JOÃO PEDRO ABREU DE OLIVEIRA

PARAMETRIC ANALOG SIGNAL AMPLIFICATION APPLIED TO NANOSCALE CMOS WIRELESS DIGITAL TRANSCEIVERS

LISBOA
2010

JOÃO PEDRO ABREU DE OLIVEIRA

PARAMETRIC ANALOG SIGNAL AMPLIFICATION APPLIED TO NANOSCALE CMOS WIRELESS DIGITAL TRANSCEIVERS

Thesis presented in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the subject of Electrical and Computer Engineering by the Universidade Nova de Lisboa, Faculdade de Ciências e Tecnologia.

LISBOA
2010

At quite uncertain times and places,
The atoms left their heavenly path,
And by fortuitous embraces,
Engendered all that being hath.
And though they seem to cling together,
And form 'associations' here,
Yet, soon or late, they burst their tether,
And through the depths of space career.

James Clerk Maxwell (1874)

Acknowledgments

This page would be always small to cope for the acknowledgments that I would like to publicly express to many people that have followed me along this journey.

First and foremost, I strongly acknowledge my supervisors Prof. A. Steiger Garção and Prof. João C. Goes.

I am sincerely grateful to Prof. A. Steiger Garção whose expertise, guidance, support, and patient have made this work possible. His personal and original view on electronics have broken me barriers and opened doors into this amazing scientific and technological world.

I would like to express my personal and deep gratitude to Prof. João C. Goes for his support, availability, encouragement and motivation, deep technical competence, commitment, orientation. He his the mentor for many of the breakthrough that made possible the work to pursuit towards the final objective.

I also would like to deeply acknowledge Prof. Yannis Tsividis for his kind encouragement and useful remarks. The suggestions kindly sent by Prof. Yannis Tsividis, about the MOS Parametric Cell, have been absolutely crucial to understand the operation details of this circuit.

To my colleagues and friends, Nuno Paulino for the deep technical discussions, Rui Tavares for his CAD support and to Luís Oliveira for the fruitful discussions on radio transceivers.

To Prof. Jorge Fernandes from INESC-ID who was always a unconditional supporter for this work with very helpful contributions.

Definitively to João Ferreira, Michael Figueiredo, Edinei Santin, José Custódio, Bruno Esperança for all the support and work performed. To Erik Snelling, João Faria, from Silicon and Software Systems(S3) for the lab support.

To José Faustino from INESC-ID for the excellent chip on board work.

To the following Institutions:

- Department of Electrical Engineering of the FCT-UNL that provided the necessary conditions to realize the work,
- CTS-UNINOVA that provided the necessary conditions to realize the work,
- Fundação para a Ciência e Tecnologia that supported part of this work through the projects SPEED, LEADER and IMPACT.

I thank my parents and sister family, for their existence and unconditional friendship.

To my wonderful wife Lúcia and my son João, for their unconditional and persistent help, encouragement, motivation and support.

Sumário

A atenuação que inevitavelmente ocorre na propagação de um sinal rádio entre a antena emissora e o receptor motivou uma procura persistente no domínio da amplificação electrónica de sinal. Todavia, o indissociável ruído existente quer no canal quer no dispositivo de amplificação em si torna a tarefa de recuperação do sinal original bem mais complexa.

Uma das formas de amplificação electrónica, surgida em meados do século passado, assenta na utilização de uma capacidade variável não linear. Através desta última é possível misturar o sinal de entrada com um outro fornecido por uma fonte auxiliar e forçar a transferência de energia entre as diversas componentes espectrais entretanto originadas. Desta forma e sob determinadas condições, é possível construir-se um amplificador paramétrico cujo o processo de amplificação é intrinsecamente ausente de ruído uma vez que não é baseado numa transconductância mas sim na variação paramétrica de uma reactância.

Apesar da técnica acima referida ser conhecida há várias décadas, apenas recentemente foram apresentados alguns exemplos da sua utilização na tecnologia CMOS. Nestes exemplos incluem-se aplicações quer no domínio do tempo contínuo quer em tempo discreto. A amplificação paramétrica em tempo discreto assenta na alteração periódica da capacidade do dispositivo MOS, obtida através da alteração forçada do seu regime de funcionamento enquanto mantém aprisionadas as cargas entretanto armazenadas na porta do dispositivo. Esta alteração da capacidade, enquanto se mantém a carga constante, terá de ser compensada por uma variação correspondente da tensão aos terminais do dispositivo. A amplificação do sinal assim obtida não depende directamente da transcondutância associada ao transístor MOS o que a torna uma opção a ter em conta no projecto de circuitos analógicos em tecnologias submicrométricas. Pretende-se, com o recurso a esta técnica, potenciar a utilização de células de amplificação com ganho reduzido e em malha aberta como forma de contornar as crescentes dificuldades de integração de amplificadores operacionais devido ao escalonamento verificado na tecnologia CMOS.

A presente dissertação debruça-se sobre a utilização da amplificação paramétrica em tempo discreto na tecnologia CMOS digital e, em particular, sobre a sua aplicação em transreceptores rádio. É neste contexto que são analisados diversos circuitos que podem integrar amplificação paramétrica, nomeadamente, um comparador, um misturador passivo e, inclusivamente, um conversor analógico-digital. Dada a funcionalidade absolutamente fundamental de um conversor analógico-digital num receptor rádio digital, demonstra-se experimentalmente a viabilidade da utilização da técnica através da integração, numa tecnologia CMOS de 130 nm, de um conversor analógico-digital de velocidade moderada com 8-bits de resolução e 120 MS/s de frequência de amostragem efectiva. Prova-se dessa forma e após caracterização experimental, que é possível construir um conversor analógico-digital que apenas utiliza dispositivos MOS e sem recurso a amplificadores operacionais de alto desempenho e consumo. Destes factos resulta a consolidação de uma técnica alternativa de processamento analógico de sinal que vem contribuir para o desenvolvimento de circuitos no seio das mais recentes tecnologias CMOS nanométricas.

Abstract

Signal amplification is required in almost every analog electronic system. However noise is also present, thus imposing limits to the overall circuit performance, e.g., on the sensitivity of the radio transceiver. This drawback has triggered a major research on the field, which has been producing several solutions to achieve amplification with minimum added noise. During the Fifties, an interesting out of mainstream path was followed which was based on variable reactance instead of resistance based amplifiers. The principle of these parametric circuits permits to achieve low noise amplifiers since the controlled variations of pure reactance elements is intrinsically noiseless. The amplification is based on a mixing effect which enables energy transfer from an AC pump source to other related signal frequencies.

While the first implementations of these type of amplifiers were already available at that time, the discrete-time version only became visible more recently. This discrete-time version is a promising technique since it is well adapted to the mainstream nanoscale CMOS technology. The technique itself is based on the principle of changing the surface potential of the MOS device while maintaining the transistor gate in a floating state. In order words, the voltage amplification is achieved by changing the capacitance value while maintaining the total charge unchanged during an amplification phase.

Since a parametric amplifier is not intrinsically dependent on the transconductance of the MOS transistor, it does not directly suffer from the intrinsic transconductance MOS gain issues verified in nanoscale MOS technologies. As a consequence, open-loop and opamp free structures can further emerge with this additional contribution.

This thesis is dedicated to the analysis of parametric amplification with special emphasis on the MOS discrete-time implementation. The use of the latter is supported on the presentation of several circuits where the MOS Parametric Amplifier cell is well suited: small gain amplifier, comparator, discrete-time mixer and filter, and ADC. Relatively to the latter, a high speed time-interleaved pipeline ADC prototype is implemented in a standard 130 nm CMOS digital technology from United Microelectronics Corporation

(UMC). The ADC is fully based on parametric MOS amplification which means that one could achieve a compact and MOS-only implementation. Furthermore, any high speed opamp has not been used in the signal path, being all the amplification steps implemented with open-loop parametric MOS amplifiers. To the author's knowledge, this is first reported pipeline ADC that extensively used the parametric amplification concept.

The dynamic performance of this ADC was experimentally evaluated and confirmed with 3 chip samples that were mounted the printed circuit board using a direct-bonding technique. Experimental results show that the measured DNL and INL errors at 120 MS/s are within 0.8/+1.4 LSB and ±2.0 LSB, respectively. The measured FFT for a 20 MHz input signal frequency (f_{in}) and 120 MS/s sampling frequency (F_s), demonstrates that the circuit achieves a peak SNR of 39.7 dB, a SFDR of 49.3 dB and a peak THD of 47.5 dB, corresponding to an ENOB of 6.2 bits. The ADC features an active area below 0.12 mm^2 and dissipates less than 14.5 mW at 120 MS/s and 1.2 V supply, resulting in standard figure of merit (FOM) better than 191 [fJ.mm^2 *per* conversion].

List of Symbols and Acronyms

γ	Body effect coefficient
μ	Carrier mobility
μ_0	Magnetic Permeability
μ_n	Carrier mobility
ϕ	Clock phase
ϕ_F	Fermi potential
ϕ_t	Thermal voltage, 0.0259 V at 300 K
ψ_s	Surface potential
A_{IGS}	Gate leakage mismatch factor
A_{VTH}	Threshold voltage matching parameter
c_{ov}	Overlap capacitance
C_{ox}	Total Oxide capacitance
C'_{ox}	Oxide capacitance per unit of area
clk	Master clock
$dBFS$	dB Full Scale
F	Noise factor
f_{in}	Input signal frequency
F_s	Sampling frequency
f_T	Transition frequency

g_{d0}	Drain to Source conductance at zero bias
g_{ds}	Drain to Source conductance
g_m	Transconductance
k	Boltzmann's constant
L	Channel length
l_{ov}	Overlapped length
N_A	Acceptor concentration
N_D	Donor concentration
p	Intrinsic parametric gain factor
q	Unity charge
S	Elastance
T	Absolute Temperature
t_{ox}	Oxide thickness
V_t0	Threshold voltage with zero bias
V_th	Threshold voltage
V_{FB}	Flat band voltage
v_{sat}	Velocity saturation
W	Channel width
3G	3rd Generation for Mobile Communications
A/D	Analog-to-digital
ADC	Analog-to-digital Converter
ATG	Asymmetric Transmission Gate
BB	Baseband
BiCMOS	Bipolar and CMOS technology
BS	Bulk-switching

BSIM3	Berkeley Short-channel IGFET Model version 3
BSIM4	Berkeley Short-channel IGFET Model version 4
BTBT	Band-to-band-tunneling
BW	Signal Bandwidth
CMOS	Complementary Metal-Oxide-Semiconductor
CT	Continuous Time
D/A	Digital-to-analog
DAC	Digital-to-analog Converter
DCR	Direct Conversion Receiver
DIBL	Drain-induced-barrier-lowering
DNL	Differential Nonlinearity
DT	Discrete Time
EDT	Edge-direct-tunneling
ENOB	Effective Number of Bits
FFT	Fast Fourier Transform
FOM	Figure-of-merit
FQ	Flash Quantizer
GaAs	Gallium Arsenide
GIDL	Gate-induced-drain-leakage
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
IF	Intermediate Frequency
IGFET	Insulated-Gate Field-Effect Transistor
INL	Integral Nonlinearity
IP	Internet Protocol

LNA	Low Noise Amplifier
LSB	Least Significant Bit
LTE	3GPP Long Term Evolution
MASER	Microwave Amplification by Stimulated Emission of Radiation
MDAC	Multiplying Digital-to-analog Converter
MIM	Metal-insulator-metal
MOM	Metal-oxide-metal
MOS	Metal-Oxide-Semiconductor
MOSCAP	MOS Capacitor
MOSFET	MOS Field Effect Transistor
MPA	MOS Parametric Amplifier
MSB	Most Significant Bit
MTBA	Multiply-by-two Amplifier
NF	Noise Figure
NF	Noise figure in dB
NMOS	N-channel MOSFET
NQS	Non-quasi-static
opamp	Operational Amplifier
PAMP	Parametric Amplifier
PAN	Personal Area Network
PCB	Printed Circuit Board
PFBL	Positive Feedback Latch
PMOS	P-channel MOSFET
PSTN	Public Switched Telephone Network
PVT	Process-supply-temperature

QoS	Quality of Service
RF	Radio Frequency
S/H	Sample-and-hold
S/N	Signal-to-noise Ratio
SC	Switched Capacitor
SCE	Short Channel Effects
SDR	Software Defined Ratio
SF	Source Follower
SFDR	Spurious Free Dynamic Range
SICAS	Semiconductor International Capacity Statistics
SINAD	Signal-to-noise and Distortion Ratio
SNDR	Signal-noise-plus-distortion Ratio
SNR	Signal-to-noise Ratio
SoC	System-on-chip
TG	Transmission Gate
THD	Total Harmonic Distortion
UWB	Ultra WideBand
WiFi	Standard for wireless local area networks
WLAN	Wireless Local Area Network
WSpW	Wafers per Week

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation and Scope

Silicon-based electronics has been one of the key factor contributing for the creation and refinement of modern and sophisticated end-user applications. Examples of these achievements are reflected by the most recent mobile platforms able to carry out very complex tasks. In fact, they aggregate, in a single product, complex multi-standard and multi-mode radio transceivers (e.g., GSM, 3G, WiFi, Bluetooth) as well as digital processors able to reach high processing capabilities reinforced by significant amount of memory. But these driving forces are still strongly accelerating, and therefore pushing the technology to continue to grow at a high rate. The known Moores' law [1] for digital technology, Edholm' law [2] for access bandwidth and Metacalfe' law [3] for network value, try to quantify this technology escalation.

At the end of 2008, more than 4 billion mobile phones were estimated to exist world-wide, representing more than 60% of penetration. Another emerging market of wireless sensor networks will tend to grow significantly in the next years, which can already reach approximately 120 million of remote units by 2010. Those huge numbers of devices results from the continuous and successful increase of the digital processing capacity and also, indirectly, from the selected supporting technology: the Complementary Metal-Oxide-Semiconductor (CMOS). Alternative technologies, like Bipolar or GaAs, are less attractive for transistor density increase due to downscaling lithography issues.

In the last decades, the technology has evolved from minimum device length ranging from 10 μm in 1971 to 45 nm in 2008 [4] and 32 nm in 2010. Interestingly, Figure 1.1 shows the evolution of the installed wafer production capacity, in Wafers *per* Week

(WSpW), both for MOS and Bipolar technologies, according to SICAS data [5]. It is clear from it that the production capacity has been completely dominated by MOS technology in last years[1]. Relative process simplicity and high manufacturing capacity contributed to lower the wafer die cost of CMOS when compared to Bipolar-CMOS (BiCMOS) or other alternative compound of distinct semiconductor technologies. It

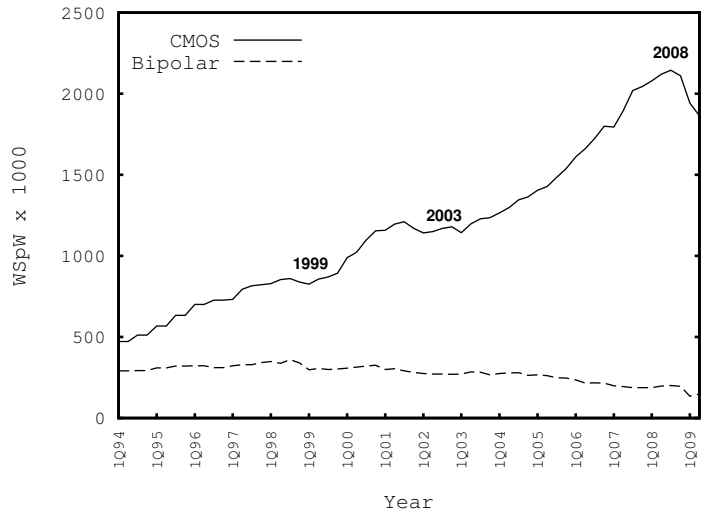


Figure 1.1: Comparative evolution of the Bipolar and MOS installed wafer production capacity, based on data available from SICAS [5].

is likely that competitive pressure will dictate that wireless applications (namely those related to a software defined radio architecture in a single chip), which can be implemented with acceptable performance in CMOS, will be effectively implemented in CMOS.

The CMOS scaling path has been evolving towards to higher integration level. In addition to this corresponding increase in device density, the process has significantly improved the switching speed, which is the result of higher transistor transition frequency (f_T) (this is a positive improvement for the integration of analog radio-frequency stages). On the other hand, the total chip power dissipation tends to be more demanding due to higher number of transistors *per* area unit. One of the selected countermeasure consists in reducing the applied power supply voltage, which will be as low as 0.5 V by 2020 [6]. Relatively to the transistor intrinsic gain (defined as the ratio between the device transconductance and the output conductance, g_m/g_{ds}), this value has decreased due to higher g_{ds}, degrading the gain achieved by operational amplifiers (opamps) and worsen the performance of closed loop configurations.

As a partial conclusion, while Moore's Law has been predictive of continuous gains

[1] As an out of scope remark, it can be detected in graphic 1.1 the impact of economic crisis in 1999, 2003 and 2008.

in digital circuit scaling, the relative performance of analog circuits has not scaled along with digital ones, and some inherent challenges intensify with reduced supply voltage. In order to overcome some of these issues, new design methodologies both at the circuit and system levels have to be addressed together for the design of optimum wireless transceivers in submicron and nanoscale CMOS.

As digital circuitry has scaled down, it has become practical to use digital processing in conjunction with analog functions to offload some of the already known bottlenecks, making the track of digitally assisted analog an important one. At the analog circuit level, amplifier configurations stages with gain directly dependent on the g_m/g_{ds} ratio will experience a performance degradation due lower size MOS transistors. Alternatively, new amplification approaches have to be found or recovered from earlier electronics, one of them being the Parametric Amplification.

In a Parametric Amplifier (PAMP), the amplification is governed by varying, with time, the reactance value of a capacitor or an inductor. In a traditional amplifier, the gain is mostly dependent on the transconductance and output conductance, and therefore, relies in a "resistive" type of configuration. Since the intrinsic gain process of a PAMP does depend on a time-varying reactance rather than on a "resistance", it is intrinsically noiseless. Moreover, due to its simple nature, it adapts well to a low power supply, and since this reactance can be implemented using standard digital MOS devices, it is a promising technique to be used in digital nanoscale technologies. In order words, this contributes for the task of designing a full transceiver integration in a pure digital technology towards a low-cost MOS-only implementation.

1.2 Contributions of this work

The main objective of this work is to demonstrate the use of the parametric amplification technique as an alternative approach to overcome some of the difficulties of designing traditional analog circuits, as CMOS technology scale evolves into nanoscale range. This is achieved by a set of research contributions [7, 8, 9, 10, 11, 12] that are described ahead in more detail.

- A modified MOS parametric cell amplifier is proposed, [9]. Instead of a single MOS device, this modified cell uses two half-sized MOS devices which are connected in parallel and with one of the tied terminals left floating. As a consequence it is shown that it becomes possible to decrease the effect of extrinsic gate parasitic capacitance during amplification phase, and therefore reducing the

loading effect on the effective gain. Main expressions are derived for this class of MOS parametric amplifier cells.

- Application of this amplifier cell in some of the most important modern transceiver building blocks. It is shown how this cell can be used in an original passive sampling mixer [12], a comparator [7] (for low resolution flash ADC) and in multiplying digital-to-analog (D/A) converter (MDAC[2]) for multi-step, algorithmic or pipeline ADCs, [10].

- Both for the comparator and MDAC/residue-amplifier, analytical expressions are derived and are related with offset, gain accuracy, noise. This supports a design methodology for this type of cells. They demonstrate that it is possible to design medium resolution (5 to 8 bit) ADCs [11] using this parametric technique and without the need of post-processing digital calibration.

- An 8-bit 120 MHz time-interleaved pipeline ADC is designed and fabricated in 0.13 μm digital CMOS technology, using only MOS devices as consequence of the extensive use of parametric amplification techniques within the comparators and MDACs. It is then proved, by experimental results obtained from three different samples, that this parametric amplification technique is well adapted for purely digital CMOS technology, [11].

To the author's knowledge, this work presents the first silicon proved pipeline ADC fully implemented using a discrete-time parametric technique, under digital CMOS technology. Therefore, benchmark with other analogous implementation is, at the moment, difficult to achieve.

1.3 Thesis Organization

This thesis is structured in seven Chapters, which are complemented by a couple of Appendixes. After this introductory Chapter the following one is dedicated to give an insight into the design space of modern wireless systems supported on modern nanoscale digital CMOS technology. As a result of the previous analysis, it becomes clear that the CMOS analog design process faces the need to use alternative circuit and transceiver system design techniques to achieve power, area and cost requirements. Within this context, Chapter 3 goes into the foundations of the parametric amplification principle,

[2]The main function of an MDAC block is to reconstruct a residue and amplifying it by a power of 2 for subsequent analog processing.

known since the middle of the 20th century. This technique is presented as a promising one to be recovered and adapted to CMOS digital technology, by means of a discrete time configuration, which is fully analyzed in Chapter 4. Latest sections of Chapter 4 focus on the use of parametric amplifiers in major digital transceiver buildings blocks, specially inside an ADC. For this last circuit, the design, implementation and testing details of an integrated prototype that is fully based on parametric amplification are given under Chapters 5 and 6.

In more detail, Chapter 2 aims to demonstrate that the wireless system design space is shaped by the increasing diversity of modern multi-standard wireless system requirements on the one hand and, on the other hand, by the evolution of CMOS technology into deep submicron range. Not only the former pushes the transceiver architecture towards a software defined one but also, the technology scaling increases the available digital processing power *per* unit of wafer area. Therefore, this Chapter presents an overview of the technology scaling and its impact on the devices performance, and also gives an overview concerning the most suited transceiver architecture for the modern CMOS based wireless environment.

The analysis in Chapter 2 concerning CMOS scaling shows the impact on device conductance, intrinsic gain, speed, noise and leakage. It reveals that the design of traditional analog blocks, e.g., operational amplifiers, are getting more difficult to achieve due to several reasons, one of them being the reduction of the available voltage headroom. This Chapter ends with an overview about present emerging circuit design techniques to overcome some of the above limitations. This Chapter is complemented by the Appendix A in which, the signal amplifier from RF to baseband, mixer and oscillator are discussed, with the emphasis on inductor less topologies as well as open-loop configurations. This first Appendix also presents a simple co-design strategy involving these three circuits and also refers to an alternative merged topology.

Interestingly, many of these blocks only need limited gain amplifiers, and therefore the use of a reactance based amplifier is an alternative that the following Chapters demonstrate to be compatible with CMOS digital technology.

Chapter 3 recovers the concept of parametric amplifiers, that rely on reactance elements to achieve signal amplification, rather resistive ones. The classical Manley-Rowe power relations are presented as a fundamental result to understand the transfer of power between the different signal frequencies involved on the circuit operation. It is also shown that the MOS variable capacitor (varactor) is a good candidate for the implementation of the continuous time parametric converter. The gain achieved by this type of structure is of limited value but since the amplification is reached by the

parametric change to the capacitance value, it has the advantage of being intrinsically noiseless.

The extension of the parametric amplification to the discrete-time domain is described in Chapter 4. This Chapter describes the principle of operation of the basic parametric amplifier cell complemented by analytical models for the gain and noise. A modified MOS parametric cell is then presented in order to reduce some parasitic capacitance effects from the original cell, [13]. The Chapter ends with the use of parametric amplification in common analog MOS circuits. The main focus goes to a discrete-time mixer and filter, a comparator and a multiplying by two amplifier. It is shown that the MOS parametric cell can be used from high to low frequency applications.

Chapter 5 is completely dedicated to give a detailed description about the design of a time-interleaved pipeline ADC for intermediate frequencies (IF) frequencies. The design is validated through pre and post layout simulations using BSIM3v3 MOS model. For the used 0.13 μm technology, this is the most advanced model available since BSIM4 is only available for lower technology nodes.

The experimental results of the pipeline ADC prototype described in Chapter 5 are presented in Chapter 6. The achieved results, based on measurements of three chip samples, show the effectiveness and reliability of using parametric amplification techniques in the design of moderate speed and medium resolution ADCs.

The last Chapter (Chapter 7) draws the most relevant conclusions of this dissertation and projects future work improvements on the application of the parametric signal processing in nanoscaled CMOS technology.

Chapter 2

Wireless system and circuit design space in modern digital CMOS technology

2.1 Introduction

Among all available processes, GaAs process has better speed footprint when compared to standard CMOS digital technology. The former presents charge carriers with greater mobility and saturation velocity, v_{sat}, thus pushing the f_T to more than 250 GHz [14]. Considering comparable generations (technology nodes), CMOS has a lower f_T, lower g_m and lower driving capabilities. Nevertheless, CMOS has lower cost since it is fabricated from Silicon and has less demanding fabrication process. Moreover, the mobility of the PMOS transistor is much higher when compared to an equivalent GaAs structure, which enables the implementation of efficient and complementary digital gates, thus reducing drastically static power consumption.

The wireless mobile digital system has been chosen in this work since it has been viewed as a good representative showcase to illustrate the modern market trends towards multi-application, multi-mode, and multi-standard encapsulation in a single box. The corresponding end-user equipment must be portable while maintaining long battery life. It should also includes components that ranges from multiple RF front-ends to baseband interactive multimedia and sensors devices. As a consequence, the interface to the analog world is then accomplished at distinct planes including high to low frequency domains. The open-source and semi open-hardware project, OpenMoko [15], original basis for the Android platform [16], is a good illustration for this concept.

While that an initial product version can use RF front-end built on high performance technology (e.g., GaAs), the success market grow of one particular solution will determine the need to lower costs. Usually, this will reflect the necessity to push analog RF components into CMOS. This technological migration step, is most of the times, not just a simple transistor replacement from traditional circuits, but frequently involves the research for new circuits and/or alternative systems as well as for new design paradigms.

In this Chapter, an overview over new and existing techniques within the context of modern multi-standard wireless systems (from low data rate sensor to high data rate multimedia applications) is presented. The strategically adopted top-down approach will aggregate the information into three designer perspectives:

- **Device Level**: lithography scaling impact on MOS device performance,
- **System Level**:

  Application/Network Level,

  End-user terminal architecture,
- **Circuit Level**: CMOS analog design options.

The design space that appears over the following sections reflects the trends to achieve a better co-design parameterization between system and constituting circuits, to reach a transceiver plus baseband integrated in a single chip (i.e., in a System-on-Chip, SoC)

One of the objectives of this Chapter is to revisit some of the key building blocks of an RF transceiver and check if they can benefit from alternative designs that are better adapted to standard digital CMOS technology. One proposal of this work is to recover the concept of Parametric Amplification (historically significant in the period from 1947 to 1970) and applying it, wherever possible, inside the digital transceiver and baseband circuitry. One of the building blocks is the ADC (the last analog block in the signal path receiver chain) which will be described in more detail in last chapters.

2.2 Wireless System Overview and Requirements

The rate at which information is transmitted assumes an important role in digital based communication systems since it represents one of the major factors that determines network type, online services and applications. Another conditioning fact is the

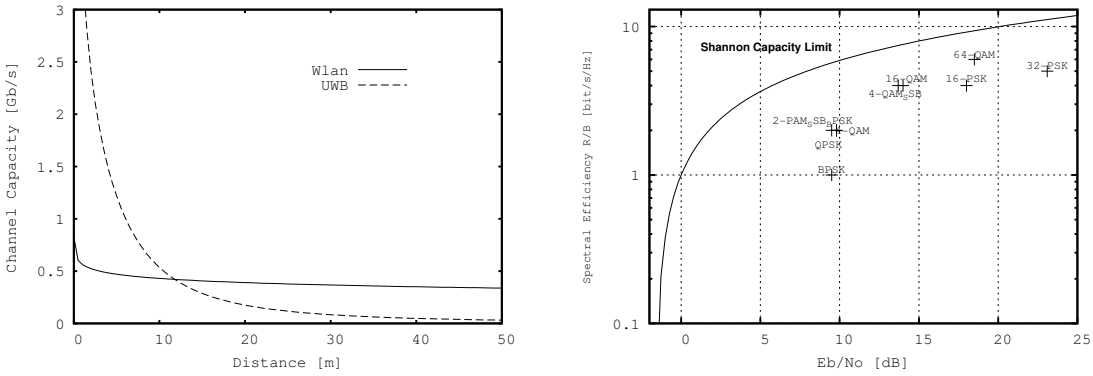
access mode which ranges from wired to wireless types. In both cases, the associated communication channel is affected by additive noise that difficults the recovery of the information at the receiver side. If the added noise is considered to have a flat power spectral density N_0 in [dBm/Hz] ($N_0 = kT$, k being the Boltzmann constant and T the temperature in Kelvin), over the channel bandwidth (BW) in [Hz], then by applying the Shannon's theorem [14] to it, the upper bound for the channel capacity C in [bits/s] is expressed by

$$R \leq C = BW \cdot log_2 \left[1 + S/N\right] = BW \cdot log_2 \left[1 + \frac{E_b}{N_0} \cdot \frac{R}{BW}\right] , \tag{2.1}$$

where S denotes the received signal power in [W], N denotes the total power of the white noise impairing the received signal and R denotes the information rate in [bits/s]. Furthermore, under these conditions, the signal power can be expressed by the product between the energy per information bit, E_b, and by the information rate R. Similarly, the total noise power N from the channel is simply given by multiplying the noise power spectral density N_0 and its bandwidth BW.

Expression (2.1) shows that higher data rates are achieved by increasing the channel bandwidth. The same result is predicted by increasing the signal power but at lower changing rate (due to the logarithm function). Interestingly, it also shows that the increase of the channel bandwidth can be used to reduce the transmission power as long as a low data rate value is kept. Results from the application of (2.1) to narrowband and wideband channels cases are presented in Figure 2.1(a), where a constant total power at the transmitter has been assumed. For both cases, the graphic represents the decrease of channel capacity when moving the receiver away from the emitter. This reflects the signal to noise ratio, S/N, degradation due to signal power attenuation which is dependent on the distance between emitter and receiver [17]. More visible than in the narrowband case, the ultra wideband (UWB) communications offers both high data rate transmission over short distances and low data-rate over long ones. One alternative way to use UWB signal is to generate short duration impulse radio signals [18, 19]. This pulse-based signal can be implemented using simple digital gates [19], a power amplifier and a wideband antenna, mounted in a topology well adapted to digital CMOS technology.

A path for approaching the Shannon's limit consists on using modulation schemes over the communication channel. Figure 2.1(b) shows the relative position of typical modulation schemes. Unfortunately, the theorem does not describe any modulation type that reaches the Shannon's limit, and most popular schemes require far greater E_b/N_0 than -1.6 dB. The use of high level modulation schemes, such as 64-QAM,



(a) Wide versus narrowband (b) Modulation and Shannon limit

Figure 2.1: Shannon channel capacity.

demands for higher S/N levels and also for more digital signal processing to recover the original digital stream. However, this overhead digital processing is well supported by CMOS digital processors that tends to get more powerful with technology downscaling.

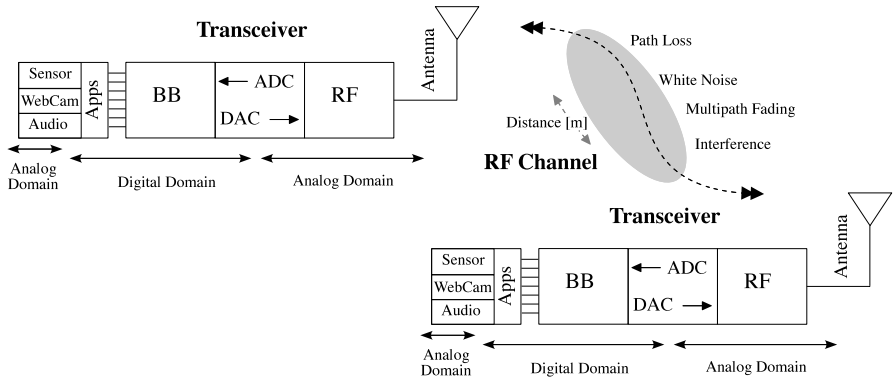


Figure 2.2: Simplified point to point communication model.

This initial discussion has intrinsically assumed the simplified wireless communication path model depicted in Figure 2.2. In this point-to-point model, the user data is delivered to the baseband processor unit that prepares and encodes the data to be sent over the modulated signal injected in the channel. Since the radiated signal by the antenna is intrinsically analog, a Digital-to-Analog (D/A) Converter (DAC) is needed at the transmitter side. Due to reasons that are clarified in section 2.4, the DAC output does not feed directly the antenna. A intermediate RF stage is used to modulate and increase signal power before radiating it.

Despite a spatial directive gain introduced by the antenna, the radiated signal suffers

enormous degradation [17] along the propagation path in direction to the receiver. Signal attenuation, additive noise, multipath fading, co-channel interferer, are just a few examples that will difficult the information recovery at the receiver side [17].

At the reception equipment the reverse operation is performed. Starting from a RF front-end block, the amplified and downconverted signal is delivered to the Analog-to-Digital (A/D) Converter (ADC) which will deliver the digital signal for the baseband (BB) processor. Its function is to recover the original data information. By using a digital transmission, the task of detecting transmission errors and correct them is most likely shifted to an algorithm and software issue, which contributes for an improved flexibility. Shifting the signal processing as much as possible to the digital domain, has pushed the development efforts towards a SoC approach by means of integrating in the same technology (CMOS in this case) and inside the same chip the RF, ADC and digital processing parts.

From the perspective of the Metcalfe's law [3] this point to point communication case is of limited interest since this law states that the value of a communication system is related to the square of the number of connected users in the network. The evolution of the wired and wireless networks are now entering into a complementary phase in which very high data rate services are supported by the former type of networks while mobility at moderate to high data rate are given by a variety of wireless network standards. Major advances have been achieved, with efforts focused primarily on voice communications over cellular networks and short-range Wireless Local Area Network (WLAN) data communications.

Today, a wireless device needs to communicate in a heterogeneous environment, including current and next-generation cellular wide area networks, medium-range WLAN, and short-range Personal Area Network (PAN). This evolution has been supported in a paradigm on how the digital stream is sent over the network. In fact, the necessity to better adapt and optimize network resources to multi-applications environment, e.g. multimedia, has forced the evolution from switched-circuit centric approach (due to historically structure of the rele-based PSTN) to a packet switched centric one. This migration has, in great extent, being promoted by the networking layer defined by the Internet Protocol (IP) communication stack, [20]. This layer has being one the most successful technologies used to interconnect all kind of networks, extending the concept of the Internet to the wireless environment.

In contrast with these ubiquitous mobile and advanced multimedia end-user applications, the emerging wireless sensor network market is based on the premise that the remote nodes have to be ultra-low power, at very low cost and should be almost

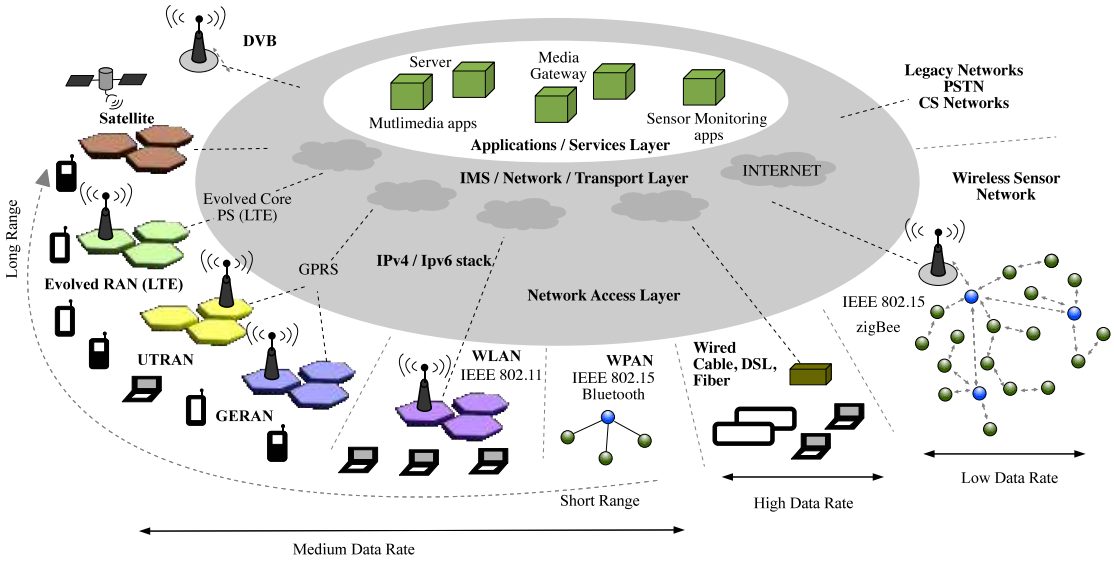


Figure 2.3: All IP network architecture approach.

Coverage	Standards	RF Frequencies	Access	Modulation	Data Rate	Application Centric
Cellular	GSM / GPRS	900 MHz, 1.8 GHz	FDM/TDMA	GMSK	170 kbps	Voice. First CMOS IC in 2001. Power amplifier in GaAs or Bipolar.
	UMTS	2.1 GHz	W-CDMA	QPSK	384 kbps, 14 Mbps (HSPDA mode)	Voice, moderate data mobile. Implementation in CMOS reduces costs.
	3GPP LTE	2.1 GHz		OFDMA, SCFDMA	360 Mbps	Multimedia, Hight data rate. Implementation in CMOS reduces costs.
Short-Medium Range	WiMax IEEE 802.16e	2-6 GHz	TDMA	OFDMA-64QAM	144 Mbps	Multimedia, High data rate, Long range. Implementation in CMOS reduces costs.
	WLAN IEEE 802.11(abg)	2.4 GHz, 5.4 GHz	CSMA-CA	DPSK, OFDM, DQPSK	280 Mbps	Multimedia, High data rate. Dominated by CMOS.
	802.15.3a UWB	3.1-10.6 GHz	OFDM	QAM	500 Mbps	Wireless personal network, multimedia. Impulse Radio UWB is a promising technique for low cost CMOS implementation.
	Bluetooth	2.4 GHZ	FHSS	GFSK	1 Mbps	Wireless personal network, short range, Audio. Dominated by CMOS.
	ZigBee	2.4 GHz	CSMA-CA	QPSK	250 kbps	Monitoring, Low data rate and Ultra-low power.Dominated by CMOS.

Table 2.1: Summary of most common wireless network standards.

maintenance free over the network live age (from 2 to 10 years). Interestingly, besides the effort to reduce energy consumption at the circuit level, e.g., by choosing alternative circuit topologies, important research is being done to find multiple low-cost solutions for local energy harvesting [14] to be integrated within the sensor nodes. The very tight specifications concerning the power consumption are balanced by less stringent data rate requirements due to the nature of the data transmitted by the majority of the sensors. Additionally, the low cost demand is pushing to a CMOS SoC solution for the sensor node. ZigBee, IEEE 811.15 are just standards specially defined for wireless sensor networks.

Figure 2.3 demonstrates that modern communications are grouped in distinct families of cable and wireless networks providing varying coverage area, data rate and quality

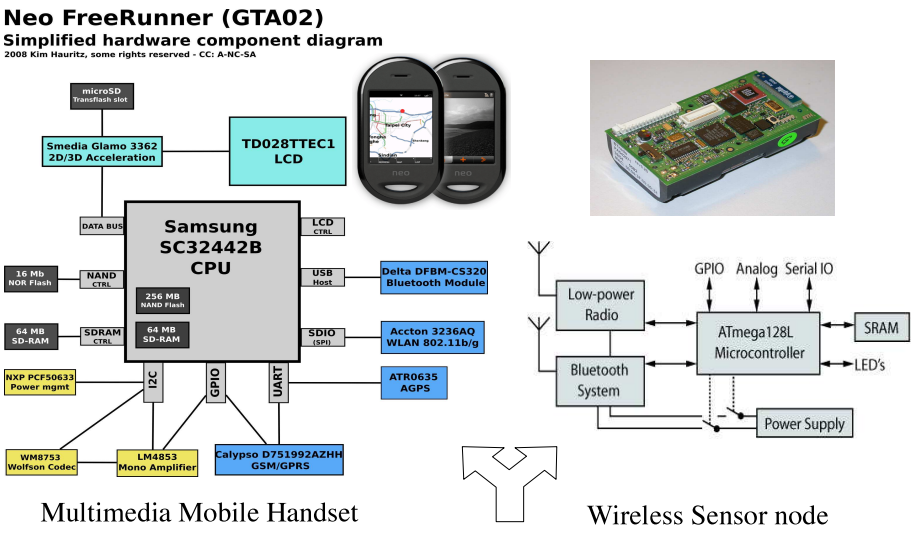


Figure 2.4: Multimedia wireless equipment (OpenMoko) [15] and wireless sensor node (BTNode) [22].

of service (QoS). Some of the key specifications of modern wireless data systems are presented in Table 2.1.

Heterogeneous networks that use several types of radio systems are more suitable for diverse user demands than homogeneous networks (that use a single radio technology). Portable devices typically have more than one type of wireless interface built-in. To satisfy the bandwidth and QoS requirements of the applications, the mobile devices need to be able to seamlessly switch among their wireless network interfaces.

Each network infrastructure tends to be divided into layers: application layer, transport and backbone network, access network and user equipment. This division will be more explicit in the next generation networks, being the 3GPP LTE one of the strongest proposed path [21]. Interestingly, the integration of several radio access modes have been implemented in first place at the user terminal side. An excellent example is the modern smart phone OpenMoko [15]. The OpenMoko block diagram is shown in Figure 2.4 to illustrate the integration of GSM/GPRS, Wi-Fi and Bluetooth radio modules in the same platform.

New paradigms are emerging, in order to better use the available bandwidth conditioned by the objective to maximize the full integration in CMOS technology. Two directions have been defined. The first one, where a multistandard, multimode and high data date approach is followed to enhance end user mobile multimedia applications. The second one is to reduce to the minimum the energy consumption for wireless sensor systems, as the BTNode [22] shown in Figure 2.4. Both high and low data rates, with different end applications specifications, requires the integration of multiple functions,

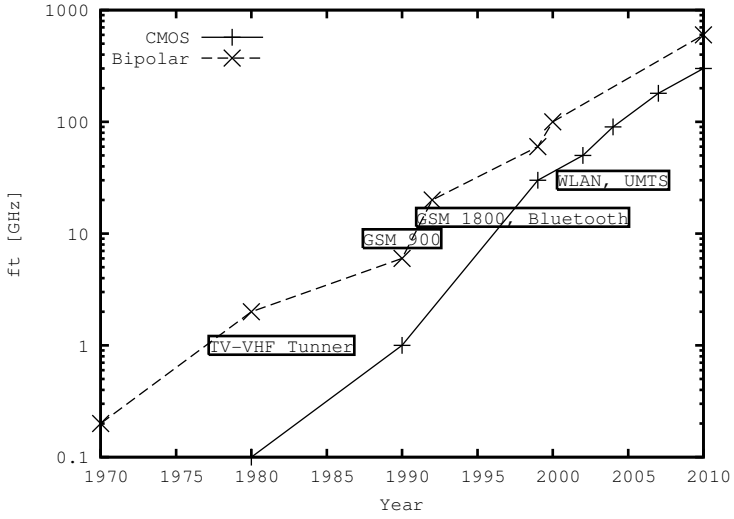


Figure 2.5: Transit frequency comparison between CMOS and Bipolar technology.

preferentially, in a single chip. As stated before, several technologies are available (e.g. GaAs, Bipolar) but cost dictates CMOS. The problem is that the scaling of CMOS has a global negative impact in analog circuits due to reasons explained next, in section 2.3. This demands new circuit design techniques, but, most likely, this is still not enough to solve entirely all the issues. Instead, a more closed and jointly design between system and circuit is gaining momentum. A good example is the implementation of routing energy aware protocols [14] in wireless sensor networks.

