques must be developed, as already said. These methods should in turns lead to improve the circuit design methodologies.

Those items will be discussed in details in the next chapters. In **chapter 2**, some nonlinear characterization methods are introduced and developed. We highlight the integral function method and the Volterra series. The methods are used to analyze the linearity of FD and PD SOI devices. The evolution of the distortion as a function of bias and frequency is investigated. We explain how simple tools are efficient in most of application in the low GHz range. The linearity of a narrowband LNA is also investigated.

The oscillator circuit is studied in **chapter 3**. The limitations of linear models are discussed. To face the complexity of nonlinear models, the output signal amplitude is evaluated by quasi-linear approximators. The design techniques and the trade-offs in the RF analog circuit design are detailed. Small and large signal characterization of tunable capacitors for VCOs in SOI technology are provided.

Tunable capacitors in MEMS technology are also investigated in a **fourth chapter**. The design of these devices is introduced and the fabrication techniques are widely discussed. Specific dry etching techniques are developed to pattern silicon

with high precision. The silicidation of polysilicon is investigated as a mechanical material. The residual stress is characterized. The silanization process is investigated to decrease the probability of the stiction failure.

References

[1] Behzad Razavi, *RF Microelectronics*, Prentice Hall, Upper Saddle River, NJ 07458, 1998.

[2] Thomas H. Lee, *Design of CMOS Radio Frequency Integrated Circuits*, Prentice hall, 1994.

[3] I. Vassiliou, K. Vavelidis, T. Georgantas, S. Plevridis, N. Haralabidis, G. Kamoulakos, C. Kapnistis, S. Kavadias, Y. Kokolakis, P. Merakos, J.C. Rudell, A. Yamanaka, S. Bouras, and I. Bouras, "A single-chip digitally calibrated $5.15-5.825$-GHz 0.18 μm CMOS transceiver for 802.11a wireless LAN," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 12, pp. 2221 – 2231, December 2003.

[4] Y.P. Tsividis, *Operation and Modeling of the MOS Transistor*, McGraw-Hill, second edition, 1999.

[5] E.P. Vandamme and L.K.J. Vandamme, "Critical discussion on unified $1/f$ noise models for MOSFETs," vol. 47, no. 11, pp. 2146 – 2152, November 2000.

[6] Yuhua Cheng, Chih-Hung Chen, M. Matloubian, and M.J. Deen, "High-frequency small signal AC and noise modeling of MOSFETs for RF IC design," vol. 49, no. 3, pp. 400 – 408, March 2002.

[7] R. van Langevelde, L.F. Tiemeijer, R.J. Havens, M.J. Knitel, R.F.M. Ores, P.H. Woerlee, and D.B.M. Klaassen, "RF-distortion in deep-submicron CMOS technologies," *IEEE International Electron Device Meeting*, pp. 807–810, 2000.

[8] Sanghoon Kang, Byounggi Choi, and Bumman Kim, "Linearity analysis of CMOS for RF application," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 3, pp. 972–977, March 2003.

[9] Piet Wambacq and Willy Sansen, *Distortion analysis of analog integrated circuits*, Kluwer, Dordrecht, 1998.

[10] Leland Chang, Yang-Kyu Choi, J. Kedzierski, N. Lindert, Peiqi Xuan, J. Bokor, Chenming Hu, and Tsu-Jae King, "Moore's law lives on [CMOS transistors]," *IEEE Circuits and Devices Magazine*, vol. 19, no. 1, pp. 35 – 42, January 2003.

[11] D. Flandre, S. Adriaensen, A. Akheyar, A. Crahay, L. Demeûs, P. Delatte, V. Dessard, B. Iniguez, A. Nève, B. Katschmarskyj, P. Loumaye, J. Laconte, I. Martinez, G. Picun, E. Rauly, C. Renaux, D. Spôte, M. Zitout, M. Dehan, B. Parvais, P. Simon, D Vanhoenacker, and J.-P. Raskin, "Fully-depleted SOI

CMOS technology for heterogeneous micropower, high-temperature or RF microsystems," *Solid-State Electronics*, vol. 45, pp. 541–549, February 2001.

[12] J.-O. Plouchart, N. Zamdmer, J. Kim, M. Sherony, Y. Tan, A. Ray, M. Talbi, L. F. Wagner, K. Wu, N.E. Lustig, S. Narasimha, P. O'Neil, N. Phan, M. Rohn, J. Strom, D. M. Friend, S. V. Kosonocky, D. R. Knebel, S. Kim, K. A. Jenkins, and M. M. Rivier, "Application of an SOI 0.12 μm CMOS technology to SoCs with low-power and high-frequency circuits," *IBM Journal of Research and Development*, vol. 47, no. 5/6, pp. 611 – 629, September/November 2003.

[13] Jean-Pierre Colinge, *Silicon-on-insulator technology: Materials To VLSI*, Kluwer Academic Publisher, 1991.

[14] J.-P. Raskin, A. Viviani, D. Flandre, and J.-P. Colinge, "Substrate crosstalk reduction using SOI technology," vol. 44, no. 12, pp. 2252–2261, December 1997.

[15] H.C. Nathanson, W.E. Newell, R.A. Wickstrom, and J.R. Davis, "The resonant gate transistor," vol. ED-14, no. 3, pp. 117–133, March 1967.

[16] K.E. Petersen, "Silicon as a mechanical material," *Proceedings of the IEEE*, vol. 70, no. 5, pp. 420–457, May 1982.

CHAPTER 2
NONLINEAR CHARACTERIZATION OF SOI MOSFETS

2.1 Introduction

The development of submicron CMOS devices and the ever increasing performance required for electronic applications put the focus on the nonlinear analysis. In short-range telecommunication applications, where CMOS is considered as the appropriate low-cost technology, distortion is responsible for the generation of spurious frequency bands (Fig. 2.1). In the presence of large blocking signals, the linearity requirements can become the limiting specification. Indeed, the generation of intermodulation tones pollutes the signal down-conversion in receivers (Fig. 2.1): if the signal at angular frequency ω_1 is the wanted signal and the other one an interferer in an adjacent channel, the product of intermodulation may fall in the band of interest and disturb the transceiver working. The same phenomena occurs when two strong interferers are close to the wanted signal (Fig. 2.2). This chapter examines the linearity of silicon-on-insulator (SOI) MOSFETs and investigates its effect in a Low-Noise Amplifier (LNA).

The origin of nonlinearities is explained by the semiconductor physics. The DC drain current exhibits a highly nonlinear characteristic as a function of the applied voltages [1]. In new technologies dedicated to high-frequency (HF), the dimensions shrink and other linearity degradations linked to short channel effects and parasitic appear [2]. So carrier velocity saturation, channel length modulation, series resistances and mobility degradation have to be taken into account for an accurate large signal modeling of the MOSFET [3]. At high frequency of operation, the transistor behavior still depends on the DC current characteristics, but in addition, the reactances affect the behavior of the device. Nevertheless, previous experimental work showed that the distortion is dominated by the current-voltages ($I-V$) characteristics up to radio frequencies (RF) [4]. The DC measurement of the $I-V$ characteristics is then sufficient to evaluate the distortion of a device or system up to a certain frequency. However, this frequency has not yet been determined by

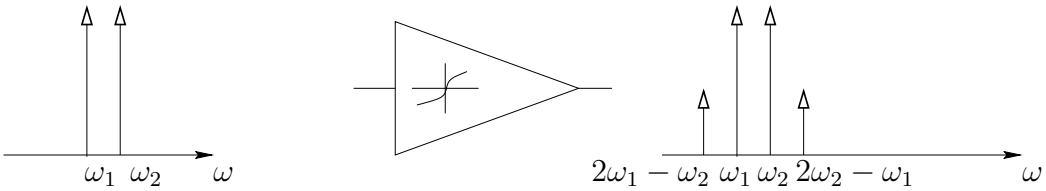


Figure 2.1: Intermodulation in a nonlinear amplifier.

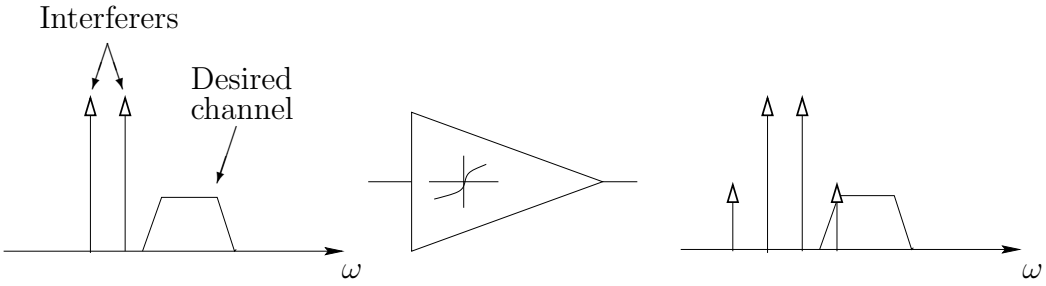


Figure 2.2: Two interferers corrupting the output of a nonlinear system.

analytical tools. Furthermore, the differentiation of the measurements required for the distortion analysis may become intractable because of the large amount of measurement noise. Recently, an integral function method (IFM) has been proposed to avoid this feature [5, 6]. Unfortunately, this method is restricted to low frequency.

If the measurement of the DC $I-V$ characteristics is today common, performing nonlinear measurements at microwave is a complex task. Instrumentation for nonlinear systems must allow simultaneous multiple tone measurements and requires a calibration of all separate incident and reflected waves instead of single frequency measurements and a relative calibration for linear systems [7]. Specific instrumentation setups and experiment design are required. Measurements results involving spectrum analyzer, sampling oscilloscopes and vectorial network analyzers are found in the literature. The recent development of large-signal network analyzers (LSNA) permits an accurate characterization of devices and systems at microwaves [8, 9], at the price of an expensive and non widespread setup.

On the other hand, the extraction of the main figures of merit from the measurements is not necessarily straightforward, as it relies on both the setup specifications and the hypotheses that are later made for the extraction. These hypotheses define the validity range of the considered method. The set of assumptions generally depends on the measurements conditions. Let us consider the case of low cost DC $I-V$ measurements. Despite the drain current of a MOSFET depends on both the gate and the drain voltages, it is widely assumed for the extraction of linearity factors of merit that only the gate voltage affects the linearity of the device [3]. More accurate characterization at LF includes the dependence on the output conductance and impedance values.

Analytical formulation of the distortion presents on the other hand an interest to understand the transistor nonlinear mechanisms, and to the circuit designers to deal with the different trade-offs involved in their design. At low frequency, a Taylor series analysis is generally used for that purpose, while the Volterra series are used at high frequency when inductors and capacitors are present [2, 10]. In these cases, the nonlinear elements are described in terms of the Taylor series expansion of their current-voltage or charge-voltage characteristics. This limits the validity range of these techniques, as the derivatives of the characteristic around any bias point must remain constant over the AC voltage and current deviation from that bias point. To ensure the validity of the Taylor series, the nonlinearity must be weak enough and the excitation signals small enough. We speak about *weak* nonlinear systems. We will use a more restrictive definition of a weak nonlinearity [2]: A circuit behaves weakly nonlinearly if, for the applied input signal, it can be accurately described by the first three terms of its converging Volterra series. This corresponds roughly to the range of input power for which there is no compression. For the MOSFETs used later in this chapter, this means that the input voltage amplitude should stay below $0.3 \ldots 0.4$ V depending on the bias conditions, to ensure the weak nonlinear behavior of the transistor.

Different techniques are investigated in this work for the RF characterization of the nonlinearity of SOI MOSFETs. Both weak and large excitations are considered. The aim of our analysis is to give the simplest nonlinear characterization technique that ensures good accuracy and is that adapted to the application under consideration. The validity range of the techniques are discussed and if possible enlarged.

The linearity analysis presented here includes the influence of drain and gate voltages, as well as the load impedance and the signal amplitude. We propose a method to apply IFM at high frequency. Analytical tools, based on the Volterra series formalism, are used to determine the frequency limitation of the DC characterization techniques and LSNA measurements are used to validate the analysis.

This chapter is organized as follows. Section 2.2 gives the theory required to analyze nonlinear systems. Simple analytical formulas of the MOSFET distortion are determined with help of the Volterra series. In section 2.3, the linearity of an SOI nMOSFET is detailed and the different characterization techniques compared. Since SOI technology provides several kinds of devices, a fair comparison of their nonlinearity is provided in section 2.4. In the last section, the analysis of different transistor configurations is given in order to characterize the linearity of an LNA.

2.2 Theory

The usual way for dealing with nonlinearities is to construct a linear model that approximates the nonlinearity. The linear term of a Taylor series expansion around a fixed operating point is commonly used for weak nonlinear systems. We will introduce in section 2.2.3 another kind of linearization that is more adequate in the case of *strong* nonlinearities. Due to their simplicity, linearized models are not satisfying to describe properly some properties of real systems, such as harmonic and intermodulation distortion.

We call *distortion* of a system any difference between the shape of the output waveform versus input waveform, except for a scaling factor. We classify the kind of distortion following the kind of the transfer characteristic. *Linear distortion* is caused by the excitation of a linear circuit by a signal with a non constant amplitude or phase characteristic as a function of the frequency. For instance, a square wave is distorted at the output of a linear low-pass filter. On the other hand, *nonlinear distortion* is caused by a nonlinear transfer characteristic, for instance the $I-V$ characteristics of a MOSFET. The following of this text deals with nonlinear distortion only.

The waveform at the output of a nonlinear system may be studied either in the time or in the frequency domain. Nevertheless, electrical engineers prefer to analyze the spectral components in the frequency domain. The Fourier transform is used for the conversion between one domain to another. It is the more general way to extract nonlinearities from measurements. However, since this technique gives poor insight of the nonlinear mechanisms involved, and because the Fourier transform is not so easy to implement numerically, other characterization techniques are required.

Among them, it was mentioned in the introduction of this chapter that a Taylor series expansion of the $I-V$ characteristic may be used for low-frequency, weak nonlinear systems. For a strong nonlinearity, the number of terms considered in the expansion should be as high that it renders the analysis practically intractable. In this case, the large-signal transfer function of the system must be calculated and the IFM is more appropriate because it depends directly on the signal amplitude. The IFM was initially restricted to low frequency characterization, but we provide in section 2.2.4 the extension of this method to HF. However, IFM is to date only able to calculate intermodulation distortion when it depends on harmonic distortion and when the influence of capacitance and/or inductances can be neglected [6]. Weak nonlinear systems may on the other hand be analytically studied at HF by the

Table 2.1: Range of application of the characterization techniques used in this chapter.

	Weak Nonlinearity	Strong Nonlinearity
Low Frequency	Taylor	IFM
High Frequency	Volterra	IFM extended

Volterra series approach [11, 2]. Table 2.1 summarizes the domain of validity of the different techniques used in the following of this work.

2.2.1 Basic concepts and Taylor series analysis

The definition of the main distortion factors will be introduced in this section with the help of the Taylor expansion of a nonlinearity. Both single- and multi-tone(s) excitation cases are discussed.

2.2.1.1 Single tone analysis

Consider a circuit excited with one sinusoidal source of amplitude A at angular frequency ω. When the amplitude A of the input signal is small enough, the output spectrum of the circuit only contains one frequency component above the noise floor, namely the response corresponding to the circuit linear behavior. This signal is at the same frequency as the input signal, called the *fundamental frequency*. The amplitude of this signal changes proportionally with the input amplitude.

When the amplitude is increased, the output spectrum contains signals at the angular frequencies 2ω and 3ω. These signals, called the *second* and *third harmonics*, originate from the second- and third-order nonlinear circuit behavior, respectively. Higher harmonics are found with increasing amplitude A. We define the *harmonic distortion of order n* (HD_n) as the ratio of the n^{th} harmonic Y_n to the fundamental response Y_1:

$$HD_n = \frac{|Y_n^2|}{|Y_1^2|}. \tag{2.1}$$

The *Total Harmonic Distortion* (THD) is defined as the sum of the Root Mean Square (RMS) values of the higher harmonics, relative to the fundamental component:

$$THD = \sqrt{\frac{\sum_{n=2}^{\infty} Y_n^2}{Y_1^2}}. \tag{2.2}$$

Consider the following relationship of a memoryless circuit between the input

signal $x(t)$ and the output signal $y(t)$:

$$y(t) = f(x(t)) = K_0 + K_1 x(t) + K_2 x^2(t) + K_3 x^3(t) + \dots \tag{2.3}$$

If we assume that the right-hand side of relation (2.3) converges to $f(x(t))$, the coefficients K_n can be identified with the coefficients of a Taylor series of $f(t)$:

$$K_n = \frac{1}{n!}\frac{d^n f}{dx^n}. \tag{2.4}$$

When the circuit is excited by a sinusoidal source $A\cos(\omega t + \phi)$, we find that

$$\begin{aligned} y(A\cos(\omega t+\phi)) &= \left(K_0 + \frac{K_2}{2}A^2\right) + A\left(K_1 + \frac{3K_3}{4}A^2\right)\cos(\omega t+\phi) \\ &\quad + \frac{K_2}{2}A^2\cos(2\omega t + 2\phi) + \frac{K_3}{4}A^3\cos(3\omega t + 3\phi) + \dots \end{aligned} \tag{2.5}$$

The harmonic distortion of order two or three are then calculated from relation (2.1):

$$HD_2 = \left|\frac{2K_2 A}{4K_1 + 3K_3A^2}\right| \tag{2.6}$$

$$= \frac{A}{2}\left|\frac{K_2}{K_1}\right| \qquad \text{if } K_1 \gg \frac{3}{4}K_3A^2, \tag{2.7}$$

$$HD_3 = \left|\frac{K_3A^2}{4K_1 + 3K_3A^2}\right| \tag{2.8}$$

$$= \frac{A^2}{4}\left|\frac{K_3}{K_1}\right| \qquad \text{if } K_1 \gg \frac{3}{4}K_3A^2. \tag{2.9}$$

It appears from (2.5) that the even order harmonics give rise to a DC shift, i.e., a variation of the component at 0 Hz. The bias of a nonlinear circuit may then change as a function of the signal amplitude. This property is called *self-biasing*.

Fig. 2.3 depicts the fundamental and the third harmonic levels at the output of such nonlinear circuit as a function of the input signal amplitude. At low input level, the amplitude of the n^{th} harmonic increases with the n^{th} power of the input amplitude. At high input amplitudes, this is not true anymore and the fundamental response may increase faster or slower than linear (*gain expansion* and *gain compression*, respectively). In (2.5) the gain $\left(K_1 + \frac{3K_3}{4}A^2\right)$ is modulated by A and the gain compression occurs if K_3 is negative. The 1-*dB* or 3-*dB compression point* ($P_{-1\text{dB}}$ and $P_{-3\text{dB}}$, respectively) is a measure of the gain compression, expressing the input level for which the fundamental response is 1- or 3- dB lower than the extrapolation of the response at lower amplitudes which is caused by linear behavior only.

The *input dynamic range* is the range of input amplitude over which the circuit

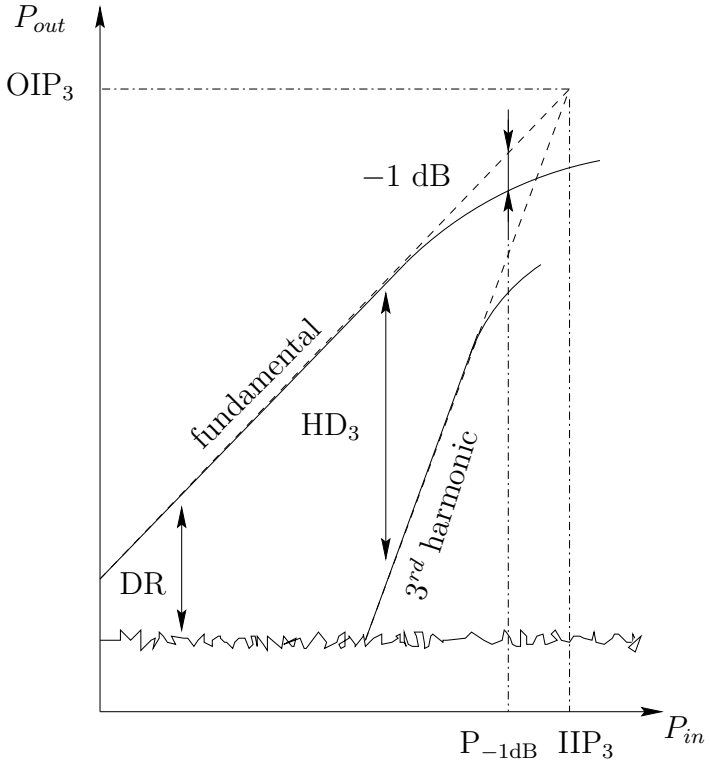


Figure 2.3: Output power at fundamental and third harmonic *vs* input power on a logarithmic scale.

can be used without a too large signal degradation. It is given by the difference between the input compression point and the noise floor.

At low input amplitude, the second harmonic increases as the square of the input amplitude. Its extrapolation must therefore cross the extrapolation of the linear gain. The intersection of those extrapolated curves is defined as the *second order intercept point* IP_{2h}. The IP_{2h} usually refers to the input level and it is sometimes denoted IIP_{2h} while the output level corresponding to the IP_{2h} is written OIP_{2h}. When the IIP_{2h} is given in Volts, we speak about *Voltage Intercept Point of order* 2 (VIP_{2h}). The *third order intercept point* IP_{3h} is defined similarly. From the previous example (eq. (2.3)), assuming a low level of distortion i.e., that $K_1 \gg \frac{3}{4}K_3A^2$, we have

$$VIP_{2h} = 2\left|\frac{K_1}{K_2}\right|, \tag{2.10}$$

$$VIP_{3h} = 2\sqrt{\left|\frac{K_1}{K_3}\right|}. \tag{2.11}$$

2.2.1.2 Multi-tones analysis: intermodulation

In the RF and microwave field of applications, harmonic distortion of transceivers should be as low as possible, but the main nonlinear problems come from the intermodulations of an interferer with the wanted signal (Fig. 2.2).

Consider that we have the signal $x(t) = A_1 \cos \omega_1 t + A_2 \cos \omega_2 t$ at the input of the system (2.3). Suppose that the signal at angular frequency ω_1 corresponds to the wanted signal, the other one being the interferer. At the output of the system, we have

$$\begin{aligned} y(t) &= \left(K_1 A_1 + \frac{3}{4} K_3 A_1^3 + \frac{3}{2} K_3 A_1 A_2^2 \right) \cos \omega_1 t \\ &\quad + \left(K_1 A_2 + \frac{3}{4} K_3 A_2^3 + \frac{3}{2} K_3 A_2 A_1^2 \right) \cos \omega_2 t + \dots \end{aligned} \tag{2.12}$$

The amplitude of the interferer may be much higher than the wanted signal ($A_2 \gg A_1$). In this case, the gain κ is given by

$$\kappa = \left(K_1 + \frac{3}{2} K_3 A_2^2 \right). \tag{2.13}$$

If the system saturates, $K_3 < 0$ and the gain decreases with A_2. We say that the signal is *blocked* when A_2 is so important that the gain becomes equal to zero. Furthermore, the variations of A_2 modulate the output signal (see (2.13)), we speak about *cross modulation*.

When two (or more) signals are applied, the output signal contains the fundamental and the harmonics, but also signals at others frequencies, resulting from the cross-products which are usually unwanted. The distortion introduced by these terms is called the *intermodulation distortion* (IMD). In communication circuits, these unwanted products are often denoted as *spurious responses*. For instance, at the output of the system (2.3) excited by $x(t) = A_1 \cos \omega_1 t + A_2 \cos \omega_2 t$ we find e.a. the following terms:

$$K_2 A_1 A_2 \left[\cos((\omega_1 + \omega_2)t) + \cos((\omega_1 - \omega_2)t) \right], \tag{2.14}$$

$$\frac{3 K_3 A_1^2 A_2}{4} \left[\cos((2\omega_1 + \omega_2)t) + \cos((2\omega_1 - \omega_2)t) \right], \tag{2.15}$$

$$\frac{3 K_3 A_1 A_2^2}{4} \left[\cos((\omega_1 + 2\omega_2)t) + \cos((\omega_1 - 2\omega_2)t) \right]. \tag{2.16}$$

The distortion products at $2\omega_1 + \omega_2$ and $\omega_1 + 2\omega_2$ will usually fall outside the frequency band of interest. However, signals at $2\omega_1 - \omega_2$ and $\omega_1 - 2\omega_2$ may fall in this band,

corrupting the transmission. The same mechanism occurs when a weak signal is close to two important interferers. In this case, one of the third-order intermodulation products falls in the band of interest and corrupts the wanted signal (Fig. 2.2).

When the input amplitudes are taken equal, the *n^{th}-order intermodulation distortion* IMD_n is defined as the ratio of the n^{th} intermodulation product to the fundamental response. So, IMD_2 and IMD_3 are given by

$$IMD_2 = \left|\frac{Y_{1,\pm1}}{Y_{1,0}}\right| \tag{2.17}$$

$$IMD_3 = \left|\frac{Y_{2,\pm1}}{Y_{1,0}}\right| = \left|\frac{Y_{1,\pm2}}{Y_{1,0}}\right| \tag{2.18}$$

where $Y_{n,m}$ is the response of the system at $n\omega_1 + m\omega_2$. The last equality in (2.18) is only valid at low frequencies. Indeed, sidebands asymmetries are observed at high frequency [12]. From the previous example, assuming low-distortion and identical amplitude for both signals ($A_1 = A_2 = A$), we have

$$IMD_2 = \left|\frac{K_2}{K_1}\right| A = 2 \cdot HD_2 \tag{2.19}$$

$$IMD_3 = \frac{3}{4}\left|\frac{K_3}{K_1}\right| A^2 = 3 \cdot HD_3 \tag{2.20}$$

As the output tone at $2\omega_1 - \omega_2$ (or identically, at $\omega_1 - 2\omega_2$) increases faster than at the fundamental, we may define a third-order intercept point IP_{3i} similarly as in the harmonic distortion case: it is the intersection between the extrapolated curves at ω_1 and at $2\omega_1 - \omega_2$ from low amplitudes. The *intercept point of order n^{th}* is defined as the input amplitude for which the extrapolated intermodulation distortion figures of order n are equal to one ($IMD_n = 1$). From equations (2.17) and (2.18), and under the same assumptions, we find

$$IP_{2i} = \left|\frac{K_1}{K_2}\right| = \frac{IP_{2h}}{2} \tag{2.21}$$

$$IP_{3i} = 2\sqrt{\left|\frac{K_1}{3K_3}\right|} = \frac{IP_{3h}}{\sqrt{3}} \tag{2.22}$$

where the index i refers to intermodulation intercept point and index h to harmonic intercept point.

2.2.2 Volterra series analysis

In the analysis of small signals FET amplifiers, such as LNAs, we are interested in the weakly nonlinear behavior of the circuits, because the excitations are weak signals. The Volterra series analysis is the most appropriate tool for that purpose. This approach provides analytic relations of the nonlinear behavior that include memory effects coming from inductances and capacitances.

2.2.2.1 Generalities

The most common implementation of Volterra analysis uses a technique called *the method of nonlinear currents*. In this method, each nonlinear circuit element is converted into a linear element in parallel with current sources that represent the element nonlinearities. The current at each order of nonlinearity depends on the element voltages at all lower orders, in such a way that the currents may be calculated recursively. The second order current is obtained from the first-order voltages. The circuit is then excited by the second-order current, resulting in the second-order voltage. The third-order current are then computed from the first- and second-order voltages, and so on.

The Volterra series describe the output of a nonlinear system as the sum of the responses of different operators, each operator representing a certain order of nonlinearity. Every operator is described either in the time domain or in the frequency domain by a *Volterra kernel*. The frequency domain representation is usually preferred for circuit analysis.

Any relationship between the input $x(t)$ and the output $y(t)$ of a nonlinear, time invariant system can be expressed with the following series [2]:

$$y(t) = \sum_{n=1}^{\infty} y_n(t) = \mathbf{H_1}[x(t)] + \mathbf{H_2}[x(t)] + \mathbf{H_3}[x(t)] + \dots \tag{2.23}$$

in which $\mathbf{H_n}$ is the n^{th}-order *Volterra operator* given by the following convolution:

$$\mathbf{H_n}[x(t)] = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} h_n(\tau_1, \tau_2, \dots, \tau_n) x(t-\tau_1) x(t-\tau_2) \dots x(t-\tau_n) d\tau_1 d\tau_2 \dots d\tau_n \tag{2.24}$$

where the functions $h_n(\tau_1, \tau_2, \dots, \tau_n)$, called the *Volterra kernels of the system*, satisfy the causality property. Frequency domain representation is obtained by the

multi-dimensional Laplace transform of the kernels:

$$H_n(s_1, s_2, \ldots, s_n) = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} h_n(\tau_1, \tau_2, \ldots, \tau_n) e^{-(s_1\tau_1 + \ldots + s_n\tau_n)} d\tau_1 d\tau_2 \ldots d\tau_n \tag{2.25}$$

The input-output spectral relationship is obtained by Fourier transforming both sides of (2.23):

$$Y(f) = \sum_{n=1}^{\infty} \mathbf{Y_n}(f) \tag{2.26}$$

with

$$\mathbf{Y_n}(f) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} H_n(f_1, \ldots, f_n)\delta(f - f_1 - \ldots - f_n) \cdot \prod_{i=1}^{n} X(f_i) df_i. \tag{2.27}$$

where $X(f)$ is the Fourier transform of the input signal $x(t)$. The previous relationships show that the first order kernel $H_1(s_1)$ is the transfer function of the linearized system. The Volterra series permits to calculate all mixing products of the excitations (harmonics, intermodulation, etc.). The order of the output component can be thought of as the number of contributing input frequencies. More precisely, the order O of the mixing product $f = mf_1 + nf_2 + pf_3 + \ldots$ is the sum of the coefficients of the input frequencies: $O = |m| + |n| + |p| + \ldots$ There are many cases where different order nonlinearities can result in responses at the same frequency.

To illustrate this feature, let us take the example of a second-order nonlinear system. Its operation can be seen as the combination of two signals, which eventually are identical. When two sine waves at frequency f_1 and f_2 are applied, $X(f) = A\delta(f - f_1) + A^*\delta(f + f_1) + B\delta(f - f_2) + B^*\delta(f + f_2)$, the first order response is given by (2.27) for $n = 1$:

$$\mathbf{Y_1}(f) = AH_1(f_1)\delta(f-f_1)+A^*H_1^*(f_1)\delta(f+f_1)+BH_1(f_2)\delta(f-f_2)+B^*H_1^*(f_2)\delta(f+f_2) \tag{2.28}$$

This corresponds to the response of the linearized system. The second-order response is similarly obtained from (2.27) for $n = 2$, producing nine terms at frequencies $f_1 \pm f_1$, $f_2 \pm f_2$ and $f_1 \pm f_2$. A representative term in $\mathbf{Y_2}(f)$ at the frequency $f_1 + f_2$ is of the form

$$[H_2(f_2, f - f_1) + H_2(f_1, f - f_2)]\delta(f - f_1 - f_2). \tag{2.29}$$

In contrast to the Taylor series, the Volterra series can describe weakly nonlinear systems with inductors and capacitors. Note that for a memoryless system, the Volterra series reduces to the Taylor series.

Table 2.2: Different responses at the output of a nonlinear system described by Volterra kernels of order 1, 2 and 3, excited by $A_1 \cos \omega_1 t + A_2 \cos \omega_2 t$.

Order	Frequency of response	Amplitude of response	Type of response
1	ω_1	$A_1\|H_1(j\omega_1)\|$	Linear
1	ω_2	$A_2\|H_1(j\omega_2)\|$	
3	$\omega_1 = \omega_1 + \omega_2 - \omega_2$	$\frac{3}{2}A_1A_2^2\|H_3(j\omega_1, j\omega_2, -j\omega_2)\|$	Third-order
3	$\omega_2 = \omega_1 - \omega_1 + \omega_2$	$\frac{3}{2}A_1^2A_2\|H_3(j\omega_1, -j\omega_1, j\omega_2)\|$	desensitization
3	$\omega_1 = 2\omega_1 - \omega_1$	$\frac{3}{4}A_1^3\|H_3(j\omega_1, j\omega_1, -j\omega_1)\|$	Third-order expansion
3	$\omega_2 = 2\omega_2 - \omega_2$	$\frac{3}{4}A_2^3\|H_3(j\omega_2, j\omega_2, -j\omega_2)\|$	or compression
2	$\omega_1 + \omega_2$	$A_1A_2\|H_2(j\omega_1, j\omega_2)\|$	Second-order
2	$\|\omega_1 - \omega_2\|$	$A_1A_2\|H_2(j\omega_1, -j\omega_2)\|$	IM products
2	$2\omega_1$	$\frac{1}{2}A_1^2\|H_2(j\omega_1, j\omega_1)\|$	Second
2	$2\omega_2$	$\frac{1}{2}A_2^2\|H_2(j\omega_2, j\omega_2)\|$	Harmonics
2	0	$\frac{1}{2}A_1^2\|H_2(j\omega_1, -j\omega_1)\|$	DC
2	0	$\frac{1}{2}A_2^2\|H_2(j\omega_2, -j\omega_2)\|$	shift
3	$2\omega_1 + \omega_2$	$\frac{3}{4}A_1^2A_2\|H_3(j\omega_1, j\omega_1, j\omega_2)\|$	
3	$\|2\omega_1 - \omega_2\|$	$\frac{3}{4}A_1^2A_2\|H_3(j\omega_1, j\omega_1, -j\omega_2)\|$	Third-order
3	$\omega_1 + 2\omega_2$	$\frac{3}{4}A_1A_2^2\|H_3(j\omega_1, j\omega_2, j\omega_2)\|$	IM products
3	$\|\omega_1 - 2\omega_2\|$	$\frac{3}{4}A_1A_2^2\|H_3(j\omega_1, -j\omega_2, -j\omega_2)\|$	
3	$3\omega_1$	$\frac{1}{4}A_1^3\|H_3(j\omega_1, j\omega_1, j\omega_1)\|$	Third
3	$3\omega_2$	$\frac{1}{4}A_2^3\|H_3(j\omega_2, j\omega_2, j\omega_2)\|$	harmonics

The different responses at the output of a nonlinear system described by Volterra kernels of order one, two and three, excited by two sinusoids $A_1 \cos(\omega_1 t)$ and $A_2 \cos(\omega_2 t)$ are given in table 2.2. For instance, the HD_3 can be calculated from that table:

$$HD_3 = \frac{A_1^2}{4} \left| \frac{H_3(j\omega_1, j\omega_1, j\omega_1)}{H_1(j\omega_1)} \right| \tag{2.30}$$

The remaining question is: How to evaluate the kernels? The most common way to compute the kernels is the so-called *nonlinear current method*, widely explained in [11, 13, 2]. This method is summarized as follows and will be applied to a MOSFET in the next subsection:

1. Solve for the first-order response $y_1(t)$ which is the response of the linear part of the circuit to the excitation $x(t)$ as if the nonlinearity were removed from the circuit.

2. When the first-order voltage across the nonlinearity $y_1(t)$ is found, compute the nonlinear current $i_{NL2}(t)$. In general, compute the nonlinear current $i_{NLn}(t)$

Table 2.3: Nonlinear current sources [2]. K_{ij} is the nonlinear coefficient of order i as given by (2.3) for a conductance ($j = g$) or a capacitance ($j = C$). H_{ij} is the Volterra kernel off order i evaluated at the circuit node j.

Type of basic nonlinearity	Nonlinear current source of order two	Nonlinear current source of order three
(trans) conductance	$K_{2g1}H_{1k}(s_1)H_{1k}(s_2)$	$K_{3g1}H_{1k}(s_1)H_{1k}(s_2)H_{1k}(s_3)$ $+\frac{2}{3}K_{2g1}\left[H_{1k}(s_1)H_{2k}(s_2,s_3)\right.$ $\left.+H_{1k}(s_2)H_{2k}(s_1,s_3)+H_{1k}(s_3)H_{2k}(s_1,s_2)\right]$
capacitance	$(s_1+s_2)K_{2C}$ $H_{1k}(s_1)H_{1k}(s_2)$	$(s_1+s_2+s_3)K_{3C}H_{1k}(s_1)H_{1k}(s_2)H_{1k}(s_3)$ $+\frac{2}{3}K_{2C}(s_1+s_2+s_3)\left[H_{1k}(s_1)H_{2k}(s_2,s_3)\right.$ $\left.+H_{1k}(s_2)H_{2k}(s_1,s_3)+H_{1k}(s_3)H_{2k}(s_1,s_2)\right]$
two-dimensional conductance (only cross terms)	$\frac{1}{2}K_{2g1\&g2}\left[H_{1k}(s_1)H_{1m}(s_2)\right.$ $\left.+H_{1k}(s_2)H_{1m}(s_1)\right]$	$\frac{1}{3}K_{2g1\&g2}\left[H_{1k}(s_1)H_{2m}(s_2,s_3)+H_{1k}(s_2)H_{2m}(s_1,s_3)\right.$ $H_{1k}(s_3)H_{2m}(s_1,s_2)+H_{2k}(s_1,s_2)H_{1m}(s_3)$ $\left.+H_{2k}(s_1,s_3)H_{1m}(s_2)+H_{2k}(s_2,s_3)H_{1m}(s_1)\right]$ $+\frac{1}{3}K_{3,2g1\&g2}\left[H_{1k}(s_1)H_{1k}(s_2)H_{1m}(s_3)+\right.$ $\left.H_{1k}(s_1)H_{1k}(s_3)H_{1m}(s_2)+H_{1k}(s_2)H_{1k}(s_3)H_{1m}(s_1)\right]$ $+\frac{1}{3}K_{3,g1\&2g2}\left[H_{1k}(s_1)H_{1m}(s_2)H_{1m}(s_3)\right.$ $\left.+H_{1k}(s_2)H_{1m}(s_1)H_{1m}(s_3)+H_{1k}(s_3)H_{1m}(s_1)H_{1m}(s_2)\right]$

when each $y_{n-1}(t)$ is determined. Nonlinear currents sources of order 2 and 3 are found in table 2.3.

3. Determine $y_n(t)$ from the linear differential equation

$$L[y_n(t)] + i_{NLn}(t) = 0, \qquad n = 2, 3, \ldots \tag{2.31}$$

in which L denotes symbolically the linear circuit operations and $i_{NLn}(t)$ is the nonlinear current source computed from the knowledge of $y_1(t), \ldots, y_{n-1}(t)$.

4. Finally, the total response $y(t)$ is the sum of all the components:

$$y(t) = \sum_{n=1}^{\infty} y_n(t). \tag{2.32}$$

2.2.2.2 Volterra analysis of a MOSFET

In this section, we will apply the method of nonlinear currents to a MOSFET in common-source (CS) configuration. The simple equivalent circuit presented in Fig. 2.4 will be used. The drain current I_D is assumed to be:

$$I_D = I_{D0} + g_{m1}V_G + g_{m2}V_G^2 + g_{m3}V_G^3 + g_{d1}V_D + g_{d2}V_D^2 + g_{d3}V_D^3 \tag{2.33}$$

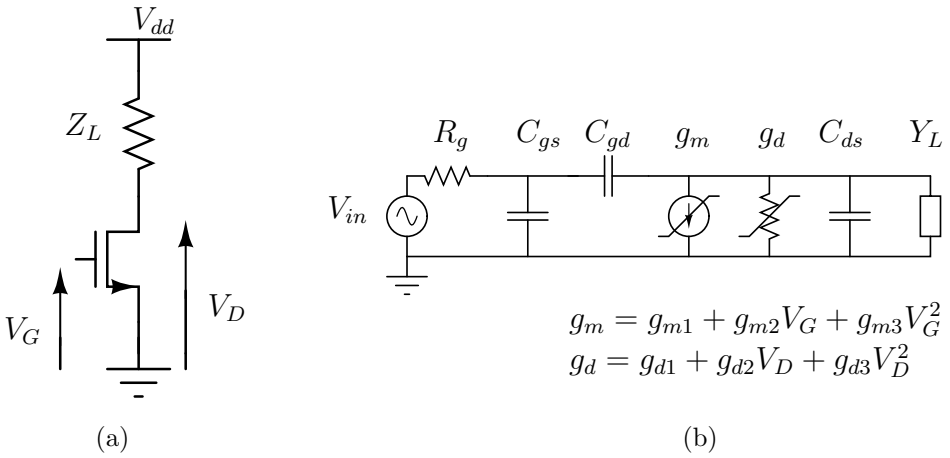


Figure 2.4: (a) Common-source MOSFET configuration. (b) its equivalent schematic used in the Volterra series analysis.

This means that the cross-derivatives are not taken into account and we only assume that the transconductance g_m and the output conductance g_d are nonlinear. As discussed in [14], this model is well appropriated to analyze the distortion of a MOSFET in the saturation region. In the linear region however, the cross-modulation terms are no more negligible. Furthermore, the capacitive elements do not generate any significant harmonics in saturation. Nevertheless, the linear capacitances are included in the model. C_{gd} influences the harmonics by the feedback; C_{ds} reduces the output impedance at HF, and C_{gs} filters the inputs. Compared to BJTs, MOSFETs have more linear capacitances, which is an advantage of this technology.

First order response The first order response corresponds to the linear case (Fig. 2.5(a)). The considered independent source is $V_{in} = A\sin(\omega t)$. We can write the circuit equations from the Kirkkhoff's relations:

$$\begin{pmatrix} (C_{gs}+C_{gd})s + 1/R_g & -C_{gd}s \\ g_{m1} - C_{gd}s & Y_L + g_{d1} + (C_{ds}+C_{gd})s \end{pmatrix} \cdot \begin{pmatrix} V_G \\ V_D \end{pmatrix} = \begin{pmatrix} A/R_g \\ 0 \end{pmatrix}$$

The first-order Volterra kernels are obtained by the normalizing the independent source A to one:

$$\begin{pmatrix} (C_{gs}+C_{gd})s + 1/R_g & -C_{gd}s \\ g_{m1} - C_{gd}s & Y_L + g_{d1} + (C_{ds}+C_{gd})s \end{pmatrix} \cdot \begin{pmatrix} H_{11}(s) \\ H_{12}(s) \end{pmatrix} = \begin{pmatrix} 1/R_g \\ 0 \end{pmatrix}$$

$H_{11}(s)$ and $H_{12}(s)$ are the solution of this linear set of equations. Note that in the notation H_{xy}, the index x refers to the order of the kernel and y to the appropriate node of the circuit (1 is node G and 2, node D).

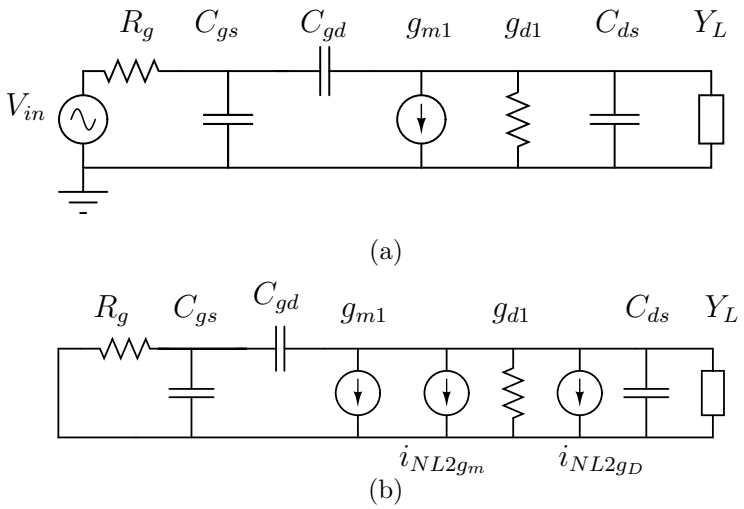


Figure 2.5: Circuit that has to be solved for the calculation of (a) the first and (b) the second order kernels of the circuit in Fig. 2.4.

Second-order response The first order kernels being known, we may calculate the nonlinear current sources of order two, corresponding to the linear response on second order coefficient of g_m and g_d (table 2.3):

$$i_{NL2g_m} = g_{m2} \cdot H_{11}(s_1)H_{11}(s_2) \tag{2.34}$$

$$i_{NL2g_d} = g_{d2} \cdot H_{12}(s_1)H_{12}(s_2) \tag{2.35}$$

$$i_{NL2tot} = i_{NL2g_d} + i_{NL2g_m} \tag{2.36}$$

Those are the only sources to be considered in the analysis, as we assumed that V_{in} was a pure sine wave. The second-order system is represented by (Fig. 2.5(b)):

$$\begin{pmatrix} (C_{gs}+C_{gd})(s_1+s_2)+1/R_g & -C_{gd}(s_1+s_2) \\ g_{m1}-C_{gd}(s_1+s_2) & Y_L+g_{d1}+(C_{ds}+C_{gd})(s_1+s_2) \end{pmatrix} \cdot \begin{pmatrix} H_{21}(s_1,s_2) \\ H_{22}(s_1,s_2) \end{pmatrix} = \begin{pmatrix} 0 \\ -i_{NL2tot} \end{pmatrix} \tag{2.37}$$

Third order response Similarly as for the second order, we may write the third order equations of the system as follows:

$$\begin{pmatrix} (C_{gs}+C_{gd})(s_1+s_2+s_3)+1/R_g & -C_{gd}(s_1+s_2+s_3) \\ g_{m1}-C_{gd}(s_1+s_2+s_3) & Y_L+g_{d1}+(C_{ds}+C_{gd})(s_1+s_2+s_3) \end{pmatrix} \cdot \begin{pmatrix} H_{31}(s_1,s_2,s_3) \\ H_{32}(s_1,s_2,s_3) \end{pmatrix} = \begin{pmatrix} 0 \\ -i_{NL3tot} \end{pmatrix} \tag{2.38}$$

with

$$
\begin{aligned}
i_{NL3tot} &= i_{NL3g_m} + i_{NL3g_d} && (2.39)\\
i_{NL3g_m} &= g_{m3} \cdot H_{11}(s_1)H_{11}(s_2)H_{11}(s_3) + \frac{2}{3}g_{m2}(H_{11}(s_1)H_{21}(s_2,s_3) \\
&\quad + H_{11}(s_2)H_{21}(s_1,s_3) + H_{11}(s_3)H_{21}(s_1,s_2)) && (2.40)\\
i_{NL3g_d} &= g_{d3} \cdot H_{12}(s_1)H_{12}(s_2)H_{12}(s_3) + \frac{2}{3}g_{d2}(H_{12}(s_1)H_{22}(s_2,s_3) \\
&\quad + H_{12}(s_2)H_{22}(s_1,s_3) + H_{22}(s_3)H_{22}(s_1,s_2)) && (2.41)
\end{aligned}
$$

Nonlinear performance parameters The large-signal figures of merit are evaluated from the Volterra kernels. The expressions are later approximated as follows [15]. The approximation consists of eliminating the high order terms in frequency and load. Further, we opted for a poles and zeros form:

$$HD_2(f) \approx HD_{2DC}\frac{1 + jf/f_{zHD_2}}{1 + jf/f_{pHD_2}} \tag{2.42}$$

with

$$
\begin{aligned}
HD_{2DC} &= \frac{A}{2}\left|\frac{g_{m2}}{g_{m1}} + \frac{g_{d2}g_{m1}}{Y_L'^2}\right| && (2.43)\\
f_{zHD_2} &= \frac{1}{4\pi}\left[R_g(C_{gd}+C_{gs}) + \frac{g_{m2}Y_L'(C_{ds}+C_{gd}) - g_{d2}g_{m1}C_{gd}}{g_{m2}Y_L'^2 + g_{d2}g_{m1}^2}\right]^{-1} && (2.44)\\
f_{pHD_2} &= \frac{1}{2\pi}\left[3R_g\left(C_{gd}\left(\frac{g_{m1}}{Y_L'}+1\right)+C_{gs}\right) + 3\frac{C_{ds}+C_{gd}}{Y_L'} - \frac{C_{gd}}{g_{m1}}\right]^{-1} && (2.45)\\
Y_L' &= g_{d1} + \frac{1}{Z_L} && (2.46)
\end{aligned}
$$

and

$$HD_3(f) \approx HD_{3DC}\frac{(1 + jf/f_{z1HD_3})(1 + jf/f_{z2HD_3})}{(1 + jf/f_{p1HD_3})(1 + jf/f_{p2HD_3})} \tag{2.47}$$

with

$$
\begin{aligned}
HD_{3DC} &= \frac{A^2}{4}\left|\frac{(-g_{d3}Y_L' + 2g_{d2}^2)g_{m1}^2}{Y_L'^4} + \frac{2g_{d2}g_{m2}}{Y_L'^2} + \frac{g_{m3}}{g_{m1}}\right| && (2.48)\\
f_{z1HD_3} &= \frac{2g_{d2}^2 - Y_L'g_{d3}}{4\pi\left[(ag_{d3} - 2R_g g_{d2}^2)C_{gd} + g_{d3}C_{ds} + (-2R_g g_{d2}^2 + g_{d3}Y_L'R_g)C_{gs}\right]} && (2.49)\\
f_{z2HD_3} &= \frac{1}{6\pi R_g C_{gd}} && (2.50)
\end{aligned}
$$

$$f_{p1HD_3} = \frac{-Y_L^{'}/2\pi}{\left[7C_{gd}\left(a - \frac{Y_L^{'}}{7g_{m1}}\right) + 4(C_{ds} + Y_L^{'}R_gC_{gs})\right]} \tag{2.51}$$

$$f_{p2HD_3} = \frac{-g_{m1}}{2\pi C_{gd}} \tag{2.52}$$

$$a = 1 + R_g(g_{m1} + Y_L^{'}) \tag{2.53}$$

Note that (2.43) and (2.48) are the factors given by a Taylor analysis. The third-order distortion is not only generated by the third-order nonlinear coefficient g_{m3}, but also by the combination of lower-order terms (see (2.48)).

Following the same methodology, a two-tone analysis may be performed. In order to simplify the expressions, C_{gd} was neglected. This means that a perfect isolation between output and input is supposed. The error introduced by this simplification will be discussed in section 2.3.3. It was also assumed that the tones are close together with regard to the operation angular frequency ω and that the global output conductance $G_o(\omega)$ verifies $G_o^*(\omega) = G_o(-\omega)$. This hypothesis is verified in the case of the considered MOSFET as $G_o(\omega) = Y_L + g_{d1} + j\omega C_{ds}$ is linearly related to the frequency. Then, the intermodulation distortion of order 3 is given by:

$$IMD_3 = \frac{3}{4}A^2\frac{1}{1 + \omega^2 R_g^2 C_{gs}^2}\,|IM_3| \tag{2.54}$$

where

$$\begin{aligned} IM_3 = {} & \frac{g_{m3}}{g_{m1}} - \left(\frac{g_{m1}}{G_o(\omega)}\right)^2 \frac{g_{d3}}{G_o^*(\omega)} \\ & + \frac{2}{3}g_{d2}g_{m2}\left(\frac{1}{G_o(2\omega)G_o^*(\omega)} + \frac{2}{G_o^*(\Delta\omega)G_o(\omega)}\right) \\ & + \frac{2}{3}\left(\frac{g_{m1}}{G_o(\omega)}\right)^2 \frac{g_{d2}^2}{G_o^*(\omega)}\left(\frac{1}{G_o(2\omega)} + \frac{2}{G_o^*(\Delta\omega)}\right) \end{aligned} \tag{2.55}$$

It follows from (2.54) and (2.55) that the IMD_3 not only depends on the third-order nonlinearity of the transconductance g_{m3} as commonly assumed, but also on the output conductance at low-frequency $G_o(\Delta\omega)$ and at the second harmonic $G_o(2\omega)$. Also, the only action of C_{gs} is a lowering of the signal magnitude at the input of the device. As the output admittance of a Partially-Depleted (PD) SOI MOSFET is frequency dependent, we expect from those last relations to measure a variation of IMD_3 as a function of the tones separation for that device [16]. This will be discussed in section 2.4.

2.2.3 Describing functions

A classical procedure for dealing in a practical way with nonlinear systems is to construct a linear model to approximate the nonlinearities. A small-signal linearization consists for instance of expanding the nonlinear function in a Taylor series around some operating point and retaining only the linear term in the analysis. If the excitations are not small enough to permit this simple form of linearization, one can do better by letting the linear approximation depends on the input. The idea behind describing functions (DF) is to extend the classical transfer function concept by finding a quasi-linear approximation of a nonlinearity that corresponds to a specific input signal [17].

Because the DF is a kind of linearization of the system, it is not used to compute HD and IMD. Nevertheless, it is introduced here as it will be demonstrated in the following that it is powerful to study the influence of the amplitude on different circuit parameters.

Suppose that the input of a nonlinear operator N is $x(t) = \sum_i x_i(t)$, and its output is $y(t)$. Then, we look for the approximation $y_a(t) = \sum_i w_i(t)x_i(t)$ of $y(t)$ so that the error is minimized in the least mean square sense. It can be shown that the set of coefficients w_i which minimizes the mean-square approximation error is that set which equates, over the non-negative range of their arguments, the input-output cross-correlation function for the nonlinearity and its approximation [17]. This cross-correlation equivalence is required for each component of the input.

If the components of the inputs $x_i(t)$ are statistically independents, the optimal set of w_i must satisfy the following condition [17]:

$$\int_0^\infty w_i(\tau_2)\phi_{ij}(\tau_1 - \tau_2)d\tau_2 = \overline{y(t)x_i(t-\tau_1)} \qquad \tau_{1,2} \geq 0,\ i,j = 1,2,3,\ldots \tag{2.56}$$

where ϕ_{ij} is the cross-correlation function between i and j.

Suppose that the input of the system has a sinusoidal form whose phase φ is not a priori known: $x(t) = A\sin(\omega t + \varphi)$. Then, (2.56) becomes:

$$\begin{cases} \frac{A}{2}\int_0^\infty w_1(\tau_2)\cos(\omega\tau_2)d\tau_2 = \overline{y(0)\sin\varphi} \\ \frac{A}{2}\int_0^\infty w_1(\tau_2)\sin(\omega\tau_2)d\tau_2 = -\overline{y(0)\cos\varphi} \end{cases}$$

and the approximation N_A of the nonlinear operator N writes:

$$N_A = \frac{2}{A}\overline{y(0)\sin\varphi} + j\frac{2}{A}\overline{y(0)\cos\varphi} \tag{2.57}$$

$$= \quad n_p + j \cdot n_q \tag{2.58}$$

The describing function N_A is then a linear gain, chosen in such a way that it renders similar responses for the nonlinearity and its approximation, in the mean-square error sense, to the same sinusoidal input. The form (2.58) of N_A underlines the fact that the gain has in-phase and in-quadrature terms. In the case of memoryless, single-value nonlinearity, the quadrature gain is always zero.

On the other hand, the DF for a sinusoidal input component may be viewed as the amplitude and phase relationship between an input sinusoid and the fundamental harmonic component of the expectation of the output of the nonlinearity taken with respect to all statistical parameters except φ:

$$N_A(A,\omega) = \frac{j}{\pi A}\int_0^{2\pi} y(A\sin\phi, A\omega\cos\phi)e^{-j\phi}d\phi \tag{2.59}$$

where $\phi = \omega t$. The DF may be equivalently defined as the ratio between the phasor representation of the output component at frequency ω and the phasor representation of the input component at frequency ω:

$$N_A(A,\omega) = \frac{A_1(A,\omega)}{A}e^{j\phi_1(A,\omega)} \tag{2.60}$$

where A_1 and ϕ_1 are the amplitude and phase of the fundamental Fourier component.

The DF is sometimes referred as harmonic linearization, due to the form (2.59) of N_A. It will be used in the analysis of oscillators in the next chapter.

2.2.4 Integral function method

In the previous sections, we applied Taylor or Volterra series expansion of the measured output DC curves of a MOSFET to get its non-linear behavior at DC. The drawback of this technique is that it requires the calculation of high-order derivatives of the $I_D(V_D)$ and $I_D(V_G)$ characteristics which is sensitive to the noise introduced by the measurements. Recently, an integral function method (IFM) [5, 18] was proposed with the advantage to be less sensitive to the measurements noise than Fourier series analysis. Furthermore, as it was the case for the DF, the IFM depends on the signal amplitude. Mathematical basis of this method was presented in [19]. The method is summarized in section 2.2.4.1. As stated before, this method applies in that way on DC characteristics. We provide the extension of IFM to high frequency operation in section 2.2.4.2.

2.2.4.1 Principle

The IFM is based in the following features:

- the region of interest in the nonlinear DC transfer characteristic $Y(X)$ is selected; at the same time the input bias point X_o and the input signal amplitude A are selected.

- this characteristic, $Y(X)$, is normalized in both axis from 0 to 1: $y(x)$ (Fig. 2.6).

- the resulting square of area equal to one (Fig. 2.7), is divided into two areas, one above (AREA 2), and another below (AREA 1) the normalized curve. The integral function D is defined as the difference between these two areas

$$D = 2\int_0^1 y(x)dx - 1 \tag{2.61}$$

- The resulting function D is proportional to the total harmonic distortion including the DC shift [19], $THD_o = 1.06D$.

- It is preferable to use the modulus of the normalized characteristic is used, defining the D_s function, to avoid integral cancellation when $y(x)$ crosses the $y = x$ curve.

IFM computes the output difference $Y_r(V) = Y(X) - Y(-X)$ in order to eliminate the even harmonics in the output signal. Normalizing $Y_r(V)$ in the same way as above, function $D_r = 4\int_0^{0.5} y_r(x)dx - 0.5$ is calculated. Function D_r is defined as D_{rs} if the modulus of the characteristics is taken into account. Function D_r is equal to the third harmonic distortion HD_3 if the orders of nonlinearity higher than three are neglected. Under that assumption, the principal harmonic parameters can be calculated directly through the integral function D_s and D_{rs} as follows:

$$THD = \sqrt{\frac{(1.06D_s)^2}{2} + \frac{D_{rs}^2}{2}} \tag{2.62}$$

$$HD_2 = \sqrt{\frac{(1.06D_s)^2}{2} - \frac{D_{rs}^2}{2}} \tag{2.63}$$

$$HD_3 = D_r \tag{2.64}$$

If the input signal is equal to $X_o + A\sin(\omega t)$, D_s, D_r, and D_{rs} are functions of X_o and A, giving the possibility to analyze the harmonic distortion as a function of the bias voltage and the input signal amplitude.

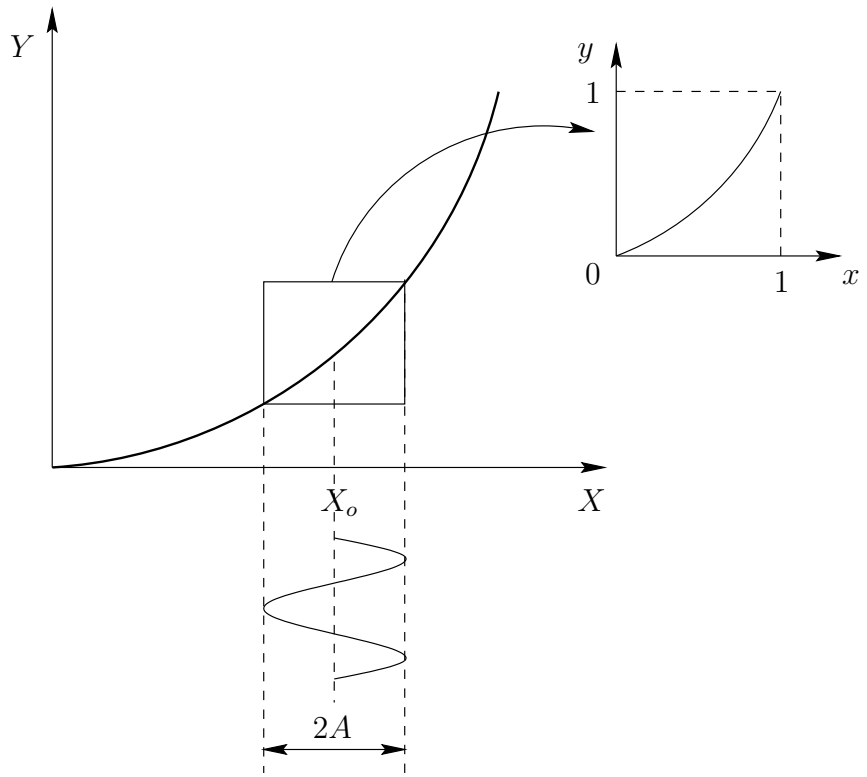


Figure 2.6: Normalization procedure in IFM.