2.3 CMOS Scaling Impose Challenges for Analog Design

MOS devices are scaled down with the objective of enhancing speed performance and inserting more units within the same Silicon wafer (to reduce cost *per* transistor and increase digital processing capacity). As referred previously, for the last three decades, the feature size of the MOS transistor has been reduced due to significant improvements in lithography at a rate of 0.7 times every three years. When the technology progressed to the turning point where high volume process nodes provided transition frequencies, f_T, on the range of 30-40 GHz (final of the 1990 decade), it begun to be possible to integrate the design of RF circuits sharing the same CMOS substrate that was used by the baseband circuits. Figure 2.5, shows a speed performance (in terms of f_T) comparison between CMOS and Bipolar transistors. The relation with some major wireless standard data bearers is also indicated.

Parameter	Constant Field Scaling	Constant Voltage Scaling
Dimensions (W, L)	$1/\alpha$	$1/\alpha$
t_{ox}	$1/\alpha$	$1/\alpha$
$_{VDD}$	$1/\alpha$	1
Fields	1	α
V_{th}	$1/\alpha$	1
Idrive	$1/\alpha$	α
C_g	$1/\alpha$	$1/\alpha$
Delay time	$1/\alpha$	$1/\alpha^2$
Power.Delay	$1/\alpha^3$	$1/\alpha$
Power / Circuit	$1/\alpha^2$	$1/\alpha$
Power / Area	1	$1/\alpha^3$
Line Resistance	α	α
RC	1	1
I.R/V_{DD}	α	α^2

Table 2.2: Theoretical CMOS scaling impacts, [24].

The scaling of MOS devices is performed in both the vertical and the lateral directions. The lateral shrinking is performed to obtain a shorter gate and a higher packing density, while the vertical is scaled accordingly to maintain the MOSFET functionality. Considering a classical constant field scaling approach, the supply voltage is scaled along with the oxide such that the electric field in the oxide remains constant. Since the MOS transistor is mainly driven by a current drift physical phenomena, the current density is strongly related to the electric field. Hence, if the same scaling factor, α, is applied to all dimensions and voltages, the electric field and current density remain constant and the DC characteristics are unchanged. The doping level of the substrate is increased by α so that the depletion layer thickness scales down with α. Under these conditions the circuit gets faster by a factor α, the power *per* circuit is reduced by α^2, the power delay product improves by α^3, and the density power remains constant. An alternative constant voltage scaling approach was also followed in the past but this approach ended when the gate length reached 0.5 μm and t_{ox} near 10 nm [23], since at this point the high fields and high currents tend to damage the gate oxide and leaded to device deterioration.

Despite the operational speed increase, facilitated by new generation of CMOS, the referred reduction of the voltage power supply, due to density power dissipation issues, is one of the major design constraints that affects the performance of analog circuits in CMOS. In analog CMOS circuit design, the transistor is preferentially used in saturation region for amplifiers circuits but must be in ohmic region for the implementation of switches. For the former case it is the transistor intrinsic gain that is important while for the switches is the lowest ON resistance.

2.3.1 g_m, g_{ds}, and transistor's intrinsic gain

One important parameter for the design of analog circuits is the gain achieved by an active device. In general, for a linear transconductance based device, as the case of a MOS transistor, the intrinsic gain is given by equation 2.2, where g_m is transconductance and g_{ds} is the output transistor conductance.

$$A_{vi} = \frac{g_m}{g_{ds}} \tag{2.2}$$

To obtain amplification it is usual to bias the MOS device into saturation where, assuming an approximate square law for the drain current (i_{DS}), the small signal transconductance is given by equation

$$g_m = \mu C'_{ox} \left(\frac{W}{L}\right) (V_{GS} - V_t) \tag{2.3}$$

where μ is the carrier mobility, C'_{ox} is the oxide capacitance per unit of area, W and L is width and the length of the transistor, respectively, V_t is threshold voltage and V_{GS} is the biasing voltage between the gate and the source of the MOS transistor.

Short channel effects (SCE) are becoming more important with the decreasing of the transistor lateral dimensions. When entering in carrier velocity saturation state, the transconductance is no longer dependent on the channel length, as shown in,[25],

$$g_m = \frac{1}{2} W \frac{\epsilon_{ox}}{t_{ox}} \upsilon_{sat} \,. \tag{2.4}$$

The corresponding transconductance efficiency, g_m/I_{DS}, in strong inversion is given by

$$\frac{g_m}{I_{DS}} = \frac{2}{V_{GS} - V_t} = \frac{2}{V_{ov}} \tag{2.5}$$

which reduces to

$$\frac{g_m}{I_{DS}} = \frac{1}{V_{ov}} \tag{2.6}$$

when the device suffers from carrier velocity saturation. Interestingly, g_m and g_m/I_{DS}, in strong inversion are independent of the transistor sizing and process parameters. They mainly depend on the DC drain current biasing I_{DS} and the overdrive voltage V_{ov} $(= V_{GS} - V_{th})$. Another interesting result is that, to a first order analysis, the transconductance efficiency halves the value when the device enters in velocity saturation case.

While the primary contributor to MOS drain-source conductance for typical conditions

is channel length modulation, drain-induced barrier lowering (DIBL) effect due to SCE also has to be considered [25]. In short channel devices, since V_t decreases when increasing V_{DS}, the overdrive voltage, V_{ov}, increases thus increasing the drain current. This increase of drain current with the increase of V_{DS} gives rise to an additional component to g_{ds}. The transistor's output conductance increase as V_{DS}/L increases. This ratio is growing because V_{DD} is not scaling as fast as the gate length. Furthermore, beyond 90 nm, gate leakage restricts the scaling of the oxide so that L and t_{ox} cannot be reduced proportionally. Despite the predicted increase of the device transconductance with scaling, the intrinsic gain is in fact reducing instead of growing, mainly due to the increase of the g_{ds}, as explained. High gain opamps are therefore harder to design, since not only the intrinsic gain of the active devices are lowering but also due to small voltage headroom available at lower power supply voltage. As a result, cascode techniques are getting more difficult to be employed. Alternatively, system topologies less dependent on complex opamp structures have been under research, being this work an example of this investigation.

In another important circuit block, the switch, the ohmic operation region of the MOS device is used instead of saturation. The ON and OFF states quality depends greatly on the output conductance of the transistor not only due to its nominal value but also due its dependence on the input signal. As technology scales, maintaining the linearity of the switches is a challenging task [24]. Determined for $V_{DS} = 0$ V, the parameter given by

$$g_{ds} = \mu_n C'_{ox} \frac{W}{L} (V_{GS} - V_t) \tag{2.7}$$

represents the small signal output conductance of a MOS transistor. Despite the result expressed by (2.7) is used for small signal analysis, a few remarks can be extrapolated with respect to MOS scaling effect in switches. Since the threshold voltage does not scale linearly with the supply voltage, the ON resistance tends to increase in smaller size CMOS nodes, which leads to longer settling times in Switched Capacitor (SC) circuits if not correctly addressed.

In a short conclusion, technology scaling tends to achieve a higher g_{ds} in saturation which degrades the gain but, at the same time, due to limited voltage range, the g_{ds} in linear region tends to be lower, degrading the ON resistance of the switch.

2.3.2 f_T, f_{max}, and parasitic capacitances

The decrease on the minimum length of the transistor has remarkably pushed the CMOS into the RF circuit design range. Usually the device cutoff frequency, f_T, is

Node	**nm**	**250**	**180**	**130**	**90**	**65**
L	nm	180	130	92	63	43
t_{ox}	nm	6.2	4.45	3.12	2.2	1.8
V_{DD}	V	2.5	1.8	1.5	1.2	1
V_{th}	V	0.44	0.43	0.34	0.36	0.24
g_m (peak)	μS/μm	335	500	720	1060	1400
g_{ds}	μS/μm	22	40	65	100	230
g_m/g_{ds}	-	15.2	12.5	11.1	10.6	6.1
f_T	GHz	35	53	94	140	210

Table 2.3: IBM digital CMOS technology characterization, from [26].

used to characterize the frequency response of an active transconductance device and it is defined as the frequency at which the current gain of the transistor is unity. Transposing this concept to the MOS transistor, results in a f_T given by

$$f_T = \frac{g_m}{2\pi C_{gg}} \tag{2.8}$$

where g_m is the transconductance value and C_{gg} is the total MOS gate capacitance.

It is visible in Table 2.3, that the f_T has entered in the hundred GHz range, which is an indication that the MOS transistor can be used in the RF domain.

The f_T mark is of limited use since in the majority of amplifier cases it is useless to use a device at the frequency where it produce an unity current gain. Instead, other parameters can be defined to complement this frequency characterization, as the maximum operation frequency [27].

Another useful parameter is the maximum oscillation frequency at which the maximum power gain is one [27, 14, 28]. This frequency mark is calculated by

$$f_{max} = \frac{f_T}{2\sqrt{g_{ds}\left(R_{g,poly} + R_i + R_s\right) + 2\pi f_T R_g C_{gd}}} \tag{2.9}$$

where $R_{g,poly}$ and R_s are the gate and source resistance, respectively. R_i is the real part of the input impedance due to Non-Quasi-static effects [29], and C_{gd} is gate-drain capacitance.

The analysis of (2.9) reveals that this frequency parameter is significantly dependent on the gate resistance and parasitic capacitances of the MOS transistor. Therefore, the device layout will play an important role on the calculation of f_{max} and there is no fixed trend for scaling it. Reported results from [14] show that f_{max} can be as high as 150 GHz using a 65 nm CMOS technology.

2.3.3 Noise and dynamic range

At RF frequencies, the noise components of a MOS device are generated thermally by the contribution of the channel and the gate. The first component, the drain channel noise $\overline{i_d^2}$, is due to the charge carrier fluctuation over the channel. These thermal fluctuations capacitively couple through gate-channel capacitance resulting in an induced gate noise current, $\overline{i_g^2}$. Clearly these two components are correlated and can not be treated independently. A third, statistically independent, noise source is thermally generated by the gate resistance and it can be modeled simply by

$$\overline{v_g^2} = kTR_g \triangle f \,, \tag{2.10}$$

where R_g is the distributed gate resistance, k is Boltzmann's constant, T is the temperature in Kelvin, and $\triangle f$ is the noise bandwidth. Comparatively with the others noise sources, the height of this one can be reduced if a multi finger structure is used for the gate layout.

The most dominant transistor noise component comes from the channel. As usual, this power current noise can be input referred just dividing by the square of the transconductance

$$\overline{v_{g,id}^2} = \frac{\overline{i_d^2}}{g_m^2} = 4kT\gamma \triangle f \frac{g_{d0}}{g_m^2} \,, \tag{2.11}$$

where g_{d0} is the zero drain-source voltage conductance of the channel, γ reflects both non-uniform charge distribution and high-field effects. For a long channel, γ (which reflects the excess drain noise) is theoretically equal to 2/3, but it may increase with the growth of the electrical field within the channel, when velocity saturation begins to dominate, [25].

The impact of the CMOS scaling on this noise component is primarily modeled by the term g_{d0}/g_m^2 and therefore the noise performance should improve with successive technology nodes.

From classical two-port theory, if a certain optimized noise driven matching strategy is followed, the circuit Noise Figure (F) reduce to a minimum value, F_{min} which is directly associated with the active device. F_{min} of a MOS transistor can be estimated using (2.12), [30, 14],

$$F_{min} = 1 + 2\,\frac{f}{f_T}\sqrt{g_m R_g \frac{\gamma}{\alpha}} \;>\; 1 + 2\,\frac{f}{f_T}\sqrt{\frac{1}{5}\frac{\gamma}{\alpha}} \,, \tag{2.12}$$

where the gate resistance R_g accounts for the material resistance and the Non-Quasi-

Static (NQS) gate resistance ($R_g = R_{g,poly} + R_{NQS}$), α assumed to be $g_m/gds0$. The lower bound has been determined by assuming that a multi-finger device was used to reduce $R_{g,poly}$ and therefore the total gate resistance is approximately given by R_{NQS}, which is estimated to be roughly equal to $1/5g_m$, [29]. This lower bound for the noise figure indicates that for a given operating frequency f, increasing f_T tends to improve noise performance, as reflected in (2.12). In Figure 2.6 the NF_{min} versus f_T is presented for several signal frequencies.

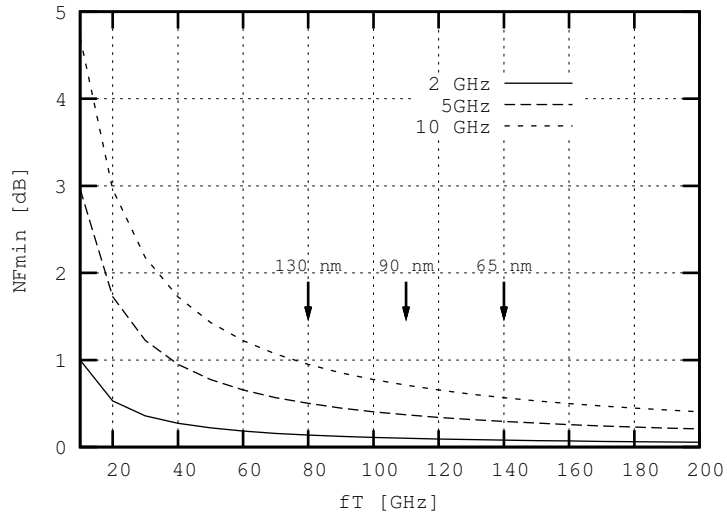


Figure 2.6: NF_{min} *versus* f_T, [14].

An additional noise contribution appears in the drain current mainly due to fluctuations of channel-free carriers (electrons in an NMOS device) as the result of random capture and emission of charge carriers by interface traps located at the Si-SiO_2 interface. An unified theory [31], indicates that this random capture and emission of carriers not only causes fluctuations in carrier number but also in their mobility. This latter fact introduces some level of dependence on gate biasing, which was not predicted in the previous results. Nevertheless, a $1/f$ noise spectrum is predicted for this noise component, known as flicker noise, if the trap density is uniform in the oxide. This noise component is dominant at low frequencies but it directly reflects on the phase noise of an RF VCO as well of mixers and, naturally on baseband circuits. The drain current noise power spectral density, S_{id}, is obtained by the trap density, the density of channel carriers and by the gate area. A few number of models have been developed, one of them being described by (2.13), [30], where K is a technology parameter for the device.

$$S_{id} = \frac{K\, g_m^2}{C_{ox}'^2 WL} \frac{1}{f} \tag{2.13}$$

Assuming that the trap density can be maintained constant over technology scaling,

the model predicts small changes in the drain current flicker noise component as long as the transistor dimensions are fixed. This is the result of the proportional variations of both the g_m and C_{ox}. However if the channel length of the transistor is scaled to minimum dimension it is expected an increase of the $1/f$ noise.

2.3.4 Leakage and OFF state current

With the oxide becoming too thin, due to quantum mechanical effects (QME) associated to the oxide potential barrier, [31], the probability of carriers to tunnel between gate and channel increases with CMOS scaling. Moreover, gate tunneling occurs both at the channel region and at the gate overlap regions over the source and drain regions. Besides the oxide thickness, the total tunnel gate current also depends on the gate voltage and gate area, as explicitly indicated in, [25],

$$\begin{aligned} I_G &\propto WL \cdot \frac{V_{oxide}}{t_{ox}} \cdot Q'_{INV} \cdot P_{tunnel}\left(V_{oxide}\right) \\ &= WL \cdot \frac{V_{oxide}}{t_{ox}} \cdot Q'_{INV} \cdot \left[e^{-\frac{E_B t_{ox}}{V_{oxide}} \cdot \left(1 - \sqrt[3/2]{1 - \frac{V_{oxide}}{X_B}}\right)} \right] \end{aligned} \tag{2.14}$$

where Q'_{INV} represents the inversion charge density in the channel below the gate, V_{oxide} is the voltage of across the oxide and $P_{tunnel}\left(V_{oxide}\right)$ the tunneling probability through the gate oxide which depends on a characteristic electric field E_B and on the oxide voltage barrier X_B, [25]. To model the impact of the gate leakage, the theoretical capacitive input gate impedance model has been complemented with a parallel tunnel conductance g_{gs} (representing an equivalent gate to source conductance), [32, 25]. This compound presents a conductance input type at low frequency and a capacitive at high frequency. The turning point between these two values occurs at frequency f_{gate}, which is given by, [32],

$$f_{gate} = \frac{g_{gs}}{2\pi C_{gsi}} = \begin{bmatrix} 1.5(\text{NMOS}) \\ 0.5(\text{PMOS}) \end{bmatrix} \cdot 10^{16} \cdot v_{GS}^2 \cdot e^{t_{ox}(v_{GS} - 13.6)} \tag{2.15}$$

where t_{ox} is given in [nm], C_{gsi} represents the intrinsic gate to source capacitance. With the decrease of the oxide thickness, t_{ox}, this leakage frequency related parameter tends to be higher, meaning that not only the gate current leakage increase (due to higher g_{gs}) but also the resistive behavior occupies a wider bandwidth. Data from [32], shows that f_{gate} changes from 0.1Hz for a 0.18 μm process to 1 MHz in 65 nm CMOS, showing an increase on the minimum operating frequency. Gate leakage affects circuit performance when implemented in processes with t_{ox} around 2 nm and below [25].

In other type of circuits, like switches, the behavior of the transistor in OFF-state is also affected by current leakages. Non zero leakages present in OFF-state generates an increase of the stand-by power, which is a major concern in long battery life applications such as wireless sensor nodes.

When in OFF-state conditions, the gate to channel leakage current no longer exists (as long a g_{GS} is zero), but the ones due to gate-drain overlap area are still present. Besides these components, the contributions that were neglected when in ON-state conditions have now to be added in the calculation of the total OFF state drain current, I_{off}, specially in deep submicron nodes (e.g., 45 nm and below). They include:

- DIBL and subthreshold current: subthreshold leakage current is mainly present in OFF-state conditions and increases exponentially due to threshold voltage reduction,[25],

$$I_{off,sub}|_{V_{GS}=0} = 2n\mu_0 C'_{ox}\phi_t^2 \cdot \frac{W}{L} \cdot e^{-\frac{V_{th}}{n\phi_t}} \tag{2.16}$$

  where, n is a substrate related factor and μ_0 is the carrier mobility. Not only the technology down-scaling is reducing, accordingly, the threshold voltage but also the SCE associated with the drain-induced barrier lowering (DIBL). The latter is explained from shorter drain to source distance and higher V_{DS} voltage that results in deeper channel depletion region, which lowers the potential barrier and therefore decreasing V_{th} and increasing I_{DS}.

- Gate related currents: the overlap between the gate and the drain/source produces a direct tunneling leakage current, gate edge-direct-tunneling (EDT) current (I_{EDT}), [33], and induce a junction current to this substrate, gate-induced drain-leakage (GIDL) current (I_{GIDL}). Also, in addition to GIDL, the injection of hot-carrier from the substrate to the gate can happen if high electric fields are present in the gate-drain overlap region. All these contributions depend directly from oxide thickness and therefore will increase as t_{ox} is downscaled.

- Junction related: If the drain voltage is high enough, the parasitic diode between the drain and substrate could be strongly reverse biased originating a reverse leakage current. This component is strongly dependent on temperature and, for a fixed sized transistor, it is inversely proportional to the technology scaling factor. The second junction related component is the band-to-band-tunneling (BTBT) leakage. The mechanism behind this component takes place at depletion regions close to the heavily doped drain junction. A sufficiently strong electric field applied in this region may originate a current flow between the drain and the substrate due to band to band tunneling effect. This component increases

exponentially with the decrease of the effective gate length, due to higher lightly doped-drain or pocket-doping concentrations, [33].

- Punch-through: another SCE effect (in complement to DIBL) that arise when the depletion regions associated to the source and drain PN junction touch each other. This mechanism create a second path between the drain and source enabling current to flow. This component varies inversely with the channel length, i.e., with the CMOS downscaling.

2.3.5 Transistor's Matching

Different levels of gate leakage currents are expected among transistors due to t_{ox} variations and other process parameters. This contributes for the overall transistor mismatch model, [34], with an additional term given by, [25],

$$\sigma\left(\Delta I_G\right) = A_{IGS} \cdot J_{GS}\left(V_{GS}, V_{th}\right) \cdot \sqrt{WL} \tag{2.17}$$

where A_{IGS} is a gate leakage mismatch factor (corresponding to parameter X_{IGS} in [32] being approximately equal to 0.03), and J_{GS} is the gate leakage current density. Combining this mismatch contribution with the conventional one due to threshold-voltage mismatch [34] and neglecting the contribution of the transconductance factor mismatch (if the transistor is not in strong inversion, [25]), results in a global relative drain current mismatch model described by,

$$\sigma^2\left[\frac{\Delta I_{DS}}{I_{DS}}\right] = \overbrace{\left(\frac{A_{VTH}}{\sqrt{WL}} \cdot \frac{g_m}{I_{DS}}\right)^2}^{\text{Vth related}} + \overbrace{\left(\frac{A_{IGS}}{\sqrt{WL}} \cdot WL \cdot \frac{J_{GS}\left(V_{GS}, V_{TH}\right)}{I_{DS}}\right)^2}^{\text{Gate leakage related}}, \tag{2.18}$$

where A_{VTH} is a technology-dependent coefficient used to model the threshold voltage matching. Being approximately proportional to the oxide thickness, the variation of A_{VTH} due to technology downscale implies a reduction in the first term of (2.18), if a transistor fixed size is assumed between technological nodes. The traditional design approach to improve matching, by increasing the transistor area, has now to be corrected by the opposite impact of the gate leakage matching. The latter term degrades with the increase of transistor area or more precisely with the increase of the device length, as it is evidenced in,

$$\sigma^2\left[\frac{\Delta I_{DS}}{I_{DS}}\right] = \overbrace{\left(\frac{A_{VTH}}{\sqrt{WL}} \cdot \frac{2}{V_{ov}}\right)^2}^{\text{Vth related}} + \overbrace{\left(\frac{A_{IGS}}{\sqrt{WL}} \cdot \frac{2}{V_{ov}^2} \cdot \frac{J_{GS}\left(V_{GS}, V_{TH}\right)}{\mu C'_{ox}} \cdot L^2\right)^2}^{\text{Gate leakage related}} \tag{2.19}$$

which was obtained under saturation conditions.

2.3.6 Passive components: MIM Capacitors and Inductors

In mixed-mode CMOS chip design, the effect of CMOS downscaling does not necessarily originate a chip area reduction at the same rate as the technology scaling factor. This happens primarily due to the analog part since the digital one scales well with technology evolution. In addition to the transistor scaling issues, the use of passive components also limits the degree of analog circuit area scaling.

Passive on-chip components set includes capacitors, resistors, and inductors. From these three passive elements, the capacitor is the element that better fits to CMOS scaling, [35, 36], since its size can be reduced without affecting the associated quality factor, Q,

$$Q_{C,MIM} = \frac{1}{\omega C_{MIM} R} = \frac{t_{ox}}{\omega \varepsilon \rho K\left(t_m\right)} . \tag{2.20}$$

In (2.20), t_{ox} is the dielectric thickness, ε is the dielectric constant, ρ is metal resistivity and $K(t_m)$ is related with the contact resistance to the metal and the metal thickness. It has been considered that the capacitance value of an metal-insulator-metal (MIM) capacitor is approximately given by

$$C_{MIM} = \frac{\varepsilon\, A}{t_{ox}} , \tag{2.21}$$

where A is the capacitor area. Additionally, the equivalent series resistance is dominated by the metal resistance, which is given by

$$R = \frac{\rho K\left(t_m\right)}{A}. \tag{2.22}$$

Planar on-chip inductors are implemented using metal lines shaped in a spiral format. The inductance value of such elements depends on the number of spiral turns and dimensions [37],

$$L = \mu_0 K_1\, n^2 \left(\frac{d_{avg}}{1 + K_2 \rho}\right) \tag{2.23}$$

where μ_0 represents the magnetic permeability, K_1 and K_2 are coefficients determined by the geometry and n is the number of turns. In 2.23, d_{avg} refers to the average diameter of the inductor,

$$d_{avg} = \frac{d_{outer} + d_{inner}}{2} \tag{2.24}$$

and ρ is a fill factor given by

$$\rho = \frac{d_{outer} - d_{inner}}{d_{outer} + d_{inner}} \tag{2.25}$$

where d_{outer} and d_{inner} are the outer and inner diameter of the inductor. The inductor model described by (2.23) shows that the reduction of each dimension of the planar inductor by the same scaling factor will result in a lower inductance value. Simultaneously, if the inductor geometry is maintained, the equivalent series resistance will increase and therefore contributing to the quality factor degradation, [35]. In conclusion, the chip-area used by on-chip inductors does not scale down with the CMOS technology evolution. Despite the research on innovative 3D geometry inductors, an alternative RF circuit design approach consists on using inductor-less circuit topologies in order to save chip area. However, innovative noise canceling techniques might be necessary [38].

2.4 Architecture Considerations for RF Transceivers in Submicron CMOS

As a consequence of CMOS technology scaling, the time-domain resolution of a digital signal edge transition has been increasing in comparison to voltage resolution of analog signals which is getting lower (due to power supply reduction). Since the received signal is essentially analog and continuous-time, this means that somewhere inside the transceiver a sampling process has to be done and this action is getting faster and more accurate in time. The distinction between digital transceiver topologies remains on where the sampling process is made and at what speed. After this step, the acquired sample is then digitized in amplitude by the ADC to be digitally processed at baseband.

In the ideal digital receiver, the ADC would connect directly to the antenna, see Figure 2.7. This architecture would be the most suitable for full CMOS integration since the entire spectrum range is directly sampled and digitized by the high-resolution ADC and all processing is made in the digital processing unit. By pushing all the processing into digital domain, not only the receiver adapts better to the multi-standard radio environment but also allows to introduce new services and features with just software updates.

Despite the benefits added by the ideal digital receiver, the requirements needed for the ADC are not practicable with the available technology. In fact, to receive signals up to 6 GHz (to cover major wireless radio standards) the required ADC, clocked at 12 GHz, would need more than 16 bit of resolution, due to the stringent signal-to-noise ratio

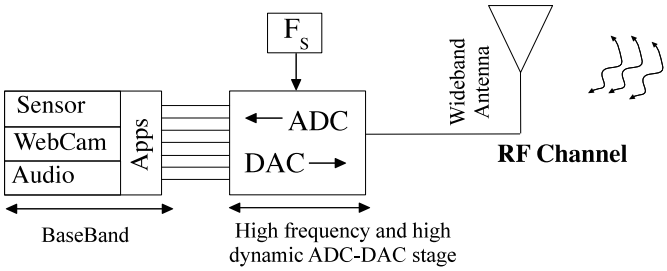


Figure 2.7: Ideal digital receiver.

(SNR) specifications. Additionally, for a capacitor based ADC, the dynamic energy consumed by this building block would be very high, and therefore, not suitable for mobile applications.

Since the useful information is not spread over all full spectrum input band, the required ADC specifications can be relaxed if some analog signal processing is made prior to the ADC input. To reduce dynamic range and limit the bandwidth, different receiver architectures approaches have been used but only a few of them are truly suitable for CMOS implementation. Yet, the importance of reaching a flexible transceiver that can dynamically adapt in a multi-standard radio environment is of paramount importance in modern and future wireless communications equipment. To accomplish this, the ideal digital receiver has been adapted towards a Universal Software-Defined Radio (USDR) where the analog building blocks have digitally programmable capabilities.

At some extent, the RF transceiver architecture involves the transfer of a high frequency spectral component down to a lower frequency range and simultaneously amplifying it. In practice, this down-conversion step is accomplished by means of multiplying mixer or by direct sampling or sub-sampling techniques.

2.4.1 SuperHeterodyne Receiver

The original superheterodyne receiver, known since 1918 [30], has been adapted to the modern digital baseband unit, Figure 2.8. This architecture is still used due to its high sensitivity and selectivity as long as a good image rejection is achieved. In this topology, before the Low Noise Amplifier (LNA), an RF bandpass filter selects the wanted RF band, which is also used as an image rejection filter for the first mixing stage. By its turn, this last block mixes the amplified signal from the LNA with the signal generated by a first local oscillator. The resulting downconverted signal, with its main component at an intermediate frequency f_{IF}, is then feed into a channel selection

bandpass filter. To obtain the signal at the baseband, a second mixing stage is used which is driven by a second local oscillator. Since in most of the cases an In-phase (I) and Quadrature-phase (Q) modulation scheme is used, the second mixing stage needs two branches to recover both components. The number of downconversions stages is the result of a trade-off between image rejection and channel-selection requirements, being the most common the dual stage implementation represented in Figure 2.8. The

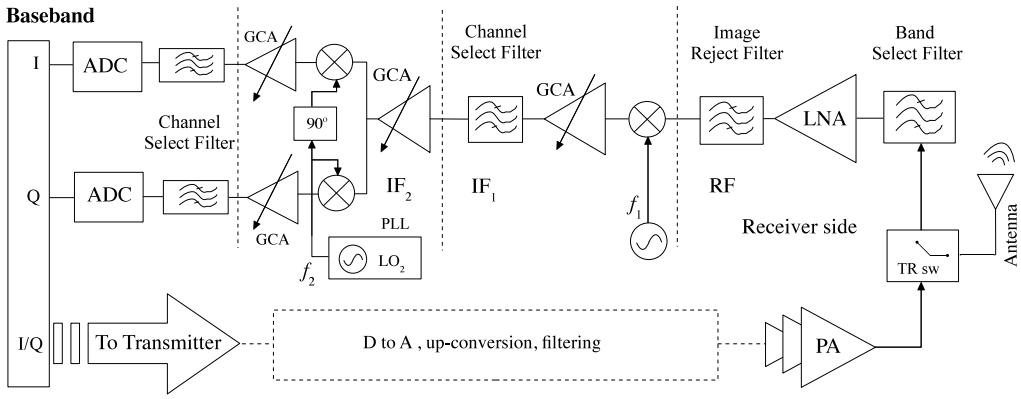


Figure 2.8: Heterodyne receiver, [30].

superheterodyne architecture needs high-Q RF and intermediate frequency band pass filters which are usually implemented with off-chip high-Q filters (such SAW filters). Since the design and chip integration of such high-Q filters are difficult to achieve, due to the low-Q and high area inductor available in a standard digital CMOS process, alternative implementations has been followed in the past years.

2.4.2 Zero-IF and Low-IF receivers

To avoid the integration of high-Q and high-frequency filters, the transceiver architecture can be modified in order to transfer them to a lower frequency, coincident or close enough to baseband, where these inductors based filters can be more easily replaced by active ones. Examples of these blocks ranges from continuous-time gm-C to discrete-time implementations such as switched-capacitor based filter or switched-current, (SI), , which are compatible to digital standard CMOS processes.

In one basic configuration, based on Figure 2.9, this direct conversion receiver type (DCR) converts directly the RF modulated signal to baseband where it is applied to the ADC through two paths: In-phase and Quadrature-phase. Typical digital modulation usually form two sidebands around the RF carrier, both containing relevant information. To recover the original stream at the baseband, it is necessary to use a quadrature

down-conversion to obtain I and Q components, otherwise loss of information would occur [39].

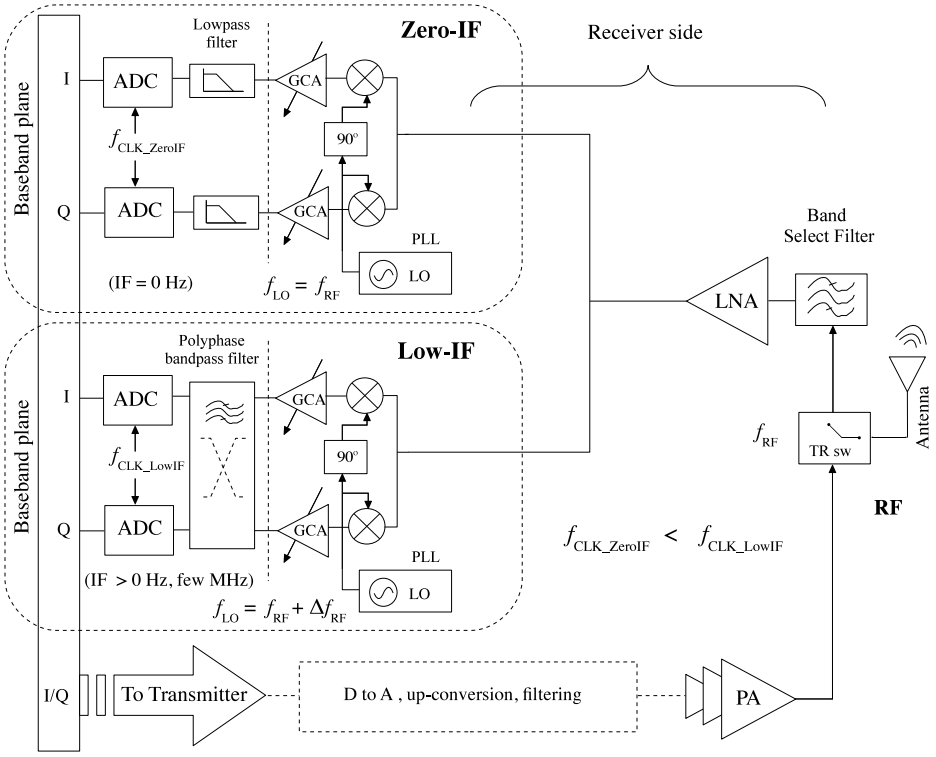


Figure 2.9: Low-IF or Zero-IF receiver.

Despite being one of the most suitable architecture for submicron CMOS, due to its simplicity and nonexistence of high-Q and high-frequency filters, several drawbacks have to be addressed before reaching a good design. The first one is related with the mismatch between I and Q paths, originated by gain errors and phase inbalance of the Local Oscillator (LO) quadrature outputs [30]. These effects degrade the SNR due to increase in distortion, but it is not as serious as in the image-reject architecture, [30]. Even-order intermodulation effects are usually neglected in typical RF digital heterodyne receivers, but can significantly degrade the zero-IF architecture performance. This is because second order products from the intermodulation of two strong interferers close to the channel, can corrupt the baseband signal of interest due to feedtrough effects in the mixer. Additional corruptions in the baseband can occur as the result of DC offsets and flicker noise, since both components fall within the same band as the desired signal. DC offsets appears in the zero-IF receivers due to LO leakage and self-mixing in the conversion mixer stage, [14]. These DC offsets can also saturate the following baseband stages (e.g., ADC).

The Zero-IF receiver introduce less gain before reaching the baseband stage, when compared to the heterodyne one (which distributes the gain across the RF and inter-

mediate stages). This increase significantly the negative impact of baseband flicker noise added by MOS transistors. As seen in section 2.3, flicker noise power spectral density is inversely proportional to gate area, suggesting that more attention has to be taken when sizing the front-end transistors of the baseband blocks in deeper submicron CMOS.

Although the performance degradation due to baseband flicker noise is reduced by using wider bandwidths (which tends to be more common in high data rate applications), the application of Zero-IF topology for narrowband case suffers from serious SNR degradations problems. One alternative to bypass the flicker noise $1/f$ corner is to reuse the heterodyne concept while maintaining the simplicity of Zero-IF architecture. As a result, a Low-IF structure is derived where the second conversion falls very near baseband, thus avoiding DC offset, flicker noise and LO leakages issues. Comparatively to the Zero-IF, the Low-IF receiver is more sensible to image rejection and I/Q mismatch. Band pass filtering is used after the I/Q mixers, instead of low pass one in Zero-IF, for channel selection requirements, but their operation can be combined to achieve, simultaneously, interferers cleaning and image rejection. A complex polyphase filter can accomplish this task. Moreover, since the complex polyphase filter can be implemented in a digital processing unit, instead of analog implementation, then the same Zero-IF architecture can be used, where only the LO frequency need to be adjusted accordingly, [14].

2.4.3 Universal Software Defined Radio (USDR)

The need to cope with the demanding of multiple radio standards has been a major research driving force towards a flexible, radio adaptable and software controlled transceiver architecture. This implicitly implies a wideband coverage for the RF front-end and also that all building blocks should be digitally controlled.

The original proposal for a software radio [40], in which the ADC is directly connected to the antenna, impose very demanding requirements for this block that are still hard to achieve in present technologies. Moreover, the performance of such ADC would be limited by the dynamic range, power of the interferers as well as it would consume significant amount of energy. As an example, a 12 bit ADC running at 10 GHz [41] would dissipate several Watt. This means that some degree of frequency down-conversion has to be done before digitalization occurs. This can be done at baseband, low-frequency or at IF stage.

One consensual type of architecture to implement a USDR, is depicted in Figure 2.10,

which resembles to a Zero-IF topology. However the differences are significant since all the blocks are digitally controlled and the RF front-end should operate over a wide frequency range. This is specially true for the LNA and for the mixer. Moreover, depending on the ADC performance and reconfigurability, this structure can implement a Zero-IF, Low-IF and even a heterodyne receiver. In this last configuration the second down-conversion is performed in the digital domain. The switch between modes and radio standards is performed by software means. The frequency scanning and spectrum

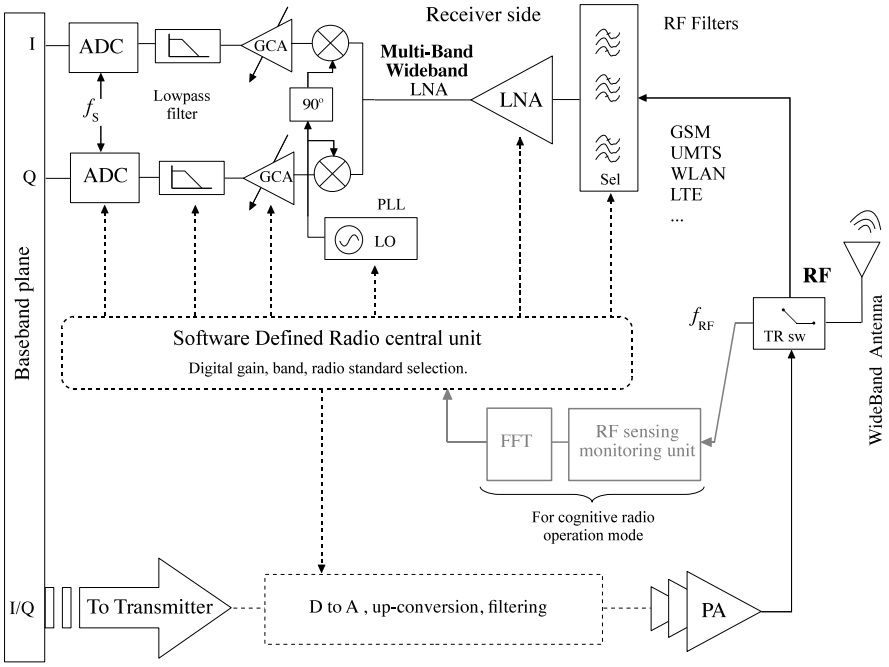


Figure 2.10: Universal software defined radio.

sensing block included the block diagram is used to cognitive radio techniques [42] which is a promising technique to increase spectrum allocation efficiency. The basic idea, when in cognitive radio mode, the transceiver dynamically choose free spectrum regions were it can transmit and received with minimum of interference. This process goes periodically and therefore contributing for radio interferer avoidance.

The USDR is a very suitable and flexible architecture but relies greatly on high digital processing power, only available in submicron and nanoscale CMOS.

2.4.4 Receiver with (sub)sampling

Designing a digital receiver, means that a quantization and sampling processes have to be done at some point along the signal path. This task, which is performed jointly by a sample-and-hold (S/H) and an ADC, can be done at RF, IF or at baseband. The rate

that samples are acquired determines the type of sampling and the following digital processing. Moreover, the Nyquist theorem defines the relation between this sampling rate, f_s, and the bandwidth, B, of the signal to be sampled and B_{RF} the full RF band. Sampling processes can be

- at Nyquist rate, when $f_s = 2B_{RF}$;
- oversampled, when $f_s > 2B_{RF}$;
- undersampled, when $f_s < 2B_{RF}$. In this case aliasing will occur, but under special conditions the signal information is not corrupted.

It has been implicitly assumed, till now, that sampling is at the Nyquist rate. However, the other two options can also be used in a CMOS digital transceiver. Using an oversampling technique within a negative feedback loop (resulting in a Delta-Sigma conversion approach), the noise can be shaped away from the baseband producing a better SNR, and therefore, relaxing filtering requirements.

Most of the previous data communications considerations falls into narrowband scenario, meaning that only a small part of the RF-band is of interest and therefore undersampling (subsampling) can be used. In order words, the Nyquist criterion has to be fulfilled for the channel bandwidth rather to the entire spectrum. These subsampling receivers, see Figure 2.11, differentiates from the previously described ones by the fact that no mixer is used to downconvert the RF/IF signal to baseband. Instead the RF/IF signal is directly sampled (at a lower rate) and downconverted to baseband. Besides the application of Nyquist criteria with respect to the signal bandwidth B,

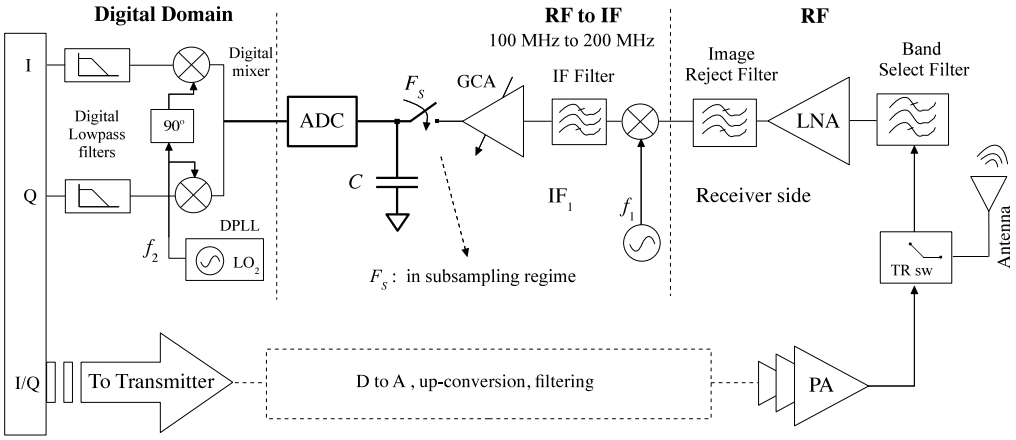


Figure 2.11: IF subsampling receiver architecture.

the following conditions should be verified to avoid the overlap of images generated by

passband uniform sampling [43]

$$\frac{2 \cdot f_H}{n} \leq f_s \leq \frac{2 \cdot f_L}{n-1} \tag{2.26}$$

where n is the integer given by

$$1 \leq n \leq \frac{f_H}{f_H - f_L} = \frac{f_H}{B}\,, \tag{2.27}$$

where f_H and f_L represent, respectively, the upper and lower band frequency. As previously stated, under these conditions destructive aliasing is avoided and a low frequency replica appears after subsampling. Therefore, since a frequency downconversion is performed, the baseband signal can be recovered through an appropriate low pass filter.

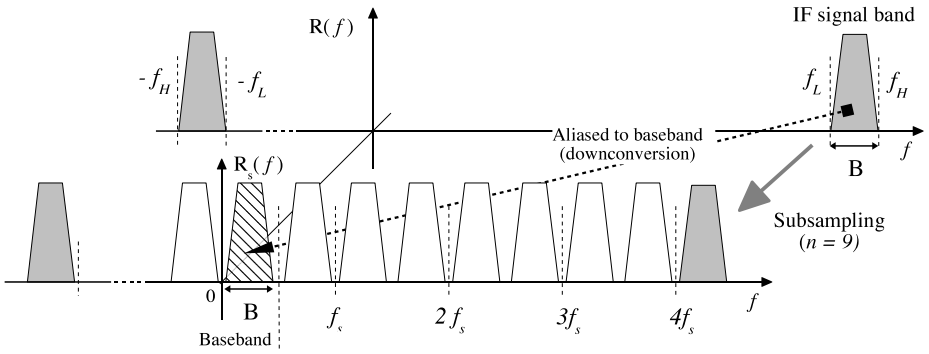


Figure 2.12: Spectrum transformation in a subsampling receiver.

Uniform bandpass subsampling degrades the SNR ratio due to, essentially, two main reasons. The first one is related to the out-of-band noise which is folded back (aliased) into baseband thus increasing its power by a factor of $2 \cdot M$ (where M represents the subsampling factor, i.e., $M = [(f_H - f_L)/2]/f_s$), [44]. The kT/C noise added by the sampler, during the subsampling process, is one of the major noise sources.

The second reason is related to timing jitter associated with the sampling clock phase noise. An increase of M^2 is expected in its power spectral density after subsampling [44]. The reduction of those noise levels is achieved not only by applying high-Q filtering before sampling (to limit out-of-band noise) but also by selecting low-phase noise oscillator (used in the sampling clock). Nevertheless, assuming that the receiver is dominated by these noise sources, the overall performance is strongly dependent on the subsampling factor, M. This ratio can be particularly high in a narrowband scenario. But, interestingly this is not the case for UWB receiver where M is small. Moreover, since UWB signals are characterized by bandwidths higher than 500 MHz, the total

noise is mostly dominated by the in-band component. Therefore not only lower Q filter are needed but also the sampling jitter constraint is reduced by the lower ADC dynamic range [19]. As consequence, pushing subsampling of UWB signals directly within the RF stage is an option.

Following the UWB approach, one can reach a wireless transceiver with substantially less hardware than a traditional narrowband case, as can be observed in Figure 2.13. An

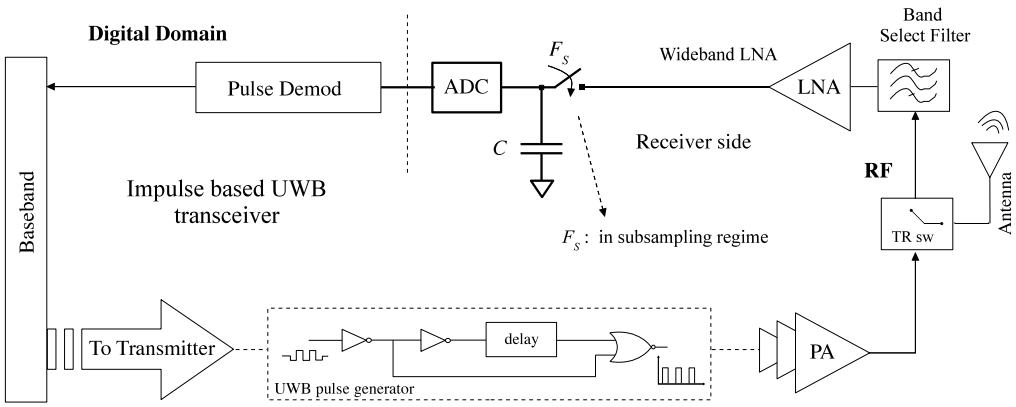


Figure 2.13: UWB receiver with direct subsampling.

impulse based UWB emitter can be easily implemented with a few digital gates driving the antenna. The latter defines the radiated pulse shape. In other hand, in an impulse radio UWB non-coherent communication case, the data stream can be recovered at the receiver by using a simplified energy detector circuit followed by a high-speed low-resolution ADC. Both facts contributes to a simpler transceiver structure which is very well adapted to CMOS technology, specially in the submicron range where short pulses (below 2 ns) are more easily generated. These types of non-coherent receivers are limited to low to moderate data rate but this is not a serious limitation in wireless sensor networks where ultra low cost and ultra averaged low power consumption are more important issues.

2.4.5 Energy Detector based Receiver

Simpler but less performant receiver can be designed by removing the local oscillator from the circuit and replacing it with a self-mixing technique. The solution is well known for decades [30], and consists on using a signal squarer followed by an integrator, which can be viewed has an envelope detector/demodulator, as represented in Figure 2.14. This envelope detector block relies on the non linear nature of a semiconductor device, being the diode the traditional option. However, a MOS transistor biased in

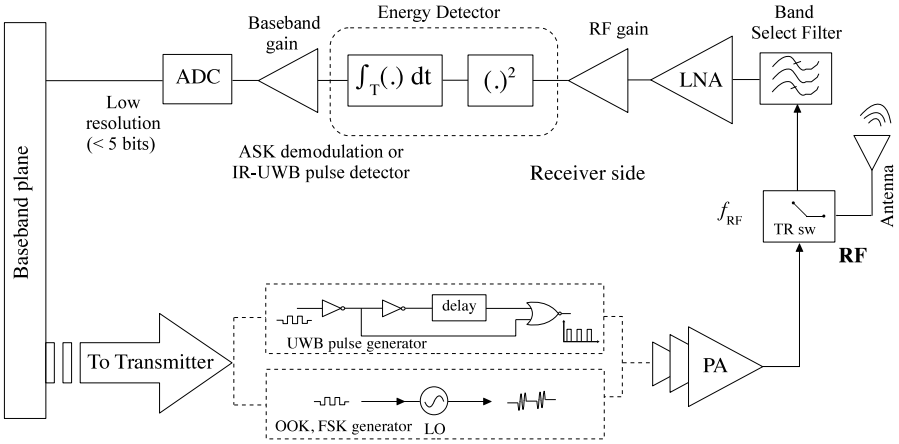


Figure 2.14: Envelope detection receiver architecture.

saturation (specially in weak inversion) as well as in triode region can also and easily implement this non linear function, which has already being proposed, [25]. The self-mixing operation is insensitive to phase and frequency variations, which restrict the applicability domain to AM modulation or to the digital counterpart the ASK, or simply OOK. Binary FSK (and MFSK), can also be adapted to this structure but requires frequency selectivity that has to be provided through narrowband filtering directly at RF, which is a major drawback. Nevertheless, this is still an attractive approach for full CMOS integration for low-data rate transceiver applications and mode recently for the implementation of simple, low-cost, and non-coherent impulse radio UWB based transceivers, [14].

2.5 Emerging circuit design strategies for CMOS nanoscale wireless transceivers

2.5.1 Digital assisted analog approach

With the evolution of CMOS technology, it has been verified that the power reduction on the digital part reduces at a faster rate than the analog counterpart, [45]. Therefore, a research path has been conducted in the direction of transferring some of the analog function to the digital domain, complementing mechanisms of pre and post digital processing in order to assist the analog part. This transfer can be optimized in order to reduce the total power. To illustrate this issue, the energy budget that is involved with an ADC of 10-bit that consumes 0.25 mW *per* MHz, spend around 0.26 nJ *per* conversion. Interestingly, with this energy, it is possible to toggle 100 K logic gates in

90-nm CMOS technology.

2.5.2 Open Loop and reactance based amplifiers

Although depending on the wireless system standard, it is usually requested for the receiver to be able to process very low signal from the antenna (as low a few μV). Being the first amplifying stage, the voltage gain achieved by the LNA is usually not enough to correctly adjust the signal level to the ADC input. Therefore, the total receiver gain is split over the signal path. Since amplification is harder to achieve at RF frequencies, the remain analog amplification is performed either at IF and/or at baseband. Additionally, RF systems are usually characterized by large variations on the strength of signal detected by the receiving antenna. This variability depends on the radio channel conditions, e.g., the physical distance between the two radio peers. To adapt this input signal dynamic range to the ADC input, the receiver design has to include programmable signal gain amplifiers (PGA) in the main signal path. Once again, it is easier to implement a PGA at lower frequencies.

The classical approach for the PGA design, uses a high gain opamp connected in a closed loop topology where the gain is programed, for example, by an array of capacitors, [46]. Furthermore, the opamp is also extensively used in discrete time signal processing blocks, such as SC filters or ADCs. As CMOS scale down large opamp gain is getting difficult to achieve at low supply voltage without affecting bandwidth and/or power dissipation, [14]. This performance degradation reduces the accuracy of charge transfers in traditional SC circuits.

Instead of increasing the number of stages inside the opamp to reach higher gain under low power supply conditions, alternative solutions and circuit techniques have been recently proposed and applied, for example, in ADCs. All of them start from the principle of using simpler circuits. Examples include low opamp gain, open-loop amplification [47], comparator based technique [7] and dynamic source-follower based amplifier [48, 49]. The last one is based on the circuit shown in Figure 2.15 with which a more efficient charge processing is accomplished when compared to an high power opamp based solution and with less thermal noise injection. The discrete time operation of this circuit is divided in two non-overlapping phases. During the sampling phase, the charge corresponding to the input voltage sample is accumulated in the transistor, which is connected in a MOS capacitor configuration.

During the amplification phase the gate is left floating, the drain is connected to V_{DD} and the transistor, acting as a source follower, drives the previously discharged

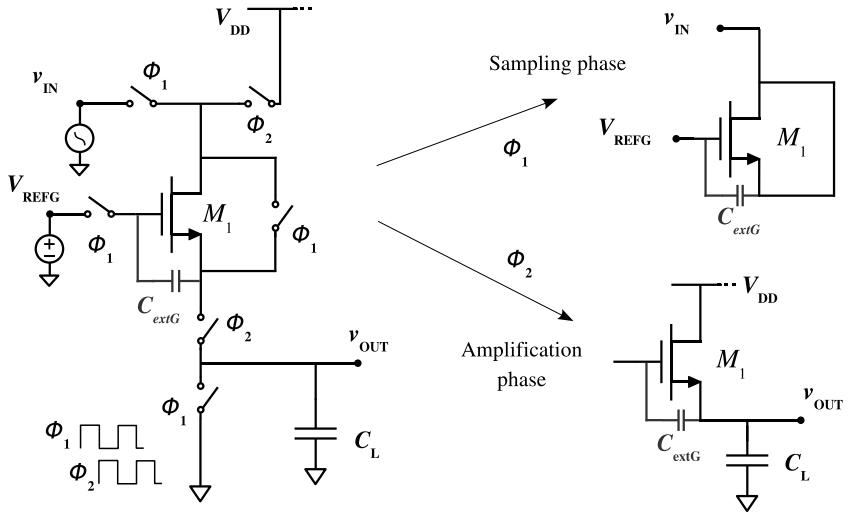


Figure 2.15: Dynamic Source Follower Amplifier.

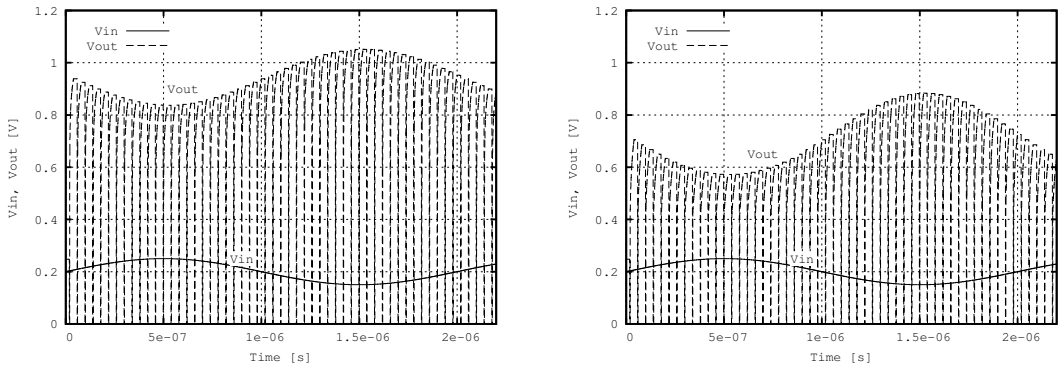
capacitive load. Due to charge conservation at the transistor gate, the signal dependent charges move entirely to the gate-drain capacitance since the gate to source voltage settles to a value close to the threshold voltage, $V_{gs} = V_{th}$. From a small signal analysis, a capacitance reduction is achieved which results in a signal voltage amplification.

Attending that the gate is left floating during the amplification phase, the charge redistribution that occurs between the gate parasitic capacitances respects the charge conservation principle, thus resulting an input to output voltage relation given by

$$v_{OUT} = \underbrace{-\frac{C_{Gtotal,\phi1}}{C_{gd,\phi2}}}_{\text{signal gain}} \cdot v_{IN} + \underbrace{V_{DD} + V_{REFG}\frac{C_{Gtotal,\phi1}}{C_{gd,\phi2}} - V_{th}\frac{C_{Gtotal,\phi2}}{C_{gd,\phi2}}}_{\text{signal independent component}}, \tag{2.28}$$

where the total gate capacitance during sampling phase $\phi1$, $C_{Gtotal,\phi1}$, is approximately given by the sum of the overlap capacitances from the gate to the drain and source, plus the gate to bulk capacitance which is equal to C_{ox} in series with the depletion layer capacitance (the device is maintained in depletion region during $\phi1$ by the DC bias V_{REFG}). Assuming almost complete settling in the amplification phase $\phi2$, the gate to drain capacitance, $C_{gd,\phi2}$, is also approximately equal to the overlap parasitic capacitance C_{ov}. Considering that the $C_{gd,\phi1}$ is small, since the transistor is depleted, this structure can achieve a gain of 2. Nevertheless, second-order and body effects lower the predicted gain given by (2.28). A way to achieve an extra gain adjustment function is to add an extra capacitance between the gate and the source, which increase the total gate capacitance during sampling phase but has reduced impact during the amplification phase since V_{gs} settles towards to the threshold voltage V_{th}.

Figure 2.16 shows the simulated input and output voltage waveforms using a standard 1.2 V 130 nm CMOS digital technology and BSIM3V3 models, being the NMOS device sized with aspect ratio 4/0.13 μm. The results shows that without the extra capacitor the circuit reaches a gain around 2 but can have an improvement of 50 % just by adding a 3 fF extra capacitor. Considering the body effect, by connecting the bulk to ground,



(a) Without C_{extG} (b) With $C_{extG} = 3$ fF

Figure 2.16: Voltage waveforms at the input and output of the Dynamic Source Follower Amplifier, without body effect ($V_{sb} = 0$).

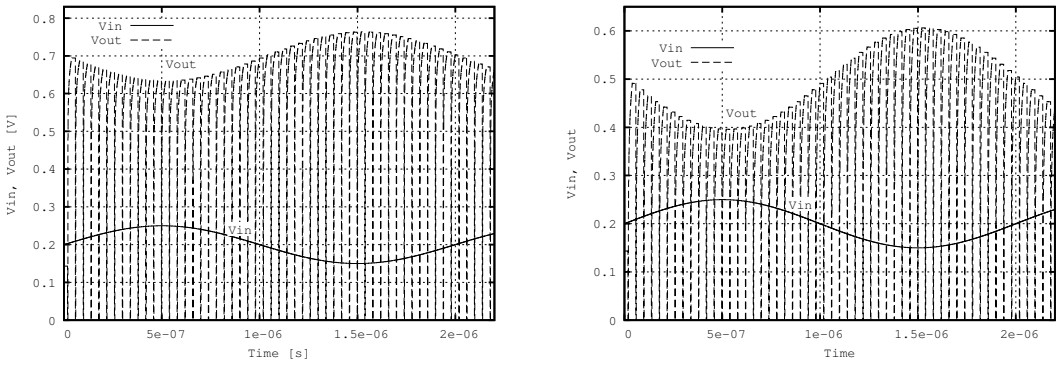
the circuit is significantly affected as it can be seen in Figure 2.17, where the gain has reduced to 1.4 without C_{extG} and 2.3 for $C_{extG} = 3$ fF.

On the main advantages of this class of amplifier circuits is that they do not depend on cascode structures nor on high intrinsic transistor gain g_m/g_{ds}, making it very adaptable to technology scaling. Following this track of open loop and MOS based discrete time amplifiers, an alternative approach supported on reactance amplification and less sensitive to the body effect, is presented and used in the different building blocks of a pipeline ADC in the following Chapters.

2.6 Summary

In this chapter a global overview was given about the wireless data communications systems, not only at the network level but also at the transceiver architecture level. To bound and contextualize the target technological implementation, a review of the current trends of the CMOS evolution has been reflected on the analysis.

The dominant CMOS technology is entering at deep nanoscale age but, as presented in this Chapter, the design of analog circuits are getting more difficult due to the lower power supply, growing parasitic effects due to the reduction of the oxide. Not only



(a) Without C_{extG}

(b) With $C_{extG} = 3$ fF

Figure 2.17: Voltage waveforms at the input and output of the Dynamic Source Follower Amplifier, with body effect ($V_{sb} \neq 0$).

the analog technique has to evolve, but also the research of new system level design technique is being developed. In complement to this Chapter, Appendix A presents one simple co-design approach for the RF front-end receiver.

At the receiver level software radio approach along with digital assisted analog circuits, is a promising technique since it relax the requirements of the analog building blocks.

Chapter 3

Parametric Signal Amplification in Continuous Time Domain

3.1 Introduction

Preceding the availability of high frequency, high gain and sufficiently low noise semiconductor transistors, the need to increase the radio receiver sensitivity as led to a concerted work among several research teams during the mid-fifties, [50]. One of the objectives was to reduce the impact of the signal loss and noise added at the mixing stages. To achieve this goal, LNAs were developed, designed and inserted between the antenna and the mixer stage of a conventional receiver architecture. As a result, the equivalent receiver noise temperature, during that period, has reduced from 3000 K to 1000 K when using traveling-wave tube, or to 10 K in the case of a MASER, [50, 51].

An alternative technique, which gained great popularity at that time, consisted on the implementation of parametric amplifiers (PAMP). The parametric amplifier operates through a nonlinear process of mixing and frequency conversion in a nonlinear reactance. Despites some ferrite based (inductor like) amplifiers were presented, the most used mixing reactance was the nonlinear variable capacitor (varactor). Since the process of amplification is based on the capacitance parameter change over time, rather than a transresistive/transconductance approach, it was expected to achieve very low noise level during the amplification step.

Nowadays, there are not many cases exemplifying the CMOS implementation of such parametric based circuits, despites the fact it is a well suited technology to implement varactors. This is justified by the continuous evolution of the MOS transistor, which

has boosted the design of transconductance based signal amplifiers, which are fully and easily integrable on a chip. However, with the degradation of the transistor intrinsic gain due to the continuous CMOS scaling, described in Chapter 2, the parametric amplifier alternative technique is recovering some interest, particularly in the microwave domain.

This Chapter is dedicated to the review of the continuous time parametric signal processing applied to signal amplification and frequency conversion. Also a short reference is done on the application of this technique in traveling wave amplifiers.

Targeting design of parametric amplifiers in CMOS technology, an introduction of the available MOS varactor configurations is given. After this preliminary study on the continuous time parametric amplification, Chapter 4 extends the analysis towards the discrete-time version of this technique.

3.2 Using a reactance to build a transistor-free amplifier

In a transistor based amplifier, the flow of additional energy delivered to the load, which is needed to achieve a power gain higher than one, is supplied from a DC power source. In simplistic terms, the input signal modulates the injection of energy coming from the DC power supply towards the load by means of a transconductance operation. This feature of the transistor is usually (at least in the MOS case) dependent on the DC biasing current and thus impacting both, efficiency and power dissipation. Additionally, if one recognizes that a resistive and, therefore, a dissipative property is intrinsically embedded into this amplification process, a non negligible thermal noise component has to be taken into account.

Interestingly, the delivery of energy to the load does not need to be performed directly from the DC power supply but, instead, can be transfered from an AC power source. Instead of a resistive (conductance) and dissipative element, a non-linear energy storage element can be used to perform this energy transfer. One advantage that seems to emerge in first place is the fact that the amplifier output thermal noise can be significantly reduced by using a low-loss reactance (in theory, a noise free amplification process could be achieved with an ideal lossless reactance). The amplifier built around a reactance element is usually referred as a parametric amplifier (PAMP) since its operation relies on the controlled time variation of a reactance parameter. When using a nonlinear capacitor the variable parameter is the capacitance value (the same applies

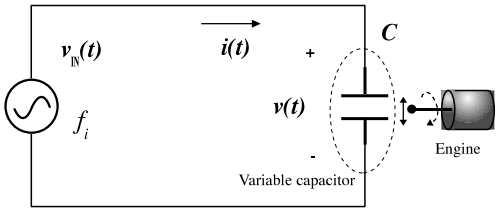


Figure 3.1: Simple model for the parametric capacitor.

for the inductance property of an inductor).

To achieve power gain from this parametric type network one must follow the condition set in [52] or [53], which states that a necessary condition to achieve power amplification from a nonlinear network, whose individual elements are passive, is the use of one nonlinear reactive device and one AC power source. A classical approach to explain the operation of a parametric amplifier is to analyze the energy balance of a capacitor excited by an electrical time varying signal and simultaneously by a mechanical periodic 'pump' force, as shown in Figure 3.1. The 'pump' force is applied to change the distance, y, between the capacitor plates (thus changing the capacitance value). Applying the analysis from [54], when in presence of a time variation on the capacitance, $C(t)$, it can be shown that the instantaneous charge stored in the capacitor $q(t)$, represented in Figure 3.1, is given by

$$q\left(t\right) = C\left(t\right) \cdot v\left(t\right) \,, \tag{3.1}$$

where $v(t)$ is the total voltage applied to the plates of the capacitor. Besides the usual capacitive current generated by $v(t)$, the total current, $i(t)$, through the capacitor also reflects the variation of the capacitance created by the mechanical operation of the virtual/imaginary engine since

$$i\left(t\right) = \frac{dq\left(t\right)}{dt} \,. \tag{3.2}$$

As usual, the energy stored in the electrical field is given by

$$W_{cap}\left(q, C\right) = \frac{1}{2} \cdot \frac{q^2}{C} \,. \tag{3.3}$$

Under these conditions, the total energy stored in the capacitor can be determined by applying the derivate chain rule to Equation 3.3,

$$P_{cap} = \frac{dW_{cap}}{dt} = \left(\frac{\partial W_{cap}}{\partial q}\right)_C \frac{dq}{dt} + \left(\frac{\partial W_{cap}}{\partial C}\right)_q \frac{dC}{dt} \,, \tag{3.4}$$

resulting in

$$P_{cap} = \frac{q}{C} \cdot i + \left(\frac{\partial W_{cap}}{\partial C}\right)_q \frac{dC}{dy} \cdot \frac{dy}{dt} = v \cdot i + \underbrace{\left(\frac{\partial W_{cap}}{\partial y}\right)_q}_{\text{force}} \underbrace{\frac{dy}{dt}}_{\text{velocity}} . \tag{3.5}$$

A deeper analysis of (3.5) reveals that the first part of the equation represents the electrical power resulting from $v(t)$ while the renaming one reflects the mechanical power transfered from the external imaginary engine,

$$P_{cap} = \underbrace{P_{e,v}}_{\text{due to } v(t)} + \underbrace{P_m}_{\text{mechanical}} . \tag{3.6}$$

Furthermore, this result indicates that the capacitor can be used as an energy converter between different domains or, as it will be shown in Section 3.4, between signals with different frequency. This is in fact the basis of the parametric operation, where the capacitor is used as an active element.

To transpose the initial schematic into a more practical and fully electrical circuit, one must find a way to electrically change the capacitance value instead of doing it by mechanical means. Such operation can be reached through the use of a modified capacitor that reacts non-linearly with respect to the applied voltage. The resulting mixing effect that occurs in the nonlinear capacitor permits an energy transfer and a frequency combination between a weak signal (e.g., coming from an RF antenna) and a relatively strong pump source (e.g., a local oscillator). Both signals are superimposed across the nonlinear capacitor, see Figure 3.2. The mixing that occurs between the input signal $v_i(t)$ with frequency f_i and the pump source $v_p(t)$ with frequency f_p generates a set of harmonics with frequencies located at $f_0 = m\, f_p \pm n\, f_i$. Usually only one of them is somehow extracted from the circuit. The following analysis assumes that the output signal corresponds to the harmonic with frequency $f_0 = f_p \pm f_i$. A generalized result is presented in Section 3.4.

When using a nonlinear capacitor, the charge $q(v)$ is also a nonlinear function of the applied voltage $v(t)$ and therefore may be expanded in a Taylor series given by

$$q(v) = \alpha_1 v + \alpha_2 v^2 + \alpha_3 v^3 + \cdots . \tag{3.7}$$

Considering only the first two terms of Equation 3.7, as long as the voltage across the capacitor is not too large [55], the current flowing in the nonlinear capacitance is

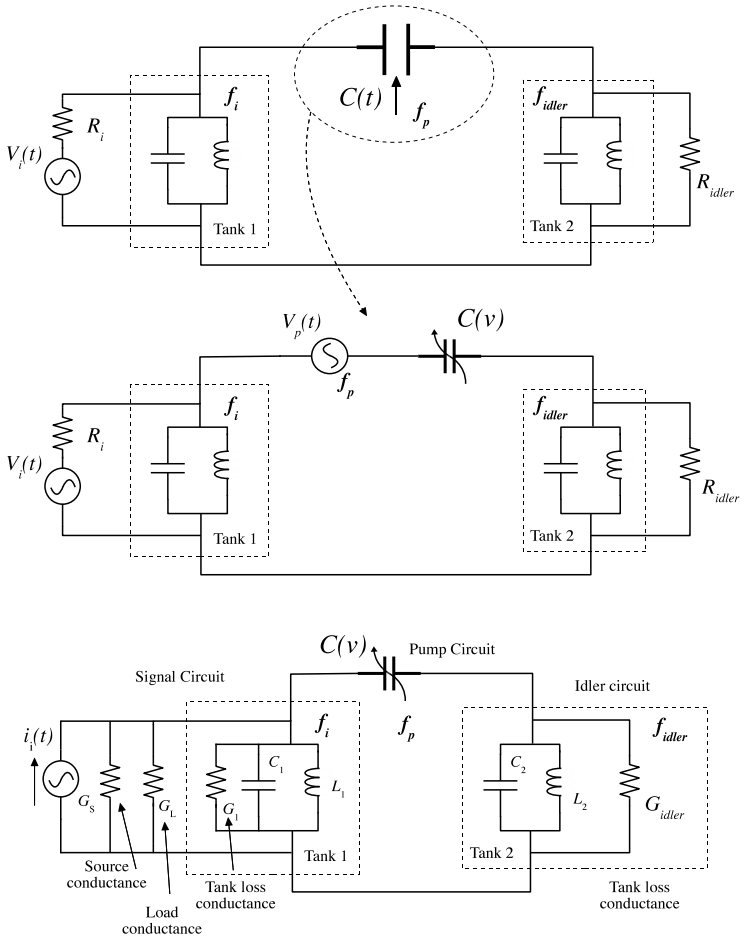


Figure 3.2: Equivalent model for a two-tank parametric amplifier.

obtained from

$$i(t) = \frac{dq\,(t)}{dt} = \underbrace{\alpha_1}_{C_0} \frac{dv\,(t)}{dt} + \underbrace{2\alpha_2 v\,(t)}_{C_v(t)} \frac{dv\,(t)}{dt} = (C_0 + C_v(t))\,\frac{dv(t)}{dt}\,. \tag{3.8}$$

The result from (3.8) confirms that, if properly driven by a pump signal, the nonlinear capacitor behaves like a time-varying linear capacitance thus validating its use in a parametric circuit configuration.

Figure 3.2 represents a possible configuration for a parametric amplifier based on a nonlinear capacitor C whose capacitance value is changed by a pump signal of frequency f_p. Associated with the input signal with frequency f_i is a circuit tank consisting of L_1, C_1 and $G_{t1} = G_S + G_L + G_1$ represents the total loading due to source, load resistance and internal tank losses. Similarly, the sideband frequency f_{idler} has an associated resonant tank circuit formed by L_2, C_2 and G_{idler}.

Depending on the relation between the three involved frequencies, the circuit has slightly different designations, [56]:

- Negative-Resistance Amplifier, if the output frequency is the same as the input one;
- Down-Converter Amplifier, if the output signal frequency (taken from the idler tank) is the difference between the pump and input signal frequencies, i.e., $f_{idler} = f_i - f_p$;
- Up-Converter Amplifier, if the output signal frequency (taken from the idler tank) is the sum between the pump and input signal frequencies, i.e., $f_{idler} = f_i + f_p$.

As indicated above, if the upper sideband is filtered out, and the output is taken at the input signal frequency, then a negative resistance parametric based amplifier is built. In this case, the output signal is physically extracted from the circuit at the input tank. The remaining part, formed by the second tank plus the nonlinear capacitor, behaves as a negative resistance which is responsible for the amplifying operation of this 'reflection' type of amplifier. From [55] and [56], the negative conductance is given by

$$G = \frac{\omega_i\,\omega_{idler}\,C^2}{4\,G_{idler}} \tag{3.9}$$

where $\omega_{idler} = 2\pi f_{idler} = \omega_p - \omega_i$, and the remaining parameters have already been defined. The power gain corresponds to the ratio of the power dissipated in the load conductance G_L to the available power from the input signal source with conductance

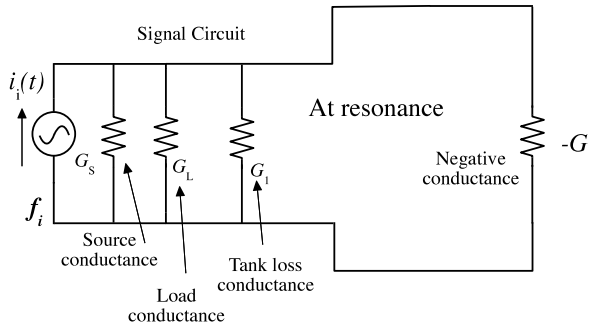


Figure 3.3: Equivalent model for a two-tank parametric amplifier at resonance.

G_S. Using the negative conductance calculated in (3.9), the power gain at resonance is determined by

$$G_{pwr} = \frac{4\,G_S\,G_L}{(G_{t1} - G)^2} \tag{3.10}$$

where G_S and G_L are the loading effect due to the input source and output load, respectively. Both are calculated at the frequency of the input signal. Additionally, (3.10) demonstrates that an oscillating regime is possible and occurs when $G = G_{t1}$.

The F of the negative resistance amplifier can be determined from the ratio between the output and input SNR. To calculate F at resonance state, the circuit model represented in Figure 3.3 can be used.

$$F = \frac{1}{G_{pwr}} \cdot \frac{1}{kTB} \cdot N_o = \frac{1}{4kTB} \cdot \frac{(G_{t1} - G)^2}{G_s G_L} \cdot N_o \tag{3.11}$$

In (3.11), B is the bandwidth, T is the noise temperature in Kelvin, k the Boltzman's constant, N_o is the total output referred noise, and the remaining parameters have already been defined. A detailed analysis is done in [56], in which all noise sources were added at the output, thus resulting in the closed expression given by

$$F \approx 1 + \frac{G_1}{G_S} + \frac{G}{G_S}\frac{f_i}{f_{idler}} \ . \tag{3.12}$$

Remarkably from (3.12), the lowest possible F obtainable is equal to $1 + f_i/f_{idler}$. Therefore, the use of an higher difference frequency $f_{idler} = f_p - f_i$ contributes to lower F.

Several configurations were used in the past to implement parametric amplifiers. Despite some reported amplifiers were based on ferrite/ferromagnetic (as the variable reactance), [57], the use of semiconductor solid-state diode was definitively the primary choice when designing a parametric amplifier at that time, [50]. In fact, in the former case the amplifier needed a large pump power while with the varactor diode case

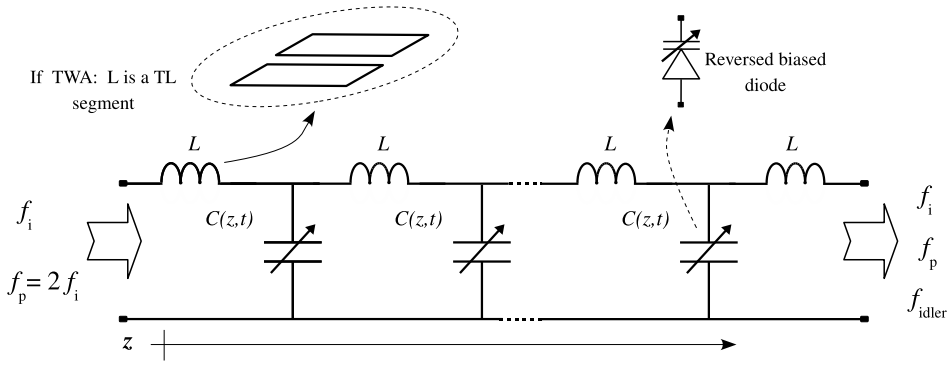


Figure 3.4: Model for a traveling wave parametric amplifier.

the requirements for the pump power are less stringent. The utilization of diode varactor also permitted the construction of more 'compact' and less costly amplifier that could achieve low F even at room temperature. The effective noise temperature of a parametric diode based amplifier could be as low as 30 or 50 K (corresponding to a 1.3 dB Noise Factor), [57]. It was used in several applications. (e.g., military radars) ranging from 200 MHz to 35 GHz operation frequency and achieving a single stage gain ranging from 5 dB to 20 dB, [51, 58, 59, 60].

Up to now, the described parametric amplifier uses only a single variable reactance element and one or more resonant circuits. With this configuration, it is possible to obtain a reasonable gain with small Noise Figure footprint. Nevertheless, the use of high Q tanks squeeze the amplifier bandwidth [61]. An alternative for the parametric amplifier construction utilizes a distributed approach, being the traveling-wave structure one of the most important, [62, 63, 64, 61, 65, 54]. In this type of structure, several varactors are distributed along a propagating wave circuit (e.g., transmission line), contributing each one of them with a small amount of signal gain. Amplification of the signal power is then obtained in the form of a growing wave propagating along this structure of shunt varactors separated by series of inductances or transmission line sections. A general distributed amplifier model is shown in Figure 3.4. When the lumped inductors of a distributed amplifier are replaced by transmission lines, the new circuit configuration is generally designated by Traveling Wave Amplifier (TWA), [66].

In Figure 3.4, the variable capacitors $C(z,t)$ are varactors (e.g., reversed biased diodes) whose capacitance is a function of position and time. The capacitance depends on its position since the amplifier acts as a terminated transmission line when considering the wavelength of the pump signal. The operation is supported on two propagating modes. One mode is excited by the input signal and, the other mode is used as the idling circuit. Additionally, the structure processes the traveling wave generated by

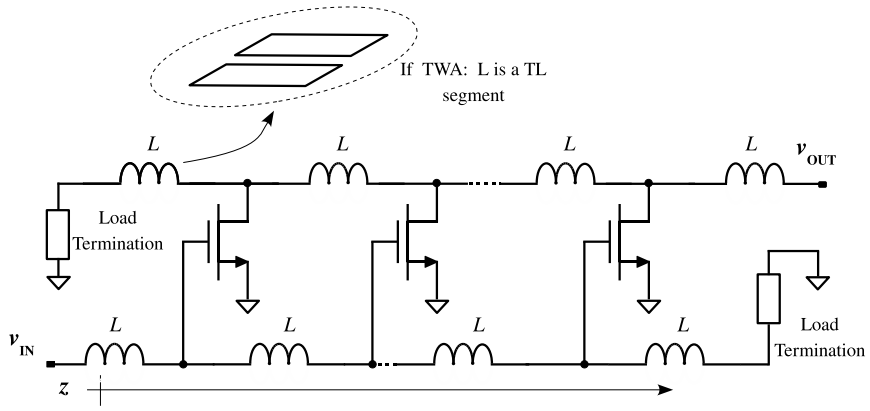


Figure 3.5: Simplified model of a MOSFET distributed amplifier.

the pump signal which provides, by modulating the capacitance of the varactors, a time varying coupling between the two propagating modes. This coupling behavior originates the necessary AC energy transfer for the parametric amplification to work. One of the first implementations of a TWA [64] using varactor diodes achieved more than 10 dB of gain, a noise temperature of 74 K (NF of 1.9 dB), 100 MHz of bandwidth, with a 10 mW pump signal at 890 MHz. Similar results were obtained in [61, 65, 54].

Distributed amplifiers have been already integrated in CMOS technology [67, 66, 68]. A simplified schematic of the usual reported approach is depicted in Figure 3.5. As shown in Figure 3.5, the MOS devices, that are separated by inductors or transmission line (TL) segments, are used in a common source configuration. The drains and sources are connected by a two separated TL. The original signal is fed into the input TL. Each gate is then affected by it as it will propagate along the line until it reaches the matched termination. At the drain a current is correspondingly injected into the output TL. If the phase constant of both lines are similar, it is expected that these drain currents add constructively at the output, [66]. A recent implementation of a 7 stage structure CMOS TWA, [68], in a standard 130 nm CMOS digital technology, achieved 8.5 dB of gain up to 40 GHz, a 3 dB bandwidth of more than 50 GHz and a Noise Figure below 7 dB.

The use of MOS varactors in TWA is not common. However, to reduce the Noise Figure, a slightly modification can be made in the circuit of Figure 3.5 which consists on replacing the active conductance transistors by MOS varactors, which are energized by a common pump oscillator.

3.3 Varactors in CMOS technology

By the time the parametric amplifier was strongly developed, the semiconductor diode played an important role as a simple and nonlinear variable capacitor element. This component continues to be available today, namely, within most of the standard CMOS generations and therefore a short description is given next.

Creating a junction between a P-type and N-type doped regions in a Silicon (Si) environment, [69] originates a distinct electronic element. The resulting diode semiconductor has a nonlinear operation, which results from the balance between a few physical phenomenons that appear across the junction. The most important ones are diffusion of carriers and the corresponding and opposite built-in electric field, [29].

By applying an external voltage higher than a threshold voltage, the mobile carriers gain sufficient momentum to cross the junction barrier thus creating the necessary conditions for electrical current to flow. The diode is said to be in a forward biased state.

In the opposite case, a negative voltage is applied to the PN junction, putting the diode in reverse-biasing condition in which a negligible DC current flows. This is due to the extension of the depletion region that surrounds the junction interface. Inside the depleted region, the majority mobile carrier are a scarce resources (all of them where moved apart by the external electrical field). Under these conditions, the P-type and N-type regions behave like capacitor plates while the depletion region acts as the 'dielectric' between them. The thickness of this depletion region defines the distance between the capacitor plates, directly affecting the equivalent capacitance value. Knowing the depletion region length, d_j, one can determine the equivalent small signal capacitance value of the reverse biased diode, [29], resulting in

$$C_j(V_R) = \frac{\varepsilon_{si}\, A}{d_j} = \frac{A \cdot \sqrt{2q\varepsilon_{si}\frac{N_A N_D}{N_A+N_D}}}{2\sqrt{V_R+\phi_{bi}}} = \frac{C_{j0}}{\left[1+\frac{V_R}{\phi_{bi}}\right]^{\alpha_j}}\,, \tag{3.13}$$

where A the junction cross-sectional area, N_A and N_D are the acceptor and donor densities, ϕ_{bi} is the built-in potential of the junction (given by the difference between the Fermi potentials of the N and P sides, [29]), C_{j0} denotes the capacitance value for zero reverse diode voltage V_R. The value of α_j reflects the doping profile which is equal to 1/2 for an abrupt one and 1/3 for a linearly graded junction.

The obtained result (3.13) reveals that the reversed biased diode capacitance is a nonlinear function of the applied voltage thereby validating its use in a continuous

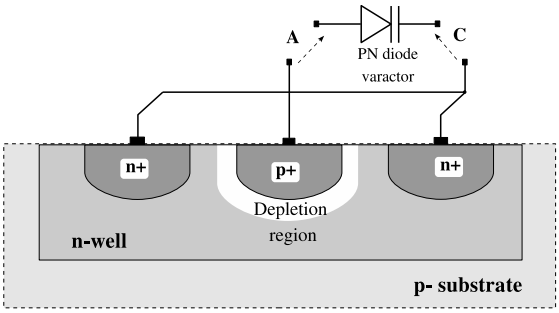


Figure 3.6: P+ to n-well diode based varactor.

time parametric amplification.

To characterize the varactor performance, a few specifications have been defined. One of them reflects the capacitance tuning range of the device and is usually described by the ratio between the maximum and minimum capacitance values,

$$C_r = \frac{C_{max}}{C_{min}} \ . \tag{3.14}$$

Another important figure of merit to measure the varactor operation is the quality factor, [70], defined by

$$Q = 2\pi \frac{\text{Stored energy}}{\text{Dissipated energy per cycle}} = \frac{1}{\omega RC} \ , \tag{3.15}$$

where R and C are the values of the series resistance (loss) and capacitance.

Varactor based on junction diodes can be used in standard CMOS technology, but here only source/drain to well junctions are available. An example of diode structure is depicted in Figure 3.6. Another configurations are possible and can be seen in [70]. This type of PN diode can reach high quality factor (higher than 20) when implemented in MOS technology, [71, 72], but the tuning range is small, lower than 2, [72]. Furthermore, since the charge *versus* voltage (C-V) diode characteristic is not very sharp, the tunning range tends to worsen due to MOS technology downscaling and to the decrease of the voltage dynamic range. Alternatively one can use two type of integrated MOS varactors: an Accumulation mode MOS varactor (AMOS) and an Inversion mode MOS varactor (IMOS).

In a P-type substrate MOS technology, the accumulation NMOS varactor (ANMOS), shown in Figure 3.7, is implemented in a N-well region. The controlling varactor voltage is applied between the drain/source N+ implants areas and the other varactor terminal, the gate. The variation of the capacitance set by the controlling voltage is obtained by changing the state of the device from accumulation to depletion and vice-

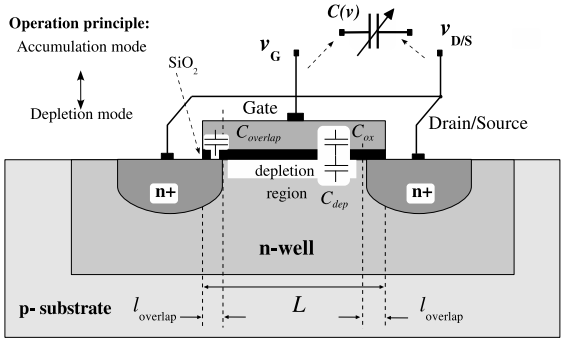


Figure 3.7: Accumulation mode MOS varactor.

versa. When in depletion the total gate capacitance is given by the series of the oxide capacitance C_{ox} and the depletion capacitance C_{dep}, resulting in a small value (due to C_{dep}). On the opposite direction, the varactor reach a maximum value of C_{ox} when in deep accumulation state. Despite the sharper C-V characteristic when compared with the diode based varactor, the accumulation mode varactor also suffers from the reduced control voltage range imposed by the advanced CMOS technology nodes. In fact, the C-V sharpness, for accumulation MOS varactor, is not sufficiently high to maximize the tunning range in case of controlling voltage lower than 1 V, [73].

It can be observed in the C-V characteristic of a MOS transistor configuration that, the transition between depletion to inversion is much faster than from depletion into accumulation. This is due to the distinct physical phenomenons that happens in each case. When the MOS device enters in weak inversion region, the fast capacitance change towards C_{ox} is due to the exponential growth of the charge at the interface Si-SiO_2 (corresponding to the channel below the gate area). In order to facilitate the use $C-V$ characteristic segment from depletion to inversion, a simpler configuration based on a standard MOS transistor can be adapted. In fact, it is just needed to setting up a short-circuit between the drain and source and designate the gate terminal as the other capacitor plate. Figure 3.8 illustrates the resulting Inversion Mode MOS varactor (IMOS), implemented from a NMOS device.