The voltage intercept point of order 3 as a function of the derivatives under the condition of low amplitude of A, can also be defined as function of D_{rs} using relations (2.9) and (2.11):

$$VIP_{3h} = \frac{A}{\sqrt{D_{rs}}} \tag{2.65}$$

It is worth noting that the present version of IFM evaluates the intermodulation factors of merit under the same assumptions that the Taylor approach does. In other words, IFM computes the HD and IMD is next evaluated through relations such as (2.20).

2.2.4.2 Extension to HF

At HF, the transfer function differs from the DC characteristic. The presence of capacitors being charged and discharged produces indeed memory effects or hysteresis. In order to include this feature in the analysis, a time-domain simulation has to be performed, giving both the input $X(t)$ and output $Y(t)$ waveforms. Eliminating the time variable, the relationship between output and input ($Y(X)$) for each analyzed fundamental frequency is obtained (Fig. 2.8).

IFM can be used for the HD analysis of a large HF signal, if function D, defined in (2.66), is applied to the HF $Y(X)$ characteristics. Noting that the first term in

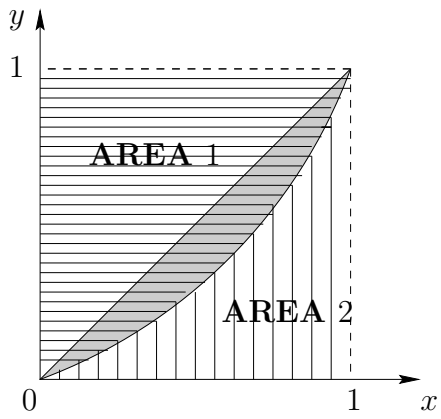


Figure 2.7: The difference between the normalized characteristic $y(x)$ and the $y = x$ curve is proportional to the THD.

(2.61) is indeed the integral of the normalized output as a function of the normalized input over a whole period of excitation, we have at HF:

$$D = \left[\int_0^1 y(x)dx\right]_{up} + \left[\int_0^1 y(x)dx\right]_{down} - 1 \tag{2.66}$$

where the *up* (*down*) subscript refers to the increasing (decreasing) half period of the excitation i.e., the upper (lower) part of the HF curve in Fig. 2.8. It is worth noting that relation (2.66) reduces to (2.61) in the case of a memoryless systems. The general expression (2.66) of the D function may then be used at HF as well as at LF and provide a new tool for the characterization of HF large signal nonlinearities.

2.2.5 Large signal transfer function

A general way to analyze distortion is to start directly from the large-signal transfer function (LSTF). The effects of g_m, g_d and the load impedance Z_L should be included in this characteristic. Indeed, from (2.42)-(2.54), the load impedance highly affects the distortion factors-of-merit. The characterization presented in this section is useful for the design of power amplifiers, as it allows treating large signal amplitudes.

2.2.5.1 Low frequency case

The transfer function is found from the $I_D(V_G, V_D)$ relationship of the considered device. This two-dimensional problem may be reduced to a one-dimensional one

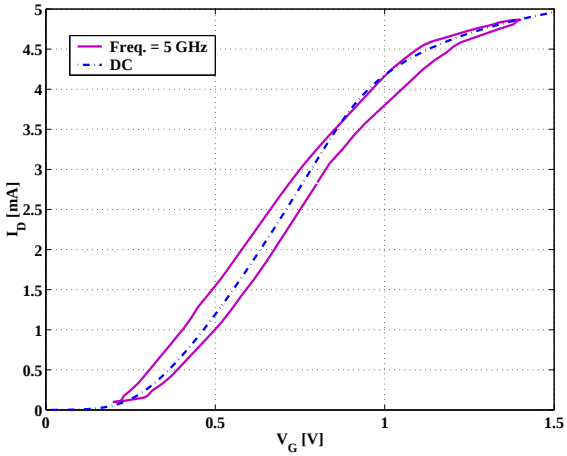


Figure 2.8: Transfer characteristics of a common-source 8x2 μm/0.14 μm FD SOI MOSFET from OKI corp. at DC and at HF for $A = 0.6$ V, $V_G = 0.8$ V, $Z_L = 200\ \Omega$, $V_{dd} = 1.2$ V (BSIM3v3 model provided by the founder). The 5 GHz frequency refers to the frequency of the input sinusoidal signal.

by adopting appropriated boundary conditions [20]. Indeed, for the common-source MOSFET in Fig. 2.4(a), the following relationship is always verified:

$$(V_{dd} - V_D) \cdot Y_L = I_D(V_G, V_D) \tag{2.67}$$

where $Y_L = 1/Z_L$ is the load admittance. Solving this equation for each V_G, we determine both the drain voltage as a function of the gate voltage and the large-signal current characteristic that includes the effect of the load $I_D(V_G, Z_L)$. The distortion may then be computed by IFM, Fourier or Taylor series expansion on that characteristic. Indeed, the $I(V_G, Z_L)$ characteristics may also be expressed as:

$$I(V_G, Z_L) = G_{m1}V_G + G_{m2}V_G^2 + G_{m3}V_G^3 + \dots \tag{2.68}$$

where G_{mi} are obtained by i^{th} differentiation of the $I(V_G, Z_L)$ characteristics; they may also be obtained from the g_{mi} and g_{di} coefficients as discussed in [21]. The effect of both transconductance and output conductance, as well as gain compression are then included in the HD obtained from the ratio of G_{mi}. The advantage of such method its simplicity since the transition from triode to saturation is well taken into account. It is interesting to evaluate the G_{mi} because the minima of HD_i correspond to the values that nullify G_{mi}.

It is interesting to remark that the effective gain G_{m1} is easily obtained as a function of the signal amplitude from the $I - V$ by the describing function (relation (2.59)). The higher-order coefficient G_{m2} and G_{m3} may also be obtained as a function of the signal amplitude without any derivative by the IFM. Indeed, the D_r

function is equal to HD_3, which is the ratio of G_{m3} to G_{m1}. Once G_{m1} is found by the DF approach, G_{m3} is obtained. It must be underlined that the G_{mk} obtained for A tending to zero are those obtained by differentiation of the $I(V_G, Z_L)$ characteristic.

On the other hand, it as been reported that minima of distortion (*sweet-spots*) depend on the input signal power [20]. For a correct representation of this phenomenon, a magnitude dependent technique, as IFM, has to be performed.

Designers are mostly interested in the linearity performances of an amplifier that has a given output swing. As we get from the transfer characteristics (equation (2.67) for a CS MOSFET) the relation between V_G and V_D, this technique allows to predict how to bias the MOSFET as a function of the load for a given output dynamic range and the associated distortion level.

2.2.5.2 Extension to high frequency

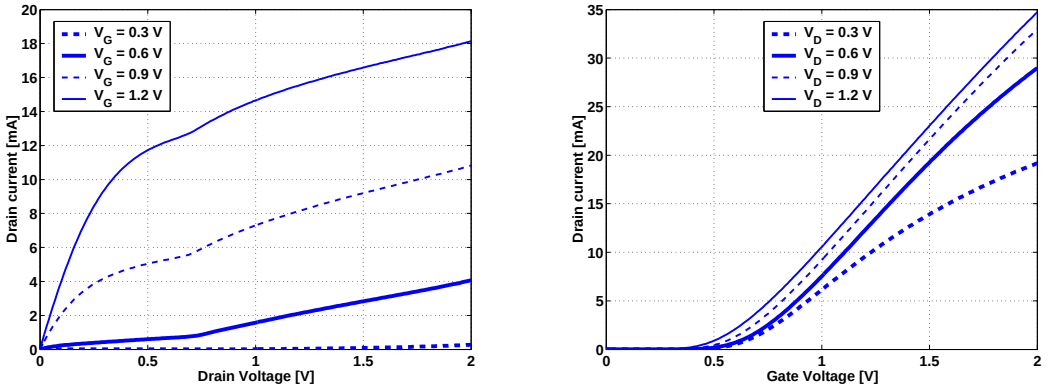
As previously stated, the Volterra series is widely used for the analysis of distortion at HF. The validity domain of the expressions given in section 2.2.2 is limited to the biases for which the FET is saturated i.e., where capacitances do not vary much. If a large signal is applied to the transistor of Fig. 2.4(a), numeric simulations showed that a third order Volterra model is valid if the LSTF obtained from section 2.2.5.1 is used in conjunction to a bias-dependent model of the capacitors. Unfortunately, this approach leads to huge expressions which give poor insight to the designer.

Best is to use the LSTF concept at HF. In this case, the LSTF differs from the DC one. Indeed, as memory effects are present, hysteresis appears in the LSTF. In order to include them into the analysis, a time-domain simulation has to be performed, giving both $I_D(t)$ and $V_G(t)$. By the elimination of the time variable, the relationship between output and input ($I_D(V_G)$) is obtained (Fig. 2.8). This function is the frequency dependent LSTF of the system. As previously stated, the IFM may be applied to this LSTF.

2.3 Nonlinear characterization of SOI nMOSFETs

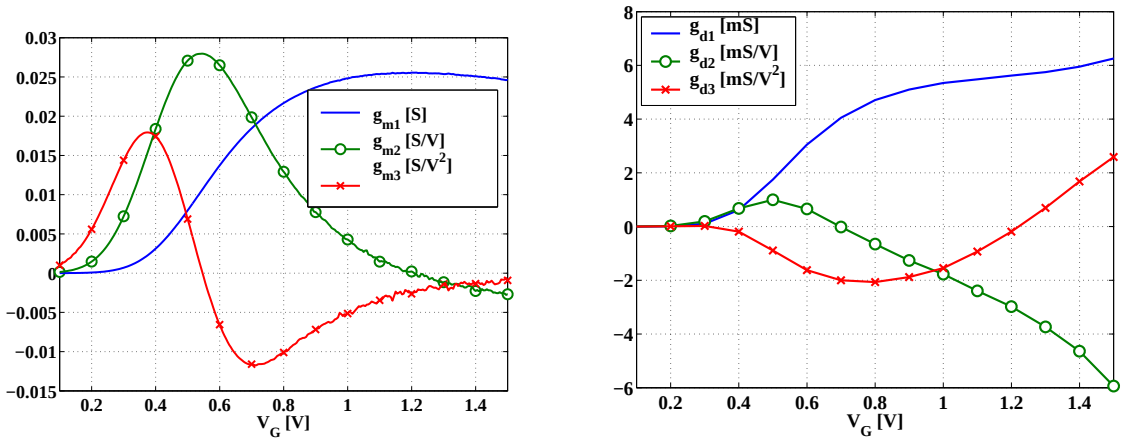
The characterization techniques described in the previous section are applied in the present section to a MOSFET in the common-source configuration. The influence of bias, load and frequency is discussed in details. The different extraction techniques are compared with LSNA measurements.

The considered device-under-test (DuT) is a Partially-Depleted (PD) Floating-



(a) (b)

Figure 2.9: Measured DC characteristics of the considered PD SOI nMOSFET.



(a) (b)

Figure 2.10: Nonlinear coefficients of the considered PD SOI nMOSFET, obtained from differentiation of the curves in Fig. 2.9; $V_D = 0.9$ V.

Body (FB) SOI nMOS composed of 12 fingers connected in parallel, each 6.6 μm width and 0.25 μm length. The devices were fabricated at CEA-LETI, France, following a 0.25 μm process. The extracted threshold voltage V_{th} value is 0.56 V. All measurements were performed on-wafer with the source and the back-gate of the MOSFET grounded. The measured[1] $I-V$ relation of the DuT is shown in Fig. 2.9.

The nonlinear coefficients defined in (2.33) are depicted in Fig. 2.10 as a function of the gate voltage. In weak inversion, the drain current follows an exponential type behavior with respect to the gate voltage, as diffusion mechanism dominates. The g_{m3} coefficient is positive in this region and as a consequence, the device experiences gain expansion. In the strong inversion region, however, the mechanism for drain

[1] An HP 4156 setup was used for that purpose.

current is mainly due to drift and the drain current characteristics follows roughly a square-law. In reality, short-channel effects reduce the exponent to values lower than two. Also, the mobility modeling is crucial for a good understanding of the distortion mechanisms. As a result, g_{m3} is nonzero and negative in strong inversion. The device experiences then gain compression. An important consequence of this feature is that g_{m3} passes through zero in the moderate inversion region. At this point, the device acts as an ideal square-law device and do not experience any third-order distortion. Higher order distortion components nevertheless also influence the third-order distortion, and even if it experiences a minimum, its value differs from zero. The existence of a low distortion region close to V_{th}, where g_{m3} is null, is thus inherent to the CMOS device and is independent of short-channel effects [22].

2.3.1 LSNA measurements

Performing nonlinear measurements of a device at RF and microwaves is a complex task. Recently, the LSNA setup was developed to facilitate the measurement and provide exhaustive information on the DuT that is needed for accurate nonlinear modeling at microwaves. The setup [8, 23, 9] injects sinusoid signals at different power levels and measures both phase and magnitude of the transmitted and reflected waves at the fundamental and the harmonics (Fig. 2.11). Signals up to 20 GHz can be detected, so for a 900 MHz fundamental frequency, the 22 first harmonics can be measured.

A first calibration (in both magnitude and phase) was done using a calibration kit on alumina substrate in order to put the reference planes at the end of the probes tips. Further de-embedding is required if we want to get rid of the parasitic introduced by the access pads and lines. For that purpose, an on-wafer calibration kit has to be measured and adequate algorithms applied. We checked that the input power level has no influence on the passive devices used for the calibration. For instance, the variation of the measured scattering parameters of a short-circuit is shown in Fig. 2.12. The standard deviation of $|S_{11}|$ is $3 \cdot 10^{-4}$. Therefore, linear measurements available from classical linear vectorial network analyzers (VNA) may be used for de-embedding purpose. We used two algorithms to de-embed our DuTs. The first one was proposed by Vandamme [24] (method A). The other one (method B) computes the scattering matrices $\underline{\underline{S}}_p$ of the parasitic structures from the calibration kit measurements, as described in [25, 26, 27]. Next, the matrices $\underline{\underline{S}}_p$ are converted into chain matrices, assuming a 50 Ω reference impedance of the equipment. Then the voltages and currents at both ports of the DuT are computed from those matrices

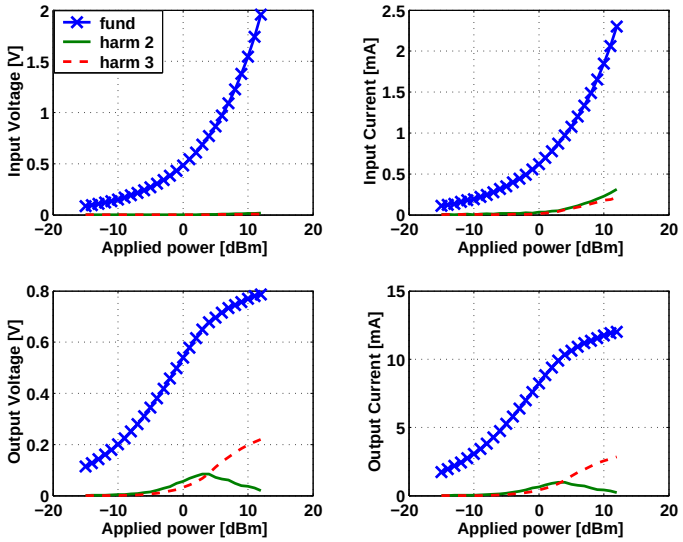


Figure 2.11: Magnitude of input and output current and voltage of our DuT measured as a function of the power delivered by the generator (LSNA measurements). $V_D = 0.9$ V and $V_G = 1.0$ V; $f = 900$ MHz.

and the measured current and voltages. To be accurate, this second method requires a good evaluation of the characteristic impedance of the on-wafer access lines. The two methods give results in agreements in the studied frequency band (Fig. 2.13). In the following, the de-embedding proposed in [24] (method A) will be considered, as it does not require the evaluation of the characteristic impedance of the access lines.

2.3.2 *Low frequency characterization*

At low frequencies, the capacitances may be neglected. The drain current I_D of the MOSFET can be easily obtained as a function of the gate and drain voltages by simulations or measurements. The harmonic distortion of order n (HD_n) of the characteristics are given by (2.7) and (2.9) where $K_i = g_{mi}$ or $K_i = g_{di}$ (the g_{mi} and g_{di} coefficients were defined in (2.33)).

As the effect of the transconductance is often the only nonlinearity taken into account [3], the distortion of the $I_D(V_G)$ relationship is investigated first.
The HD_2 of our DuT evaluated by (2.7), by the IFM applied on $I_D(V_G)$ and by LSNA measurements at 900 MHz is reported in Fig. 2.14 as a function of V_G. The results given by IFM and (2.7) are almost identical. Since we deal with measurements data, IFM only will be used in the following. For gate voltages between 0.2 V and 1.0 V LSNA method gives results in agreement. The minimum depends on

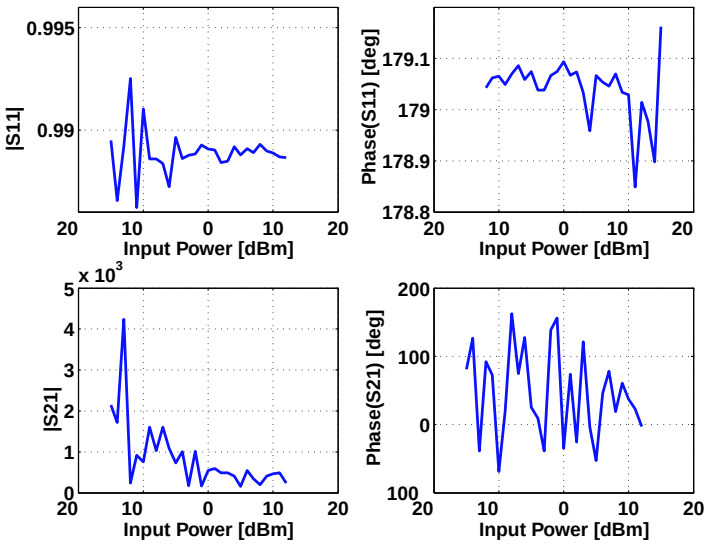


Figure 2.12: Influence of power on measured S parameters of a short circuit at 900 MHz. The input power is the power delivered by the generator.

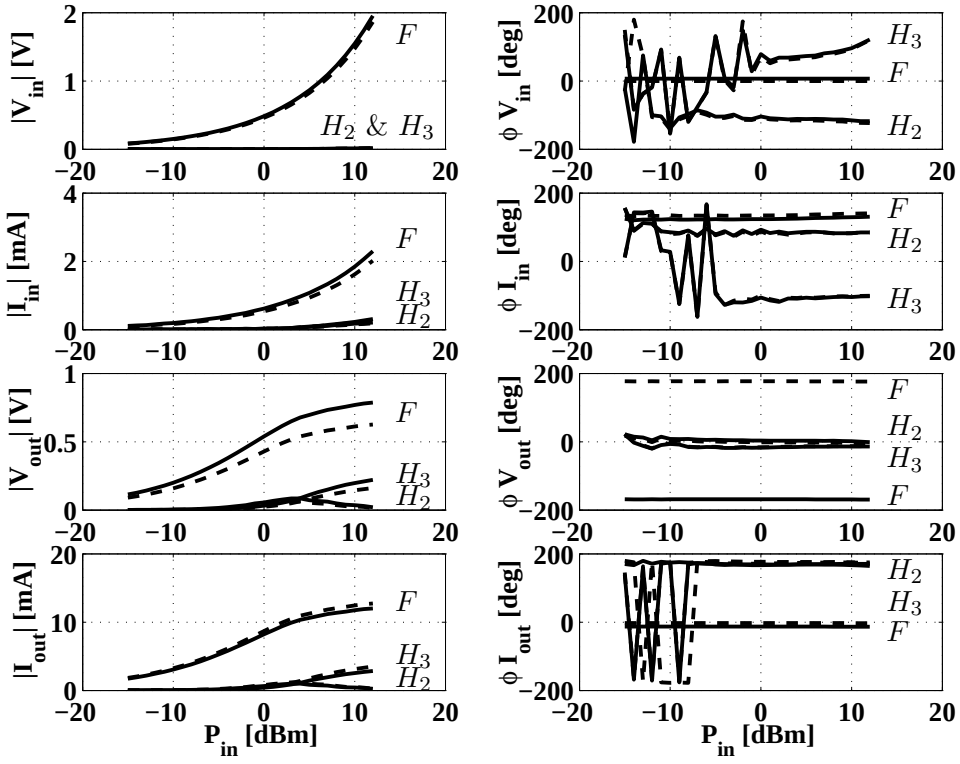


Figure 2.13: Comparison of de-embedding techniques for MOSFETs at 900 MHz. Straight lines: method B; dashed lines: method A. Fundamental (F) and the two first harmonics (H_2 and H_3) are given. PD MOSFET, P_{in} is the power delivered by the generator, $V_G = V_D = 1.2$V.

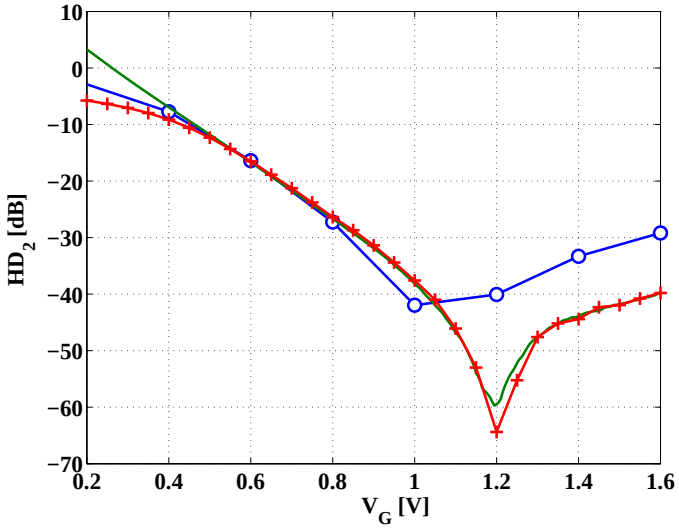


Figure 2.14: HD_2 of a 12x6.6 μm/0.25 μm PD SOI transistor measured with LSNA (circles) and calculated by IFM applied to $I_D(V_G)$ (crosses) and relation (2.7) (straight line) as a function of V_G; $V_D = 1.2$ V and the amplitude of the AC input signal A is 0.2 V. The LSNA measurements are performed on a 50 Ω setup at 900 MHz.

the drain bias condition as it is determined by the inflection point of the considered transfer function. It corresponds to the gate transconductance maximum for the methods based on $I_D(V_G)$, while from RF measurements, the gate voltage at which the minimum is reached corresponds to the maximum power gain G_P of the transistor, defined as $G_P = \left|\frac{(V \cdot I^*)_{output}}{(V \cdot I^*)_{input}}\right|_{f=f_o}$.

It may be concluded from Fig. 2.14 that analyzing the transconductance only it is not enough to characterize the nonlinear behavior of a MOSFET at high V_G in practical cases when the transistor is loaded. Indeed, if the load impedance would be zero, no dependence to the drain voltage would occur. However, a 50 Ω load is used in the LSNA measurements. The effect of the output conductance is taken into account in this section, while the effect of the capacitances will be given in the next section. Introducing the effect of the output conductance, we find for the transistor in Fig. 2.4, in the case of weak nonlinearities [28]:

$$HD_2 \simeq \frac{A}{2}\left|\frac{g_{m2}}{g_{m1}} - A_{vDC}\frac{g_{d2}}{(Y_L + g_{d1})}\right| \tag{2.69}$$

$$HD_3 \simeq \frac{A^2}{4}\left|\frac{g_{m3}}{g_{m1}} - A_{vDC}^2\frac{g_{d3}}{(Y_L + g_{d1})}\right| \tag{2.70}$$

where $Y_L = 1/Z_L$ is the load admittance and $A_{vDC} = -g_{m1}/(Y_L + g_{d1})$ is the linear DC voltage gain. Relations (2.69) and (2.70) confirm the intuition that the lower the load impedance, the lower the effect of the g_d nonlinearity.

When dealing with measurements, the evaluation of the derivatives may become tricky, as the fluctuations are amplified [22]. One could chose to fit the data to a high degree polynomial, but this lacks of physical insight. A more complicated experimental setup may also be used to circumvent this [2]. When dealing with $I-V$ measurements, as already mentioned, the IFM is an alternative that limits numerical problems if a sufficiently large amplitude is considered. Indeed, the numerical noise is averaged over the amplitude of the input sine wave.

Both LSNA and IFM allow to extract the harmonic voltage intercept points of order three (VIP_{3h}). The extraction with IFM is based on the assumption that the signal amplitude A is low enough so that the relation (2.71) holds.

$$VIP_{3h} = 2\sqrt{\left|\frac{g_{m1}}{g_{m3}}\right|} \tag{2.71}$$

The chosen value for A should however not tend to zero, because the averaging decreases, increasing the sensitivity to noise measurements. In other words, the amplitude should be taken small enough so that the system is weakly nonlinear. As shown in Fig. 2.15, there is a good agreement between the methods. At high V_G, the discrepancy comes from the fact that the effect of the output conductance was not taken into account. The IFM is more robust that the direct evaluation of (2.71) as no derivatives are needed. On the other hand, the evaluation of the VIP from the plot of output power *vs* input power measured at RF is very sensitive to the extrapolation of the linear part of the curves. Furthermore, the peak around the threshold voltage is still present at RF. As in practical cases, balanced circuits are use to cancel even order distortion, this means that a careful bias of the transistors in the moderate inversion improves the linearity of the devices. However, this bias is not often used because it is a very narrow sweet-spot that depends on the spreading of the threshold in the fabrication process. The effect of feedback on the sweet-spot, as present in most real circuit configurations, will be discussed in section 2.5.

2.3.3 Volterra model

The model presented in section 2.2.2.2 is used here to characterize our DuT. As shown in Fig. 2.16, there is a good agreement between the model and the LSNA measurements. It worth noting that the cross-derivative terms are not included in the representation of the $I-V$ characteristics. Numerical simulations showed that the presence of these terms improved the accuracy of the model compared to measurements, especially in the triode regime. Even if it is not difficult to introduce

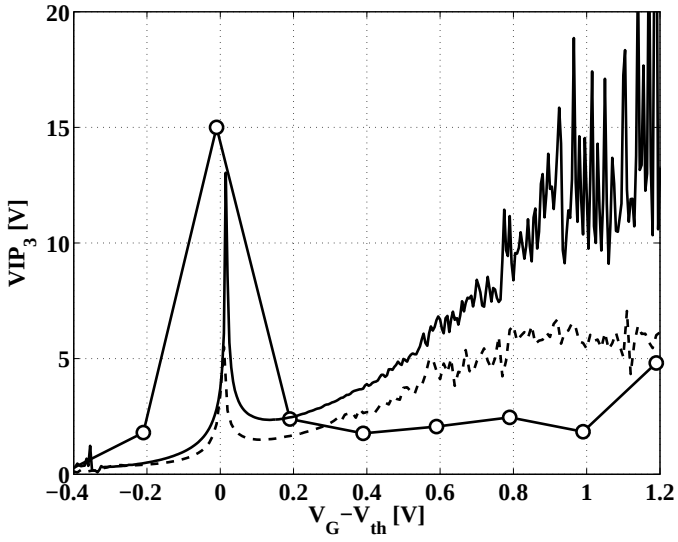


Figure 2.15: VIP_{3h} as a function of the gate-voltage overdrive, extracted from IFM (dashed line), LSNA (circles) and equation (2.71) (straight line); the drain bias is set to 1.2 V.

these terms, they will however be neglected through this section, in order to keep the simplicity of the modeling.

It appears from the previous section that the HF behavior of HD may be characterized by poles and zeros. Their values depend on the bias point of the MOSFET and on the load impedance. For our DuT and for most bias points, the zeros of (2.42) and (2.47) lie at higher frequency than their poles. Furthermore, in (2.47) the dominant pole and zero of HD_3 are respectively f_{p1HD_3} and f_{z1HD_3}. In other words, both HD_2 and HD_3 are almost constant until the frequency f_{pHD_2} or f_{p1HD_3} is respectively reached. It may be seen in Fig. 2.17 that the HD predicted by the model decreases at high gate voltage with the frequency. The linear capacitances included in the model reduce the output impedance at HF. The voltage swing also decreases, increasing the effective output conductance linearity. On the other hand, HD_2 increases with frequency at the V_G corresponding to its minimum (Fig. 2.17), due to the feedback action of C_{gd}. The dominant poles are plotted in Fig. 2.18 as a function of the load impedance Z_L. For instance, the f_{pHD2} and f_{pHD3} of our DuT loaded by $Z_L = 200\ \Omega$ are found to be about 10 GHz and 4.5 GHz, respectively. It is interesting to note that the ratio between the pole of the voltage gain A_v and the one of HD_2 (HD_3) is almost constant at relatively high value of Z_L. Furthermore, it can be concluded from relations (2.45) and (2.51) that

$$|f_{pHD_2}| \geq \frac{|f_{pA_v}|}{3} \tag{2.72}$$

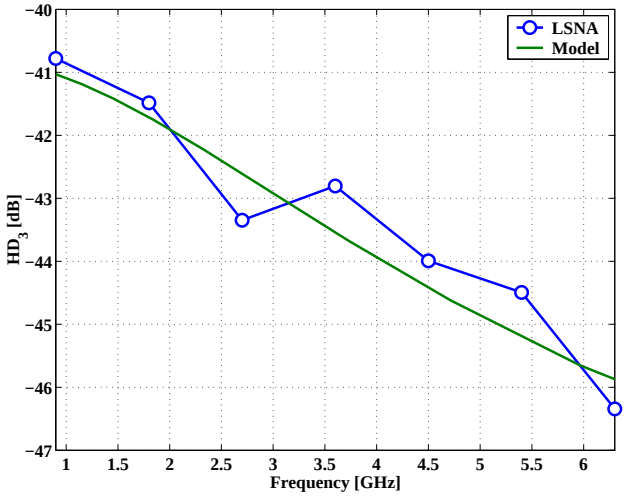


Figure 2.16: HD_3 from 900 MHz LSNA measurements and our model; $V_D = V_G = 1.2$ V; applied AC magnitude is $A = 0.3$ V.

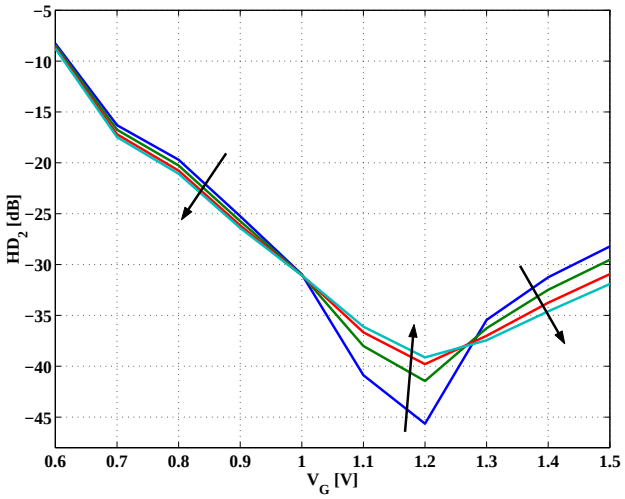


Figure 2.17: HD_2 evaluated by (2.42) for frequencies between 1 MHz and 12 GHz. The arrows indicate a frequency increase. $V_D = 1.2$ V, $A = 0.3$ V; $Z_L = 50\Omega$.

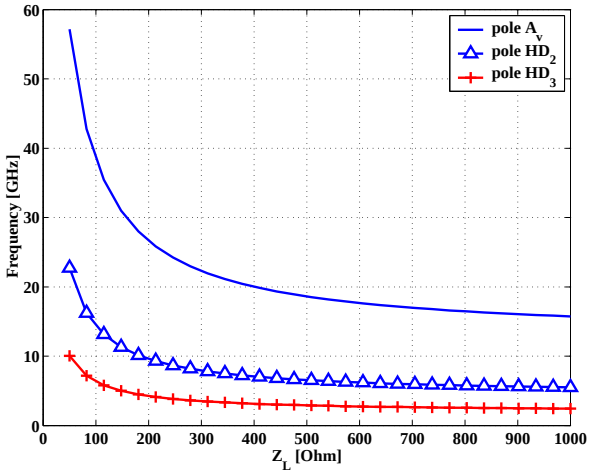


Figure 2.18: Poles of HD and voltage gain of our DuT; $V_D = V_G = 1.2$ V.

$$|f_{p1HD_3}| \geq \frac{|f_{pA_v}|}{7} \tag{2.73}$$

Those last relations give us a rule of thumb to find the frequency at which the low-frequency analysis (e.g., Taylor approach or IFM) is not valid anymore.

In order to check the assumptions of the model presented in Section 2.2.2.2, the kernels of Volterra were also calculated numerically with an extended version of the model, denoted *model B*. In model B, the LF drain current is again modeled as (2.33), but this time the three capacitances (C_{gs}, C_{gd} and C_{ds}) are also described by a third-order Taylor series. As we deal with numerical simulations, no simplifications of the kernels were done. In the following, *model A* will refer to the schematic presented in Fig. 2.4(a) i.e., the capacitances are assumed to be linear. The analytical formulation (2.42)-(2.55) is a simplified version of model A that relies on the same schematic. The importance of nonlinear capacitances is shown in Fig. 2.19 were the relative error between model A and model B is given for different bias conditions. Very good agreement is obtained for HD_3 (and IMD_3, as it uses the same kernels) and for HD_2 when the bias is below the minimum in Fig. 2.17. The analytical formulation presented in Section 2.2.2.2 was also compared to model B in Fig. 2.20. The relative error of the formulation increases with the frequency. We calculate that it stay in the complete bias range below 3% (4%) at 2 GHz for HD_2 (HD_3) and below 20% at 10 GHz for both HD_2 and HD_3.

Numerical simulations allowed to analyze the relative impact of the different elements present in the equivalent circuit in Fig. 2.4 on the nonlinear factors of merit. At low frequency, the importance of both g_m and g_d has already been discussed

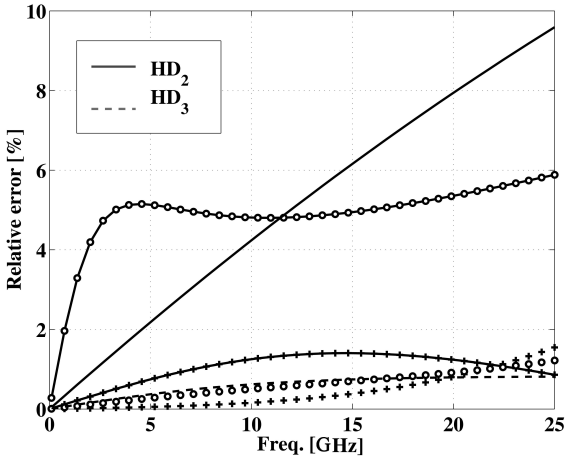


Figure 2.19: Relative error between model A and model B (reference = model B); $V_D = 1.2$ V and $A = 0.2$ V. Dashed lines refers to HD_3 and continuous lines to HD_2. Symbols refer to different gate voltages: $V_G = 1.4$ V no symbol, $V_G = 1.2$ V circles, $V_G = 1.0$ V crosses.

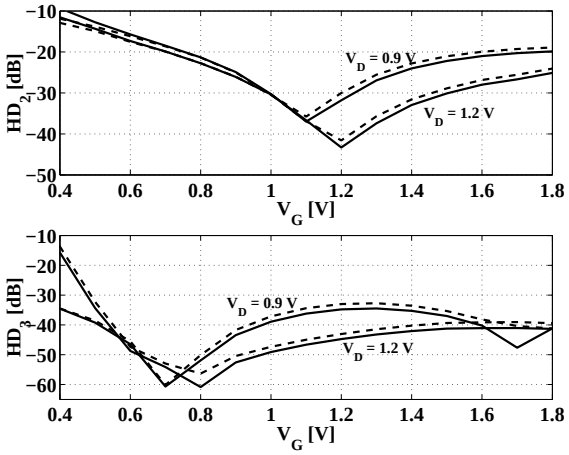


Figure 2.20: Proposed analytical expressions (straight lines) compared to model B (dashed lines); frequency = 10 GHz; $V_D = 1.2$ V; $A = 0.2$ V; $V_{th} = 0.5$ V.

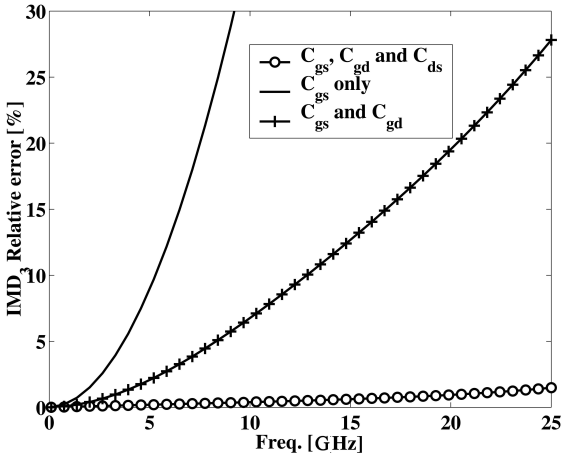


Figure 2.21: Relative error compared to model B due to the model simplifications; $V_G = 1.0$ V; $V_D = 1.2$ V; $A = 0.2$ V.

previously. If the drain current representation of (2.33) is adopted, we showed above that a good approximation of the frequency behavior may be obtained if only the linear capacitances are taken into account. However, if they are not included in the model, a 30% error relative to the complete model is obtained around 15 GHz for HD_2 and around 9 GHz for HD_3. The gate capacitance C_{gs} acts as a lowpass filter that decreases the amplitude of the input signal. Thus, a high value of C_{gs} leads to a lower distortion level but, as the distortion is the ratio between output and input signals, it is not much affected by C_{gs}. Furthermore, as the C_{gs} value of the DuT is in the order of several tens of femto-Farads, the pole created by this capacitance lies above the considered frequency band. The distortion is thus not much affected by C_{gs}, even around the threshold voltage. Simulations showed that C_{ds} also has a minor impact on the distortion factors of merit, as it does not change much with the bias. Furthermore, it contributes to lower the distortion level, as it filters the output admittance. Only C_{gd} has a strong influence on the HD because it models the feedback in the transistor. Numerical results confirm this discussion for linear capacitances. The relative error to model B is plotted in in Fig. 2.21 when only C_{gs} is present, when C_{gs} and C_{gd} are present, and when the three capacitances are taken into account.

The model highly relies on the extraction of the small-signal elements in Fig. 2.4. The procedure presented in [29] was used for that purpose. A sensitivity analysis was performed to see the impact of a wrong evaluation of the elements of the equivalent circuit. Numerical simulations showed that if these elements vary within 10% of their nominal value, the relative error is in any case inferior to 20%. The error is dominated by the drain current model of (2.33). Typically, depending on the bias

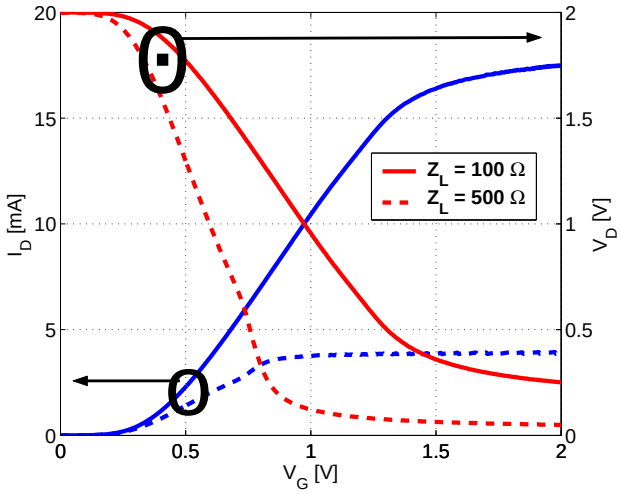


Figure 2.22: DC characteristics of the common-source PD device as a function of the gate bias given for several load impedances; $V_{dd} = 2$ V.

conditions, a maximum value of 10% error on HD_3 at 5 GHz is evaluated for a 10% variation of g_{di} and g_{mi}, g_{m1} and g_{d3} having the highest impact. The formulation is almost independent of a variation of C_{gs} and C_{ds} i.e., the error is smaller than 1%. However, the error due to C_{gd} is more important ($< 5\%$) as it is situated in the feedback loop.

2.3.4 Large signal analysis of a SOI nMOSFET

The IFM applied to the LSTF of SOI transistors will be analyzed in this section.

2.3.4.1 Low frequency case

From section 2.2.5, the DuTs large signal DC characteristics $I_D(V_G, Z_L)$ may be obtained as a function of the load impedance (Fig. 2.22). The distortion is computed following relations (2.7) and (2.69), and from the LSTF presented in Fig. 2.22. The results show that the saturation of the characteristics due to the limited supply voltage is well taken into account for the IFM applied on $I_D(V_G, Z_L)$ (Fig. 2.23). This method gives more accurate results compared to those from the Taylor approach, even if the effect of the output conductance is taken into account.

Impact of the load impedance In order to analyze the impact of the load, HD_3 is plotted in Fig. 2.24 as a function of V_G for different values of Z_L. First, note that following (2.69) and (2.70), the transconductance is the dominant nonlinear source at low load impedance levels, while for high values of Z_L, the output conductance nonlinearity dominates. Moreover, the curves exhibit several minima of HD_3. They

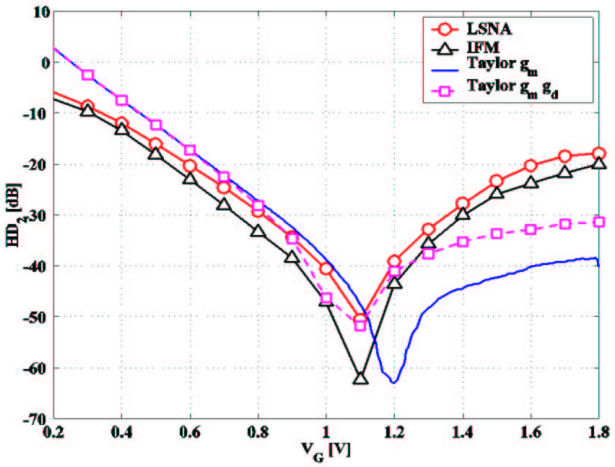


Figure 2.23: Comparison of the techniques used to extract HD_2. The drain voltage was set to 1.2 V for the curves obtained by Taylor series extraction, while the other ones follow the $I_D(V_G, Z_L)$ relationship with $Z_L = 50\ \Omega$; $A = 0.2$ V. Circles: LSNA measurements at 900 MHz; triangles: IFM on LSTF; squares: relation (2.7); straight line: relation (2.69).

correspond to the value of V_G that nullify G_{m3}. Fig. 2.25 shows the dependence of these minima on V_G and Z_L. The first minimum (close to $V_G = 0.3$ V) occurs at bias near pinch-off i.e., at the inflection point of G_{m1}, which is close to the inflection point of g_{m1} and almost independent on the load value. This sweet spot does not indicate a high linearity of the transfer function but its nearly quadratic shape. This bias seems attractive to the design of high-linear differential small-signal amplifiers with high efficiency (the g_m/I_D is high), but unfortunately the price to pay is a low transconductance (and thus gain) value. The second sweet spot appears at high V_G when the transistor enters in the triode mode. The third minimum may also be exploited for the design of high linear low-voltage circuits but tend to disappear at high input power level (Fig. 2.27).

Impact of the signal magnitude Typical input-output power plots present sweet-spots (Fig. 2.26) at given input signal amplitude. Simplified models as those given previously from a Taylor or Volterra series expansion do not include this feature. Indeed, a shift of the curves (on a log scale) is expected from the relations of section 2.2.2.2 for which only a few terms in (2.33) are taken into account. As IFM is however directly related to the magnitude of the signal, it does include the evolution of the sweet-spots as a function of the input level (Fig. 2.27).

On the other hand, the self-bias effect, as defined in page 20, is not taken into account with the analysis described in this section. As depicted in Fig. 2.28, this

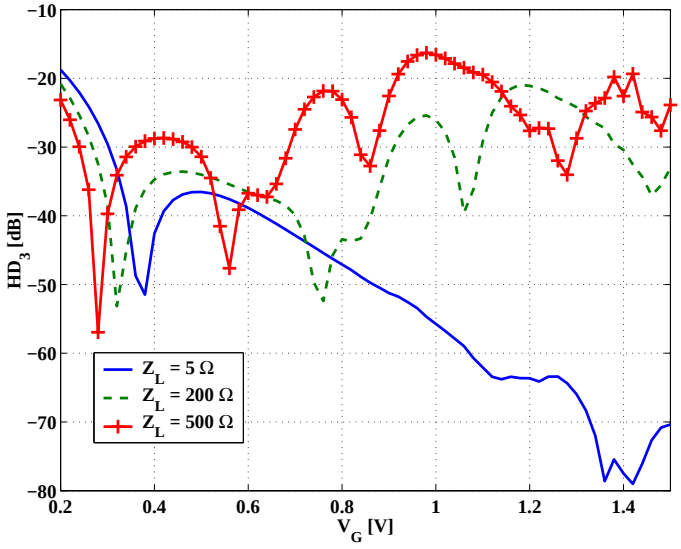


Figure 2.24: HD_3 as a function of the gate voltage for several load conditions extracted by IFM on LSTF; $V_{dd} = 2$ V; $A = 0.2$ V.

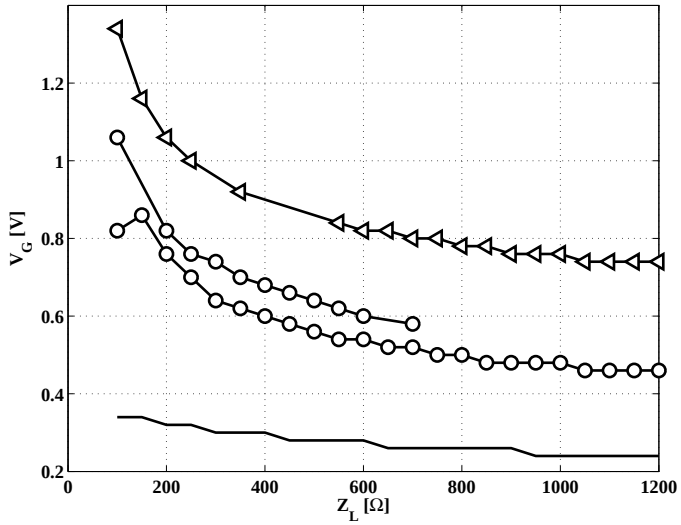


Figure 2.25: Locus of the minima of HD_3 in the (V_G, Z_L) plane extracted by IFM on LSTF; $A = 0.2$ V; $V_{dd} = 2$ V. Straight line: sweet-spots around pinch-off; circles: sweet-spots close to triode; triangles: sweet-spot related to the saturation of the LSTF at supply voltage.

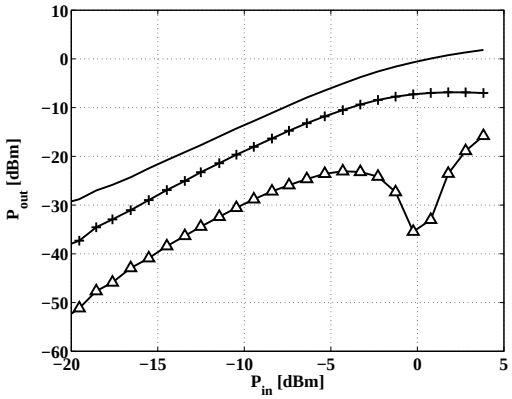


Figure 2.26: Measured output power versus input power at 900 MHz; straight line: fundamental, plus sign: harmonic two, triangles: harmonic three; $V_G = 0.4$ V; $V_D = 0.9$ V.

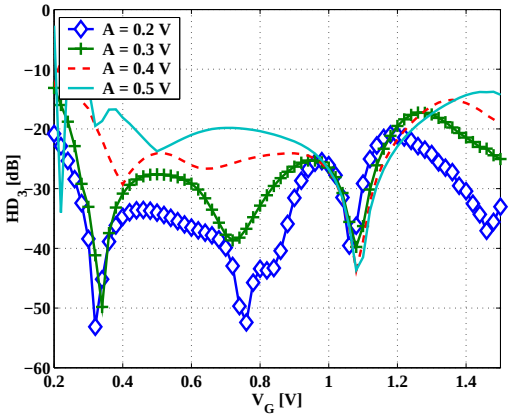


Figure 2.27: HD_3 as a function of the gate voltage for several input amplitudes extracted by IFM on LSTF; $Z_L = 200\ \Omega$; $V_{dd} = 2$ V.

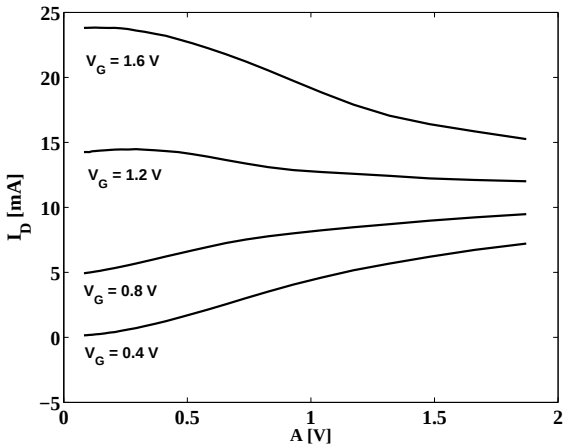


Figure 2.28: Self biasing effect: evolution of the drain bias current as a function of the excitation amplitude A; LSNA measurements at 900 MHz; $V_D = 0.9$ V.

effect may be important at high input levels.

2.3.4.2 High frequency case

The validity of the IFM method at HF is check by both measurements and numerical simulations. In Fig. 2.29 the HD_2 extracted from LSNA measurements is compared to IFM as given by (2.66). The time domain input waves needed for the IFM are provided by the LSNA measurements. A good agreement between simulation and measurements is observed.

To check the validity of the IFM applied to the frequency dependent LSTF, numeric simulations were also performed from a BSIM3v3 model with ELDO RF. For that purpose, another DuT is considered: a Fully-Depleted (FD) SOI nMOSFET

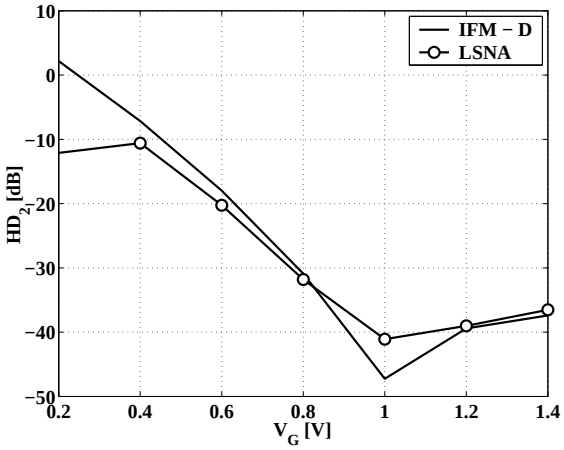


Figure 2.29: HD_2 of the PD MOSFET at 4.5 GHz from the IFM and the LSNA measurements; $V_D = 1.2$ V; $A = 0.15$ V.

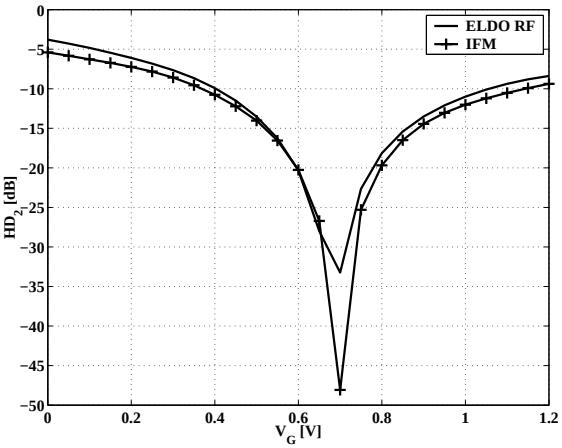


Figure 2.30: HD_2 of the FD MOSFET at 5 GHz from the IFM and the BSIM model; $Z_L = 200\ \Omega$; $A = 0.3$ V.

composed of 8 finger of 2 μm width and 0.14 μm long each, available from OKI corp. Fig. 2.30 shows the HD_2 obtained from a fast-Fourier transform of the time domain waves given by ELDO RF and the IFM. The method presented in section 2.2.4.2 is validated by numerical examples.

2.4 SOI devices comparison

The basic characteristic of SOI technology is the separation of the top active region from the underlying mechanical substrate by a thick insulator layer. Depending on the thickness of the top silicon layer, the film may be fully- or partially depleted. The nonlinear behavior of a PD SOI MOSFET has been analyzed up to now. FD MOSFETs present several advantages over PD ones [30]. However, a fair comparison of their nonlinear behavior in saturation has not been done previously and will be

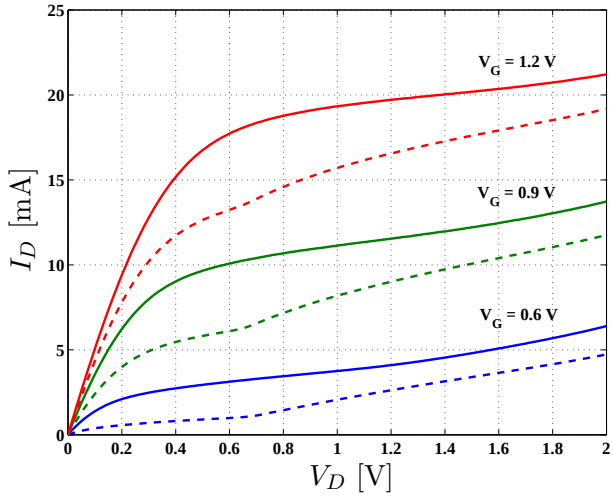


Figure 2.31: Current drain *vs* drain voltage of the considered PD (dashed line) and FD (plain line) SOI nMOSFET.

examined in this section. The comparison uses a FD device[2] of same dimensions as the PD one studied before i.e., 12x6.6 μm/0.25 μm.

2.4.1 Partially- and fully-depleted SOI transistors

Fig. 2.31 shows the $I_D(V_D)$ characteristics of both devices. Two main differences appear: first, the FD device does not present any kink, while the PD transistor does. Second, the gain of the FD is higher than the one of the PD transistor. The kink effect can be explained as follows [30]. When the drain voltage of a thick-film PD nMOSFET is high enough, electrons can acquire sufficient energy in the high electric field zone near the drain to create electron-hole pairs, due to an impact ionization mechanism. The generated electrons rapidly move into the channel and the drain, while the holes (which are the majority carriers in the p-type body) migrate towards the place of lowest potential i.e., the floating body. The injection of holes into the floating body forward biases the source-body diode. The body potential increases, giving rise to a decrease of the threshold voltage. This in turn induces an increase of the drain current. However, in thin-film FD devices, the source-to-body diode is already forward biased, due to the full depletion of the film. The holes can readily recombine in the source without having to raise the body potential.

It is well-known that this kink is frequency-dependent. Indeed, the high impedance is shunted through capacitances at moderate frequencies [31]. The output conductance of our PD DuT is reported as a function of the frequency in Fig. 2.32.

In order to compare the two devices, let us define the α_{FET} coefficient that

[2]LETI 0.25 μm technology.

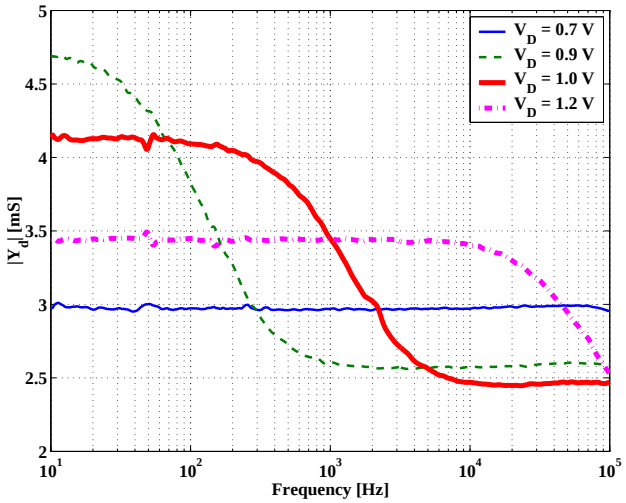


Figure 2.32: The output conductance of the PD device varies at low frequencies; $V_G = 1.0$ V.

measures the efficiency of the coupling between the gate potential and the potential of the channel of the MOSFET at the Si-SiO_2 interface. This coefficient can be expressed as a function of the total gate capacitance. The gate capacitance of the FD is composed of three series capacitances, namely the front gate oxide capacitance C_{ox1}, the capacitance of the silicon film C_{Si}, and the buried oxide capacitance C_{ox2}. The gate capacitance of the PD is the front gate oxide capacitance C_{ox1} connected in series to the depletion capacitance of the silicon film C_{depl}. The factor α_{FET} is then obtained as follows [30]:

$$\alpha_{FET,FD} = \frac{C_{ox2}C_{Si}}{C_{ox1}(C_{ox2}+C_{Si})} \tag{2.74}$$

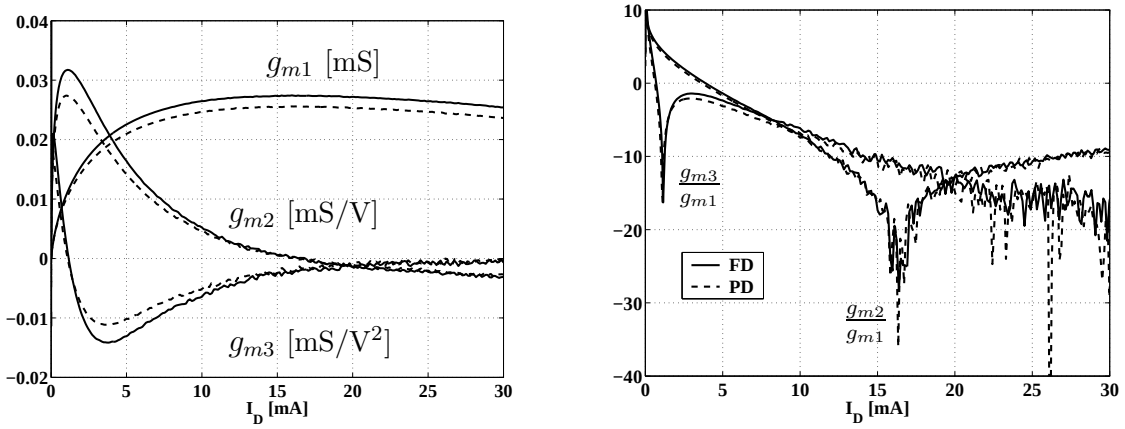
$$\alpha_{FET,PD} \simeq \frac{C_{depl}}{C_{ox1}} \tag{2.75}$$

where $\alpha_{FET,FD}$ ($\alpha_{FET,PD}$) is the α_{FET} coefficient of a FD (PD) SOI MOSFET. Better coupling is obtained when α_{FET} is small. For a FD MOSFET, α_{FET} is typically equal to zero, and to $0.3 \ldots 0.5$ for a PD MOSFET. It is worth noting that the value of α increases for small devices.

Ideally, the saturation current of long channels SOI MOSFETs is [32]

$$I_D = \frac{W}{L}\frac{\mu C_{ox1}}{2(1+\alpha)}(V_{gs}-V_{th})^2 \tag{2.76}$$

where W and L are the width and the length of the transistor, μ is the mobility, V_{gs} is the gate to source voltage, and V_{th} is the threshold voltage. From that equation,



(a) (b)

Figure 2.33: (a) Measured g_{mi} of a FD (straight lines) and a PD (dashed lines) transistors of same size (12x6.6 μm/0.25 μm) and technology. (b) ratio of the g_{mi} on a log scale. $V_D = 1.0$ V.

the g_{mi} are easily obtained:

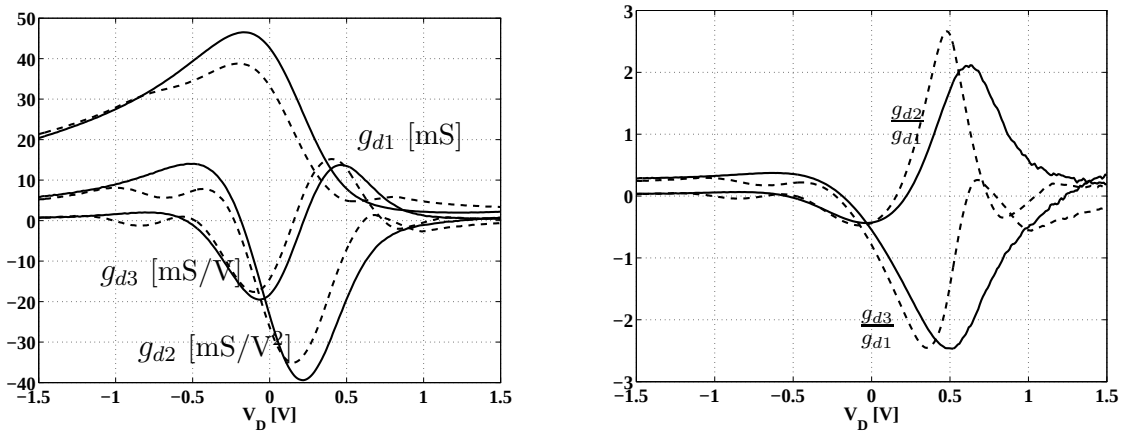
$$g_{m1} = \frac{W}{L}\frac{\mu C_{ox1}}{(1+\alpha)}(V_{gs} - V_{th}) \tag{2.77}$$

$$g_{m2} = \frac{W}{L}\frac{\mu C_{ox1}}{(1+\alpha)} \tag{2.78}$$

$$g_{m3} = 0 \tag{2.79}$$

From these ideal equations, it can be deduced that the FD transistor will have a higher current drive and transconductance than the PD transistor, with the same technological parameters, provided that the coefficient α is lower. However, harmonic distortion factors of merit will not be affected as they are defined from the *ratios* of the g_{mi}. This feature is depicted in Fig. 2.33 for our DuTs. Contrary to the ideal model, third-order distortion is generated because the mobility is strongly voltage dependent. Furthermore, the exponent is not exactly 2 due to short-channel effects. The ratio between $g_{m2,3}$ and g_{m1} is however not much affected by the technology. Because the HD is mainly determined by this ratio (see (2.70)), both devices have similar linearity. Nevertheless, it is expected from Fig. 2.33(b) that the PD has a small linearity improvement (within 1 dB) in comparison with FD.

Similarly, the linearity of the output conductance g_d may be analyzed through the g_{di} coefficients. They are plotted in Fig. 2.34 as function of the drain voltage. Because of the kink effect, the $g_{m2,3}$ of the PD transistor nullify were the FD does not. Sweet-spots are induced by this way. Nevertheless, as from Fig. 2.32, the kink effect disappears when the frequency increases, those biases are not interesting for improving the linearity in real circuit implementations.



(a) (b)

Figure 2.34: (a) Measured g_{di} of a FD (straight lines) and a PD (dashed lines) transistors of same size (12x6.6 μm/0.25 μm) and technology. (b) ratio of the g_{di} from (a). $V_G = 1.0$ V.

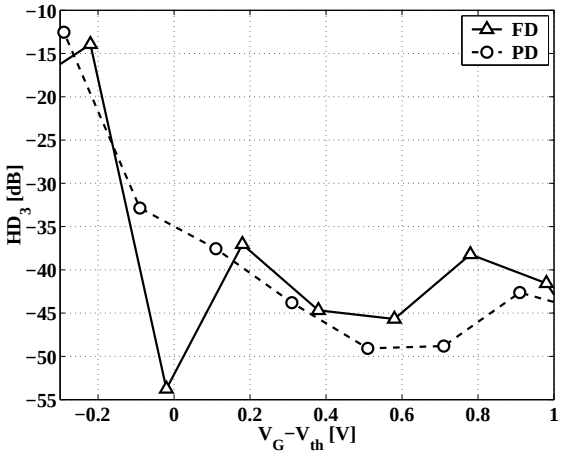


Figure 2.35: FD and PD devices exhibit same level of HD_3; $V_D = 1.2$ V, $A = 0.2$ V. LSNA measurements at 900 MHz.

2.4.2 Harmonic distortion

As expected from the previous comparison of g_{mi}, LSNA measurements in Fig. 2.35 show a comparable level of distortion for both devices. The minimum of the curve for the PD device is not visible due to a lack of experimental points, but LF analysis assesses its presence. The devices in CS configuration behave however slightly differently as a function of the load impedance. Indeed, Fig. 2.36 shows the minima of HD_3 calculated from the LF LSTF of each device. For a high load value, the influence of the output conductance nonlinearity gets more importance and the behavior of the devices differs. It is worth noting that at HF when the device does not experience the kink effect anymore, the difference between the devices decreases.

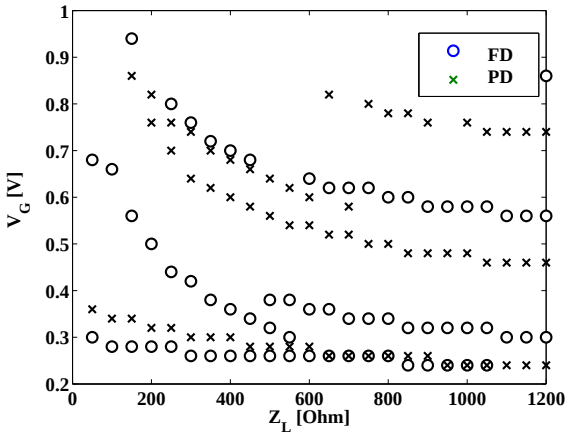


Figure 2.36: Locus of the sweet-spots in the load impedance-gate voltage plane; $A = 0.2$ V.

2.4.3 Intermodulation distortion

In the previous section, it was shown that HD was not much affected by the type of device considered. This is due to the similar shape of drain current, which is the main source of nonlinearity, at HF were the kink vanishes. However, we expect a IMD difference between the devices even at HF from the previous Volterra-based model. Indeed, equation (2.55) shows that IMD_3 is not only dependent on $G_o(\omega)$ but also on $G_o(\Delta\omega)$. The G_o dependence on frequency was depicted in Fig. 2.32.

Two-tones tests were performed by LSNA and compared to simulations obtained by the Volterra-based model (Fig. 2.37). It appears that, for bias points where the $I - V$ exhibits a kink ($V_D = 0.8$ V for our DuT), IMD_3 changes with the tone separation. Even if the model describes this only qualitatively, the phenomenon is more important when the impedance load is higher. Measurements on FD devices showed that IMD_3 is independent on the tone separation, as expected from the model.

Perspectives: other SOI devices We provided a complete description of both PD and FD devices. It was shown that only the IMD differs from one device to another because of relation (2.55). It is worth noting that this relation also depends on $G_o(2\omega)$. This means that a variation of IMD_3 versus the tone separation is expected at a certain frequency ω for devices whose output conductance changes at twice this frequency. In reality, only few devices present a kink at HF, but there are. The dynamic-threshold MOSFET (DTMOS), deeply studied in [29] shows such kink in g_m and g_d around 900 MHz. The linearity around 900 MHz, but also around 450 MHz may then be affected. The study of the linearity of these new SOI devices is however beyond the scope of this work.

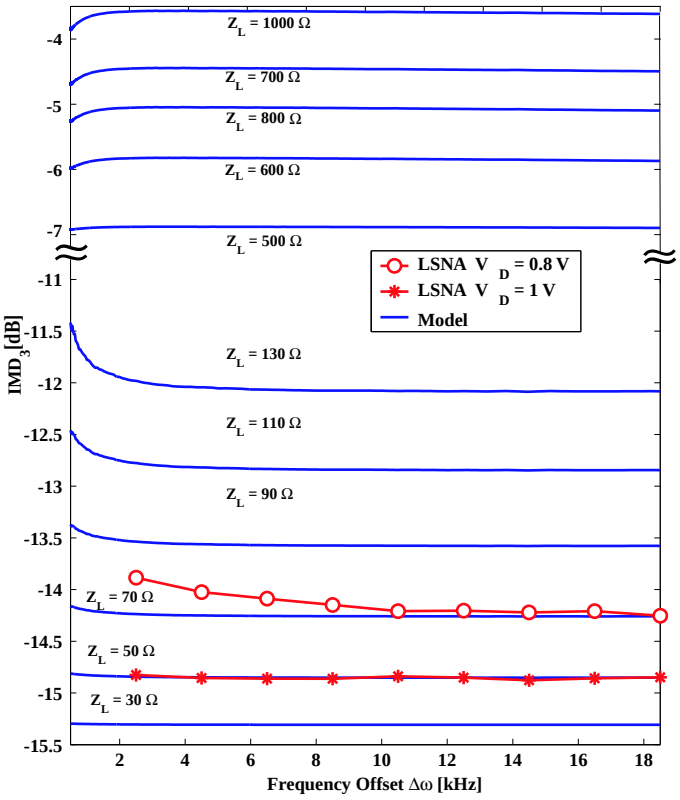


Figure 2.37: IMD_3 *vs* tone separation; PD MOSFET; LSNA measurements were performed on a 50 Ω set-up at 900 MHz. $V_G = 1$ V, $A = 0.6$ V. The drain bias is set to $V_D = 1.0$ V in the simulations.

2.5 Linearity analysis of a low-noise-amplifier

Since the LNA is the first gain stage in the receive path, it determines the sensitivity of the whole receiver. As consequence, noise is the major concern in the design of such stage. If the noise figure of the LNA may be minimized by appropriated choice of source impedance, this impedance does not necessary -and most of the time does not- correspond to the impedance required for power matching, which is typically 50 Ω. This is the first trade-off in the design of LNAs. Engineers have fortunately found design strategies to face this [33, 34]. Typical specifications on LNAs are set to the gain, the power consumption, the noise and power matching, the stability, and also the linearity (IP_3), as spurious should be avoided as much as it could be. The present section is dedicated to the linearity analysis of a typical CMOS LNA.

The LNA architecture of Fig. 2.38 is widely adopted as the source degeneration offers a real input impedance Z_{in} at resonance frequency $\omega_o = [(L_g + L_s)C_{gs}]^{-0.5}$:

$$Z_{in} \simeq \frac{g_m L_s}{C_{gs}} \tag{2.80}$$

The value of the inductances at the gate and at the source of the transistor (L_g and L_s, respectively) are tuned to satisfy the power matching condition at resonance, $Z_{in} = R_s = 50\ \Omega$ (R_s is the source impedance typically set to 50 Ω). High isolation between output and input is provided by the cascode stage. In the following, the linearity of these two subsystems will be analyzed separately. Then, the IMD of the whole amplifier is obtained by their combination. The Volterra expressions for the cascade combination of nonlinear systems is given in [2]. If the second-order effects are neglected, the IMD_3 of a cascaded system is dominated by the last stage because of the amplification provided by the first stages.

Because feedback is present in most circuit stages, its influence on linearity will be first introduced with the help of the Volterra formalism.

2.5.1 Nonlinear feedback systems

Let us consider the nonlinear negative feedback system in Fig. 2.39 where the feedback operator $F(s)$ is linear. The nonlinear operator H is described by a third-order Volterra series. Then, the first order transfer function of the whole system i.e., the linearized system, is $Q_1(s_1) = H_1(s_1)R(s_1)$ with $R(s) = (1 + T(s))^{-1}$ and $T(s) = H_1(s)F(s)$, the loop gain. In the usual case of high loop gain, $R(s)$ may be regarded as a gain reduction factor. The expression for the second- and third-order

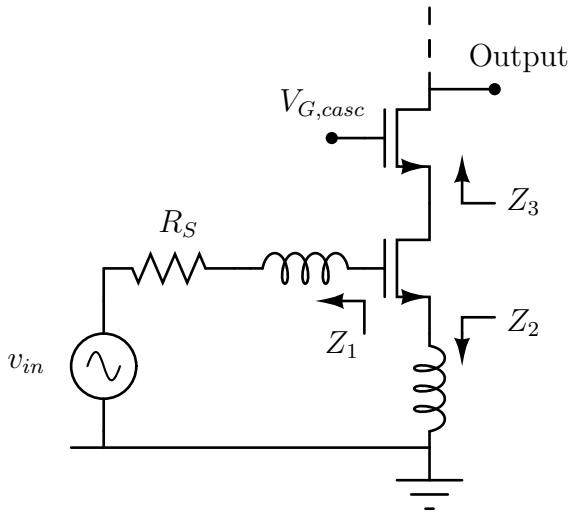


Figure 2.38: Studied LNA architecture.

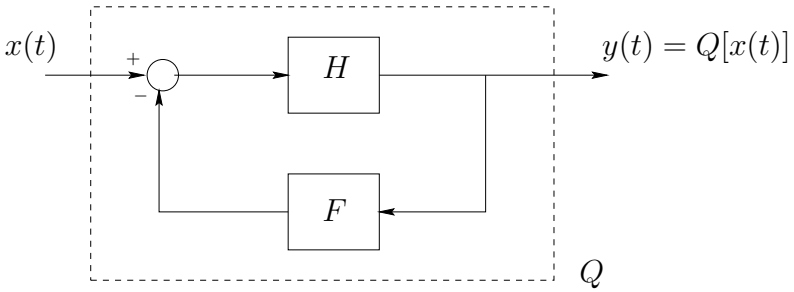


Figure 2.39: A nonlinear feedback system.

transfer functions are [2]:

$$Q_2(s_1, s_2) = R(s_1)R(s_2)R(s_1 + s_2)H_2(s_1, s_2) \quad (2.81)$$

$$Q_3(s_1, s_2, s_3) = R(s_1)R(s_2)R(s_3)R(s_1 + s_2 + s_3)[H_3(s_1, s_2, s_3) - 2H_2(s_1, s_2)F(s_1 + s_2)R(s_1 + s_2)H_2(s_3, s_1 + s_2)] \quad (2.82)$$

The feedback reduces the second-order transfer function by a factor $R(s_1)R(s_2)R(s_1+s_2)$. If, however, the reduction of the linear gain is compensated by an increase of the input level with a factor $R(s)$, then the second-order transfer function is only reduced by a factor $R(s_1 + s_2)$. Similar comment applies for the third-order transfer function. Furthermore, because of the minus sign in the product under brackets in (2.82), $Q_3(s_1, s_2, s_3)$ can be made zero by adjusting the loop gain such that the contribution of the second-order nonlinearity cancels with the contribution of the third-order nonlinearity. On the other hand, when a sweet-spot is present in the open-loop configuration, $H_3(s_1, s_2, s_3) = 0$, but due to the second-order contribution in the feedback system, Q_3 is in general different from zero.

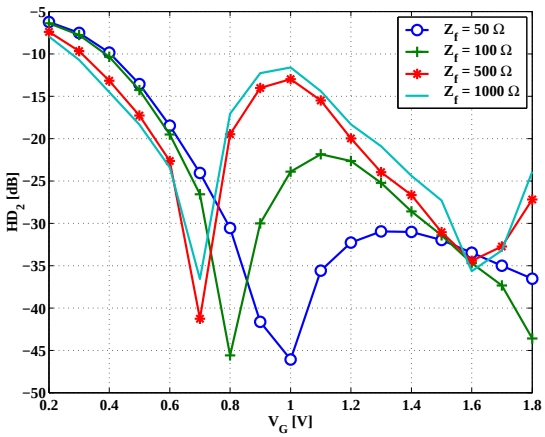


Figure 2.40: Impact of Z_f on HD_2 extracted by IFM; FD device, $A = 0.2$ V, $Z_L = 200\ \Omega$.

IFM may also be used to characterize the HD of feedback systems. For instance, if Z_f is the feedback impedance connected between the gate and the drain of the MOSFET, we can write the following relationship:

$$(V_{dd} - V_d)Y_L = I_D(V_G, V_D) + Y_f(V_d - V_g) \tag{2.83}$$

Using (2.83) instead of (2.67), the LSTF of the circuit with feedback is obtained. Fig. 2.40 shows the linearity improvement of such feedback system.

In the following, both LNA stages, the source-degenerated and the cascode, will be analyzed separately.

2.5.2 *Linearity of a source-degenerated MOSFET*

The source-degenerated configuration is an example of series-series feedback. A simplified signal flow graph is presented in Fig. 2.41. The loop gain is:

$$T(\omega) = \frac{1}{j\omega C_{gs}} \cdot g_m \cdot \frac{j\omega L_s}{R_s + j\omega L_g} \tag{2.84}$$

where it was assumed that $Z_1(j\omega) = R_s + j\omega L_g$ and $Z_2(j\omega) = j\omega L_s$. Because of relation (2.80), the loop gain of the first stage of the LNA is equal to unity at ω_o, $T(j\omega_o) = 1$. The term under bracket in (2.82) will then get more importance at resonance. The sweet-spots are then intended to disappear even if the linearity is in general improved.

In [35], Aparin gave the Volterra formulation of the third order intermodulation product of the first stage of the circuit of Fig. 2.38 when the output conductance is

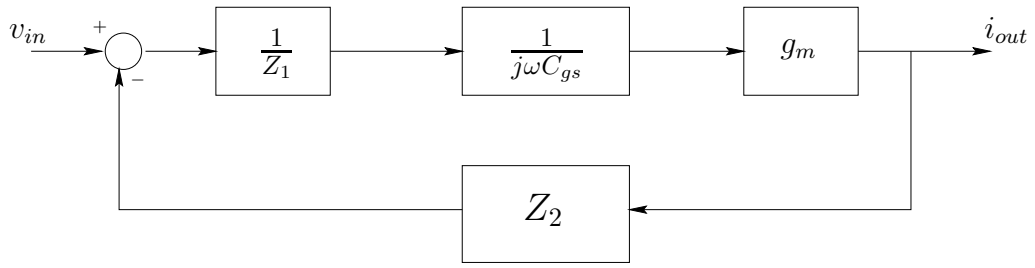


Figure 2.41: Signal flow graph of a source-degenerated stage.

supposed to be linear and $C_{gd} = 0$:

$$IMD_3(2\omega_2 - \omega_1) = \frac{3}{4}|H(j\omega)||A_1(j\omega)|^3|\epsilon(\Delta\omega, 2\omega)|A^2 \quad (2.85)$$

$$H(j\omega) = \frac{1 + j\omega C_{gs}(Z_1 + Z_2) + j\omega C_{gd}Z_1}{g_{m1} - j\omega C_{gd}(1 + Z_2(g_{m1} + j\omega C_{gd}))} \quad (2.86)$$

$$A_1(j\omega) = \frac{1 + j\omega C_{gd}Z_3}{(g(j\omega) + g_{m1})Z_x} \quad (2.87)$$

$$Z_x(j\omega) = Z_2 + j\omega C_{gd}(Z_1Z_2 + Z_1Z_3 + Z_2Z_3) \quad (2.88)$$

$$g(j\omega) = \frac{1}{Z_x}\left(1 + j\omega C_{gd}(Z_1 + Z_3) + j\omega C_{gs}(Z_1 + Z_x)\right) \quad (2.89)$$

$$\epsilon(\Delta\omega, 2\omega) = g_{m3} - \frac{2g_{m2}^2}{3}\left(\frac{2}{g_{m1} + g(\Delta\omega)} + \frac{1}{g_{m1} + g(2\omega)}\right) \quad (2.90)$$

It appears from that expression that a small C_{gs} and high g_{m1} are desirable for high linearity. Also from (2.85), linearization is possible by choosing adapted out-of band terminations [35, 36]. Fig. 2.42 shows the IMD_3 of a FD SOI transistor computed with (2.85). As expected from the feedback theory, the linearity increases with the inductance L_s value, while the sweet-spot disappears.

At resonance, the relations simplify and best IMD_3 is found when the factor $\epsilon(\Delta\omega, 2\omega)$ is minimized. This is only possible if $g_{m3} > 0$ which is only the case when the FET is biased the moderate inversion. Despite of its high efficiency (g_m/I_D), this region of operation has not been used up to now for RF circuit because of the poor value of f_T and g_m. As the MOS technology goes in deep-submicron, the active devices exhibit enough performances for few-GHz low-power circuits in moderate and weak inversion regions [37]. Therefore, it is a good idea to bias low-power circuits in moderate inversion, even at RF.

2.5.3 Linearity of a cascode stage

The cascode transistor (Fig. 2.43) is often used in amplifiers to increase the reverse isolation of a CS stage and to improve stability. Ideally, the cascode transistor passes

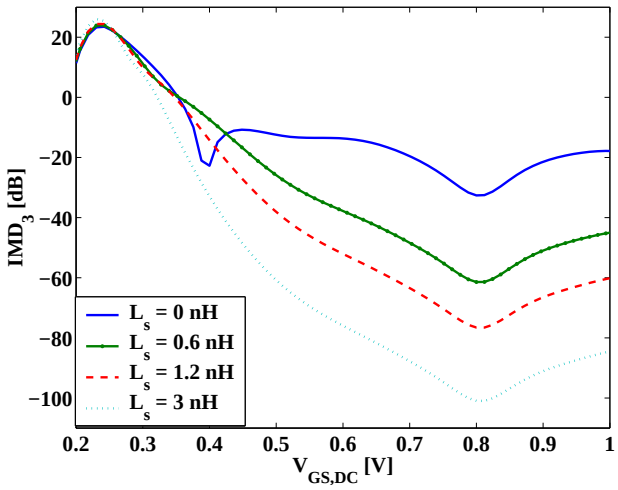


Figure 2.42: IMD_3 of a source degenerated stage computed by (2.85) as a function of the source inductance. FD SOI nMOSFET 80x2 μm/0.16 μm; $A = 0.2$ V; $V_D = 0.6$ V;$L_g = 6.7$ nH; $f = 2.45$ GHz; $\Delta f = 10$ kHz; $Z_L = 50\ \Omega$.

the current of the input transistor (i_{in}) to the load Z_L with a unity gain. Thus, no distortion is expected. In reality, both the body effect and the output conductance deteriorate the linearity. In SOI technology, the body effect (or, more accurately, the back-gate effect) is small and it will be neglected in order to simplify the analysis. The distortion of the common-gate transistor is thus similar to the one of the CS.

The Volterra analysis of the circuit in Fig. 2.43 by the method of nonlinear current leads to the following expression for IMD_3:

$$IMD_{3,cascode} \simeq \frac{A^2}{4}\left|\frac{Y_1(\omega)(Y_2(\omega)+Y_3(\omega))^2(Y_2^*(\omega)+Y_3^*(\omega))\cdot N_1(2\omega,\Delta\omega)}{(g_{m1}+Y_3(\omega))D(\omega)^2D(2\omega)D^*(\omega)D^*(\Delta\omega)}\right| \tag{2.91}$$

$$\begin{aligned} N_1(2\omega,\Delta\omega) &= 2g_{m2}^2(Y_2(2\omega)D^*(\Delta\omega)+2Y_2^*(\Delta\omega)D(2\omega)) \\ &\quad +3g_{m3}D^*(\Delta\omega)D(2\omega) \end{aligned} \tag{2.92}$$

$$D(\omega) = g_{m1}Y_2(\omega)+Y_T(\omega) \tag{2.93}$$

$$Y_T(\omega) = Y_1(\omega)Y_2(\omega)+Y_1(\omega)Y_3(\omega)+Y_2(\omega)Y_3(\omega) \tag{2.94}$$

with $Y_1(\omega) = j\omega C_{gs}+G_T$, $Y_2(\omega) = j\omega C_{gd}+Y_L$, and $Y_3(\omega) = j\omega C_{ds}+g_{d1}$ and A the amplitude of the input current i_{in}. It was assumed that the output conductance g_d and the capacitors were linear. Fig. 2.44 gives the IMD_3 of the different configurations under the same hypothesis ($g_{d2} = g_{d3} = C_{gd} = 0$), using the previously derived Volterra-based formulation.

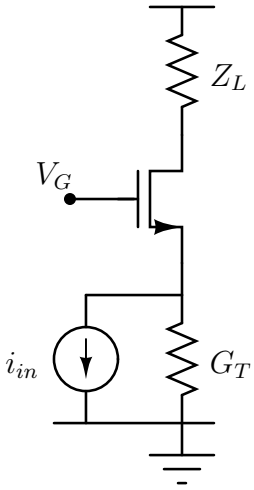


Figure 2.43: Cascode topology.

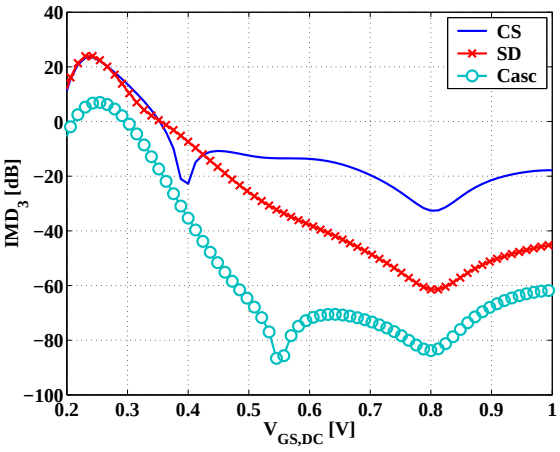


Figure 2.44: Comparison of IMD_3 for different MOS configurations (Volterra model). FD SOI nMOSFET 80x2 μm/0.16 μm; $A = 0.2$ V; $V_D = 0.6$ V; $L_s = 0.6$ nH; $R_s = 50\ \Omega$.

2.5.4 Linearity of a narrowband LNA

In the case of the circuit of Fig. 2.38, as the gain of the first stage should be higher than one, we may expect that the second stage dominates the linearity behavior. This feature is interesting, as linearity and noise optimizations may be decoupled [38]. However, from Fig. 2.44, because of the different order of magnitude involved, and because second order terms also affect the third order response, we cannot conclude directly that the linearity of the LNA will be determined by the second stage only.

The circuit was realized on the 0.14 μm FD SOI technology from OKI. The gate inductance L_g was set to 6.7 nH and the source inductance L_s to 0.6 nH in order to offer a real 50 Ω input impedance at 2.45 GHz. Transistors of same sizes were used in both stages. The transistors are biased in the moderate inversion in order to improve the g_m/I_D ratio and thus lower the current consumption. A large W/L ratio is thus used, namely the transistors are composed of 80 fingers, 2 μm width and 0.16 μm length each. The threshold voltage is 0.32 V. The supply voltage is set to 1.2 V and the gate bias of the cascode is 0.9 V. The frequency offset between the tones considered here is 20 kHz. If not mentioned, these settings are used in this chapter.

LSNA measurements The measured IMD_3 and IP_3 are shown respectively in Fig. 2.45 and Fig. 2.46. Let us compare the LSNA measurements of the circuit with those of the CS MOSFET used in this circuit in Fig. 2.47. It is seen that the level of distortion are about the same. Furthermore, the sweet-spot of the NMOS curve around the threshold voltage is less marked for low input levels, as expected from the source-degeneration analysis. Because of the self-bias, this sweet-spot tend disappear at high input levels.

Input Impedance Measurements showed that the input impedance varies with the magnitude of the signal (Fig. 2.48). This property is not really meaningful for a LNA because of the low level involved in the receive path, but the degenerated source architecture is also used in the design of power amplifiers. Note that the describing function theory explained in the section 2.2.3 may also be used to analyze this feature from the $V - I$ characteristic.

Integral Function Method The $I - V$ characteristic of the LNA was measured at DC in order to apply IFM. As we showed before, the linearity is indeed mainly dependent on the LF current characteristic. At low frequency, the inductances are

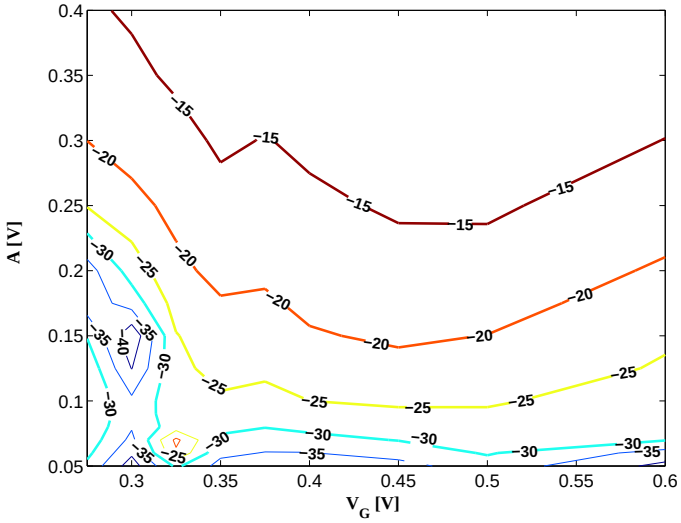


Figure 2.45: Measured LNA IMD_3 as a function of the input bias V_G and the amplitude A of the input signal.

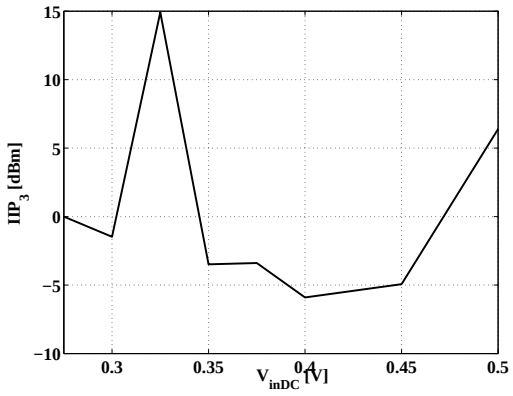


Figure 2.46: Measured IP_{3i} of the LNA; $V_{dd} = 1.2$ V; $V_G = 0.9$ V; $V_{th} = 0.32$ V.

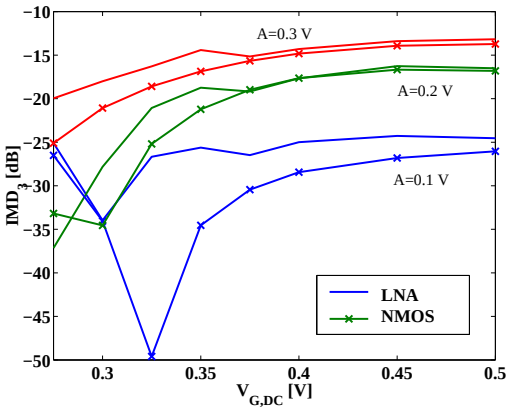
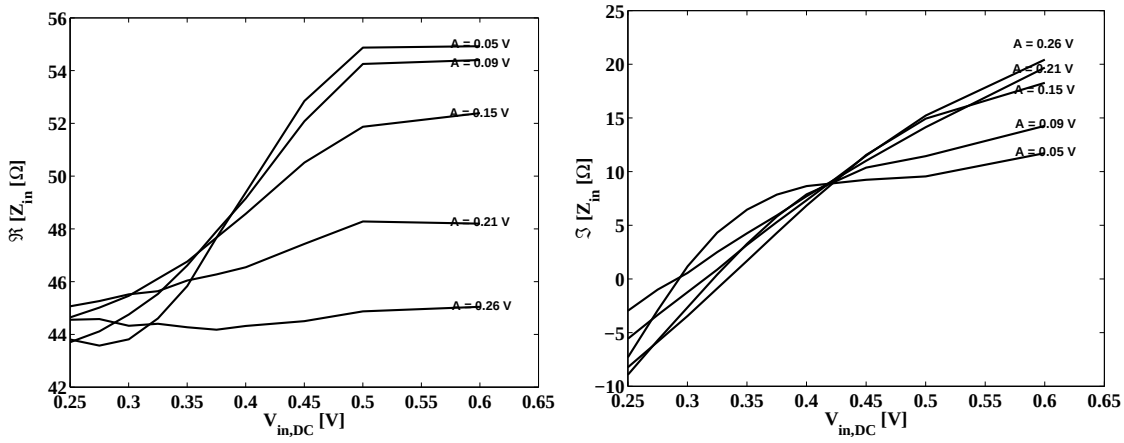


Figure 2.47: Comparison of IMD_3 of the LNA and the NMOS used in this circuit. LNA: $V_{dd} = 1.2$ V and NMOS: $V_D = 0.6$ V.



(a) (b)

Figure 2.48: Input impedance of the LNA as a function of the input bias V_G and the amplitude A of the input signal (LSNA measurements); $f = 2.45$ GHz; $V_{dd} = 1.2$ V; $V_{G,casc} = 0.9$ V. (a) real part; (b) imaginary part.

shorted and the effect of source degeneration is not visible. Furthermore, even if the weak nonlinearity assumption holds so that we have $IMD_3 = 3 \cdot HD_3$, the variation of IMD with the tone separation explained previously is not taken into account in DC methods. Nevertheless, Fig. 2.49 compares results from IFM with the measurements. The sweet-spot around the threshold is well present in the IFM curve as the feedback process is not taken into account. In the measurements, the sweet-spots migrate towards lower V_G value for which the second and third order compensate as suggested in (2.90).

We showed in section 2.2.4.2 that the IFM may be used when the time domain input and output waves are known. In practice, this technique is not used because only LSNA measurements or accurate simulations provide these requirements.

Discussion We provided linearity analysis of both stages of the LNA. Depending on the gain of the first stage and thus on the level of the signal present at the input of the second stage, the linearity will be determined by the first or the second stage. Indeed, when attacked by a current, the cascode transistor shows a high degree of linearity. Simplified formulas were given, allowing a separate optimization of each stage. In principle, it should then be possible to include linearity specification in designs flows such as presented in [33, 34]. However, as it will be discussed now, the design process should hold in all regimes of operation, from weak to strong inversion.

Indeed, the previous analysis showed that the moderate inversion is interesting for the design of LNA. As the dimensions shrink to deep-submicron, we are no more restricted to low-frequency applications in that regime. Moreover, the f_T of the

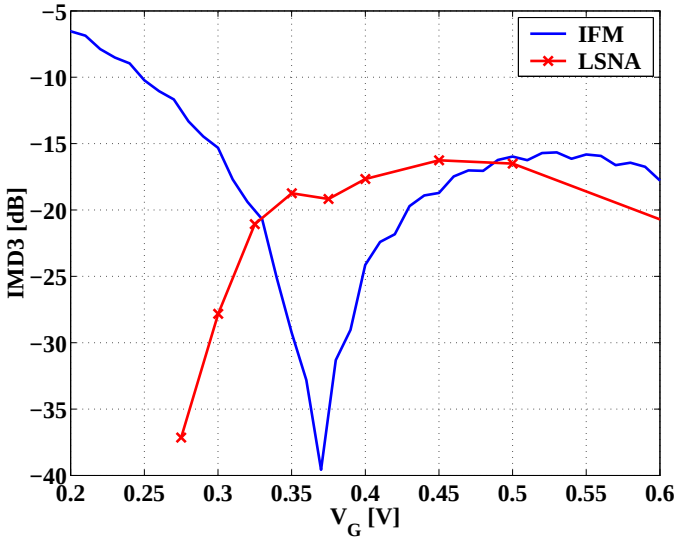


Figure 2.49: IMD_3 evaluated by IFM compared to LSNA measurements. $A = 0.2$ V; $V_{dd} = 1.2$ V; $V_{G,casc} = 0.9$ V.

device, while proportional to the inverse of the length L of the device in velocity saturation, is proportional to L^{-2} in moderate inversion. Secondly, the efficiency (that may be measured by the g_m/I_D ratio) of the device is increased. Thirdly, the present analysis on linearity showed a sweet-spot in this regime. Even if this linearity improvement point is less effective due to feedback (see (2.82) with $H_3 = 0$), the only possibility for getting a sweet-spot is to bias the transistor in moderate inversion (see (2.90) and (2.92)). In conclusion, it is, when permitted, advantageous to bias the transistors in moderate inversion for low-power, low voltage operation.

2.6 Conclusions

Nonlinear distortion is key in the design of modern telecommunication circuits. On one hand, intermodulation distortion must be known for controlling the spurious generation. On the other hand, the energy transfer between the fundamental and the harmonics is important in most applications, requiring the harmonic distortion characterization.

In this chapter, several methods for determining the nonlinearity of a system are presented. From an analytical point-of-view, a Taylor series describes a frequency independent nonlinearity while the analysis of frequency dependent nonlinearities requires a Volterra series. A third order series is used to describe weakly nonlinear systems. For strong nonlinearities or large excitations, this representation fails to describe accurately the system and the IFM may be used. Initially dedicated to memoryless systems, we provide in this chapter the extension of this method to

frequency dependent systems. Since IFM depends on the signal amplitude, it has the further advantage to avoid some numerical issues that power series approaches may present.

These methods were used to characterize the nonlinearity of sub-micron SOI nMOSFETs. At low frequency, the $I-V$ characteristic is sufficient to compute the nonlinear factors of merit. Since the $I-V$ relation depends on both V_G and V_D, both the gate transconductance and the output conductance affect the distortion. Furthermore, the distortion of the device decreases for low load impedance, as it shunts the output conductance nonlinearity. It was demonstrated that the large signal transfer function of the device permits to solve a one-dimensional problem that takes the influence of the transconductance, the output conductance and the load impedance into account. In conjunction with IFM, the sweet-spots are easily predicted.

The results were confronted to state-of-the art LSNA measurements. The measurements indicated that the MOSFET HD is dominated by the nonlinearity of the DC $I-V$ characteristics up to several giga-Hertz. A Volterra-based model of the transistor confirmed analytically this property. The model allows determining the frequency up to which the capacitances affect the main factors of merit. The simplicity of the proposed modeling presents the advantages that, without losing accuracy, the simulation time is low and it gives insight in the nonlinear behavior of the MOSFET. A linear representation of the capacitances is sufficient to describe properly the behavior of the transistor up to several GHz. Compared to the other capacitances, C_{gd} has relatively more influence since it is located in the feedback loop. The cross-derivative terms could be easily introduced in the model in order to improve the accuracy and enlarge the validity range to the triode region.

The choice the method will depend on the aimed application, depending on the frequency range, the nonlinearity degree and the complexity. We tried through this work to use the simplest method as much as we could. In particular, we demonstrated that low frequency characterization (Taylor or IFM) is accurate enough for low-GHz telecommunication applications.

From the performances point-of-view, the linearity of FD and PD SOI transistors was compared. It was showed that these devices present similar HD level in saturation, while due to the kink effect, the IMD of PD transistors depends on the tone separation.

The linearity of a common-source transistor with resistive load was widely examined in this chapter. In CMOS circuits however, loads are generally capacitive.

The introduction of any reactance in the computation of the HF large signal transfer function and in the Volterra model is straightforward. In the case of memoryless approaches (Taylor or IFM DC), only the amplitude -and not the phase- of the complex load may be included in the analysis.

Due to the generality of the methods used in this work, they also apply to other transistor configurations than common-source. The Volterra formalism was used to determine analytically the IMD_3 of a cascode and a source-degenerated stages. These stages are the basic building blocks of typical narrowband LNAs. The linearity analysis of a LNA showed that the only possibility for a sweet-spot is in the moderate inversion regime. Because of the high efficiency of the FETs in this regime, and since the linearity is not that much degraded, the moderate inversion regime has to be used for the design of low-GHz LNA in sub-micron CMOS.

References

[1] Y.P. Tsividis, *Operation and Modeling of the MOS Transistor*, McGraw-Hill, second edition, 1999.

[2] Piet Wambacq and Willy Sansen, *Distortion analysis of analog integrated circuits*, Kluwer, Dordrecht, 1998.

[3] R. van Langevelde, L.F. Tiemeijer, R.J. Havens, M.J. Knitel, R.F.M. Ores, P.H. Woerlee, and D.B.M. Klaassen, "RF-distortion in deep-submicron CMOS technologies," *IEEE International Electron Device Meeting*, pp. 807–810, 2000.

[4] L.F. Tiemeijer, R. van Langevelde, and O. Gaillard, "RF distortion characterization of sub-micron CMOS," *Proc. ESSDERC*, pp. 464–467, 2000.

[5] A. Cerdeira, M. Estrada, R. Quintero, D. Flandre, A. Ortiz-Conde, and Garcia Sanchez F.J., "New method for determination of harmonic distortion in SOI FD transistors," *Solid-State Electronics*, vol. 46, pp. 103–106, 2002.

[6] A. Cerdeira, M. Alemán, M. Estrada, and D. Flandre, "Integral function method for determination of nonlinear harmonic distortion," *Solid-State Electronics*, vol. 48, pp. 2225–2234, 2004.

[7] Y. Rolain, W. Van Moer, G. Vandersteen, and J. Schoukens, "Why are nonlinear microwave systems measurements so involved?," *IEEE Transactions on Instrumentation and Measurement*, vol. 53, no. 3, pp. 726 – 729, June 2004.

[8] J. Verspecht, P. Debie, A. Barel, and L. Martens, "Accurate on wafer measurement of phase and amplitude of the spectral components of incident and scattered voltage waves at the signal ports of a nonlinear microwave device," *IEEE International Microwave Symposium Digest*, September 8-11 1995.

[9] Wendy Van Moer, *Development of New Measuring and Modelling Techniques for RFICs and their Nonlinear Behaviour*, Ph.D. thesis, Vrije Universiteit Brussel, Brussels, Belgium, June 2001.

[10] Willy Sansen, "Distortion in elementary transistor circuits," *IEEE Transactions on Circuits and Systems—Part II: Analog and Digital Signal Processing*, vol. 46, no. 3, pp. 315–325, March 1999.

[11] Julian J. Bussgang, Leonard Ehrman, and James W. Graham, "Analysis of nonlinear systems with multiple inputs," *Proceedings of the IEEE*, vol. 62, no. 8, pp. 1088–1119, August 1974.

[12] J. Brinkhoff and A.E. Parker, "Effect of baseband impedance on FET intermodulation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 3, pp. 1045 – 1051, March 2004.

[13] S.A. Maas, *Nonlinear Microwave Circuits*, Artech House, Norwood, MA, 1988.

[14] Sanghoon Kang, Byounggi Choi, and Bumman Kim, "Linearity analysis of CMOS for RF application," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 3, pp. 972–977, March 2003.

[15] B. Parvais and J.-P. Raskin, "Analytical expressions for distortion of SOI MOSFETs using the Volterra series," *European Gallium Arsenide and other Compound Semiconductors Application Symposium 2004*, October 11-15 2004, Amsterdam, the Netherlands.

[16] Alberto O. Adan, Toshihiko Yoshimasu, Shoichi Shitara, Noriyuki Tanba, and Masayuki Fukumi, "Linearity and low-noise performance of SOI MOSFETs for RF applications," *IEEE Transactions on Electron Devices*, vol. 49, no. 5, pp. 881–888, May 2002.

[17] Arthur Gelb and Wallace E. Vander Velde, *Multiple-input describing functions and nonlinear system design*, McGraw Hill, New York, 1968.

[18] A. Cerdeira, M. A. Aleman, M. Estrada, D. Flandre, B. Parvais, and G. Picun, "The integral function method: A new method to determine the non-linear harmonic distortion," *SBMicro 2003, Sao Paulo, Brazil*.

[19] A. Cerdeira and M. Estrada, "Mathematical basis of the expressions used by the integral function method for the determination of nonlinear harmonic distortion in devices and circuits," *Proc. ICSICT 2004*, October 2004, Bejing, China.

[20] Nuno Borges de Carvalho and José Carlos Pedro, "Large- and small-signal IMD behavior of microwave power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, no. 12, pp. 2364–2374, December 1999.

[21] Emigdio Malaver, José Angel Garcia, Antonio Tazon, and Angel Mediavilla, "Characterizing the linearity sweet-spot evolution in FET devices," *Proc. European Gallium Arsenide and other Compound Semiconductors Application Symposium 2003*.

[22] Bill Toole, Calvin Plett, and Mark Cloutier, "RF circuit implications of moderate inversion enhances linear region in MOSFETs," *IEEE Transactions on Circuits and Systems—Part I: Fundamental Theory and Applications*, vol. 51, no. 2, pp. 319 – 328, February 2004.

[23] Dominique Schreurs, *Measurement based modeling of heterojunction field-effect devices for non-linear microwave circuit design*, Ph.D. thesis, Katholieke Universiteit Leuven, Leuven, Belgium, 1997.

[24] Ewout P. Vandamme, Dominique M. Schreurs, and Cees van Dinther, "Improved three-step de-embedding method to accurately account for the influence of pad parasitics in silicon on-wafer RF test-structures," *IEEE Transactions on Electron Devices*, vol. 48, no. 4, pp. 737–742, April 2001.

[25] Holger Heuermann and Burkhard Schiek, "Robust algorithms for Txx network analyser self-calibration procedures," *IEEE Transactions on Instrumentation and Measurement*, vol. 43, no. 1, February 1994.

[26] Hermann-Josef Eul and Burkhard Schiek, "A generalized theory and new calibration procedures for network analyser self-calibration," *IEEE Transactions on Microwave Theory and Techniques*, vol. 39, no. 4, pp. 724–731, April 1991.

[27] Holger Heuermann and Burkhard Schiek, "Procedures for the determination of the scattering parameters for network analyser calibration," *IEEE Transactions on Instrumentation and Measurement*, vol. 42, no. 2, pp. 528–531, April 1993.

[28] B. Parvais, A. Cerdeira, D. Schreurs, and J.-P. Raskin, "Harmonic distortion characterization of SOI MOSFETs," *European Gallium Arsenide and other Compound Semiconductors Application Symposium 2003*, October 6-7 2003, Munich, Germany.

[29] Morin Dehan, *Characterization and Modeling of SOI RF Integrated Components*, Ph.D. thesis, Université catholique de Louvain, Louvain-la-Neuve, Belgium, September 2003.

[30] Jean-Pierre Colinge, *Silicon-on-insulator technology: Materials To VLSI*, Kluwer Academic Publisher, 1991.

[31] Rupert Howes and William Redman-White, "A small-signal model for the frequency-dependent drain admittance in floating-substrate MOSFET's," *IEEE Journal of Solid-State Circuits*, vol. 27, no. 8, pp. 1186–1193, August 1992.

[32] Jean-Pierre Colinge, "Fully-depleted SOI CMOS for analog applications," *IEEE Electron Device Letters*, vol. 46, no. 5, pp. 1010–1016, May 1998.

[33] Derek Shaeffer and Thomas H. Lee, "A 1.5-V, 1.5-GHz CMOS low noise amplifier," *JSSC*, vol. 32, no. 5, pp. 745–759, May 1997.

[34] Johan Janssens and Michiel Steyeart, *CMOS Cellular Receiver Front-Ends from Specification to Realization*, Kluwer Academic Publishers, 2002.

[35] Vladimir Aparin and Charles Persico, "Effect of out-of-band terminations on intermodulation distortion in common-emitter circuits," *IEEE International Microwave Symposium Digest*, pp. 977–980, 1999.

[36] Vladimir Aparin and Lawrence E. Larson, "Linearization of monolithic LNAs using low-frequency low-impedance input termination," *European Solid-State Circuit Conference (ESSCIRC)*, September 2003.

[37] M. Steyaert, M. Borremans, J. Craninckx, J. Crols, J. Janssens, and P. Kinget, "RF CMOS design, some untold pitfalls," *Analog Circuit Design*, pp. 63–88, 1996, Kuwer Academic Publisher.

[38] Wei Guo and Daquan Huang, "Noise and linearity optimization methods for a 1.9-GHz low noise amplifier," *Proc. of* 3^{rd} *Int. Conf. on Microwave and Millimeter Wave Technology*, pp. 923–927, 17-19 August 2002.

CHAPTER 3
DESIGN OF INTEGRATED OSCILLATORS

Oscillators are machines designed to transform DC energy into RF energy by creating a stable periodic signal. For that purpose, a time reference must exist somewhere in the machine. The time reference will determine the type of oscillator. For instance, crystal oscillators make use of the very stable eigen frequency of crystals, which is used as time reference. In this work, the time reference consists of an inductor-capacitor (LC) resonator. This time reference has two complex poles and thus a preferred frequency. Indeed, the energy flows from one tank (C) to the other (L) at a specific speed determined by the time reference itself.

Whereas fully integrated ring oscillators are being widely used for digital and data applications, fully integrated LC-tuned oscillators have been used only for few years in wireless products. Performance concern as well as large area[1] still inhibit the widespread acceptance of integrated tuned oscillators. However, the reduction of the number of RF interfaces in the package, the ease of use of fully integrated parts and the compact board size outweigh the extra die cost for large volume wireless terminals.

The competition between different semiconductor technologies to achieve integrated oscillators is not only based on single transistor performances, but is also related to other on-chip components. Compared to bipolar technology, CMOS realizations of fully integrated LC oscillators showed better overall performances [1] and are thus good candidates for the integration of short-range transceivers. Because of its reduced junction capacitances and high substrate isolation, SOI may offer some advantages over bulk CMOS. A comparison of integrated oscillators in both CMOS bulk and CMOS SOI done in [2] showed competitive performances and that SOI offers the ability of very low power consumption, while bulk has better spectral purity.

Much technological efforts have been made in the past few years to improve the performances of the passive devices embedded in CMOS integrated oscillators. In the design of microwave integrated oscillators however, both the resonator and

[1]Integrated spiral inductors occupy large chip area, especially for applications in the low GHz range, and their quality is low compared to external inductors. On the other hand, area consuming input/output pads are necessary to connect off-chips components.

the transistors must be designed. This contrasts with the traditional approach in microwave engineering that consists in designing the passive part for a given active amplifier. This also contrasts with the approach of micro-electronics engineers who design the transistors for a given resonator. Only few design methodologies [3] have been given in the recent literature for the optimization of the circuits. Therefore, we propose in this chapter a methodology for the design of integrated LC oscillators.

As stated in the introduction chapter, voltage-controlled-oscillators (VCOs) are used in transceivers inside a phase-locked-loop (PLL). The main specifications on the design of VCOs are [1]:

- *Center frequency*: the output frequency f_o of the VCO with the control voltage at its center value.
- *Tuning range*: the range of output frequencies that the VCO oscillates at over the full range of the control voltage.
- *Tuning sensitivity*: the change of output frequencies per unit change in the control voltage, typically expressed in [Hz/V].
- *Spectral purity* around the oscillation: in the frequency domain, the spectral purity is specified in term of phase noise; it is surely the most important specification in the design of oscillators.
- *Load pulling*: the sensitivity of the output frequency to changes in its output load; this specification depends strongly on the output stage in the oscillator.
- *Supply pulling*: the sensitivity of the output frequency to changes in the power supply, expressed in [Hz/V].
- *Power consumption*: the DC power drained by the oscillator.
- *Output power*: the power delivered by the oscillator to a specified load.
- *Harmonic suppression*: specifies how much smaller the harmonics of the output spectrum are compared to the fundamental component, typically expressed in [dBc].

All those specifications will be discussed in the following of this chapter in the case of integrated LC oscillator in SOI technology. The chapter is organized as follows. In Sections 3.1 and 3.2, the theory of oscillators and phase noise are respectively introduced. It will be explained how nonlinearities affect the design. This work focuses on the well known LC oscillator topology using a differential crossed-coupled

pair as negative resistance, but a single-ended Colpitts oscillator is also briefly analyzed. In section 3.3, an adequate design methodology based on describing functions is provided and the trade-offs are discussed. The tuning range and the sensitivity of VCOs are studied in details in section 3.4. Small- and large signal characterizations of SOI varactors are provided. The role of the technology will finally be studied experimentally in the last section.

3.1 Theory of oscillators

Linear modeling of oscillators is widely used for its simplicity and because it highlights the governing mechanisms. However, linear models predict an infinite signal power at the oscillation frequency, which is not physically acceptable. Circuit nonlinearities explain the limitation of the output signal. Oscillators are thus inherently nonlinear circuits. Even if some designs may work in class A, because the best spectral purity is obtained for high amplitude level, practical designs are far from the validity range of linear approximation. Despite, a linear model is introduced hereafter to facilitate the understanding of the oscillation mechanisms. The determination of the output power will then be discussed in section 3.1.2.

3.1.1 Linear modeling of oscillators

The feedback (two-port) model Oscillators may be modeled as closed-loop systems (Fig. 3.1). The transfer function of such a system writes

$$\frac{Y(s)}{X(s)} = \frac{H(s)}{1 - F(s)H(s)} \tag{3.1}$$

The feedback loop may be a unity return, $F(s) = 1$, as in the case of ring oscillators. To improve the oscillation stability, a filter (a resonator) may be introduced in the feedback loop. In order to provide a stable oscillation at a given angular frequency ω_o, the system must satisfy the Barkhausen conditions

$$\begin{cases} |F(j\omega_o)H(j\omega_o)| &= 1 \\ \angle F(j\omega_o)H(j\omega_o) &= n \cdot 2\pi \end{cases} \tag{3.2}$$

where n is any integer. The Barkhausen conditions tell us that the poles of an oscillating system are placed on the imaginary axis of the complex plane. In practical implementations however, the gain value is set higher than unity, placing the poles in the right half of the complex plan. Indeed, if the effective loop gain would be

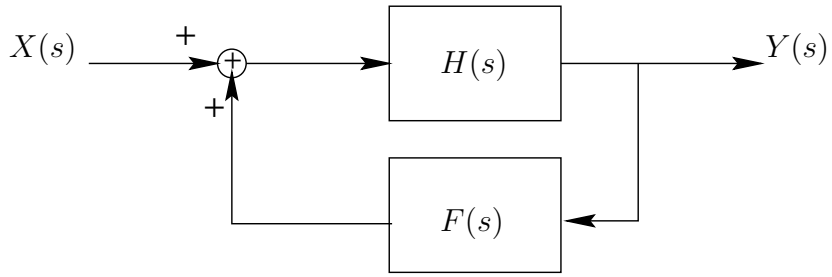


Figure 3.1: Two-ports model of oscillator: block diagram.

exactly unity for all signal levels, an initial low level oscillation would never build up in amplitude. Some nonlinearity must be present to reduce the effective loop gain to unity at the full output level. During each RF cycle the instantaneous voltage will sweep over the full range and thus will normally experience this non-linearity. From the technological point-of-view, the addition of gain is advantageous as it decreases the sensitivity to process variations.

The negative resistance (one-port) model In the previous paragraph, the oscillator was described in terms of two-ports (an amplifier and a resonator in the feedback loop). The oscillator may in general also be described by a one port model. Indeed, the system may be viewed as a resonator tank whose losses are compensated by a negative resistance (Fig. 3.2). If the impedance of the resonator is $Z_f = R_f + jX_f$ and the impedance of the active circuit (the negative resistance) is $Z_a = R_a + jX_a$, then the system must satisfy the following conditions to give a oscillation at the pulsation ω_o:

$$\begin{cases} R_a(\omega_o) + R_f(\omega_o) & < & 0 \\ X_a(\omega_o) + X_f(\omega_o) & = & 0 \end{cases} \tag{3.3}$$

Those relations mean again that the poles of the system have to be placed in the right half complex plan. Note that relations (3.3) and (3.2) are necessary but not sufficient conditions [4].

The frequency $f_o = \omega_o/2\pi$ given by the linear approach presented here is not necessary the frequency at which the oscillator works in regime. However, if the quality factor of the resonator is high enough, f_o is close to the regime frequency [4].

Link between the one- and the two-port models Consider an LC tank connected to the drain of a transistor (Fig. 3.3 (a)). Following the two-port model, a null phase shift is necessary to obtain an oscillation. As at resonance, the tank

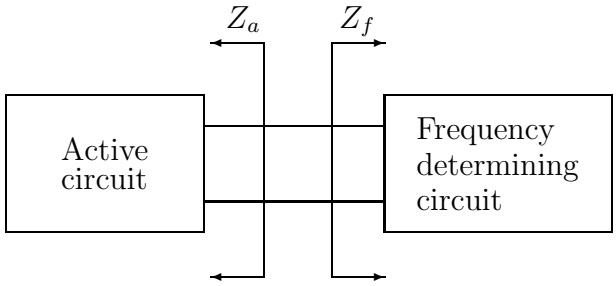


Figure 3.2: One-port model of an oscillator.

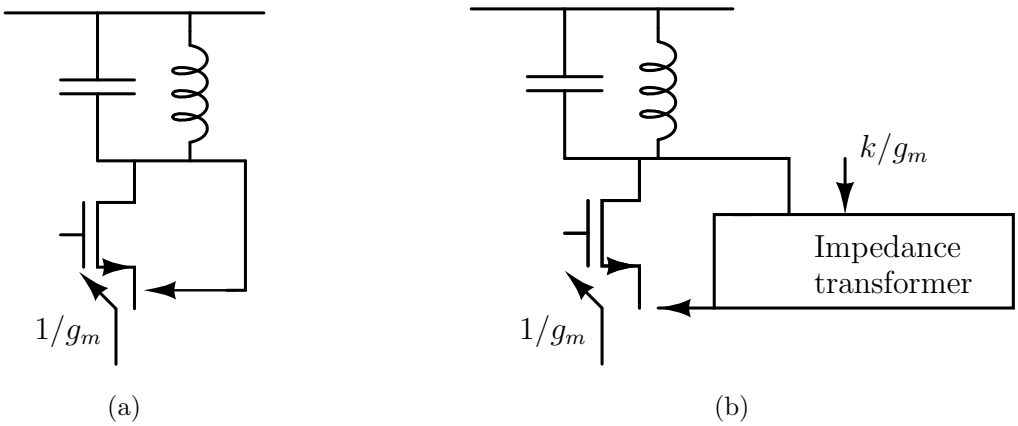


Figure 3.3: Impedance conversion in LC oscillators.

impedance is real, the drain of the transistor should be connected to the source. But the impedance seen from the source is so low (about $1/g_m$, with g_m the gate transconductance of the MOSFET) that the loop gain is below one. An impedance transformer is thus needed. This transformer can be implemented at least by three ways, leading to three well-known families of oscillators: Colpitts (Fig. 3.4 (a)), Hartley (Fig. 3.4 (b)), and active (Fig. 3.4 (c)). In the following, two of those structures will be studied. First, a single-ended Colpitts oscillator using a common-source FET configuration (Fig. 3.5). Second, an active feedback differential topology (Fig. 3.6). The one-port model will be used in the next paragraphs in the case of these two circuits, filling the gap between the one- and the two-port models.

A single-ended Colpitts oscillator The oscillator topology presented in Fig. 3.5 presents the advantage that the parasitic capacitances arising from the MOSFET add to the capacitances $C_{1,2}$ of the resonator. The analysis is thus simplified.

The impedance Z between the gate and drain of the transistor is

$$Z = \frac{-\omega^2(a(C_1+C_2)+b(g_{m1}+g_{d1}))}{\omega^2(a^2+\omega^2 b^2)} + j\cdot\frac{\omega(a(g_{m1}+g_{d1})-\omega^2 b(C_1+C_2))}{\omega^2(a^2+\omega^2 b^2)} \quad (3.4)$$

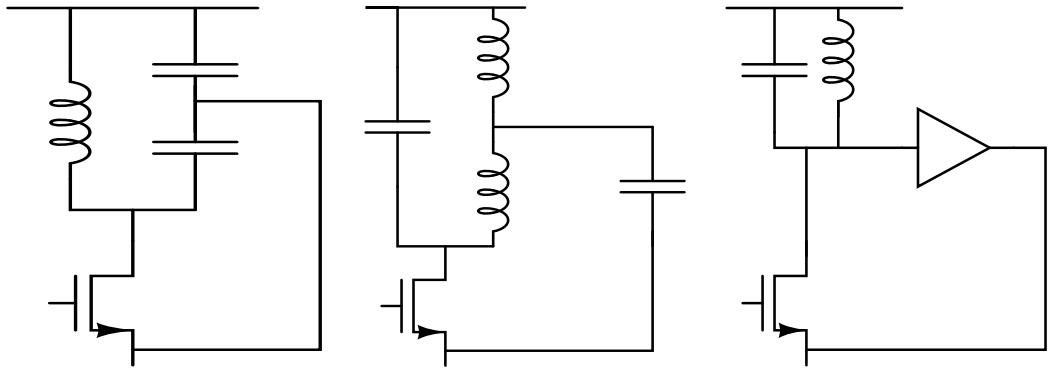


Figure 3.4: Several kinds of LC oscillators: (a) Colpitts (b) Hartley (c) Active feedback.