The variation in capacitance is greater in inversion and accumulation MOS varactor than in PN junction type. Consequently, higher tunning range is achievable with the former type, considering the same device area. With respect to the quality factor, the PN type varactor reaches better values due to their lower resistance, [70]. The higher losses in the MOS varactor compared to PN-junction varactor come partly from the ohmic loss of the polysilicon gate, partly from the high series resistance of the lightly-doped area near the source and drain regions, and also from the channel conductance (when in inversion state), [74, 75]. Nevertheless, MOS varactors are much

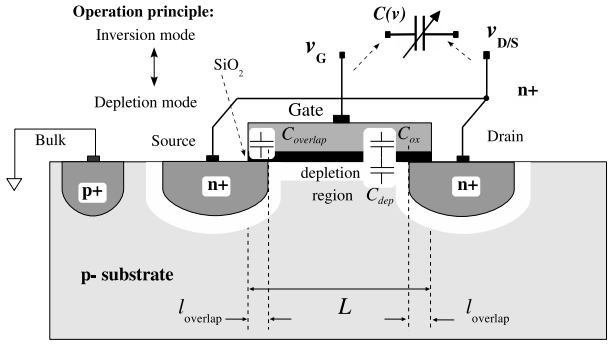


Figure 3.8: Inversion mode MOS varactor.

more area-efficient (in terms of absolute capacitance values) and show much sharper C-V characteristics, which is very important for low voltage applications, [73]. Additional degradation from parasitic factors contributes to the lower the quality factor of the MOS varactor as losses in the substrate and the well, the ohmic losses of metal and polysilicon interconnections, and the parasitic capacitances, [72, 70]

Reference [72] presents a comparative analysis between the three varactor structures, based on experimental results obtained from an old 0.5 μm CMOS technology. Table 3.1 summarizes the experimental reported results. Results for more recent technologies, [70, 73, 76, 77] show similar trends. Inversion MOS-based varactors are gaining more

Varactor Type	C_{max}/C_{min}	Q_{min}	Q_{max}
Diode ($p^+$to n-well)	1.32	18.0	22.6
Diode ($p^+$to n-well, small size)	1.23	94.5	109.0
Accumulation NMOS	1.69	33.2	38.3
Inversion NMOS	2.15	25.8	34.3

Table 3.1: Performance comparison between varactor in MOS technology, [72].

interest over the reverse-biased diodes due to their wider tuning range and lower voltage range of operation, both of which improve with every new process generation. In particular, the discrete-time amplification technique that is used throughout this work, uses an inversion based MOS varactor as the parametric reactance device.

3.4 Manley-Rowe power relations for nonlinear reactances

Only several year later from its initial work, [50], Jack Manley and Harrison Rowe have finished the development of a number of fundamental equations concerning the

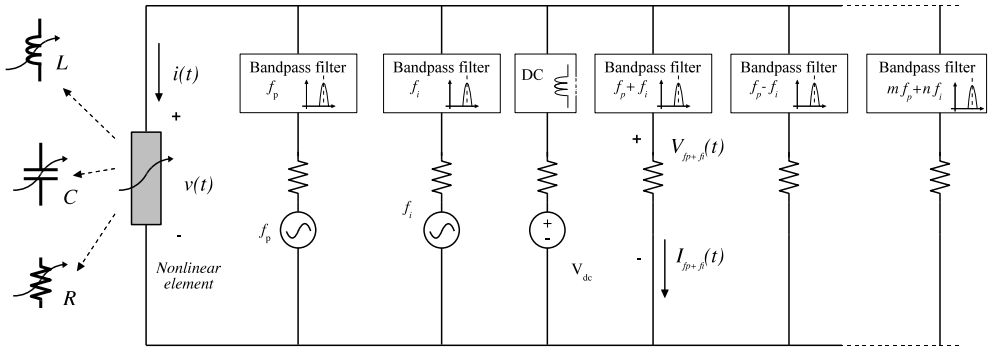


Figure 3.9: Circuit model for the Manley-Rowe relations.

flow of energy between the different signals frequencies components originated when a nonlinear and lossless reactance is driven by distinct signals simultaneously. The power relations between these signals are described by the designated Manley-Rower relations, [78, 79].

When applying simultaneously a strong local oscillator, or pump signal, of frequency f_p and an input signal of f_i to a nonlinear capacitor, the set of new frequencies that will appear in the circuit includes harmonic, sum and difference combinations in the form of $f_{m,n} = m\,f_p + n\,f_i$, with m and n integer positive, negative, or zero. Additionally, the two fundamental frequencies f_p and f_i are considered to be positive and incommensurable (leading to independent variables) [78] and the other ones are the resulting sidebands.

The determination of the power relations is based on the analytical model represented in Figure 3.9, where resistive loads are connected in series with an ideal bandpass filter. Each resistive branch is then connected in parallel to the nonlinear capacitor. As mentioned before, it is assumed that the nonlinear element is lossless and does not present any hysteresis in the characteristic charge-voltage, C-V, in case of a capacitor, or inductance-current, L-I, in case of an inductor. Since the nonlinear capacitor is purely reactive, it does not dissipate energy, that is, no net power into or out of this element will be possible. In other words, the lossless nonlinear capacitor do store energy but it does not dissipate it. Adding the fact that the fundamental frequencies are considered to be incommensurable, the time-average power due to interacting harmonics will be null, [69]. In particular, the form taken by the principle of energy conservation in this case is expressed by

$$\sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} P_{nm} = 0\ , \tag{3.16}$$

where the sum extends to all signal frequencies generated at the nonlinear reactance.

P_{nm} is the average power at frequency $f_{m,n} = m\, f_p + n\, f_i$ and in the nonlinear reactance case, only the real power P_{nm} appears, [78].

$$P_{n,m} = V_{n,m} I^*_{n,m} + V^*_{n,m} I_{n,m} = 1/2\; Re\left[V_{n,m} I^*_{n,.m}\right] \tag{3.17}$$

The $I_{n,m}$ and $V_{n,m}$ components in Equation 3.17, are the coefficients of the double Fourier series associated to the total voltage and current present at the nonlinear reactance (a capacitor in this case), which are described in Equations 3.18 to 3.20.

$$v = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} V_{n,m}\, e^{j(n\omega_p t + m\omega_i t)} \tag{3.18}$$

For signal $v(t)$ to be real, one must satisfy $V_{m,n} = V^*_{-m,-n}$ and $V_{-m,-n} = V^*_{m,n}$, for the complex amplitudes. The coefficients are given by

$$V_{n,m} = \frac{1}{4\pi^2} \int_0^{2\pi} dy \int_0^{2\pi} dx \cdot v \cdot e^{-j(nx+my)} \;, \tag{3.19}$$

where $x = \omega_p t$ and $y = \omega_i t$. In the same way, the total current flowing in the capacitor can be expanded in double Taylor series as well, resulting in

$$\begin{aligned} i = \frac{dq}{dt} &= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} j\,(n\omega_p + m\omega_i)\; Q_{n,m}\, e^{j(n\omega_p t + m\omega_i t)} \\ &= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} I_{n,m}\, e^{j(n\omega_p t + m\omega_i t)} \;. \end{aligned} \tag{3.20}$$

Note that the current calculation in the capacitor has assumed a nonlinear function of the charge with respect to the applied voltage $q = f(v)$. Furthermore, it was assumed that this function meets all conditions, namely single-value, to be expanded in Taylor series. From 3.16 and after some manipulation, the Manley-Rowe relations are derived, [78, 79, 69], resulting in

$$\sum_{n=0}^{\infty} \sum_{m=-\infty}^{\infty} \frac{n P_{n,m}}{n\omega_p + n\omega_i} = 0 \tag{3.21}$$

$$\sum_{m=0}^{\infty} \sum_{n=-\infty}^{\infty} \frac{m P_{n,m}}{n\omega_p + n\omega_i} = 0 \tag{3.22}$$

where n and m are integers and $P_{n,m}$ represents the power series coefficients. Although the condition of energy conservation (i.e., the varactor with no losses) is necessary for the previous derivation, it also depends on the long term averaged power of a component to be proportional to its frequency, [80]. The later condition occurs in the capacitor case, where the instantaneous power is given by the product of v and $dq/dt = i$. In fact when differentiating q with respect to time, the term becomes multiplied by the

corresponding frequency $n\,\omega_p + m\,\omega_i$.

A general power relations for nonlinear resistive elements can be found in [81] and the extension for more than 2 exciting sources can be found in [80]. An example of using the Manley-Rowe results is the up-converter parametric circuit topology, where it is only allowed to flow the currents associated to f_i, f_p and $f_i + f_p$. Using the results of 3.21 and 3.22, one can obtain the following relations between signal frequencies:

$$\frac{P_{1,0}}{\omega_i} + \frac{P_{1,1}}{\omega_i + \omega_p} = 0, \tag{3.23}$$

$$\frac{P_{0,1}}{\omega_p} + \frac{P_{1,1}}{\omega_i + \omega_p} = 0. \tag{3.24}$$

The supplied power is provided at frequencies f_i and f_p, which means that the $P_{1,0}$ and $P_{0,1}$ component power are positive. As a consequence the power at the sideband harmonic component is negative meaning that the power is delivered to the load from the nonlinear reactance at the frequency f_i+f_p. This results in the maximum achievable power gain given by

$$-\frac{P_{1,1}}{P_{1,0}} = \frac{f_p + f_i}{f_i} = 1 + \frac{f_p}{f_i}\ . \tag{3.25}$$

Similar relations can be obtained for downconverter parametric circuits or just parametric amplifiers (where $f_{out} = f_i$), [69].

3.5 CMOS parametric amplification with frequency conversion in continuous time domain

In a receiver chain, the frequency downconversion stage can be accomplished by a parametric based mixing circuit. Not only a reduced noise addition is expected from this operation (due to the low noise parametric characteristic) but, if the local oscillator frequency is conveniently chosen, the parametric mixer can also deliver an effective extra power gain. In fact, if one choose to use a local oscillator frequency $f_p = f_{LO}$ higher than the RF carrier frequency $f_i = f_c$ (but lower than $2\,f_c$), from the Manley-Rowe expressions one could expect an effective conversion power gain implicitly described by (3.26) and (3.27). The negative power expressed in (3.27) means that the nonlinear reactance is delivering power to the IF load. The energy is coming from the local oscillator but to reach an IF power higher than the input signal, the power of the

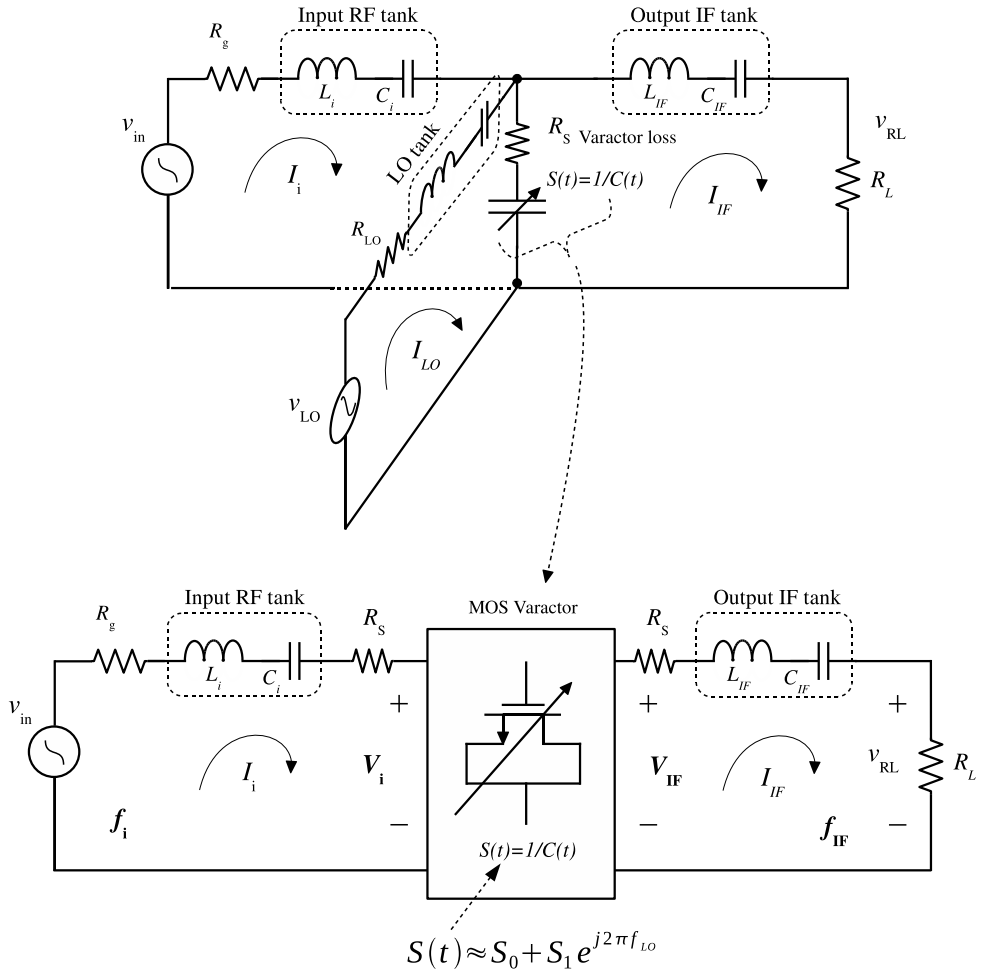


Figure 3.10: Equivalent simplified model for the parametric mixer.

local oscillator must be higher than the one of the input signal.

$$P_{IF} = \frac{f_{IF}}{f_i} \cdot P_i \tag{3.26}$$

$$P_{IF} = -\frac{f_{IF}}{f_{LO}} \cdot P_{LO} \tag{3.27}$$

The Manley-Rowe relations establish a maximum limit of the power conversion gain under loss-less conditions and, interestingly, they depend on frequency relations rather than explicit circuit implementation details. However, to obtain the actual power gain expression, a two-port analysis can be adapted for the mixer case. To support this approach, a conversion matrix (admittance or impedance type) is filled up [82, 69, 83] under the conditions of strong LO signal (in comparison to the input RF signal). This reasonable requirement means that only the oscillator signal is responsible for the variations of the nonlinear reactance element. A complementary analysis based on a coupled-mode theory can be found in [83].

The mixer conversion matrix describes the relationships (and couplings) between voltages and currents, in a phasor type notation (assuming sinusoidal regime), coming from different circuit sections that operate at different frequencies, as shown in Figure 3.10. Under this construction, one must also account for the losses due to the varactor. Considering the typical structure of a MOS varactor, a resistor R_s might be included in series with the variable reactance, [84]. Moreover, to represent the voltages at the varactor two-port model (including losses), it is better to represent the variable capacitor by its elastance rather than the capacitance [54, 84, 85]. Therefore,

$$v\left(t\right) = R_s \cdot i\left(t\right) + \int S\left(t\right) \cdot i\left(t\right)\, dt\,, \tag{3.28}$$

where $S(t)$ represent the elastance of the varactor as a function of time,

$$S\left(t\right) = \frac{1}{C\left(t\right)} = \sum_{p=-\infty}^{+\infty} \left(S_p\, e^{jp\omega_{LO}t}\right)\,. \tag{3.29}$$

Inside the mixer circuit three frequencies coexist and all the remaining ones are assumed that are filtered out by tunned LC tanks. These three frequencies include the input signal frequency f_i, the local oscillator f_{LO}, and the intermediate signal frequency f_{IF}. Moreover, it is considered that all of them are coupled together only through the time-varying component of the varactor. Therefore, the equivalent circuit model, when only the input and IF (output) sections are explicitly shown, is illustrated in Figure 3.10.

Based on the previous model, the conversion impedance matrix, [83],

$$\begin{bmatrix} V_i \\ V_{IF}^* \end{bmatrix} = \begin{bmatrix} R_s + \frac{S_0}{j2\pi f_i} & -\frac{S_1}{j2\pi f_{IF}} \\ \frac{S_1^*}{j2\pi f_i} & R_s - \frac{S_0}{j2\pi f_{IF}} \end{bmatrix} \cdot \begin{bmatrix} I_i \\ I_{IF}^* \end{bmatrix}\,, \tag{3.30}$$

is obtained for the donwconversion parametric mixer case where the local oscillator frequency is higher than the input RF signal frequency. This latter fact establishes a direct coupling to the negative intermediate frequency which is shown as conjugate voltages and currents phasors. In 3.30, S_0 is the varactor average value while S_1 represents the elastance variation at the LO frequency.

The conversion matrix reflects a linearization process applied around the nonlinear varactor that involves currents and voltages at different mixing frequencies (which are coupled through the variable component of the varactor). Since, under these conditions, linear circuit theory holds true for each matrix element, Kirchhoff's laws can be applied to connect the circuit elements thus originating closed form for the power conversion gain between RF and IF signals. Considering that the mixer is in a tunning state, the

midband conversion power gain is determined from the output power developed in R_L, $1/2\,|I_{IF}|^2\,R_L$, and from the available power at input, $V_{in}^2/8R_g$. The final result of the gain at IF, [86, 83], is given by

$$G_{IF} = \frac{4 \cdot R_L R_g \cdot |S_1|^2}{\left[2\pi f_i \cdot (R_L + R_s)\,(R_g + R_s) - \frac{|S_1|^2}{2\pi f_{IF}}\right]^2}\,, \tag{3.31}$$

where the input source resistance R_g and load resistance R_L have also been included.

The two-port circuit model is also used to find the NF performance of the mixer. In this case, the resistive loss of the varactor contributes with noise both at RF and IF frequencies, which is reflected by

$$e_{nIF}^2 = e_{ni}^2 = 4kTR_s\Delta f\,. \tag{3.32}$$

Additionally, the resistive loss at the RF input also constitutes a source of noise which is expressed by

$$e_{n,g}^2 = 4kTR_{ip}\Delta f\,. \tag{3.33}$$

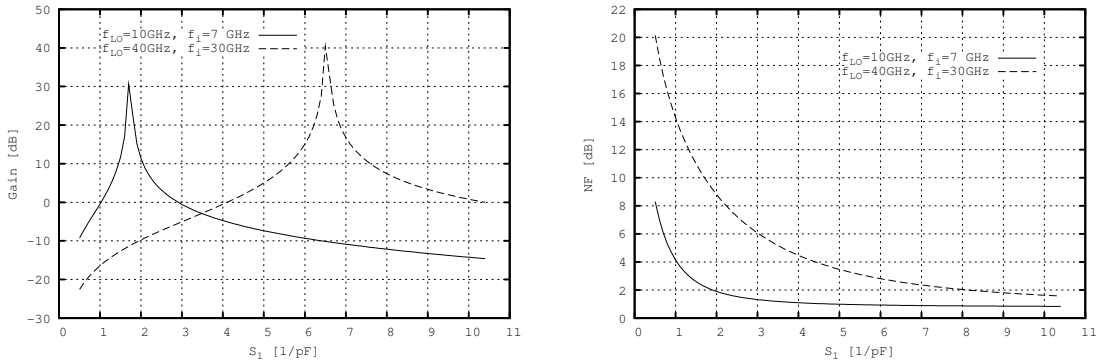
The final noise factor associated with this class of mixers is given by

$$F_{IF} = 1 + \frac{e_{ni}^2}{e_{n,s}^2} + \frac{e_{nIF}^2}{e_{n,s}^2 \cdot \left[\frac{|S_1|}{2\pi f_i \cdot (R_{ip}+R_s)}\right]^2} \tag{3.34}$$

where it can be observed the strong dependence on the S_1, as expected.

The elastance variation range is dependent on the device but also on the pumping conditions. It has been demonstrated, [86, 83], that the elastance parameter S_1 can reach up to $10^{12}\,F^{-1}$ in a 130 nm CMOS with reasonable sized device and oscillator amplitude (200 mV). Based on these assumptions, the graphics represented in 3.11 trace the evolution of the transducer power gain and NF for two combinations. In the first one, it was considered a 10 GHz LO, a 7 GHz RF input signal and a 3 GHz IF. For the second case, a higher LO frequency is used (40 GHz) with an RF input of 30 GHz. It is clear from the results that S_1 has to be carefully chosen in order to obtain an effective power gain higher than 1. Moreover, two possible value of S_1 can reach the same gain level. The tradeoff is then made at the NF result. If a lower NF is desired, the highest S_1 has to be chosen. Interestingly, it can also be observed that pushing upward the frequencies of the LO and of the input RF signals must be compensated by an increase of S_1 (by changing the LO amplitude and/or the MOS varactor area).

The use of an MOS varactor as parametric mixer is still in its emerging phase. One



(a) Power transducer gain

(b) Noise Figure

Figure 3.11: Parametric downconvertion mixer gain and NF.

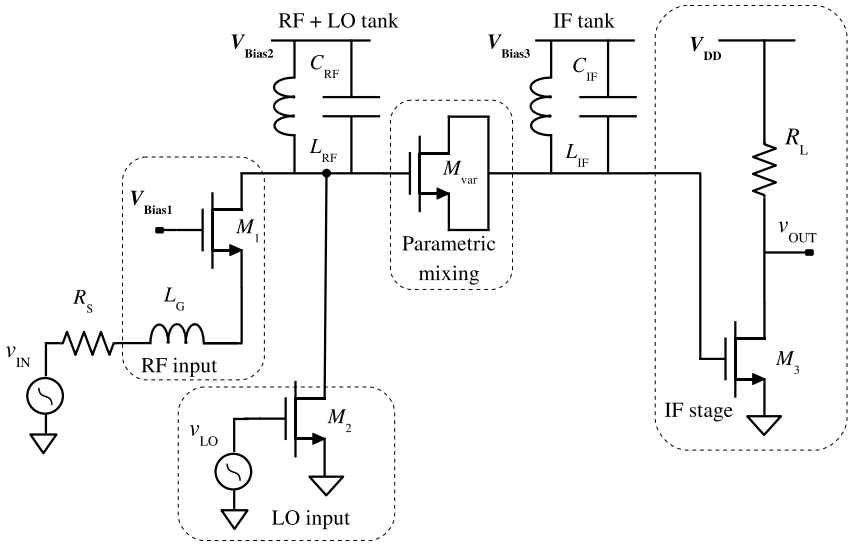


Figure 3.12: Simplified model of a MOSFET parametric down converter, [83].

possible implementation is the circuit is proposed in [83] which can act as the first microwave/millimeter wave frequency conversion step in a heterodyne receiver. The circuit is shown in Figure 3.12 and includes a MOS varactor as the parametric element, two-tank, which resonate at the input/LO and idler frequencies. The input and LO signals are superimposed through the addition of the current coming from the input common gate stage, formed by M_1, and the current generated by the LO signal at M_2. The stronger LO signal will modulate the MOS varactor which is biased by V_{bias2} and V_{bias3}. The use of the common gate input M_1 ensures high bandwidth and facilitates the input 50 Ω matching since its input impedance is roughly given by $1/g_m$. The mixer circuit, designed in a 130 nm technology, is able to downconvert an RF signal at 30 GHz into an IF of 10 GHz with a conversion voltage gain higher than 4 dB and a NF less than 1.8 dB. The MOS varactor is usually used as two terminal device. However,

the gate-bulk capacitance can be controlled independently by a third terminal formed by the D/S connection. This three terminal MOS varactor device can be used in a discrete-time configuration as described in next Chapter. However, a continuous time operation can also be envisaged.

In complement of MOS continuous-time parametric downconverter mixer, [87, 85, 86, 83, 88], another example is the use parametric approach in a frequency multiplier circuit [89]. The analyzed MOS mixer is to be used in a receiver chain. Nevertheless, the parametric conversion can also be very efficient in the upconversion stage of the emission chain. A first example has been proposed in a FET technology to build a Parametric MOS RF power amplifier, [90].

3.6 Summary

The amplification and frequency conversion based on parametric variations of reactive elements is extensive and the Chapter has focused in the most important results from the supporting theory. In addition to the referred bibliography, more detailed analysis of the parametric analog signal approach can be found in [84, 91, 92].

Being the varactor the key element for the parametric approach, rather than the alternative inductor option, an overview over the varactor structures in CMOS technology was made. Of special importance is the MOS inversion mode varactor that will be the basis of the discrete-time amplification that is discussed in next Chapter.

Chapter 4

Discrete Time Parametric Amplification in digital CMOS technology

4.1 Introduction

The mixing action that occurs in a nonlinear capacitor, subject of discussion in Chapter 3, can be used to transfer energy from a pump source to the signal of interest, thus enabling the implementation of amplifiers without requiring transconductance based configurations. In the resulting circuitry, one or more tanks are used to filter or to act as temporary energy storage element (e.g., the idler block). Usually, the design of these tanks includes inductors which are needed to achieve resonance modes.

Some characteristics of the parametric circuits, analyzed in the previous Chapter, fit well to submicron MOS technology. Despites having the advantage of reaching higher frequency with lower intrinsic noise signature, the need of inductors pushes the design of an integrated parametric amplifier towards an alternative approach.

A closer look of the continuous-time operation cycle, [85], of the parametric circuit reveals that it starts by gathering the signal energy into the reactive element while, in a second moment, the signal energy is increased by a synchronized pump AC source. After this pumping effect, the reactive element delivers the energy to the load and, only afterwards, it suffers a reset to an initial known state. This description shows that this continuous time operation can be discretized in time through the use of clock driven switches inserted among the signal path, thus transforming the original circuit

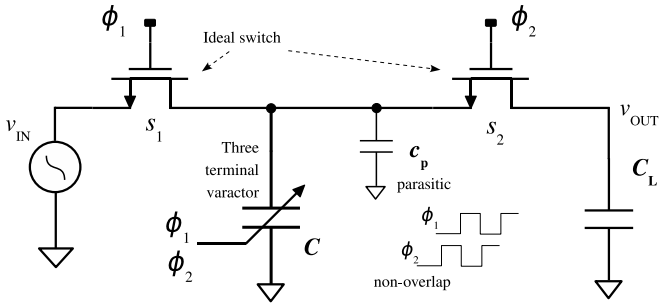


Figure 4.1: Conceptual discrete-time parametric amplifier.

into a discrete-time parametric amplifier (DT-PAMP). This Chapter will show that the resulting circuit does not need inductors, thus facilitating the integration in standard MOS technology.

The Chapter begins by presenting the principles of a DT-PAMP operation. In Section 4.2, the MOS based DT-PAMP (MPA) implementation is analyzed while the performance study is separately given in the subsequent Sections. In last Sections, a modification of the original MPA cell is proposed as well as several applications of circuits with embedded MPA.

4.2 Conceptual operation of a discrete-time parametric amplifier

The discrete-time operation of a parametric based amplifier is reached by adding switching elements, and associated clock signals, around a nonlinear reactance. The conceptual schematic of this amplifier is depicted in Figure 4.1 as well as the discrimination of the main operating phases. The conceptual DT-PAMP diagram is composed of an input voltage source v_{IN}, input and output switches, a capacitive load C_L, and an electrically variable MOS capacitor C which is controlled by a squared voltage pump signal with frequency f_p, [93, 14], thah corresponds to the sampling frequency F_s. Additionally, a generic parasitic capacitance c_p is also included in the model. For this simplified analysis it is considered that the ideal switches are driven by two-phase non-overlapping clocks (ϕ_1, ϕ_2), and f_p is set by ϕ_2.

During ϕ_1 a sample of the input signal is acquired by C through switch S_1 (for simplicity reasons it is considered that the duration of ϕ_1 is long enough to fully charge the capacitor). During this sampling phase, the output switch S_2 is opened and the controlling pump signal is such that the varactor C reaches a high capacitance value,

$C_{\phi 1}$. Therefore, at the end of ϕ_1, the total stored charge due to the sampled input voltage is given by

$$Q_{\phi 1} = Q_{C,\phi 1} + Q_{cp,\phi 1} = C_{\phi 1} \cdot v_{IN} + c_p \cdot v_{IN} \,, \tag{4.1}$$

corresponding to a total stored energy, [14], in capacitors C and c_p of

$$E_{\phi 1} = \frac{1}{2} \cdot Q_{\phi 1} \cdot v_{IN} = \frac{1}{2} \cdot (C_{\phi 1} + c_p) \cdot v_{IN}^2 \,. \tag{4.2}$$

At beginning of ϕ_2, the input switch is OFF but the output one remains also OFF for some instants while the pump source forces the capacitance value of the varactor to change to $C_{\phi 2}$. During this short interval, the charge of C remains at the value that was sampled at the end of ϕ_1. Then, neglecting c_p (for simplicity) and due to charge conservation, the voltage across C has to change accordingly to

$$v_{C,\phi 2} = \frac{C_{\phi 1}}{C_{\phi 2}} \cdot v_{C,\phi 1} = p \cdot v_{C,\phi 1} \,, \tag{4.3}$$

where p is the parametric gain factor. To achieve an effective voltage gain, p must be higher than one thus indicating that the varactor capacitance value has to be reduced during the transition between phase ϕ_1 to ϕ_2.

At some instant during ϕ_2, the output switch closes thus originating a charge re-distribution between all three capacitors, namely, C, c_p and C_L, which will set the final output voltage. Since the total charge conserves, meaning that $Q_{\phi 2} = Q_{\phi 1}$, the voltage gain achieved by the structure between ϕ_1 and ϕ_2 is found to be,

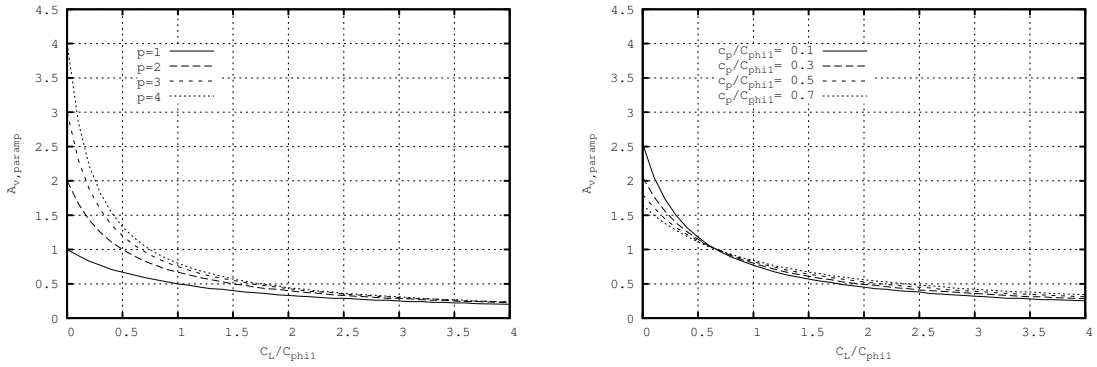
$$A_{v,dtpamp} = \frac{v_{OUT}}{v_{IN}} = \frac{v_{C,\phi 2}}{v_{C,\phi 1}} = \frac{C_{\phi 1} + c_p}{C_{\phi 2} + c_p + C_L} = \frac{C_{\phi 1} + c_p}{\frac{C_{\phi 1}}{p} + c_p + C_L} \,. \tag{4.4}$$

Starting from the ideal case where the voltage gain is given just by the parametric factor, the introduction of the parasitic and load capacitances lowers the total gain. The analysis of (4.4) reveals a strong dependence on the load capacitor. In fact, to achieve a gain higher than one, for a given p, the load capacitor should be lower than,

$$A_{v,dtpamp} \geq 1 \;\Rightarrow\; C_L \leq C_{\phi 1} \cdot \left(\frac{p-1}{p} \right) \,, \tag{4.5}$$

where c_p is neglected.

Figure 4.2 shows the achievable gain *versus* the normalized load capacitance C_L/C_{ϕ_1}, for distinct values of p and different normalized parasitic capacitance $c_p/C_{\phi 1}$. As



(a) For different p and no parasitics

(b) For $p = 3$ and different values for the normalized c_p/C_{ϕ_1}

Figure 4.2: Voltage gain versus normalized load capacitor ($C_L/C_{\phi 1}$).

expected, increasing the parametric gain factor reduces the restriction for C_L which can get closer to $C_{\phi 1}$. At the limit, if p is much higher than 1 and c_p is neglected, the total gain tends to the ratio given by $C_{\phi 1}$ to C_L, which is an interesting result.

Without a varactor, the circuit acts as a standard Sample and Hold (S/H), [94], with a low frequency voltage gain, $A_{v,SH}$, is given by

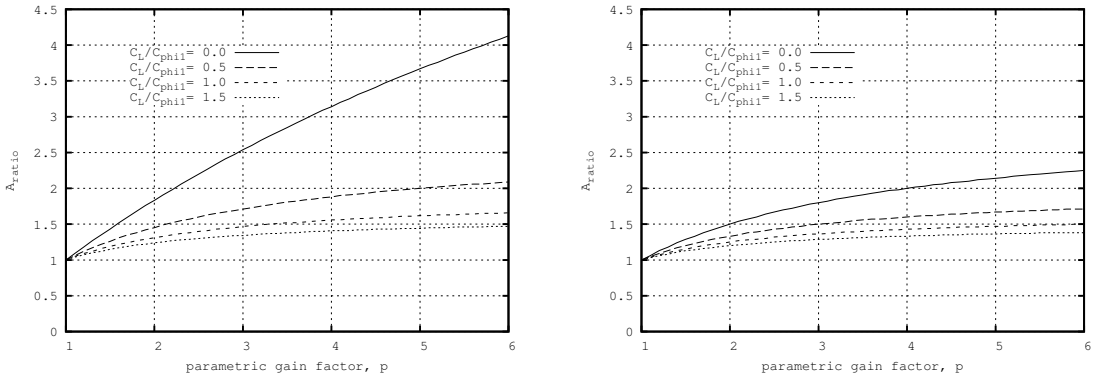
$$A_{v,SH} = \frac{C_{SH} + c_p}{C_{SH} + c_{_p} + C_L}\,, \tag{4.6}$$

where C_{SH} represents the sampling capacitor with a fixed capacitance value. From 4.4 and 4.6, one can conclude that, changing the capacitance value of the sampling capacitor from sample to hold phases can be used to compensate the voltage gain drop introduced by the load capacitor. This is confirmed through the ratio between the voltage gain of both structures, which is given by

$$A_{ratio} = \frac{A_{v,dt-pamp}}{A_{SH}} = \frac{C + c_p + C_L}{\frac{C}{p} + c_p + C_L} = \frac{1 + \frac{C}{c_p + C_L}}{1 + \frac{1}{p}\frac{C}{c_p + C_L}}\,, \tag{4.7}$$

where $C_{SH} = C_{\phi 1} = C$ is considered. As stated before, the upper limit is just given by the relation between the load capacitor and the sampling one. These results are also visible in the graphical representation of Figure 4.3, which is obtained calculating (4.7) versus the intrinsic parametric factor p, for various normalized values of C_L.

A similar analysis concerning the energy transfer reveals that under specific conditions, the structure can achieve an effective voltage gain even if the total energy delivered to the load is lower. Assuming that the capacitor load is reset during ϕ_1, at the end of



(a) For different $C_L/C_{\phi 1}$ and $c_p/C_{\phi 1} = 0.1$ (b) For different $C_L/C_{\phi 1}$ and $c_p/C_{\phi 1} = 0.5$

Figure 4.3: Gain improvement by applying parametric amplification to a S/H.

ϕ_2 the stored energy in this reactive element corresponds to the total energy that is delivered by the circuit. Therefore, recognizing that C_L has a voltage v_{OUT} at the end of ϕ_2, its energy can be calculated by, [95],

$$E_{C_L} = \frac{1}{2} C_L v_{OUT}^2 = \frac{1}{2} C_L \left(A_{v,dtpamp} v_{IN} \right)^2 \ . \tag{4.8}$$

Replacing $A_{v,dtpamp}$ by (4.4) in (4.8), results in

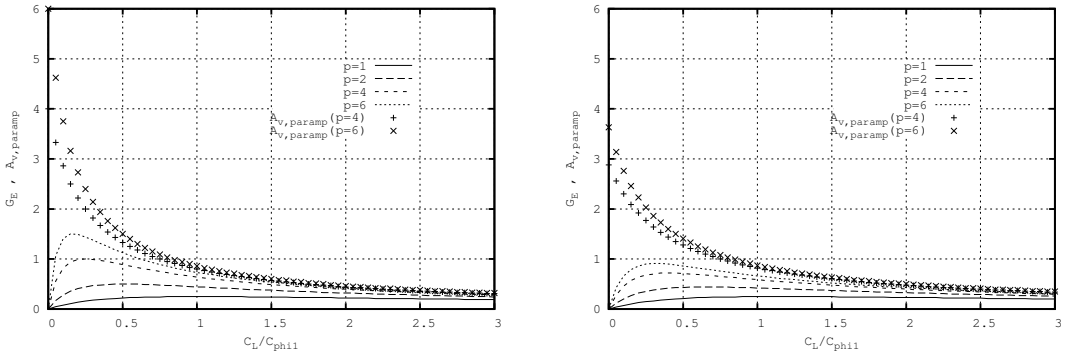
$$E_{C_L} = \underbrace{\left[\frac{1}{2} \cdot (C_{\phi 1} + c_p) \cdot v_{IN}^2 \right]}_{E_{\phi 1}} \cdot \left[\frac{C_L (C_{\phi 1} + c_p)}{\left(\frac{C_{\phi 1}}{p} + c_p + C_L \right)^2} \right] = E_{\phi 1} \cdot G_E \tag{4.9}$$

where G_E represents an energy gain factor, which can take values higher than one depending on the load capacitance and the parametric cell gain $A_{v,dtpamp}$. This result is represented in (4.10), that was obtained from (4.9) after some rearrangements.

$$G_E = \frac{C_L}{\frac{C_{\phi 1}}{p} + c_p + C_L} A_{v,dtpamp} \tag{4.10}$$

The result expressed by (4.9) indicates that a voltage gain higher than one is not a sufficient condition to reach an effective energy gain but, additionally, it is necessary to choose a correct value for C_L. Furthermore, depending on the circuit parametrization, this value might not exist, meaning that the load will receive less energy.

A deeper insight comes up after determining the coordinates of the maximum achievable G_E. As usual, this is found by calculating the point where the derivative of (4.11) is null. For the sake of simplicity and convenience, c_p is assumed to be small and, therefore,



(a) For different p and no parasitics.

(b) For different p and $c_p/C_{\phi 1} = 0.15$.

Figure 4.4: Energy and voltage gain versus normalized load capacitance $C_L/C_{\phi 1}$.

negligible.

$$\left.\frac{dG_E}{dC_L}\right|_{C_{L,GEmax}} = \frac{C_{\phi 1}\left(\frac{C_{\phi 1}}{p}+C_L\right)^2 - 2\cdot C_{\phi 1}C_L\left(\frac{C_{\phi 1}}{p}+C_L\right)}{\left(\frac{C_{\phi 1}}{p}+C_L\right)^4} = 0 \tag{4.11}$$

The resulting value of C_L to maximize G_E is given by

$$C_{L,GEmax} = \frac{C_{\phi 1}}{p} = C_{\phi 2} \tag{4.12}$$

which coincides to the capacitance of the varactor C during phase ϕ_2.

A comparable result is obtained when determining the maximum power transfer from a source with internal resistance R_s to a resistive load R_L. Under the previous conditions, the energy delivered to the load reaches a maximum gain of

$$G_E\left(C_{L,GEmax}\right) = G_{E,max} = \frac{p}{4}, \tag{4.13}$$

while the voltage gain is higher by a factor of two,

$$A_{v,paramp}\left(C_{L,GEmax}\right) = \frac{p}{2}. \tag{4.14}$$

Figure 4.4 represents the variation of the voltage and energy gain as a function of the normalized load capacitance for several values of the parametric gain factor p and normalized parasitic capacitance. The voltage gain monotonically decreases with growing C_L and is higher than the energy gain. This graphic confirms that for an intrinsic parametric voltage gain of four, the energy gain is only one. For higher values of p, the overshoot on the plot of G_E corresponds to an effective energy gain at the

load capacitor. Similarly to the continuous time parametric amplification, governed by the Manley-Rowe relations presented in Chapter 3, the extra energy in ϕ_2 is injected in the circuit by the pump source.

4.3 Implementation of a Discrete-Time MOS Parametric amplifier

Attending to the previous analysis, a varactor device can be setup as a DT-PAMP as long as its capacitance changes between two distinct values by the action of a pump electrical source. On the other hand, and considering the results from Section 3.3, the MOS device is a natural varactor choice mainly due to the sharp transition of its C-V characteristic between depletion and inversion regions. The alternative accumulation-mode MOS varactor, presents a smoother transition and requires less standard process fabrication, [75, 77, 72]. Therefore, only the Inversion-mode MOS varactor will be considering in the following analysis.

The MOS varactor has been essentially used in circuits with tuning requirements, [75, 96, 30]. Using it as a parametric amplifier or a variable coupling device is not so common, but one of the first exceptions was published in 1972, [97]. In the circuit that was then proposed, a MOS variable capacitor is used to compensate threshold losses in dynamic MOS logic digital circuits.

Based on the circuit shown in Figure 4.5, the dynamic logic MOS circuit relies on the charge storage tailored by the parasitic capacitance at node G and by the high input impedance at the MOS gate terminal of T_2. Binary logic "0" and "1" is then stored by setting previously the voltage at node G, respectively, to a threshold voltage below and higher than V_{th} of the MOS device, during a memorization phase. When transistor T_2, acting as a diode switch, is ON, the voltage a node G is set while when is in OFF mode G is surrounded by high impedance devices and therefore the voltage is kept at reasonable constant value. Nevertheless, due to leakage currents, it experiences a small voltage decay, that can turn OFF T_1 under certain conditions, thus changing the logic level. Furthermore, due to the MOS operation, node v_{OUT} cannot rise to voltage higher than two V_{th} below v_{G1}. The last two issues can be alleviated by changing (increase negatively in the PMOS case) the node G voltage during the memory phase (after the pre-charge). This might be achieved by introducing a capacitive coupling between nodes G and D, and simultaneously inject a periodic bootstrap pulse voltage, as shown in Figure 4.5. Furthermore, if the capacitive coupling is supported by a varactor, C_{var},

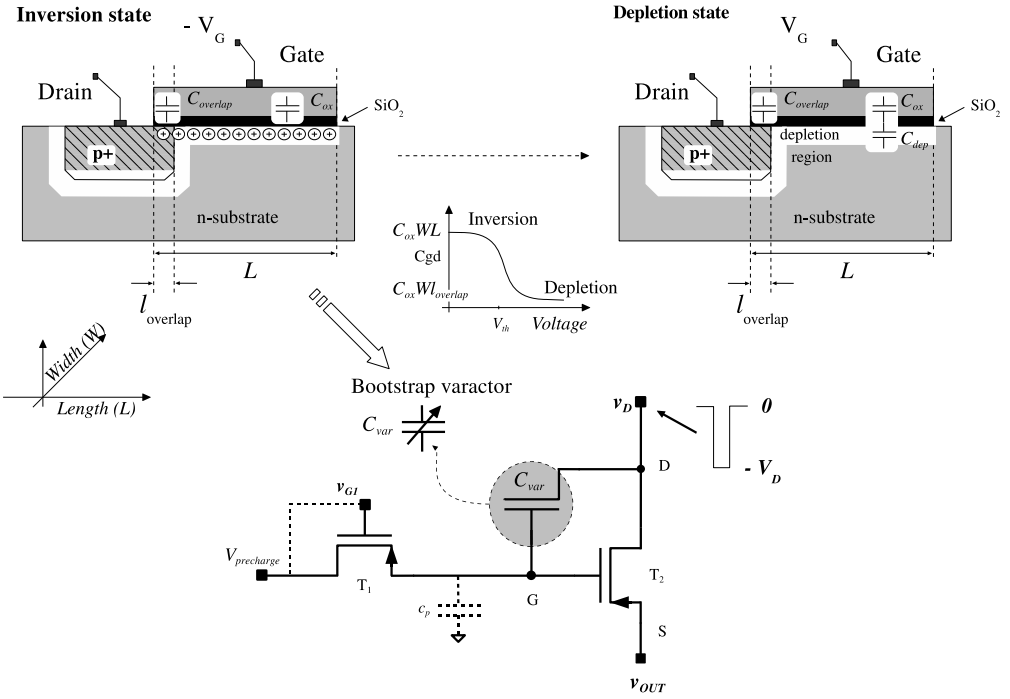


Figure 4.5: Dynamic MOS logic with bootstrap varactor, based on [97].

the level of charge injection into node G is not only determined by the v_D step but also by the state of the variable capacitor, which can change from inversion to depletion. The level of charge injection is reflected on the bootstrap ratio, [97], that is given by

$$\triangle v_G = \underbrace{\frac{C_{var}}{c_p + C_{var}}}_{r} \cdot \triangle v_D \ , \tag{4.15}$$

where c_p is the parasitic capacitance at node G, and $\triangle v_D$ is the step voltage at the drain of T_2.

The varactor structure used in [97], which is depicted in Figure 4.5, has only two terminals: one drain and gate, while the other channel peer terminal is not connected. Despite this apparent difference, the structure is fully compatible with CMOS. In inversion, when v_G is higher than the threshold voltage, an inverted channel exists below the gate and is directly connect to the drain region. This conductive area forms one plate of the capacitor while the other one is the gate itself. In this state, the capacitance is high and approximately equal to C_{ox}. When below the threshold level, the channel charges flow towards the drain area letting the area below the gate completely depleted. In this case, the capacitance between the gate and the drain is reduced to just the overlap one, see Figure 4.5, which is much smaller.

The implementation of a MOS varactor is not restricted to two terminals since the

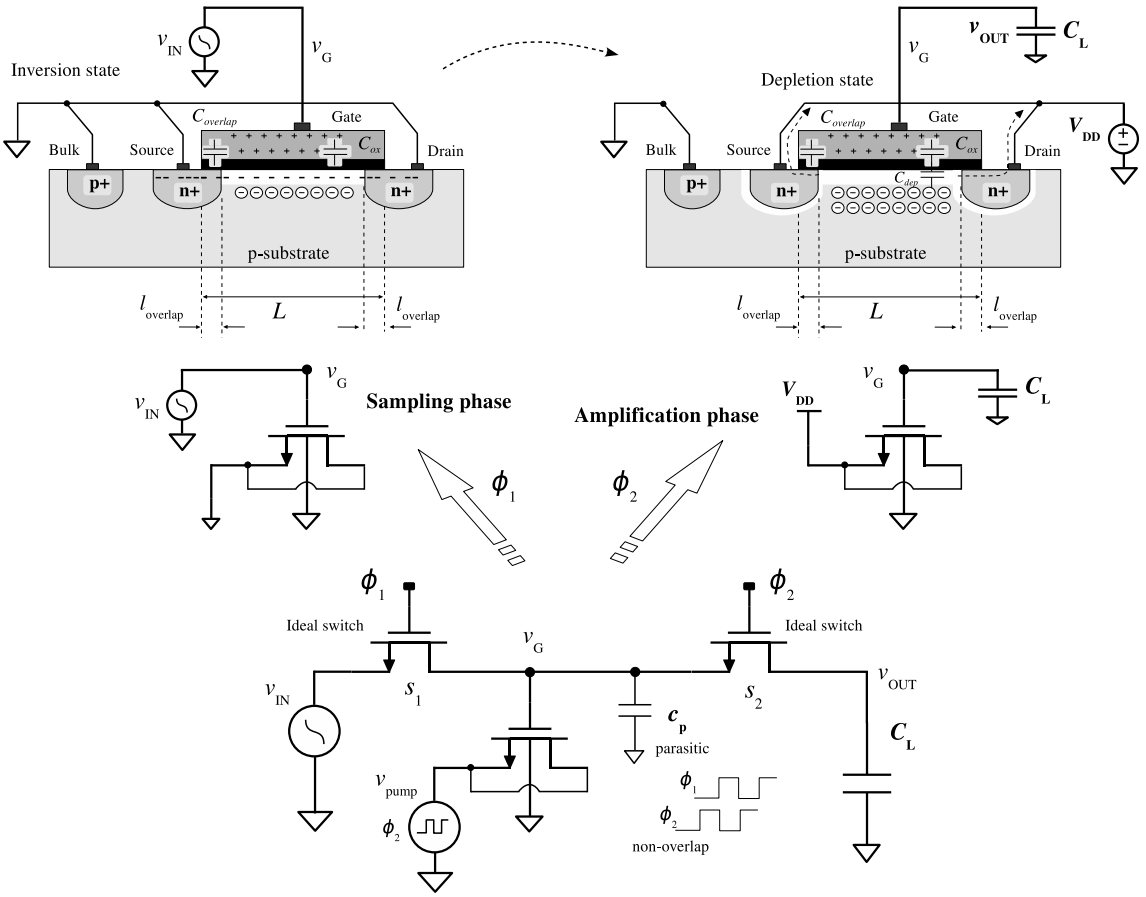


Figure 4.6: MOS parametric amplifier cell, based on [99].

MOS device is, in reality, a four-terminal device, [29]. It might be used to compose a three terminal configuration which can be further explored in the sense of separating the input signal from the terminal that modulates/controls the variations of the MOS capacitance. In [98], some proposals in that direction were presented, one of them being the use of a three terminal MOS varactor as a DT-PAMP. The work has been complemented with the CMOS integration of an 100 kHz signal amplifier with 3 μW power dissipation, [13, 99, 100]. The original configuration of a NMOS DT varactor based parametric circuit, [99, 100, 101], is shown in Figure 4.6, where the three terminals device is obtained by connecting directly the drain and source into only one D/S terminal and maintaining the remaining ones (gate and bulk). As in the usual S/H case, the discrete-time operation is achieved by the application of the clock signals to the input and output switches. Additionally, these clocks are also used as the pump signal that is applied to the D/S connection. The modulation of the device capacitance is then achieved by controlling the voltage at the D/S terminal. In fact, the design of the circuit is such that the clock signal is able to change the device state from inversion into depletion mode and vice-versa.

The principle of operation of this MOS Parametric Amplifier (MPA) cell, [13, 99, 100], is illustrated in Figure 4.6, where the bulk is always connected to ground.

During phase ϕ_1, the input signal is sampled by the total gate capacitance (including parasitic capacitances). Considering that the input voltage is high enough (at least higher than V_{th}) and the D/S terminal is connected to ground, a thin layer (sheet) of inverted charges (electrons in the NMOS case) is formed just below the gate/substrate interface. The device is then in strong inversion state if the input voltage is sufficiently higher than V_{th}, [29]. This conducting sheet extends over the gate region, connecting the two drain/source implant areas. This forms a kind of capacitor plate, while the other plate is the gate material. These two plates are separated by the oxide capacitance with a "sandwich" type structure. It is clear now that the total gate capacitance when the device is in strong inversion is approximately given by the oxide capacitance $C_{ox} = C'_{ox}WL$ plus two overlap parasitic capacitances c_{ov}, [29].

In phase ϕ_2, the sampling switch opens and the controlling terminal D/S is connected to the highest supply voltage, V_{DD}, forcing the transistor into the depletion operating region. In fact, by connecting the positive power supply of the D/S terminal, the inverted charges that sustained the channel during ϕ_2 are now pulled-out by this source, hence destroying the conduction sheet below the gate and letting this area depleted. Under these conditions the device enters into the depletion state, in which the gate to bulk (which is connected to ground) capacitance is roughly given by the series of the oxide capacitance C_{ox} and the capacitance associated with the depleted region which is usually small. Therefore, during ϕ_2, the total gate capacitance is significantly reduced.

Interestingly, if the gate is left floating during the initial instants of ϕ_2, the change of the gate capacitance occurs without affecting the amount of charge sampled in the gate plate. Due to charge conservation, the structure has to respond to this capacitance decrease with a gate voltage increase. Voltage amplification is then achieved. Naturally, parasitic capacitances and load capacitance will affect the overall gain, as verified in Section 4.2, thus requiring a careful design.

4.4 Analysis of the discrete-time MPA cell as an amplifier

To contribute for a better design one must determine the equations that govern the MPA circuit cell. The next subsections addresses this objective. The key for solving the circuit operation is the application of the principle of charge conservation at the

gate node in conjunction with the device operation model, [29].

4.4.1 Amplification Gain

The parametric operation of the amplifier relies on the way that the surface potential is affected by the competitive action generated by the voltage applied at the D/S terminal and the gate voltage. In other words, the surface potential is determined by the balance between the lateral and vertical electric fields that jointly shape the bulk charge density distribution inside the region near the oxide surface interface. The surface potential, ψ_s, is then defined as the potential change from the surface interface oxide/substrate and the electrically neutral substrate region, [29].

Due to the structural construction of the MOS device, the surface potential is not the only one that is involved in the voltage loop around the gate and bulk. This potential balance is given by,

$$v_{GB} = \psi_{ox} + \psi_s + \phi_{MS} \ , \tag{4.16}$$

where ψ_{ox} is the potential drop across the oxide and ϕ_{MS} is the sum of the contact potentials around the loop. In this case, ϕ_{MS} is just given by

$$\phi_{MS} = \phi_{\text{bulk material}} - \phi_{\text{gate material}} = -\phi_F - 0.56 \ [V] \ , \tag{4.17}$$

for a n^+ polysilicon gate, and ϕ_F is the Fermi voltage (determined by the thermal voltage ϕ_t, acceptor concentration N_A in case of p-type substrate and intrinsic carrier concentration n_i in a semiconductor,[29]).

A corresponding result can be established for the balance of charge across the same the MOS structure, since each charge has a potential associated with it. The total charge at the gate terminal, Q_G, can be expressed by

$$Q_G = -\left(Q_o + Q_I + Q_B\right) \tag{4.18}$$

which is related to the oxide potential through the oxide capacitance, in the form of $Q_G = C_{ox}\psi_{ox}$. Q_o is the parasitic charge within the oxide and oxide-substrate interface and is mainly originated during fabrication process and can be assumed to have a fixed value. This parameter is embedded in the so called Flat Band voltage, [29],

$$V_{FB} = \phi_{MS} - \frac{Q_o}{C_{ox}} \, . \tag{4.19}$$

The total charge in the substrate region below the gate, is built by the charge Q_I at

the thin/sheet inversion layer (if exists) and by the Q_B charge of the depletion region of depth X_d, which is also dependent on the surface potential. Furthermore, both Q_I and Q_B are, in reality, a function of the surface potential ψ_s and the source to bulk voltage, V_{SB}.

To simplify the analysis, lets consider the case where the capacitive load C_L and parasitic capacitance (namely due overlap) do not exist. Under these conditions, a sample of the input voltage v_{IN} is acquired at the gate during the tracking phase (ϕ_1), while the S/D terminal is connected directly to the bulk (i.e., $V_{SB} = 0$). Assuming that this sample voltage is high enough to put the MOS device in strong inversion, then

$$v_{IN} = \psi_{ox,\phi 1} + \psi_{s,\phi 1} + \phi_{MS} \,. \tag{4.20}$$

Similarly, during the amplification phase ϕ_2, the voltage at the gate is given by

$$v_{OUT} = \psi_{ox,\phi 2} + \psi_{s,\phi 2} + \phi_{MS} \,. \tag{4.21}$$

Meanwhile, the gate is left floating at ϕ_2 meaning that the charge Q_G is maintained and consequently ψ_{ox} is unchanged (note that this is only true if C_L and c_p are not considered). Reflecting this result in (4.20) and (4.21), the obtained transfer function is determined by

$$v_{OUT} = v_{IN} + \psi_{s,\phi 2} - \psi_{s,\phi 1} \,, \tag{4.22}$$

which implicitly determine v_{OUT} since the surface potential during ϕ_2 ($\psi_{s,\phi 2}$) is strongly dependent on v_{OUT} as indicated in (4.23) (the device is in depletion mode during this phase), [29, 99, 102].

$$\psi_{s,\phi 2} \approx \left(-\frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + v_{OUT} - V_{FB}} \right)^2 \tag{4.23}$$

In (4.23), parameter γ is the body coefficient, [29].

During ϕ_1 the D/S to bulk voltage is set to zero and the input voltage is high enough to put the device in strong inversion mode with a thin conductive layer well defined beneath the gate. In strong inversion the potential at the surface substrate/oxide is approximately determined by

$$\psi_{s,\phi 1} \approx \phi_0 + V_{SB} = \phi_0 \,, \tag{4.24}$$

where $\phi_0 = 2\,\phi_F + n\,\phi_t$ (n varies between 0 and 6 depending on the process, [29]).

Combining (4.22), (4.23) and (4.24), results in an explicit format for the v_{OUT} and v_{IN}

relation which may be expressed by

$$v_{OUT} = \frac{v_{IN}^2}{\gamma^2} + v_{IN}\left[1 - \frac{2}{\gamma^2}\left(V_{FB} + \phi_0\right)\right] + \frac{\left(V_{FB} + \phi_0\right)^2}{\gamma^2} - \phi_0 \ . \tag{4.25}$$

The quadratic form of (4.25) indicates a non linear relation and, therefore, to obtain a better insight, the classical approach is to obtain the small signal operation of the circuit around a known and fixed value. Assuming the input signal in the form $v_{IN} = V_{IN} + v_{in}$, where the first term represents the DC signal component and the second term is the input small signal variations, and replacing it into (4.25) results in

$$\begin{aligned} v_{OUT} = \underbrace{\frac{1}{\gamma^2}}_{a_2} v_{in}^2 + \underbrace{\left[\frac{2V_{IN}}{\gamma^2} + 1 - \frac{2}{\gamma^2}\left(V_{FB} + \phi_0\right)\right]}_{a_1} v_{in} \\ \underbrace{+\frac{V_{IN}^2}{\gamma^2} + \left[1 - \frac{2}{\gamma^2}\left(V_{FB} + \phi_0\right)\right] V_{IN} + \frac{\left(V_{FB} + \phi_0\right)^2}{\gamma^2} - \phi_0}_{a_0} . \end{aligned} \tag{4.26}$$

Note that the same result can be obtained by calculating the corresponding Taylor series. The analysis reveals that a constant voltage a_0 at the output, and that the linear component of the small signal gain is given by a_1. The latter is not dependent on the physical dimensions of the device but rather on the input DC voltage, oxide thickness and temperature. However, the total device area does play an important role to minimize the effects of the load, parasitic and overlap capacitances, as it will be shown in the following analysis.

Figure 4.7 shows the spice simulation results of the total gate capacitance and total gate charge when the D/S terminal of the NMOS is connected to ground (representing the sampling phase) and when this terminal is changed to a pump source V_{DD} (representing the amplification phase). The NMOS device area is $W = 14$ μm and $L = 1$ μm, and is modeled by a BSIM3v3 model of an 130 nm CMOS technology node. It is clear from Figure 4.7 that for each pump value, the ramp evolution of the total gate charge versus the gate voltage (v_{GB}) has two regions with different slopes. In the first one, it increases more slowly with v_{GB}, with a ramp slope given by the gate to bulk capacitance c_{gb} (since the device is in depleted mode), while the higher slope of the second region is dominated by oxide capacitance C_{ox} (since the device is now in inverted mode). It is also visible that the value of the voltage pump source shifts the curve to the right, meaning that the v_{GB} range for which the device is depleted has increased. As referred before, it is this capacitance change that is used to achieve the parametric

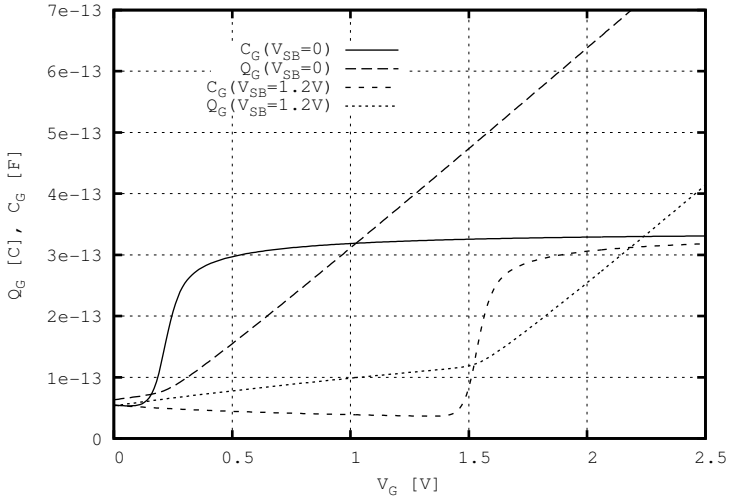


Figure 4.7: Total gate charge and capacitance for a pump source of 0 V and 1.2 V.

amplification, as indicated by

$$\frac{v_{out}}{v_{in}} = \frac{v_{gb,\phi 2}}{v_{gb,\phi 1}} = \frac{C_{ox}}{c_{gb}} = \frac{C_{ox}}{\frac{C_{ox}}{\sqrt{1+\frac{4}{\gamma^2}(V_{GB}-V_{FB})}}} = \sqrt{1+\frac{4}{\gamma^2}\left(V_{GB}-V_{FB}\right)} \tag{4.27}$$

which is valid for small signal variations and where the gate to bulk capacitance, c_{gb}, has been replaced by its relation with C_{ox}, [29].

The implicit linearization reflected in (4.27), results from a piecewise linear approach with two segments, which is demonstrated in Figure 4.8. A closer observation from the model depicted in Figure 4.8 highlights three distinct amplifier operating regions, [101]:

- Depleted region: if the input voltage is lower than V_{t0} (threshold voltage at V_{SB} = 0 V), then during the sampling phase the device is in depleted state, since the voltage is not high enough to form the inverted channel beneath the gate. The gate capacitance coincides with the gate to bulk capacitances which is small. During the amplification phase the device remains depleted and therefore the small signal voltage gain is one, i.e, $v_{out} = v_{in}$.
- Amplification region: if the input voltage is between $v_{IN,min} = V_{t0}$ and the transition point defined by $v_{IN,max}$, then during the sampling phase the device is inverted with the maximum capacitance at the gate, C_{ox}, reducing to the depleted gate to bulk capacitance during the following phase. Since the gate is is left floating and the charge is conserved between phases, the gate voltage has to increase, meaning that signal amplification is achieved.

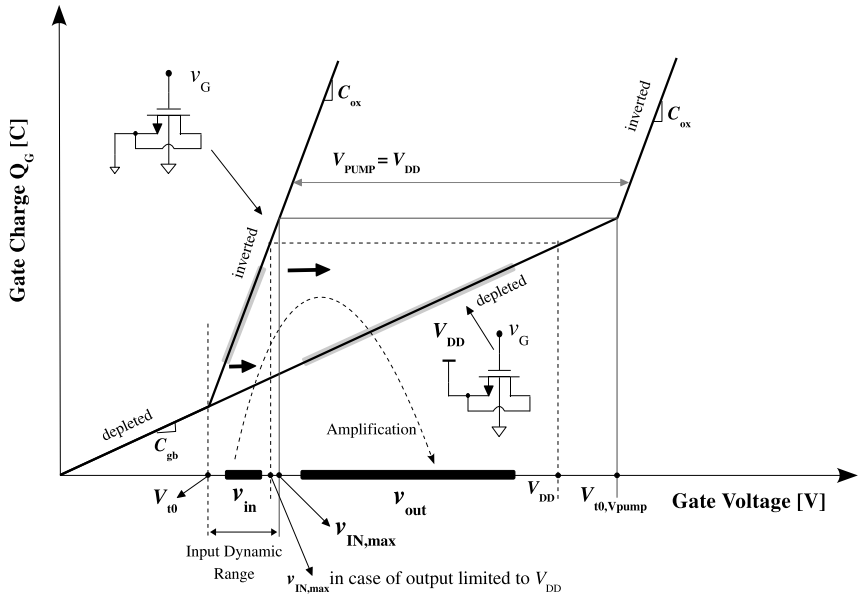


Figure 4.8: Model for the variation of the total gate charge, Q_G, for a pump source of 0 V and V_{DD}.

- Inversion region: if the output voltage tends to be higher than $V_{t0,Vpump}$, the vertical electric over inverted channel formed during sampling phase is strong enough to resist to the action of the lateral field (self-induced). As a consequence of this competitive action, the inverted channel does not disappear during the second phase and the device keeps itself in inverted state, with the same C_{ox} at the gate. In this case the output and input voltage are separated only by a fixed voltage equal to step value of the pump signal (V_{DD}, in this case). The small signal voltage gain is again equal to one.

These observations indicate that the MPA has a limited input range, which varies roughly from V_{t0} to $v_{IN,max}$, where it really acts as a signal amplifier. The threshold voltage at zero V_{SB} is approximately given by

$$V_{t0} = V_{FB+}\phi_0 + \gamma\sqrt{\phi_0} \ , \tag{4.28}$$

while $v_{IN,max}$ has to be found.

At $V_{\mathrm{t0,Vpump}}$, the charge stored in c_{gb} is the equivalent to the nonlinearly accumulated one at the gate capacitance with $V_{SB} = 0$, when the input reaches $v_{IN,max}$. On the other hand, previous results show that when the device keeps its inverted state, the input voltage is only shifted by an amount of V_{PUMP}. Then, both $v_{IN,max}$ and $V_{\mathrm{t0,Vpump}}$

are determined from

$$\begin{cases} c_{gb}V_{\mathrm{t0,Vpump}} = c_{gb}V_{t0} + C_{ox}\left(v_{IN,max} - V_{t0}\right) \\ V_{\mathrm{t0,Vpump}} - v_{IN,max} = V_{\mathrm{PUMP}} \end{cases} . \tag{4.29}$$

Under normal conditions, the $V_{\mathrm{t0,Vpump}}$, given by

$$V_{\mathrm{t0,Vpump}} = V_{\mathrm{t0}} + \frac{C_{\mathrm{ox}}}{C_{\mathrm{ox}} - c_{\mathrm{gb}}}V_{\mathrm{PUMP}} = V_{\mathrm{t0}} + \frac{p}{p-1}V_{\mathrm{PUMP}}\,, \tag{4.30}$$

is usually higher that V_{DD} which has to be handled carefully during the design process. The intrinsic parametric gain factor p used in (4.30) is given by

$$p = \frac{C_{ox}}{c_{gb}}\,. \tag{4.31}$$

The corresponding maximum input voltage that assures that during the amplification phase the device will not enter in inversion state is given by $v_{IN,max}$

$$v_{\mathrm{IN,max}} = V_{\mathrm{t0}} + \frac{c_{\mathrm{gb}}}{C_{\mathrm{ox}} - c_{\mathrm{gb}}}V_{\mathrm{PUMP}} = V_{\mathrm{t0}} + \frac{1}{p-1}V_{\mathrm{PUMP}}\,. \tag{4.32}$$

However, preventing the gate voltage to go higher than the power supply is reflected by choosing a lower input signal range, and (4.32) is reduced to (4.33).

$$v_{\mathrm{IN,maxVDD}} = V_{\mathrm{t0}} + \frac{V_{\mathrm{DD}} - V_{\mathrm{t0}}}{p} \tag{4.33}$$

Both (4.27) and (4.26) are only valid within the amplification zone, i.e., for the input voltage range roughly defined between V_{t0} and $v_{IN,max}$. Using both equations to determine the expected intrinsic gain that can be achieved with an 130 nm digital CMOS technology allow us to plot the curves shown in Figure 4.9. The difference between them is visibly small thus validating the use of the simplified approach represented by (4.27).

The gain expressed by (4.27) does not depend on the width and length of the device, i.e., it is independent of the total device area size. However, it is strongly and nonlinearly dependent on the oxide thickness through $\gamma = t_{ox}\cdot\sqrt{2qN_A\varepsilon_S}/\varepsilon_{ox}$ body factor, making it very tied to the used technology node. This points out that the t_{ox} reduction associated with the evolution of the CMOS technology into nanoscale range, see Chapter 2, tends to improve the intrinsic achievable intrinsic parametric gain (p) as demonstrated in Figure 4.10, which was obtained with the data from another foundry[1] (IBM) gathered

[1]Results shown in Figure 4.9 were not obtained using IBM 130 nm CMOS technology node data.

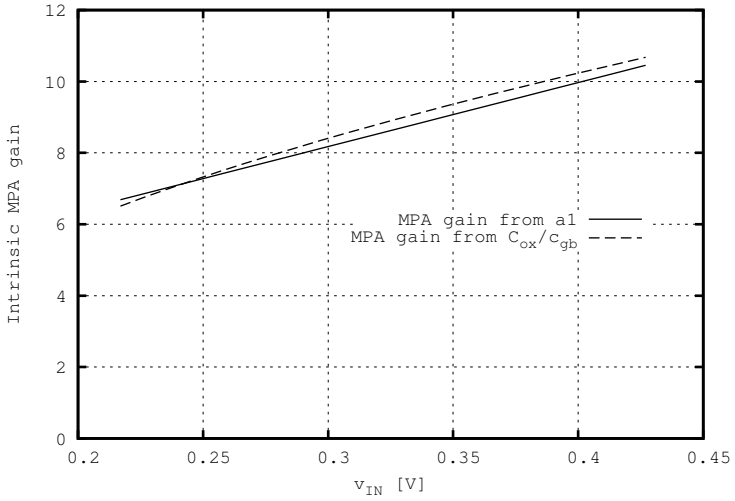


Figure 4.9: Comparison of the intrinsic parametric amplification gain obtained from (4.27) and (4.26).

in Table 2.3. Nevertheless, other small size effect, e.g. gate leakage (see Chapter 2), counteracts this gain improvement. The analysis done so far does not include the effect

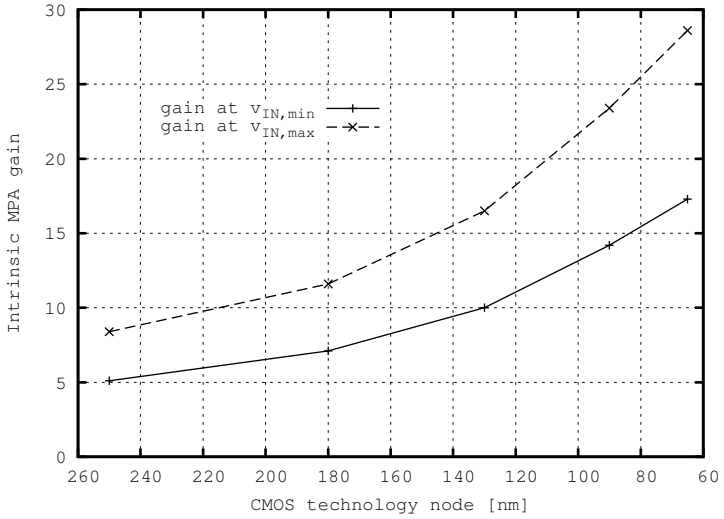


Figure 4.10: MPA cell gain calculated using data from nanoscale IBM CMOS technology.

of any extrinsic and load capacitors. Disregarding the high frequency effects, at this phase, the most important extrinsic capacitances to count for are the load capacitor C_L, the parasitic capacitance c_p (which represents, the total of internal and external parasitics from the gate to the ground, e.g, due contacts or routing lines) and the overlap capacitance associated with the MOS device, c_{ov}. Figure 4.11 clarifies their location. Including these extrinsic capacitors in the gain model is important since they will affect the effective gain value, distortion and speed achieved by the original

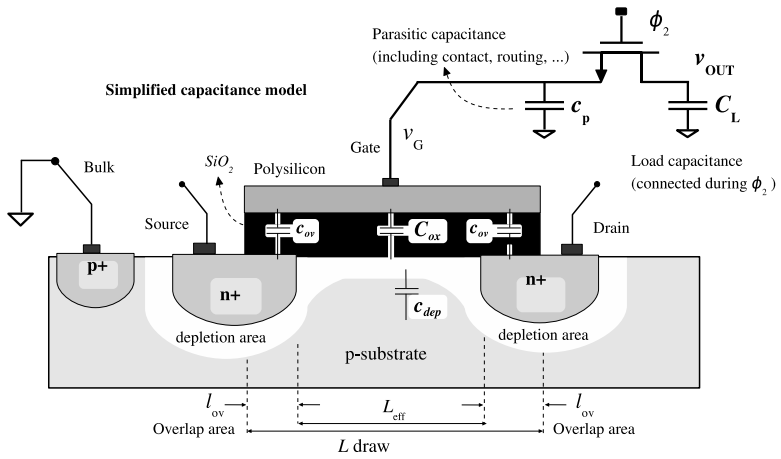


Figure 4.11: Extrinsic capacitance model of the MPA cell.

intrinsic MPA cell, specially if the device is very small in size. To model their effects on the output signal, each individual contribution is determined from the first order small signal capacitance model depicted in Figure 4.11.

Accounting only the parasitic capacitance c_p results in (4.34) which indicates a gain reduction due to the charge sharing that occurs with the gate capacitance.

$$\frac{v_{out}}{v_{in}} = \frac{v_{gb,\phi 2}}{v_{gb,\phi 1}} = \frac{C_{ox} + c_p}{c_{gb} + c_p} \quad < \quad \frac{C_{ox}}{c_{gb}} \tag{4.34}$$

Proceeding in the same way for the load capacitor C_L, results in

$$\frac{v_{out}}{v_{in}} = \frac{v_{gb,\phi 2}}{v_{gb,\phi 1}} = \frac{C_{ox}}{c_{gb} + C_L} \quad < \quad \frac{C_{ox}}{c_{gb}}. \tag{4.35}$$

The comparison between (4.34) and (4.35) indicates that for the same value of c_p and C_L, the latter originates greater gain reduction, see Figure 4.13. This is explained by the fact that c_p samples the input signal as well.

The third extrinsic capacitor type to be included in the model represents the overlap capacitances due to the overlap area between the gate and D/S diffusion areas. Note that the overlap capacitance exists both at the drain and source and, therefore, has to be accounted twice. However, in this analysis, the associated outer and top fringing capacitances, [29], are not included. Applying the same charge conservation methodology for all extrinsic capacitances included in the model, results in a closed form for v_{out},

$$v_{out} = \frac{p \cdot \frac{2c_{ov}}{C_{ox}}}{1 + p\left(\frac{c_p}{C_{ox}} + \frac{C_L}{C_{ox}} + \frac{2c_{ov}}{C_{ox}}\right)} V_{PUMP} + \frac{p \cdot \left(1 + \frac{c_p}{C_{ox}} + \frac{2c_{ov}}{C_{ox}}\right)}{1 + p \cdot \left(\frac{c_p}{C_{ox}} + \frac{C_L}{C_{ox}} + \frac{2c_{ov}}{C_{ox}}\right)} v_{in}\,, \tag{4.36}$$

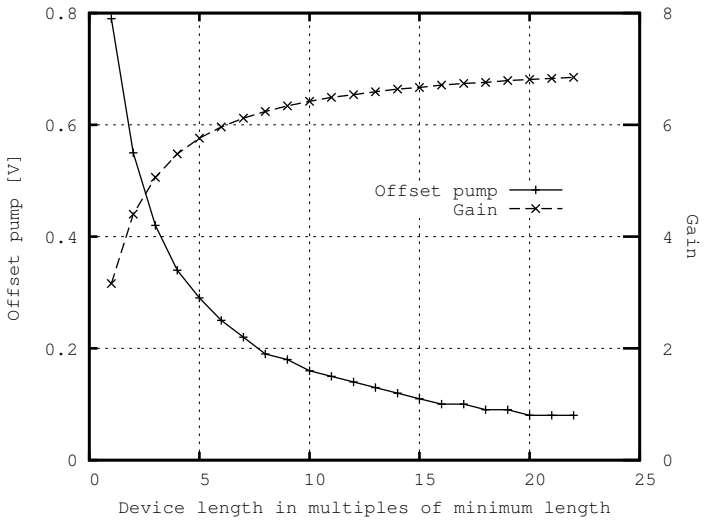


Figure 4.12: The impact of the device length on the MPA cell gain.

where p is the intrinsic MPA cell gain already defined.

A closer look to (4.36), reveals that, besides lowering the total gain, the overlap capacitance is responsible for an offset injection which is proportional to the voltage pump. Under the restriction of keeping the gate voltage below V_{DD}, this offset jump reduces even further the input dynamic range of the amplifier. This has to be addressed during the design process. The c_{ov}/C_{ox} ratio reduces to the relation between the overlap extent and the effective device length $L_{eff} = L - 2\,l_{ov}$, as shown in (4.37), given that the overlap capacitance is approximated by $c_{ov} \approx l_{ov} C'_{ox} W_{eff}$ (a more detailed expression can be found in [29], pg. 408).

$$\frac{c_{ov}}{C_{ox}} \approx \frac{l_{ov}}{L_{eff}} \tag{4.37}$$

Consequently, the cell total gain depends on the device area and, particularly on the length of the device (assuming that the overlap length is fixed for a given technology). The curves depicted in Figure 4.12 represents the offset pump and the voltage gain, only due to the overlap capacitance (i.e., $c_p = C_L = 0$), against the total device length (given in multiples of the minimum length). The results show that the minimum length device should be avoided since both the offset pump is increased and the voltage gain is significantly affected. However, a total length higher than 7 times the minimum length, permits to reach 90% of the gain obtained when this overlap capacitance is not considered. A cell configuration improvement to reduce the effect of the overlap capacitance is proposed in Subsection 4.4.5.

To complement the above analysis the individual and the joint effects of the three type of extrinsic capacitors have been calculated and represented in Figure 4.13. The

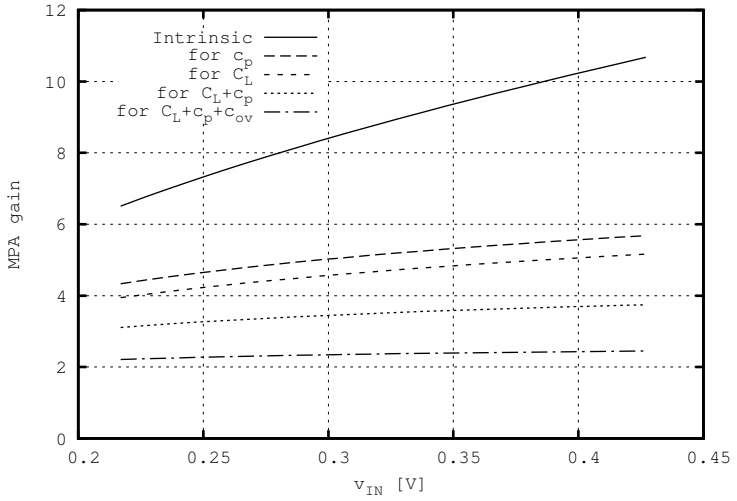


Figure 4.13: MPA gain obtained for different combination of parasitic, overlap and load capacitances, in an 130 nm CMOS technology.

technological parameters of an 130 nm digital CMOS technology has been used and for a better comparison it was considered that both C_L and c_p are 10% of C_{ox} (which is not always true). For the c_{ov} is was assumed a minimum length device, corresponding to a 13% C_{ox} of overlap capacitance.

Analyzing Figure 4.13, one can highlight that, despites the reduction on the gain value, the joint action of C_L and the other capacitances allows a gain flattening over the input dynamic range. The latter is an interesting results for application where this MPA cell can be used as a low gain open loop amplifier. That is, by controlling the amount of the load capacitance (in a certain extent) can be used to improve the gain accuracy, in a digital assisted analog calibration configuration. As a partial conclusion, the maximum gain achieved by the MPA cell is limited by the intrinsic one which is dependent on technology and common mode voltage but not on the device size. On the other hand, the effective gain is determined by the relative size of the parasitic, load and overlap capacitance, with respect to the total oxide capacitance. Also the overlap capacitance puts some restrictions when sizing the MPA device length.

A simple NMOS MPA cell has been simulated in the 130 nm technology with a device size of $W = 14$ μm and $L = 2$ μm. The SPICE simulations used a BSIM3v3 model and all the signal and clock sources as well as the switches are considered ideals in order to not contaminate the results of the MPA cell by itself. A clock signal of 25 MHz and a pump voltage equal to $V_{DD} = 1.2$ V are used. The input signal is a ramp varying between 0 V and $V_{DD} = 1.2$ V.

Figure 4.14(a) presents the simulation for the unloaded case, i.e., without C_L. The

incremental gain ($^{\triangle v_{out}}/_{\triangle v_{in}}$) inside the amplification zone is around $^{512\,\mathrm{mV}}/_{73\,\mathrm{mV}} \approx 6.5$, while outside that area it approaches the unit. The simulated gain is very close to the predicted by the previous analysis. The observed offset comes from the MPA operation by itself but also from the overlap (coupling) capacitor. This effect is clear on the depletion region of the curve. A zoom inside the amplification area is shown in Figure 4.14(b), in which one can observe the sampling and the amplification phases of the MPA operation.

The effect of a 40 fF load capacitor is depicted in 4.14(c). Although the gain is reduced to $^{\triangle v_{out}}/_{\triangle v_{in}} = {}^{606\,\mathrm{mV}}/_{150\,\mathrm{mV}} \approx 4$ (which is close to the predicted one), it confirms that the input range for which the MPA cell is able to amplify has correspondingly increased.

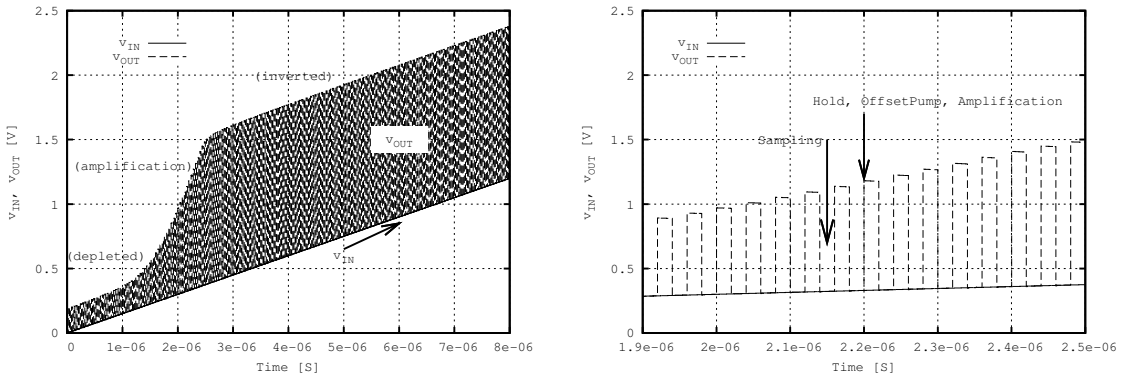
In order to better check the coupling effect due to the overlap capacitance, the device length was reduced to the minimum length permitted by the referred 130 nm CMOS technology. However, to maintain the same C_{ox}, from previous simulations, the width of the device has increased accordingly to $W = 200$ μm. The simulation shown in 4.14(d) confirms that, in this case, the overlap coupling is responsible for a high offset at the gate output voltage, thus virtually vanishing the parametric amplification.

4.4.2 Harmonic distortion

The quadratic form of the output voltage obtained for the intrinsic MPA cell operating inside the amplification region (without C_L, c_p and c_{ov}), reflected by (4.25), suggests that the circuit is intrinsically and only affected by second harmonic distortion, when in presence of a pure sinusoidal input signal. Following the approach of [102] and the definition of the harmonic components, the second harmonic component can be determined by

$$HD2 \approx \frac{V_{vin,A}}{2} \cdot \left|\frac{a_2}{a_1}\right| = \frac{V_{vin,A}}{2} \cdot \frac{\frac{1}{\gamma^2}}{\frac{2V_{IN}}{\gamma^2} + 1 - \frac{2}{\gamma^2}\left(V_{FB} + \phi_0\right)} \tag{4.38}$$

where $V_{vin,A}$ is the input signal amplitude and a_1, a_2 are the first order and second order coefficient terms, respectively. The traditional constant capacitance S/H circuit, when operating in an unloaded conditions, does not add additional harmonics beyond those associated with the sampling process itself. What (4.38) shows is that for the unloaded case, the MPA differs from the previous traditional S/H not only because of the gain but also due to the second harmonic. The latter is independent of the device size but strongly depends on the used technology and common mode voltage. This second harmonic comes from the gate capacitance change from an approximately



(a) Transient for $C_L = 0$

(b) Transient for $C_L = 0$: zoomed view

(c) Transient for $C_L = 40$ fF

(d) Transient for minimum length device,$C_L = 0$

Figure 4.14: SPICE simulations obtained with NMOS MPA cell, in an 130 nm CMOS technology.

constant capacitance (C_{ox}), during sampling, to a depletion capacitance whose value is nonlinearly determined by the final gate voltage during the amplification phase.

However, a more detailed analysis shows that the total gate capacitance when in strong inversion depends slightly and non-linearly on the applied voltage, due to the dependence of the surface potential with respect to the gate to bulk voltage V_{GB}, [29, 31, 102]. Hence, (4.24) is adapted into (4.39) and, as a consequence, the output voltage signal v_{OUT} of the intrinsic MPA cell will show up high order harmonics other than the second.

$$\psi_{s,strong,\phi 1} \approx 2\phi_F + \phi_t \cdot F_{ns}\left(V_{GB}, V_{SB}\right) + V_{SB} \tag{4.39}$$

The inclusion of the load capacitance C_L as well as c_p and c_{ov}, is also responsible for the arising of higher order harmonics as a result of the charge sharing with the non-linear MOS device capacitance, even when (4.24) is considered instead of (4.39). However, this does not necessarily implies an higher total harmonic distortion (THD). In fact, a constant value capacitor (represented by c_p, in this model case) which is connected at both sampling and amplification phases widen the gap between the second harmonic and the fundamental frequency amplitudes. Naturally, these load capacitances reduces the overall linear gain, as seen before.