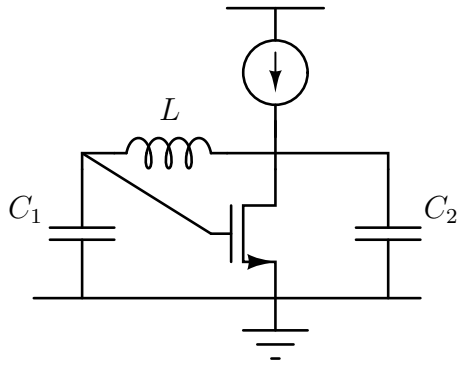


Figure 3.5: A single ended common source Colpitts oscillator.

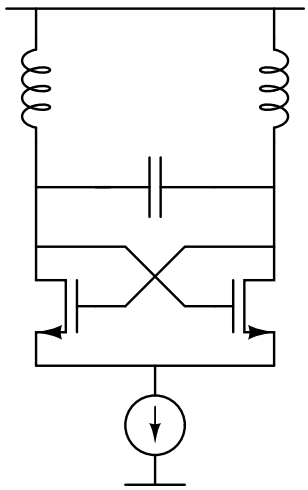


Figure 3.6: A differential active feedback oscillator.

where $a = C_3(g_{m1} + g_{d1}) + g_{d1}C_1$, $b = C_1C_2 + C_1C_3 + C_2C_3$, g_{m1} is the transistor linear gate transconductance, g_{d1} its linear output conductance and C_3 the total capacitance between gate and drain. The oscillation conditions (3.3) implies in this case that

$$\begin{cases} \Im(Z) + L\omega &= 0 \\ \Re(Z) + R &< 0 \end{cases} \tag{3.5}$$

where R represents the equivalent series resistance of the inductor L. Those relations simplifies in the case where $C_3 = g_d = 0$ to:

$$\omega_o^2 = \frac{1}{L\left(\frac{C_1C_2}{C_1+C_2}\right)}, \tag{3.6}$$

$$g_m > \frac{R(C_1 + C_2)}{L}. \tag{3.7}$$

In such Colpitts topology the transistor may work either in class A or in class C. We provide hereafter a more exact description of the Colpitts oscillator using a nonlinear representation of the drain current I_D [5]:

$$I_D(t) = \frac{\beta}{2}(V_G(t) - V_{th})^2 \cdot g(t) \tag{3.8}$$

$$= \frac{\beta}{2}(V_m \cos(\omega_o t) + V_B - V_{th})^2 \cdot g(t) \tag{3.9}$$

where

$$\begin{aligned} g(t) &= 1 \quad \text{for } V_m \cos(\omega_o t) > (V_{th} - V_B) \\ &= 0 \quad \text{for } V_m \cos(\omega_o t) \leq (V_{th} - V_B) \end{aligned} \tag{3.10}$$

and V_m is the oscillation amplitude, V_G th gate voltage, V_B the DC gate bias in steady state and V_{th} the threshold voltage below which the transistor cuts off, β is a coefficient that depends on the technology and the transistor size. The average current I_o and the fundamental current component I_{ω_o} at ω_o are then given by

$$I_o = \frac{\beta V_m^2}{2}\left(1 + 2\left(\frac{V_{th} - V_B}{V_m}\right)^2\right) \tag{3.11}$$

$$I_{\omega_o} = \beta V_m^2\left(\frac{V_B - V_{th}}{V_m}\right) \equiv g_{mc}V_m \tag{3.12}$$

where g_{mc} is the critical transconductance for oscillation. The oscillation conditions in class A lead to

$$2\beta I_o \leq \frac{3}{2}g_{mc}^2. \tag{3.13}$$

The maximum amplitude is then equal to g_{mc}/β. The condition (3.13) only leaves

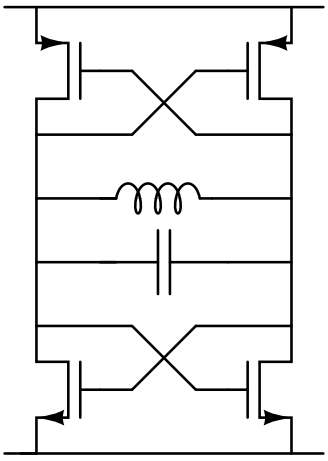


Figure 3.7: Active feedback LC oscillator with double differential pair.

a small range for nominal transconductance; an increase in 20% in g_{mc} can easily kill the oscillation. So designers prefer to use the class C mode of operation. In this regime, Huang [5] showed that, as long as large current is available, at the limit case when the current in the FET is a series of Dirac impulses, the amplitude is $V_m = \frac{2I_o}{g_{mc}}$, independently of the switch implementation. In this case, the power of the fundamental component is maximized, which gives best spectral purity performances to the design [6].

A differential oscillator with active feedback The architecture presented in Fig. 3.6 is popular among CMOS designers for their relaxed startup conditions. Furthermore, its differential output is an advantage for its integration in practical transceivers. The circuit has been widely studied in the literature e.g., in [7, 8, 9, 10, 11]. Several modifications to this basic topology have been proposed to improve the performances or functionality. A double cross pair (Fig. 3.7) improves gain, symmetry and phase noise without draining more current [12]. Coupling of circuits gives differential quadrature outputs [8, 13]. The implementation may use or not a current bias source. This source may be a pMOSFET or a nMOSFET (Fig. 3.8). The pMOS tail current is usually preferred for two reasons. First, with respect to n-core VCO, the pMOS current source taps the tank common-mode point which has less variation from AC perspective compared to the common-source point tapped by the nMOS source. Second, a pMOS has lower flicker noise level for a given current bias.

The losses of the *LC* tank resonator are compensated by the active cross-coupled

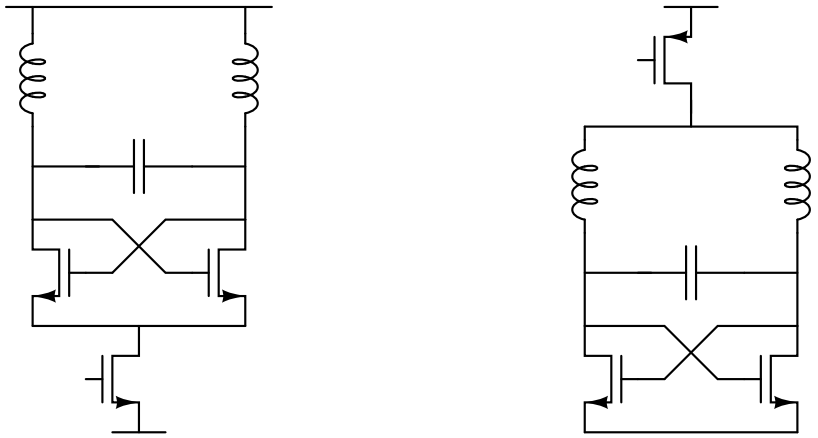


Figure 3.8: Two implementation styles for an active feedback oscillator. (a) nMOS current source. (b) pMOS current source.

differential pair. The admittance $Y(\omega)$ at the drains of the transistors is:

$$Y(\omega) = \frac{g_d - g_m}{2} + j\omega\frac{C_{gs}}{2} \tag{3.14}$$

with C_{gs} the gate-to-source capacitance. As in practical cases $g_d \ll g_m$, the real part of $Y(\omega)$ is negative, as required by the one-port oscillator model.

Because of the positive feedback involved in the differential pair, the current quickly switches from one side of the pair to another. The differential pair may then be modeled as a current source switching between $+I_B/2$ and $-I_B/2$. The amplitude of the output voltage increases then linearly with the tail current I_B, as it will be shown in the next section. This mode of operation is referred to the *current-limited* regime. For increasing amplitudes, the transistors will enter in the triode regime and the tank voltage will finally be clipped at ground (and at the supply voltage for the architecture in Fig. 3.7). Since the tail transistor is in triode region, the tail current does not stay constant anymore. This mode of operation is called the *voltage-limited* regime [12].

3.1.2 Amplitude of the signal

A linear model may not predict the amplitude of the signal by itself. Indeed, the limitation of the signal amplitude comes from the saturated gain of the amplifier. We provide in this section some techniques to find easily the output level of oscillators.

The amplitude of the output signal of the circuit presented in Fig. 3.6 may be evaluated as follows [12, 14]. In the current-limited regime, the current $I(t)$ in the transistors may be approximated by a square wave of amplitude I_B. From

the filtering action of the resonator, this current may be approximated by its first-order Fourier series. Then, the amplitude of the differential tension V_d is given by the Ohm's law, provided that the imaginary parts of the impedances cancel out at resonance:

$$V_d \simeq \frac{2}{\pi} I_B R \tag{3.15}$$

where R is the tank equivalent resistance. The amplitude behavior is summarized as follows

$$V_d = \begin{cases} \frac{2}{\pi} R I_B & \text{current-limited regime} \\ V_{limit} & \text{voltage-limited regime} \end{cases} \tag{3.16}$$

Volterra series From a more general point of view, the amplitude of the signal may be found if the amplifier is described as a nonlinear system. As explained in chapter 2, a third order Volterra series may represent a weakly nonlinear amplifier. The Barkhausen criteria becomes [15]

$$H_1(j\omega) + \Omega_1(j\omega) A^2 = 1 \tag{3.17}$$

where $H_k(j\omega)$ is the k^{th}-order Volterra Kernel,

$$\Omega_1(j\omega) = \frac{1}{4} \left(\mathcal{H}_3(j\omega, j\omega, -j\omega) + \mathcal{H}_3(j\omega, -j\omega, j\omega) + \mathcal{H}_3(-j\omega, j\omega, j\omega) \right), \tag{3.18}$$

and

$$\mathcal{H}_3(a,b,c) = H_2(a+b,c) \frac{H_2(a,b)}{1 - H_1(a+b)} + H_2(a,b+c) \frac{H_2(b,c)}{1 - H_1(b+c)} + H_3(a,b,c). \tag{3.19}$$

Both the amplitude A of the signal and the oscillation frequency may be obtained from the complex relation (3.17).

Let us take the example of the active feedback oscillator. Suppose that the transconductance $g_m = g_{m1} + g_{m2} V_G + g_{m3} V_G^2$ is the only nonlinearity and is frequency independent. As a consequence, only the linear GLC tank is frequency dependent. In this case, the Barkhausen criterion (3.17) becomes

$$\begin{cases} \frac{3}{4} A^2 \frac{L^2 \omega^2}{D} g_{m2}^2 (4\omega^2 LC - 1) + \omega^2 LC - 1 = 0 \\ g_{m1} - G + \frac{3}{4} A^2 \left(g_{m3} + 4 \frac{L^2 \omega^2}{D} g_{m2}^2 (G - g_{m1}) \right) = 0 \end{cases} \tag{3.20}$$

with $D = (1 - 4\omega^2 LC)^2 + 4\omega^2 L^2 (G - g_{m1})^2$.

Describing Functions If the nonlinearity is too hard to be described by a simple (third-order) Volterra series, the describing function approach presented in the previous chapter is an elegant alternative. The advantage of this technique is that it allows the inclusion of the nonlinear behavior of a system while maintaining the simplicity associated with the linear system analysis. Besides, it allows to evaluate the distortion of the generated signal in some cases [16]. Nevertheless, the analysis is based on the assumption that the signal is close to a sine wave. This is the case for high-Q LC oscillators.

Let us consider the two-port model of oscillators depicted in Fig. 3.1, but $H(s)$ being this time nonlinear. This nonlinear transfer function may be approximated by a describing function H_{DF} from (2.59). Then the oscillation conditions are again given by the non-trivial solution of the system:

$$H_{DF}(A,\omega)\cdot F(\omega) - 1 = 0 \tag{3.21}$$

where $F(\omega)$ is the transfer function of the linear filter in the feedback path. This equation is the Barkhausen criteria in which the describing function of the amplifier replaces its classical linear transfer function. Practical use of describing functions in the design of oscillator will be demonstrated in section 3.3.

The van der Pol approximation B. van der Pol, in the late 1920s, evaluated electronic oscillators using vacuum tubes that provide a negative conductance characteristic, and obey to the following nonlinear differential equation [17]:

$$LC\frac{d^2v}{dt^2} + L\frac{d}{dt}[Gv + f(v)] + v = 0. \tag{3.22}$$

With the variables changes $T = t/\sqrt{LC}$ and $\xi(v) = Gv + f(v)$, equation (3.22) becomes

$$\frac{d^2v}{dT^2} + \sqrt{\frac{L}{C}}\frac{d}{dT}[\xi(v)] + v = 0. \tag{3.23}$$

Van der Pol gave the solution of this equation when the function $\xi(v)$ is approximated by a third order polynomial of the form $F(v) \approx -a_1 v + b_1 v^3$. The analysis lead to

$$v(t) = \frac{V_{max}}{\sqrt{1+\exp\left(\frac{(t_o-t)\eta}{\sqrt{LC}}\right)}}\cos\left(\frac{t}{\sqrt{LC}} + \Phi_o\right), \tag{3.24}$$

$$V_{max} = \sqrt{\frac{4a_1}{3b_1}}, \tag{3.25}$$

$$\eta = \sqrt{\frac{L}{C}} a_1. \tag{3.26}$$

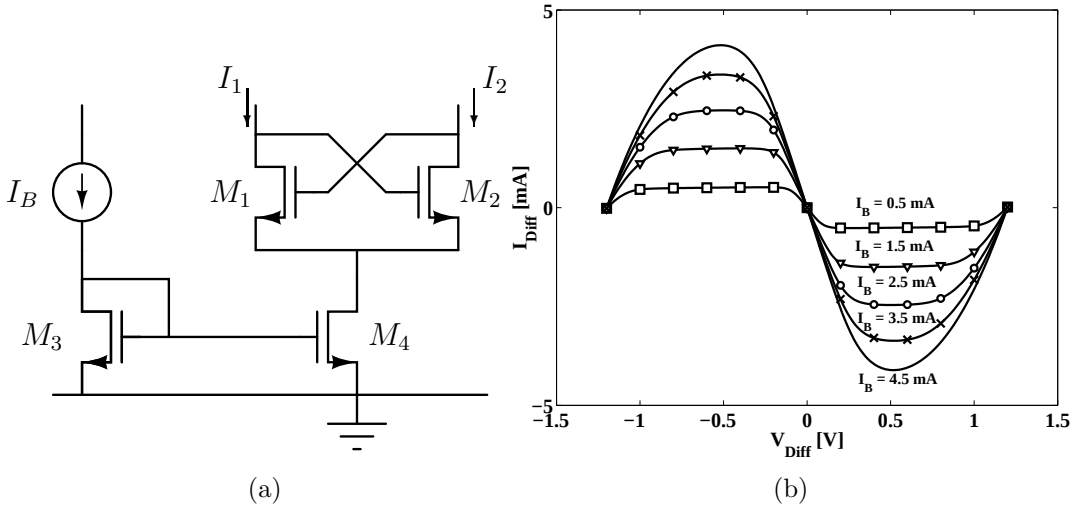
where Φ_o is the phase of the wave $v(t)$ at the instant $t = t_o$. Because some CMOS implementations follows the same law, this analysis is not restricted to vacuum tubes [18].

Discussion We compared the different methods for calculating the amplitude of the output signal for the cross-coupled differential transconductor drawn in Fig. 3.9(a). Its $I_{Diff} - V_{Diff}$ characteristics (i.e., differential current as function of differential voltage) is shown in Fig. 3.9(b). It was obtained by numerical simulation using the BSIM3v3.1 model provided by the foundry. Transistors M_1 and M_2 have 8 fingers, 2 μm width and 0.14 μm long each, while M_3 and M_4 keep the same W/L ratio but the finger length is 2 μm. The $I - V$ plot consists of three main sections and their symmetric, which denote the operating states of the devices. Near the common-mode, both devices operate in saturation and $|I|$ increases linearly with $|V|$. The slope is $-g_m/2$ according to (3.14). When $|V|$ continues to increase, one of the devices slips into triode. When $|V|$ is large enough, this device falls deeper into triode while the other shuts off as its gate bias drops below threshold.

The accuracy of each method depends on the accuracy of the used approximation. Typically, a piecewise approximation holds in the current-limited regime, while at higher biases, a third-order polynomial is a suitable approximation of the transconductor. A comparison of the methods is shown in Fig. 3.10. We considered the transconductor of Fig. 3.9 and the loaded Q of the oscillator was fixed to 9. The transition in real circuits between current- and voltage-limited regimes is obviously not as sharp as relation (3.16) indicates, but it gives a good first order approximation. In the case of a differential crossed-pair, a third order approximation of the Volterra Kernels is not sufficient and the results are not in agreement. However, at higher voltages the van der Pol's approximation becomes more accurate and gives results in agreement with the describing function method. This last method is the more accurate as it is directly related to the first-order Fourier coefficient of the transconductor.

3.2 Phase noise

The output spectrum of oscillators is ideally a Dirac function. In practical cases, a spreading of the signal occurs due to the action of noise. Noise affects both amplitude and phase of the signal. The output amplitude modulation may be corrected by



(a) (b)

Figure 3.9: Crossed-coupled differential transconductor. (a) Schematic. (b) Simulated characteristics versus current bias; $I_{Diff} = I_1 - I_2$; V_{Diff} is the voltage between the drains of M_1 and M_2; Bsim3v3.1 model; FD SOI nMOS, M_1 and M_2: 8x2 μm/0.14 μm; M_3 and M_4: 8x28 μm/2 μm; OKI technology.

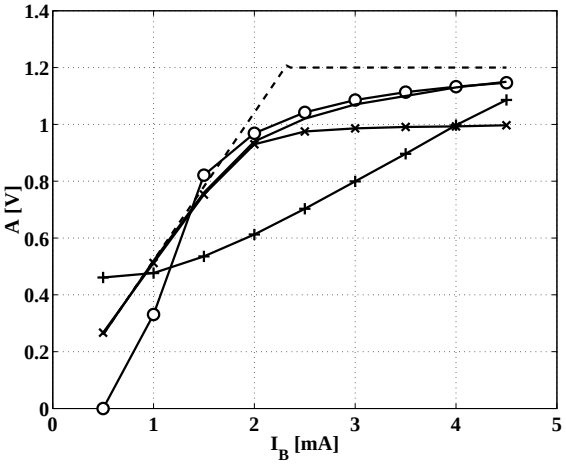


Figure 3.10: Amplitude of the output signal of an LC oscillator using the transconductor of Fig. 3.9(b) and $Q_T = 9$ evaluated by different methods: DF (straight line), relation (3.15) (dashed line), van der Pol (circles), Volterra (plus), Eldo RF simulations (crosses).

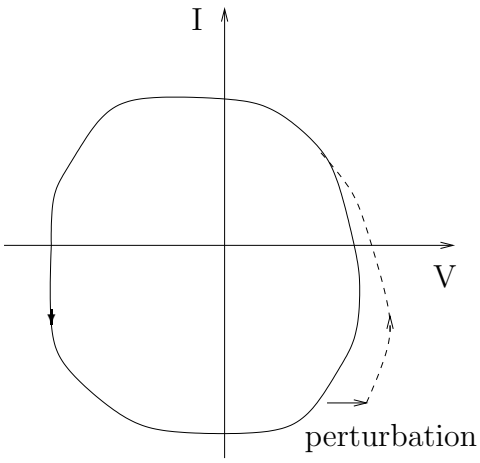


Figure 3.11: Limit cycle in the $V - I$ state space of an LC oscillator.

automatic amplitude control loops, but phase noise cannot be corrected. Indeed, an oscillator is an autonomous system for which no phase reference does exist. Phase noise quantifies the signal broadening.

The dynamics of a LC oscillator may be described by a plot of the voltage V across the capacitance and the current I across the inductance (Fig. 3.11). Regardless of its starting point, the trace for unperturbed oscillator in regime forms a closed curve, namely its *limit cycle*, since this waveform is perfectly periodic. When a stable oscillator is perturbed, the trajectory remains within a small band around the limit cycle. Radial perturbations remain small due to the tendency of the state to return to the limit cycle. However, fluctuations in the direction along the limit cycle do not experience any restoring force to return the phase to its original value [19]. Any fluctuation in the phase of the oscillation persists indefinitely. Consequently, in the presence of noise, the state point walks randomly along the limit cycle, or, the phase undergoes a diffusion process. Both physical [20] and mathematical [18] analysis showed that the associated constant of diffusion is the parameter that characterizes phase noise.

The single sideband-to-carrier ratio (SSCR) is a suitable measure of the phase noise (Fig. 3.12). It corresponds to the ratio between the output noise power in a 1-Hz bandwidth at the frequency offset $\Delta\omega$ from the carrier and the power of the carrier:

$$L(\Delta\omega) = 10 \log \left(\frac{P_{sideband}(\omega_o + \Delta\omega, 1 \text{ Hz})}{P_{carrier}} \right) \tag{3.27}$$

where $P_{sideband}(\omega_o + \Delta\omega, 1 \text{ Hz})$ represents the power of a 1 Hz bandwidth contained in a lateral band of the output spectrum. This power is measured at a $\Delta\omega$ shift from the center frequency ω_o. $L(\Delta\omega)$ is usually expressed in [dBc/Hz] i.e., decibels in a

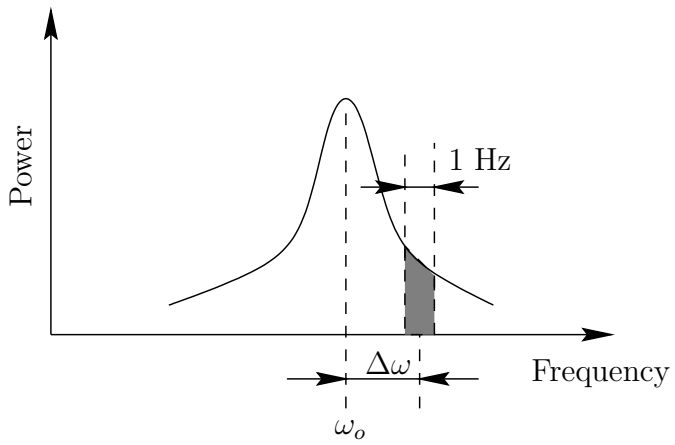


Figure 3.12: Definition of Phase Noise: the ratio between the output noise power in a 1-Hz bandwidth at the frequency offset $\Delta\omega$ from the carrier and the power of the carrier.

1 Hz bandwidth relative to the carrier. Note that both amplitude and phase noises are included in (3.27). The relation (3.27) also relates an important trade-off: the higher the signal amplitude, the lower the phase noise, but the higher the amplitude also means the higher the power consumption in most practical realizations.

Phase noise is one of the major concerns for the CMOS transceiver designers. Indeed as shown in Fig. 3.13, phase noise limits the channel density. Telecommunication standards fix the specification on phase noise. For instance the GSM[2] imposes that a -102 dBm signal must be detected even if a -43 dBm signal is located in the neighbor channel. The channels are 200 kHz width and 600 kHz spaced. For a 9 dB demodulated SNR, definition (3.27) states that the maximal phase noise is -121 dBc/Hz at 600 kHz offset from the carrier.

Oscillator phase noise has been studied from several angles, ranging from the mathematical physics [21] to CAD-oriented methods [22, 19, 23] and design-oriented approaches [24, 25, 26, 27], to name a few. We will restrict us in the following to the design-oriented approaches. The simplest modeling makes use of linear time-invariant (LTI) model (also known as the Leeson's model [24]) and has been widely used due to its simplicity. Leeson showed that phase noise is proportional to noise-to-carrier ratio and inversely proportional to the square of resonator quality factor. However, this model leaves the constant of proportionality unspecified and the close-in phase noise due to the upconversion of low frequency noise sources is only described by an empirical approach. Since any oscillator is a periodically time-varying system, its time-varying nature has been taken into account in Hajimiri's

[2]global system for mobile communication.

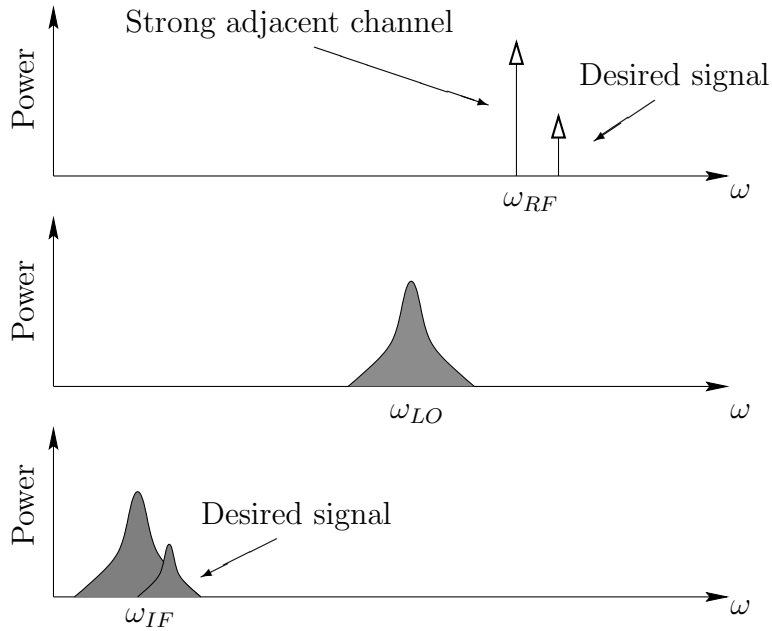


Figure 3.13: Phase noise effect on a transceiver. Top: signals spectra at the input of the transceiver; center: spectrum of the local oscillator of the considered transceiver; bottom: output (down-converted) spectrum.

model [26, 28], leading to better accuracy and insight to the oscillator design. This linear time-varying (LTV) modeling uses an impulse-sensitivity function (ISF) that measures the phase perturbation caused by small impulses injected into the circuit at any time during a period of operation. Deriving this function directly from circuit properties is not straightforward however. Approximation of the ISF reduces to the LTI model and, for more accuracy SPICE-like simulator has to be used to estimate the phase response. The method has been shown to be incorrect, but not significantly inaccurate, in some cases [29]. Despite their lacks, linear models are analytically easily treatable and give good insight of the trade-offs between the design parameters. More rigorous analysis requires the use of nonlinear models. The switching behavior of oscillators was included in the noise factor appearing in the Leeson's formula in [30, 27]. On the other hand, harmonic balance tools [31] or conversion matrix [32] approaches embedded in CAD softwares give accurate results but poor insight to the designer. Some authors bridged the gap between CAD-oriented and design-oriented nonlinear models, for instance reference [33] in the case of Colpitts oscillator and reference [18] for the crossed-coupled LC oscillator.

In the next subsection, the phase noise mechanism is explained with the LTI model [34] while the effect of nonlinearities are introduced in section 3.2.2.

3.2.1 Linear time invariant model

The model is based on the hypothesis that the amplifier is linear and time invariant i.e., its behavior does not change much for several periods of oscillation.

The two-port model of oscillators is used to model the phase noise (Fig. 3.1). The noise arises from the tank losses and from the active devices. If the noise is injected at the signal access port, the output spectrum results from the modulation of the input noise power density $S_{n,in}$ by the square of the norm of the transfer function of the system, according to the Wiener-Kintchine theorem:

$$S_{n,out}(\omega) = \left|\frac{Y(\omega)}{X(\omega)}\right|^2 S_{n,in}(\omega) \tag{3.28}$$

where X and Y were defined in Fig. 3.1. Let us then evaluate $\left|\frac{Y}{X}\right|^2$ starting from the transfer function of the closed-loop system (3.1). Since the circuit oscillates in regime at the angular frequency ω_o, and as we are interested in the behavior close to the carrier, we may approximate $H(j\omega_o + \Delta\omega)$ by its first-order Taylor series $H(j\omega_o + \Delta\omega) \simeq H(j\omega_o) + \Delta\omega\frac{dH}{d\omega}$ with the derivative $\frac{dH}{d\omega}$ evaluated at $\omega = \omega_o$. Then, the describing equation becomes

$$\frac{Y[j(\omega_o + \Delta\omega)]}{X[j(\omega_o + \Delta\omega)]} \simeq \frac{H(j\omega_o) + \Delta\omega\frac{dH}{d\omega}}{1 - H(j\omega_o) - \Delta\omega\frac{dH}{d\omega}}. \tag{3.29}$$

Because $H(j\omega_o) = 1$ at the oscillation frequency, we have

$$\left|\frac{Y[j(\omega_o + \Delta\omega)]}{X[j(\omega_o + \Delta\omega)]}\right|^2 \simeq 1 + \frac{1}{(\Delta\omega)^2 \cdot |\frac{dH}{d\omega}|^2}. \tag{3.30}$$

A more intuitive version of this relation is obtained if the polar form of $H(j\omega) = A(\omega) \cdot e^{j\phi(\omega)}$ is used. Indeed, we have $\frac{dH(j\omega)}{d\omega} = (\frac{dA(\omega)}{d\omega} + jA(\omega)\frac{d\phi(\omega)}{d\omega}) \cdot e^{j\phi(\omega)}$ and (3.30) writes

$$\left|\frac{Y[j(\omega_o + \Delta\omega)]}{X[j(\omega_o + \Delta\omega)]}\right|^2 = 1 + \frac{1}{(\Delta\omega)^2[(\frac{dA(\omega_o)}{d\omega})^2 + (\frac{d\phi(\omega_o)}{d\omega})^2]}. \tag{3.31}$$

Using the definition (A.7) of the quality factor in appendix A, we have

$$\left|\frac{Y[j(\omega_o + \Delta\omega)]}{X[j(\omega_o + \Delta\omega)]}\right|^2 = 1 + \frac{1}{4Q^2}(\frac{\omega_o}{\Delta\omega})^2 \tag{3.32}$$

$$\simeq \frac{1}{4Q^2}(\frac{\omega_o}{\Delta\omega})^2. \tag{3.33}$$

This relation underlines the importance of the quality factor. The higher the Q, the lower the phase noise. This is intuitively satisfying as following (A.7), a high Q

means that the oscillator is strongly opposed to a change in its output frequency. Relation (3.33) also indicates that the phase noise due to white noise decays by -20 dB per decade.

Following (3.27), the phase noise finally writes

$$L(\Delta\omega) = \frac{\overline{v_n^2}}{V_{RMS}^2} \cdot \frac{1}{4.Q^2} \left(\frac{\omega_o}{\Delta\omega}\right)^2 \tag{3.34}$$

where $\overline{v_n^2}$ represents the noise spectral power density [V^2/Hz] and V_{RMS} is the root-mean-square amplitude of the output signal. As we have $\overline{v_n^2} = 4k_BT \cdot R$ for a resistor R (k_B is the Boltzman's constant and T is the temperature expressed in Kelvin), Leeson stated that the phase noise of an oscillator may be expressed by

$$L(\Delta\omega) = \frac{4k_BT \cdot F \cdot R}{V_{RMS}^2} \cdot \frac{1}{4.Q^2} \left(\frac{\omega_o}{\Delta\omega}\right)^2 \tag{3.35}$$

where F is a suitable *noise factor*. This factor F is analogous to the noise factor N of other RF circuits. Indeed, it represents the total oscillator phase noise normalized to the phase noise due to the resonator loss. The difference with N is that N represents the total noise normalized by the system characteristic impedance. In the following section, it will be explained how to evaluate this factor in the case of differential active feedback oscillators.

3.2.2 Circuit analysis and impact of nonlinearities

From the previous section, the phase noise performances are highly improved by increasing the Q of the resonator. Much technology efforts have been achieved to improve the Q of the resonator. Among them, micromachining techniques will be studied in the next chapter. In CMOS technology, the tank quality factor is mainly determined by the Q of the inductors. In the last ten years, some technological and design techniques have been proposed to improve the quality of the inductors. Among them, bonding wires inductors showed high Q but lack of reproducibility for industrial manufacturing [35], the optimization of spiral inductors was also discussed in [35]. Thick copper layers inductors are used to decrease the series losses [36]. Various micromachining techniques are used to avoid substrate losses [37]. Patterned ground shield [38], high resistivity substrates [39], symmetry [40] and tapped [41] designs also help to increase the performances. But the technological improvements do not explain the difference in phase noise reported in the literature. An adequate (nonlinear) analysis of oscillators has thus to be performed.

Relation (3.35) shows that the higher the signal amplitude, the lower the phase noise. The spectral purity may then be improved at the expense of power consumption. For a given resonator and fixed bias condition, the only way to improve performances is to find the best transconductor. As the phase noise would be minimized when the signal power is maximized, transconductors which have the maximum Fourier coefficient value for the fundamental will have the best spectral purity. Wang [6] showed that the characteristic that corresponds to this optimum is a train of impulses. So, the narrower is the impulse current, the lower is the phase noise. It is worth noting that same result is obtained by Hajimiri's LTV model. This property explains why the Colpitts topology have better phase noise performances than the active feedback architecture.

The topology presented in Fig. 3.14(a) is now analyzed. It is today well understood that the large signal periodic switching of this structure underpins the noise factor F [30]. Let us assume that the output current is a square wave. Then, noise at input of the differential pair modulates the instants of zero crossing, giving rise to phase shifts. The randomly pulse-width modulated current at the output may be decomposed into the original periodic square wave in the absence of noise, superimposed with pulses of constant height but random width. In turn, these pulses may be approximated by a train of impulses at twice the oscillation frequency. The differential pair acts then as a mixer, up- or down-converting noise components around the harmonics. From that observations, Rael [27] gave closed form formulation of the phase noise for this topology that gives good design insight. This model will be summarized hereafter.

First of all, remind that noise affects both phase and amplitude. Decomposition of the perturbation in amplitude and phase modulation (AM and PM, respectively) lead to a first correction of relation (3.35): the right-hand side should be divided by a factor 2 [27]. The analysis of harmonic transfer in nonlinear oscillators provided in [30] lead to the same result.

In the topology of Fig. 3.14, the noise arises from three main sources:

1. the tank losses $\overline{v}_{nT}^2 = 4k_BT \cdot R$.

2. the differential pair:

   - channel thermal noise $\overline{i}_{nNR,C}^2 = 4k_BT \cdot PG_m$ with the noise factor P, defined in [42], is similar to the widespread channel noise factor γ. The difference between P and γ is that P is referenced to the transconductance of the transistor, while γ is to the output conductance at zero drain-to-source voltage. For long

channel devices, value for γ and P are close together and $\gamma = \frac{2}{3+nU_t\frac{g_m}{I_D}}$ [43]. When velocity saturation or hot carrier effects occur, the noise temperature increases and γ becomes higher than unity (2.5 is a usual value). Even if some short-channel effects are included in recent thermal noise models [44], accurate modeling of γ in sub-micron technologies remains an important issue in analog design.

- channel flicker noise $\overline{i}^2_{nNR,1/f} = K_f\frac{I_D}{A_D}\frac{1}{f}$ where K_f is a technological noise constant, I_D is the drain current, A_D the channel area and f the frequency.

- induced gate noise $\overline{v}^2_{nNR,G} = 4k_BT\cdot\delta\frac{\omega^2C^2_{gs}}{5g_m}$ where $\delta = 4/3$ for long channel devices. This noise is correlated to the thermal drain noise, as it has the same physical origin. The induced gate noise will not be considered in this work.

- Gate resistance noise. This contribution to the phase noise may be evaluated as proposed in [30] for the base noise in bipolar transistors. Numerical simulations showed that the gate resistance has a minor impact. As a consequence, it will be neglected in the following for the sake of simplicity.

3. the tail current noise, including the tail transistor noise and the noise originating in the whole bias circuit.

Each of these sources has a proper impact on the phase noise, because different mechanisms are involved. Furthermore, as the conversion mechanisms are nonlinear, the relative emplacement of equivalent sources in the circuit leads to a difference phase noise amount [45]. The model of section 3.2.1 directly applies for the tank losses, as the resonator filters the higher harmonics. Other sources are investigated hereafter, following Rael [27].

Channel thermal noise Rael's analysis gives:

$$F_{pair} = \frac{2PI_BR}{\pi A} \tag{3.36}$$

As the amplitude is given by (3.15) in the current-limited regime, F_{pair} reduces to P in this case. However, in the voltage-limited regime, F_{pair} increases with the bias current I_B because the amplitude A do not grow up anymore. An optimum of phase noise is thus situated at the limit between current- and voltage-limited regimes, for a given inductor.

Channel flicker noise In section 3.2.1, it was explained that additive noise gives rise to a phase noise spectrum with a -20 dB/decade slope close to the oscillation frequency f_o. In practice, a -30 dB/decade slope is observed at small offset from f_o. This slope is associated with the flicker noise of the active devices in the oscillator. The Leeson's formula is then modified in consequence:

$$L(\Delta\omega) = \frac{4k_BT \cdot F \cdot R}{V_{RMS}^2} \cdot \left(1 + \left(\frac{\omega_o}{2Q\Delta\omega}\right)^2\right)\left(1 + \frac{\omega_{1/f^3}}{|\Delta\omega|}\right) \tag{3.37}$$

where ω_{1/f^3} is the angular frequency of the corner between the $1/f^2$ and $1/f^3$ regions. The LTI model is however unable to predict how flicker noise appears around f_o. As the bias point varies over the period of oscillation, the flicker noise is upconverted to f_o. The upconverted noise enters then into the oscillator loop according to Leeson's model. The understanding of the upconversion of the $1/f$ noise needs either LTV [26] or non-linear models [46, 27].

The LTV model demonstrates that a symmetry improvement lowers the flicker noise upconversion [12, 14]. Circuit topology (Fig. 3.7), sizing [47] and layout may be adapted accordingly. This model allows to evaluate the ω_{1/f^3} corner: $\omega_{1/f^3} = \omega_{1/f} \cdot (c_o^2/\Gamma_{rms}^2)$ where $\omega_{1/f}$ is the $1/f$ corner of the flicker noise, Γ_{rms} is the root-mean-square value of the ISF and c_o its DC Fourier coefficient.

Tail current noise The switching action of the differential pair commutates noise in the tail currents like a single-balanced mixer. Noise originating in the tail current at $\Delta\omega$ upconverts to $\omega_o \pm \Delta\omega$. Similarly, noise at $2\omega_o \pm \Delta\omega$ downconverts to $\omega_o \pm \Delta\omega$. It was shown that the upconversion only produces AM, while the downconversion results in PM. The phase noise contribution due to thermal noise originally at $2\omega_o$ is:

$$F_{tail} = \frac{4}{9} P_B R g_{mB} \tag{3.38}$$

where P_B is the drain noise factor of the tail current source transistor; g_{mB} is its transconductance. As only even-order harmonics influence the phase noise, designs techniques had been proposed to filter the noise from the tail current source [48, 49]. Even if only AM noise results from upconversion, AM-to-PM may occurs in the varactor [50]. The tail current generator is generally considered the most significant source of flicker noise in differential active feedback oscillators. This noise is usually higher than the noise associated with the tail transistor alone since the diode and its associated bias network also contribute to fluctuations [51].

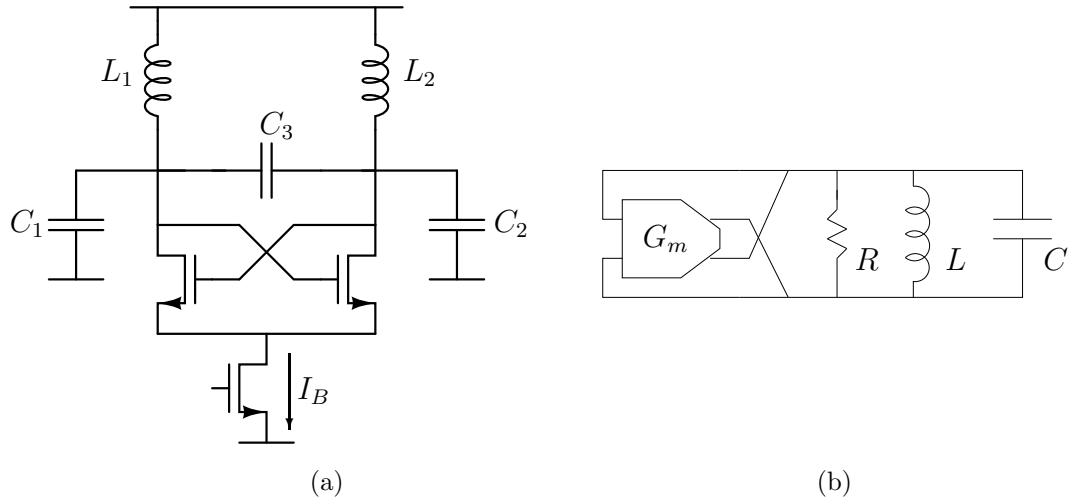


Figure 3.14: An oscillator with active feedback and its equivalent representation.

Total Noise Factor In the case of the differential active feedback oscillator, neglecting the flicker noise contribution, the total noise factor F is given by the sum of each contributions previously calculated.

$$F = 1 + \frac{2PI_BR}{\pi A} + \frac{4}{9}P_B g_{m,B} R \tag{3.39}$$

The relative contribution of each of these terms will be discussed in section 3.3.4.

3.3 Design of active feedback oscillators

The oscillator topology in Fig. 3.14(a) is investigated in this section. First, the relations to get the equivalent representation in Fig. 3.14(b) are provided. Second, the transconductor is analyzed and finally, a design methodology is given. The methodology is based on simple relations that give insight of the main trade-offs. When applied, this methodology leads to near-optimum design that has to be simulated with an appropriated numerical tool in order to tune finely the different design parameters.

3.3.1 Resonator

Because losses are present in each fabricated reactance, we may associate to $L_{1,2}$ and $C_{1,2,3}$ depicted in Fig. 3.14 a resistance. Considering these lossy devices, the resonator of the circuit in Fig. 3.14(a) can be re-drawn as in Fig. 3.15. Using the

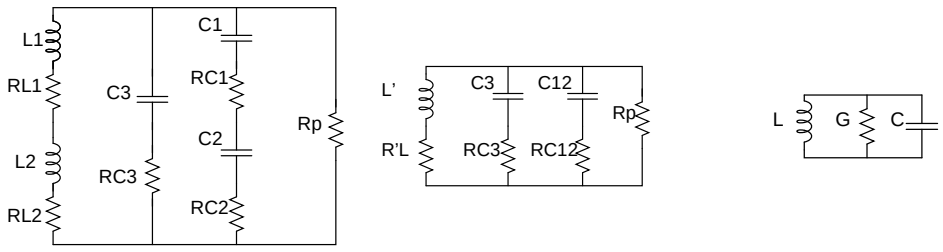


Figure 3.15: Three equivanlent representations of the lossy resonator used int he oscillator in Fig. 3.14(a).

formulas given in appendix A, we may write:

$$L' = L_1 + L_2 \tag{3.40}$$

$$R'_L = R_{L1} + R_{L2} \tag{3.41}$$

$$C_{12} = \frac{C_1 C_2}{C_1 + C_2} \tag{3.42}$$

$$R_{12} = R_{C1} + R_{C2} \tag{3.43}$$

$$Q_{C12} = \frac{Q_{C1} Q_{C2}}{\frac{Q_{C1} C_1 + Q_{C2} C_2}{C_1 + C_2}} \tag{3.44}$$

and

$$L = L' \left(1 + \frac{1}{Q_L^2}\right) \tag{3.45}$$

$$C = \frac{C_3}{1 + \frac{1}{Q_{C3}^2}} + \frac{C_{12}}{1 + \frac{1}{Q_{C12}^2}} \tag{3.46}$$

$$G = \frac{1}{R} = \frac{1}{R'_L(1 + Q_L^2)} + \frac{1}{R_{C3}(1 + Q_{C3}^2)} + \frac{1}{R_{C12}(1 + Q_{C12}^2)} + \frac{1}{R_p} \tag{3.47}$$

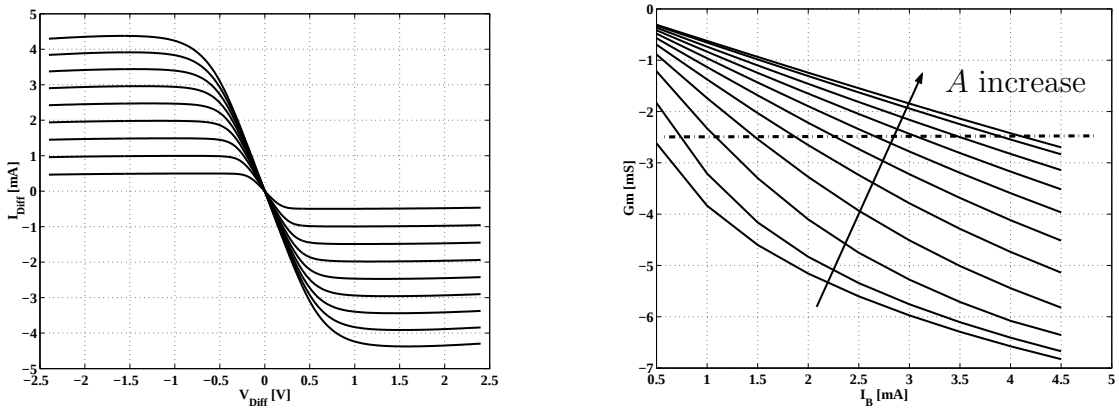
$$Y(\omega) \triangleq \frac{1}{j\omega L} + j\omega C + G \tag{3.48}$$

where Q_x is the quality factor of element x and R_p is the resistor that represents the output conductances of the transistors. The resonance angular frequency ω_o is given when the imaginary part of the tank admittance $Y(\omega)$ is set to zero:

$$\omega_o = \sqrt{\frac{1}{LC}}. \tag{3.49}$$

Note that for low values of Q_L and Q_C, the resonance frequency does not only depends on the inductance and capacitances values, but also on the dissipative elements. The tank quality factor Q_T is also obtained from the previous relations:

$$Q_T = \frac{1}{Z_0 \cdot G} = \left(\frac{Z_0}{R_p} + \frac{1}{Q_L} + \frac{1}{Q_{C3}} + \frac{1}{Q_{C12}}\right)^{-1} \tag{3.50}$$



(a) (b)

Figure 3.16: (a) $I-V$ characteristics of the transconductor depicted in Fig. 3.9 in current-limited regime. I_B from 5 mA to 45 mA by 5 mA step. (b) its effective gain as a function of the bias current I_B; A from 0.1 V to 1.9 V by 0.2 V step.

where we defined $Z_0 = \omega L = \frac{1}{\omega C} = \sqrt{\frac{L}{C}}$, the characteristic impedance of the tank.

3.3.2 Transconductor analysis

The transconductor considered here is a cross-coupled differential pair constituted of FD SOI nMOS from the 0.14 μm OKI technology (Fig. 3.9). Each transistor has 8 fingers of 2 μm width and 0.14 μm length. In the current-limited regime, the current $I_D(t)$ in the transistors may be approximated by a square wave of amplitude I_B. In this case, the $I-V$ relationship for a nMOS pair is provided in Fig. 3.16(a). The effective transconductance $G-m$ is found as a function of the signal amplitude by the describing function of its gain, as shown in Fig. 3.16(b). The horizontal dashed line represents the losses of a given tank resonator. Since the effective gain must compensate these losses in the oscillator, a stable oscillation is only possible for biases located at the cross-punt between the dashed and the continuous lines. This plot gives thus the relation between the bias current I_B and the amplitude A of the signal. When I_B increases, the overall gain also increases but in the velocity saturation region, the increase of G_m is slower and the signal amplitude tends to saturate. From the phase noise point of view, this means that the velocity saturation region should be avoided, as A does not increase, while P does [42]. In reality, the transconductor working is not limited to the current-limited, and its $I-V$ characteristics is given in Fig. 3.9(b).

3.3.3 Phase Noise and Design trade-offs

The active part of the oscillator should provide a negative conductance g_{mc} that compensate the losses of the resonator:

$$g_{mc} = \Re e\{Y(\omega_o)\} = G = \frac{1}{Q_T \cdot Z_0} \tag{3.51}$$

From (3.14), the transconductor provides a negative conductance $g_{mNR} \cong -g_m/2$. In practice, this transconductance is over-dimensioned to ensure start-up, $g_{mNR} \equiv -\alpha g_{mc}$ with $\alpha \geq 1$. The current consumption is then obtained from (3.50) and (3.51):

$$\frac{g_m}{2} = \left(\frac{g_m}{I_D}\right)\frac{I_D}{2} = \alpha G = \frac{\alpha}{Q_T Z_0} \tag{3.52}$$

$$\Rightarrow I_D = \frac{2\alpha}{Q_T Z_0}\frac{1}{(g_m/I_D)} \tag{3.53}$$

The MOSFETs current consumption I_D decrease when the tank quality factor increases. Also, low power consumption is possible if the transistor is biased in weak or moderate inversion (i.e., for high g_m/I_D).

As previously mentioned, the power supply V_{DD} limits the largest achievable differential amplitude V_d to $2 \cdot V_{DD}$ when the tail current source FET shuts off at each negative peak of oscillation. At the limit between the two regimes, the tank amplitude is $V_d = 2 \cdot V_{DD}$ and is still approximately given by (3.15). The drain current may then be approximated by

$$I_{D0} \equiv I_B(V_d = 2V_{DD}) = \frac{\pi V_{DD}}{R} = \frac{\pi V_{DD}}{Q_T Z_0} = \frac{2\alpha}{Q_T Z_0}\frac{1}{(g_m/I_D)} \tag{3.54}$$

We deduce from this equation that, when $V_d = 2 \cdot V_{DD}$,

$$\frac{g_m}{I_D} = \frac{2\alpha}{\pi V_{DD}}. \tag{3.55}$$

Nevertheless, in real circuit implementations, it is better to keep the bias transistor in saturation to let it works as a current source. The maximum differential amplitude is then given by

$$V_{d,max} = 2 \cdot (V_{DD} - V_{DSsat}) \tag{3.56}$$

$$= 2 \cdot \left(V_{DD} - \frac{2}{(g_m/I_D)_B}\right) \tag{3.57}$$

where V_{DSsat} is the minimum drain-to-source voltage that keeps the transistor in

saturation and $(g_m/I_D)_B$ is the transconductance-to-drain current ratio of the bias transistor. The last equation is only valid if the transistor is biased in strong inversion region.

From the analysis presented in the previous section, the phase noise writes:

$$L(\Delta\omega) = \frac{1}{2}\left(\frac{\omega_0}{2Q_T\Delta\omega}\right)^2 \cdot \frac{4kTRF}{V_{RMS}^2} \tag{3.58}$$

$$= \frac{1}{2} \cdot \frac{kTF}{V_{RMS}^2}\frac{Z_0}{Q_T}\left(\frac{\omega_o}{\Delta\omega}\right)^2 \tag{3.59}$$

with the noise factor F given by (3.39). From this equation, it is clear that high Q_T, high V_{RMS} and low F are desirable. As $V_{RMS} = \frac{\sqrt{2}}{\pi} I_B Q_T Z_0$ in current limited regime and with the help of (3.53), we can re-write (3.59) as follows:

$$L(\Delta\omega) = \frac{kT\pi^2}{16}\frac{F}{\alpha^2}\frac{Z_0}{Q_T}\left(\frac{g_m}{I_D}\right)^2\left(\frac{\omega_o}{\Delta\omega}\right)^2 \tag{3.60}$$

$$= \frac{kT\pi^2}{8}\left(\frac{F}{\alpha}\right)\frac{1}{Q_T^2}\frac{(g_m/I_D)}{I_B}\left(\frac{\omega_o}{\Delta\omega}\right)^2 \tag{3.61}$$

Despite (3.61) seems to indicate that the phase noise is to the first order independent on Z_0, there is an optimum inductance value for a given ω_o. Indeed, the R should be maximized, or from (3.47), the product LQ_L should be maximized. The optimization for low phase noise thus depends on the LG product, as explained in [3]. On the other hand, a large start-up gain α increases both the negative resistance noise contribution and the carrier power. The noise attributed to the negative resistance then dominates and $L(\Delta\omega) \sim \frac{P(g_m/I_D)}{I_B Q_T^2}$. On the other hand, for a given I_B, the higher the Z_0, the higher the V_{RMS} and α, or the lower the g_m/I_D, which is desirable for low phase noise. But low g_m/I_D also results in low α (see (3.53)). The g_m/I_D should be determined by numerical simulation because of the inter-dependence of the parameters in the previous relationships. Furthermore, the transconductance is linearized for high Z_o, which results in a decrease of the oscillation amplitude for the same bias current I_B. To decrease the phase noise in current limited regime, the bias current should be high to maximize V_{RMS}.

To get high spectral purity and low power consumption, we should minimize the product $L(\Delta\omega)P_d$ ($P_d = I_B \cdot V_{DD}$ is the dissipated power). We should then keep the voltage supply as low as possible, but it is related to the current (see (3.54)) and the output power level. The voltage supply is besides often determined by the considered technology.

Including the trade-off between phase noise and dissipated power, the following

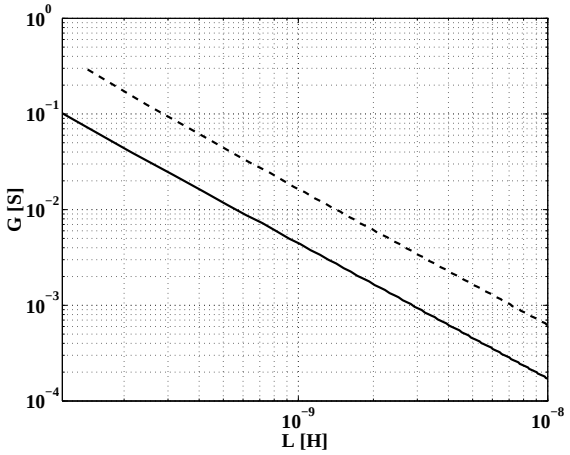


Figure 3.17: Equivalent parallel losses G of the inductors versus the inductance for designs at 5 GHz (straight line) and 2.5 GHz (dashed line).

figure of merit FOM is defined in order to compare the performances of different oscillators:

$$FOM = 10\log_{10}\left(\left(\frac{\omega_o}{\Delta\omega}\right)^2 \frac{1}{L(\Delta\omega)P_d}\right). \tag{3.62}$$

3.3.4 Design strategy

The design of oscillators for wireless communications consists in the determination of appropriate values for I_B, L, C and transistor sizes so that power, frequency and phase noise requirements are achieved. Further specifications are set on the tuning range in the case of VCOs.

From the above analysis, we implemented design methodologies for the topology of Fig. 3.14(a) that meets these specifications. We suppose that the circuit is perfectly symmetric i.e., $L_1 = L_2$ and $C_1 = C_2$. The procedure starts from a given inductor (i.e., the inductor value $L_1 = L_2$ and Q_L are known). The model presented in reference [52] was used to find the inductor parameters. Limitations on the chip area may be introduced in the design flow as this model gives the physical sizes of the inductors[3]. Among the different elements in the resonator, the inductor Q_L has the poorer quality factor. It was thus assumed for sake of simplicity that Q_L equals the tank quality factor $Q_T = Q_L$. The tank losses obtained from that model are plotted in Fig. 3.17 as a function of the inductance value. In order to find the best inductor, the procedure described hereafter has to be re-iterated for each assumed inductor.

In the following, two design methodologies are provided. The goal is to give

[3]Inductors are the most area consuming devices in VCOs (see for instance Fig. 3.36).

an accurate view of the trade-offs. A last optimization by numerical simulations is nevertheless required as the proposed methodology only gives near-optimum solution for a given set of specifications. The first method is based on the set of equations of section 3.3.3. The second one is more general and accurate as the amplitude of the signal -which is determinant for the phase noise calculation-, is evaluated by the describing functions. Both methods make use of the unique relationship between the normalized drain current and the transconductance over drain current ratio g_m/I_D of a MOSFET. This relationship has been shown to be very effective in the design of amplifiers [53].

In the first method [54], we force the oscillator to operate at the limit between the current- and voltage-limited regimes. From the previous analysis, this bias corresponds to a minimum of phase noise (if LG increases with L [3]). The g_m/I_D of the FETs in the differential pair is found by (3.54) and (3.57):

$$\frac{g_m}{I_D} = \frac{2\alpha}{\pi} \frac{1}{V_{dd} - \frac{2}{(g_m/I_D)_B}} \tag{3.63}$$

where the start-up gain α and the bias of the current source transistor ($(g_m/I_D)_B$) are inputs of the method. Then the current consumption I_B is obtained through (3.52):

$$I_B = \frac{2\alpha G}{g_m/I_D} \tag{3.64}$$

The width-to-length (W/L) ratio of the transistor are given by the ratio of I_B to the normalized current corresponding to the g_m/I_D value.

A numerical example in the 0.25 μm PD technology from LETI is given in Fig. 3.18. For that technology, the g_m/I_D versus $I_D/(W/L)$ curve is depicted in Fig. 3.19. The other parameters required to obtain Fig. 3.18 is given as follows. The transconductance-to-drain current ratio of the current source transistor $(g_m/I_D)_B$ was fixed to 5. The noise parameters P and P_B were taken constant and equals to 1.5 and 1, respectively [42]. The voltage supply is 1.2 V. The desired frequency of oscillation is 5 GHz.

When the width-to-length ratio is chosen, the gate length is then determined by assuring that f_o is lower than the maximal oscillation frequency of the MOSFETs (f_{max}). The size of the transistor will also determine the part of the tank capacitance due to the MOSFETs parasitic. The extra-capacitance is finally obtained from the oscillation frequency and desired tuning range. The process must be re-iterated by adjusting the tank quality factor including the losses in the capacitance (see (3.50)) for more accuracy.

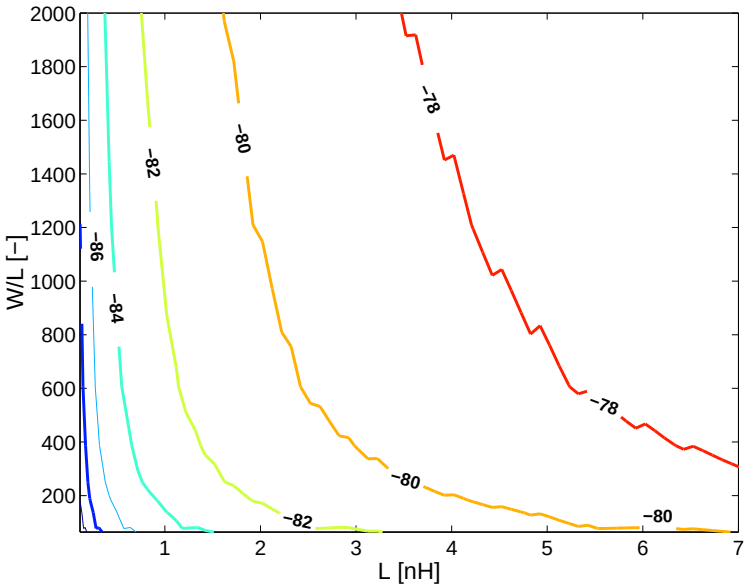


Figure 3.18: Iso-phase noise curves in the inductance / width-to-length ratio plane. LETI 0.25 μm PD SOI technology; $V_{dd} = 1.2$ V; $P = 1.5$; $P_B = 1$; $(g_m/I_D)_B = 5$; $f_o = 5$ GHz.

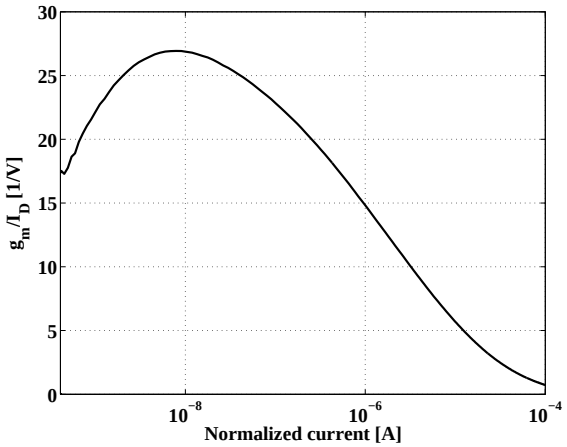


Figure 3.19: Transconductance-to-drain current ratio versus normalized current of the considered PD MOSFET (LETI technology, 12x6.6 μm/0.25 μm); $V_D = 1.1$ V.

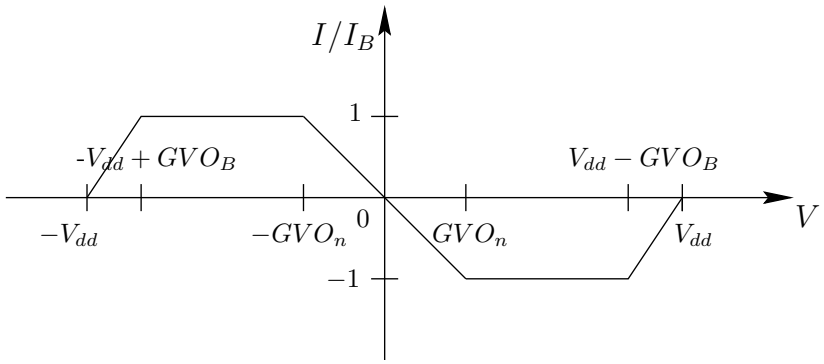


Figure 3.20: Piecewise approximation of the normalized transconductor.

In the second method, it is no more assumed that the circuit operates at the corner between current- and voltage-limited regimes. The trade-offs are thus more completely described and optimization for low power consumption is possible.

The method starts by the approximation of the transconductor by a piecewise. More exactly, the $I - V$ relationship is normalized to the bias current I_B and described as in Fig. 3.20. We further assume the that the transistors are in strong inversion so that the gate-voltage overdrive GVO is $GVO = \frac{2}{g_m/I_D}$. The normalized gain of this amplifier g_A is then calculated from that curve as a function of the signal amplitude A by relation (2.59). To make the oscillation possible, the gain of the transconductor should equals the tank losses G_T:

$$g_A\left(\frac{g_m}{I_D}, A\right) = \frac{G_T}{I_B} \tag{3.65}$$

For a given g_m/I_D, the relationship between the signal amplitude A and the current bias I_B is thus obtained. The width-to-length ratio of the transistor are given by the ratio of I_B to the normalized current corresponding to the chosen g_m/I_D. The phase noise is finally easily calculated by (3.59). All the trade-offs might then be discussed graphically.

This method was applied to a 0.25 μm PD SOI technology (same example as for the first design method). The corresponding g_m/I_D versus $I_D/(W/L)$ curve is shown in Fig. 3.19. The transconductance-to-drain current ratio of the current source transistor $(g_m/I_D)_B$ was fixed to 5. The noise parameters P and P_B were taken constant and equals to 1.5 and 1, respectively [42]. The voltage supply is 1.2 V. The desired frequency of oscillation is 5 GHz. Using these parameters, the gain g_A is plotted in Fig. 3.21. The noise factor F is calculated from (3.39) and depicted in Fig. 3.22. The second term in (3.39) splits the curves in Fig. 3.22 in three part. This term is equals to P when the amplitude of the signal is given

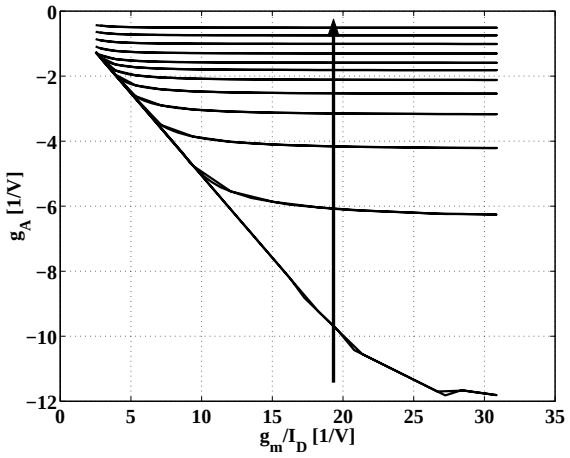


Figure 3.21: Gain of the transconductor normalized by I_B as a function of the g_m/I_D ratio and signal amplitude. The arrow indicates an amplitude increase (A from 0.1 V to 1.9 V by 0.2 V step).

by (3.15) (middle part of Fig. 3.22). However, for amplitudes smaller than GVO_n, the model considers a linear gain $g_m/2$, and the ratio of (3.15) to A increases (first part in Fig. 3.22). The same mechanism appears at high values of A when the voltage limited mode is reached (third part in Fig. 3.22). There is thus a value of bias current that minimizes the phase noise (Fig. 3.23), as explained in [27]. In this example, the main contribution comes from the transconductor noise, as from 3.39, the contribution from the tank losses is normalized to unity. The contribution of the tail transistor is almost independent of the bias and from that model, it does not depend on the signal amplitude. However, it depends on the tank quality factor. In the example in Fig. 3.22, its contribution is however smaller than the contribution from the tank losses. Fig. 3.24 shows the impact of g_m/I_D and the inductor on the phase noise while Fig. 3.25 gives the consumption and phase noise as a function of the sizes of the MOSFETs.

Our second method presents some advantages. First, both current- and voltage-limited mode of operation are considered in a single take. Second, it gives the range g_m/I_D or equivalently the transistor sizes, that makes the circuit oscillates or not, for a given inductor and bias. Furthermore, the start-up gain α is a consequence of the desired tank amplitude and no more a design parameter.

3.4 Tunable capacitors

In modern telecommunications systems, large and accurate tuning ranges are two desirable features. A large tuning is obtained through a bank of digitally switched capacitors [55], while a fine tuning is realized inside a Phase-locked loop (PLL) by

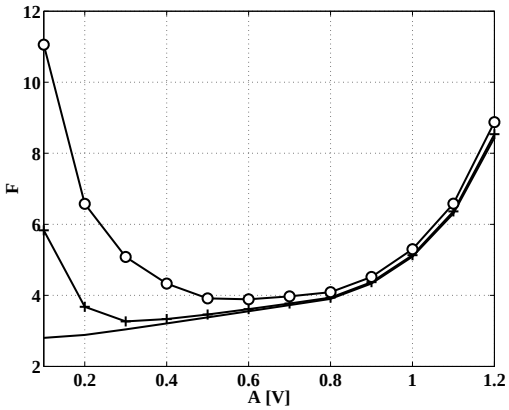


Figure 3.22: Noise factor F as a function of the signal amplitude. $L = 1$ nH. Straight line: $g_m/I_D = 27.2$ V^{-1}; plus sign: $g_m/I_D = 8.7$ V^{-1}; circle sign: $g_m/I_D = 4.2$ V^{-1}.

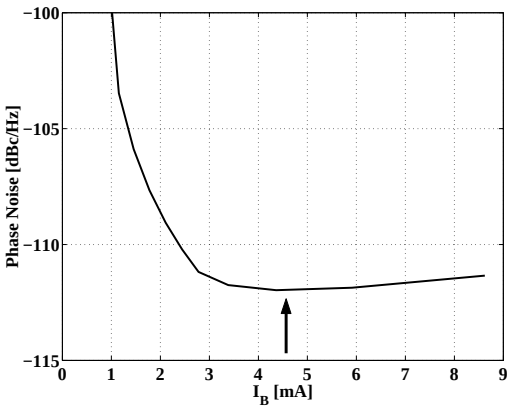


Figure 3.23: Phase noise at 600 kHz offset as a function of the bias current. The arrow indicates the optimum current. $L = 1$ nH; $g_m/I_D = 8$.

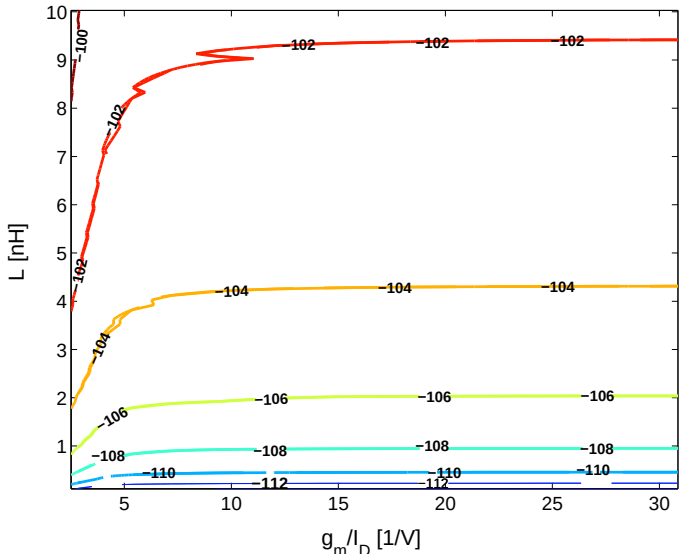


Figure 3.24: Iso-phase noise curves in the g_m/I_D-inductance plane for a fixed signal amplitude ($A = 0.5$ V).

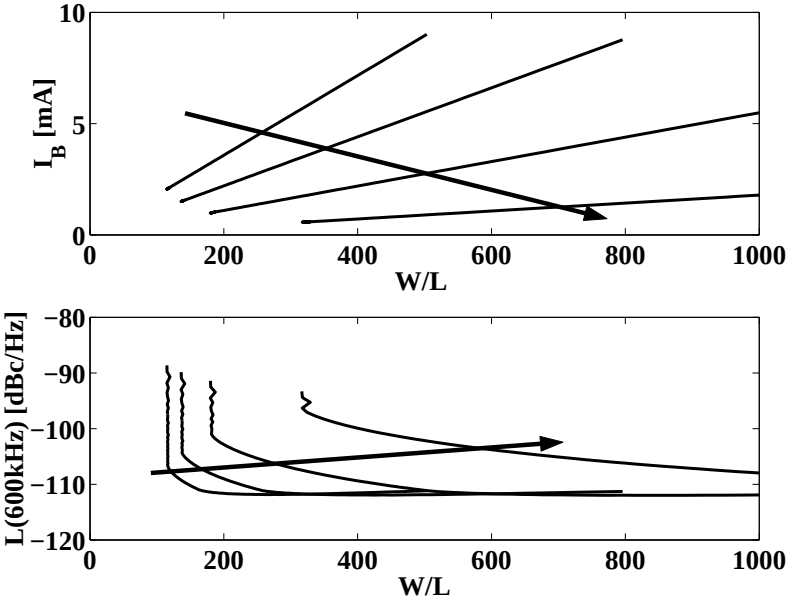


Figure 3.25: Impact of the sizing on the current consumption and phase noise as a function of the g_m/I_D. The arrows indicate a g_m/I_D increase; the plotted g_m/I_D are $4.2, 5.7, 8.7, 15.3, 27.2$. $L = 1$ nH.

a single varactor. The large tuning may be used to skip from one frequency band to another in multi-standards applications.

Large tuning Large tuning is generally done by digitally switching with MOSFETs, metal-insulator-metal (MIM) capacitors that are connected in parallel to the resonator. It was reported in [55] that the tank quality factor was affected by the use of a bank of digitally switched varactors or capacitors, due to the leakage current in the switches. To avoid this, one may take advantage of the emerging micromachining technology. Indeed, no DC current is flowing through a MEMS tunable capacitor. Furthermore, the devices have high Q, making them a candidate of choice for RF applications (Q values of 60 at 1 GHz has been demonstrated in [56] while in [57], the Q reached 660 at 30 GHz). As it will be explained in chapter 4, we believe that MEMS technology is more adapted for the large tuning than for the fine tuning of oscillators. However, reliability and cost will decide the designer to chose between standard CMOS or MEMS technology for the implementation of his/her VCO. Technical details on micromachining will be given in the next chapter of this work.

Fine tuning The following of this section concerns fine tuning only. A reverse-biased diode junction capacitance may be used to implement the varactor but it gives limited tuning range and poor quality factor Q. Therefore, to achieve higher

performances, a MOS varactor in accumulation mode (AMOS) is commonly preferred [58].

SOI technology offers several advantages over bulk technology: no junctions lowering the quality factor Q and injecting substrate noise into the resonator; device symmetry for differential tuning; and a reduction of the parasitic capacitances, increasing the VCO tuning range [58, 59].

The varactors sensitivity k_v is described by the following relation:

$$C(V_{GS}) = C_o + k_v \cdot (V_{GS} + v_n) \tag{3.66}$$

where C_o is the zero bias capacitance, V_{GS} the control voltage and v_n the noise voltage that is modulated by k_v. As we will show in section 3.4.2, this relation depends on the amplitude A of the input signal, k_v decreasing as A increases. Since in a VCO, the varactor is driven by a large signal, it is crucial to understand and to model this behavior. Indeed, k_v can directly impact the VCO voltage-to-frequency characteristics and the phase noise, as well as the PLL stability [60, 61].

Phase noise related to a varactor As stated before, the tail current I_B has the main contribution to phase noise. This contribution is proportional to the square of the sensitivity $\frac{d\omega_o}{dI_B}$ of the oscillation frequency change to a change in I_B. The bias current influences the oscillation frequency by two ways [50]. Firstly, the common-mode level is modulated. Differential tuning techniques [62] have been used to improve the common-mode noise rejection. Secondly, the amplitude modulation due to the bias current converts to frequency modulation in the varactor. The part of the sensitivity $\frac{d\omega_o}{dI_B}$ related to this mechanism is proportional to k_v. In other words, the higher the varactor sensitivity (i.e., the steeper the tuning curve), the higher the phase noise.

The trade-off between tuning range and spectral purity is more crucial in new technologies, as the lowering of the supply voltage goes with the scaling of the MOS transistors to deep-submicron. As consequence, the control voltage of the varactors is limited to low values, requiring a large varactor sensitivity k_v in order to maintain enough tuning range.

This section provides a complete analysis of the evolution of k_v between small- and large- signal regimes and its consequences to the VCO design. Furthermore, several SOI devices were investigated, showing that zero-Vt accumulation-mode varactors have symmetrical characteristics and high Q.

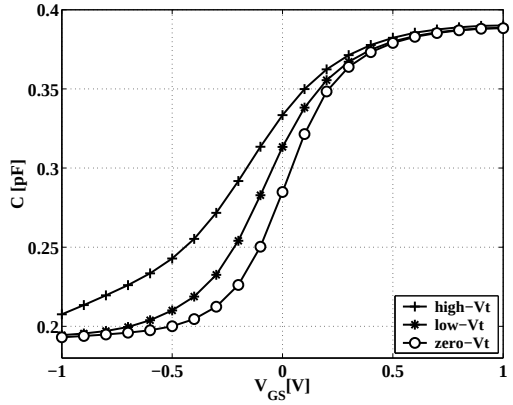


Figure 3.26: C-V curves for accumulation-mode varactors with high-V_{th}, low-V_{th} and zero-V_{th} channel doping levels at 5 GHz.

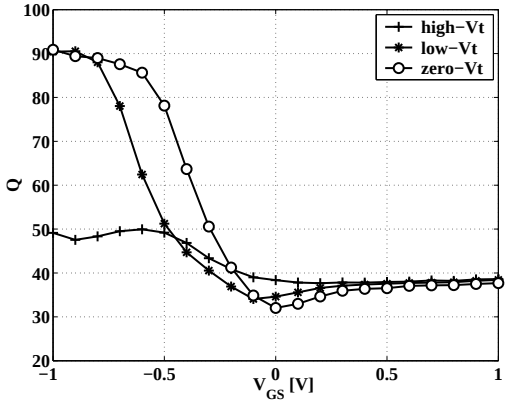


Figure 3.27: Quality factor of accumulation-mode varactors with high-V_{th}, low-V_{th} and zero-V_{th} channel doping level as a function of V_{GS} at 5 GHz.

3.4.1 Varactor small-signal characterization

Accumulation-mode N+/N-/N+ devices, offering the best varactor performances [59], have been designed using the high-V_{th}, low-V_{th} and zero-V_{th} channel doping levels of the 0.15 μm Fully-Depleted SOI CMOS process from OKI. Zero-V_{th} transistors are realized with the very low SOI intrinsic doping ($10^{15} - 10^{16}$ cm^{-3}). So low channel doping levels are not achievable in bulk [63]. The silicon film and gate oxide thicknesses are 40 nm and 2.5 nm, respectively. The devices use the process minimum gate length to achieve high Q. Small-signal wide-band frequency measurements have been carried out between 40 MHz and 40 GHz. De-embedded[4] results give the capacitance (Fig. 3.26) and Q (Fig. 3.27) versus the gate-to-source voltage V_{GS} at 5 GHz. For positive gate-to-source voltages V_{GS} (the source is connected to the drain), the devices are in accumulation mode and the Q is dominated by the channel resistance. When V_{GS} is going towards negative values, the zero-V_{th} varactors deplete rapidly, giving a sharp C-V curve, while for higher doping levels the transition is smoother (Fig. 3.26). This gives to intrinsic varactors a better symmetry with respect to V_{GS}, which is beneficial for many circuit applications. The much higher Q of zero-V_{th} and low-V_{th} varactors in the depletion region (Fig. 3.27) can be explained as follows: for these devices, the film totally depletes and the total capacitance is, to the first order, C_{ox} in series with a depletion capacitance, while for high-V_{th} devices a quasi-neutral region remains, adding a resistance in series with the depletion capacitor and degrading the quality factor.

[4]The open de-embedding method was used.

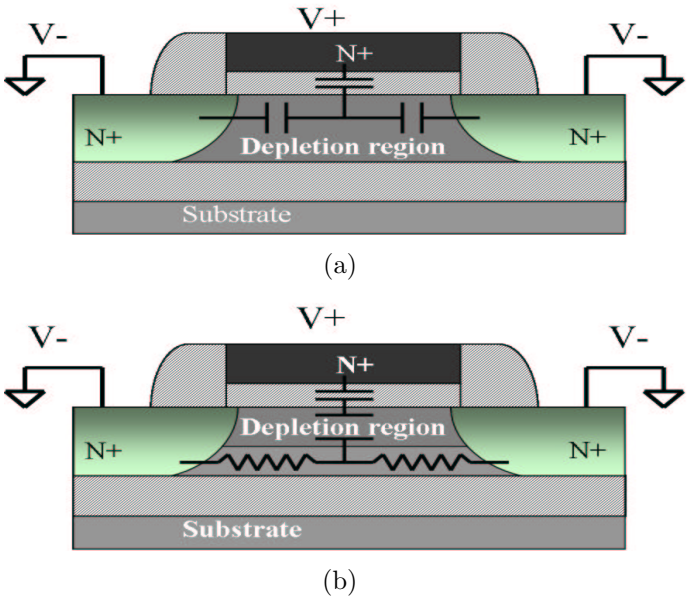


Figure 3.28: Accumulation-mode varactor with negative bias for low- or zero-V_{th} level (a) and for high-V_{th} doping level (b).

3.4.2 Varactor large-signal characterization

The varactor C-V curve was also measured with a Large Signal Network Analyzer (LSNA) at 900 MHz (Fig. 3.29), showing the influence of the amplitude A of the input signal. The derivative of these curves give k_v (Fig. 3.30, straight lines), showing a decrease of k_v with A. In order to evaluate k_v as a function of A without having to make costly LSNA measurements, the describing function (DF) of the nonlinear capacitor can be computed from the small-signal C-V curve plotted in Fig. 3.26. This DF model (Fig. 3.30, crosses) shows an excellent agreement with the measurements (Fig. 3.30, straight lines). On the other hand, we observed a decrease of the quality factor with A, indicating that more energy is dissipated during one cycle. Also, as expected form Fig. 3.29, the maxima of HD_3 lie at higher $|V_{GS}|$ for increasing amplitudes (Fig. 3.31).

3.4.3 Impact on the VCO design

The varactor sensitivity k_v will directly impact the VCO but also the PLL design. Indeed the PLL stability depends on the VCO gain which is directly related to k_v. The decrease of k_v with the signal amplitude will make easier the stabilization of the PLL, assuming that C_{max}/C_{min} does not change for the given V_{GS} range. Moreover, as explained before, the VCO phase noise also depends on k_v, the VCO having the worst phase noise in the middle of the tuning range where k_v is the highest. The

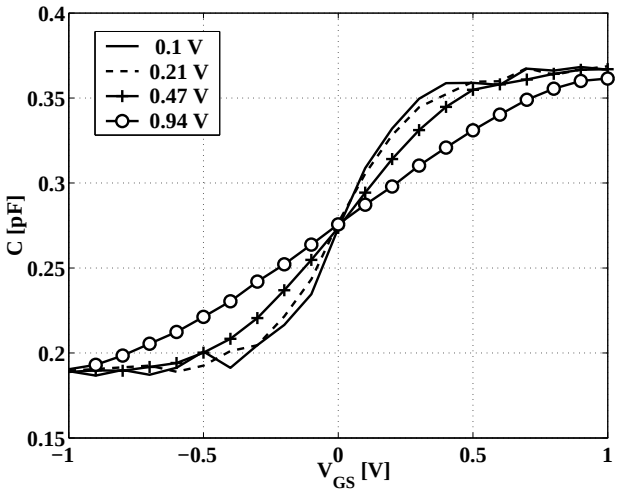


Figure 3.29: Measured varactor (intrinsic doping) tuning curves as a function of the input signal amplitude.

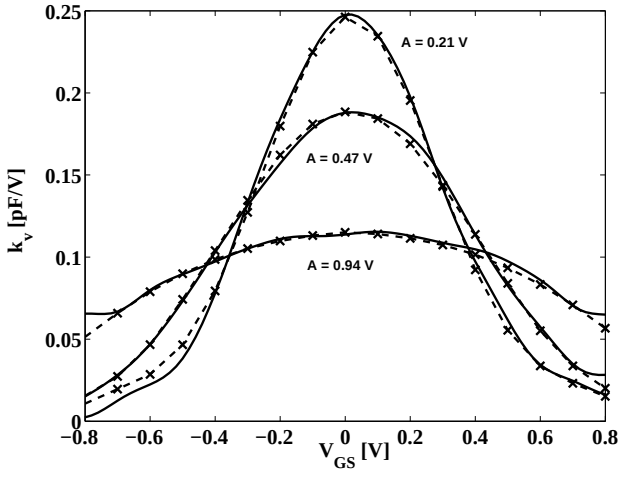


Figure 3.30: Varactor sensitivity versus the control voltage V_{GS} and the signal amplitude. Measurements (-) are compared to the DF model (x).

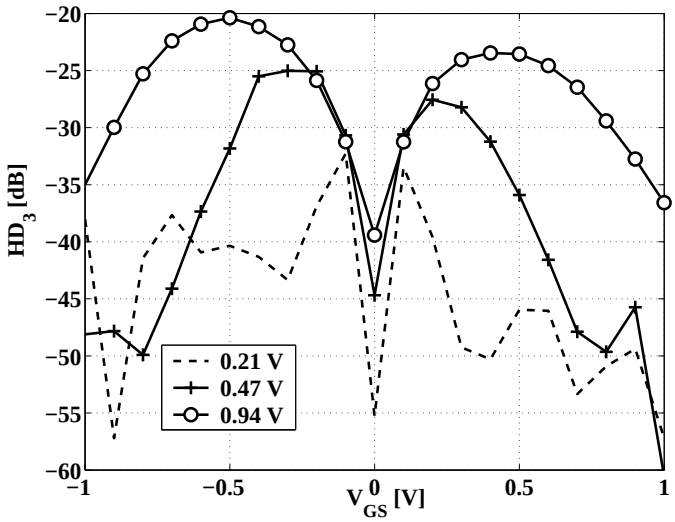


Figure 3.31: Measured HD_3 of the zero-V_{th} varactor as a function of the control voltage V_{GS} and signal amplitude A.

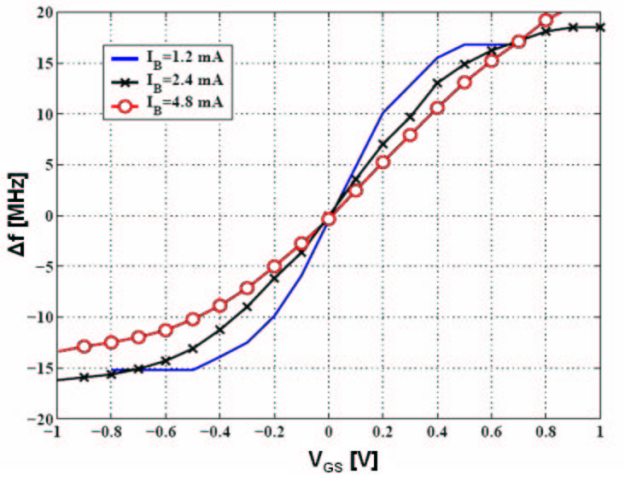


Figure 3.32: Measured VCO variation of frequency versus the control voltage V_{GS}, for different current biases.

fine-tuning curve of a 2.45 GHz LC differential VCO[5] including the characterized varactor was measured as a function of the bias current I_B (Fig. 3.32). As the signal amplitude is proportional to I_B, this plot confirms that, as expected from Fig. 3.29, a smoother variation of the oscillation frequency is obtained for a higher signal amplitude. The measured phase noise at 2.4 mA bias and a 1.3 V supply voltage is -111 dBc/Hz @ 600 kHz. It does not vary significantly over the whole tuning range[6] (Fig. 3.33). This is explained by the small variable capacitance with respect to the overall tank capacitance.

[5]A double differential pair was used, as depicted in Fig. 3.7.

[6]in the validity range of the used experimental setup (see appendix B).

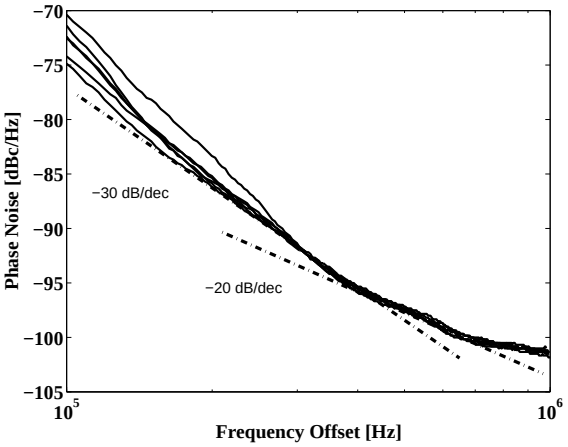


Figure 3.33: Measured phase noise for tuning voltages between 0 and 1 V. The supply voltage is 1 V and the measured core current is 1.2 mA.

3.5 Realized circuits

Oscillators were designed on the 0.25 μm PD SOI technology from LETI, France. Depending on the design, the body of the transistors is either floating or contacted to the source. The connection causes a diminution of the effective gate width. Also, both high- and low resistivity substrates (i.e., $>$ or $<$ 1000 Ωcm, resp.) were used. The two topologies presented in Fig. 3.34 are studied. The nMOS transistors used in these non-optimized designs are composed of 12 fingers connected in parallel, each of 6.6 μm width and 0.25 μm long. The pMOS of the current mirrors have same width-to-length ratio as the transistors from the core, but higher gate length (1 μm) to improve the output conductance. The g_m/I_D of the transistors of the mirror is fixed to 5. It would however be preferable to bias these transistors more deeply in weak inversion. Indeed, it has been shown recently that high g_m/I_D gives better matching between the transistors [64], which is desirable for an accurate control of the bias current. A high g_m/I_D value presents the further advantage to increase the size of the transistors, and thus their capacitance. While this is most of the time avoided in RF designs, this capacitance may contribute to filter the harmonics, which is favorable to decrease the phase noise associated to the tail current noise.

On-wafer measurements were performed with an HP8562A spectrum analyzer (details of the setup are provided in appendix B). It is important to note that it is not possible to measure correctly phase noise close to the carrier with this setup. Moreover, the low sensitivity of the setup does not allow to measure small variations in phase noise. Measurements were rendered even more difficult as on-chip decoupling capacitors were not present at the bias inputs of these non-optimized designs. This results in some cases in relatively large oscillation frequency variation

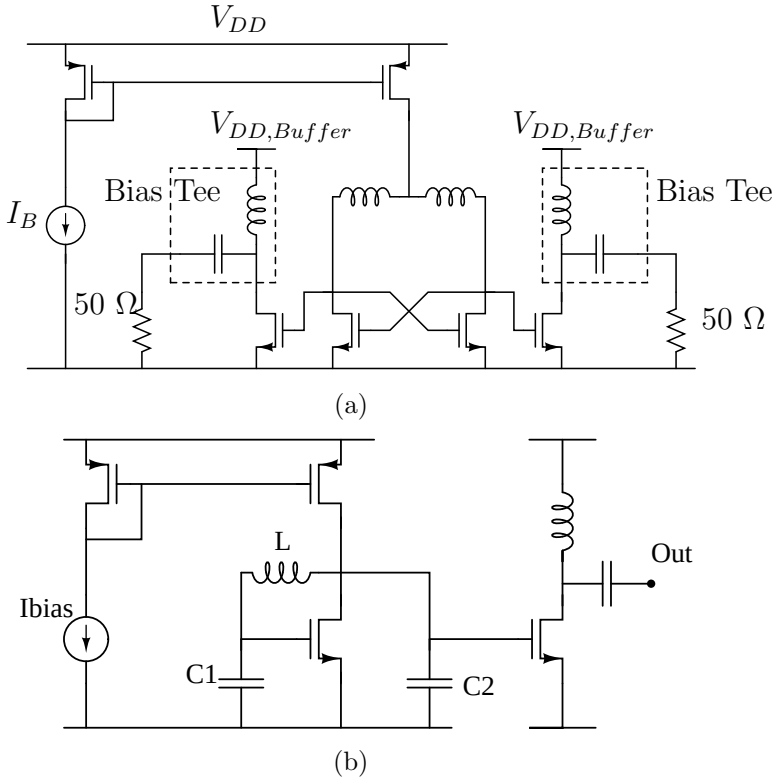


Figure 3.34: Realized circuits. (a) Differential active feedback oscillator. (b) Single-ended Colpitts oscillator.

within one measurement sweep time. Losses in the cables and bias tee are 3 dB at 5.8 GHz and 7 dB at 11 GHz, respectively.

3.5.1 Inductor

Two identical square spiral inductors are used in the tank resonator of the circuit in Fig. 3.34(a). Much attention was paid on the symmetry of the layout of the two inductors. Each inductor has 3.5 turns, 10 μm width, 1 μm spacing between lines, 50 μm inner diameter providing a nominal inductance of 1.6 nH. As already mentioned, since the quality factor of the inductor in CMOS technologies is relatively small, it determines the total tank quality factor. This quality factor is mainly limited by the technology. The quality factor Q of the two inductors in series is reported in Fig. 3.35. The Q is improved by staking several metal layers, as the series resistance is decreased, and by the use of high resistivity substrate, as the shunt conductance is increased.

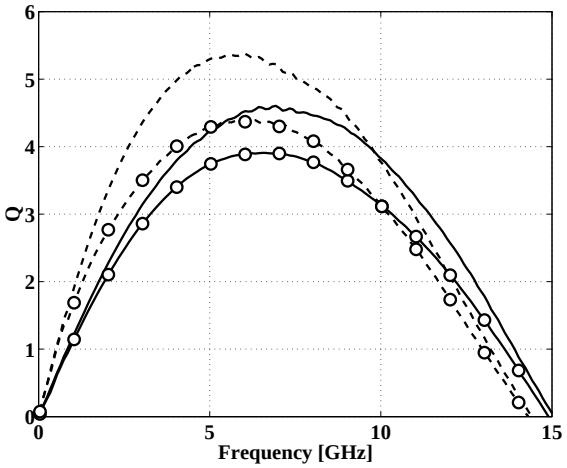


Figure 3.35: Measured inductor Q. Symbols: low resistivity substrate, no symbols: high resistivity substrate; straight lines: one metal layer, dashed lines: two metal layers.

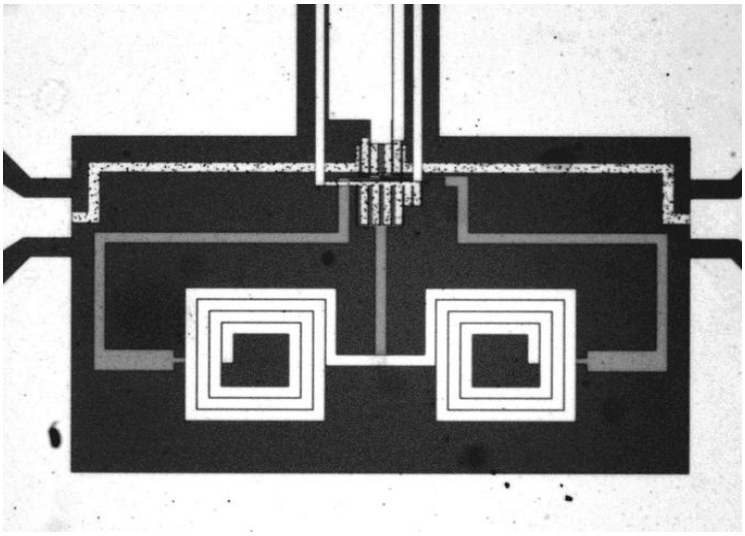


Figure 3.36: Die photograph of the active feedback oscillator.

3.5.2 Oscillators

The differential active feedback topology is first investigated. A special care was taken to ensure high layout-symmetry to avoid any additional asymmetry in the oscillation waveform. The tank capacitance consists of only the MOSFETs parasitic capacitances. Fig 3.36 shows a die micrograph of the circuit.

Performances The circuit was designed to oscillate at 5.8 GHz. When the oscillator is biased with 1.4 V supply voltage, it drives 1.48 mA and has a phase noise of -101 dBc/Hz at 1 MHz offset from the carrier which is in agreement with equation (3.61). In that case, the FOM reaches a value of 202.5. This design is able to start and oscillate with only 1.2 mA under a 0.8 V voltage supply, at the price of

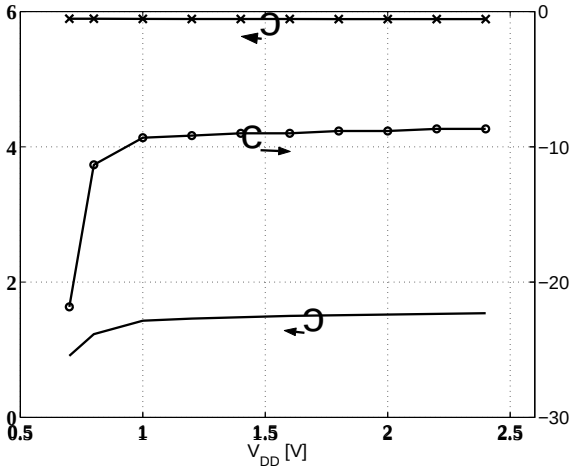


Figure 3.37: Supply pulling. Crosses: oscillation frequency [GHz]; circles: output power [dBm]; straight line: DC core current [mA].

a phase noise degradation. Furthermore, increasing the supply voltage above 1.2 V does not lead to any change in the oscillation frequency neither in the output power (Fig. 3.37). On the other hand, a decrease of the oscillation frequency was observed at high current value (Fig. 3.38). This can be explained by two different mechanisms. First, the bias current I_B determines the operating point of the transistor capacitances used in the resonator and the tank amplitude that acts in turn on the capacitance value, as explained previously. Second, because the amplitude of the signal grows with I_B, the harmonic content increases. The influence of the content of harmonics on the frequency variation was studied by Groszkowski [65]: If the oscillations are purely sinusoidal, the energy distribution between the L and C arms in the resonator is equal. When harmonics appear, the currents corresponding to them flow chiefly through the low-impedance C arm. Therefore, they increase the electrostatic energy of this arm in comparison with the arm L. In order to keep the energy equals in both arms, the fundamental frequency must slightly diminish itself. The current in arm L then increases slightly, in order to allow its electromagnetic energy to increase suitably.

To achieve higher frequency of oscillation (> 10 GHz), high impedance transmission lines were used to obtain sufficient inductance in the resonator. Fig. 3.39 reports the measured phase noise of this circuit. The phase noise is relatively high compared to the circuits found in the literature. This is attributed to the fact that the DC accesses were not properly decoupled in this non-optimized design. A large amount of supply noise is then upconverted.

Table 3.5.2 compares FOM of various integrated giga-Hertz CMOS oscillators using LC tank presented recently in the literature. This shows that our design

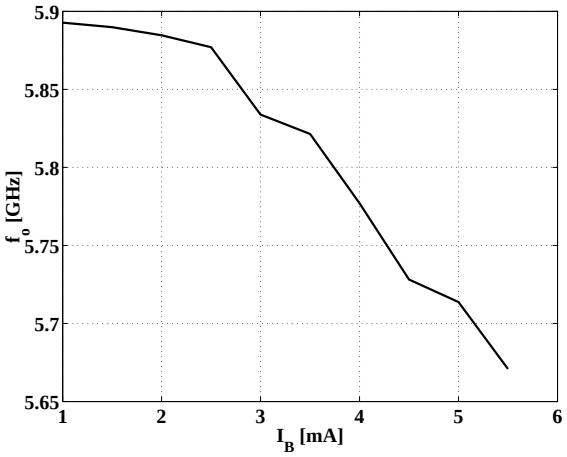


Figure 3.38: Oscillation frequency decrease as a function of the bias current. $V_{dd} = 1.4$ V.

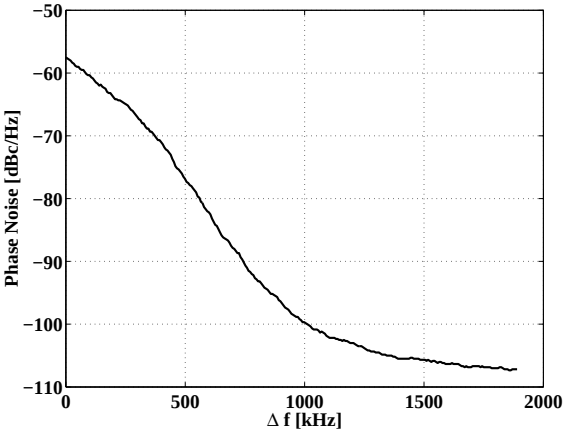


Figure 3.39: Measured phase noise versus frequency offset. $f_o = 11.2$ GHz, $V_{DD} = 1.5$ V, $I_B = 1.5$ mA.

Table 3.1: Comparison of performance of CMOS oscillators recently published.

Reference	Frequency [GHz]	Dissipated power [mW]	FOM	Supply Voltage [V]	Technology
[66]	9.8	12	187	1.3	0.35 μm, bulk
[7]	1.8	9	182.5	3.0	0.7 μm, bulk
[9]	5.2	10.8	174	2.7	0.35 μm, bulk
[58]	40	11.25	178.5	1.5	0.13 μm, SOI
[11]	50	13	181.8	1.3	0.25 μm, bulk
[67]	10.55	28	200.4	1.5	0.35 μm, bulk
[67]	5	9	201.7	1.5	0.35 μm, bulk
This work	5.8	2.1	202.5	1.2	0.25 μm, SOI
This work (Colpitts)	11	1.85	188.1	1.0	0.25 μm, SOI

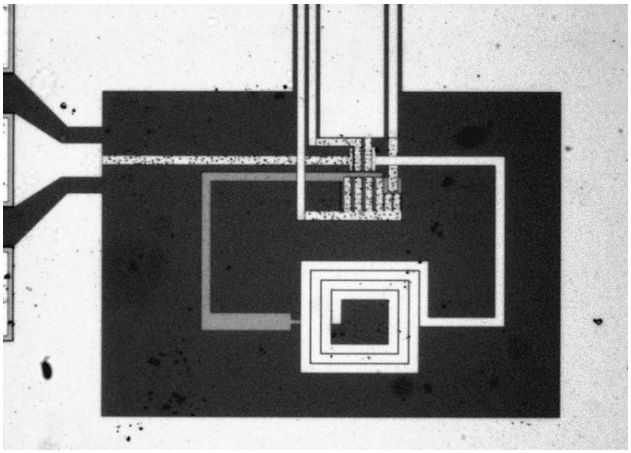
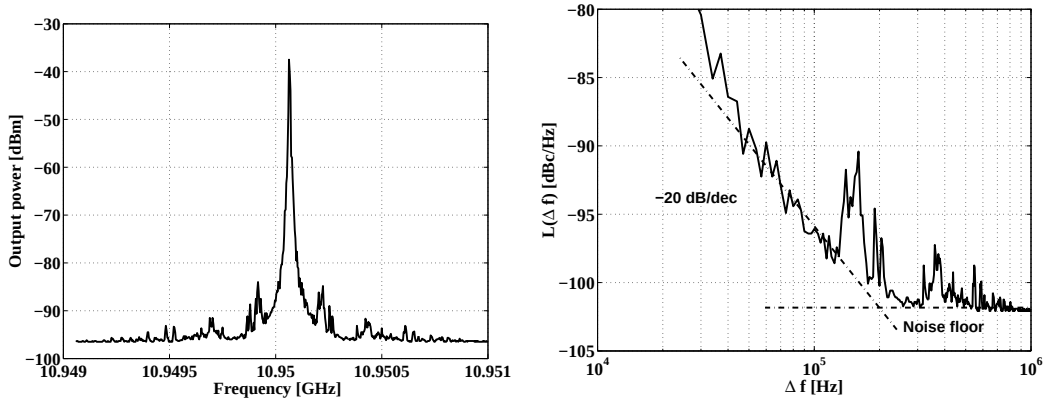


Figure 3.40: Die photograph of the Colpitts oscillator.

presents interesting oscillation behavior under very low power consumption without degrading the FOM. SOI is then well-appropriated for low-power designs. Low-voltage operation is possible thanks to low threshold voltage of SOI transistors (V_{th} = 0.4 V in the present case). For comparison purpose, we added in this table the results obtained for a 11 GHz Colpitts single-ended oscillator. It was fabricated in the same technology (schematic in Fig. 3.34(b) and micrograph in Fig. 3.40). The measured spectrum and associated phase noise is presented in Fig. 3.41.

Impact of the technology Even if the quality factor of the resonator is improved by the use of high-resistivity substrate (Fig. 3.35), our measurements did not indicate significant improvements in phase noise in this case.

Furthermore, circuits using floating-body and body-tied PD devices showed comparable amount of phase noise. However, phase noise degradation is expected at low frequency, for the circuit using FB devices, because of the kink in its DC



(a) (b)

Figure 3.41: Output of the Colpitts oscillator. Voltage supply is 2.5 V; $I_{bias} = 1.6mA$; Resolution bandwidth is 100 Hz. (a) spectrum. (b) phase noise.

characteristic. Even if the kink effect can be suppressed at high frequency operation [68] as shown in the previous chapter, it does affect the low-frequency noise because the low frequency components of the signal experience high source-body junction impedance. This induces a low-frequency noise overshoot superimposed on the "pure" 1/f noise [69]. After upconversion, this leads to phase noise overshoot compared to BT PD or fully-depleted devices. We were not able to highlight this effect due to the restrictions imposed by our experimental setup. In design of SOI integrated oscillators, better phase noise response is thus expected for circuits that use body-tied transistors at the cost of extra capacitance.

3.6 Conclusions

The design of low-power integrated VCOs for wireless communications remains today a challenging task. There is indeed a need for low-power and low-phase noise circuits. In this chapter, the oscillators were carefully studied in order to improve their performances. The example of a differential active feedback LC oscillator was used through the chapter, but most tools apply to other circuit architectures.

An oscillator is intrinsically a nonlinear circuit that provides at the output a large (sinusoidal) signal. Linear models of oscillators lack therefore to predict some properties, among which the oscillation amplitude. An accurate evaluation of the signal amplitude is necessary to estimate correctly both the output power and the phase noise. Sophisticated nonlinear models were previously developed to face the drawbacks for the linear ones. Unfortunately, because they lack from insight, they are not widely used by circuit designers. In order to predict the output amplitude,

while maintaining the simplicity of linear modeling, we proposed to use a quasi-linear approximation of the nonlinearity in the oscillator. The describing functions formalism was used for that purpose, as it gives the best amplitude-dependent linear approximation of a nonlinearity to the mean squares sense. It is a powerful tool for the analysis and the design of oscillators.

We introduced in the beginning of this chapter the basic features of oscillators with linear models. Simple analytical tools were provided in order to understand the different mechanisms in LC oscillator and to give insight of the main trade-offs between the design parameters. Several methods to determine the oscillation amplitude were discussed. We showed that the describing function formalism is therefore a simple, accurate and efficient tool. It was also demonstrated how phase noise depends on the tank quality factor and the power consumption. Based on a nonlinear analysis, it was explained how the signal amplitude affects the oscillator noise factor.

Following this general analysis, there are two ways for improving the performances.

Firstly, *technological efforts* improve the tank quality factor. The potentiality of SOI technology for microwave oscillators was investigated. First non-optimized realizations showed that, for a given figure of merit FOM, SOI is the adequate technology for the design of low-power low-voltage integrated microwave oscillators. Tunable capacitors were investigated in details. Capacitors in SOI technology are considered in this chapter while micromachined devices are investigated in chapter 4. Fully-depleted accumulation-mode varactor with intrinsic doping provides best performances, as it has a symmetrical tuning curve and a high quality factor. The varactor sensitivity is another important parameter since it impacts both the phase noise and the oscillator frequency tuning. Large-signal RF characterization showed that the varactor sensitivity varies with both the signal amplitude and the control voltage. This effect may be modeled accurately from the describing function of the varactor small-signal tuning curve.

Secondly, *design efforts* improve the performances by tuning optimally the different parameters. Two design strategies were developed based on the linear and the nonlinear analysis of the oscillator. In the first methodology, the oscillator works at the limit between the current- and the voltage-limited regimes since for a given inductor, the best phase noise performance is expected at this bias. The second methodology is independent of the mode of operation of the oscillator and is therefore more general. This method is also more accurate since the evaluation of the

signal amplitude by the describing function is embedded in the methodology. Moreover, it gives the range of bias and sizing that makes the circuit oscillates or not. However, the methods have still to be validated experimentally.

References

[1] Peter Kinget, "Integrated GHz voltage controlled oscillators," in *Analog Circuit Design*, Willy Sansen, J. Huijsing, and R. van de Plassche, Eds., pp. 353–381. Kluwer Academic Publishers, 1999.

[2] Tae Youn Kim, A. Adams, and N. Weste, "High performance SOI and bulk CMOS 5 GHz VCOs," *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*, pp. 93 – 96, 8-10 June 2003.

[3] Donhee Ham and Ali Hajimiri, "Concepts and methods in optimization of integrated LC VCOs," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 6, pp. 896–909, June 2001.

[4] Nhat M. Nguyen and Robert G. Meyer, "Start-up and frequency stability in high-frequency oscillators," vol. 27, no. 5, pp. 810–820, May 1992.

[5] Qiuting Huang, "Phase noise to carier ratio in LC oscillators," *IEEE Transactions on Circuits and Systems—Part I: Fundamental Theory and Applications*, vol. 47, no. 7, pp. 965–980, July 2000.

[6] HongMo Wang, "A solution for minimizing phase noise in low-power resonator-based oscillators," *IEEE International Symposium on Circuits and Systems*, pp. III–53–57, May 2000.

[7] Jan Craninckx and M.S.J. Steyaert, "A 1.8-GHz CMOS low-phase noise voltage-controlled oscillator with prescaler," *IEEE Journal of Solid-State Circuits*, vol. 30, no. 12, pp. 1474–1482, December 1995.

[8] Ahmadreza Rofougaran, Jacob Rael, Myriam Rofougaran, and Asad Abidi, "A 900 MHz CMOS LC-oscillator with quadrature outputs," *IEEE International Solid-State Circuits Conference*, pp. 392–393, February 1996.

[9] Peter Kinget, "A fully integrated 2.7 V 0.35 μm CMOS VCO for 5 GHz wireless application," *IEEE International Solid-State Circuits Conference*, 1998.

[10] Behzad Razavi, "A 1.8 GHz CMOS voltage-controlled oscillator," *IEEE International Solid-State Circuits Conference*, pp. 388–389, 1997.

[11] HongMo Wang, "A 50 GHz VCO in 0.25μm CMOS," *IEEE International Solid-State Circuits Conference*, 2001, paper 23.8.

[12] Ali Hajimiri and Thomas H. Lee, "Design issues in CMOS differential LC oscillators," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 5, pp. 717–724, May 1999.

[13] Jae-Hong Chang, Yong-Sik Youn, Mun-Yang Park, and Choong-Ki Kim, "A new 6 GHz fully integrated low power low phase noise CMOS LC quadrature VCO," *IEEE International Microwave Symposium Digest*, pp. 295–298, 2003.

[14] HongMo Wang, "Comments on Design issues in CMOS differential LC oscillators [and reply]," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 2, pp. 286–287, February 2000.

[15] Y. Hu, J.J. Obregon, and J.-C. Mollier, "Nonlinear analysis of microwave FET oscillators using Volterra series," *IEEE Transactions on Microwave Theory and Techniques*, vol. 37, no. 11, pp. 1689 – 1693, November 1989.

[16] E. Vidal, A. Poveda, and M. Ismail, "Describing functions and oscillators," *IEEE Circuits and Devices Magazine*, vol. 17, no. 6, pp. 7 – 11, November 2001.

[17] Donald O. Pederson and Kartikeya Mayaram, *Analog Integrated Circuits for Communication: Principles, Simulation and Design*, Kluwer, 1991.

[18] Sebastian K. Magierowski and Stefan Zukotynski, "CMOS LC- oscillator phase-noise analysis using nonlinear models," *IEEE Transactions on Circuits and Systems*, vol. 51, no. 4, pp. 664–676, April 2004.

[19] A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase noise in oscillators: a unifying theory and numerical methods for characterization," *IEEE Transactions on Circuits and Systems—Part I: Fundamental Theory and Applications*, vol. 47, no. 5, pp. 655 – 674, May 2000.

[20] D. Ham and A. Hajimiri, "Virtual damping and Einstein relation in oscillators," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 3, pp. 407 – 418, March 2003.

[21] M. Lax, "Classical noise. V. noise in self-sustained oscillators," *Physical Review*, vol. CAS-160, no. 2, pp. 290–307, 1967.

[22] F.X. Kaertner, "Determination of the correlation spectrum of oscillators with low noise," *IEEE Transactions on Microwave Theory and Techniques*, vol. 37, no. 1, pp. 90–101, January 1989.

[23] Angelo Brambilla, "Method for simulating phase noise in oscillators," *IEEE Transactions on Circuits and Systems—Part I: Fundamental Theory and Applications*, vol. 48, no. 11, pp. 1318–1325, November 2001.

[24] D. Leeson, "A simple model of feedback oscillator noise spectrum," *Proceedings of the IEEE*, vol. 54, pp. 329–330, February 1966.

[25] Behzad Razavi, "A study of phase noise in CMOS oscillators," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 3, pp. 331–343, March 1996.

[26] Ali Hajimiri and Thomas H. Lee, "A general theory of phase noise in electrical oscillators," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 2, pp. 179–194, February 1998.

[27] J.J. Rael and A.A. Abidi, "Physical processes of phase noise in differential LC oscillators," *IEEE Custom Integrated Circuits Conference*, pp. 569–571, 2000.

[28] Ali Hajimiri and Thomas H. Lee, "Corrections to a general theory of phase noise in electrical oscillators," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 6, pp. 928 – 928, June 1998.

[29] G.J. Coram, "A simple 2-D oscillator to determine the correct decomposition of perturbations into amplitude and phase noise," *IEEE Transactions on Circuits and Systems—Part I: Fundamental Theory and Applications*, vol. 48, no. 7, pp. 896 – 898, July 2001.

[30] Carlo Samori, Andrea L. Lacaita, Francesco Villa, and Franco Zappa, "Spectrum folding and phase noise in LC tuned oscillators," *IEEE Transactions on Circuits and Systems—Part II: Analog and Digital Signal Processing*, vol. 45, no. 7, pp. 781–790, July 1998.

[31] S. Ver Hoeye, A. Suarez, and J. Portilla, "Techniques for oscillator nonlinear optimization and phase-noise analysis using commercial harmonic-balance software," *IEEE International Microwave Symposium Digest*, vol. 1, pp. 95 – 98, June 2000.

[32] Beaudoin Tamen, *Modélisation du bruit dans les composants en régime de fonctionnement grand signal: Application la conception de circuits intégrés non linéaires faible bruit pour les télécommunications*, Ph.D. thesis, IEMN, Lille, France, September, 27 2000.

[33] Q. Huang, "On the exact design of RF oscillators," *IEEE Custom Integrated Circuits Conference*, pp. 41 – 44, 11-14 May 1998.

[34] Behzad Razavi, *RF Microelectronics*, Prentice Hall, Upper Saddle River, NJ 07458, 1998.

[35] Jan Craninckx and Michel Steyaert, *Wireless CMOS frequency synthesizer design*, Kluwer Academic Publishers, Boston, 1998.

[36] J.W.M. Rogers, V. Levenets, C.A. Pawlowicz, N.G. Tarr, T.J. Smy, and C. Plett, "Post-processed Cu inductors with application to a completely integrated 2-GHz VCO," vol. 48, no. 6, pp. 1284 – 1287, June 2001.

[37] C.L. Chua, D.K. Fork, K. Van Schuylenbergh, and Jeng-Ping Lu, "Out-of-plane high-Q inductors on low-resistance silicon," *IEEE/ASME Journal of Microelectromechanical Systems*, vol. 6, no. 12, pp. 989 – 995, December 2003.

[38] Thomas H. Lee, *Design of CMOS Radio Frequency Integrated Circuits*, Prentice hall, 1994.

[39] K. Benaissa, Jau-Yuann Yang, D. Crenshaw, B. Williams, S. Sridhar, J. Ai, G. Boselli, Song Zhao, Shaoping Tang, S. Ashburn, P. Madhani, T. Blythe, N. Mahalingam, and H.S. Shichijo, "RF CMOS on high-resistivity substrates for system-on-chip applications," vol. 50, no. 3, pp. 567 – 576, March 2003.

[40] Yeung-Bun Choi, Tee-Hui Teo, and Huailin Liao, "Symmetrical spiral inductor design and optimization for VCO design in 0.35 μm CMOS technology," *ASIC, 2003. Proceedings. 5th International Conference on*, vol. 2, pp. 1066 – 1069, 21-24 October 2003.

[41] Ping Wing Lai, L. Dobos, and S. Long, "A 2.4 GHz SiGe low phase-noise VCO using on chip tapped inductor," *European Solid-State Circuit Conference (ESSCIRC)*, pp. 505 – 508, 16-18 September 2003.

[42] G. Pailloncy, B. Iniguez, G. Dambrine, J.-P. Raskin, and F. Danneville, "Noise modeling in fully depleted SOI MOSFETs," *Solid-State Electronics*, vol. 48, no. 5, pp. 813 – 825, May 2004.

[43] Jean-Paul Eggermont, *Study and Realization of SOI Operational Amplifiers for High-Temperature or High-Frequency Applications*, Ph.D. thesis, Microelectronics Lab., U.C.L., Louvain-la-Neuve, Belgium, 1998.

[44] G. Knoblinger, P. Klein, and H. Tiebout, "A new model for thermal channel noise of deep-submicron MOSFETs and its application in RF-CMOS design," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 5, pp. 831 – 837, May 2001.

[45] O. Llopis, J.B. Juraver, B. Tamen, F. Danneville, M. Chaubet, A. Cappy, and J. Graffeuil, "Nonlinear noise modeling of a PHEMT device through residual phase noise and low frequency noise measurements," *IEEE International Microwave Symposium Digest*, vol. 2, pp. 831 – 834, 20-25 May 2001.

[46] Heinz J. Siweris and Burkhard Schiek, "Analysis of noise upconversion in microwave FET oscillators," *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-33, no. 3, pp. 233–242, March 1985.

[47] N.H.W. Fong, J.-O. Plouchart, N. Zamdmer, Duixian Liu, L.F. Wagner, C. Plett, and N.G. Tarr, "Phase noise improvement of deep submicron low-

voltage VCO," *European Solid-State Circuit Conference (ESSCIRC)*, pp. 811–814, 24 - 26 September 2002, Florence, Italy.

[48] E. Hegazi, H. Sjland, and A. Abidi, "A filtering technique to lower LC oscillator phase noise," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 12, pp. 1921–1930, December 2001.

[49] Pietro Andreani and Henrik Sjland, "Tail current noise suppression in RF CMOS VCOs," *IEEE Journal of Solid-State Circuits*, vol. 37, no. 3, pp. 342–348, March 2002.

[50] S. Levantino, C. Samori, A. Bonfanti, S.L.J. Gierkink, A.L. Lacaita, and V. Boccuzzi, "Frequency dependence on bias current in 5 GHz CMOS VCOs: impact on tuning range and flicker noise upconversion," *IEEE Journal of Solid-State Circuits*, vol. 37, no. 8, pp. 1003–1011, August 2002.

[51] B. De Muer, M. Borremans, M. Steyaert, and G. Li Puma, "A 2-GHz low-phase-noise integrated LC-VCO set with flicker-noise upconversion minimization," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 7, pp. 1034 – 1038, July 2000.

[52] S.S. Mohan, M. del Mar Hershenson, S.P. Boyd, and T.H. Lee, "Simple accurate expressions for planar spiral inductances," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 10, pp. 1419 – 1424, October 1999.

[53] F. Silveira, D. Flandre, and P.G.A. Jespers, "A g_m/i_d based methodology for the design of CMOS analog circuits and its application to the synthesis of a silicon-on-insulator micropower OTA," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 9, pp. 1613–1619, September 1996.

[54] Bertrand Parvais, Michael Goffioul, Danielle Vanhoenacker, and Jean-Pierre Raskin, "Oscillateurs RF intégrés en technologie CMOS/SOI," *Journées Nationales Microondes*, Mai 2001, Poitiers (France).

[55] A. Kral, F. Behbahani, and A.A. Abidi, "RF CMOS oscillators with switched tuning," *IEEE Custom Integrated Circuits Conference*, pp. 555–558, May, 11-14 1998.

[56] D.J. Young and B.E. Boser, "A micromachined variable capacitor for monolithic low-noise VCOs," *Solid-State Sensors and Actuators Workshop*, pp. 86–89, June 1996, Hilton Head.

[57] Nicolas Scott Barker, *Distributed MEMS transmission Lines*, Ph.D. thesis, University of Michigan, Ann Arbor, Michigan, USA, 1999.

[58] N.H.W. Fong, J.-O. Plouchart, N. Zamdmer, Duixian Liu, Wagner, P. Garry, C., and N.G. Tarr, "A 40 GHz VCO with 9 to 15% tuning range in 0.13 μm SOI CMOS," *VLSI Circuits Digest of Technical Papers*, pp. 186 – 189, 13-15 June 2002.

[59] N.H.W. Fong, J.-O. Plouchart, N. Zamdmer, Duixian Liu, L.F. Wagner, C. Plett, and N.G. Tarr, "A 1-V 3.8 - 5.7-GHz wide-band VCO with differentially tuned accumulation MOS varactors for common-mode noise rejection in CMOS SOI technology," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 8, pp. 1952 – 1959, August 2003.

[60] J.W.M. Rogers, J.A. Macedo, and C. Plett, "The effect of varactor nonlinearity on the phase noise of completely integrated VCOs," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 9, pp. 1360 – 1367, September 2001.

[61] R.L. Bunch and S. Raman, "Large-signal analysis of MOS varactors in CMOS -G_m LC VCOs," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 8, pp. 1325 – 1332, August 2003.

[62] M. Tiebout, "Low-power low-phase-noise differentially tuned quadrature VCO design in standard CMOS," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 7, pp. 1018 – 1024, July 2001.

[63] M. Harada, T. Tsukahara, and J. Yamada, "0.5-1 V 2 GHz RF front-end circuits in CMOS/SIMOX," *IEEE International Solid-State Circuits Conference*, pp. 378 – 379, February, 7-9 2000.

[64] L. Vancaillie, F. Silveira, B. Linares-Barranco, T. Serrano-Gotarredona, and D. Flandre, "MOSFET mismatch in weak/moderate inversion: model needs and implications for analog design," *European Solid-State Circuit Conference (ESSCIRC)*, September 16-18 2003, Lisbon, Portugal.

[65] J. Groszkowski, "The interdependence of frequency variation and harmonic content, and the problem of constant-frequency oscillators," *Proc. of the IRE*, vol. 21, no. 7, pp. 958–981, July 1933.

[66] HongMo Wang, "A 9.8 GHz back-gate tuned VCO in 0.35 μm CMOS," *IEEE International Solid-State Circuits Conference*, pp. 406 – 407, 15-17 February 1999.

[67] A.H. Mostafa, M.N. El-Gamal, and R.A. Rafla, "A sub-1-V 4-GHz CMOS VCO and a 12.5-GHz oscillator for low-voltage and high-frequency applications," *IEEE Transactions on Circuits and Systems—Part II: Analog and Digital Signal Processing*, vol. 48, no. 10, pp. 919 – 926, October 2001.