Adapting for capacitance c_p to the analysis made in [102], where a third order Taylor expansion was used to determine the harmonic distortion, results in

$$HD2 = \frac{k_2 k_3^2 V_{in,A}}{8\left(k_3 V_{IN} + k_4\right)\left(2k_0\sqrt{k_3 V_{IN} + k_4} - k_2 k_3\right)} \tag{4.40}$$

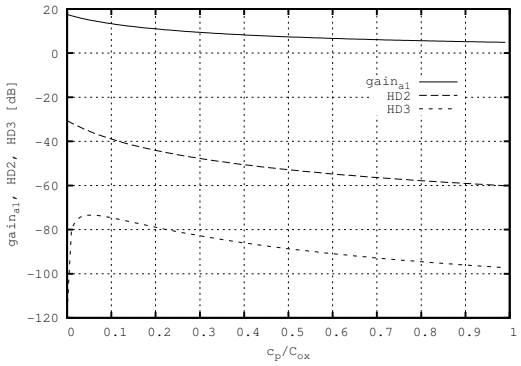
for the second order harmonic while for the third one the outcome is

$$HD3 = \frac{k_2 k_3^3 V_{in,A}^2}{32\left(k_3 V_{IN} + k_4\right)^2\left(2k_0\sqrt{k_3 V_{IN} + k_4} - k_2 k_3\right)} \ . \tag{4.41}$$

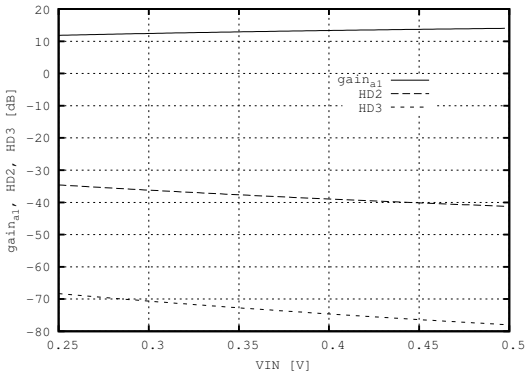
In (4.40) and (4.41), the $k_{0,1,2,3,4}$ factors are determined under a charge conservation approach and are expressed by

$$\begin{cases} k_0 = 1 + \frac{C_{ox}}{c_p} \\ k_1 = \frac{C_{ox}}{c_p}\frac{\gamma^2}{2}\left(1 + \frac{C_{ox}}{c_p}\right) - \frac{C_{ox}}{c_p}\left(V_{FB} - \phi_0\right) \\ k_2 = \frac{\gamma}{2}\frac{C_{ox}}{c_p^2} \\ k_3 = 4c_p\left(c_p + C_{ox}\right) \\ k_4 = \gamma^2\left(C_{ox} + c_p\right)^2 - 4V_{FB}\left(c_p^2 + c_p C_{ox}\right) - 4c_p C_{ox}\phi_0 \end{cases} . \tag{4.42}$$

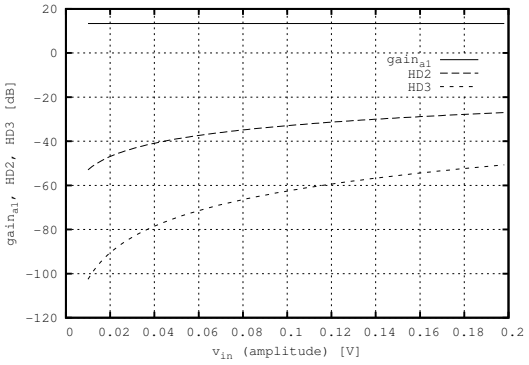
Figure 4.15 presents a set of harmonic distortion results obtained from the above analytical model, using the transistor parameter data available from the BSIM3v3 model for the 130 nm CMOS technology. The transistor size is considered to be 14 μm wide and 2 μm long and the effects of the variation have been calculated separately for c_p, input common mode voltage, V_{IN} and input signal amplitude $V_{vin,A}$. The results



(a) Harmonic distortion *versus* c_p/C_{ox}, for $V_{IN} = 0.4$ V and $v_{in,amplitude} = 50$ mV



(b) Harmonic distortion *versus* V_{IN}, for $c_p/C_{ox} = 0.1$, and $v_{in,amplitude} = 50$ mV



(c) Harmonic distortion *versus* $v_{in,amplitude}$, for $c_p/C_{ox} = 0.1$, $V_{IN} = 0.4$ V

Figure 4.15: Harmonic distortion analysis, using parameters from a 130 nm standard CMOS technology.

obtained from this simplified analytical model predicts that:

- the charge sharing between the MPA intrinsic cell and the load fixed capacitor c_p originates third order harmonic distortion;
- the load capacitance c_p reduce the gain but, at the same time, it has a positive impact on the total harmonic distortion;
- within the amplification region of the amplifier, a small improvement can be achieved by using a higher common mode voltage, V_{IN};

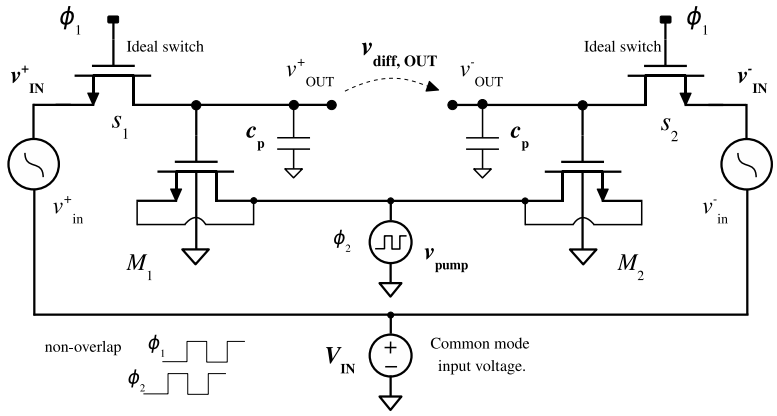


Figure 4.16: Differential structure for the MPA cell based amplifier.

- to keep $HD3$ below -50 dB, the input signal amplitude must be smaller than 200 mV.

The distortion analysis reveals that, under certain conditions, the dominant second order distortion can be quite high in the MPA cell (can reach -20 dB), while the third one, $HD3$, is generally at least 30 dB below the second one. The total harmonic distortion (THD), can be approximately determined by,

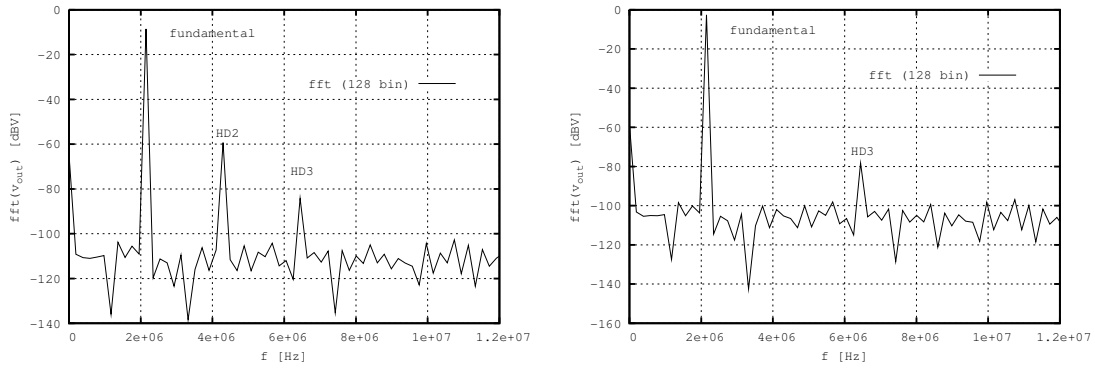
$$\text{THD} \approx 10\log\left(HD2^2 + HD3^2\right) = 20\log\left(HD2\right) + 10\log\left(1 + \alpha_{HD32}^2\right) \ , \qquad (4.43)$$

where the relation between second and third harmonic is defined by,

$$\alpha_{HD32} = \frac{HD3}{HD2} = \frac{V_{in,A}}{4\left(V_{IN} + \frac{k_4}{k_3}\right)} \, . \qquad (4.44)$$

This last result, indicates a gap between the third and the second harmonic that depends on the input signal amplitude amplitude and its common mode DC component (since for the used technology k_4/k_3 tends to be much lower than V_{IN}). The results obtained in [99] suggests a similar trend. It also reveals that the second harmonic is the dominant one, since α_{HD32} is lower than one (for the used CMOS technology). Hence a significant improvement is expected if a pseudo-differential structure is used which is well known by its ability to eliminate distortion coming from even order harmonics. This is confirmed by the spectrum composition of the output signal, shown in Figure 4.17, obtained from the FFT calculation over a transient SPICE electrical simulation with $W/L = 14/2$ μm MOS devices. It is observed that the -60 dB $HD2$ in the single ended case was completely vanished in the differential scenario. In practice, any mismatch between the two branches weakens the even harmonics cancellation effect of the differential structure.

The FFT was determined from coherently sampled data, which is a method that increases the spectral resolution and eliminates the need for window sampling. Coherent sampling defines the sampling of a periodic signal, where an integer number of cycles fits into a predefined sampling time window. In order words, this can be defined as $f_{in}/F_s = M/N$, where f_{in} is the input signal frequency, F_s the sampling/clock frequency, M is an integer number of signal periods within the sampling window and N, the number of data points in the sampling window or FFT. By choosing M to be prime, one ensure that the samples within the unit time interval are unique and the N samples are then repeatable. For efficient FFT calculation (performed using MATLAB script), N is usually chosen to be a power of two. Applying the previous strategy, the FFTs of Figure 4.17 are calculated from $N = 128$ data points, and obtained from a transient SPICE simulation longer than a sampling time window of $5.12\mu s$, corresponding to $M = 11$ (prime number) signal periods of $f_{in} = 2.1484375$ MHz input signal, which was sampled at $F_s = 25$ MHz. The MPA simulated circuit includes 14 μm by 2 μm NMOS device, a 300 fF (corresponding to $c_p/C_{ox} \approx 1$) load capacitor c_p and all switches and sources are ideal ones. The input signal has an amplitude of 200 mV over a DC level of 0.6 V. It is clear from previous analysis and simulations that



(a) Harmonic distribution for single MPA cell (b) Harmonic distribution for a MPA differential structure

Figure 4.17: 128 point based FFT obtained from SPICE simulations data, in a 130 nm CMOS technology.

the differential structure better deals with resolutions higher than 8 bits. Therefore, during this work this type of topology is exhaustively employed.

The results presented in Figure 4.17 are valid under the capacitance model number 2 (BSIM3v3 parameter $CAPMOD = 2$). By using this capacitance model 2 it is granted the continuity of all the capacitance derivatives and that the surface potential is constant, [31]. This option is compatible with the analytical model presented, which is confirmed by comparing the results obtained from both approaches, see Table 4.1.

Type	Gain [dB]	HD2 [dB]	HD3 [dB]
FFT from SPICE simulations	4.99	-50.78	-75.67
Analytical Model	5.43	-50.88	-77.77

Table 4.1: Comparative harmonic distortion results.

However, a more realistic simulation should set $CAPMOD$ option parameter to level 3, specially for nanometer scale technology. Relatively to the second capacitance model level, this third model enters into account with quantum mechanical effects (channel carriers do not reside right on the oxide/semiconductor interface, but at a certain distance deep into the Silicon) and it also correctly calculates the surface potential as function of biases, [31]. As a consequence the final total distortion is higher and a difference as high as 10 dB might be expected, depending the input common mode voltage and input signal amplitude.

4.4.3 Time response (speed)

Several timing races occur during one operation cycle of the MPA cell and, therefore, the minimum time needed for the circuit to work is a combination of those individual contributions. For example, after the sampling phase the D/S terminal of the MOS device is subjected to a step voltage $V_{PUMP} = V_{DD}$ to pull out the charges from the previously formed inverted channel. The amplification time can be determined by the time needed to settle the D/S node to V_{DD} together with the time required to remove the channel since charges will need some nonzero time to flow into the D/S node. A qualitative representation of the charge balance between gate (Q_G), channel (Q_I) and bulk (Q_B), is outlined in Figure 4.18. It is shown that, since the gate charge is maintained at constant value during the amplification phase, the removal of the channel charges has to be compensated by the increase of the bulk charge. It is well known that the charge variations due to changing the electrode voltages in a MOS structure can be represented by a capacitor dynamic model schematic, [29, 27, 31, 30], which is partially depicted in Figure 4.19. By inserting this capacitance model into the MPA cell, it becomes possible to estimate the total response time using an RC time constant approach. Referring to Figure 4.19, the oxide material beneath the gate terminal constitutes an effective isolation layer with a total capacitance value $C_{ox} = C'_{ox} \cdot WL$. However, the access to this capacitance greatly depends on the device operating bias point, which depends on the existence of the channel and the depletion layer beneath the gate area. When the channel is formed, it shields the gate from the bulk and therefore the total capacitance from the gate to the D/S terminal is roughly given by C_{ox}. In this case the depletion layer that isolates the thin channel from the bulk

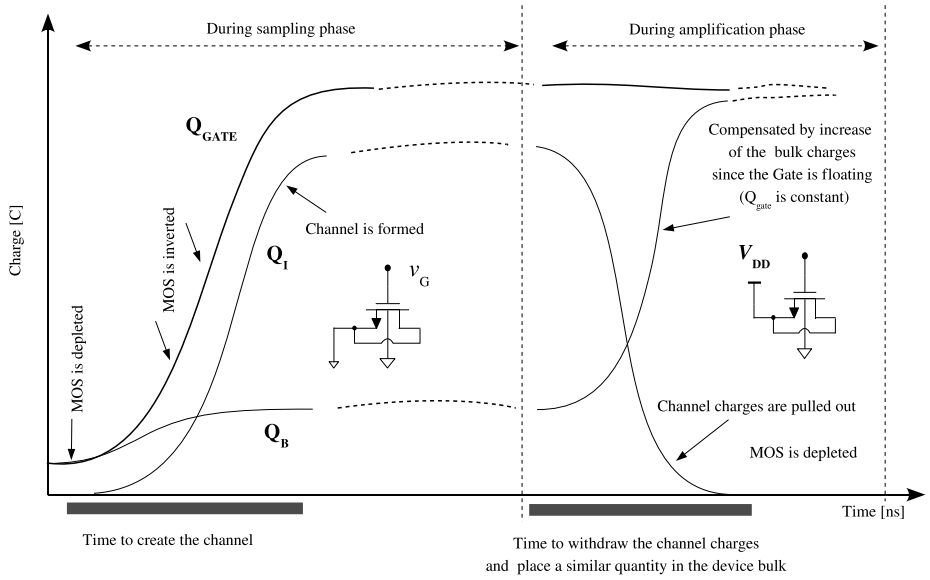


Figure 4.18: MPA cell time response model.

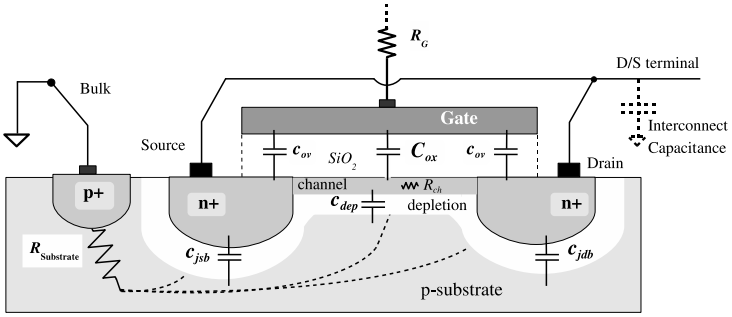


Figure 4.19: MOS capacitance model.

behaves like a junction type of capacitor, c_{dep} which is usually split 50/50 between the drain and source parasitic capacitance model (considering linear region of operation), [29]. The c_{dep} is approximately given by, [30],

$$c_{dep} = \frac{\varepsilon_{si}}{x_d} WL \ , \tag{4.45}$$

where ϵ_{si} is the permitivity of the Silicon, and the x_d is the depth of the depletion layer, whose value depends on the surface potential as shown in Equation 4.46.

$$x_d = \sqrt{\frac{2\varepsilon_{si}}{qN_{sub}} \left|\psi_s - \phi_F\right|} \ . \tag{4.46}$$

When the MOS device is in depletion mode, the channel is not formed thus removing the connection between C_{ox} to the D/S terminal. In this case the total capacitance seen at the gate is approximately given by the series of C_{ox} and c_{dep} which is much lower than C_{ox}.

It is also clear from Figure 4.19, that the source and drain areas form a kind of reverse-biased PN junction with the substrate. Therefore, standard junction capacitances c_{jsb} and c_{dsb} are added to the model, which are determined by the area and perimeter of the source and drain regions, [29]. For the 130 nm technology used so far, these are in range of a few fF.

It has already been shown that the overlap capacitances do play an important role on the gain performance of the MPA cell, justifying their inclusion on the device capacitance model. They are caused by the lateral material diffusion during fabrication, extending the source and drain regions underneath the gate electrode by some amount l_{ov}. The overlap capacitance is roughly given by, [30],

$$c_{ov} \approx \frac{\varepsilon_{ox}}{t_{ox}} W l_{ov} \, . \tag{4.47}$$

Many other capacitances associated with the MOS device have to be added to the model specially if the device is to be used in the GHz range. A more detailed analysis can be found in [29, 27, 31, 30].

Contributing not only to the MPA cell transient response but also to the overall noise budget, both channel and substrate resistances have to be included as well. When the drain is directly connected to the source terminal, it can be proven, [74, 75, 73], that by applying the Telegraphers, [74], equation to the channel extent, the equivalent channel resistance is

$$R_{eq,ch} = \frac{R_{ch}}{12} = \frac{1}{12} \cdot \frac{L}{k_n \cdot W \left(V_{GS} - V_{th}\right)} \ . \tag{4.48}$$

As a primary conclusion when the MPA cell is in strong inversion, the impedance seen between the gate and the D/S nodes is the series of the total oxide capacitance with a resistor $R_{ch}/12$.

A resistive path exists between the intrinsic MOS body node and the substrate contacts, which is generally described by a resistive network. This is a complex task due to the distributed and bi-dimensional nature of the physical structure since each point of the device sees multiple, unequal paths through the substrate to the contacts, see Figure 4.20. Several models have been proposed to model this resistive network [103, 27, 104, 105, 106, 107, 108, 109], reflecting a strong dependence on the number of fingers and layout configuration. Among these, a ring of substrate contacts around the device tends to produce a smaller substrate resistance. One common conclusion among the substrate resistive networks proposed in the literature, is that this substrate resistance is weakly dependent on the operating bias point.

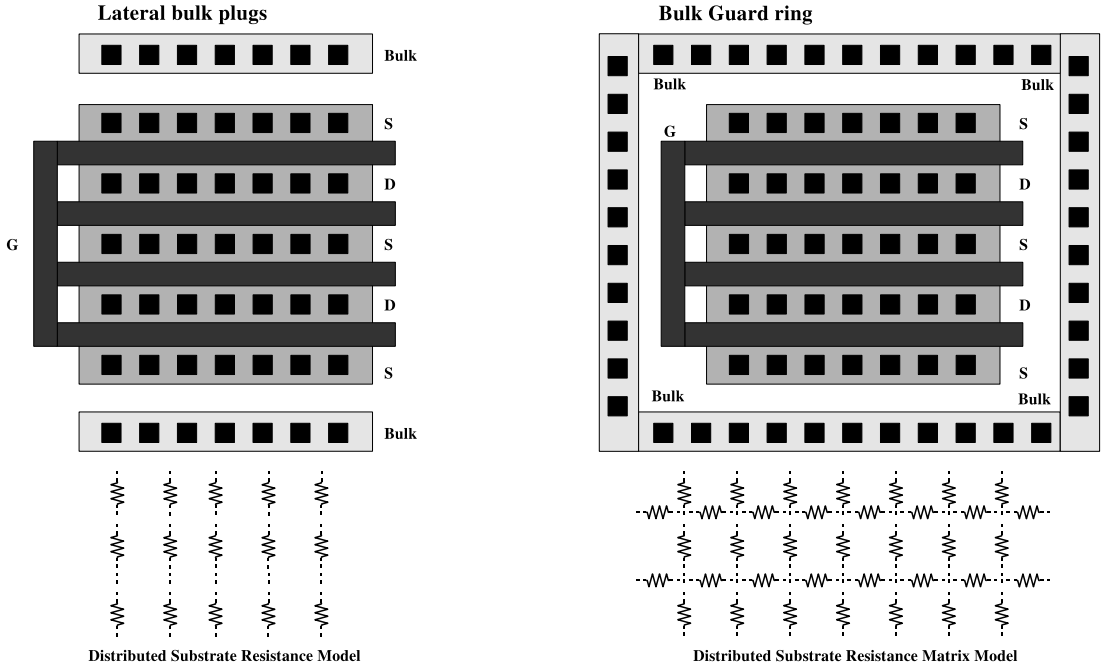


Figure 4.20: Layout schematic for substrate resistance model.

Next, the settling time estimation analysis is presented. Starting from a zero charged gate case and, assuming that the initial input voltage is high enough to settle the MOS device into strong inversion mode, one can identify some of the most important time delays presents in the overall process. In fact, during the sampling phase, the following timings have to be taken into account:

- time needed for the D/S to go from V_{DD} to ground, which can be modeled by a

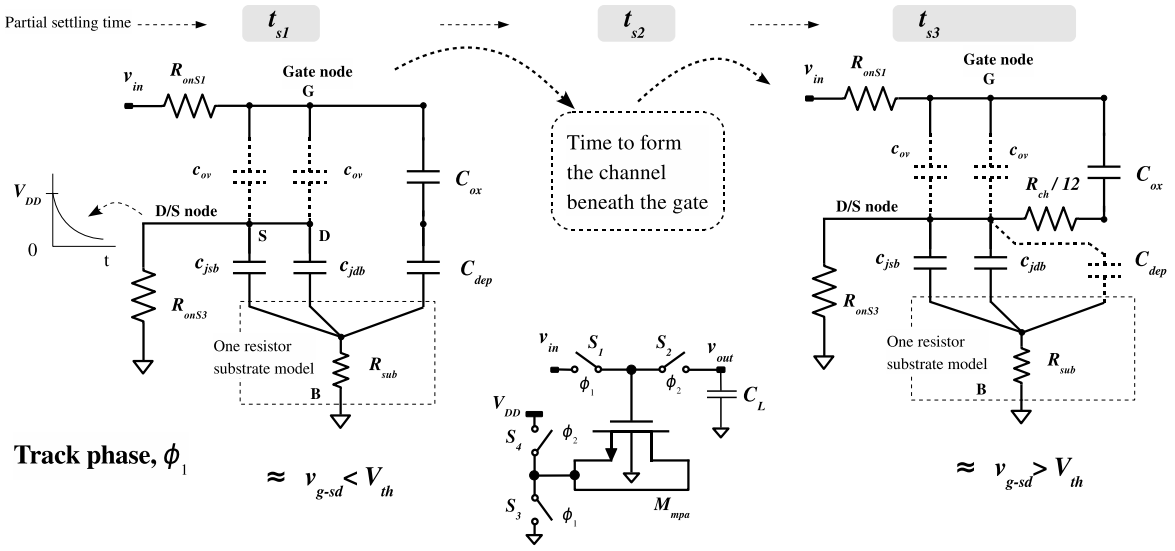


Figure 4.21: MPA cell during track time period.

RC network formed by the node parasitic capacitance and the ON-resistance of the pulling switch;

- time needed to move charges inside the bulk during the initial instants when the MOS device is still in depletion and after being inverted, the time needed to fill up the sheet of inverted charges beneath the gate (to form the channel). A model similar to [110], can be used to estimate this time slice.

- the settling time associated with the input RC network, dependent on the ON-resistance of the input switch and the total capacitance at the gate node which includes C_{ox} (assuming that the MOS device moves quickly into inversion mode) and the load capacitances c_{ov} and c_p. This is greatly determined by the MOS device size and the width of the input sampling switch;

Meanwhile, the timings during hold/amplification phase have to include:

- time required for the D/S terminal to go from ground to V_{DD}, which can be modeled by an RC network formed by the node parasitic capacitance and the ON-resistance of the pulling switch;

- time need to remove the inversion charges from the channel, i.e., time needed to move the device from strong inversion into depletion.

Figures 4.22 and 4.21 details all the referred timings during tracking and amplification phases, respectively.

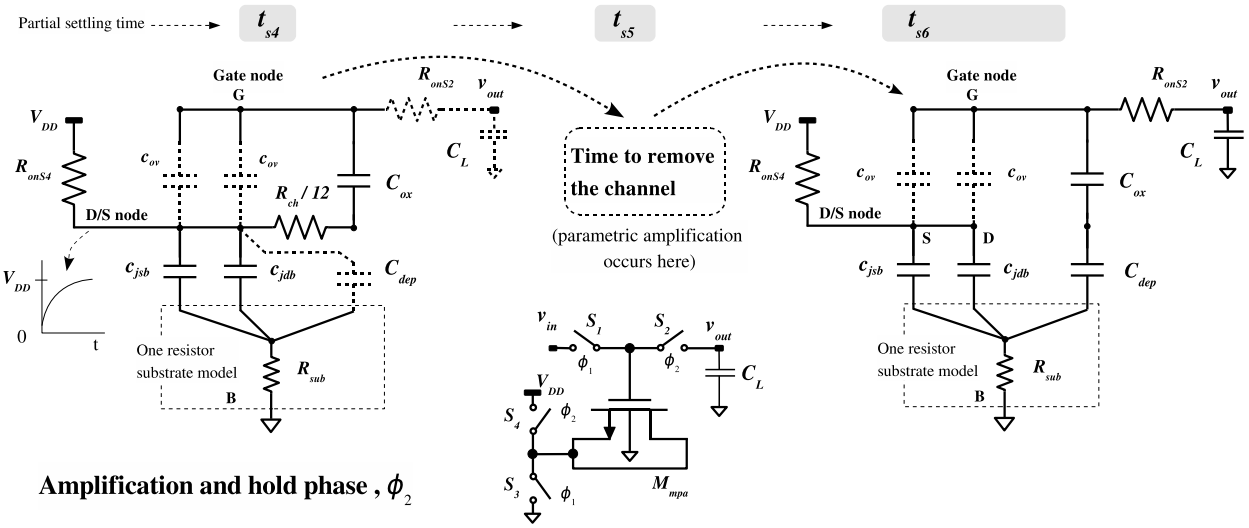


Figure 4.22: MPA cell during amplification and hold time period.

One of the referred timing terms reflects the channel removal which is attained by pulling up the D/S terminal to V_{DD}. The application of this voltage, causes a potential well to form at the D/S terminal and the electrons in the vicinity of this Well will flow into it by field-assisted diffusion, [97, 111] or alternatively designated by self-induced drift, [112]. The exact determination of this complex mechanism is difficult but an approximated analysis can be obtained by adapting the surface charge transport in silicon sheet model. It originates a diffusion equation type given by, [111, 97],

$$\begin{aligned} \frac{\partial \rho(x,t)}{\partial t} &= \frac{\partial}{\partial x}\left(\frac{\mu_n}{C_{ox}+c_{dep}}\cdot\rho(x,t)\,\frac{\partial \rho(x,t)}{\partial x}\right) \\ \Rightarrow\ \frac{\partial f(x,t)}{\partial t} &= \frac{\partial}{\partial x}\left(\frac{\mu_n}{C_{ox}+c_{dep}}\cdot C_{ox}\left(V_G-V_{th0}\right)\cdot f(x,t)\,\frac{\partial f(x,t)}{\partial x}\right) \end{aligned} \tag{4.49}$$

where the effective diffusion coefficient is a function of the carrier concentration, $\rho(x,t)$, across the channel lengt and $f(x,t)$ represents a channel charge density function. Based on 4.49, a timing estimate of the gate voltage bootstraping during the amplification phase can be obtained by calculating the time needed to transfer the charge channel after a step voltage being applied to the D/S terminal. The time needed to reach 90% of its final value is approximately given by, [97],

$$t_{90\%} \approx 6.2\cdot\frac{1}{\mu_n}\cdot\frac{C_{ox}+c_{dep}}{C_{ox}}\cdot\frac{L^2}{V_G-V_{th0}} \qquad [s]\ , \tag{4.50}$$

where μ_n is the electron mobility (for the NMOS device), V_G is the gate voltage at the beginning of the amplification phase and c_{dep} is the depletion layer capacitance. As a partial conclusion, the time taken by the parametric amplifier to amplify is the

sum of the time to pull the source voltage to V_{DD} and the time it takes to remove the inversion charges formed under the gate. The latter depends on device dimensions (the L) which can be made extremely small. It may, however, not be possible to simulate the frequency limitations of the parametric amplifier unless the models incorporate high-frequency effects, like NQS [29].

4.4.4 Noise analysis

The MPA cell analyzed so far behaves as a S/H circuit, which can achieve an intrinsic gain without any active amplifier. Furthermore this parametric amplification is intrinsically noiseless. However, the overall circuit includes components, as switches, power sources, clock generators and buffers that are responsible to add noise to it. Even the MPA MOS device itself also contributes to add noise through the substrate resistance. Some of these noise sources are depicted in Figure 4.23. As it happens in a

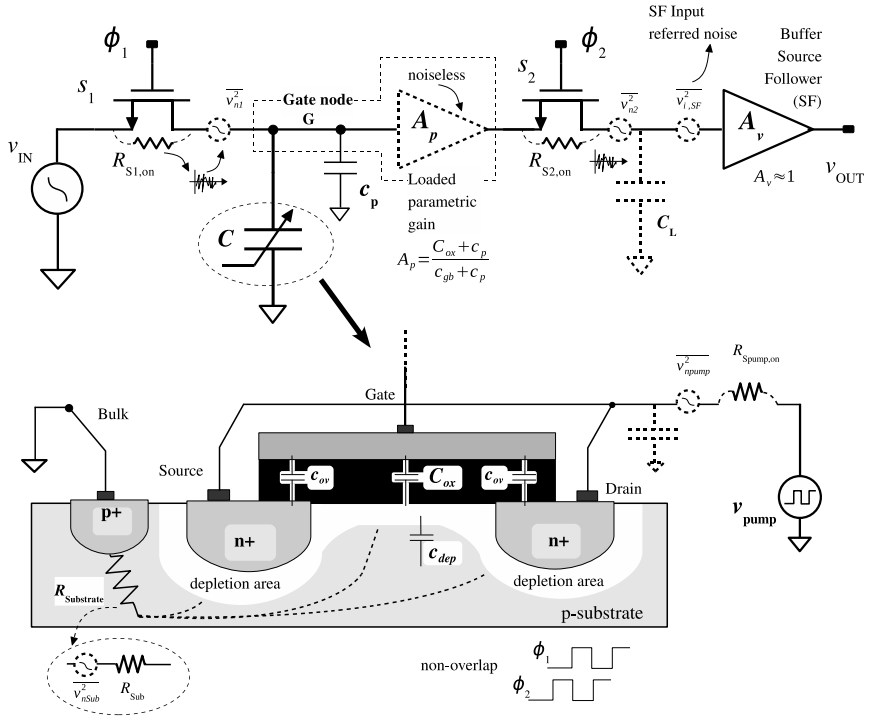


Figure 4.23: Simplified noise model for the MPA cell.

conventional SC S/H circuit, during the sampling phase the circuit behaves as a simple RC network, formed by the switch ON-resistance and the total capacitance at the gate node G which is approximately given by $C_{TH} = C_{ox} + c_p$ (it is assumed that the input voltage is high enough to put the device in strong inversion). Due to this nonzero value of the ON-resistance, a thermal noise component is sampled altogether with the input signal.

The random nature of the noise signal makes easier to use a power spectral density analysis. Taking the thermal noise power spectral density generated by the switch resistance given by $S_R(f) = 4kTR_{on}$, it is reasonable to consider that it is low pass filtered during sampling phase by the RC circuit with a bandwidth of $1/\tau$, where $\tau = R_{on}C_{TH}$ is the circuit constant time. Remembering that the noise is also submitted to the sampling process, the thermal noise power density is folded into the frequency band, up to half the sampling rate, [113], resulting in the well known kT/C_{TH} [V^2] noise power. Note that the determination of the last result not only assumes a 50% duty cycle but also that the circuit constant time is much lower than the avaliable sampling/aquisition time δts which is given rougly by $T_s/2$ (under the above conditions), thus permiting the complete settling. Nevertheless, a more accurate result can be found in [114], where the sampled noise power is determined to be

$$\overline{v_{n,TH}^2} = \int_0^\infty 4kTR_{on}\left|\frac{1-e^{-(j\omega+1/\tau)\cdot\delta ts}}{1+j\omega\tau}\right|^2 df = \frac{kT}{C_{TH}}\cdot\left(1-e^{-2(\delta ts/\tau)}\right). \tag{4.51}$$

It is clear from (4.51) that if it is given enough time for circuit to settle, the noise sample is kT/C_{TH} while for incomplete settling, interestingly, this sampled noise component is reduced. However, in the remaining discussion it is considered that the settling is long enough to cope for the desired resolution.

During sampling phase the D/S terminal is connected to ground through a switch, in case of an NMOS MPA cell. Since the channel is formed beneath the gate this switch ON-resistance, $R_{Spump,on}$, appears in series with C_{ox} (the device is in inversion mode). On the other plate of C_{ox} (the gate terminal in this case), the G node is connected to c_p and to the input switch, S_1, which, by its turn, also contributes with thermal noise. Considering that these noise sources are uncorrelated, they can be added together (in power) while maintaining the sampling capacitor given approximately by the former C_{TH}.

During amplification/hold phase the noise sample is present at the gate node G, which suffers a parametric voltage amplification together with the sampled signal. Therefore, the intrinsic parametric amplification maintains the SNR (since, as referred before, the parametric amplification is intrinsically noiseless). Nevertheless, direct noise sources add at the output node G. Two of them are the noise associated with the substrate and the noise associated with the ON-resistance of the switch that connects the D/S terminal to V_{pump}. If this voltage is high enough to remove completely the channel charges, the switch noise source can couple to node G through the overlap capacitances (these are the only two available paths to G node). Effectively, it is important to reduce the ON-resistance of the switches associated with the pump circuit branch, not only to

reduce the RC time constant (to speed up the channel charges push and pull phases) but also to reduce their noise contributions. Since the pump voltages used for the push and pull actions are fixed, these single transistor switches can be sized with the appropriate width to reach the desired ON-resistance.

As indicated previously, the RC branch formed by the depleted capacitance and the substrate resistance is another source of noise, in addition of one more constant time in the circuit. The implementation strategy is then lowering as much as possible the equivalent substrate resistance by surrounding the MPA MOS device with a guard ring of a maximum number of contacts to substrate, [99]. This noise contribution at node G is then much lower during sampling, since the channel sheet charges establish a direct connection to the low resistance D/S connection to ground thus isolating the substrate.

Neglecting the direct noise components during amplification and hold phase, one can estimate the expected SNR for an input sinusoidal signal with amplitude A_{in}, through

$$\mathrm{SNR}_{\mathrm{MPA,dB}} = 10 \cdot \log\left(\frac{A_{in}^2/2}{k \cdot T/C_{TH}}\right) . \tag{4.52}$$

However, the distortion performance achieved by a single-ended MPA cell which was analyzed in a previous Section, suggests that the THD shall also be accounted in the calculation of the MPA cell effective resolution. The usual parameter normally used for this is the Signal-to-Noise-and-Distortion ratio (SINAD[2]) which can be determined by, [115],

$$\mathrm{SINAD}_{\mathrm{MPA,dB}} = -10 \log\left(10^{-\mathrm{SNR}_{\mathrm{MPA,dB}}/10} + 10^{\mathrm{THD}_{\mathrm{MPA,dB}}/10}\right) . \tag{4.53}$$

For reasons that will be clarified later on, in Chapter 5, the equivalent resolution, expressed in bits, for a given SINAD may be obtained from the Equation 4.54, which is the result of considering the quantization noise in an ideal ADC, [116].

$$\mathrm{ENOB}_{\mathrm{MPA}} = \frac{\mathrm{SINAD}_{\mathrm{MPA,dB}} - 1.76}{6.02} \quad [bits] \tag{4.54}$$

Table 4.2 summarizes the values obtained from the application of the above expressions for two sized MPA cells, while maintaining the same remaining conditions, i.e., the common mode input voltage set to $V_{IN} = 0.4$ V, input signal amplitude set to $A_{in} = 200$ mV, a load parasitic capacitance set to $c_p = 30\%$ of C_{ox} and for the 130 nm CMOS technology. It is clear from the previously shown results that the harmonic distorion is the dominant factor that degrades the effective resolution of the MPA cell. Therefore

[2]Often used in defining the effective resolution of data converters.

MOS	MPA single unit		ENOB (based on SINAD)		Observation
size	SNR [dB]	THD [dB]	Single [bits]	Differential [bits]	For a 200 mV signal
$14/2\,\mu m$	60.4	-35.7	5.6	8.8	HD2 dominates for single
$2/2\,\mu m$	55.0	-35.6	5.6	8.1	HD2 dominates for single

Table 4.2: SNR and THD for two sized single-ended MPA devices, combined or not, in a pseudo-differential structure.

for reaching resolutions from 6 to 8 bits it is preferred to go to a pseudo-differential topology despite the duplication of the total noise power in the sampling operation.

The insertion of the basic MPA cell into a more complex system do need most likely a second output switch at the gate node as well as a source follower, SF, stage in order to better control the loading effect. This means that the total output noise must include the additional contributions coming from the additional switches and the buffer itself. This represents a degradation of the output SNR, but when referring it to the input, one must divide by the parametric cell gain.

Besides the thermal noise sources already discussed, one must also consider the pump voltages source, references voltages, power supply and ground connection since they also constitute entry points for noise. To reduce them, a careful design and layout has to be performed and they should be conveniently filtered out by large decoupling capacitors.

Since there is no DC current flowing thorough the circuit, one could expect that the MPA cell to be free of $1/f$ noise. In analogous passive mixers, which also does not have biasing current, flicker noise has been observed as well. It is found that the magnitude is proportional to the input signal amplitude and inversely proportional to the slope of the gate voltage at the switching part, [117, 118]. Nevertheless, the level of flicker noise of the MPA cell is expected to be substantially lower than that of the active amplifiers. Therefore, these noise source will be neglected in the MPA cell through the text since the following development concentrates in the design of applications in the tens of MHz range, which will hopefully be far from the low-frequency flicker noise corner.

4.4.5 MPA cell design

Based on the previous analysis, the design process of a MPA cell has to take into account several factors:

- the channel length of the MOS device should not be kept at the minimum value due to the overlap capacitance effects;

- the channel length of the MOS device should not be too high since this will slow down the overall cell. The cell maximum sampling rate is inversely dependent to L^2;
- to reduce the load effect on the gain of the cell, it is preferred to increase the width of the cell instead of its channel length;
- the layout of the cell should include a maximum number of substrate contacts ring to decrease the negative impact of the substrate resistance on speed and on noise performance;
- the sizing of the cell should only stop after the inclusion of the all parasitic elements obtained from the layout extraction and post-layout simulations;
- not only the input signal switches have to carefully sized (to achieve sufficiently low ON-resistance) but also the pull up and down control switches, since they determine the settling and noise performance of the cell, as well.

The effects of the parasitic capacitances on the MPA gain are clearly visible from the analysis carried out in previous Sections. Among all, the overlap capacitances do have an important impact specially for short channel devices. The total overlap capacitance permits a direct coupling from the pump voltage to the gate terminal. Observing the original MPA configuration shown in Figure 4.6, it can be found that if, instead of connecting both the drain and the source, only one of them is used (while the other is left floating), one can reduce the effective coupling capacitance by a factor of two. This can also be achieved by using two half-sized MOS devices instead of a single one and, simultaneously, leaving their short-circuited drains in a floating state. The modified MPA cell schematic is shown in Figure 4.24.

For the original "DS configuration" (i.e., with drain and source terminals short circuited), during the sampling phase all overlap and gate-oxide capacitances are charged by the input voltage v_{IN}, as shown in Figure 4.24. During the amplification, a charge redistribution involves the gate capacitance (now in depleted state) and the sum of all overlap capacitances given by $\sum c_{ov} = c_{ov1} + c_{ov2} + c_{ov3} + c_{ov4}$. Since the internal node A is shorted with V_{DD} during amplification phase, the drain-bulk (c_{db1}, c_{db2}) and source-bulk (c_{sb1}, c_{sb2}) junction capacitances do not directly influence the gate charge. Following the same analysis approach used for the MPA cell, a first order application of the charge conservation principle, leads to a small-signal gain (apart from any existing DC offsets), G_{v1}, given in

$$v_{out} = \frac{\sum C_{ox} + \sum c_{ov}}{\sum c_{gb} + \sum c_{ov}} = G_{v1} \cdot v_{in} \ . \tag{4.55}$$

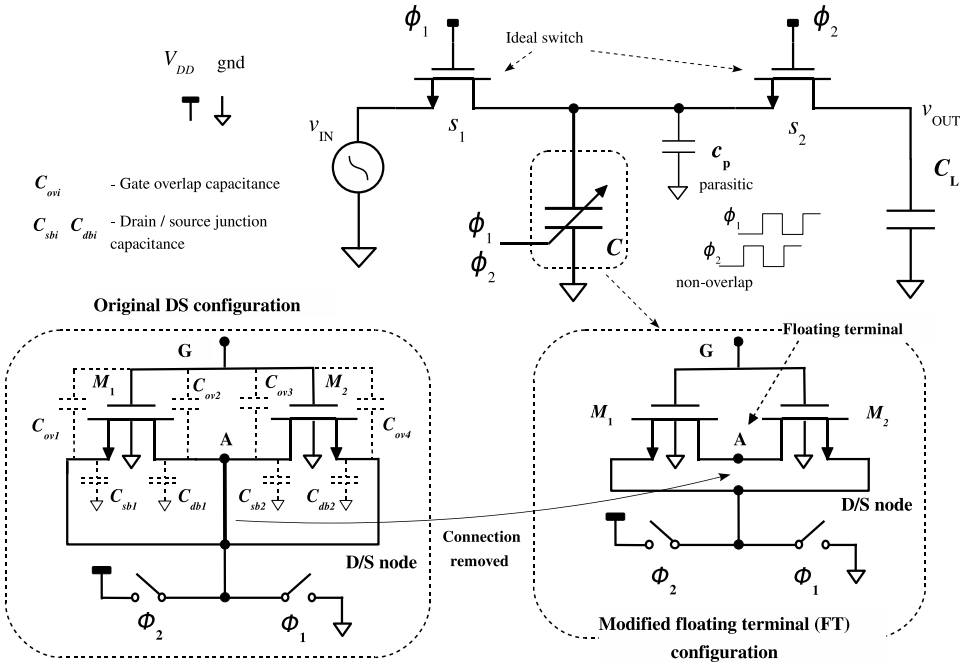


Figure 4.24: Modified MPA cell with two MOS devices: bottom-left - original "DS configuration"; bottom-right - modified "FT configuration".

For the modified "FT configuration"(i.e., with the sources in a floating state), during the sampling phase all overlap and gate-oxide of M_1 and M_2 are affected by the input voltage v_{IN}, as shown in Figure 4.24. During this phase, since the inverted channel layer exists (on M_1 and M_2), the floating terminal (node A) is pulled down to ground (0 V). During the amplification phase, there is a slightly difference from the original case, since now the drain connection between M_1 and M_2 (node A) is left floating. The voltage at this node suffers a transitory response and it changes from the initial 0 V to the final value, is such way that node A becomes the source terminal and the voltage between G and A tends to $V_{GS,M1,M2} = V_G - V_A \approx V_{t0}$ (where $v_{out} \approx v_G$ and V_{t0} is a small constant voltage). Hence, during the transitory amplification phase both, c_{ov2}, c_{ov3} and the junction capacitances (c_{db1} and c_{sb2}) experience a charge re-distribution effect. When the circuit is settled, the charges at c_{ov2} and c_{ov3} converge to approximately a fixed value. Similarly to what is suggested in [45], after the circuit has settled in the amplification phase, node A becomes approximately an AC ground (since voltage V_A just differs from V_G by a constant V_{t0} shift). Hence, the charge conservation in the amplification phase basically, involves the gate capacitance (now in depleted state) and two overlap capacitances, c_{ov1} and c_{ov4}. A first order application of the charge conservation principle, leads to (apart from any existing DC components) the result given by,

$$v_{out} = \frac{\sum C_{ox} + \sum c_{ov}}{\sum c_{gb} + \frac{\sum c_{ov}}{2}} = G_{v2} \cdot v_{in} \ . \tag{4.56}$$

	Sampling phase	Amplification phase
Gate capacitance (M_1+M_2)	C_{ox}=360 fF	c_{gb}=50 fF
c_{ov2}, c_{ov3} (at node A)	~ 2 fF (each)	~ 2 fF (each)
c_{sb1}, c_{sb2} (at node A)	10 fF (each)	0.4 fF (each)

Table 4.3: Capacitances values during sampling and amplification phases.

The relation between the gain achieved by both configurations is approximately given by

$$G_{v2} = \frac{1}{1 - \frac{\frac{\sum c_{ov}}{2}}{\sum c_{gb} + \sum c_{ov}}} \cdot G_{v1} \ . \tag{4.57}$$

In order to validate the previous results, an electrical simulation has been carried out using two equal sized devices M_1 and M_2 ($W = 7$ μm, $L = 2$ μm), under the following conditions:

- BSIM3v3 models with CAPMOD=2;
- ideal switches and ideal clock voltage source are used;
- an ideal input ramp signal is used;
- no capacitive load is connected at the gate node.