[68] Rupert Howes and William Redman-White, "A small-signal model for the frequency-dependent drain admittance in floating-substrate MOSFET's," *IEEE Journal of Solid-State Circuits*, vol. 27, no. 8, pp. 1186–1193, August 1992.

[69] Y.-C. Tseng, W.M. Huang, V. Ilderem, and J.C.S. Woo, "Floating body induced pre-kink excess low-frequency noise in submicron SOI CMOSFET technology," *IEEE Electron Device Letters*, vol. 20, no. 9, pp. 484 – 486, September 1999.

CHAPTER 4
SURFACE MICROMACHINING TECHNIQUES

4.1 Introduction

The present applications in communication systems require very low weight and volume, as well as low power consumption. Highly integrated RF front-ends circuits are thus desired. Continuing chip scaling has contributed a lot to this goal, but today a situation has been reached where the presence of rather bulky, expensive, off-chip passive RF components, such as high-Q inductors, ceramic resonators, SAW filters, varactor diodes and PIN diodes switches play a limiting role. Emerging MEMS technology meets the needs of small, integrated, high performances passive devices. The micromachining technologies firstly developed for sensors and actuators applications are now used in the RF and microwave field to build high performances components and to fabricate system-on-chip [1].

The MEMS fabrication technologies are based on the well-controlled microelectronics technologies. Nevertheless, the fabrication of three dimensional devices requires some specific steps that will be detailed in this chapter.

Three micromachining families are commonly considered: molding, bulk and surface micromachining.

In the *molding* processes, mechanical elements are created by deposition of material into a microfabricated mold. The most widespread molding process is called *LIGA*, the acronym for the German expressions of the major process steps: LIthography, Galvanoformung (electroforming) and Abformung (molding). The basic process consists of creating a polymer mold by X-ray or E-beam lithography and then electroplating metal into the mold cavities. Small feature sizes and high aspect ratio are achievable with this process. The advantage of such molding is that it allows a wide variety of materials to be considered for MEMS, beyond the traditionally ones used in microelectronics. However, the use of non-CMOS compatible process and materials restricts the capability to achieve high degree of integration of mechanical and electrical functions. For that reason, molding is not used in this work. The interested reader is referred to chapter 6 of [2] for more details on molding processes.

Bulk Micromachining makes micromechanical devices by etching deeply into the

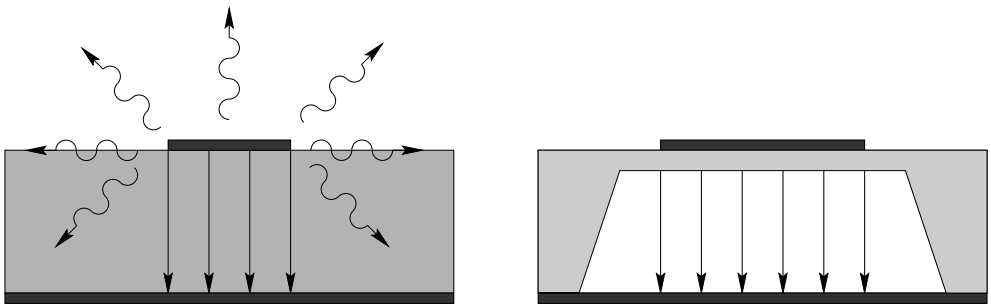


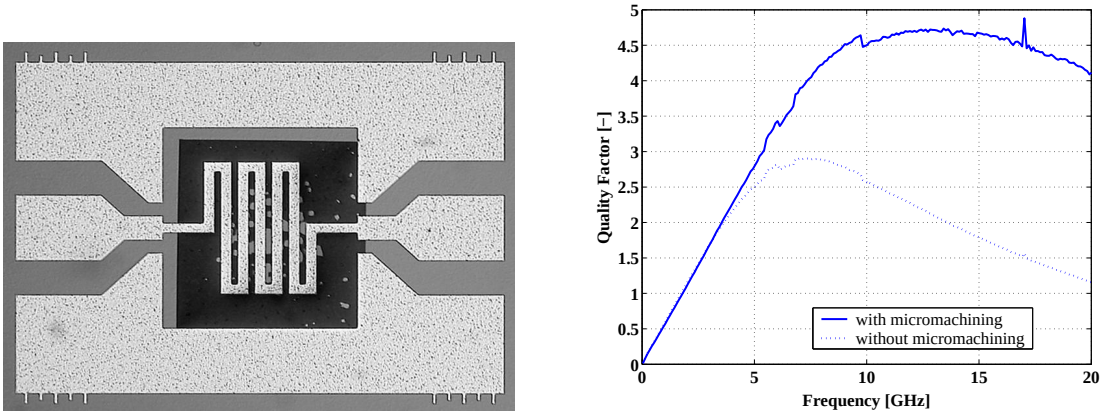
Figure 4.1: Cross-section of a microstrip line (a) without and (b) with bulk micromachining.

silicon wafer. There are several ways to etch the silicon wafer [3]. Wet anisotropic etching uses etchants like KOH (potassium hydroxide) or TMAH (tetramethylammonium hydroxide). Each crystallographic direction is etched at a different rate. Certain crystallographic planes are etched extremely slowly, and are called stop planes. The recent evolution of deep dry etching offers another way for bulk micromachining.

In the microwave field of applications, bulk micromachining has been mainly used to improve the performances of transmission lines and inductors. Indeed, attenuation in transmission lines originates from conductor, dielectric, radiation, and surface wave losses (Fig. 4.1(a)). All those losses increase with frequency. By eliminating the substrate, three advantages appear (Fig. 4.1(b)). First, the dielectric losses are reduced. Second, as the effective dielectric constant is also reduced, the surface mode losses are decreased. Third, in order to keep the characteristic impedance constant (e.g., $Z_c = 50\ \Omega$), the width of the line must be increased, which helps to decrease the conductor losses.

To demonstrate this feature, a micromachined meander inductor was fabricated. The substrate below the inductor was removed from the back side (Fig. 4.2). The 1 μm thick aluminum conductor is deposited on a three-layers dielectric membrane constituted of 400 nm silicon oxide/300 nm silicon nitride/300 nm silicon oxide stack. The silicon substrate is then etched away by the TMAH based chemistry detailed in [4]. Fig. 4.2(b) shows an improvement of the quality factor of a factor of about 2. Furthermore, the self-resonance peak is shifted to higher frequencies.

In the last micromachining family, the *surface micromachining* technology, multiple depositions and etchings are involved to build up structures at one side of the wafer. Fabrication techniques originally used in the integrated circuits industry, as polysilicon and silicon oxide or nitride chemical vapor deposition (CVD), as well as



(a) (b)

Figure 4.2: Bulk micromachined inductor. (a) Microscope picture, the dark part at the center is the region where the substrate is etched away. (b) Quality factor of the inductor in (a) before and after the substrate removing.

wet and plasma etching, are used to build three dimensional structures that have to be freed from the planar substrate to permit their displacement. Smaller devices are achievable, with lower thickness and mass than bulk micromachined ones. Lots of devices have been demonstrated in that technology: micro-engines [5], accelerometers [6], sensors and actuators in several field of applications as optics and microfluidics [7], resonators and tunable capacitors [1], etc. The micromachining process involves four central issues [8]:

- the understanding and the control of the material properties of microstructural films, such as polycrystalline silicon
- the release of the microstructure, followed by its drying and surface passivation
- the co-integration of the MEMS device with the active circuitry
- the packaging of the device.

In this chapter, some issues related to the MEMS fabrication process are discussed. The first section is dedicated to specific etching techniques using an inductively - coupled plasma (ICP) chamber. Indeed, good accuracy patterning of the mechanical structures is required for a good control of the mechanical properties such as resonance. This in turns requires adapted tools. We developed processes for deep silicon etching, as well as thin film etching with high precision. The mechanical properties of the polysilicon are discussed in section 4.3 focusing on the stress extraction. The cobalt silicidation is investigated as a mechanical material for MEMS. The third section presents some release techniques. We showed that liquid

and vapor phase silane coating increases the reliability of thin-films MEMS. Finally, it is explained how to design and fabricate a MEMS tunable capacitor that may be embedded in a VCO in section 4.5.

4.2 Dry Etching Techniques

Dry etching techniques are used in micro-engineering to face the main drawbacks of wet (purely chemical) etching, which are:

- the presence of irreproducible disturbances from bubbles, flow pattern, etc.,
- handing with corrosive and toxic materials,
- limitation of the minimum achievable dimension,
- the impossibility to obtain vertical trenches.

Dry etching may be defined as the use of radical species and/or energetic ions from a cold plasma (or ion beam source) to remove material where there is no mask material [9].

A *plasma* is a neutral ionized gas[1]. To remain ionized, the recombination of species must be avoided or at least limited. For that purpose, a constant input of energy (usually a RF field) and low pressure are needed. Indeed, the collision rate is decreased at low pressure. The formation of a plasma starts with some free electrons, which are accelerated by the applied electrical field, and collide with a gas molecule M. Three types of reaction may then occur:

1. Ionisation
$$e^- + M \rightarrow M^+ + 2e^-$$

2. Dissociation
$$\begin{aligned} e^- + M_2 &\rightarrow M^+ + M + 2e^- \\ &\rightarrow M + M + e^- \end{aligned}$$

3. Excitation
$$e^- + M \rightarrow M^* + e^-$$

[1]The Debye length λ_D is the minimum scale length for over which the plasma can be considered neutral; typically, $\lambda_D \approx 0.015\ cm$.

Ions, molecular fragments and excited molecules are also reactive. They may leave the plasma by recombination, light emission, or reaction with the chamber walls.

In dry etching applications, cold plasmas are used. A *cold plasma* is a partially ionized plasma. The ionization rate[2] lies typically in the range of $10^{-6} - 10^{-2}$. Because the electrons are not in thermal equilibrium with the neutral species, the ions and the reactor walls, chemical reactions may occur in the bulk plasma at room temperature. Detailed theory of plasma is beyond the scope of this work, the interested reader is referred to [10] for better insight to the plasma theory used in the MEMS field. In the following, some specific techniques are introduced from the process engineer point-of-view.

The removal of a substrate material in a plasma occurs when the reaction products are volatile under plasma conditions. Etching is possible thanks to both physical and chemical reactions. Chemical reactions, which are very selective, lead to an isotropic etching while physical unselective reactions leave an anisotropic profile.

As the plasma is full of mobile charge carriers, it is a electrical conductor and its potential is uniform. Therefore, it cannot be in direct contact with the chamber walls or other materials. The separation region is called *sheath*. The sheath determines how the charged particles interact with the surface being processed. Near the walls, electrons and ions escape from the plasma and are neutralized by the walls. Because of their smaller mass, electrons escape more easily. The surplus of positive ions in the plasma results in a positive charge. As the charge increases, the loss of electrons decreases due to electrical attraction, which in turns leads to an increase of positive ions. The charge of the plasma finally stabilizes at some positive value: the *plasma potential*, which lies usually in the range of $10 \ldots 200$ V. The electrical field resulting of this potential is orthogonal to the substrate. Ions are accelerated in this direction, leading to an anisotropic etch. The sheath region is darker. Indeed, because their high energy in that region, electrons are more likely to cause ionization than light-generating excitation.

Ions energy can be controlled through the applied RF power: while the ion density is determined in the bulk plasma, the voltage drop across the sheath increases with the RF power, without increasing the ion current. Note that ions are only able to follow the RF field if the frequency of the applied field is smaller than the plasma

[2]Ions and electrons densities, N_i and N_e resp., are in the range of 10^9 to 10^{13} cm^{-3} while neutral species density is in the range $10^{-6} - 10^{-2}$ cm^{-3}.

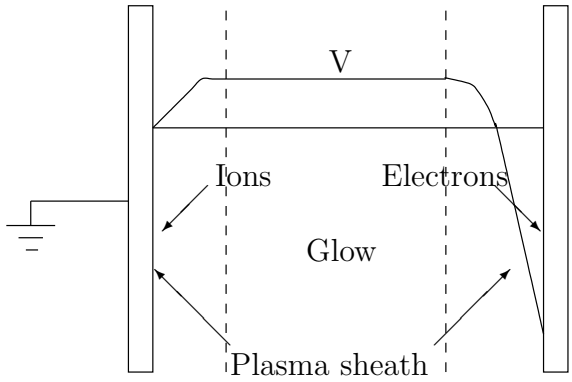


Figure 4.3: Principle of a DC plasma reactor.

frequency f_i of the ions [10]:

$$f_i = \frac{1}{2\pi}\sqrt{\frac{n_i q^2}{\varepsilon_o m_i}} \tag{4.1}$$

where n_i is the ion density, q, is the ion charge, and m_i the ion mass, ε_o is the vacuum permittivity. The standard frequency is 13.56 MHz, which corresponds to an ion density n_i of about $1.6 \cdot 10^{11}$ cm^{-3}.

4.2.1 The inductive coupled plasma reactor

The basic plasma reactor is composed of two electrodes in a low pressure chamber (11 mTorr... 1 Torr). A continuous electrical field accelerates the electrons so that molecules breakdown occurs. To sustain the discharge, the electrons have to be emitted when ions strike the cathode. In such *DC plasma reactor*, only conductive substrates can be etched. The potential distribution is not symmetric inside the chamber (Fig. 4.3). Indeed, because of the mass difference between electrons and ions, the later spent in average more time in the dark region. A too close separation of the electrodes prevents from ionizing collisions and a too far separation results in a surplus of inelastic collisions (loss of energy). The voltage needed to initiate the discharge and the gas breakdown depends on several parameters, mainly the pressure and the distance between electrodes.

The *RF plasma reactor* (diode reactor) is more popular since it allows etching conductors and dielectrics. In this chamber, the pressure is in the range of 10 to 100 mTorr. Secondary electrons are not needed to sustain the discharge.

In *Reactive Ion Etching* (RIE) reactors, both physical and chemical etching take place. The surface to be etched is placed on the powered electrode in a capacively coupled diode etching system. The electron current reaching the electrode balances the average ion current. The main disadvantage of diode reactor is the impossibility

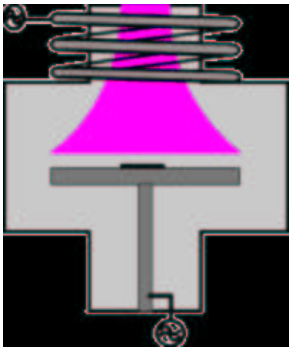


Figure 4.4: ICP plasma reactor [Oxford Plasma technology].

to control the ion energy and the ion flux independently. Even if a triode montage circumvents this, it does not answer the current needs of microfabrication i.e., high selectivity and high etch rate. This means that for MEMS applications, low ion bombardment ($< 100\ eV$) and high ions flux[3] are desired.

Inductively coupled plasma (ICP) generator meets these requirements. In this configuration, the control of plasma density is decoupled from the ion energy. The plasma density is controlled by an RF coil (typically at 13.56 MHz frequency) separated from the plasma by a dielectric wall (Fig. 4.4). The RF current in the coil induces a RF current in the plasma. This current is concentrated within a skin depth (a few cm) of the plasma surface. A plasma density of 10^{11} cm^{-3} may be obtained without high voltages on the electrodes. Compared to the diode source, the substrate sheath is thinner and there is less ions scattering.

In high density plasmas, the ion flux is determined by the coil source power and the ion energy by the bias power. Increasing the bias power (with the source power constant) increases the ions energy without changing the ion current. On the other hand, increasing the source power (with the bias power constant) leads to an increase of the ion flux and a decrease of the ion energy (Fig 4.5).

4.2.2 Silicon anisotropic etch

An ICP reactor was used to develop anisotropic Si etching processes. The different process parameters are discussed in this section. Our experimental results are given for thin-film as well as for thick-films etching.

An Oxford PlasmaLab System 100 was used for the experiments. It consists of a parallel plate reactor with a 300 W, 13.56 MHz RF source, combined with a high density ICP source with power up to 5000 W. The spacing between electrodes is

[3]about $10^{11} - 10^{12}$ electrons per cubic centimeter, which corresponds to a current of about 10 mA/cm.

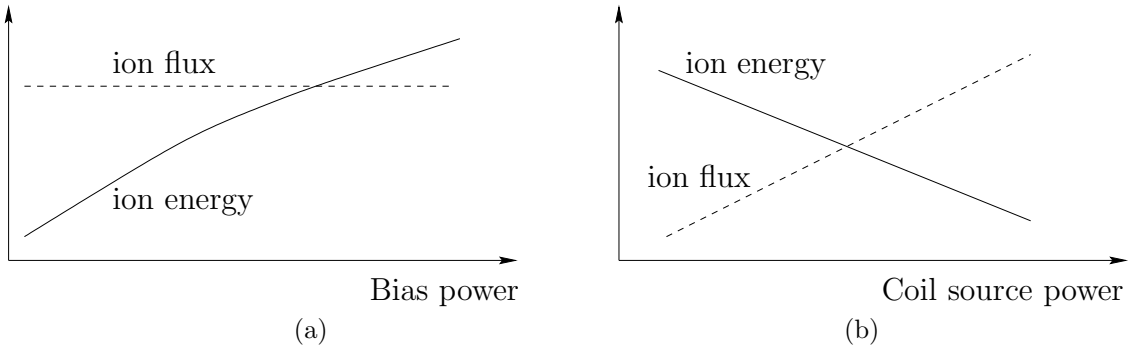


Figure 4.5: Relationship between ion flux and energy. The straight lines represent the ion energy while dashed lines the ion flux. (a) the source power is constant and larger than bias power. (b) the bias power is constant.

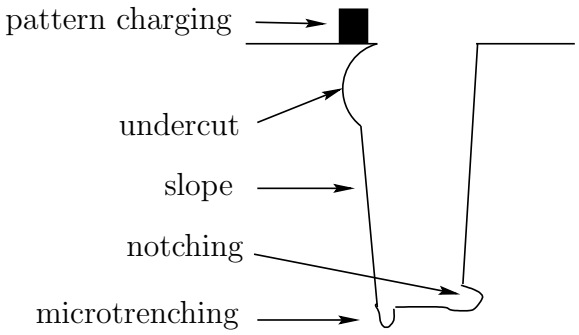


Figure 4.6: Trench characteristics.

adjustable, and the lower electrode temperature is controlled by backside cooling or heating with helium.

4.2.2.1 Etch characterization

The etch characterization includes selectivity to the mask, etch rate and depth, aspect ratio, uniformity, loading and micro-loading, and shape of the trench. Other nonidealities are described in Fig. 4.6.

Referring to Fig. 4.7, the *anisotropy* $\mathcal{A}$ is defined by

$$\mathcal{A} = 1 - \frac{u}{v} \tag{4.2}$$

where v is the etch depth and u is the maximal undercut of the mask or, in the case of outward sloped profiles, the lateral extension of the sidewall. The anisotropy is governed by the joint action of chemical and physical etch processes. A chemical reaction leads to isotropic profile, excepted in some cases where the profile follows the crystalline orientation (for example, Si etch with TMAH or KOH). The chemical reactions are highly dependent on the temperature. As there is no thermal equilibrium in a plasma reactor, the reactor temperature is not a well defined quantity.

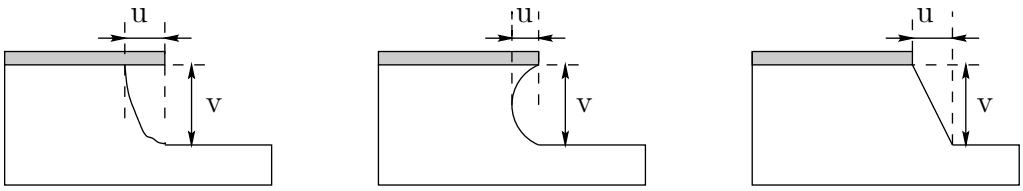


Figure 4.7: Definition of anisotropy in different etch profiles.

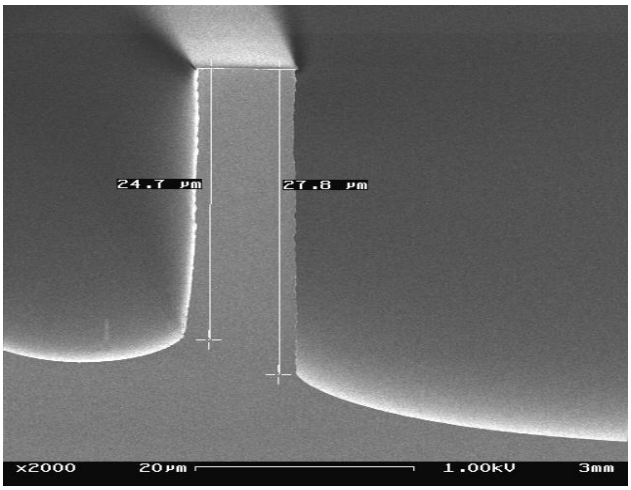


Figure 4.8: Micro-loading effect in silicon etch. The trench width at the left hand side is ten times smaller than this at the right.

When a reactive gas is used, the heat transfer is generally realized by a helium flow.

The *loading effect* characterizes the fact that etch rate and profile for a given trench depend on the etched area. This effect is more visible for recipes that use more chemical than physical etch. It is the case for instance of the Si etching recipes that use fluorine chemistry. Its origin lies in the fact that, contrary to what happens in the wet etching case, the number of radicals in the plasma is in proportion to the number of atoms to be removed. There are no exact rules and etching tests should be done with each mask design. Anyway, the larger the area exposed, the less the reaction. The loading effect gets less importance at lower pressure. A good thermal contact permits to diminish the loading effect. For uniform etching, Madou [2] states that the ratio of rate of formation of the etch product to the rate of etch gas flow should be greater than 0.1.

The etch rate also depends on the trench width, specially for widths smaller than 200 μm: this is the *micro-loading effect* (Fig. 4.8). This follows a transport phenomenon: the introduction of radicals and the extraction of the etch products are easiest in big trenches. Deep silicon etch is very sensitive to microloading. This unavoidable effect should be taken into account at the design level.

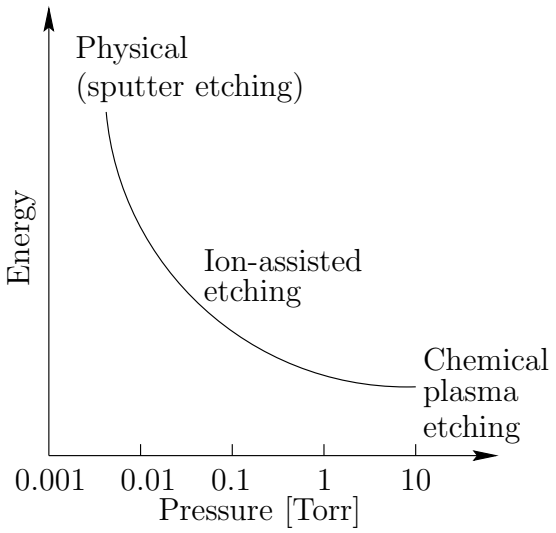


Figure 4.9: Typical energy-pressure relationship in a RIE reactor.

Table 4.1: RIE process variables and their effects.

Process variable	Effects
gas mixture (chemistry)	etch rate, profile shape
flow rate	etch rate (residence time)
pressure	etch rate, profile shape, selectivity/mask material (ions density)
RF power	etch rate (dissociation rate)
wafer temperature	profile shape, chemistry
self-bias of the chuck	selectivity, etch rate, shape profile

In general, three factors control the etch rate in a plasma reactor: the concentration of the neutral atom and free radical concentration, the ion concentration, and the ion energy. The ion and radical concentrations control the reaction rate, while the ion energy provides the necessary activation and controls the degree of anisotropy. The respective contribution to chemical and physical action of a plasma can be manipulated by varying both the voltage and the gas pressure. Etching at low pressures, with a long mean free path length of the ions, is more directional and less contaminating. On the other hand, the ion density drops off quickly at low pressure, decreasing the etch rate. The relationship between ion energy, pressure and reactions is sketched in Fig. 4.9. The different process variables and their effects are summarized in table 4.1.

4.2.2.2 Fluorine-based silicon etching

Silicon may be etched by fluorine or chlorine chemistries, as SiF_4 and $SiCl_4$ are volatile under plasma condition. As the silicon-halide volatility increases with in-

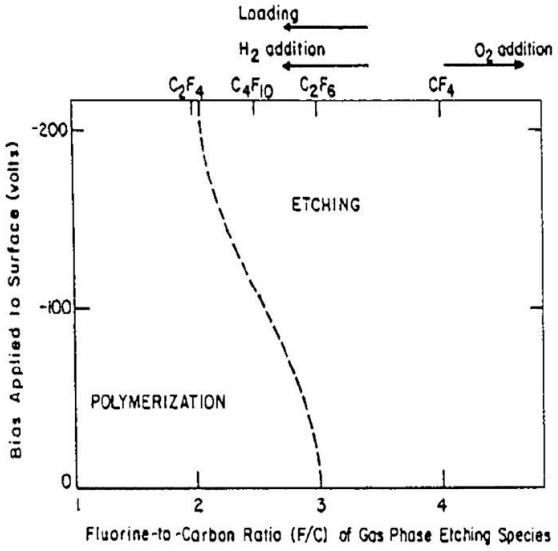


Figure 4.10: Illustrative plot of the boundary between etching and polymerization in fluorocarbon low density plasma as function of F/C ratio and ion bombardment at the surface (reprinted from [11]).

creasing halogen electronegativity, fluorine is chosen for high etch rates. The C_xF_y involved in the processes decomposes into a number of fragments, including F. The silicon is etched by the following reaction:

$$Si + 4F \rightarrow SiF_4$$

In the following, only fluorine plasmas involving SF_6 will be considered. To etch silicon, the F/C ratio should be high. Indeed, depending on the F/C ratio, either the silicon can be etched or a thin films of polymer can be deposited on the surface (Fig. 4.10). Addition of H_2 leads to polymerization, as it consumes F, while addition of O_2 consumes C and leads to Si etching. If pure fluorine chemistry is used to etch Si, an isotropic profile is obtained. Anisotropy is improved by lowering the temperature to cryogenic range (-110^oC) [12] or by using a sidewall protection as will be explained in section 4.2.2.3.

It is worth noting that the mask material may affect the process conditions, because it may act as a catalyst for SF_6 to generate fluorine. Fluorine-based plasmas also allow the etch of silicon oxide and nitride:

$$SiO_2 + 4F \rightarrow SiF_4 + O_2 \quad (4.3)$$

$$Si_3N_4 + 12F \rightarrow 3SiF_4 + 2N_2 \quad (4.4)$$

When the RF power increases, the atomic fluorine concentration significantly

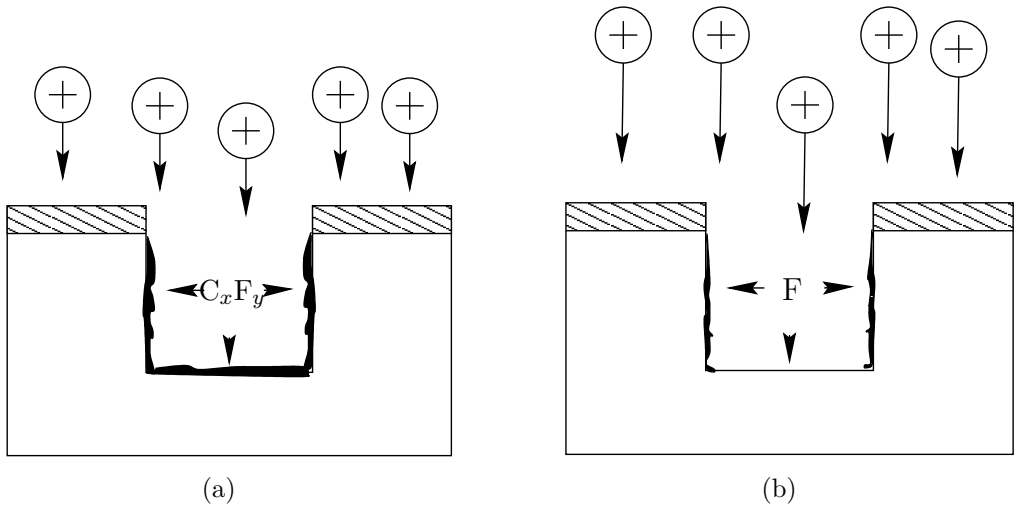


Figure 4.11: Bosch process for deep silicon etching. (a) Deposition step and (b) Etching Step.

increases, while the oxygen concentration increases only slightly. The effect of pressure on the etch rate can be summarized as follows: at low RF power the etch rate decreases with increasing pressure, while at high RF power the etch rate first increases with pressure and then decreases.

4.2.2.3 Deep RIE: the Bosch process

At room temperature, the patented Bosch process [13] is used for deep silicon etching with high aspect ratio. It takes the advantage of the fact that both etching and polymer deposition may occur in the same chamber, as depicted in Fig. 4.10. If a fluorocarbon film is deposited on the wafer, it protects it from etching. The process consists of alternatively one etching step with SF_6 and one passivation step with C_4F_8 (Fig. 4.11). The patterns are isotropically etched during the first step. Then, a C_xF_y layer is deposited on the wafer. As in the subsequent etch step, the ions are accelerated in the vertical direction, the protection layer remains on the sidewalls only, providing a high anisotropy level. The etch rate is high (several μm/min) and the selectivity toward photoresist and SiO_2 is high. The main drawback of this technique is the formation of a hard polymer of unknown composition inside the wafer. Because of the cycling process, scallops are formed at the sidewalls (Fig. 4.12, right part).

Using the PlasmaLab System 100 reactor, a Bosh-type three-steps recipe was developed (Fig. 4.12). The *Ignition step* was necessary to ensure the plasma ignition. The process should end by an *Etching step* in order to get rid of part of the fluorocarbons compounds at the bottom. The silicon etch rate obtained with this process

	Ignition Step	Deposition Step	Etching Step
SF_6	30 sccm	-	110 sccm
C_4F_8	70 sccm	120 sccm	-
RF power	25 W	5 W	25 W
ICP power	2000 W	2000 W	2000 W
Temperature	15^oC	15^oC	15^oC
Pressure	6 mTorr	20 mTorr	40 mTorr
Table height	20 mm	20 mm	20 mm
Step time	5 s	4 s	7 s

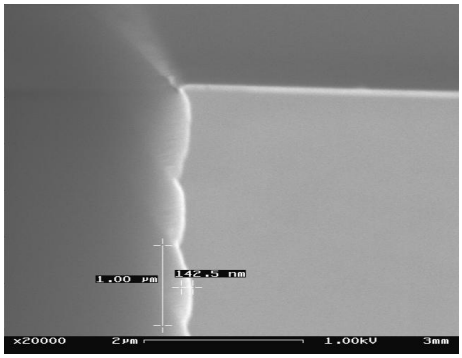


Figure 4.12: UCL Bosch process parameters and resulting profile obtained on a Si bulk wafer with a thermal SiO_2 mask.

is 4.5 μm/min i.e., 1.2 μm per cycle. The selectivity toward metallic compounds is quasi infinite (about 500 in the case of Aluminum) but relatively low toward silicon oxide (a selectivity of 63 was measured). As shown in Fig. 4.12, no undercut was observed and scallops of small depth are obtained.

4.2.2.4 Thin film etching

In the case of thin film (poly) silicon etching, the most widely used recipes involve either chlorine and/or bromine chemistry to produce anisotropic etch profile [14]. Pure SF_6 chemistry may be used, leaving rough surfaces. This drawback may be avoided by a sidewall protection generated at the same time the etch takes place. Therefore, the adjunction of CHF_3 or C_4F_8 in the reactor results in smooth vertical sidewalls at the price of a lower etch rate [15]. Indeed, if the ratio F/C$<$ 8/3, the carbon fragments polymerize to form an inert, etch-resistant film that slows the etch process.

Using that feature, we developed a recipe for the etching of nanometric devices with high precision (Fig. 4.13). High aspect ratio, 50 nm width patterns were obtained. As the selectivity to SiO_2 is poor (about 2), a metallic mask should be used for that process. The silicon etch rate is 1400 Å/min. Since mask undercut is negligible ($<$ 10 nm), this recipe may be used for the patterning of lateral nano-resonators operating in the GHz range.

We also developed an alternative for precise silicon micromachining, by adaptating of the Bosch process to thin-film technologies [16]. To overcome the scallops formation, the ICP-assisted plasma is replaced during the etch step by a low density plasma, only activated through RF coupling. This adaptation produces very

SF_6	5 sccm
C_4F_8	95 sccm
RF power	100 W
ICP power	1000 W
Temperature	5^{o}C
Pressure	5 mTorr
Table height	50 mm

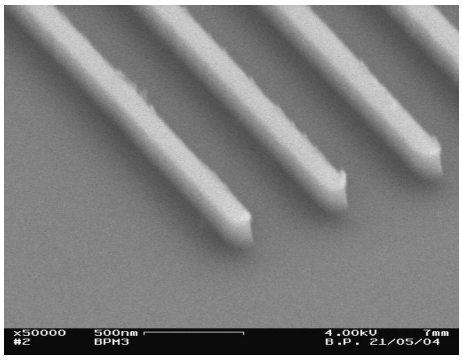


Figure 4.13: Recipe for precise silicon etching and 80 nm width wires patterned with that process on a 400 nm thick Si layer on top of a thermal SiO_2 stop layer (Cr mask).

smooth vertical sidewalls on polysilicon microstructures, while keeping a good control on etch depth with any angle of anisotropy. The polysilicon etching procedure is adapted from the standard Bosch-type process. It starts with a 15 s shallow etching step under a low pressure SF_6 plasma generated with RF source only, and continues with a 10 s passivation step enhanced via an inductively coupled plasma source during which C_4F_8 gas molecules are dissociated to form a polymeric layer over the exposed polysilicon layer. During the subsequent etch step, this layer is preferentially withdrawn on horizontal surface rather than on sidewalls because of ionic physical etching. The process parameters and the corresponding etch profile are reported in Fig. 4.14. Full details of the experiments may be found in reference [16]. The absence of undercutting is the first major advantage of this method compared with $SF_6/O_2/CHF_3$ gases mixture for example [15].

The etch step yields a mean etch rate of 2000 Å per min. Combined with passivation steps and assuming a mean value of 10 s for gases admission and pressure stabilization, effective etch rate is 1000 Å per min. Figure 4.15 presents the etch rate as a function of C_4F_8 flux during the passivation step. The effective etch rate strongly depends on the passivation efficiency which means ICP power. On the other hand, it was observed that the etch rate does not vary when the freon flow rate lies between 70 and 210 sccm.

In the case of polysilicon layers with thickness inferior to 2 μm, we measured the highest selectivity to be 13 : 1 between polysilicon and oxide hard mask. This relatively low value for fluorine-based plasma comes from the passivation step. Indeed we assert that during this step, there is an equilibrium between species useful for the polymer formation on the sidewalls, and fluorocarbon radicals chemisorbing in SiO_2 and unfortunately etching the oxide mask. The physical etching effect of ion

	Etch Step	Deposition Step
SF_6	100 sccm	-
C_4F_8	-	70 sccm
RF power	30 W	10 W
ICP power	-	4000 W
Temperature	15^{o}C	15^{o}C
Pressure	20 mTorr	20 mTorr
Table height	20 mm	20 mm
Step time	15 s	10 s

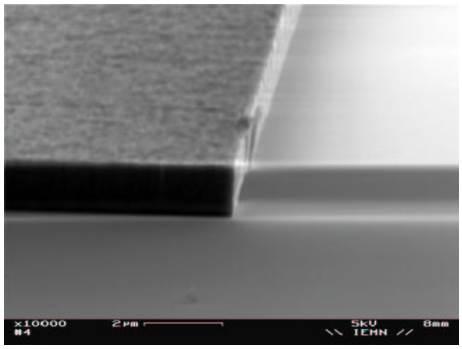


Figure 4.14: Modified Bosch process for the anisotropic etch of thin polysilicon films.

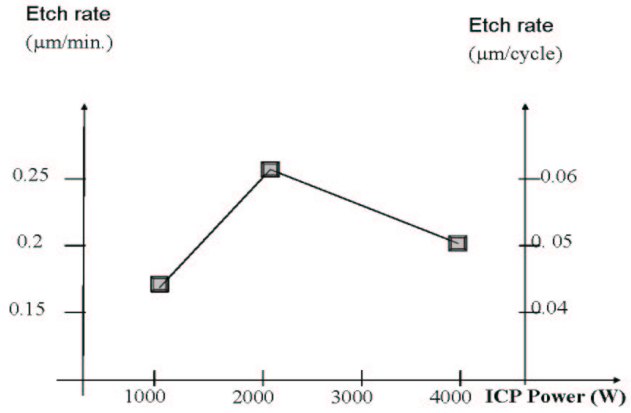


Figure 4.15: Etch rate as a function of the ICP power during the passivation step.

bombardment is also superposed to this mechanism. With the range of ICP power involved here, active CF_x species are numerous and so etch both physically and chemically the oxide mask. Furthermore, this result is all the more understandable if we consider the etch rate rather in term of thickness of etched polysilicon per etch / passivation cycle than in term of etched thickness per unit of time. Indeed, our results show that a 3000 Å can sustain around 30 successive cycles leading to a selectivity over oxide of 10 : 1. But in our case, this selectivity is deduced from etch steps enabling an etching of a polysilicon thickness as low as 50 nm. So, compared with standard Bosch process for which plasma densification through ICP is used for both steps, the typical silicon etched thickness during etch step is around 1 μm. If this value is related to the oxide etched rate measured in our case and asserting that the oxide / polysilicon selectivity is quasi infinite during the etch step, we can find back a polysilicon over oxide selectivity of 100 : 1, comforting our previous hypothesis.

An interesting feature of this process is that it allows controlling the anisotropy, which may be advantageous for microfluidic applications for instance. Indeed,

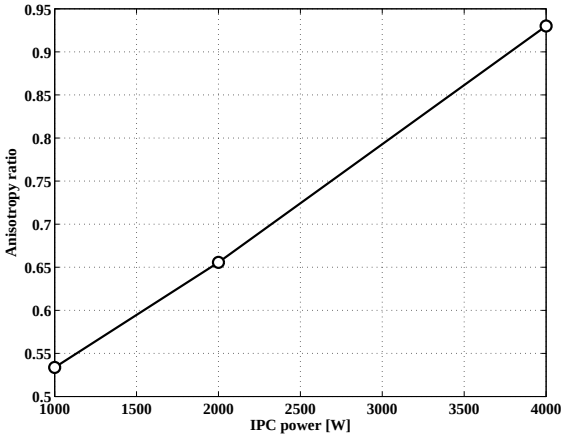


Figure 4.16: Anisotropy $\mathcal{A}$ as a function of the ICP power during the passivation step.

Fig. 4.16 shows that the anisotropy ratio is related to the equilibrium between etching and passivation i.e.,, that is determined by the ICP power. The phenomenon can be explained as follows: during the passivation, CF_x radicals are adsorbed at the polysilicon surface to form a polymeric protection layer over the whole surface. But, in the mean time, ions are accelerated towards this exposed surface leading to a purely physical etching of this thin layer, except on the sidewalls. When the ICP power is high, the high density plasma provides a high ionic surface density to etch the horizontal polymeric layer, and thus after passivation, only sidewalls remain protected. But in the case of weaker plasma, etching species are less numerous, and it remains a small extent of the polymeric layer at the foot of the sidewall, which acts as protection during the subsequent SF_6 etch step. This leads to the "scallops" effect. Compared to other methods like wet etching of single crystal silicon through KOH or TMAH solutions, this technique allows any angle of anisotropy while keeping precise transversal dimensions since no undercutting of the patterning oxide mask was observable. Note that this property can be also tuned by the time ratio between etch and passivation steps.

4.3 The control of the stress in thin polysilicon films

Polycrystalline silicon (polysilicon) is a material widely used for both electronics and micromechanical applications [17]. Cost of the process is reduced if the same polysilicon is used for the electronics and mechanics. Because the film is isotropic and has superior material properties compared to metal films, polysilicon is the structural

material of choice for surface micromachining applications. However, high conductivity materials are desirable in most MEMS applications. This work investigates the silicidation of polysilicon as a solution to obtain low-resistivity, low-stress, high-quality mechanical material. Thin films have different material properties than their bulk material counterpart. Residual stress is a major concern for the operation and the reliability of MEMS. Indeed, the device may buckle under compressive stress, while tensile stress may crack the structure.

Residual stresses of thin films highly depend on the depositions parameters. Even if the analysis of stress of has been widely reported in the literature for undoped [18, 19], *in situ* doped [20, 21] and ion-implanted [22, 23], or even doped by diffusion [24, 25] polysilicon films, the use of stress values extracted from the literature is not sufficient since materials properties can change a lot from one equipment to another.

In this section, we introduce first some background theory of elasticity and some methods to extract the residual stress in thin polysilicon films. Experimental results are then given on phosphorous doped polysilicon wafers; thermal annealing and silicidation are discussed.

4.3.1 What is residual stress?

Stress σ is defined microscopically as the force per unit area acting on the surface of a differential volume element of a solid body. Stresses perpendicular to a differential face are called *normal stresses*, while forces acting along the faces are called *shear forces* τ. Stresses can result from externally applied forces. In this case, after the load is removed, the stresses vanish. On the other hand, thin films are stressed even without the application of externally imposed forces. This kind of stress is called *residual stress*.

Strain ϵ is the differential deformation of a solid body submitted to some forces. The axial strain is then related to the displacement u in the x direction by

$$\epsilon = \frac{\partial u}{\partial x} \tag{4.5}$$

In the case of a thin film deposited on a thick substrate, the system can be assumed to be biaxial. Since in MEMS applications, the materials should stay in the elasticity domain, they follow Hooke's law, which states that they deform linearly with load:

$$\begin{pmatrix} \epsilon_x \\ \epsilon_y \end{pmatrix} = \begin{pmatrix} 1/E & -\nu/E \\ -\nu/E & 1/E \end{pmatrix} \begin{pmatrix} \sigma_x \\ \sigma_y \end{pmatrix}$$

where E is the *Young's modulus* or *elastic modulus* and ν is the *Poisson's ratio*. The Young's modulus relates how much a material is elongated for a given load. The higher the E, the lower the deformation, for a given stress. Typical E value for silicon $< 100 >$ and polysilicon are 130 GPa and 160 GPa, respectively. The Poisson ratio describes how an element expands in a direction normal to the applied stress. Typical values lie between 0 and 0.5. Since the volume expansion is proportional to $(1 - 2\nu)$, incompressible material have a Poisson ratio equal to 0.5. Values for silicon $< 100 >$ and polysilicon are 0.28 and 0.23, respectively. In the case of isotropic materials, $\epsilon_x = \epsilon_y = \epsilon$ so that the in-plane (biaxial) stress $\sigma = \sigma_x = \sigma_y$ is equal to:

$$\sigma = \left(\frac{E}{1-\nu}\right)\epsilon. \tag{4.6}$$

The ratio $E/(1-\nu)$ is called the *biaxial modulus*.

The stress in a thin film originates from different sources that may be classified into two families: intrinsic and extrinsic stresses. The latter group represents the *thermal expansion* of the film. Indeed, as the film is deposited at high temperature on the substrate that have a different *thermal expansion coefficient* $\alpha_T \stackrel{\triangle}{=} \frac{d\epsilon}{dT}$, a strain $\alpha_T \cdot \Delta T$ appears when the structure is cooled down to ambient temperature. Other sources of residual stresses are lattice mismatch, substitutional or interstitial impurities, chemical reaction while processing, etc. We call *intrinsic stress* the ensemble of stresses that are different from the thermal expansion and that are mainly dependent on the process parameters.

Residual stress can be uniform or non-uniform through the depth of a thin film. In the case of polysilicon, a stress gradient exists between the top and the bottom of the film. Indeed, when deposited at $620 - 650^o$C, the first grains of poly-Si are small and randomly oriented, while they organize in columnar coarse grains latter in the deposition. The stress is so highly dependent on the thickness of the film.

Stress in polysilicon Thin films of polysilicon typically exhibit a compressive stress that depends on the grains size. Krulevitch [19] suggested that low thickness compressive polysilicon films are made up of small grains that are misoriented [18]. The adatoms do not have enough surface mobility to find the lowest energy levels; that is, the film is amorphous at first. As the deposition time is increased, the polysilicon film begins to grow vertically in a columnar fashion and distinct islands arise, changing the initially amorphous film into a crystalline structure (Fig. 4.17). This explains the crystallographic structure of compressive thin film polysilicon a

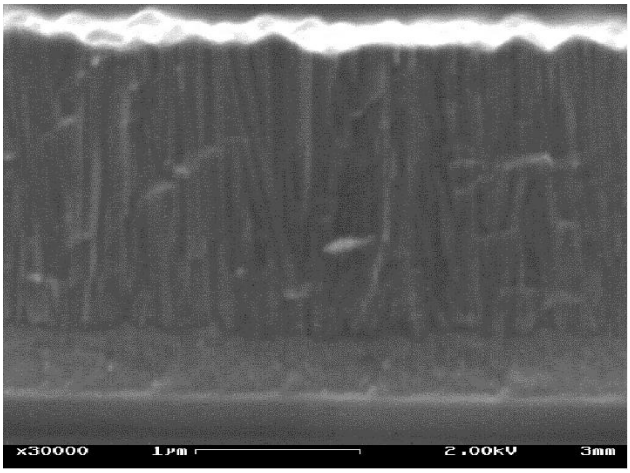


Figure 4.17: Compressive polysilicon formed at 625^o C constituted of columnar coarse grain structure on top of random small grains at the interface with silicon oxide.

higher thicknesses. Also, as the deposition is performed at relatively high temperature, a longer deposition means a longer annealing. A longer annealing time promotes the grain growth and the loss of both the defects and gases trapped in the film. As the film thickness builds up, the deposition environment gets purer and the stress becomes less compressive [24]. In conclusion, as-deposited stresses decrease when polysilicon thickness increases. Nevertheless, for thicknesses values higher than 1.5 μm, the residual stress becomes almost independent on the thickness [18].

The underlying layer influences the stress of the polysilicon. At room temperature, silicon oxide is compressive and silicon nitride tensile. For polysilicon films on oxide, the grains at the interface are smaller than if the film was deposited on a silicon substrate, which could provide nuclei for epitaxial growth. This leads to high stress value. Kim [18] reported that if the layer thickness is greater than 0.8 μm, the evolution of the stress versus the polysilicon thickness follows the same profile, whatever the layer is deposited on silicon oxide or nitride; the polysilicon on silicon nitride is a bit more compressive.

Thermal impact on stress Thermal stresses are the highest contributors to the overall stress value. Annealing reduces the stresses caused by mismatches expansion and by non-uniform nucleation of the film. For polysilicon, annealing at temperatures well above its deposition temperature should allow the atoms to find the lowest energy level, changing amorphous structures into crystalline forms [18].

Several annealing methods have been explored to control the residual stress of polysilicon films. Annealing in a conventional furnace takes a long time at high temperature ($> 1000^o$C) to relax the strain [26, 27]. These high temperature steps

are unfortunately not compatible with standard CMOS processes. This technique is then only well suited for the MEMS pre-processing. Rapid thermal annealing (RTA) avoids this high thermal budget [28]. Even if high temperatures are involved, the annealing time is so short that it does not affect the electronics. Other techniques involve *in situ* annealing directly after deposition [27, 29] at about 600^{o}C temperature for several hours.

4.3.2 Stress extraction methods

The stress characterization of thin films is usually done by two ways. First, the average stress is obtained by measuring the wafer curvature and applying Stoney's equation. Indeed, a tensile stress will bend and render the surface concave; a compressive stress renders the surface convex. At the micro-scale however, the local stress evaluation requires other techniques involving the design of specific test structures.

Stoney's equation Let us consider a thin film of thickness t_f deposited on top of a substrate of thickness t_s. Since any change in wafer shape is directly attributable to the stress in the deposited layer, it is relatively straightforward to calculate the stress by measuring these changes. The *Stoney's equation* gives the relation between the stress in the thin film and the radius of curvature R of the stack:

$$\sigma_f = \frac{1}{6R(t_f + t_s)t_f}\left(\frac{E_f t_f^3}{1-\nu_f} + \frac{E_s t_s^3}{1-\nu_s}\right) \tag{4.7}$$

where E_s (E_f) and ν_s (ν_f) are respectively the substrate (film) Young's modulus and Poisson's ratio. If the substrate is thick i.e., $t_s >> t_f$, relation (4.7) simplifies to:

$$\sigma_f = \frac{E_s}{6(1-\nu_s)} \cdot \frac{t_s^2}{t_f} \cdot \frac{1}{R}. \tag{4.8}$$

This form presents the advantage that only the substrate Young's modulus and Poisson ratio have to be known to extract the residual stress in any kind of thin film. Relation (4.8) is a good approximation for $t_s/t_f > 10$ and modifications have been proposed in [30] to enlarge the validity range up to $t_s/t_f > 2.5$.

the stress measured by the Stoney's equation is an average stress since it is obtained by integrating the stress over the thickness of the film. The Stoney's equation relies on the assumption that the substrate has transversal isotropic elastic properties with respect to the film. This argument is satisfied for silicon $< 100 >$ wafers [2]. The films should also be uniform in thickness and the stress must be

homogeneous and equi-biaxial over the entire substrate to ensure the validity of Stoney's approach.

Local stress measurements are made using *in situ* surface micromachined structures made directly out of the film of interest [31]. This technique allows characterizing the stress at each batch. Various micromachined structures have been proposed in the literature to measure the intrinsic stress. Among them, the clamped-clamped beams [32] and the vernier gauges [33, 34] techniques are described hereafter, as they are used in the experiments described in section 4.3.3.

Clamped-Clamped Beams This technique makes use of rows of clamped-clamped beams with incrementally increasing lengths to determine the critical buckling load and hence deduce the residual compressive stress in polysilicon films. The residual strain ϵ_r is obtained from the critical length L_c at which buckling occurs [32]:

$$\epsilon_r = \frac{4\pi^2}{A_S}\frac{I}{L_c^2} \tag{4.9}$$

where A_S is the beam cross-sectional area and I is the moment of inertia. This simple approach does not take into consideration additional effects such as the internal moments resulting from gradients in residual stress.

Vernier Gauges Vernier gauges present two main advantages over the clamped-clamped beam technique. First, both tensile and compressive stresses may be characterized. Second, the surface of the test structure is reduced. Fig. 4.18 sketches a schematic of such micromachined strain gauge. When the device is released, the test beam expands or contracts, depending on the sign of the residual stress, causing the compliant slope beam to deflect into a 'S' shape. The indicator beam, attached to the deforming beam at its point of inflection, rotates and the deflection λ is read out visually on the vernier scale by microscopy. For in-plane structures, the residual strain ϵ_r is given by [33, 34]:

$$\epsilon_r = \frac{2L_s\lambda}{3L_iL_t\zeta} \tag{4.10}$$

$$\zeta = \frac{1-d^2}{1-d^3} \qquad \text{with } d = \frac{w_i}{L_s} \tag{4.11}$$

where L_i, L_t and L_s are respectively the indicator, test and slope beam lengths; ζ is a correction factor due to the presence of the indicator beam; w_i is the width of the indicator beam. Note that technique is that (4.10) is independent of both the

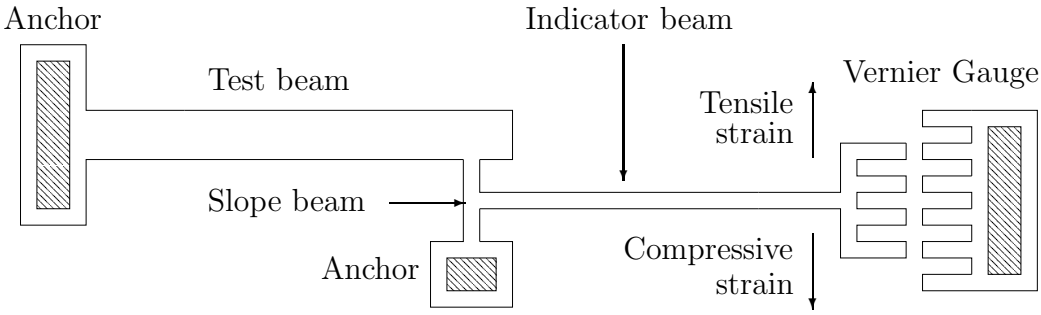


Figure 4.18: Strain gauge for the measure of tensile or compressive stress.

thickness of the deposited film and the cross-section of the microstructure.

4.3.3 Experimental results

In order to extract the stress of the doped polysilicon deposited in the UCL cleanrooms, test samples were prepared as follows. A 2 μm thick densified PECVD oxide layer is deposited on top of 3-inches p-type $< 100 >$ Silicon wafers. This corresponds to typical sacrificial layers used in silicon macromachining. The Oxford PlasmaLab System 100 was used for that deposition with the following parameters: 1 Torr pressure, 30 W RF power, chamber temperature 300^{o}C, 50 sccm SiH_4 flow, 500 sccm N_2 flow, 350 sccm N_2O flow. The oxide layer stress is -555 MPa[4]. After densification by Rapid Thermal Annealing (RTA) at 800^{o}C for 20 s, this value relaxes to -172 MPa. A 1000 Å thick SiO_2 layer was deposited at the back side using the same process parameters that were used for the front side. Then depending on the samples, a 1 or 2 μm thick polysilicon layer was deposited in an LPCVD oven at 625^{o}C with a 8 sccm SiH_4 flow. Both the polysilicon and the oxide layers were then removed at the backside. The next step consists in doping the polysilicon by CeP_5O_{14} diffusion at 900^{o}C under N_2 ambient for one hour. Then the P-glass was removed in BHF. This diffusion process was repeated two times in total. An annealing was then performed to decrease the residual stress. In order to check the effect of silicidation on the stress, the samples undergo a cobalt silicidation as follows. A 200 Å thick Co layer was deposited by E-gun, followed by a RTA at 450^{o}C for 40 s. The samples were then cleaned in a piranha mixture before a second RTA at 900^{o}C for 20 s. Control wafers constituted of silicon oxide at the top wafer side only, followed the same thermal budget were used to compute the stress using the Stoney's approach.

The curvature of the wafer was measured after each process step at room temperature using a Dektak profilometer over a distance of 50 mm with a 20 mg stylus

[4]A negative stress value refers to a compressive stress, while a positive sign refers to a tensile stress.

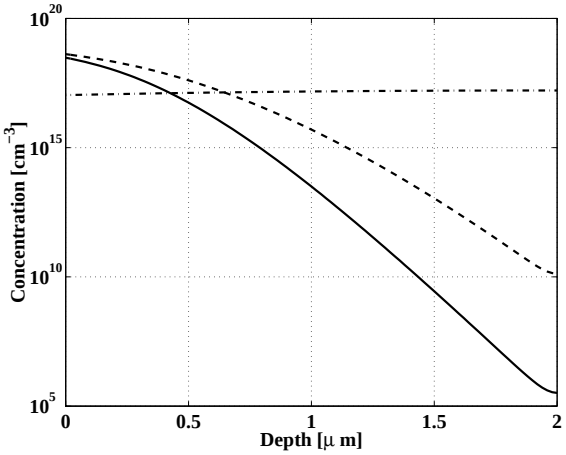


Figure 4.19: Athena simulations of the doping profile in the 2 μm thick polysilicon layer, after 60 min (straight line) and 120 min diffusion process (dashed line), and after 60 min annealing at 1100^{o}C (dashed-dotted line).

force. The sheet resistivity was measured by a four-probes station. The stress was evaluated by the Stoney's equation. Results are given in table 4.2.

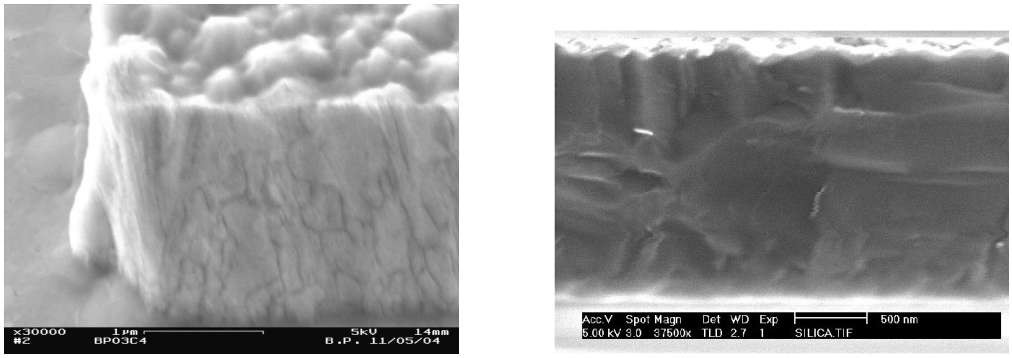
After the diffusion process, the stress in the polysilicon film becomes more compressive. At first sight, this result is not in agreement with Murarka's work [24]. But Murarka used thinner polysilicon films. We observed indeed a reduction of the stress of 4400 Å thick polysilicon from -690 MPa before doping to -203 MPa after doping. These samples were prepared the same way as described before. Nevertheless, Gianchandani [26] reported a decrease in compressive stress even for 3 μm thick polysilicon layers. The discrepancy in the results may be explained by a non uniform doping of our samples over the whole polysilicon layer. This hypothesis is verified in Fig. 4.19 by simulations using the Silvaco Athena software, where the doping P profile through the polysilicon layer is shown as a function of the process steps. A uniform doping is only obtained after the high temperature annealing. The differences observed in the polysilicon stresses before and after the diffusion, reflect the changes in the intrinsic stresses since it was reported elsewhere that the thermal expansion coefficients of the doped and undoped polysilicon are the same [24].

The stress decreases drastically during the annealing (table 4.2). Complementary experiments at 1000^{o}C showed that the stress is relaxed after one hour in the furnace and is not affected by subsequent annealing. A stress value of -190 MPa was obtained in this case for 2 μm thick polysilicon layers. This value is too high for most MEMS applications. Because the diffusion of the P species during the annealing, a more uniform doping profile is obtained after annealing. This affects in turns the electrical properties of the film.

The cobalt silicidation has minor impact on the stress budget but it contributes

Table 4.2: Stress of different polysilicon types extracted by the Stoney's equation. The minus (plus) sign refers to a compressive (tensile) stress.

	Stress		*Resistivity*	
	1 μm thick layer	2 μm thick layer	1 μm thick layer	2 μm thick layer
Undoped polysilicon	-235 MPa	-220 MPa	∞	∞
After 1 h P diffusion	-545 MPa	-362 MPa	148 Ω/sq.	74 Ω/sq.
After 2 h P diffusion	-424 MPa	-439 MPa	52 Ω/sq.	32 Ω/sq.
After 1 h annealing	43 MPa	9 MPa	14 Ω/sq.	9.5 Ω/sq.
After Co silicidation	74 MPa	43 MPa	3.5 Ω/sq.	3.1 Ω/sq.



(a) (b)

Figure 4.20: SEM picture of 2 μm thick polysilicon layers grown in a LPVCD oven at 625^{o}C (a) doped by P diffusion and (b) doped by P diffusion and after a Co silicidation.

to reduce the resistivity of the film. Experiments showed that the high temperature annealing should occur before the silicidation. If not, even if the stress does not vary much, the resistivity increases and reaches a value close to that of the samples that were not silicidated. The results in table 4.2 suggest that a 1100^{o}C annealing followed by a silicidation gives good material properties for micromachining. Fig. 4.20 shows the impact of the silicidation on the grain structure. Small grains are still present at the interface with oxide after the silicidation. The columnar structure presented in Fig. 4.17 disappears with the P doping diffusion and the high temperature annealing. Besides, the rugosity is increased after silicidation. This should not be a problem in most MEMS devices, since the higher the rugosity, the lower the capillary forces[5].

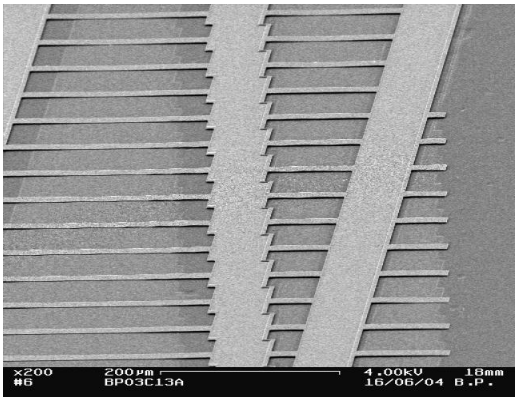
The silicidation process seems to be attractive for MEMS applications as it avoids a metalization step and proveides good mechanical properties. In reference [35], study on nickel silicide came to the same conclusion.

[5]As it will be explained in the next section, the capillary force is the main cause of stiction, a typical failure mode of the device.

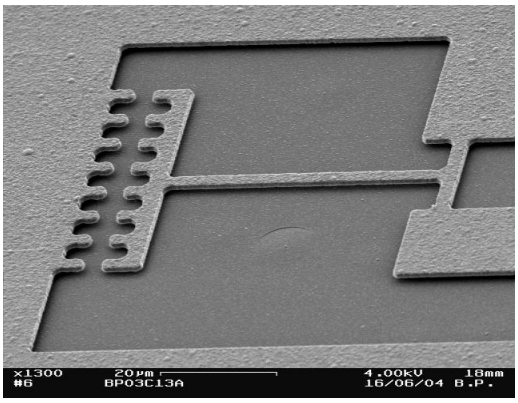
For comparison purpose, microstructures described in section 4.3.2 were built up (Fig. 4.21). The polysilicon was etched by a SF_6 based RIE and then released in concentrated hydrofluoric acid (HF) diluted in isopropanol. Subsequent rinses were performed in isopropanol and the devices were dried without any special process to avoid stiction (see next section). Indeed, it was believed that sticking of the verniers gauges is not a problem since it sticks in its relaxed state [25]. However, we observed that stiction affects the good working of the majority of the devices. This may be explained by the fact that our structures were build on a relatively small air gap (0.5 to 1.5 μm). Verniers with small indicator beams are not stuck, but they have a too small sensitivity to get accurate results. The bigger verniers suffer from stiction and the deviation in the deflection of identical devices built on the same wafer is too large to determine properly the stress with this technique. This deviation is attributed to the large forces applied on the structures when they quit the solvent. We conclude that if there is no post-process to free the verniers from stiction, this technique cannot be used to characterize the stress in situ during the batch. We further observed a lower sensitivity to stiction in the case of arrays of clamped-clamped beams. Compressive stresses were observed. This explains the lower probability of stiction, as the binding of the beam opposes itself to the attractive capillary force. The discrepancy between this method and the Stoney method remains unexplained as good agreement was found elsewhere [25, 4].

Discussion From this experimental work, the average polysilicon stress value were obtained (table 4.2). However, we were not able to compare these results with the local stress measured by micro-structures because of the stiction problem. The local stress extraction with this technique during the batch is thus not possible for MEMS with thin gaps. Specific process steps are needed to avoid stiction and will be discussed in the next section.

On the other hand, the polysilicon silicidation seems to be a low-cost mean for obtaining a material adapted to MEMS. Indeed, the stress is not much affected and the resistivity is decreased. As from table 4.2 and reference [36], about 70 nm of the polysilicon film is silicidated and that the Young's modulus of the $CoSi_2$ alloy is 160 GPa, the thermal expansion coefficient of the film lies closer to the silicon thermal expansion coefficient than metallic films, which is an advantage for the thermal stability of the MEMS.



(a)



(b)

Figure 4.21: Realized micro-test structures in polysilicon. (a) Array of beams. (b) Vernier gauge.

4.4 *Release Techniques: the* stiction *problem*

In order to make free-standing microstructures capable of motion, sacrificial layers hava to be etched away. The sacrificial layers considered in this work consist of SiO_2. The release of suspended structures is a three-steps process: etching the sacrificial layer, rinsing the etchants, and drying the structures.

The etching step is typically achieved with a selective liquid chemical, such as hydrofluoric acid (HF) or buffer HF (BHF, a mixture of NH_4F and HF).

When aluminum structures are present on the wafer, BHF or HF diluted into water should be avoided, as the Al is attacked via the following reaction:

$$2Al + 6H_3O^+ \rightarrow 2Al^{3+} + 3H_2 \uparrow + 6H_2O. \tag{4.12}$$

It is preferable to use a mixture of HF and an alcohol. The higher the concentration in HF, the higher the selectivity is, as the concentration H_3O^+ is lower. In order to get a better control on the etch rate, HF can be diluted in isopropanol as it decreases the etch rate [37]. Water should be avoided for the subsequent rinses, preventing from reaction (4.12).

The rinse liquid needs to be dried out to complete the release of the structure. During this step, several forces act on the structure: capillary, Van der Waals, electrostatic, quantum, etc. At this scale, the capillary force dominates [38]. Indeed, the surface forces become more important as the dimensions shrink, since the surface over volume ratio increases. If the capillary force is high enough, the structures are pulled down to the substrate [39]. Then, other forces (Van der Waals, electrostatic, quantum [40], etc.) cause permanent adhesion.

From a general point of view, *stiction* is defined as the unintentional permanent adhesion of micromachined surfaces. The collapse can occur during the batch process or when the device operates. Indeed, the microstructured surfaces may come into contact unintentionally for instance in a humid environment or through acceleration or electrostatic forces, or intentionally in applications where surfaces touch each other (switches, shock sensors, etc.).

In the following, we describe first how capillary forces deflect a beam. Next, the stiction of beams is discussed. Finally, some solutions to stiction are described, focusing on the grafting of silanes. We present techniques that avoid both process-induced and in-use stiction for very-thin films, increasing the reliability of MEMS devices.

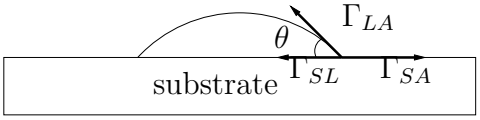


Figure 4.22: Droplet on substrate and definition of contact angle θ.

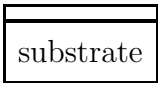


Figure 4.23: A liquid film on a substrate with $\theta = 0$.

4.4.1 Basic interfacial theory

For the sake of simplicity, we introduce here linear analysis, neglecting transient viscous forces, meniscus effects and liquid-solid interface effects.

4.4.1.1 Definition and basic concepts

Consider a gas phase and a liquid phase in contact to each other. The system can be described by a homogeneous gas, a homogeneous liquid and a superficial phase. The *Surface energy* $[J/m^2]$ is defined as the work needed to increase the area of the contact surface of an infinitesimal quantity dA. It characterizes the forces of an atom on its neighbor atoms.
Surface tension Γ [N/m] is the force by unit of length necessary to maintain into contact the two parts of the cut of an imaginary line. The surface energy is given by ΓdA.

At the interface A-B between a solid and a liquid, the *surface energy of the interface* Γ_{AB} is the difference between energies Γ_A and Γ_B provided to create free surfaces in A and in B and the bounding energy W_{AB} recovered when atoms of A and B were put into contact :

$$W_{AB} = \Gamma_A + \Gamma_B - \Gamma_{AB}. \tag{4.13}$$

For a drop of liquid placed on the surface of a solid (Fig. 4.22), the *contact angle* θ is determined by the balance between the liquid-air (γ_{LA}), the solid-air (γ_{SA}), and the liquid-solid (γ_{LS}) interfacial tensions. At equilibrium these tensions satisfy Young's equation:

$$\Gamma_{SA} = \Gamma_{SL} + \Gamma_{LA}\cos\theta, \qquad 0 < \theta < \pi. \tag{4.14}$$

If $\Gamma_{SA} > \Gamma_{SL} + \Gamma_{LA}$, it is energetically favorable for the liquid to spread on the surface of the solid with $\theta = 0$ (Fig. 4.23).

In Fig. 4.24, a drop of liquid forms a bridge trapped between two circular rigid

plates of radius r_o spaced by a gap $z << r_o$. If the liquid does not spread on the plates, the drop is confined to a radius $r_l < r_o$. The surface energy (i.e., the energy stored at the interface) is given by

$$\begin{aligned} U_S &\simeq 2 \cdot (A_{SA}\Gamma_{SA} + A_{SL}\Gamma_{SL}) && (4.15)\\ &= 2\pi(r_o^2 - r_l^2)\Gamma_{SA} + 2\pi\Gamma_{SL}r_l^2 && (4.16)\\ &= 2\pi \cdot (r_o^2\Gamma_{SA} - \Gamma_{LA}r_l^2)\cos\theta && (4.17) \end{aligned}$$

where A_{SA} and A_{SL} are the areas of the solid-liquid interfaces. To obtain this equation, we considered $A_{SA} << A_{SL}$. If $\theta = 0$, the liquid spreads, yielding

$$\begin{aligned} U_S &= 2 \cdot (A_{LA}\Gamma_{LA} + A_{SL}\Gamma_{SL}) && (4.18)\\ &= 2\pi \cdot (r_o^2(\Gamma_{SL} + \Gamma_{LA}) - \Gamma_{LA}r_l^2). && (4.19) \end{aligned}$$

In general, the stored surface energy U_S is of the form [39]:

$$U_S = C(\theta) - 2\pi\Gamma_{LA}r_l^2\cos\theta \tag{4.20}$$

where $C(\theta)$ is a constant.

It is important to note that both of these configurations are not at equilibrium since the liquid is in state of tension. If the bottom plate remains stationary, a force Q must be applied to the upper plate to maintain the equilibrium:

$$Q = -\left(\frac{dU_S}{dz}\right) = -\left(\frac{dU_S}{dr_l}\right)\left(\frac{dr_l}{dz}\right). \tag{4.21}$$

Since the volume of the drop, $V_l = 2\pi r_l^2 z$ is constant, then $dr_l/dz = -r_l/2z$, and with (4.20) and (4.21), we can write the expression of the **capillary pressure** q exerted on the rigid plate:

$$q = -\frac{Q}{\pi r_l^2} = \frac{2\Gamma_{LA}\cos\theta}{z}. \tag{4.22}$$

The term $\Gamma_{LA}\cos\theta$ is called the *adhesion tension*. Note that equation (4.22) is nothing else that the Laplace pressure applied to the case of structures with lateral dimensions much larger than the vertical spacing[6]. Pressure q is attractive or repulsive depending on the sign of $\cos\theta$ and is nonlinear with respect to the spacing z with a singularity at $z = 0$.

[6]The Laplace pressure P_L is created by the liquid meniscus at interface and $P_L = \Gamma_{LA}\left(\frac{1}{r_\parallel} + \frac{1}{r_\perp}\right)$ with $r_\parallel$ and $r_\perp$ the radii of curvature parallel and perpendicular to the beam, respectively.

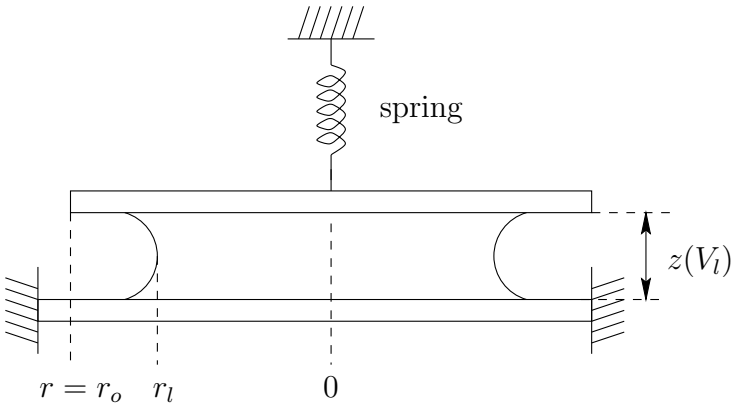


Figure 4.24: Droplet trapped between two circular plates.

In other words, from (4.22) the capillary force F becomes zero at $\theta = 90^o$ and turns to be negative for $\theta > 90^o$. On hydrophobic surfaces, for which the contact angle is greater than 90^o, the attracting process does not occur. That means that the structure is not attracted to the substrate if $\Gamma_{SA} < \Gamma_{SL}$.

4.4.1.2 Capillary forces on beams

In order to study the stability of a rigid plate attached to a spring we will refer to the simple beam model of Fig. 4.24 and follow the analysis of Mastrangelo [39]. We suppose that the weight of the plate is negligible[7] and that the original plate separation is $z = h$ when the spring (of constant k) is relaxed. A liquid bridge of volume V_l is trapped between the plate and the substrate and initially, the liquid spreads to a radius r_l so that $V_l = \pi r_l^2 z$. The maximum volume that the liquid assumes without overflowing is $V_o = \pi r_o^2 h$.

The total energy U_T of the system during the drying process can be expressed as the sum of the surface energy of the liquid U_S and the stored elastic energy in the beam U_E:

$$U_T = U_E + U_S. \tag{4.23}$$

As the liquid dries, the droplet volume V_l decreases, and the second term in (4.23) falls down. The drying process will proceed in such a way to continually lower U_T; this means that the elastic energy cannot increase faster than the decrease in surface energy.

In the case of Fig. 4.24, the surface energy of the spring-plate-liquid system is

$$U_S \simeq 2\pi r_o^2 \Gamma_{SA} - 2\pi \Gamma_{LA} r_l^2 \cos\theta \qquad \text{if } z \geq z^*$$

[7]for instance, the gravity force on a 400 μm x 400 μm, 2 μm thick polysilicon beam is equal to 7.3 nN, which is several order of magnitude smaller than typical capillary forces.

$$= U_{So} + 2\pi\Gamma_{LA}\cos\theta(r_o^2 - r_l^2) \tag{4.24}$$

$$U_S = U_{So} + \pi\Gamma_{LA}(\cos\theta - 1)(r_o^2 - r_l^2) \qquad \text{if } z \leq z^* \tag{4.25}$$

where $U_{So} \simeq 2\pi r_o^2\Gamma_{SL} + 2\pi r_l\Gamma_{LA}$ is a constant, and $z^* = V_l/(\pi r_o^2)$ is the plate separation assumed when the liquid wets completely the surface of the top plate. At $z = z^*$, U_S has a breakpoint, and for $z < z^*$, the liquid overflows. In the $z > z^*$ case, the elastic and the total initial energies are given by:

$$U_E = \frac{1}{2}k(h - z)^2 \tag{4.26}$$

$$U_T = U_{So} + 2\pi\Gamma_{LA}\cos\theta\left(r_o^2 - \frac{V_l}{\pi z}\right) + \frac{1}{2}k(h - z)^2. \tag{4.27}$$

The equilibrium plate spacing minimizes U_T. There are one or two minima. The first one develops at the breakpoint of U_S, implying that the equilibrium radius is r_o. The other one results when $dU_T/dz = 0$ i.e., from (4.23) along the curve

$$z^2(z - h)k - 2\Gamma_{LA}\cos\theta V_l = 0. \tag{4.28}$$

As the liquid volume decreases from $V_l = V_o$ to null, the reachable minimum of U_T determines a path called the *equilibrium trajectory*. This path determines the equilibrium position and final state. It depends on the control parameters changes. We call *catastrophes* sudden changes in the equilibrium for these values of the parameters at which a local minimum disappears by combining with a local maximum.

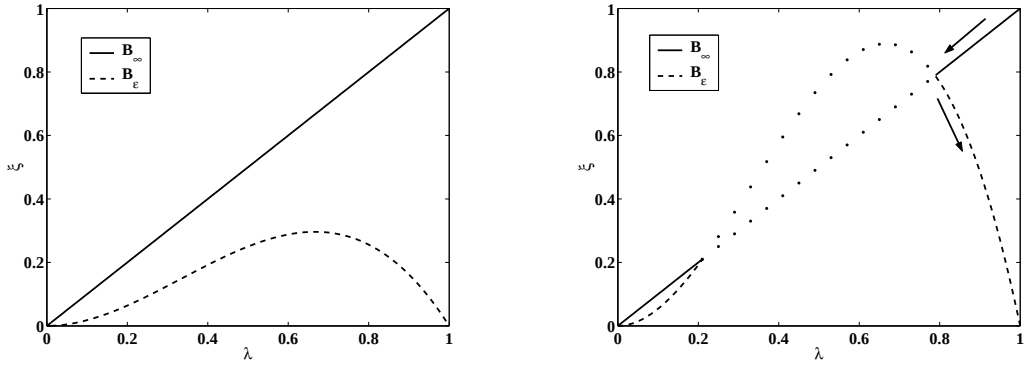
It is interesting to analyze the evolution of the local minima during the drying process. Such curve is called a *branching diagram*. Using the normalized variables $\xi = V_l/V_o$ and $\lambda = z/h$, two branches associated with the two extrema of U_T are found. The first one, B_∞, correspond to the breakpoint of U_S:

$$B_\infty : \xi = \lambda. \tag{4.29}$$

From (4.28), the second branch B_ε is:

$$B_\varepsilon : \xi = N_C(1 - \lambda)\lambda^2 \tag{4.30}$$

where $N_C = \frac{kh^2}{2\pi\gamma_{LA}\cos\theta r_o^2}$ is a nondimensional number. Branch B_ε is a minimum of U_T for $\lambda > 2/3$ and a maximum for $\lambda < 2/3$. The two branches intersect when



(a) $N_C = 2 < 4$ (b) $N_C = 6 > 4$

Figure 4.25: Branching diagram and influence of N_C on the equilibrium state.

$\xi = \lambda = N_C(\lambda - 1)\lambda^2$, i.e., for

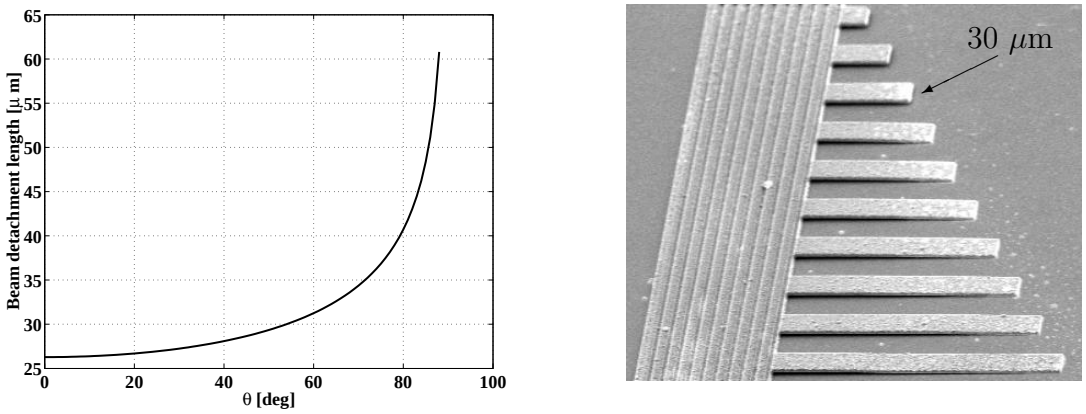
$$\lambda = \lambda_{1,2} = \frac{1}{2} \pm \sqrt{\frac{N_C - 4}{4N_C}}. \tag{4.31}$$

Depending on the value of N_C, two cases appear from (4.31). First, if $N_C < 4$, λ_1 and λ_2 are complex and the branches do not intersect (Fig. 4.25(a)). The equilibrium trajectory, starting from $(\lambda, \xi) = (1, 1)$ reaches $(0, 0)$ following the branch B_∞ because there is an energy barrier between the two branches. This means that the plate is pinned to the substrate at the end of the drying process. Second, for $N_C > 4$, the equilibrium trajectory starts from $(1, 1)$ and follows the branch B_∞ until the intersection with B_ε (Fig. 4.25(b)). Then, it is energycally favorable to follow the branch B_ε. In this case, the plate does not reach the substrate. We conclude that there is a threshold value of N_C, defined as N_T, that determines the final state ($N_T = 4$). From that analysis, it is straightforward to compute the maximum radius the plate may have to remain free. Indeed, this value is found when N_C equals N_T. In practice, the free/pinned transition is not sharp. If the plates acquire sufficient kinetic energy through agitation to overcome the energy barriers, the plate has a non-zero probability of transition between the various equilibrium states. Residual stresses have been taken into account in the calculation of the elastocapillary number N_T in [41].

Following this analysis, Mastrangelo [39] gave the *detachment length*[8] l_d of rectangular cantilevers, including a correction for the meniscus effect:

$$l_d^4 \simeq \frac{8EIh^2}{3\gamma_{LA}\cos\theta(w + t)} \tag{4.32}$$

[8]The detachment length is the maximum length that a free beam has.



(a) (b)

Figure 4.26: Detachment length of a 1 μm thick, 40 μm width silicon cantilever beam of rectangular section on a 1 μm gap calculated from (4.32). (a) versus he contact angle; Γ_{LA} was fixed to 72 mJ/m^2. (b) Experimental result, a 30 μm detachment length and 50^o contact angle were measured.

where l, t and w are respectively the length, the thickness and the width of the beam, E is the Young's modulus and I the moment of inertia. The detachment length for cantilevers is about 2.5 times smaller than that for doubly clamped beams. Note that this analysis is not valid for nonlinear displacement.

For example, the detachment length of a 1 μm thick, 10 μm width silicon cantilever beam of rectangular section on a 1 μm gap calculated by (4.32) is reported in Fig. 4.26(a) as a function of the contact angle. This simulation is in agreements with experimental results using the beams array technique (Fig. 4.26(b)). Some engineering techniques were found to limit the impact of capillary forces. For instance, higher contacts angles are obtained by increasing the rugosity [42]. Fluorocarbon film coating [43] or drying step involving critical point drying [44] also showed good results. In the two following sections, others techniques to avoid stiction will be discussed, namely the vapor HF technique and the silane grafting method.

4.4.2 Vapor HF etching technique

A way to prevent from stiction is to use the etchant in the vapor-phase instead of the wet-phase. This technique is very attractive because it replaces the sequence of etching, rinsing and elaborated drying. Furthermore, it does not requires sophisticated instrumentation. On the other hand, the price to pay is a lower etch rate compared to the wet process. The vapor HF etching technique, coupled with electron beam lithography is further an interesting alternative for etching nanometric structures, where wet and plasma etching techniques becomes inefficient [45].

In practice, condensed water appears on the etching surface. This can be avoided through heating the wafer [46], or using a HF/methanol (CH_3OH) mixture [47]. In a conventional aqueous HF process, the SiO_2 is etched when the highly polar molecule attacks the highly polar Si-O bonds by insertion between the Si and O atoms. This leads to the cleaving of the Si-O-Si bonds along with formation of thermodynamically stable products, such as SiF_4 :

$$SiO_2 + 4HF \overset{OH}{\rightarrow} SiF_4 + 2H_2O.$$

In the case of vapor HF/methanol chemistry, polarization is provided by the hydroxyl groups of methanol molecules. However, the polarization is not as strong as that provided by water, and oxide etch rates are generally slower than in the HF/H_2O system.

As in the wet phase case, it is necessary to form a condensed layer on wafer surface to initiate the etching process. When vapor HF/methanol is used, the condensed layer is expected to be thinner than in the case of vapor HF/H_2O because of the higher methanol vapor pressure [48]. However, experimental results showed that etching may take place even when condensation is not expected [48]. In this case, the etch rates are sensibly lower ($0 - 300$ Å/min) than in the presence of a condensation layer ($3000 - 12000$ Å/min in this case).

Both the partial pressure of HF and H_2O, and the mass transport of the products of the sample in the gas phase determines the condensed layer. A high etch rate is achieved at low temperature difference between the substrate and the gas. The etch mechanism is explained by the adsorption of H_2O and HF at the oxide surface. The amount of the adsorbed water influences the ionization efficiency of HF that is necessary to react with SiO_2 to form volatile SiF_4 and H_2O. The temperature of the oxide also influences the amount of adsorbed water. Anguita et al. [46] reported that for $\Delta T = 11.5^oC$ the temperature is low enough to form the initial very thin layer that allows oxide etching. This temperature is high enough to desorb the excess of water generated in the etch reaction. Therefore, the thickness of the liquid layer is kept constant, leading to a constant etch rate.

A high selectivity toward metal-based films is obtained. On the other hand, HF vapor is not a good choice if nitrides or PSG are present on the surface. This technique will be used in the next section to release successfully thin-films microstructures, without process-induced stiction damages.

4.4.3 Silanes

The vapor HF technique described in section 4.4.2 eliminates the process induced stiction. However, the in-use stiction issue remains. As explained in section 4.4.1.2, both process-linked and in-use stiction are eliminated if the surface functionality of the beams moves from hydrophilic to hydrophobic, limiting the impact of the capillary forces. For that purpose, thin organic films may be coated on the structure surfaces [49, 50, 51]. Indeed, grafting an alkane or a fluoro silane will drastically decrease the surface energy of the original (native) oxide layer and give strong hydrophobic properties to substrates. Under optimum conditions, the silanes can form a self-assembled monolayer (SAM). This technique has the further advantage to reduce wear and friction in micro-engines [52].

Various types of chemical modifications have been explored as potential anti-stiction treatment for MEMS. Nevertheless, most research in that field focuses on aliphatic chloro-silanes, such as octadecyltrichlorosilane (OTS) and 1H,1H,2H,2H - perfluorodecyltrichlorosilane (FDTS). Dialkyldichloromethylsilane (DDMS) was reported to have better chemical stability than monoalkyltrichlorosilane and simplified coating process [53]. To avoid some of the limitations imposed by the chloro-silane chemistry, alkene based monolayers were successfully coated as anti-stiction layers in [54], and [55] reports the coating of non-chlorinated films by atomic layer deposition.

Because of its simplicity and low cost, much effort has been provided to liquid phase silane coating. More recently, vapor phase anti-stiction coatings were obtained [56], presenting better reproducibility and portability. Moreover, less polymerization inducing parasitics blobs was observed. Vapor phase coating is also better adapted to the industrial requirements. This technique is only valuable to avoid in-use stiction, as the samples should be unstuck before the coating.

In the present section, both liquid and vapor phase release processes are investigated. The gas phase process combines the vapor HF release and the coating of a silane from the gas phase. It has the further advantage to release thin-film SOI structures, where the proposed liquid process failed. Thus, three different silanes were grafted for this study onto MEMS and homogeneous wafers, and two types of silanization were tested, one from the gas phase and the second one in dry isopropanol. The quality of the silane monolayers was checked by using X-ray reflectrometry (XRR) and water static contact angle.

4.4.3.1 Experiments

Trifunctional silanes (trichloro, trimethoxy, etc.) give rise to stable monolayers with a better coverage of the surface compared to monofunctional compounds. This is due to the fact that there are two chemical mechanisms involved: the hydrolysis of the silane with water and its condensation to build a siloxane bidimensional network which partially grafts onto the silanols of the oxide layer. This polymerization stabilizes the monolayer, but tridimensional gel blobs often form, surrounded by the real monolayer. Thus, the roughness and physiochemical properties are more difficult to control. As water controls the hydrolysis and subsequent polymerization, it is critical to limit the amount of water during monolayer formation. The blobs may also cause the MEMS failure. The amount water can be fixed in a glove box with a dry nitrogen or argon atmosphere, whereas adsorbed water into the oxide layer of substrate can only be removed at high temperature (i.e., $600 - 800^{o}$C) under vacuum. Unfortunately, those high temperature steps are not compatible with a CMOS post-processing.

For any chloro-silane based precursor molecules, the first step in the reaction sequence is the hydrolysis of the Si-Cl bonds. One equivalent of HCl is thus created for each Si-Cl bond that is hydrolysed. The presence of HCl in the vapor or the solution might corrode metallic surfaces, such as aluminum interconnects of the circuitry [54]. Besides, chloro-silanes are extremely sensitive to humidity and the quality of the film significantly depends on temperature and coating time, leading to reliability problems. By contrast, methoxy- or ethoxy-silanes only produce methanol or ethanol upon hydrolysis which do not damage microsystems. However, they react more slowly due to their lower sensitivity to hydrolysis.

In this work, N-propyltrimethoxysilane (1), heptadecafluoro triethoxysilane (2) and dimethylchlorosilane (3) have been tested for MEMS applications (Fig. 4.27). Compound 1 and 2 were grafted in dry isopropanol in a glove box under argon atmosphere, as ethoxy and methoxy silanes are less sensitive to water compared to chlorosilanes. The compounds 2 and 3 were also grafted from the gas phase. However, only compound 3, the cholorosilane, gave rise to neat monolayers in these conditions.

Liquid phase process The sacrificial layer of our MEMS devices consists of silicon dioxide. Before coating the silane, this layer is removed in a hydrofluoric acid (HF) solution. The device must remain wet until the anti-stiction layer is deposited, in order to avoid the collapse of the suspended structures. As already mentioned, if aluminum is present, concentrated HF should be used to improve the selectivity

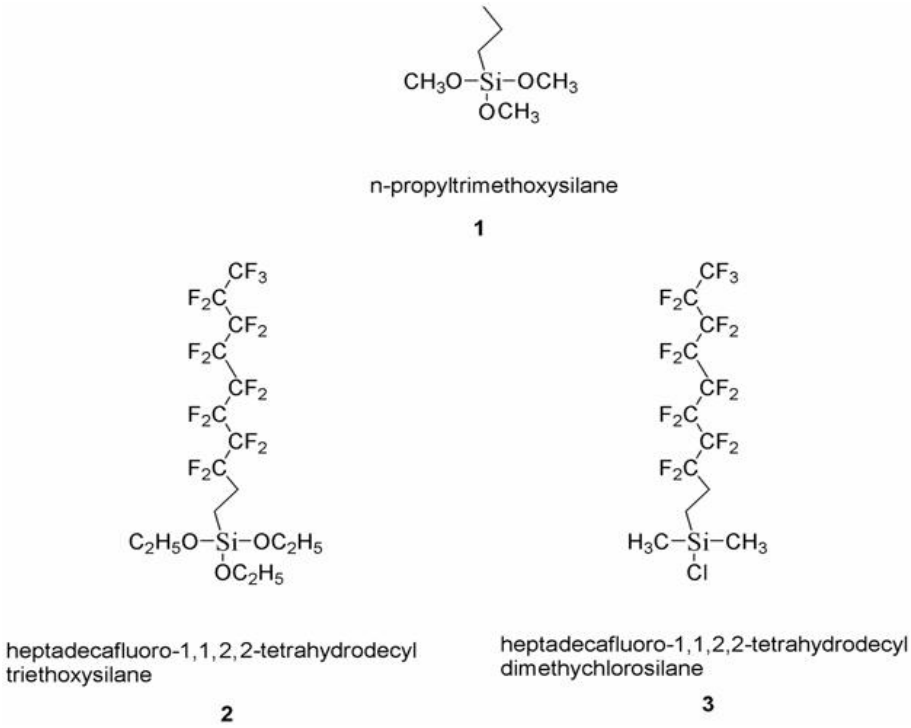


Figure 4.27: Structural formulas of the three compounds used in this study.

towards this metal. Subsequent rinses are performed in isopropyl alcohol for two reasons: it is a low-tension fluid and it prevents from attack of Al.

Compounds 1 and 2 are then deposited the same way [57]. First, the samples are oxidized through immersion in H_2O_2 for 10 min. Hydrogen peroxide is then displaced by isopropanol. Next, the wafers are put in a 1 : 600 mixture of the silane and isopropanol (dried by distillation on CaO) in a dry argon environment (glove box). The reaction is stopped by immersion in acetone. The products of polymerization are eliminated if necessary by a dichloromethane soxhlet extraction process.

This complete release process was applied to several polysilicon beams. The polysilicon was deposited in a LPCVD oven at 625^{o}C with a 90 sccm SiH_4 flow. Also, 100 nm thick monocrystalline silicon beams on a 400 nm thick SiO_2 sacrificial layer available from a standard SOI wafer were prepared as follows. The Si beams were patterned by optical lithography process and a SF_6 based RIE. On the other hand, samples of bulk silicon, LPCVD polysilicon, PECVD SiO_2, aluminum and PEVCD silicon nitride deposited on bulk silicon wafer were prepared. Those samples followed the same process as the polysilicon beams but only from the oxidation step in H_2O_2. The bulk silicon reference wafers were cleaned and oxidized through immersion in a H_2SO_4:H_2O_2 mixture before the silanization.

Gas phase process A vapor HF setup [58] is used to release the sacrificial SiO_2 layer of the SOI MEMS structures at 50^{o}C for 200 min. Native silicon oxide was then grown in air ambient. Bulk samples of several materials were also prepared. The gas phase silanization of compound 3 is performed at 70^{o}C for 24 h and is described in details in reference [59]. Special care was given to progressively and gently apply overpressure and vacuum in the gas phase reactor in order to avoid the collapse of the MEMS structures. Again, the residues of polymerization are eliminated by a dichloromethane soxhlet extraction.

4.4.3.2 Process characterization

The thickness and the quality of the coating were characterized by XRR, while the hydrophobicity level was characterized by water static contact angle measurements.

X-ray reflectrometry Homogeneous silanized Si sample were routinely characterized by X-ray reflectrometry. The experimental setup is based on a Siemens D5000 2-circles goniometer of 30 cm radius and 0.002^o positioning accuracy. Data are reported as a function of k_{z0}, the vertical component of the wave vector of the incident photons in a vacuum.

XRR data were analyzed in two ways. First, a Patterson function was computed as described elsewhere [60], and the average thickness of the monolayer was obtained from the position of the subsidiary correlation peak in this function. Second, a model of electron density was fitted to the data. The electron density profile was discretized as a succession of flat slabs of 0.5 nm width, approximately corresponding to the resolution of the experiments. The number of slabs required was initially selected from the known thickness of the monolayer determined from the Patterson function. It was progressively incremented to take into account the width of the interfaces, until the fit was deemed satisfactory as judged from the value of chi-square obtained. The fit parameters were the electron density of each slab, a general scaling parameter, and a constant background value. The reflectivity was computed using Parratt's formalism [61].

Fig. 4.28 presents X-ray reflectograms of a bare $< 100 >$ silicon wafer and silanized wafers. After one hour reaction from the liquid phase, compound 1 forms a 1 nm thick layer, much higher than the extended length of the silane ($0.3 - 0.4$ nm), indicating the formation of a gel layer as opposed to a neat monolayer. Under identical reaction conditions, compound 2 forms a 1.2 nm thick monolayer, in good agreement with the extended length of the molecule (1.25 nm). However, the electron

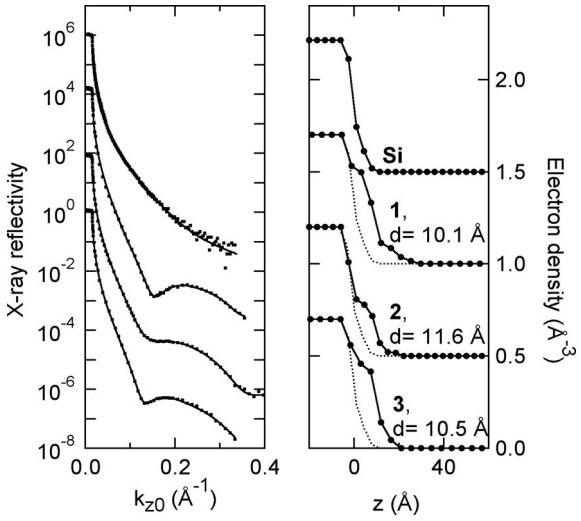


Figure 4.28: Left. X-ray reflectograms of, from top to bottom, a bare $< 100 >$ silicon wafer, and wafers silanized with compounds 1, 2 and 3. All silanization reactions were performed for 1 hour in the liquid phase, except for compound 3 (gas phase, 24 hours reaction time). Squares are experimental data, continuous lines are fits to the data based on the electron density profiles shown at the right. Right. Electron density profiles corresponding to the samples of the left panel. The dotted lines are the profile of the bare silicon wafer.

Table 4.3: Static water contact angles of various silanized materials.

	Silicon	Polysilicon	SiO_2	Si_3N_4	Al
Compound 1	54^o	-	-	-	-
Compound 2	117.3^o	118.2^o	109.7^o	108.2^o	110.3^o
Compound 3	97.9^o	-	98.2^o	96.3^o	93.2^o

density of the film is relatively low for a perfluorinated compound, suggesting a loosely packed monolayer. In contrast, when compound 3 was deposited from the gas phase (24 h reaction), a 1.1 nm thick film of high electron density is formed, corresponding to a densely packed monolayer. This suggests that mono-chlorosilanes deposited from the gas phase may form better layers than the corresponding tri-ethoxy-silane from the liquid phase.

Contact angle measurements After the coating, a high degree of hydrophobicity was obtained for compounds 2 and 3. The static contact angle was measured by deposition of a 0.5 μl ultra pure water droplet on 10 different locations on each sample. Mean values are given in Table 4.3. Note that without the monolayer, all those materials exhibit contact angles as small as 10^o (value for the oxide layer).

Temperature dependence The last processing step of any microsystems is the packaging. Any layer deposited before this step must satisfy the thermal budget requirements of the packaging. The thermal stability of the monolayer was determined by exposing it at high temperatures during at least one hour in nitrogen ambient. Experimental results showed that hydrophobicity of compound 2 and 3 is maintained up to 300^{o}C. XRR confirmed the presence of the monolayer after the annealing.

Micro-beams fabrication As already mentioned, the wet release process (compound 2) was used to fabricate polysilicon micro-beams of various lengths. One and two microns thick, up to 540 μm long cantilevers and clamped-clamped beams on a 0.5 μm SiO_2 sacrificial layer were successfully released (Fig. 4.29). Without the monolayer, even for beams as short as 45 μm, stiction problems occurs. The work of adhesion W of the coated beams was evaluated to be $W = 1.75$ μJm^{-2} using equation (4.32).

Furthermore, a complete 1 μm thick (gap 0.5 μm) polysilicon RF MEMS capacitor with aluminum interconnects realized with this process demonstrated the CMOS compatibility, as it will be shown in section 4.5. Unfortunately, experiments showed that this wet process induces the collapse of the very thin structures ($\leq$ 100 nm) from the SOI substrate. Indeed, the restoring forces on a beam may be modeled to the first order by a spring. As the spring constant is proportional to the third order of the thickness of the beam, thin beams are more sensitive to the collapse than thicker ones. The gas phase process allows to release properly such thin structures. Indeed, we observed that all MEMS structures from SOI that did not collapse during the vapor HF release step are maintained unstuck after the silane coating (Fig. 4.30). As the hydrophobicity is increased, this process allows reducing in-use stiction failures of MEMS devices.

4.5 Study-case: A MEMS Capacitor

The performances of microwave circuits as tunable filters and oscillators strongly rely on their quality factor Q. Mechanical devices have already been shown to reach higher Q than those based on semiconductors junctions [1]. Furthermore, as these devices are electrostatically actuated, the static power consumption is in theory equal to zero. Moreover, micromachined tunable capacitors are high linear and they can accommodate larger voltage swings than P-N junctions or MOSFET varactors [62]. On the other hand, the mechanical resonance frequency lies typically

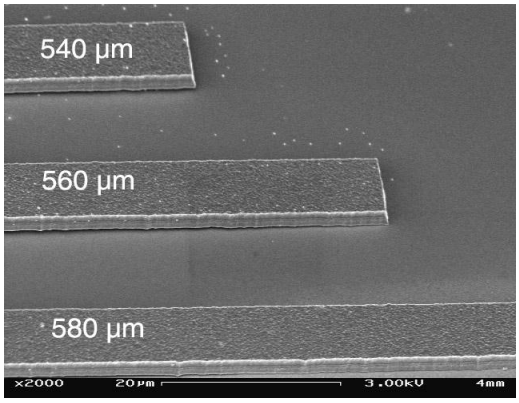


Figure 4.29: Cantilever beams released with the proposed wet process (compound 2).

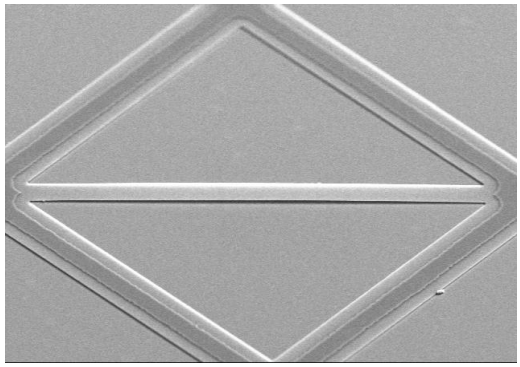


Figure 4.30: A 100 nm thick 200 μm long silicon beam successfully released with the proposed gas phase process (compound 3).

in the $10-100$ kHz band, and are not expected to respond to RF frequency for that reason. MEMS capacitors are thus a good choice for improving the performances of RF circuits, such as tunable filters or voltage-controlled oscillators (VCOs) [63].

In this section, it is explained how a MEMS tunable capacitor is fabricated with the micromachining techniques developed previously in this chapter. Before going to process explanations, the principle of operation of this device is introduced, as well as the design techniques.

4.5.1 Principle of operation

Fig. 4.31 shows a functional model of an electro-mechanically tunable capacitor that consists of two parallel plates. The top plate of the tunable capacitor is suspended by a spring with spring constant k, while the bottom plate of the capacitor is mechanically secured. When a bias voltage V_B is applied across the capacitor plates, the suspended plate is attracted toward the bottom plate due to the electrostatic force. The suspended plate moves toward the fixed plate until equilibrium between the electrostatic and the spring forces is reached. The capacitance C of such parallel plate capacitor is given by the well-known expression:

$$C = \frac{\varepsilon A_S}{d} \tag{4.33}$$

where ε is the dielectric constant of the dielectric between the two plates i.e., air in this case, and $d = d_o - x$ is the separation between the plates, d_o being the separation between the plates when the spring is in its relaxed state.

The two forces acting on the top plate are the spring force F_s and the electrostatic force F_e due to the applied DC voltage V_B between the plates. The spring force

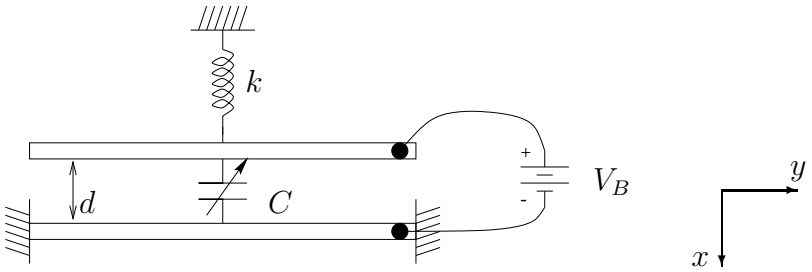


Figure 4.31: Principle of a two-plates tunable capacitor.

follows Hooke's law:

$$F_s = kx, \tag{4.34}$$

while the electrostatic force is given by the gradient of the energy stored in the capacitor:

$$F_e = \frac{1}{2}\frac{dC}{dx}V_B^2 = -\frac{1}{2}\frac{\varepsilon A_S V_B^2}{d^2}. \tag{4.35}$$

The net force on the beam is thus given by $F_{net} = F_e + F_s$. At equilibrium, we have $F_{net} = 0$, leading to

$$kx = \frac{1}{2}\frac{\varepsilon A_S V_B^2}{d^2}, \tag{4.36}$$

$$d = d_o - \frac{\varepsilon A_S V_B^2}{2kd^2}. \tag{4.37}$$

There is an implicit feedback in this equation. As the voltage is increased, the gap decreases, with the amount of decrease growing as the gap gets smaller. Thus, at some critical voltage, the system goes unstable, and the gap collapses to zero. This phenomenon is called *pull-in*. Let us now compute the voltage V_{PI} at which the pull-in occurs. For a small perturbation of the gap to $d + \delta d$, F_{net} varies as follows:

$$\delta F_{net} = \left.\frac{\partial F_{net}}{\partial d}\right|_{V_B} \delta d \tag{4.38}$$

$$= \left(\frac{\varepsilon A_S V_B^2}{d^3} - k\right)\delta d. \tag{4.39}$$

If δF_{net} is positive for positive δd, then d is an unstable equilibrium point, because a small increase δd creates a force tending to increase it further. If δF_{net} is negative, then d is a stable equilibrium point. Thus, in order to get a stable equilibrium, from (4.39), the following condition must be satisfied:

$$k > \frac{\varepsilon A_S V_B^2}{d^3} \tag{4.40}$$

At pull-in, two conditions must be satisfied: the original requirement that $F_{net} = 0$, and the new requirement that

$$k = \frac{\varepsilon A_S V_{PI}^2}{d_{PI}^3}. \tag{4.41}$$

From those requirements, we deduce that pull-in occurs when

$$d_{PI} = \frac{2}{3} d_o, \tag{4.42}$$

$$V_{PI} = \sqrt{\frac{8kd_o^3}{27\varepsilon A_S}}. \tag{4.43}$$

This property limits the tuning range of the capacitor. Indeed, a stable equilibrium only exists between $0 \geq x \geq d_o/3$. Note that the equilibrium would exist for $0 \geq x \geq d_o$ if the capacitor were biased by a charge source instead of a voltage source [64]. It is however difficult to implement in practice [65].

4.5.2 Mechanical properties of clamped-clamped beams

For small deflections, the mechanical behavior of a clamped-clamped beam can be modeled by a spring constant k. This spring constant is composed of two parts [66]. One part, k', is due to the stiffness of the bridge which accounts for the material characteristics such as Young's modulus E and the moment of inertia I. The other part of the spring constant, k'', is due to the biaxial residual stress σ within the beam. As mentioned in section 4.3, σ is a result of the fabrication process used to define the beam.

For the beam of Fig. 4.32, the constitutive equation writes [67]:

$$EI\frac{d^2y}{dx^2} = M_A + R_A x \qquad \text{for } x \leq a \tag{4.44}$$

$$y = \frac{M_A x^2}{2EI} + \frac{R_A x^3}{6EI} \qquad \text{for } x \geq a \tag{4.45}$$

$$M_A = -\frac{Pa}{l^2}(l-a)^2 \tag{4.46}$$

$$R_A = \frac{P}{l^3}(l-a)^2(l+2a) \tag{4.47}$$

where l is the length of the beam, P is the load force, M_A is the reaction moment at the left end, and R_A id the vertical reaction at the left end. The moment of inertia I of a rectangular cross-section beam of width w and thickness t is

$$I = \frac{wt^3}{12}. \tag{4.48}$$

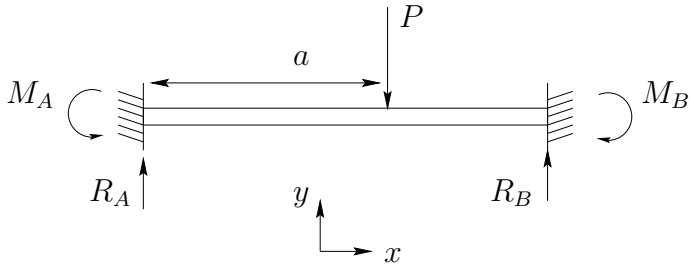


Figure 4.32: Clamped-clamped beam with concentrated vertical load P.

In MEMS applications, the load is typically distributed across the beam and the deflection of the beam at the center is used to determine the spring constant. By substituting $x = l/2$ into (4.45), the deflection at the center is found for a concentrated load at point a. Then the spring constant is deduced from the deflection:

$$k' = -\frac{P}{y} = \frac{32Ewt^3}{l^3}. \tag{4.49}$$

The part of the spring constant due to the biaxial residual stress within the beam in the case of tensile stress, is given by [66]:

$$k'' = \frac{8\sigma(1-\nu)tw}{l} \tag{4.50}$$

where ν is the Poisson's constant.

As a consequence, the total spring constant k is equal to $k = k' + k''$ as given from (4.49) and (4.50). A more general expression given for a bridge over a bottom electrode of width W is [66]

$$k = \frac{32Ewt^3}{l^3}\left(\frac{1}{2-\left(2-\frac{W}{l}\right)\left(\frac{W}{l}\right)^2}\right) + \frac{8\sigma(1-\nu)wt}{l}\left(\frac{1}{2-\frac{W}{l}}\right). \tag{4.51}$$

4.5.3 Design of parallel plates tunable capacitors

Two design styles are commonly used to implement tunable MEMS capacitors: parallel-plate capacitors and interdigital lateral capacitors [62]. We restrict the present analysis to the first case. The principle of operation of this device was given in section 4.5.1. It was then explained that the top plate can move to a gap height of $2d_o/3$ before it collapses on the bottom plate (d_o is the separation between the two plates at zero actuation voltage). The capacitance ratio is therefore

$$\frac{C_{max}}{C_{min}} = \frac{\frac{\varepsilon A_S}{2d_o/3} + C_f}{\frac{\varepsilon A_S}{d_o} + C_f} \tag{4.52}$$

where $C_f = \gamma\varepsilon A_S/d_o$ is the fringing-field and parasitic capacitance associated with the bridge. Most MEMS varactors have $0.15 < \gamma < 0.6$, which yields a capacitance ratio of $1.42 - 1.27$. Also, the electrostatic actuation method results in a nonlinear $C - V$ response, and most of the capacitance change occurs at the highest control voltage. The tuning range may be improved by a three-plate structure, as demonstrated in [63]. The sharp capacitance change at pull-in may be used to implement RF switches or digital tunable capacitors. In order to avoid a current to flow between the electrodes, they should be separated by an insulating layer.

We designed several kinds of parallel plates tunable capacitors using clamped-clamped beams: single beams (Fig. 4.33(a) and 4.33(b)), several beams of same size connected in parallel (Fig. 4.33(c) and 4.33(d)), several beams of different sizes connected in parallel (Fig. 4.33(e)), and four-points anchored plates (Fig. 4.33(f)). The actuation DC voltage may be applied on the same electrode than the RF signal, or on a separated electrode, depending on the design (Fig. 4.33(a) and 4.33(b)).

Small holes (3 μm x 3 μm) are visible in the top plate in Fig. 4.33. There are two reasons for that. First, the etching of the sacrificial layer is facilitated, which means the possibility for dry etching or for decreased release time. Second, the squeeze film damping is reduced. The anchors were designed to diminish the extra-spring effect as proposed in [68].

The relations derived in section 4.5.1 were used to dimension the beams (Fig. 4.34). The results were refined by numerical simulations performed with the software Sugar v2.0 [69, 70]. The gap height was fixed to 0.5 μm in order to get low pull-in voltages (i.e., < 10 V). The capacitor variation between on- and off-states is typically a few pF, depending on the surface area of the devices. The predicted resonance frequency lies in the kHz range. The device in Fig. 4.33(e) is designed as a digitally tunable bank of capacitors, with a "stair" $C-V$ characteristics (Fig. 4.35). Detailed analysis would require further numerical simulations based on finite-element method.

4.5.4 Process flow

The capacitors designed in the previous paragraph were realized in the cleanrooms of the Microelectronics Laboratory at UCL. It consists of two polysilicon electrodes with aluminum interconnects and PECVD SiO_2 sacrificial layer. A silicon nitride layer was used to insulate both electrodes. The process is summarized in Fig. 4.36.

Process description Starting from a $< 100 >$ p-type bulk silicon wafer, a 450 nm thick oxide layer is let grown to provide electrical insulation. Then a 200 nm

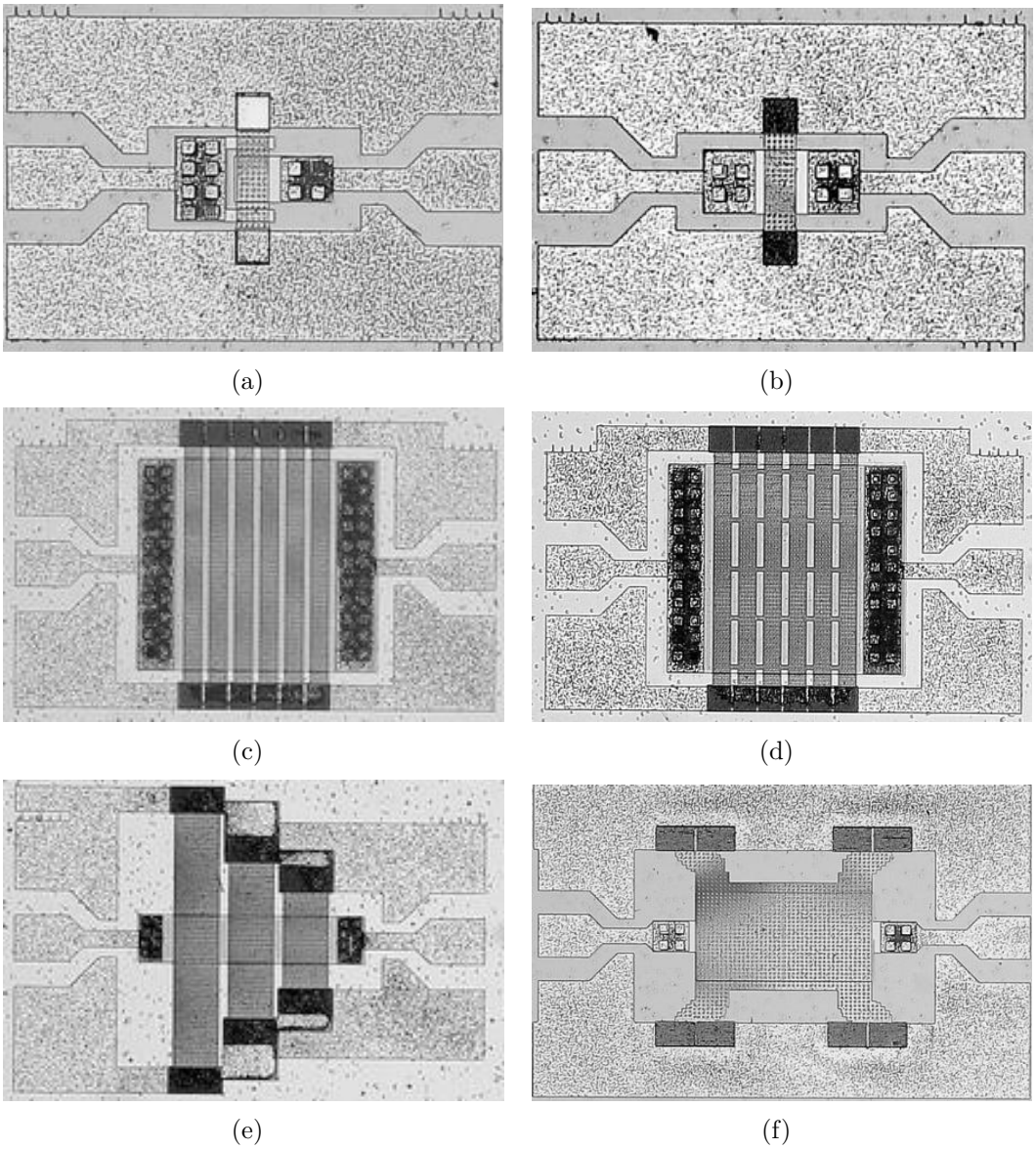


Figure 4.33: Different types of realized capacitors. (a) single beam with decoupled actuation access; (b) single beam; (c) several beams in parallel; (d) several beams interconnected in parallel; (e) bank of beams; (f) four-points anchored plate.

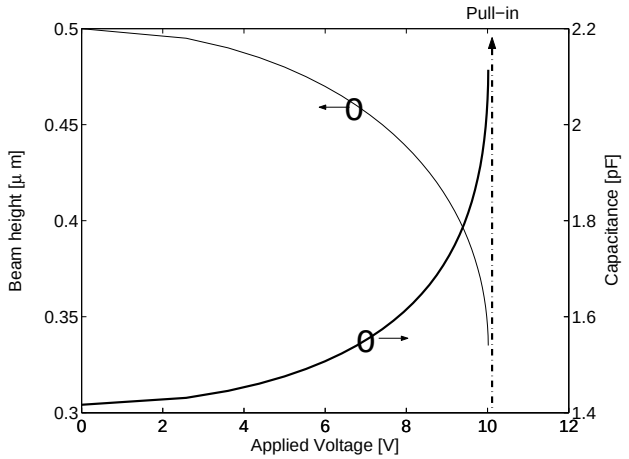


Figure 4.34: Deflection and capacitance of a Si clamped-clamped 200 μm x 400 μm beam (gap 0.5 μm), calculated by the formula of section 4.5.1.

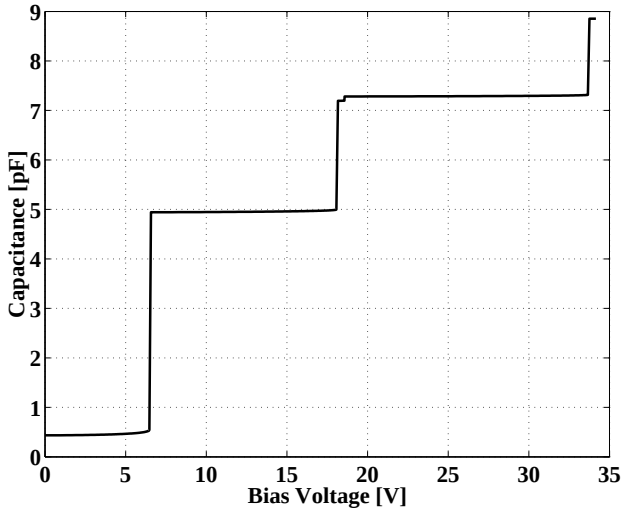


Figure 4.35: Tuning curve of the MEMS bank of tunable capacitors in Fig.4.33(e) calculated with the formula of section 4.5.1.

thick polysilicon layer is deposited at 625^{o}C in a LPCVD oven. This layer is doped by phosphorous diffusion to get a resistivity close to 100 Ω per square. The bottom electrode is then patterned by a $SiCl_4$ based RIE; some samples are silicidated with cobalt. An insulating 200 nm thick PECVD nitride layer (Fig. 4.36 (b)) and a 1 μm thick sacrificial PECVD oxide layers (Fig. 4.36 (c)) are then deposited and densified by a 20 s Rapid Thermal Annealing (RTA) step at 800^{o}C. The anchors of the beams are then determined by photolithography and plasma etching (Fig. 4.36 (d)). A CHF_3/Ar chemistry is used for etching the SiO_2 layer, while a SF_6 chemistry is used for the nitride etch. Next, the top polysilicon electrode is deposited by LPCVD and patterned by an SF_6 based RIE (Fig. 4.36 (f)). A 1100^{o}C annealing was performed in order to relax the stress in the polysilicon layer (see section 4.3). Measured resistivity after phosphorous diffusion is about 30 (60) Ω per square for the two (one) micron(s) thick polysilicon layers. Some samples are silicided with cobalt. Contact holes are then defined to reach the bottom electrode the same way the anchors were (Fig. 4.36 (g)). The aluminum metallization is next performed in order to get electrical access to the devices (Fig. 4.36 (h)). Finally, the structures are released following the wet procedure described in section 4.4. At total, five mask levels are needed.

Discussion From the MEMS point of view, the specific and most critical processing steps are the release and the annealing to control of the residual stress. Both of them were discussed in the previous sections. If the MEMS capacitor has to be co-integrated with the electronics on the same chip, some precautions have to be taken because of the high temperature steps involved in the process. Indeed, the mechanical devices have to be processed first, but not released. Then the electronic components of the system are built up before the release is performed. Passivation of the circuitry has to be performed properly, by a silicon nitride layer for instance.

The process also involves high topography with high aspect ratio with respect to classical microelectronical processes. In particular, the Al interconnects must recover steps of one or two micrometers. We found out that a good contact with the bottom electrode was only possible if the contact holes are filled in with metal by means of a lift-off. It is worth noting that this additional step does not require any further mask. Nevertheless, the electrical contact between Aluminum and the top electrode was not satisfying, preventing from using the coplanar access structures for the measurements.