The parasitic capacitances are taken in the effective amplification region of the MPA cell. Their values, obtained from the simulation, are compiled in Table 4.3.

Using the parasitic capacitances values from Table 4.3, one can calculate the expected gain improvement from (4.57), which gives approx. $G_{v2} = 1.074$ * G_{v1}, i.e, around 7.4% more gain (single-ended).

The voltage gain achieved by each configuration, may also be obtained directly from the electrical simulation by observing the input and output signals. The simulated intrinsic voltage gain for the FT configuration is approximately 6.97 while for the DS configuration is 6.49, which indicates an enhancement of about 7% (very close to the one predicted by (4.57)).

The division into more than two devices can be used in case of higher unit capacitance values. However, care must taken in not using either large L (due to speed limitations) or minimum L (due short lengths effects). This configuration is extensively used throughout this work.

4.5 Some applications of the MPA cell

In this Section, three applications of the MPA cell will be described. It will be demonstrated the use of the MPA in a Multiply-by-Two-Amplifier (MTBA), embedded in the sampling network of a discrete-time comparator and, finally, the third circuit application reflects the use of the MPA in a discrete-time mixer than can be used in a fully integrated discrete-time digital receiver.

4.5.1 A Multiply by Two Amplifier (MTBA)

The operation of the MPA cell depends on the common-mode level of the input voltage, which reflects how well the MOS device is in inversion state during the sampling phase. Therefore, an appropriate DC level should be carefully chosen. Another approach is to modify the original structure in order to shift the amplifier input dynamic range to be around half of the voltage supply. A complementary compound structure, [102], with equally sized PMOS and NMOS devices, shown in Figure 4.25, can be used for this task. During the sampling phase the input voltage is sampled by both PMOS and NMOS device gates. During the amplification/phase the total charge at these gates is kept constant while the gate capacitances are simultaneously reduced, thus achieving the required parametric amplification. However, the main difference now is that, despite the total charge is kept constant, a movement of charges occurs between individual gates during the amplification that permits to affect the DC level shift. As demonstrated in Figure 4.26 and assuming that the PMOS and NMOS devices are

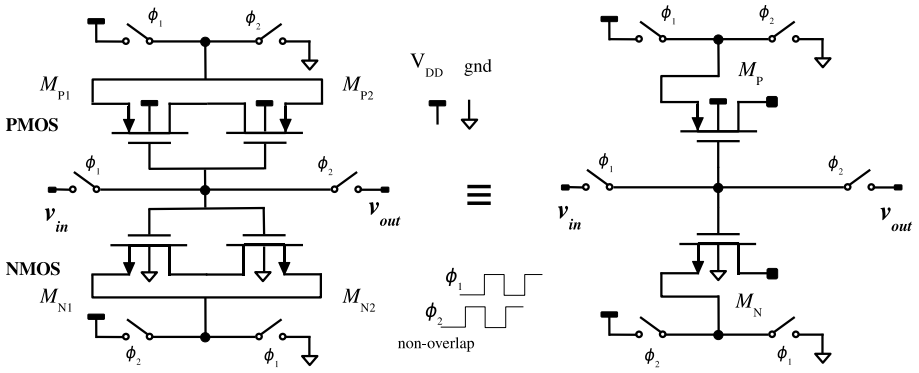


Figure 4.25: Discrete-time MPA with complementary structure, [102].

considered to be equally sized, the complementary structure does not alter the intrinsic gain of the original MPA cell. This structure permits, ideally, to extend the output range towards the full power supply range (i.e. almost rail-to-rail operation) and also to set the middle point of the input dynamic range to be $V_{DD}/2$.

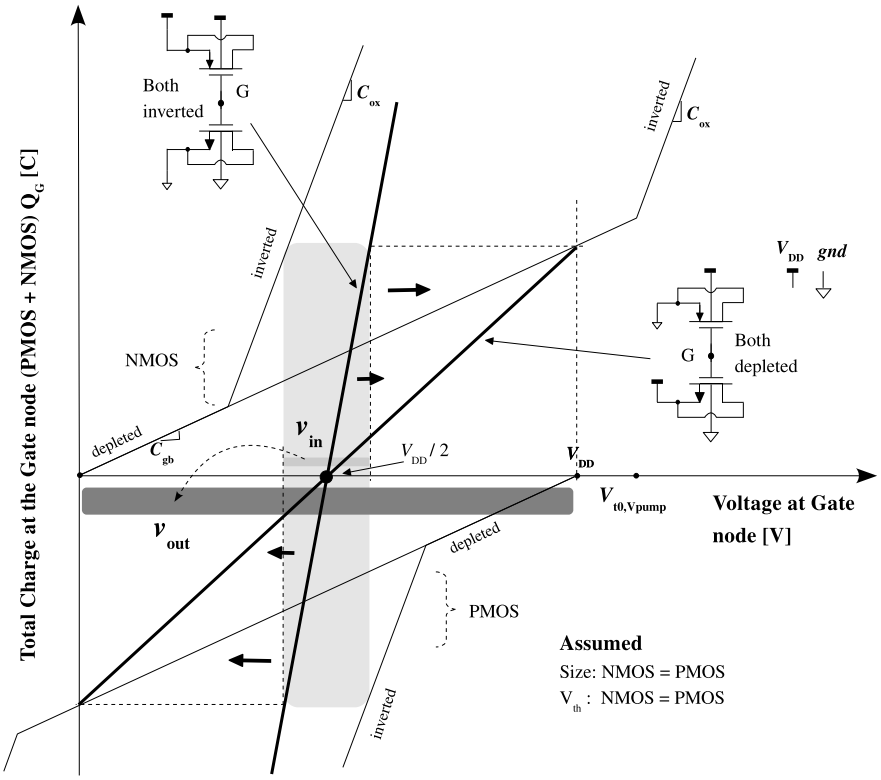


Figure 4.26: Q-V analysis of the DT MPA with complementary structure.

The previous structure can be further modified to better control the DC level shift (which tends to be higher than V_{DD}) during the amplification phase. As shown in Figure 4.27, the additional MPA PMOS device, C_{2P}, samples a fixed reference voltage, V_{refn}, instead of the input signal. Therefore, during the amplification phase, this PMOS branch produce a controlled negative DC level shifting at the gate node. A first-order charge redistribution analysis applied to the modified complementary MPA, results in an equation for the output voltage approximately given by

$$v_0 \approx \frac{\alpha_{C1N}}{k} v_i + \frac{\alpha_2 - \alpha_4}{k} V_{DD} + \frac{\alpha_4}{k} V_{refn} + \frac{\alpha_3}{k} V_{CL,\phi 1} \ , \tag{4.58}$$

where $C_{1N,\phi 1} = \alpha_{C1N} \cdot C_{1N,\phi 2}$, $C_{2P,\phi 2} = \alpha_2 \cdot C_{1N,\phi 1}$, $C_L = \alpha_3 \cdot C_{1N,\phi 2}$, $C_{2P,\phi 2} = \alpha_4 \cdot C_{1N,\phi 2}$, $k = 1 + \alpha_2 + \alpha_3$. Parameter k models the gain reduction due to the load parasitic capacitance C_L and C_{2P} during phase ϕ_2. Parameter α_{C1N} reflects the capacitance variation from phase ϕ_1 to phase ϕ_2 for C_{1N}. Parameter α_2 represents the relation between C_{2P}, during phase ϕ_2, and C_{1N} in phase ϕ_1. The remaining parameters α_3 and α_4 describe the ratio of C_L and C_{2P} in phase ϕ_2 with respect to the value of C_{1N} during phase ϕ_2. Equation (4.58) demonstrates that, besides a desired multiplying factor, the output voltage has an offset component partially controlled by the PMOS device. It also shows that the load capacitance, C_L affects the final gain of the circuit

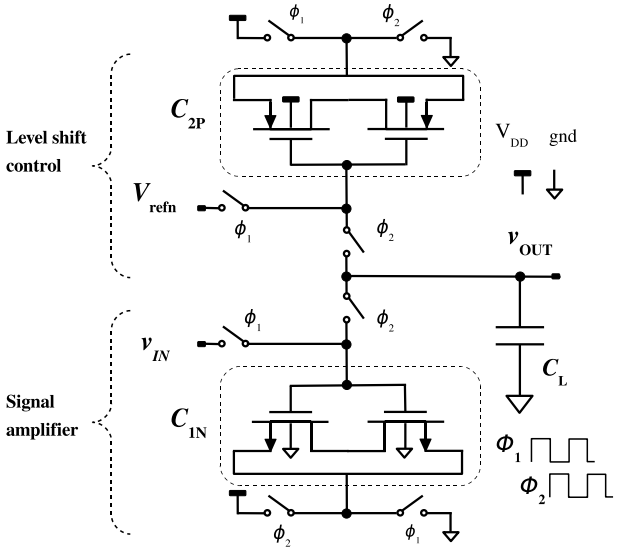


Figure 4.27: CMOS discrete-time amplifier (MPA block) with output level shift control.

(as expected) and the output offset level (depending on how the voltage is applied to C_L during ϕ_1, $v_{CL,\phi 1}$). To reduce this loading capacitance effect, a simple source-follower (SF) can be inserted at the parametric amplifier output node. This SF is also useful to buffer the output when the MPA cell is loaded by another stage or circuit.

An interesting observation is that, by combining the size of the MPA cell and the load capacitor, one can achieve a multiply-by-two amplifier (MTBA). The structure for the differential MBTA represented in Figure 4.28 comprises four single-ended MPA (+ and - type, respectively, for the positive and negative signal paths) and two SFs circuits. As stated before, these buffers are required to reduce the loading effect of the next stage (e.g., in a pipeline ADC). The schematic of each MPA block is depicted in Figure 4.29. Moreover, a time interleaved operation, in which the two SFs can be efficiently shared, may be easily implemented due to the DT nature of the MPA, meeting the requirements for high speed operation. A replica-bias circuit, which will be described later, provides the required bias voltage, V_{bias}, for the current sources in the SFs. The schematic of the MPA circuit shown in Figure 4.29 corresponds to a PMOS version. During phase ϕ_1, the input signal is sampled into C_1 and the positive reference voltage (V_{refp}) is sampled into C_2. During the parametric amplification phase, ϕ_2, the capacitance values of C_1 and C_2 decrease due to the change from inversion into depletion region and are connected to the SF input for charge re-distribution. Applying to this circuit the same methodology used for the parametric amplifier transfer function, results in a similar equation to 4.58 in which a new factor appears to include the SF gain (lower than one). Additionally, when considering a differential output voltage, the common-mode components present in 4.58 are removed, resulting in a final expression for the

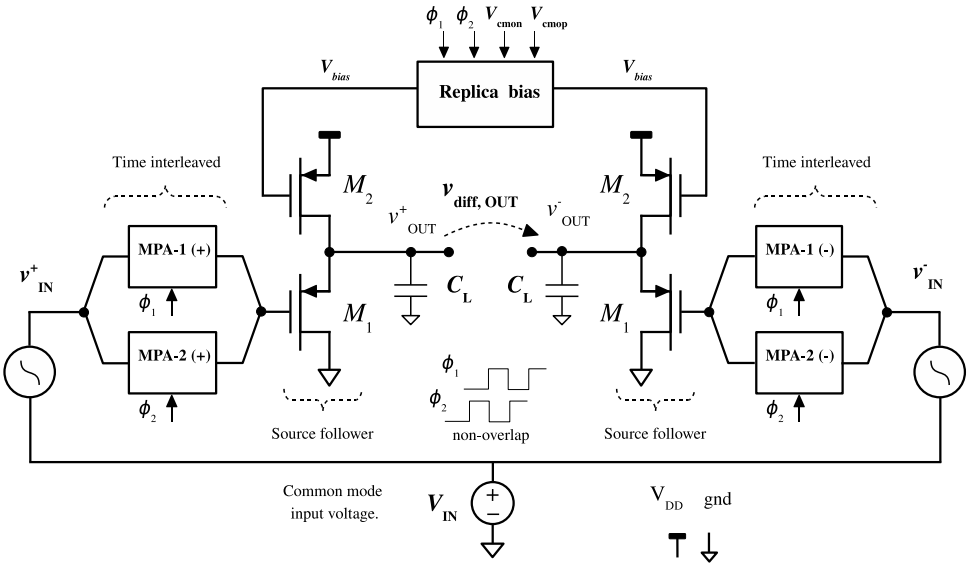


Figure 4.28: Four MPA connected in a time interleaved structure.

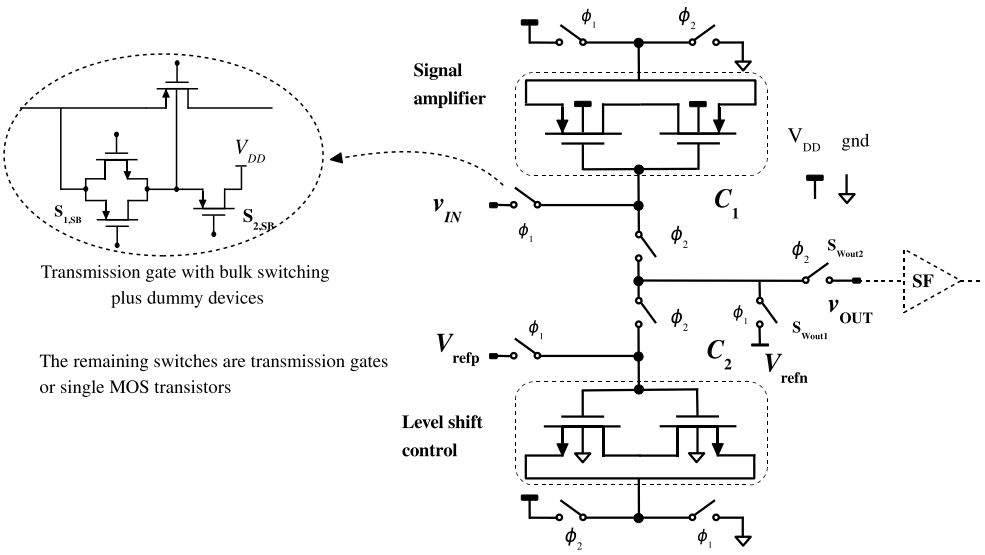


Figure 4.29: Half MPA single-ended circuit (p version).

differential voltage, v_{od}, at the outputs of the SFs given by

$$v_{od} \approx \frac{g_{m1}}{g_{m1} + g_{sb1}} \cdot \left(\frac{\alpha_{C1N}}{k} \cdot v_{id} \right) . \tag{4.59}$$

Transconductances g_{m1} and g_{sb1} represent, respectively, the main and body-effect transconductances of the common-drain of the SF (M_1).

In (4.59), it has been assumed that all MPAs have equally sized transistors and drive equal load capacitance from the SFs. Note that, due to the body-effect in the SFs, their gain is bellow the unit. Hence, to compensate this effect the complete MBTA circuit transistors are adjusted and properly sized to reach a gain factor precisely equal to two.

Switches connected to the amplifier input and output signals are implemented with asymmetric transmission-gates employing bulk-switching (ATG-BS), switches connected to the constant reference voltage (V_{refp}, V_{refn}, V_{DD}, V_{SS}) are simply MOS transistors (either PMOS or NMOS. All switches connected to the input signal have dummy structures to reduce signal-dependent charge injection. Switches S_{Wout1} and S_{Wout2} are used to reduce the memory effects that appear due to the interleaved operation.

Another important aspect is that the output common-mode voltage (V_{CMO}) at the outputs of the SFs can not vary too much in order, for example, to avoid DC accumulation errors in subsequent stages (e.g., in a pipeline ADC). A replica-bias circuit as the one depicted in Figure 4.30 is used to guarantee that V_{CMO} is restored at the MBTA output and adjusted to a proper value against process, supply and temperature (PVT) variations.

The replica-bias block comprises two time-interleaved scaled MPAs connected to a same replicated SF. The average voltage at the buffer output is set and controlled by a negative feedback loop which includes a very simple amplifier. The objective is to replicate the operation of the differential MPA and controlling the common mode voltage present at the MPAs buffers outputs through a biasing voltage V_{bias}. As referred before, stabilizing the output common-mode component at the outputs of each stage prevents accumulation of DC offsets. Note that the OTA used here does not need a high bandwidth since it is not inserted in the main signal path. Therefore, it can be designed with minimum biasing current (20 μA was used). For an input differential signal full-scale swing of 400 mV_{pp}, V_{refp} of 1 V and V_{refn} of 0 V, the values of 0.4 V and 0.7 V were found to be optimal for the common mode components V_{cmon} and V_{cmop}, respectively.

The proposed MBTA circuit was designed and simulated at the transistor level in a 130

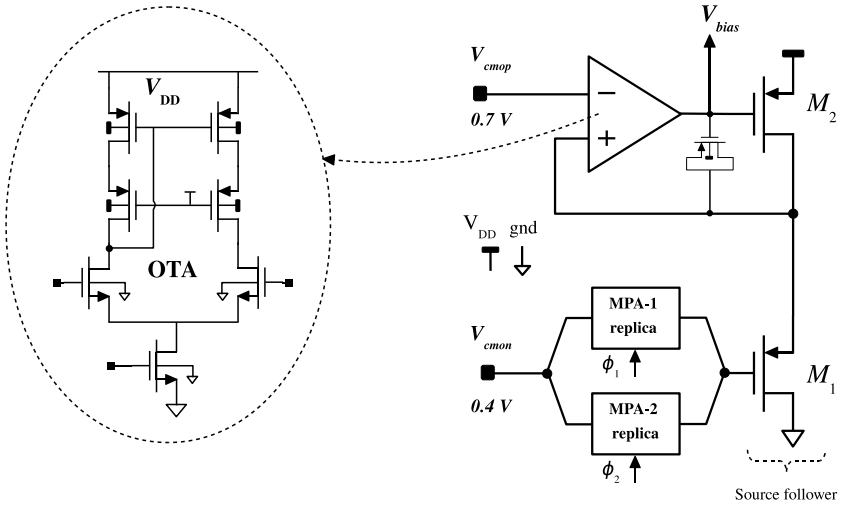


Figure 4.30: Replica bias block used to generate Vbias for the 2 source-followers used in the MBTA block.

nm standard CMOS technology. The simulation was performed in Spectre/Cadence electrical simulator using BSIM3v3 CMOS models. Figure 4.31 displays the differential signal present at the MBTA output when a differential signal of 10 MHz and 190 mV_{pp} is injected at the input. The overall MBTA circuit is sampling the input signal at a rate of 100 MHz. A gain error of ±1.62% error was obtained for typical conditions, which is compatible to a 7-bit level of accuracy. Obtained from the transient simulation, the output signal spectrum is represented in Figure 4.32. It is based in a 512-bin FFT (coherent sampling) of the amplifier output when a sampling frequency of 100.0448 MHz is used and a 10.388 MHz input differential signal of 190 mV_{pp} is injected at the input. A load capacitance of 250 fF was considered at the amplifier output node. Simulations results exhibit a THD better than -60 dB. The complete circuit dissipates

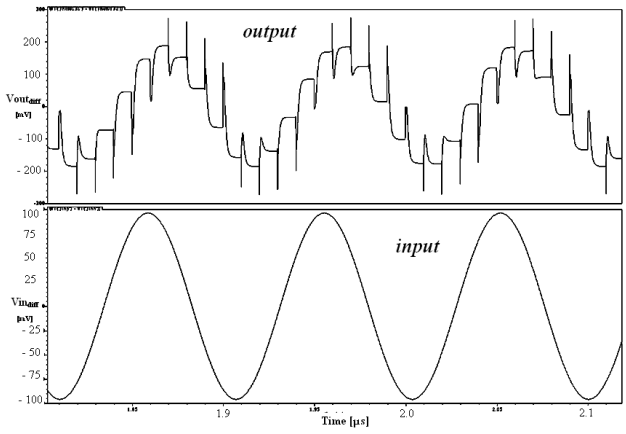


Figure 4.31: Differential output and input of the MBTA.

only 1 mW (at $V_{DD} = 1.2$ V) in typical conditions which indicates the power efficiency

achieved by the proposed MBTA circuit topology.

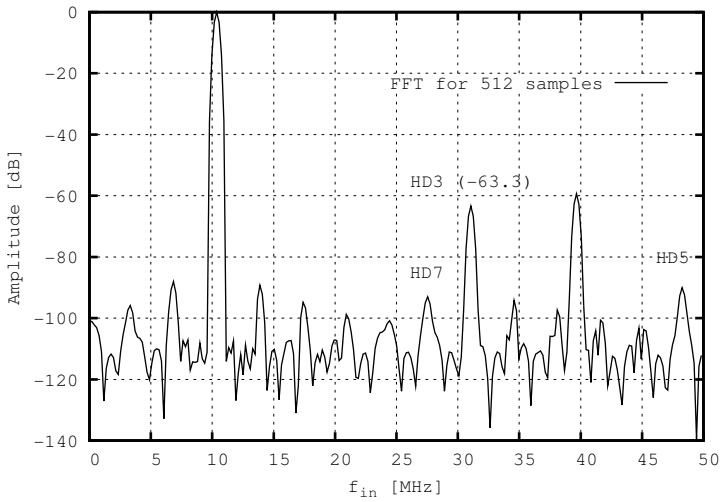


Figure 4.32: Simulated FFT output spectrum of the proposed MBTA circuit operating at 100 MHz clock rate, and for a 10 MHz input signal.

4.5.2 Comparator

In this subsection it is shown the design of a comparator with embedded parametric pre-amplification. In order to be used in the local 1.5-b quantizers of a pipeline ADC, the desired comparator [7, 8] should resolve input differences of less than ±125 mV ($\pm V_{REF}/4$), for an input dynamic range of 1 V_{pp} (differential). Moreover, a regeneration-time requirement of less than 250 ps (over PVT corners) should be obtained, in order to allow conversion times (limited by the local residue amplification used for a ADC pipeline stages) below 2 ns ($F_s > 500$ MS/s).

The block diagram of the proposed comparator circuit is shown in Figure 4.33. It is based on an input SC-network followed by a Positive Feedback Latch (PFBL) [119]. The use of a dynamic PFBL enables high-speed operation and provides no static power dissipation. However, due to its offset, a pre-amp has, normally, to be used after the input sampling network in order to avoid metastability problems and overcome the residual offset of the PFBL. To maintain low-power operation, this pre-amp can be embedded into the SC-network. In [120], the proposed charge-transfer pre-amplifier uses a dynamic source-follower to implement the required gain. Although the circuit does not have static power consumption it requires three cycles of operation. The use of a differential input structure [121], produces a proportional output voltage with a high level of rejection of the input common-mode component.

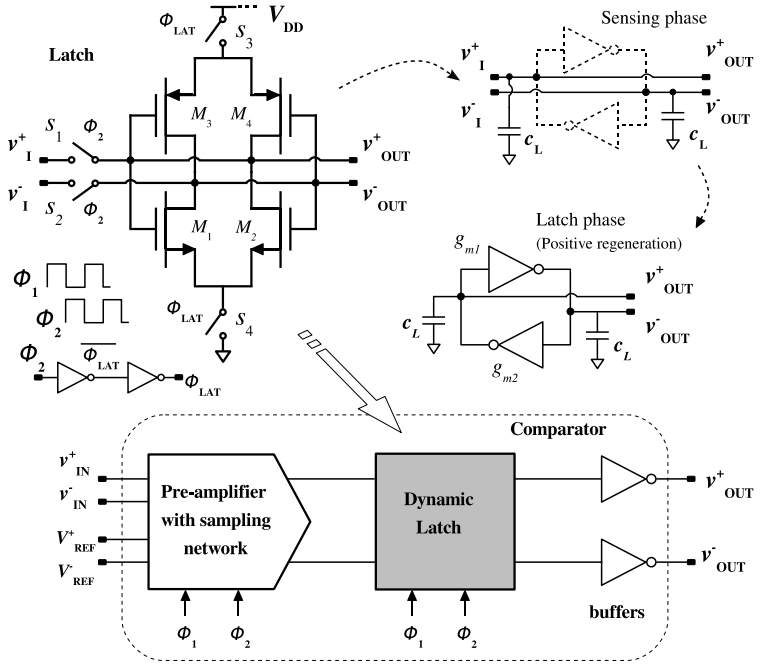


Figure 4.33: Block diagram of the parametric based comparator.

The supporting approach for the design is to define the threshold-voltage of the comparator in the sampling-phase (ϕ_1) and then obtaining the desired pre-amp gain by changing, dynamically, the capacitance values of the SC-network during ϕ_2. The use of linear amplification inside a comparator is not fully required which enables the utilization of MPAs. The modified input sampling network is illustrated in Figure 4.34. The circuit operates with a standard two non-overlapping clocks and each capacitor is implemented by means of a MOS varactor as the one described in the previous Sections. A charge redistribution analysis applied to the above circuit, results in a

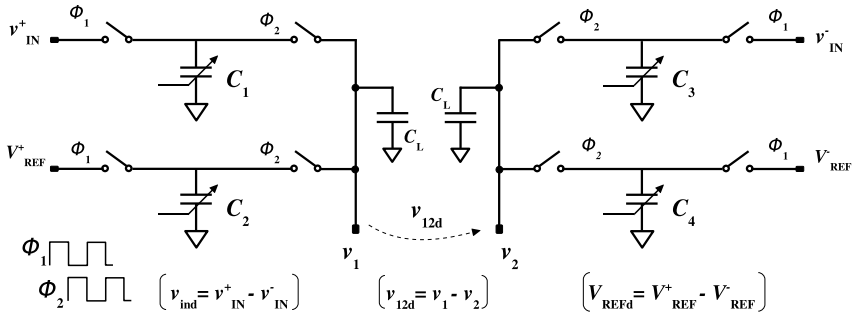


Figure 4.34: Input SC network with embedded parametric amplification.

final expression for the output differential voltage given by

$$v_{12d}^{\phi 2} = \left(v_{ind} + \frac{V_{refd}}{\alpha_c}\right) \frac{\alpha_c k_{c13}}{\alpha_c + \frac{k_{c13}}{k_{c24}}\left(1 + \frac{C_L}{C_{24,\phi 2}}\right)} \ , \tag{4.60}$$

where $C_1 = C_3 = C_{13}$, $C_2 = C_4 = C_{24}$, $C_{13,\phi1} = \alpha_c C_{24,\phi1}$, $C_{13,\phi1} = k_{c13} C_{13,\phi2}$, $C_{24,\phi1} = k_{c24} C_{24,\phi2}$. The differential threshold level does not depend directly on the load capacitance and can be adjusted by the ratio of C_1 and C_2 during ϕ_1. The load capacitance, C_L, only affects the circuit gain.

Figure 4.35 displays the complete schematic of the proposed comparator, in which a PFBL is attached to the output of the input SC-network. Also a transistor diode limiter is used to limit the maximum voltage applied at the latch input and to reduce the recovering time of the comparator in each clock cycle.

The design procedure should start by sizing the latch transistors. In order to minimize its input capacitance (C_L), which degrades the pre-amp gain, the PFBL and output buffer transistors sizes have to be kept small. As stated in [119], the latch delay-offset product suggests that, the length of the transistors should be set close to minimum length. The PMOS width is set to about 4 times higher than the NMOS counterpart, to reach similar transconductance factor and both should be higher than the minimum value due to offset requirements, [119]. Using equation (6) from [119], and accounting only for threshold and dimensions mismatches, a theoretical input systematic offset error of 7 mV for the latch is obtained. Adding the contribution of the charging injection mismatch between the switches at the latch input, determines the latch input resolution, which was set to 40 mV. For a pre-amplifier gain between 2 to 4, the expected offset comparator offset reduces to 10 to 20 mV in the worst-case.

Considering a comparison level of $V_{refd}/4$ and an input latch capacitance, $C_L = 50$ fF, from the MPA gain, a pre-amp gain of 2.5 can be achieved with $C_{13,\phi1} = 4C_{24,\phi1}$, and capacitance variation factors (CVF) k_{C13} and k_{C24} of 6 and 5.8, respectively. These CVFs can be obtained by squared transistors with $WL = 2\text{x}2\,\mu$m and $1\text{x}1\,\mu$m, respectively for C_{13} and C_{24}.

Although the input sampling switches are implemented with transmission-gates type, the remaining ones use only single devices (details given in Appendix B). The adopted DC voltages for the common-mode, positive and negative references voltages are, $V_{CM} \approx V_{DD}/2 = 0.6$ V, $V_{REF}^{-} = 0.3$ V and $V_{REF}^{+} = 0.8$ V.

The time the latch takes to produce a digital output during the regenerative phase, can be estimated from the regeneration time constant, given by C_L/g_m, [121, 119], where g_m is the total transconductance of the CMOS inverter. Considering an output swing of 1.2 V (V_{DD}) and a capacitance of 50 fF at each latch output node, the PFBL takes approximately 31 ps to reach its final state from an amplified input signal of 125 mV. The regeneration time can be further reduced by using two reset switches

connecting the latch inputs to V_{REFN} during the sampling phase. The comparator

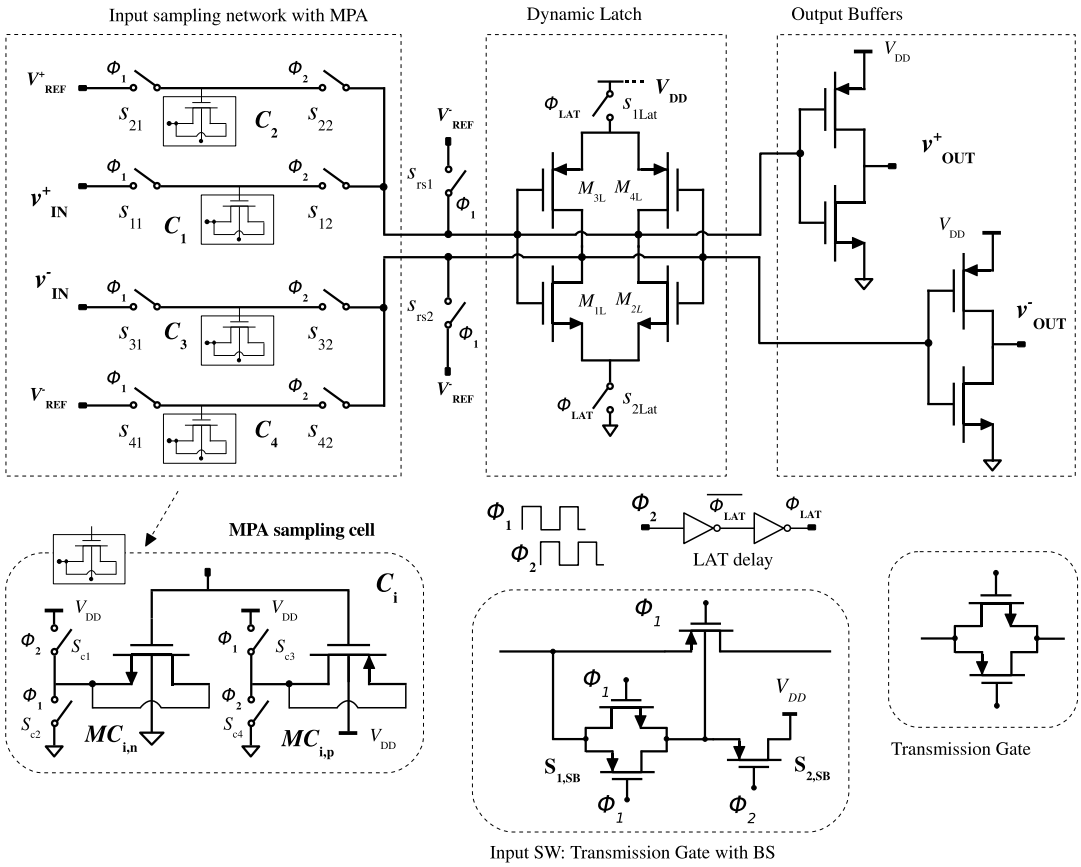


Figure 4.35: Comparator circuit schematic with embedded parametric amplification in the input SC network.

core topology was designed to eliminate any static power dissipation. The remaining dynamic power dissipation is obtained from the contributions of the input sampling pre-amplifier, regenerative latch and output buffers. This sizing results in a estimated total RMS power of 138 μW for a sampling rate of 500 MS/s at 1.2 V power supply.

In order to obtain the random input offset estimation, a Monte Carlo simulation process has been carried out, using the BSIM3v3 models available for the used 130 nm CMOS technology. A Gaussian distribution was applied for each device parameter (W, L, V_{thN} and V_{thP}), independently for each transistor, based on the statistic information provided by the foundry. The offset histogram of the comparator, based on 800-case Monte-Carlo simulations, is presented in Figure 4.36. The obtained offset average is -3.6 mV with a standard (σ) deviation of 20.3 mV. The supply current used by the comparator core is 118 μA in average, which corresponds to a total average power dissipation of about 142 μW. From the time response of the comparator, Figure 4.37, the amplification and latch regeneration time periods can be observed. The total worst-case regeneration time is 210 ps. The comparator efficiency is 0.14 pJ per comparison

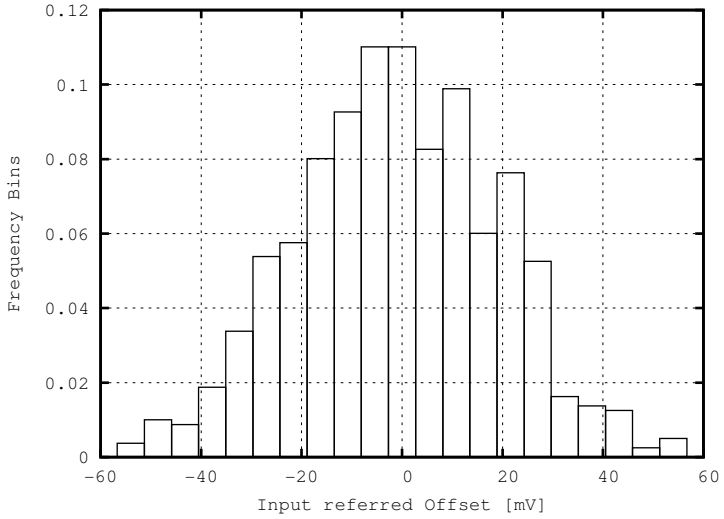
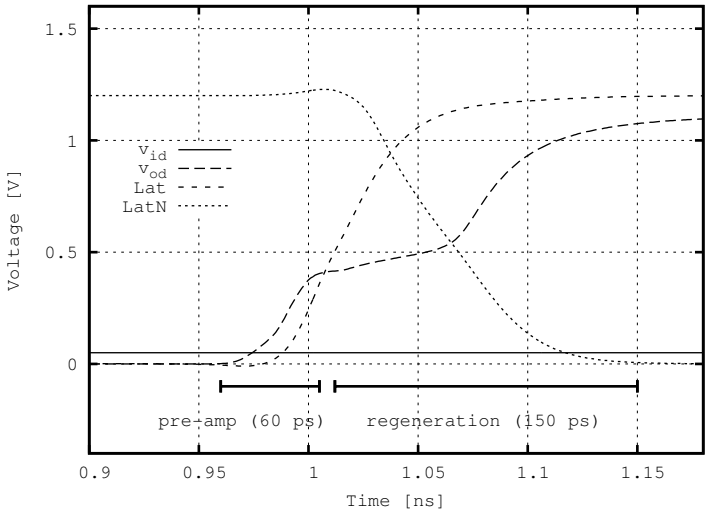


Figure 4.36: Monte Carlo simulations (800) to determine the input-referred offset of the comparator.

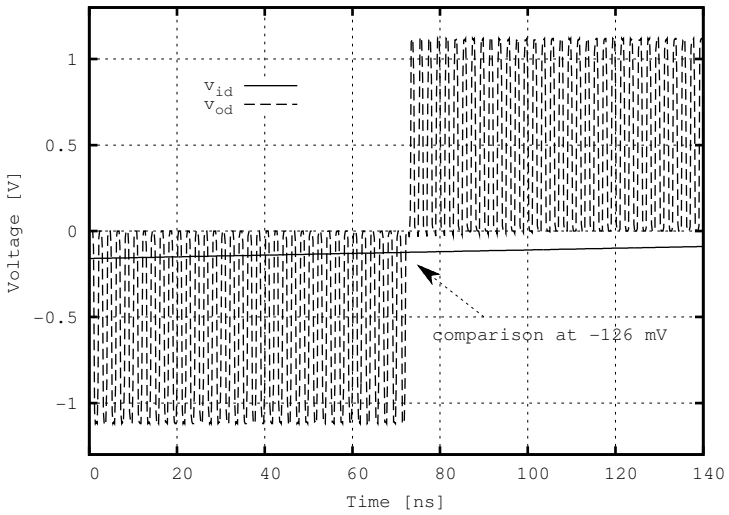
and its single-ended input capacitance is 85 fF.

4.5.3 A Mixer for a discrete-time receiver

Apart from the use in a comparator, the DT MOS parametric amplification technique can be integrated wherever a moderate to small gain is needed. There are only just a few examples of reported uses of the discrete-time MPA cell. Besides the pioneering work from [98, 13], the circuit presented in [122] demonstrates the application of the technique into a discrete-time charge-domain filter with FIR spectrum type of response. Figure 4.38 shows the block diagram of the reported three-stage FIR filter. It consists of a cascade of second and third order *sinc* filters to achieve large attenuation in the stopband, while simultaneously performing a decimation operation. The unit sampler cell needed for this type of filters is usually supported on a passive MIM capacitor and, therefore, the attenuation in the passband is not negligible. The followed approach relies on replacing these passive capacitors by modified MPA cells in order to compensate this attenuation. The modified MPA cell proposed in [122] and depicted in Figure 4.38 is based on a double-complementary MOS parametric amplifier (DCMPA) structure, and shows great similarity to the one that is described in Section 4.5 of this work. The main difference is that two of the MOS devices are interchanged. Both structures can minimize the common-mode DC offset voltage but the second one tends to be less sensitive to parasitics and therefore is able to reach a higher gain, as it will be demonstrated at the end of this Section. The experimental results reported in [122], indicate that an overall bandpass gain of the cascade can achieve a value as high as 30



(a) Latch regeneration time



(b) Comparison level

Figure 4.37: Electrical simulations to evaluate the worst-case time response of the comparator.

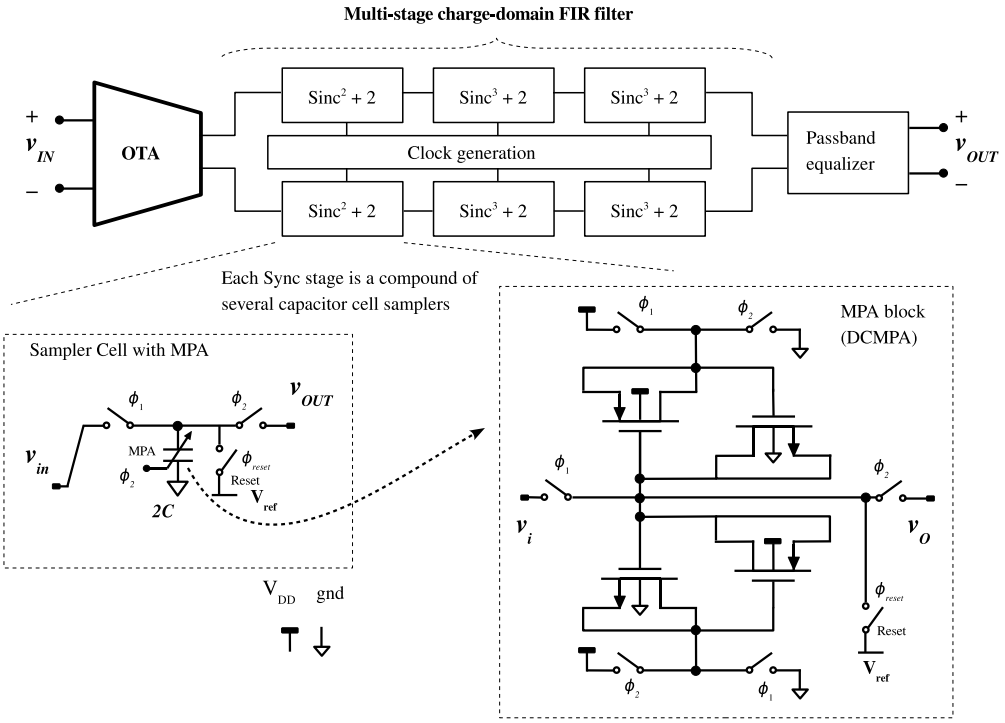


Figure 4.38: Discrete-Time and charge based FIR filter employing a MPA technique, from [122].

dB for a 80 MHz clock rate, 14 MHz of bandwidth and a stopband attenuation higher than 60 dB.

The previous circuit can be used in a discrete-time receiver, [123, 124, 125], which has some interesting characteristics when compared with the continuous-time version. One of them is its programmability feature which is very attractive for SDR applications. The DT receiver samples the incoming signal, i.e., discretizes the signal in time but performs additional analog processing before delivering it to the ADC and subsequent digital processing. The RF input sampler can be performed with respect to charge or to voltage, [126]. However, in both cases RF continuous-time pre-filters are needed to prevent noise and interferers around harmonics of the sampling clock from folding to baseband. To overcome this problem, [126, 42] propose a DT harmonic-rejection mixing (with embedded IIR filtering and decimation) architecture that relaxes the requirements for the RF filter, and allows reducing the noise folding.

The front-end receiver architecture, [12], is depicted in Figure 4.39. The SC section which implements the Mixer/IIR filter performs three functions, [42]: an oversampling, with $F_s = 8\,f_c$ (F_s and f_c being the sampling and carrier frequencies, respectively); two quadrature (I/Q) DT mixers for downconversion; two low-pass IIR filters. The low-IF quadrature signals are connected to output SFs. The Mixer/IIR filter block has

eight interleaved sampling cells that are controlled by 8 non-overlapping clock phases ($\phi_1 - \phi_8$). Each individual cell operates at f_c and, thus, the effective overall sampling frequency is $8\,f_c$. The DT mixing is achieved though the output switches, due to the transfer of charges from the sampling capacitors to the buffer input capacitor (c_b). The charge sharing between the sampling and buffer capacitors complements the low-power IIR filtering.