Another mis-processing happened while etching the bottom polysilicon electrode: the etching was not completed. This leads to a shunt conductance between the electrodes. Even if the device may be actuated, this conductance affects tremendously

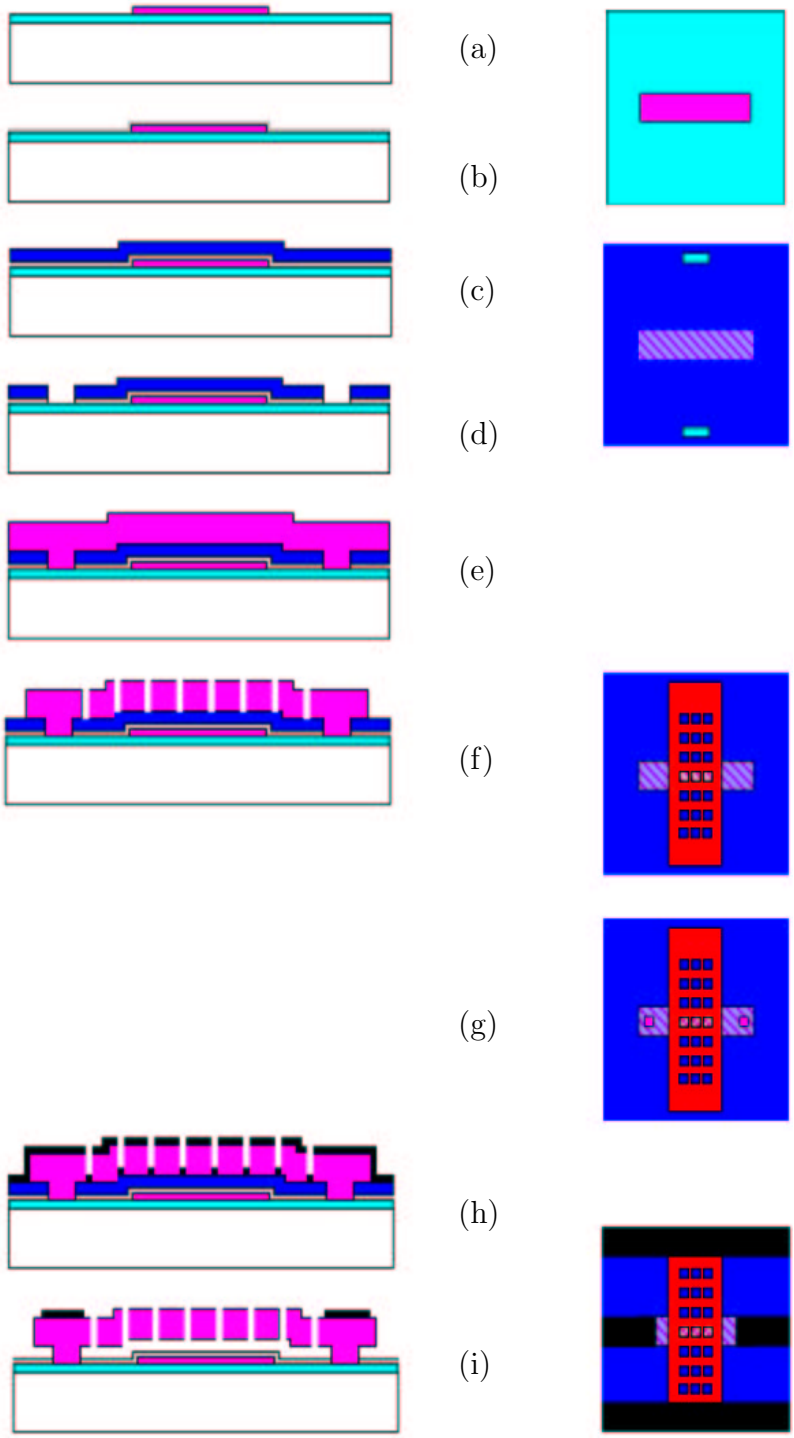


Figure 4.36: Main process steps. Left: cross-section; right: top view.

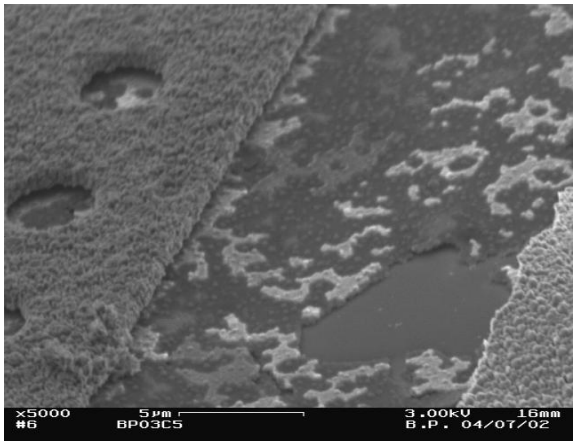


Figure 4.37: Effect of Co silicidation on top nitride layer after release.

its quality factor.

Cobalt silicidation was under investigation in this study. We showed in section 4.3.3 that silicidation is interesting for the fabrication of the top electrode. On the other hand, it was observed that the bottom silicidated electrode reacts with the PECVD silicon nitride to make white marks and pyramidal clusters that were difficultly etched by RIE. This lead to bad state of surface that in turns affected the SiO_2 layer deposited on it. We observed that this oxide layer was etched in the HF/IPA solution more rapidly in the regions above the silicidated polysilicon than elsewhere. As a consequence, over-etch occurred during the release step for the PECVD oxide and nitride in region surrounding the silicidated bottom electrode; the insulating oxide was then attacked. Since the stress in this layer was important because of its small thickness and lattice misfits with silicon nitride, cracks were observed on the bottom electrode (Fig. 4.37). In conclusion, Co silicidation should be avoided for the fabrication of the bottom electrode, while it may be used for the top electrode.

4.5.5 Characterization

SEM pictures confirmed the compatibility of non-cholinated silane coating with aluminum interconnects and that the release happens without stiction (Fig. 4.38).

As the polysilicon bottom electrode was not etched completely, both electrodes are connected. This parallel conductance permits a current to flow. The quality factor is then not characterizable. Nevertheless, we observed by microscopy that the plates are well actuated by DC voltages. The measured pull-in voltages have higher values than the simulations predicted, due to the presence of the parasitic shunt conductance and compressive stresses.

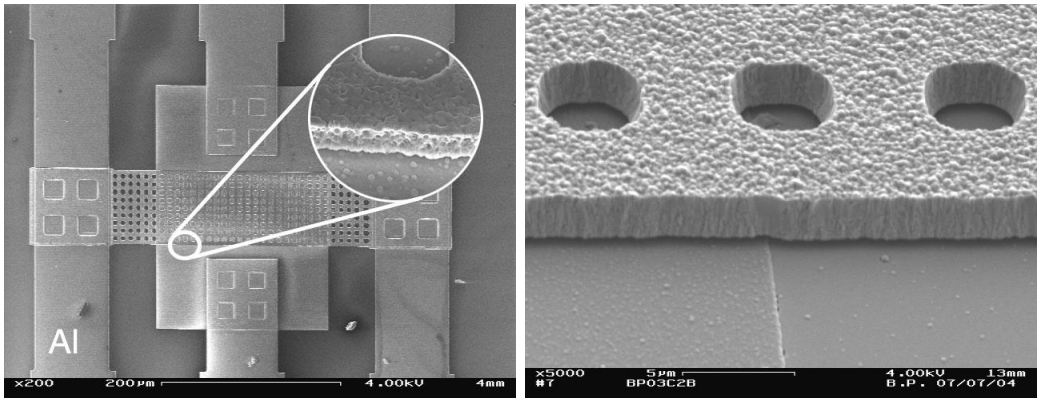


Figure 4.38: The 0.5 μm gap MEMS capacitors were properly released.

The device in Fig. 4.39 was nevertheless measured by an Agilent 4284A LCR meter at 1 MHz. This capacitor is constituted of 6 beams connected in parallel, each of 40 μm width and 500 μm long (same as in Fig. 4.33(d)). The results are plotted in Fig. 4.40 and Fig. 4.41 for two different devices processed on the same wafer. The plots show that the reproducibility is low, as the pull-in voltages varies from 15 V to 30 V, depending on the device. The spreading in the processing is thus high. The decrease in capacitance after the collapse of the beam (Fig. 4.41) is attributed to a short circuit between the electrodes that becomes in contact. At low voltages, a variation of the total capacitance is observed (Fig. 4.41(b)). In order to explained this phenomena, we represented the device by the circuit in Fig. 4.42, derived from its physical structure, including the parasitics. The access capacitances C_{access} and the fringing capacitance C_f may be calculated from the layout geometry. C_{ins} is the capacitance associated with the insulating layer (Si nitride) that determines the maximal total capacitance. The beam resistances R_{poly} may be evaluated from the resistivity values extracted section 4.3.3. The MOS capacitance C_{MOS} originates in the Si substrate-oxide-polysilicon stack and C_{MOS} may be modeled as in [71]. This lumped model allows indicating the relative influence of the parasitics over the wanted effect. Following that model, the MOS effect should not be present at the measurement frequency (1 MHz). However, the model can not be applied directly in our case, because of the process error. Indeed, the surface area of the a bottom polysilicon electrode remains unknown, even if a higher value than the mask design is expected.

4.5.6 MEMS variable capacitors for VCOs

The realized tunable capacitor may be included in a VCO. When properly fabricated, the expected advantages of a MEMS capacitor over a MOS one are mainly the high

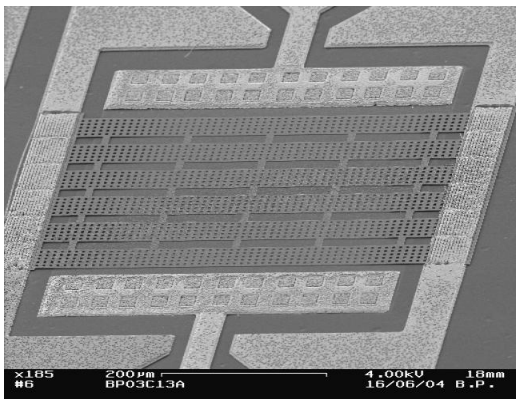


Figure 4.39: SEM picture of the measured tunable capacitor; 0.5 μm gap, 2 μm thick top electrode.

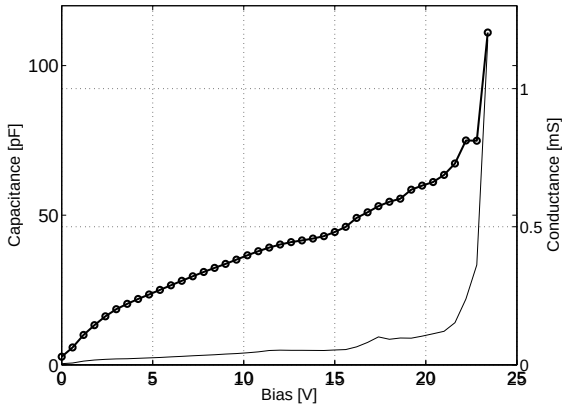
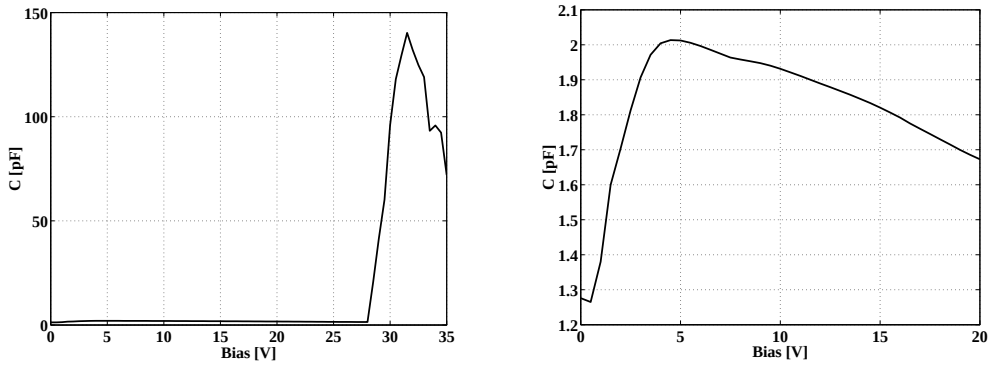


Figure 4.40: Measured capacitance (straigh line) and conductance (circles) of the capacitor shown in Fig. 4.39.



(a) (b)

Figure 4.41: Measured capacitance of the capacitor shown in Fig. 4.39. (b) is a detail of (a).

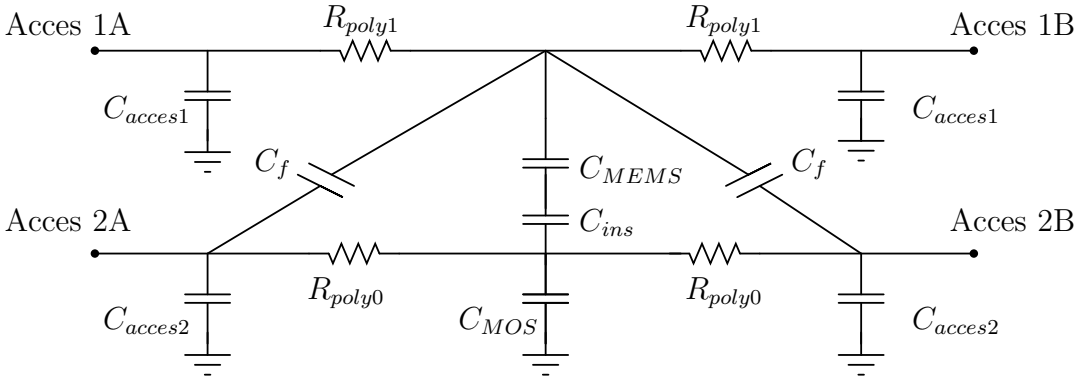


Figure 4.42: Equivalent circuit for a MEMS capacitor and its parasitics.

Q and the high linearity. However, the tuning range is limited at 1.5, and its reliability and temperature behavior may issue the working of the circuit. The cost of the complete circuit is another important point that will determine for a given application if a MEMS capacitor will be embedded in the circuit. Those items will be shortly discussed in this section.

Co-integration with the CMOS circuit There are several ways for connecting a MEMS to the silicon circuit. The choice of one or another option depends highly on the fabrication and assembly costs. The silicon cost is relatively low compared to the assembly cost. In a first option, two different chips are realized and interconnected. The MEMS may be fabricated on a different substrate type (glass, for instance) than the transistors. The bonding wires interconnection is a well-established, fast and high-yield process, but it suffers from a quite large parasitic inductance. A less-well established process, the flip-chip mounting, does not suffer from this electrical drawback, but it has a lower yield. The quality of the flip-chip mounting is moreover difficult to check. In a second option, the MEMS component is processed on the same substrate that the active circuit (single chip solution). The MEMS components may be pre- or post-processed on the silicon chip. The advantages are a smaller lost area, the absence of parasitic from the connection and the absence of additional mounting operation. However, for a given area, the silicon cost is roughly one order of magnitude higher than the MEMS cost and the quality factor of the micro-machine might be lower than in the case of a low-loss substrate (e.g., glass) were used for the MEMS. This second option requires more design time -and thus cost- since the processing of both the silicon and the MEMS must be compatible.

The capacitors designed in this work should be included in a VCO following the single chip option. Due to the high temperature steps involved during the MEMS processing, the mechanical part (without the last release step) has to be realized

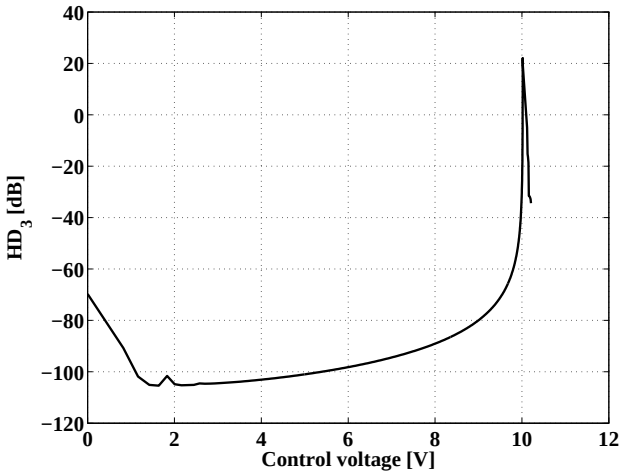


Figure 4.43: HD_3 of a 200 μm x 400 μm MEMS capacitor obtained by the Taylor method based on the tuning curve presented in Fig. 4.34; the capacitance is 27 pF for bias voltages beyond pull-in. $A = 0.2$ V.

before the standard CMOS process. The release step should end the fabrication.

Linearity The linearity of a MEMS capacitor decreases with increasing bias voltage and the capacitor becomes extremely nonlinear near its pull-in voltage (Fig. 4.43). The linearity analysis of MEMS tunable capacitors based on the Volterra series was performed by Innocent [72]. He showed how a MEMS capacitor depends on the applied bias and frequencies. In a two-tones test, the linearity depends on both the average and the difference of the applied frequencies. Two tones that are close together can generate large intermodulation products, even if they are at very high frequency. On the other hand, when the tones are widely separated, little distortion is produced. This is different from other varactors like the MOS structure studied in chapter 3. Furthermore, the VIP_3 is proportional to its pull-in voltage, which implies a trade-off. In the case of single-tone excitation, the linearity of the device increases with the frequency, which is an advantage for an oscillator working at microwaves.

Innocent showed further that when the capacitor is used in a tunable LC tank, the IP_{3i} of the tank is proportional to the IP_{3i} of the tunable capacitor. The IP_{3i} of the tank has a minimum at its resonance frequency and this minimum decreases as the quality factor of the LC tank increases. This is another trade-off.

Phase noise upconversion The mechanisms that generate phase noise in a VCO are not different whenever a MEMS or another tunable capacitor is used. There are however some differences in performances since the tuning curves have not the same shape. First of all, the brownian noise associated with the mechanical device

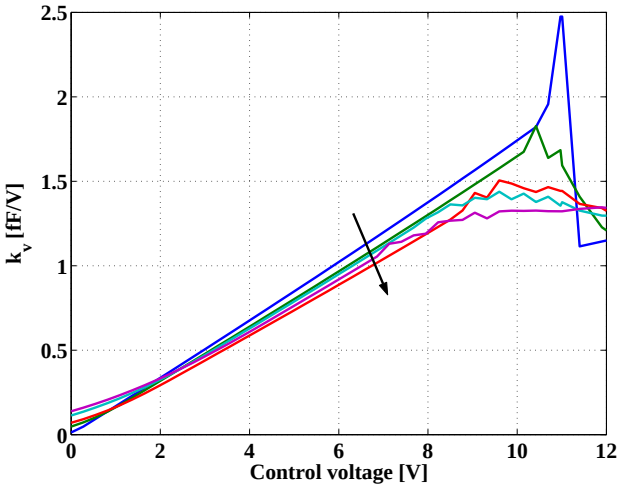


Figure 4.44: Varactor sensitivity k_v of a 200 μm x 400 μm MEMS capacitor obtained describing function method based on the tuning curve presented in Fig. 4.34; the capacitance is 27 pF for bias voltages beyond pull-in. The arrow indicates an amplitude A increase; the different values of A lies between 0.3 V and 4.3 V, by 1 V step.

contributes to phase noise [62]. The main contribution to phase noise in the oscillator sketched in Fig. 3.14(a) is the upconverstion of the flicker noise present in the current bias I_B source. This contribution to the phase noise is related to the varactor sensitivity k_v as explained before. The k_v of a typical MEMS capacitor is plotted in Fig. 4.44 as a function of the control voltage and the signal amplitude. The MEMS capacitor k_v is linearly related to the control voltage below pull-in and it reaches its maximal value at this point. On the other hand, k_v is less sensitive to the amplitude of the signal, compared to a MOS varactor (see Fig. 3.30), which is an advantage.

Compared to a MOS varactor, the tuning curve is more flat since a higher tuning voltage is required. The oscillation voltage is thus small compared to the tuning voltage. As a consequence, a lower of phase noise upconversion occurs. There is thus a trade-off between low phase noise and low pull-in voltage.

Close to pull-in, the phase noise grows as k_v increases. It is furthermore not recommended to work in this region to avoid an accidental pull-in of the structure. However, beyond pull-in, k_v decreases and the capacitor may be designed to work in the switch mode. A bank of digitally switchable MEMS capacitors may be realized in that way for the large tuning of VCOs.

On the other hand, variations of I_B at a frequency above the mechanical cut-off frequency (typically in the range $1-100$ kHz) do not modulate the output frequency. This means that the region of the bias current with the highest noise (flicker noise) is the least attenuated by the MEMS capacitor.

Temperature dependence Since the spring constant of the beam changes as a function of the initial strain and that the strain varies with the temperature (see section 4.3), the parameters of the MEMS capacitor change with the temperature. In the case of a MEMS device, we are not just interested in the expansion of the beam with the temperature, but we are interested in the difference in expansion between the beam and the substrate below since it causes the stress in the beam. In our case, the mismatch between the thermal expansion coefficients of the silicon of the silicon beam $\alpha_{T,Si} = 2.33 \cdot 10^{-6}$ K^{-1} and the silicon dioxide substrate$\alpha_{T,SiO_2} = 0.35 \cdot 10^{-6}$ K^{-1} is not too important and contributes to a tensile stress, which is even beneficial for the considered device.

A MEMS tunable capacitor cannot work under compressive stress. Indeed, when the initial length of the beam is larger than the distance between the anchors, the beam can arbitrary bend upwards or downwards. Changing from one state to the other requires a considerable force. The pull-in voltage is then increased and the device works in the switch-mode instead than as a tunable capacitor.

The temperature range in which the device should work depends on the application. For consumer products, the devices should operate between 0 and 70^{o}C, while military grade requires a range between -55^oC and 125^{o}C. There are three means to keep the stress tensile over the temperature range [65]. First, a structure can be added to the spring to relax the stress. This will make the spring more flexible, reduce the pull-in voltage and the mechanical resonance frequency, and increase the sensitivity to stress gradients. Second, the initial strain may be increased further such that it will not become compressive. This leads to very high values of the stress at low temperatures and large variation in the spring constant, and thus in the pull-in voltage. Third, another material with an expansion coefficient close to this of the substrate may be used.

Discussion Compared to on-chip MOSFET varactors, the quality factor Q of MEMS tunable capacitors is expected to be somewhat higher, but still in the same order of magnitude. This high Q improves the phase noise originating in thermally induced noise. Since the steepness of the tuning curve is small, the sensitivity to upconvert noise originating in the tail transistor is lower than for MOS varactors. The price to pay is however a lower tuning range and a higher tuning voltage. The tuning range of a MEMS based oscillator is poorly reproducible and temperature dependent. The reliability of the device is furthermore not well-established and the packaging more difficult. The expected gain from using a MEMS capacitor in a VCO is in conclusion small compared to the additional problems it introduces.

The higher cost of micromachined capacitors is another disadvantage. As a result, MEMS capacitors are not expected to be used in VCOs for main stream applications. We believe the smartest way to use MEMS capacitors in a VCO is for the large tuning. The tuning curve of the bank of capacitors presented in Fig. 4.35 is indeed much more reproducible as the MEMS operates in the switch mode. Tunable filters might on the other hand be a driver application, due to the high linearity of the micromachined capacitors.

4.6 Conclusions

In this chapter, micromachined tunable capacitors that could be embedded in voltage-controlled oscillators were investigated. The research was motivated by some previous realizations that demonstrated the high Q and the high integration level of these devices. The development of microsystems is however not straightforward. Indeed, new technological problems appear with the fabrication of small-dimension 3D devices. For instance, due to the increase of the surface-to-volume ratio with the scaling of the devices, the influence of surface forces become dominant. If not properly treated, this effect may cause the failure of the MEMS devices. In this chapter, we were interested in the development of a surface micromachining fabrication process at UCL.

Some interesting solutions were given to the present limitations of the micro- and nano-fabrication techniques. However, we were not able to investigate experimentally the circuit implications of this technology with our realizations (linearity, noise AM-to-PM for VCO applications, etc.). The reproducibility and the reliability of the devices are questionable. MEMS tunable capacitors are not expected to be used in VCOs for main stream applications, as the performance improvement is small compared to the issues the technology introduces.

As a consequence, we developed technological tools for improving the reliability of the MEMS. Thin-film technology was considered for its potentiality in the realization of high sensitivity sensors and for its integration with state of the art SOI electronics. The SOI substrate offers a high quality monocrystalline Si film that may be exploited for thin-film micromachining. This technology could be used for instance for the fabrication of the nano-scale mechanicals resonators used in RF filter stages.

Micromachining (as well as microelectronics) fabrication techniques require high precision patterning. While E-beam may be used in the lithography step, the etchings need for the development of specific plasma techniques. The Bosch process is

widely used for thick film etching. By controlling the passivation step, it may be modified in order to etch thin films and to control the profile anisotropy. On the other hand, the undercut may be limited by the addition of a fluoro-carbon to the SF_6 RIE, which is a desirable feature in most laterally driven MEMS.

Other fabrication steps are more specific to microsystems. Indeed not only the electrical but also the mechanical properties of the films have a huge influence on the device behavior. For a tunable capacitor application, a low-stress and low-resistive film is desired. The residual stress in the UCL polysilicon was characterized. Stress characterization methods at the micro-scale should however be improved. Experiments showed that the silicidation of the movable polysilicon plate is a low-cost way to get low-stress and low-resistive material for micromachining.

A special attention must be paid to the release step in order to avoid stiction. Not only process induced, but also in-use stiction is a typical failure of MEMS. Therefore, the coating of silanes prevents from stiction, improving the reliability of the devices. We demonstrated that the reliability of thin-film CMOS compatible MEMS may be improved by suitable silane coating in vapor or liquid phase. The stiction failure is however not totally eliminated, as it has other origins than capillarity (electrostatic forces from charges dielectric layers, for instance).

References

[1] C. T.-C. Nguyen, L.P.B. Katechi, and G.M. Rebeiz, "Micromachined devices for wireless communications," *Proceedings of the IEEE*, vol. 86, no. 8, pp. 1756–1768, August 1998.

[2] Marc Madou, *Fundamentals of Microfabrication*, CRC Press, 1997.

[3] G.T.A. Kovacs, N.I. Maluf, and K.E. Peterson, "Bulk micromachining of silicon," *Proceedings of the IEEE*, vol. 86, no. 8, pp. 1536–1551, August 1998.

[4] Jean Laconte, *Thesis*, Ph.D. thesis, Université catholique de Louvain, Louvain-la-Neuve, Belgium, 2004.

[5] E.J. Gacia and J.J. Sniegowski, "Surface micromachined microengine as the driver for micromechanical gears," *Digest Int. Conf. on Solid-State Sensors and Actuators*, pp. 365–368, June 1998.

[6] L.J. Ristic, R. Gutteridge, B. Dunn, D. Mietus, and P. Bennett, "Surface micromachined polysilicon accelerometers," *Techn. Digest IEEE Solid-State Sensors and Actuators Workshop*, pp. 118–121, 1992.

[7] Liang-Hsuan Lu, Kee Suk Ryu, and Chang Liu, "A magnetic microstirrer and array for microfluidic mixing," *IEEE/ASME Journal of Microelectromechanical Systems*, vol. 11, no. 5, pp. 462–469, 2002.

[8] J.M. Bustillo, R.T. Howe, and R.S. Muller, "Surface micromachining for micromechanichal systems," *Proceedings of the IEEE*, vol. 86, no. 8, pp. 1552–1575, August 1998.

[9] Cyrille Hibert, "Dry etching techniques and applications," FSRM course, Lausanne, June 2001.

[10] M. Elwenspoek and H. Jansen, *Silicon Micromachining*, Cambridge University Press, 1998.

[11] J.W. Coburn and H.F. Winters, "Plasma etchinga discussion of mechanisms," *Journal of vacuum science and technology*, vol. 16, no. 2, pp. 391–403, March 1979.

[12] K Murakami, Y. Wakabayashi, K. Minami, and M. Esashi, "Cryogenic dry etching for high aspect ratio microstructures," *IEEE Int. Conf. on MEMS*, pp. 65 – 70, Feb. 1993.

[13] F. Laermer and A. Schilp, "Method of anisotropically etching silicon," 1996, US patent No 5,501,893.

[14] D.M. Manos and D.L. Flamm, *Plasma Etching: An Introduction*, Academic publisher, Boston, 1989.

[15] R. Legtenberg, H. Jansen, M. Boer, and M. Elwenspoek, "Anisotropic reactive ion etching of silicon using $SF_6/O_2/CHF_3$ gas mixtures," *Journal of the Electrochemical Society*, , no. 142, 1995.

[16] E. Quévy, B. Parvais, J.P. Raskin, L. Buchaillot, D. Flandre, and D. Collard, "A modified Bosch-type process for precise surface micro-machining of polysilicon," *Journal of Micromechanics and Microengineering*, vol. 12, pp. 328–333, May 2002.

[17] P.J. French, "Polysilicon: a versatile material for microsystems," *Sensors and Actuators*, vol. A 99, pp. 3–12, 2002.

[18] W. Kim, Thomas, Bishnu Gogoi, Ken. G. Goldman, Rivette Nicholas J. Garling Stephen E. McNeil, Andrew C., and Daniel J. Koch, "Substrate and annealing influences on the residual stress of polysilicon," *Solid-State Sensor and Actuator Workshop*, pp. 237–240, June 8-11 1998, Hilton Head Island.

[19] P. Krulevitch, G.C. Jonhson, and R.T. Howe, "Stress and microstructure in phosphorus doped polycrystalline silicon," Material Research Society Symposium, 1992, vol. 276, pp. 79–84.

[20] M. Biebl, G.T. Mulhern, and R.T. Howe, "*In situ* phosphorus-doped polysilicon for integrated MEMS," *Transducer'95- Eurosensors IX*, , no. 42-A5, pp. 198–201, 1995.

[21] P. Lange, M. Kirsten, W. Riethmller, B. Wenk, G. Zwicker, J.R. Morante, F. Ericson, and J.A. Schweitz, "Thick polycristalline silicon for surface micromechanical applications: deposition, structuring and mechanical characterization," *Transducers 95*, pp. 202–205, June 1995, Stockholm.

[22] Min S. Choi and Eric W. Hearn, "Stress effect in boron-implanted polysilicon films," *Journal of the Electrochemical Society*, vol. 131, no. 10, pp. 2443–2446, October 1984.

[23] X.J. Ning, "Distribution of residual stress in boron doped p^+ silicon films," *Journal of the Electrochemical Society*, vol. 143, no. 10, pp. 3389–3393, October 1996.

[24] S.P. Murarka and T.F. Jr. Retajczyk, "Effect of phosphorus doping on stress in silicon and polycrystalline silicon," *Journal of Applied Physics*, vol. 54, no. 4, pp. 2069–2072, April 1983.

[25] L. Elbrecht, R. Catanescu, J. Zacheja, and J. Binder, "Highly phosphorus-doped polysilicon films with low tensile stress for surface micromachining using $POCl_3$ diffusion doping," *Sensors and Actuators*, vol. A 61, pp. 374–378, 1997.

[26] Yogesh B. Gianchandani and Khalil Shinn, Meenam n Najafi, "Impact of high-thermal budget aneals on polysilicon as a micromechanical material," *IEEE/ASME Journal of Microelectromechanical Systems*, vol. 7, no. 1, pp. 102–104, March 1998.

[27] H. Guckel, D.W. Burns, C.C.G. Visser, H.A.C. Tilmans, and D. Derdoo, "Fine-grained polysilicon films with built-in tensile strain," *IEEE Transactions on Electron Devices*, vol. 35, no. 6, pp. 800–801, June 1988.

[28] Xin Zhang, Tong-Yi Zhang, Man Wong, and Yitshak Zohar, "Rapid thermal annealing of polysilicon thin films," *IEEE/ASME Journal of Microelectromechanical Systems*, vol. 7, no. 4, pp. 356–364, December 1998.

[29] P.J. French et al., "The development of a low-stress polysilicon process compatible with standard device processing," *IEEE/ASME Journal of Microelectromechanical Systems*, vol. 5, no. 3, pp. 187–196, September 1996.

[30] C.A. Klein, "How accurate are Stoney's equation and recent modifications," *Journal of Applied Physics*, vol. 88, pp. 5487–5489, November 2000.

[31] M. Mehregany, R.T. Howe, and S.D. Senturia, "Novel microstructures for the in situ measurement of mechanical properties of thin films," *Journal of Applied Physics*, vol. 62, pp. 3579–3584, 1987.

[32] H. Guckel, T. Randazzo, and D.W. Burns, "A simple technique for the determination of mechanical strain in thin films with applications to polysilicon," *Journal of Applied Physics*, vol. 57, pp. 1671–1675, 1985.

[33] L. Lin, A.P. Pisano, and R.T. Howe, "A micro strain gauge with mechanical amplifier," *IEEE/ASME Journal of Microelectromechanical Systems*, vol. 6, no. 4, pp. 313–321, December 1997.

[34] Chi Shiang and Wensyang Hsu, "A microstructure for in situ determination of residual strain," *IEEE/ASME Journal of Microelectromechanical Systems*, vol. 8, no. 2, pp. 200–207, June 1999.

[35] Ming Qin, M.C. Poon, and C.Y. Yuen, "A study of nickel silicide film as a mechanical material," *Sensors and Actuators*, vol. 87, pp. 90–95, 2000.

[36] K. Maex and M. (editors) Van Rossum, *Properties of metal silicides*, IEE, INSPEC, 1995.

[37] P.T.J Gennissen and P.J. French, "Sacrificial oxide etching compatible with aluminum metalization," *Int. Conf. on Solid-state sensors and Actuators*, pp. 225–228, June 16-19 1997, Chicago.

[38] Roya Maboudian and Roger T. Howe, "Critical review: Adhesion in surface micromechanical structures," *Journal of vacuum science and technology*, vol. B15, no. 1, pp. 1–20, Jan/Feb 1997.

[39] C.H. Mastrangelo and C.H. Hsu, "Mechanical stability and adhesion of microstructures under capillary forces – part I : Basic theory," *IEEE/ASME Journal of Microelectromechanical Systems*, vol. 2, no. 1, pp. 33–43, March 1993.

[40] F. Michael Serry, Dirk Walliser, and G. Jordan Maclay, "The anharmonic Casimir oscillator (ACO)– The Casimir effect in a model microelectromechanical system," *IEEE/ASME Journal of Microelectromechanical Systems*, vol. 4, no. 4, pp. 193–205, December 1995.

[41] M.-J. Lin and R. Chen, "Effect of gradient residual stress on adhesion for a center-anchored circular plate and its underlying substrate," *Sensors and Actuators*, vol. A 115, no. 1, pp. 109–117, September 2004.

[42] Y. Matsumoto, T. Shimada, and M. Ishida, "Novel prevention method of stiction using silicon anodization for SOI structure," *Sensors and Actuators*, vol. A72, pp. 153–159, 1999.

[43] Yoshinori Matsumoto and Makoto Ishida, "The property of plasma-polymerized fluorocarbon film in relation to CH_4/C_4F_8 ratio and substrate temperature," *Sensors and Actuators*, vol. 83, pp. 179–185, 2000.

[44] G.T. Mulhern, D. Soane, and R.T. Howe, "Supercritical carbon dioxide drying for microstructures," *Transducers'93*, pp. 296–299, 1993.

[45] T.K. Whidden, J. Allgair, J.M. Ryan, and D.K. Kozicki, M.N. Ferry, "Catalyzed HF vapor etching of silicon dioxide for micro- and nanolithographic masks," *Journal of the Electrochemical Society*, vol. 142, no. 4, pp. 1199–1205, April 1995.

[46] J. Anguita and F. Briones, "HF/H_2O vapor etching of SiO_2 sacrificial layer for large-area surface-micromachined membranes," *Sensors and Actuators*, vol. A 64, pp. 247–251, 1998.

[47] Y.-I. Lee, K.-H. Park, J. Lee, C.-S. Lee, H.J. Yoo, C.-J. Kim, and Y.-S. Yoon, "Dry release for surface micromachining with HF vapor-phase etch-

ing," *IEEE/ASME Journal of Microelectromechanical Systems*, vol. 6, no. 3, pp. 226–233, September 1997.

[48] Y. Ma, M. L. Green, L.C. Feldman, L. Sapjeta, K. J. Hanson, and T. W. Weidman, "Vapor phase SiO_2 etching and metallic contamination removal in an integrated cluster system," *Journal of vacuum science and technology*, vol. B 13, no. 4, pp. 1460–1465, July/August 1995.

[49] U. Srinivasan, R. Houston, Howe R.T., and R. Maboudian, "Alkyltrichlorosilane-based self-assembled monolayers films for stiction reduction in silicon micromachines," *IEEE/ASME Journal of Microelectromechanical Systems*, vol. 7, no. 2, pp. 252–260, June 1998.

[50] W.R. Ashurst, C. Carraro, and R. Maboudian, "Self-assembled monolayers as anti-stiction coating for MEMS: Characteristics and recent progress," *Sensors and Actuators*, vol. A 82, pp. 219–223, 2000.

[51] S. Pamidighantam, W. Laureyn, A. Salah, A. Verbist, and H.A.C. Tilmans, "A novel process for fabricating slender and compliant suspended poly-si micromechanical structures with sub-micron gap spacing," *MEMS 2002*, pp. 661–664, January, 20-24 2002, Las Vegas, U.S.A.

[52] K. Komvopoulos, "Surface engineering and microtribology for microelectromechanial systems," *Wear*, vol. 200, pp. 305–327, 1996.

[53] B.H. Kim, T.C. Chung, C.H. Oh, and K. Chun, "A new organic modifier for anti-stiction," *IEEE/ASME Journal of Microelectromechanical Systems*, vol. 10, no. 1, pp. 33–40, March 2001.

[54] W.R. Ashurst, C. Yau, C. Carraro, R.T. Howe, and R. Maboudian, "Alkene based monolayer films as anti-stiction coatings for polysilicon MEMS," *Proceedings of Solid-State Sensor and Actuator Workshop, Hilton Head 2000,*, pp. 320–323, June 2000.

[55] C.F. Herrmann, F.W. DelRio, V.M. Bright, and S.M. George, "Hydrophobic coatings using atomic layer deposition and non-chlorinated precursors," *IEEE Int. Conf. on MEMS*, pp. 653–656, 2004.

[56] W.R. Ashurst, C. Carraro, and R. Maboudian, "Vapor phase anti-stiction coating for MEMS," *IEEE Transactions on Device and Materials Reliability*, vol. 3, no. 4, pp. 173–178, 2003.

[57] B. Parvais, A. Pallandre, A.M. Jonas, and J.-P. Raskin, "A fluoro-ethoxysilane-based stiction-free release process for submicron gap MEMS," *Proc. of the Nanotech 2003 Conf. Vol. 1, San Francisco CA*, pp. 522–525, 2003.

[58] A. Witvrouw, B. Du Bois, P. De Moor, A. Verbist, C. Van Hoof, H. Bender, and K. Baert, "A comparison between wet HF etching and vapor HF etching for sacrifical oxide removal," *SPIE*, 2000.

[59] Antoine Pallandre, K. Glinel, Jonas A.M., and B. Nysten, "Binary nanopatterned surfaces prepared from silane monolayers," *Nano Letters*, pp. 365–371, 2004.

[60] J. Als-Nielsen, F. Christensen, and P. S. Pershan, "Smectic-a order at the surface of a nematic liquid crystal: Synchrotron X-ray diffraction," *Phys. Rev. Lett.*, vol. 48, no. 16, pp. 11071110, 1982.

[61] L. G. Parratt, "Surface studies of solids by total reflection of X-rays," *Phys. Rev.*, vol. 95, pp. 359–369, 1954.

[62] Gabriel M. Rebeiz, *RF MEMS Theory, Design, and Technology*, John Wiley and sons, Hoboken, New Jersey, 2003.

[63] Aleksander M. Dec, *Design and Analysis of Microwave Oscillators*, Ph.d. thesis in the graduate school of arts and science, Columbia University, 1998.

[64] Stephen D. Senturia, *Microsystem Design*, Kluwer academic publishers, Boston/Dordrecht/London, 1998.

[65] Manuel Innocent, *MEMS variable capacitors for multi-GHz radio applications*, Ph.D. thesis, Katholieke Universiteit Leuven - IMEC, Leuven, Belgium, October 2003.

[66] Nicolas Scott Barker, *Distributed MEMS transmission Lines*, Ph.D. thesis, University of Michigan, Ann Arbor, Michigan, USA, 1999.

[67] R.J. Roark and W.C. Young, *Formulas for Stress and Strain*, Mc Graw Hill, New York, 6^{th} edition, 1989.

[68] J. J.-Y. Gill, L.V. Ngo, P.R. Nelson, and C.J. Kim, "Elimination of extra spring effect at the step-up anchor of surface-micromachined structure," *Technical Proceedings of the Fourth International Conference on Modeling and Simulation of Microsystems*, vol. 7, no. 1, pp. 114–121, March 1998.

[69] Simulation research for MEMS, "Sugar," *http://www-bsac.eecs.berkeley.edu/cadtools/sugar/sugar/*.

[70] Z. Bai, D. Bindel, J.V. Clark, J. Demmel, K.S.J. Pister, and N. Zhou, "New numerical techniques and tools in sugar for 3d mems simulation," *Technical Proceedings of the Fourth International Conference on Modeling and Simulation of Microsystems*, pp. 31–34, March 19-21 2001, Hilton Head Island.

[71] J.-P. Raskin, *Modeling, Characterization and Optimization of MOSFET's and Passive Elements for the Synthesis of SOI MMIC's*, Ph.D. thesis, Université catholique de Louvain, Louvain-la-Neuve, Belgium, December 1998.

[72] M. Innocent, P. Wambacq, S. Donnay, H.A.C. Tilmans, W. Sansen, and H. De Man, "An analytic Volterra-series-based model for a MEMS variable capacitor," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 22, no. 2, pp. 124–131, February 2003.

CHAPTER 5
CONCLUSIONS

The increasing demand for wireless RF applications asks today for high performances systems-on-chip. 'High performances' means low power consumption, high quality of the data transmission, and low volume and weight. If the CMOS scaling contributed to performances improvements, we are now confronted to short-channel parasitics and to the relatively poor quality of the passive integrated components.

There are several ways to face these limitations.
First, a better knowledge of the components permits to use them efficiently. This requires the development of adapted modeling and characterization tools. Second-order effects must be taken into account and the tools should be efficient enough to follow the rapid evolution of the technologies.
Second, one can move from standard bulk CMOS to other technologies. In that case, the development cost is high, but the expected gain is also high.
Finally, if high quality components help to improve the quality of the transceivers, the performances also highly rely on the systems and circuits architecture. The design optimization is another key in RF analog IC design.

The goal of the present work is to investigate these three ways for improving the transceivers performances. On one hand, characterization tools for the linearity analysis are developed. On the other hand, SOI CMOS technology is used in the frame of this work as an alternative to bulk CMOS technology. Micromachining technologies were used to fabricate tunable passive components. Finally, circuit design techniques for oscillators are investigated.

The quality of the wireless transmission depends on the system linearity. Due to the devices physics, most circuits are nonlinear. Some of them, oscillators for instance, rely on the nonlinearity, but this feature is unwanted in other cases e.g., in amplifiers. A classical procedure for dealing in a practical way with nonlinear systems is to construct a linear model to approximate the nonlinearities. A small-signal linearization consists for instance of expanding the nonlinear function in a Taylor series around some operating point and retaining only the linear term in the analysis. If the excitations are not small enough to permit this simple form of linearization, one can do better by letting the linear approximation depends on the input. We speak in that case of quasi-linearization. Unfortunately, linearized

models do not explain the generation of tones at different frequencies than the input frequency. The harmonic and intermodulation distortions quantify this effect in the case of single or two-tones excitation, respectively.

In this work, several tools for characterizing the distortion are provided. They are confronted to state-of-the-art LSNA measurements. The choice of one of these tools depends on the application.

For weak linearities i.e., typically for a MOSFET excitation below 0.3 V, the nonlinearity may be represented by a third-order polynomial. While the Taylor approach is used for frequency-independent systems, the Volterra analysis applies to frequency-dependent nonlinearities. A Volterra series based model is useful to understand the different mechanisms involved in the nonlinear processes at different operating frequencies. The simple Volterra based model developed in this work for the MOSFET demonstrates that the distortion of the transistor is dominated by its current-voltages $I-V$ relationship. The MOSFETs have a frequency limit under which the distortion characteristics behave like those at low frequency. This is a useful property of CMOS technology since the distortion analysis may be simplified for low GHz applications.

The integral function method is more appropriate for the analysis of strong nonlinearities, since this method is directly related to the excitation amplitude. For that reason, the sweet-spots are easily obtained by IFM. This method is well-adapted to treat low-cost $I-V$ measurements, as no derivatives are required. Initially dedicated to memoryless systems, we provide the extension of this method to frequency dependent systems. The present version of IFM however does not apply directly to multi-tones characterization.

The technology affects the linearity of devices. For instance, micromachined tunable capacitors are more linear than their MOSFET counterpart. In SOI, FD and PD transistors present similar HD level in saturation. However, the IMD of PD transistors depends on the tone separation, as a consequence of the kink effect. The analysis provided in this work is limited to FD and PD devices, but it could be enlarged to other SOI devices, such as DTMOS.

The nonlinear analysis is not restricted to devices. The example of a narrowband LNA is examined in chapter 2. Because of the high efficiency of the FETs in the moderate inversion mode of operation, and since the linearity is not that much degraded, this mode of operation is preferred for the design of low-GHz LNA in sub-micron CMOS.

Contrary to LNAs, oscillators rely on nonlinearities. Indeed, it is the oscillator nonlinearity that determines the output amplitude. However, since these circuits are built to generate sinusoidal waves, the designer is much more interested in the phase noise performance than in the distortion. A good evaluation of the phase noise requires a good evaluation of the output amplitude. The designer has then to choose between the insight provided by linear tools and the accuracy of nonlinear models. We propose to use a quasi-linear approximation of the oscillator nonlinearity in order to solve this trade-off. The describing functions formalism is used for that purpose, as it gives the best amplitude-dependent linear approximation of a nonlinearity to the mean squares sense. It is a powerful tool for the analysis and the design of oscillators.

Based on that formalism and on the oscillator analysis, a top-down design methodology is proposed to optimize the performances of an active feedback differential VCO. Even if the method remains to be validated experimentally, it improves the accuracy of the first-order designs and allows to deals with the trade-off opposing power consumption and spectral purity.

Technology is another way for improving the VCO performances, as it determines the resonator quality factor. In SOI, fully-depleted accumulation-mode varactor with intrinsic doping gives the best performances, as it has a symmetrical tuning curve and a high quality factor. The low-power, low-voltage capabilities of SOI are demonstrated experimentally in chapter 3. A more detailed experimental analysis would require measurements close to the carrier, which is not possible with the considered setup. Compared to their MOS counterpart, micromachined tunable capacitors have a slightly higher quality factor, a higher linearity and lower phase noise conversion, but, due to their low reliability and reproducibility, as well as their high cost and high temperature dependence, MEMS capacitors are not expected to be used in VCOs for main stream applications.

Surface micromachining techniques are nevertheless widely studied in chapter 4, as they are promised to improve the performances of the RF ICs in the next years. Whenever not used in VCOs, other applications (tunable filters, intelligent systems, ...) should indeed benefit from the advantages this technology offers, in particular its potentiality to build a complete system on a silicon chip. The evolution towards integrated microsystems renders the design of transceivers more multi-disciplinary than ever, requiring skills at several levels, from the development of fabrication tools to the design of transmission system.

The mechanical functions in MEMS require to use adapted materials. Thin-film technology is considered for its potentiality in the realization of high sensitivity sensors and for its integration with state of the art SOI electronics. Silicided polysilicon is used in this work as a low-cost material adapted to MEMS. Indeed, the silicidation process allows decreasing the resistivity, without affecting much the residual stress.

The main concerns related with microtechnologies, apart from the development of efficient CAD tools, are today linked to the fabrication and the reliability of the devices. Some material properties may differ at the micro-scale since the volume to surface ratio decreases. The stiction is a widespread failure mode that is mainly dependent on the capillarity forces. An anti-stiction process should be CMOS compatible and adapted to thin-film technologies. The silanes coating in liquid or vapor phase, as studied in chapter 4, prevents from both process-induced and in-use stiction. The stiction failure is however not totally eliminated, as it has other origins than capillarity (electrostatic forces from charges in dielectric layers, for instance).

APPENDIX A
QUALITY FACTOR OF A RESONATOR

A.1 Definitions of the quality factor Q

The most widespread definition of a quality factor comes from the RLC resonator analysis. In this case, the quality factor Q is defined as

$$Q = \frac{\omega_o}{\Delta\omega} \tag{A.1}$$

where ω_o is the resonance angular frequency and $\Delta\omega$ is the -3 dB bandwidth. The Q represents the selectivity of the circuit.

A more general definition is

$$Q = 2\pi \frac{\text{Stored energy per cycle}}{\text{Dissipated energy per cycle}}. \tag{A.2}$$

This definition is applied to several circuits in Fig. A.1. For instance, let us take the RLC circuit (Fig. A.1, top-right). When the circuit is excited by a sine wave $V(t) = V_A \sin(\omega t)$, the maximal voltage V_o across the capacitance is reached at time t when the current through the inductance is null. At this instant, the numerator of (A.2) is equal to

$$E_{peak} = \frac{C \cdot V_A^2}{2}. \tag{A.3}$$

The dissipated energy in the resistance is:

$$E_{loss} = \int_0^{2\pi/\omega} \frac{(V_A \cdot \sin(\omega t))^2}{R_p} \cdot dt = \frac{\pi \cdot V_A^2}{\omega \cdot R_p}. \tag{A.4}$$

Then the Q is by definition

$$Q = R_p \cdot \omega C = R_p \cdot \sqrt{\frac{C}{L}}. \tag{A.5}$$

The right-hand side equality is obtained at resonance i.e., for $\omega^2 = 1/LC$. Note that this results is in agreement with the first definition.

In the microwave field, it is familiar to define the quality factor of a two-ports

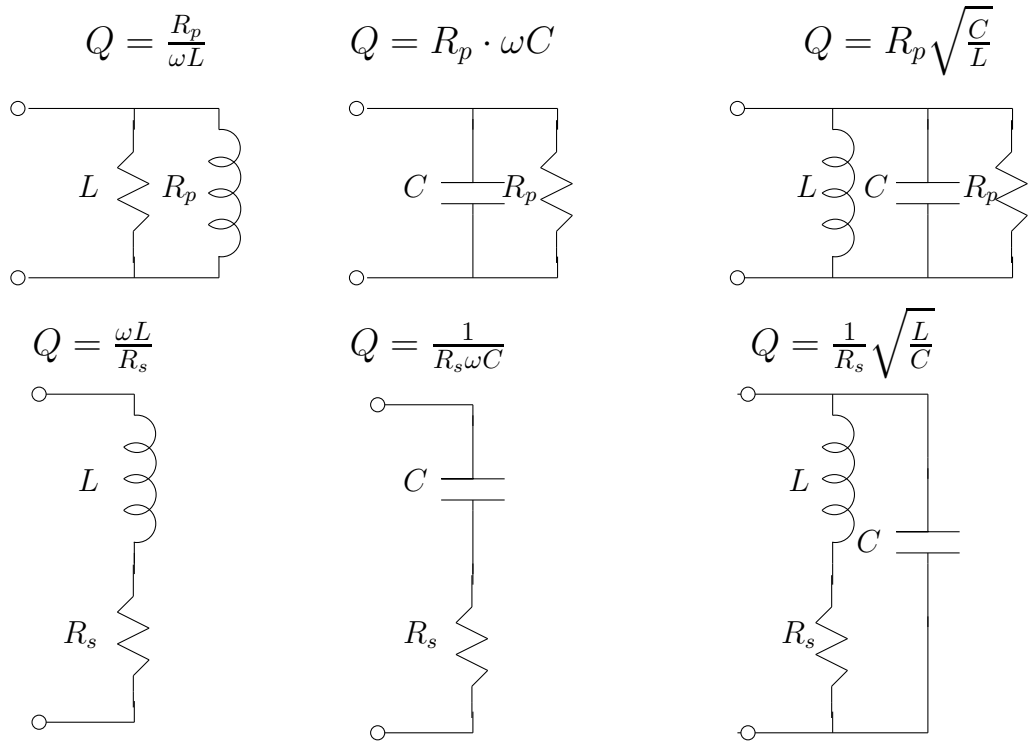


Figure A.1: Quality factors of some circuits.

from its admittance matrix $\begin{pmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{pmatrix}$:

$$Q = \frac{\Im Y_{11}}{\Re Y_{11}}. \tag{A.6}$$

From the oscillator analysis, a fourth definition is introduced hereafter [34]. Following the two-ports model of an oscillator, the transfer function writes $H(j\omega) = A(\omega).e^{j\phi(\omega)}$ in its polar form. The Q is then defined as

$$Q = \frac{\omega_o}{2}\sqrt{\left(\frac{dA(\omega)}{d\omega}\right)^2 + \left(\frac{d\phi(\omega)}{d\omega}\right)^2}. \tag{A.7}$$

Let us suppose that at the oscillation frequency, $\frac{dA(\omega)}{d\omega}$ is equal to zero. In this case,

$$Q = \frac{\omega_{nom}}{2}\left|\frac{d\phi(\omega)}{d\omega}\right| \tag{A.8}$$

The Barkhausen's conditions state that the phase difference of a closed-loop system should be an integer multiple of 2π to allow the system to oscillate. If the signal frequency shifts slightly from $f_o = \omega_o/2\pi$, the feedback tries to recover f_o. This return is more effective if the slope $\frac{d\phi(\omega)}{d\omega}$ is high. Thus, with the last definition, the Q indicates how an oscillator opposes itself to a frequency change. The higher Q,

the more stable the oscillation.

A.2 Series-parallel connections

Consider an inductor connected to several capacitors C_i. The total quality factor is

$$\frac{1}{Q} = \frac{1}{Q_L} + \sum_i \frac{1}{Q_{Ci}} \left(\frac{C_i}{C} \right) \tag{A.9}$$

with $C = \sum_i C_i$, Q_L the inductor quality factor and Q_{Ci} the quality factor of capacitor C_i.

Series connection of capacitors:

$$Q = \frac{\prod_i Q_i}{\sum_i Q_i \left(\frac{C_i}{C} \right)}. \tag{A.10}$$

A.3 Series to parallel transformation

Close to resonance, the circuits of Fig. A.2 (a) and (b) are related by [63]:

$$L_b^{'} = L_b \left(1 + \frac{1}{Q_b^2} \right) \tag{A.11}$$

$$C_a^{'} = \frac{C_a}{1 + \frac{1}{Q_a^2}} \tag{A.12}$$

$$R_c^{'} = R_c \| (R_a(1 + Q_a^2)) \| (R_b(1 + Q_b^2)) \tag{A.13}$$

with $Q_a = \frac{1}{\omega L_a R_a}$, $Q_b = \frac{\omega L_b}{R_b}$ and where the symbol $\|$ denotes the parallel connection. The circuit in Fig. A.2 (a) resonates when the imaginary part of Y_{11} is equal to zero i.e., at the angular frequency ω_o given by

$$\omega_0^2 = \frac{1}{L_b C_a} \left(\frac{1 - \left(\frac{R_b}{C_a} \right) R_b C_a}{1 - \left(\frac{R_a}{L_b} \right) R_a C_a} \right). \tag{A.14}$$

The total resonator quality factor is then

$$Q_T = \frac{R_c^{'}}{Z_0} \tag{A.15}$$

$$\frac{1}{Q_T} = \frac{Z_0}{R_c} + \frac{1}{Q_a} + \frac{1}{Q_b} \tag{A.16}$$

Figure A.2: Series to parallel transformation of a RLC circuit.

where $Z_0 = \frac{1}{\omega C_a} = \omega L_b = \sqrt{\frac{L_b}{C_a}}$.

APPENDIX B
PHASE NOISE MEASUREMENTS WITH A SPECTRUM ANALYZER

The measurement setup used in this work to measure the phase noise is a HP8562A spectrum analyzer with an option allowing measurements up to 22 GHz. A DC block is placed at the input of the analyzer. The cable length should be as low as possible, even if a calibration is not necessary as phase noise is a ratio of power waves. The biasing of the circuits, especially for on-wafer measurements, have to be highly decoupled. If not, noise from the supplies contributes to the oscillator phase noise as explained before.

Measurement Procedure The two quantities present in the definition (B.1) of phase noise are measured separately.

$$L(\Delta\omega) = 10\log\left(\frac{P_{sideband}(\omega_o + \Delta\omega, 1\text{Hz})}{P_{carrier}}\right) \tag{B.1}$$

where $P_{sideband}(\omega_o + \Delta\omega, 1\text{Hz})$ represents the power of a 1 Hz bandwidth contained in a lateral band of the output spectrum. This power is measured at a $\Delta\omega$ shift from the center frequency w_o.

$P_{carrier}$ is measured with a large Resolution Bandwidth (RBW) in order to have the whole peak inside the filter bandwidth. Typical RBW value is 2 MHz. The Video Bandwidth (VBW) should also be as high as possible (the *auto mode* may be used in most cases). In order to protect the setup, the internal attenuator should be put on its maximum value for the first measurements. Then, in order to increase the Signal-to-Noise Ratio (SNR), it should be turned to the minimum allowable value.

$P_{sideband}$ is evaluated using a low RBW. Indeed, when a high RBW is used, the filter shape is measured and not the oscillator spectrum itself. However, the smaller the RBW, the longer the sweep time. This may become an issue in the case of free-running oscillators. Grinbergs [1] obtained best results by setting the RBW to 1 kHz. Friedrich [2] advices to keep the ratio between RBW and VBW equal to 100.

In order to get $P_{sideband}$, the power is measured inside the RBW and then normalized to a 1 Hz bandwidth. A correction factor ϑ linked to the accuracy of the

Table B.1: Correction factor in function of the IF filter accuracy [3].

RBW range	Filter Accuracy	ϑ
1 Hz... 300 kHz	$\pm 10\%$	1.1
300 kHz... 1 MHz	$\pm 25\%$	1.25
2 MHz... 2 MHz	$+50\% - 25\%\%$	1.4

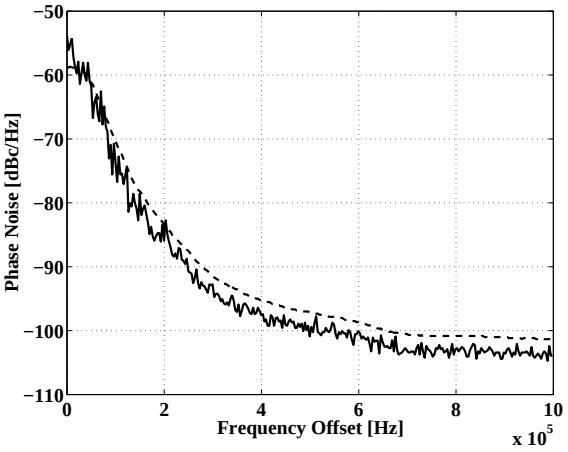


Figure B.1: Comparison of phase noise extracted by (B.2) (Straight lines) and by the Marker Noise option of the spectrum analyzer (dashed lines).

IF filter should be added [1] (table B.1):

$$L(\Delta\omega) = 10\log\left(\frac{P_{sideband}(\omega_o + \Delta\omega, 1Hz)}{P_{carrier}}\frac{1}{\vartheta RBW}\right). \tag{B.2}$$

On the other hand, it is suggested in [1] to add another correction factor of 2.5 dB if the logarithmic display mode is used.

The HP8562A permits to evaluate the noise power directly. Indeed, the *Marker Noise* option of the analyzer is used to get $P_{sideband}$. The correction factors introduced previously are automatically taken into account. This function activates a 32 points averaging. Fig. B.1 compares (B.2) to the data from the marker noise option.

The data acquisition makes use of a Lab-View interface developed in our laboratory.

Main Limitations

- Both the amplitude and the phase noise are measured. The amplitude noise should be at least 10 dB smaller than the phase noise for enough accuracy.
- Measurements close to the carrier ($\Delta f < 1$ kHz) are not possible. A good accuracy is obtained for $\Delta f > 10$ kHz.

Table B.2: Phase noise of the HP8562 Spectrum Analyzer for frequencies in the range 0 . . . 1 GHz [3].

Δf	PN [dBc/Hz]
100 Hz	< -88
1 kHz	< -97
10 kHz	< -113
100 kHz	< -117

- The phase noise of the circuit under test should be higher than the one of the spectrum analyzer (table B.2).

References

[1] Oliver Grinbergs, "Determination of microwave oscillator phase noise by direct measurement with a spectrum analyzer," 1998, available from *http* : $//ee.eng.usf.edu/WAMI-1/StudentProjects/Phase_Noise_SA_Method.pdf$.

[2] Thomas Friedrich, "Direct phase noise measurements using a modern spectrum analyzer," *Microwave Journal*, pp. 94–104, August 1992.

[3] Agilent Corp., "Catalogue," 2000.

INDEX