Distinctively to the sampler cell proposed in [42], the one shown in Figure 4.39 replaces the MIM capacitor by a MPA cell with the objective of adding effective gain to the mixer circuit, [12]. This implies using an additional clock phase *per* cell, to control the MPA operation. The full topology is show in Figure 4.39. The DT mixer circuit

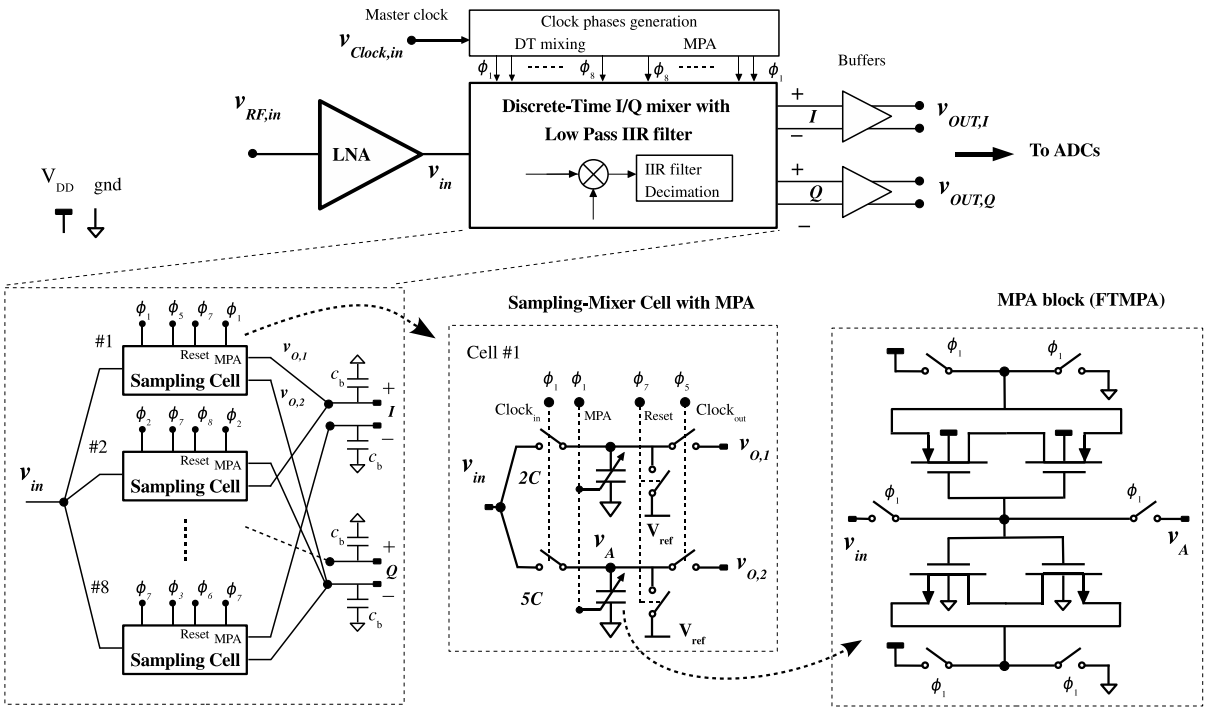


Figure 4.39: Discrete-Time Mixer with embeded MPA technique.

represented in 4.39 has been fully designed in an 130 nm CMOS technology and, for comparison purposes 3 scenarios have been analyzed:

- original DC mixer, i.e., with the MPA function disabled;
- DT Mixer with MPA activated and the circuit cell is the DCMPA from [122] (see Figure 4.38);
- DT Mixer with MPA activated and MPA cell based on the floating terminal MPA configuration (FTMPA), seen in subsection 4.4.5, and also represented in Figure 4.39;

For comparison purposes, a ratio between capacitances of the two branches was kept at 2/5 value. The absolute capacitance value was chosen to be 100 fF and 250 fF, respectively (sizing mainly imposed by KT/C noise constraints).

As stated before, the channel lengths (L) of the MOS devices must be chosen with care: a large L compromises the maximum achievable speed but a minimum one leads to short-channel effects. To find how long the MOS devices should be, a set of simulations (using a parameterized L) for the gain and noise figure of the MPA configurations was made, starting with the minimum length ($L = 120$ nm, in the adopted 130 nm 1P8M technology, with $V_{th} \approx 0.32$ V) and ending with $L = 0.36$ μm. The simulation results

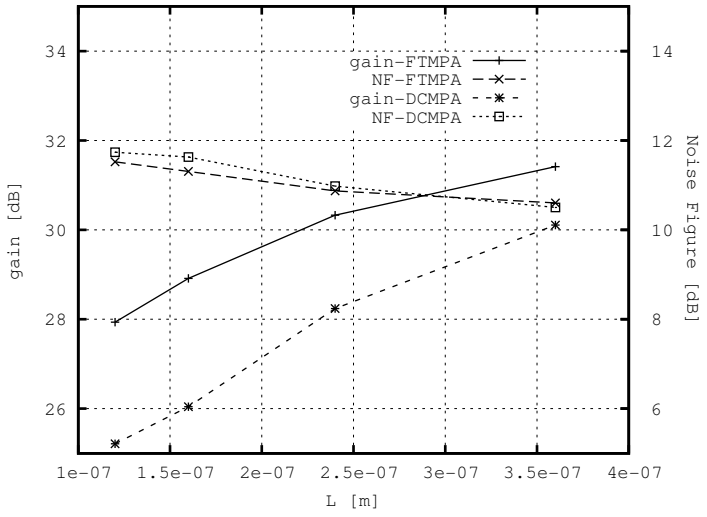
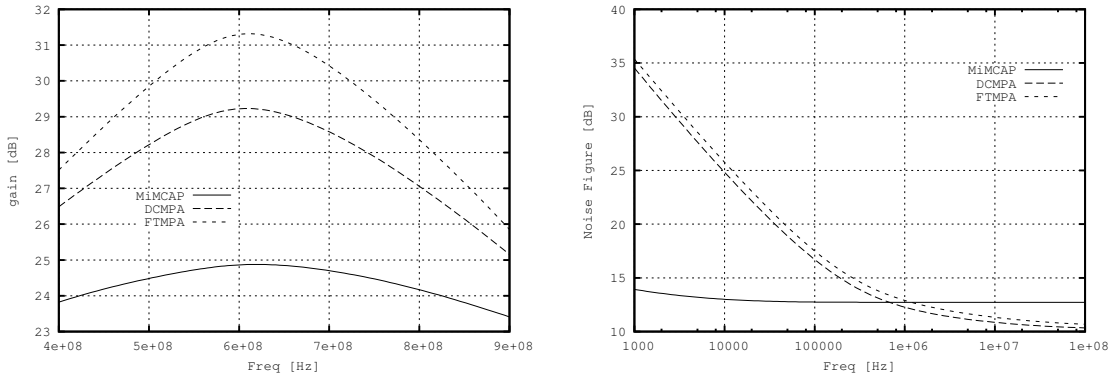


Figure 4.40: The effects of L variation on the DT-mixer gain and NF, for DCMPA and FTMPA circuits.

shown in Figure 4.40, in which the widths have been kept constant, indicate that higher gain and lower NF are easier to achieve with longer channels. Nevertheless, long channels have been avoided due to speed limitations which are related to the issues analyzed in Section 4.4.3. Trading off between short and long channel effects, an $L = 240$ nm (twice the minimum length) has been chosen. The adopted widths are $W_1 = 9.32$ μm and $W_2 = 23.3$ μm, respectively for C_1 and for C_2. The total conversion gain obtained for each of the three cases is plotted in Figure 4.41. The results show that the presented FTMPA solution achieves, as predicted, a higher gain than the other designs. In terms of noise figure, the results obtained for the same cases are shown in Figure 4.41. It can be seen that outside the flicker noise band, circuit with MPA achieve a lower noise figure when compared to the circuit with MIMCAPs. Table 4.4 gives key performance parameters. The Mixer/IIR filter circuits using MPAs have around 2.6 to 4.5 dB more intrinsic gain and 1 to 2 dB improvement in the NF. Despite having higher gains, both the 1 dB compression and third-order intercept points



(a) Conversion gain frequency response for the three DTmixer cases.

(b) NF frequency response for the three DTmixer cases.

Figure 4.41: Discrete-Time frequency response obtained from SPECTRE simulator, using an 130 nm CMOS technology.

Parameter (@ IF = 10 MHz)	units	MIMCAP	DCMPA	FTMPA
Conversion Gain$_{\text{(Mixer+IIR filter)}}$	[dB]	-1.7	2.6	4.5
NF$_{\text{total}}$	[dB]	12.7	11.0	10.5
P$_{-1dB}$	[dBm]	-20	-25	-26
IIP3	[dBm]	-1.5	-10.0	-6.2
Power $_{\text{(Mixer+IIR filter)}}$	[mW]	4.8	8.0	7.8

Table 4.4: Comparative results for the three sampling cell cases.

decrease accordingly when compared to the circuit with MIM capacitors, meaning that the distortion is not higher. As expected, the power dissipation is higher for the MPAs circuits, as a consequence of better performance. Flicker noise below 100 kHz is also visible but not relevant in Low-IF receivers.

4.6 Summary

Complementing the continuous-time parametric MOS amplification discussion introduced in the previous Chapter, in this one, the extension of this technique to the discrete-time domain is carried out. It has the advantage of not using LC filters, making possible to implement an open loop low-gain amplifier using only MOS device.

The performance achievable from the MPA cell has been deeply analyzed, in which it was concluded that the surrounding circuitry plays a major role on the gain obtained with this type of structure. Similar to the continuous counterpart, the MPA cell is intrinsically noiseless. However, switches and also the substrate resistance add noise to the operation. It has been demonstrated that an accuracy between 6 to 8 bit is

possible to achieved with an MPA cell.

The Chapter ends with the presentation of several circuit where the MPA can be employed. A multiply-by-two amplifier is presented as well as a comparator. The last circuit applies directly to a digital receiver where the MPA cell has been included in the complete implementation of DT RF front-end mixer/IIF filter.

Chapter 5

Design of a pipeline ADC fully based in MOS parametric amplification

5.1 Introduction

In the previous Chapters it has been demonstrated that the MOS parametric based amplifier can be effectively used in an RF receiver, both in continuous-time or discrete-time configurations, namely in the frequency down-conversion stage. Moreover, the applicability of the technique using digital CMOS technology has been also highlighted.

As already referred, an ADC is a cornerstone in modern receiver architectures. In this Chapter, it is demonstrated that the discrete-time parametric amplification technique can be extensively applied in the design of a pipeline ADC, as well. As a consequence, this technique opens the possibility of building an ADC that only makes use of MOS devices, i. e., without the need of any MIM or MOM capacitors, resistors, thus facilitating the full integration in digital CMOS technology with high area efficiency.

One of the main objectives for the proposed parametric based ADC is to reach a moderate resolution range (6 to 8 bits). For this resolution, the flash (parallel) architecture is less attractive due to area and input capacitance constraints. Therefore, in Section 5.2 the overall architecture of a pipeline ADC is discussed as well as the desired target specifications. Time-interleaving techniques are used to further increase the conversion rate. The following Sections detail all the major building blocks of the proposed ADC. In Section 5.5, a system level noise model analysis is presented. In Section 5.6, the

FFT results obtained from electrical simulations show that the ADC can reach more than 6-bits of ENOB, without self-calibration. At last, some conclusions are drawn in Section 5.7.

5.2 Description of the ADC architecture

The most straightforward ADC architecture is the flash converter which is based on a bank of comparators working, simultaneously and in parallel at the same speed (i.e. at the full ADC sampling rate, F_s), thus requiring each one to have the full ADC accuracy. In this type of Nyquist-rate converters, the number of required comparators is $2^N - 1$, where N (in bits) is the total resolution of the A/D converter. However, selecting this type of architecture even for medium resolutions implies the use of many comparators and, consequently, the converter area and power efficiency is seriously degraded. Hence, alternative A/D architectures have been addressed.

When medium to high conversion rates and medium to high resolutions are envisaged, the ADCs usually employ pipelining to relax the speed and accuracy requirements of the analog components. As shown in Figure 5.1, a cascade of stages is used, each consisting of a low resolution flash quantizer (FQ) and a low resolution multiplying digital-to-analog converter (MDAC), which computes and amplifies the residue to be quantized by the following stages. The DC offsets in the comparators (included in the stage FQ) do not affect the overall linearity of a pipeline ADC, if redundancy and proper output encoding are adopted, [127]. Hence, the accuracy is mainly limited by the gain-error of the MDAC in the first stage[1], caused, in conventional SC implementations, by mismatch errors associated with the capacitors and by the finite gain of the amplifier. The accuracy required for this first MDAC is that of the overall ADC and it is progressively relaxed in the following stages [128].

The number of stages in a pipeline ADC depends on the effective number of bits that each stage delivers. Among all available combinations, the 1.5-bit stage configuration proved to be very efficient in terms of design complexity, power and area for full ADC resolutions up to 8 to 10 bits, [127]. Moreover, the 1.5-bit stage resolution reveals to be well suited for the application of the MPA principle since the required residue amplification gain is only two and the offset requirements of the comparators in the local FQ are much relaxed. These non-stringent requirements for the FQ mean that the pre-amplification stage of each comparator, does not need to have a high value and, therefore, parametric amplification can be readily applied, as demonstrated in Section

[1] A minimum resolution *per* stage of 1.5-bit is assumed here.

4.5.2.

As previously explained, the pipelined structure relaxes the accuracy requirements of each individual analog building block. However, their speed requirements can be further relaxed by employing a time-interleaved structure, where two or more ADCs are connected in parallel but operating, sequentially in time. That is, for an overall conversion rate F_s, each individual ADC (designated by channel) operates at F_s/M clock, where M is the number of the time-interleaved channels. Time-interleaved pipeline ADCs have been used to achieve low/moderate resolutions at high sampling rates, [24, 129]. Moreover, additional improvements on area and power efficiency can be achieved by sharing some common building blocks between channels, with similar sizing and operating with 180^0 phase shift.

Despite the speed gains obtained by the time-interleaved operation, the overall resolution achieved by this structure relies also on the matching accuracy between channels. The three main mismatch errors are offset and gain mismatch errors between channels that arise from the individual ADCs, and clock-skew that arise from mismatches in the multi-phase clock generator and in the on-chip clock distribution and buffering. For example, the signal-to-noise-plus-distortion ratio (SNDR) obtained if only clock-skew mismatch errors are considered (i.e., all the other sources of errors are neglected as well as the quantization noise) is limited by, [129],

$$\mathrm{SNDR} = 20\log\left(\frac{1}{2\pi\sigma_{te}\left(\frac{\omega_i}{T_s}\right)}\right) + 3 \quad [\mathrm{dB}] \tag{5.1}$$

and the spurious-free-dynamic-range (SFDR) is determined by

$$\mathrm{SFDR} = 20\log\left(\cot\left(|t_e|\,\frac{\omega_i}{T_s}\right)\right) \quad [\mathrm{dB}]\,, \tag{5.2}$$

which is only valid for two parallelized channels and for a sinusoidal input signal with frequency ω_i sampled with period T_s. Moreover, it is assumed that the clock-skew t_e (associated to a single channel) is Gaussian with zero mean and variance σ_{te}^2. As an example, if only clock-skew mismatch effects are considered (i.e. all the others are neglected as well as the quantization noise) and for a 10 MHz input signal, a standard deviation of the clock- skew of less than 10 ps is needed when 65 dB of SNDR (e.g., in a 10-bit ADC) is targeted. This value is even lower if the input signal frequency is higher. Hence, the time-interleaved based ADC needs a careful design and layout (specially on the clock distribution network) in order to reduce clock-skew mismatches between the parallelized ADC channels. Besides these errors, time-interleaved ADCs also suffers from usual errors of any ADC, e.g, sampling jitter, nonlinearities, quantization noise,

thermal noise, among others.

The architecture of the proposed pipelined ADC, shown in Figure 5.1, is composed of a front-end S/H circuit, with nominal gain of two, followed by six 1.5-bit pipelined stages and by a 2-bit FQ. The remaining circuitry is used to synchronize and digitally correct the quantized data to produce the digital output word. In order to achieve higher operation speed, the proposed pipelined ADC uses a two-channel time-interleaved architecture type, in conjunction with a pseudo-differential structure to improve common-mode rejection, noise immunity and to significantly reduce even-order harmonics. Each

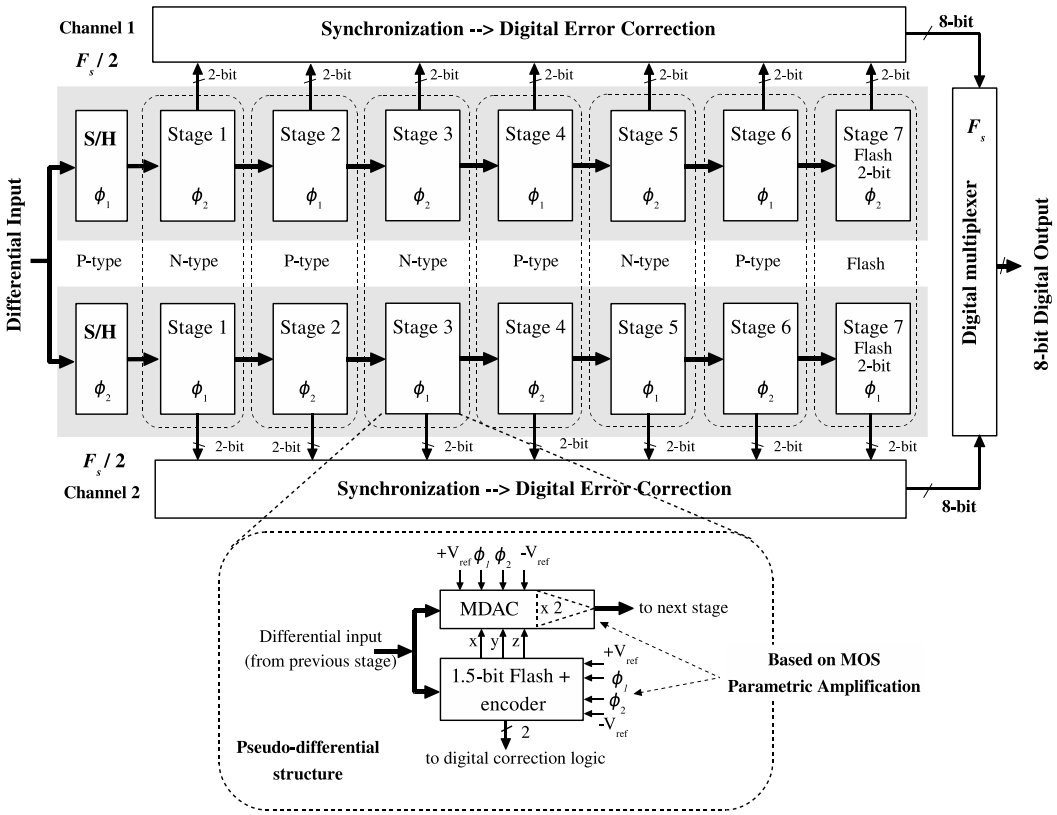


Figure 5.1: Global architecture for the ADC.

stage comprises a 1.5-bit differential MDAC and a 1.5-bit FQ (with two comparators). Both employ MOS parametric amplification either to implement a gain of two in the former and a low pre-amplification gain in the latter. As it will be explained later on, the proposed MDAC does not use any kind of high speed operational amplifiers.

The 14 output bits provided from the seven FQs are then digitally synchronized and, a net resolution of 8 bits is available at the digital output after standard digital correction is applied. As stated before, each interleaved branch operates at half of the speed of the total input sampling rate F_s.

The targeted specifications for this Nyquist type ADC are to achieve at 6 bits of

effective resolution and a maximum conversion rate of 120 MS/s. Due to cost reasons, it is also intended to reduce the area as much as possible.

5.3 Analog Building blocks

The adopted parallelized ADC structure described in the previous Section does not need extra clocking phase, besides the two non-overlapping ones that are normally used in each individual channel ADC. In fact, the pipelined stages in each adjacent channel are equally sized but they operate with a 180^0 phase shift between them. Due to its similarity, both channel stages are jointly designed forming a full time-interleaved pipelined stage, which is proposed in next Subsection. The remaining Subsections, present the most important building blocks with special emphasis on those in which MOS parametric amplification can effectively be employed. As stated before, in Chapter 4, the design strategy of this ADC is built around the extensively use of the MPA principle in the ADC, thus relying on the use of open-loop amplification.

5.3.1 Full time-interleaved pipelined stage

The full time-interleaved pipelined stage, depicted in Figure 5.2, comprises the individual stage of each channel which are stacked together with a vertical symmetry. The stage in each channel employs a 1.5-bit FQ, a pseudo-differential MDAC composed of two half-MDACs (HMDAC) and two SFs (NMOS, in this case). Four SFs are used to prevent the loading effect of the next pipelined stage. It is important to emphasize that their sizing affects the power and the attainable speed of the stage, and the MDAC gain. So, a careful and optimized size design is needed. Finally, the full pipelined stage includes a replica-bias circuit (RBC), shared by both channels, which is used to adjust the common-mode voltage at the SFs' outputs. This is justified by the need to avoid DC accumulation errors through the pipelined stages and to minimize the impact of process-supply-temperature (PVT) variations.

The buffering performed by the SFs has, however, a DC level shifting effect due to the biasing V_{GS} of the common-drain transistor. Hence, to avoid the MDAC from collapsing (due to DC saturation) in the pipeline chain, a structure that alternatively cascades P and N type stages is used. P-type stage uses a PMOS type source follower while the N-type uses an NMOS type SF. As just mentioned, these two versions are needed to deal efficiently with different input and output common-mode voltages resulting from each type of stage, which affect the attainable gain in the MPA structures used.

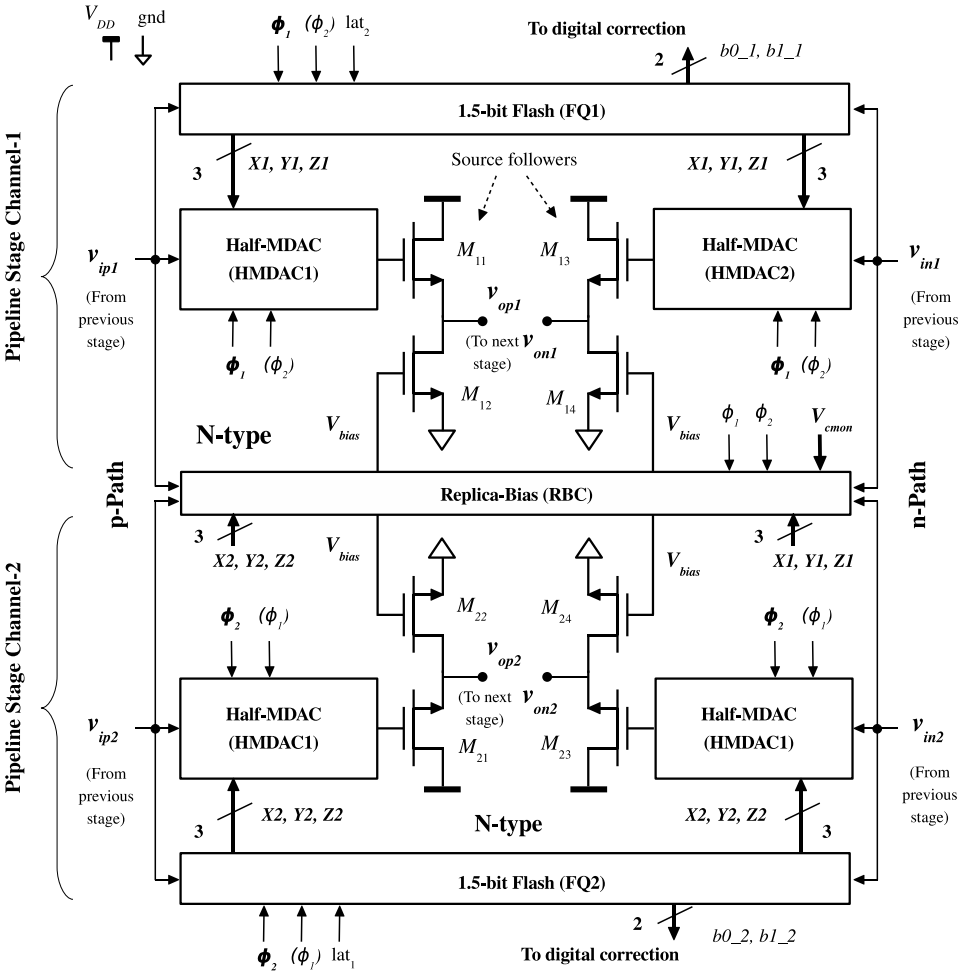


Figure 5.2: Block diagram of a pipelined stage of the ADC (N-type).

N-type stages sample an input signal with a higher common-mode voltage than the P-type stages, and then output a signal with a lower common-mode voltage. So, it is necessary to follow an N-type stage with a stage that does the inverse operation, i.e., increase the common-mode voltage. This is what P-type stages provide. As shown in Figure 5.3, the input and output common-mode levels in a N-type stage are V_{cmop} and V_{cmon}, respectively, i.e. they are inverted if compared with P-type. Therefore, a P-type stage follows an N-type stage, and *vice-versa*. In this design, the output common-mode voltage V_{cmop} for the P-Type stage is set to 0.7 V while for the N-type stage, V_{cmon}, is to 0.35 V. The full stage depicted in Figure 5.2 corresponds to the N-type. The P-type

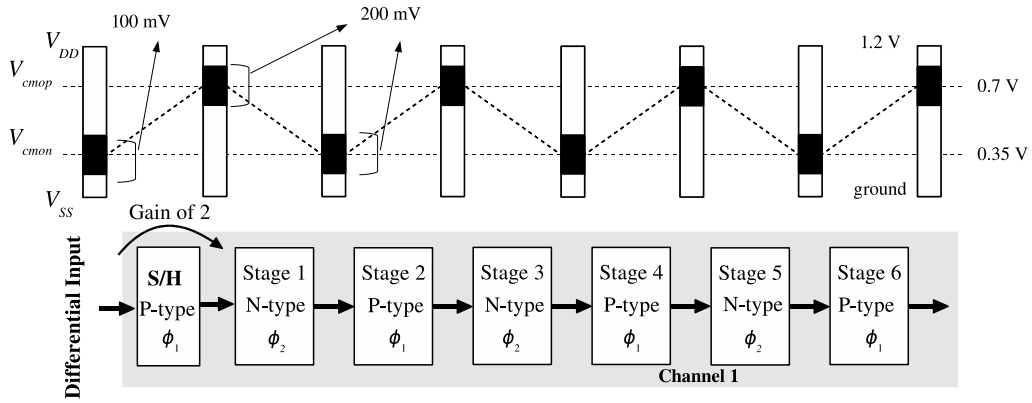


Figure 5.3: The variation of the common-mode voltage along the pipeline.

version is obtained by replacing the NMOS SF to a PMOS version and the MDACs are adapted accordingly to perform the required amplification gain for the applied input common-mode voltage.

The ideal fully differential conversions characteristic of the MDAC (e.g., of channel 1) is depicted in Figure 5.4, which may be described by

$$v_{od} = \begin{cases} 2\,v_{id} + V_{refd} & , -V_{refd} < v_{id} < -\frac{V_{refd}}{4} \\ 2\,v_{id} & , -\frac{V_{refd}}{4} < v_{id} < +\frac{V_{refd}}{4} \\ 2\,v_{id} - V_{refd} & , +\frac{V_{refd}}{4} < v_{id} < +V_{refd} \end{cases} . \tag{5.3}$$

One can observe, that the MDAC characteristic is divided in three segments, which are obtained just by adding or subtracting a reference voltage, but the slope of each one is the same and corresponds to the MDAC amplification gain of two. The selection of the segment to be used is done by the FQ, that enables one of the three control bits X, Y or Z. The digital signals X and Z provided by the FQ are exchanged between the two HMDACs to form a positive (left) and a negative (right) signal path for a fully differential MDAC.

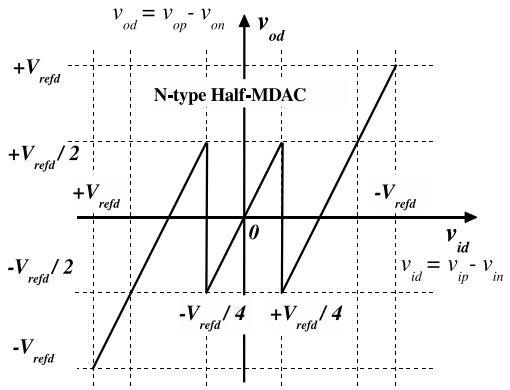


Figure 5.4: MDAC conversion characteristic for a full-scale input ramp signal.

5.3.2 MDAC for the 1.5-bit pipelined stage

As previously mentioned, the pseudo-differential MDAC of each channel is composed of two HMDACs. The schematic circuit of the HMDAC used for the positive signal path, is shown in Figure 5.5, [10], and is included in the N-type stage. The HMDAC schematic for the negative signal path is the same, except that the X and Z control signals provided by the local FQ are exchanged between the two HMDACs, [11]. The block ends with two NMOS-type SFs. The HMDAC structure includes two MPA cells,

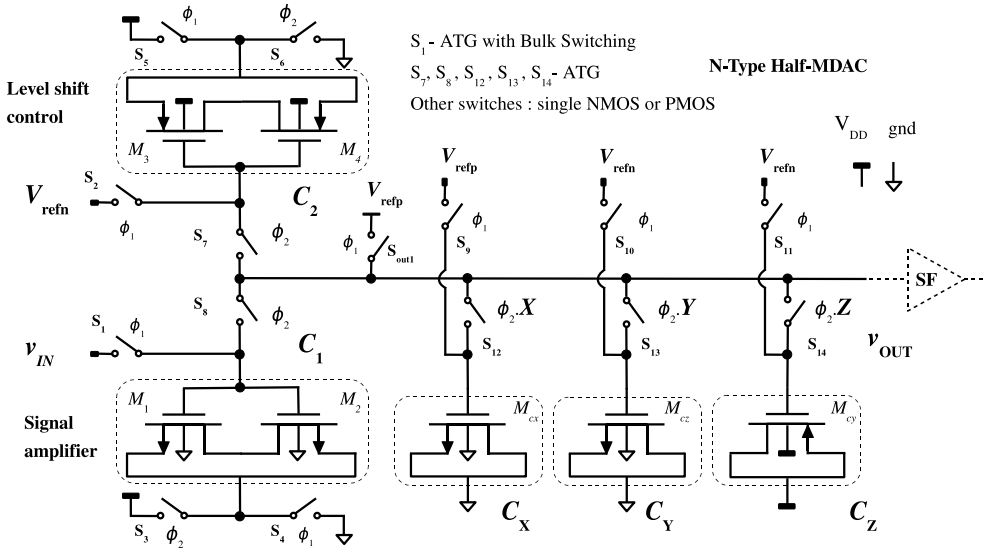


Figure 5.5: N-Type Half-MDAc schematic.

C_1 and C_2 used to perform, respectively, the residue amplification and a DC level shift control, similarly to what was explained in Section 4.5.1. Relatively to the topology analyzed in Section 4.5.1, the circuit of Figure 5.5 includes three additional MOS capacitors (C_X, C_Y and C_Z) to implement the DAC function. Therefore, these MOS capacitors, which are controlled by the FQ encoded outputs, operate in the inversion

Device	Width [μm]	Length [μm]
M_1, M_2	14	1
M_3, M_4	11.5	0.4
M_{cx}	1.33	1.33
M_{cy}	1.5	1.5
M_{cz}	1.66	1.66

Table 5.1: MOS device dimensions for the HMDAC.

region and are responsible for the addition or subtraction of a reference voltage. Only one of these capacitors is selected within a conversion cycle by means of the X, Y and Z code coming from the local FQ. C_Y (NMOS) is required to provide the same loading effect, although it does not contribute with any charge (differentially).

As previously mentioned in Section 4.5.1, due to charge conservation, the input signal applied to the HMDAC block is amplified by the MPA principle through C_1 which tends to add a positive DC level shift as well. To control this shift, i.e. preventing that it exceeds V_{DD}, a PMOS MPA cell, C_2 is used. The combined operation of all five MOS capacitors performs the MDAC function that appears at the SFs outputs, which results in a differential output ideally given by

$$v_{od} = 2 \cdot v_{id} - X \cdot V_{refd} + Z \cdot V_{refd} + Y \cdot 0 \ , \tag{5.4}$$

where v_{od} is the differential output, v_{id} the differential input, and V_{refd} is the differential reference voltage ($V_{refp} - V_{refn}$). The MPA cell has been designed to tolerate $\pm 5\%$ VDD variations and still keeping the residue amplification gain with an error below $\pm 2.5\%$ (for over 6-bit input-referred accuracy). Reference voltages $V_{refp} = 1$ V and $V_{refn} = 0$ V are adopted to minimize C_X, C_Y, C_Z and C_2. The obtained nominal dimensions for the MOS devices of the HDMAC are presented in Table 5.1. A similar structure is implement for the P-type stage.

5.3.3 Flash quantizers for the 1.5-bit and 2-bit stages

All FQs used in the pipelined stages use a pseudo-differential structure and generate an encoded bin output for the digital correction logic. The 2-bit FQ is used at the end of the pipeline chain and has one more comparator than the 1.5-bit version, which is used inside each pipelined stage. The latter also encodes and provides the X,Y and Z control bits needed to drive the digital inputs of the local MDACs.

The 1.5-bit flash quantizer circuit is shown in Figure 5.6. This FQ employs two

comparators, two gated SR latches and an additional digital circuitry to produce the desired output binary encoding (b0 and b1) as well as X, Y, Z bits. Moreover, since a differential structure is used, the only difference between the two comparators is that, their inputs are inverted, to produce a differential threshold voltage with inverted polarity, but with the same magnitude ($\pm\ V_{refd}/4 = \pm 250$ mV). An ideal 1.5-bit

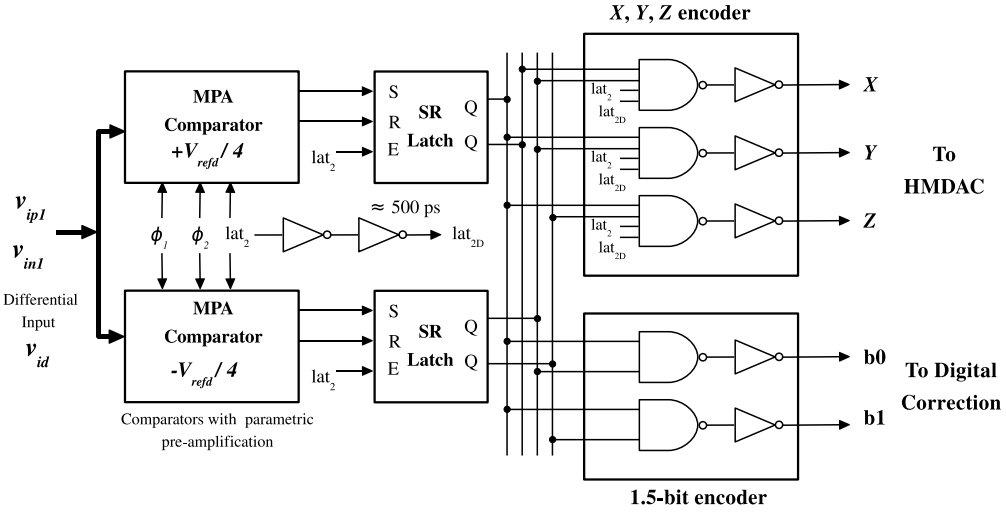


Figure 5.6: Simplified schematic of the 1.5-bit FQ.

FQ exhibits a transfer characteristic as shown in Figure 5.7. The encoded X, Y and Z outputs are also illustrated . A detail view shows that the X, Y and Z code is enabled at the rising-edge of lat_{2D} (about 500 ps delayed from lat_2) and disabled in the falling-edge of lat_2. This prevents erroneously generated X, Y and Z voltage pulses to be applied to the HMDACs, causing wrong charge redistribution due to a wrong MOS capacitor selection. The 2-bit FQ shown in Figure 5.8 is composed of three

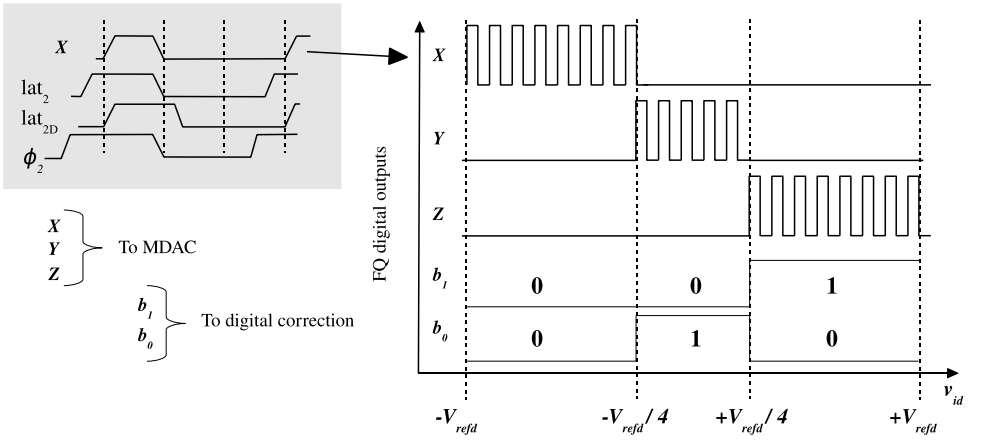


Figure 5.7: 1.5-bit flash quantizer ideal transfer function.

comparators, followed by three digital SR latches, a bubble detector and binary ROM

encoder. The last two blocks are responsible to encode the thermometer code at the output of the comparators into a binary one. In the complete pipeline ADC, two of these 2-bit FQs are used to allow time-interleaved operation.

Similarly to Figure 5.7, comparators $COMP_1$ and $COMP_3$ have exactly the same topology as the ones used for the 1.5-bit FQ, and only differs on the threshold voltage. Comparator $COMP_2$ is a zero-voltage comparator which can be simply implemented by removing the reference voltage SC sampling network.

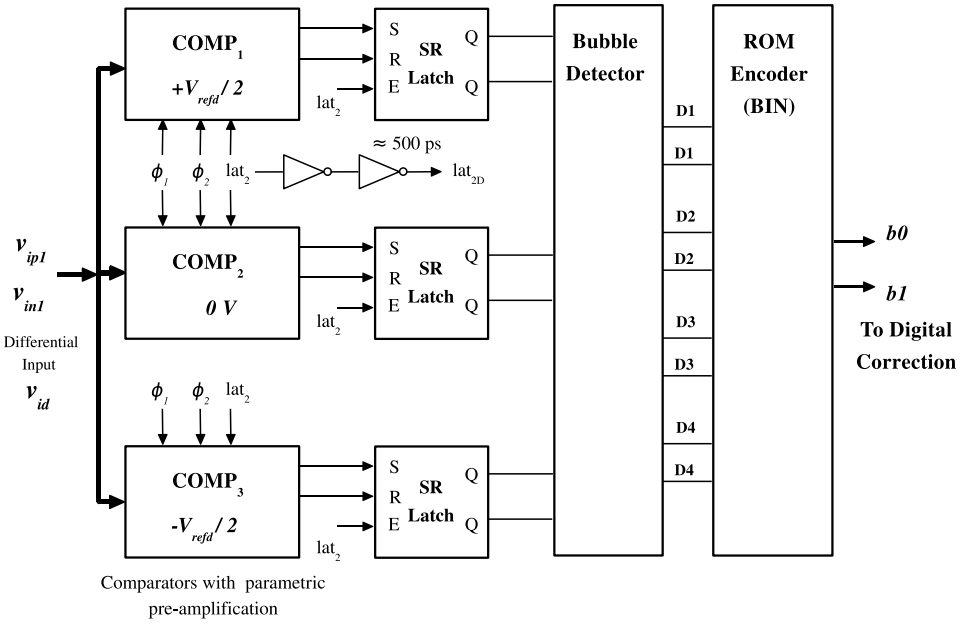


Figure 5.8: 2-bit flash quantizer.

The comparator structure used in both FQs is shown in Figure 5.9. As explained in Section 4.5.2 , each comparator comprises an input SC network followed by a PFBL [8]. The dynamic PFBL enables high-speed operation and provides no static power dissipation. However, due to the input PFBL offset, a dynamic pre-amplification step is embedded into the SC network by using the MOS parametric amplification capability, which further reduces the power dissipation (by avoiding the need of an extra pre-amplifier).

As described in the previous Chapter, the threshold-voltage of the comparator is defined by the ratio of C_1 and C_2 (nominally equal to C_3/C_4) in the sampling phase, ϕ_1, and the desired pre-amplification gain is obtained during ϕ_2 by reducing these capacitance values. The comparator resolves the output state after the rising-edge of $lat2$, which is a rising-edge delayed version of ϕ_2. This delay is about 500 ps and is needed to remove the channel inversion charges generated during the sampling phase, or in other words, the amount of time required to carry out place the required pre-amplification.

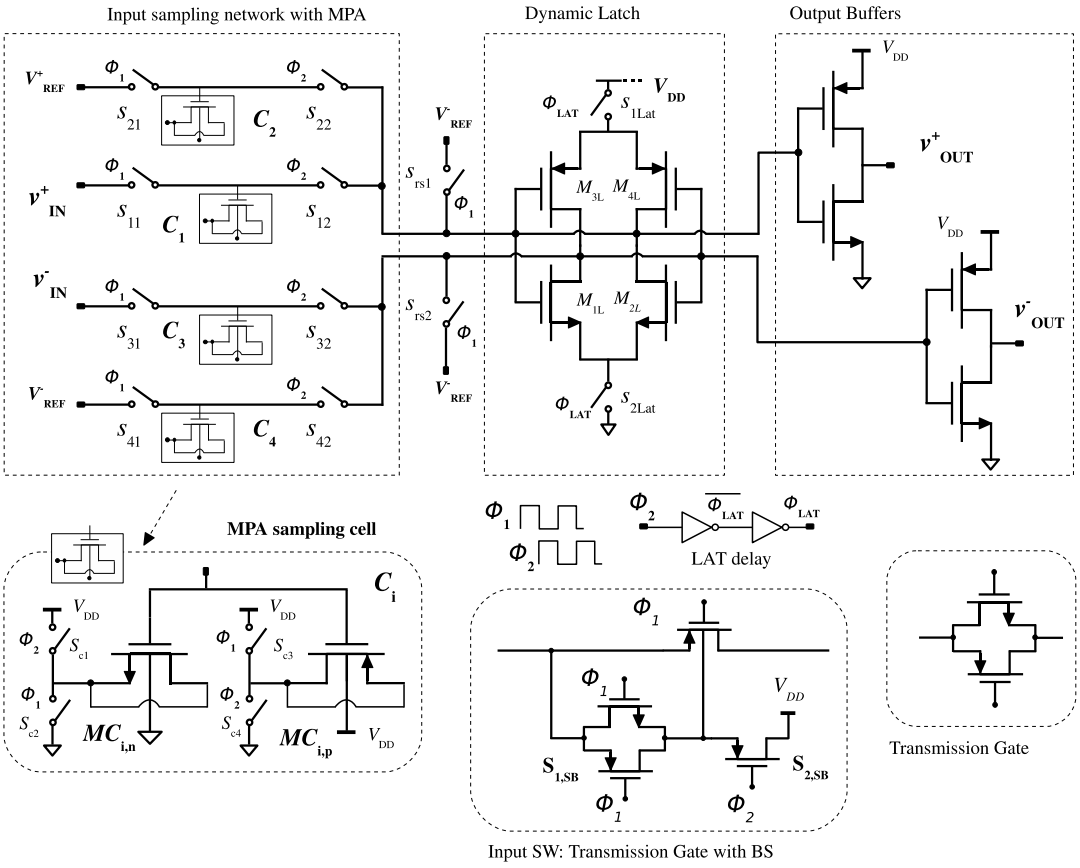


Figure 5.9: Schematic of the comparator used in the FQs.

5.3.4 Replica bias circuit (RBC)

To define the right output common-mode level at the outputs of each pipelined stage, a RBC as the one depicted in Figure 5.10 is used. Both N-type and P-type are implemented to set the V_{cmon} and V_{cmop}, respectively. As stated before, this block is needed to avoid DC accumulation errors in the subsequent pipelined stages and also to minimize the effects of PVT variations.

The N-type RBC shown in Figure 5.10 comprises four scaled N-type HMDACs, a basic amplifier (single stage OTA) and an output SF acting as a replica buffer. Both HMDACs are versions scaled-down by a factor of two from the original HMDAC of Figure 5.5, and additionally, since these HMDACs share the same output node, a switch is inserted at each scaled HMDAC output. It is intended with this RBC, to replicate the operation of the differential time-interleaved MDAC and to control the average voltage at the buffer's output through a negative feedback loop. As a result, a V_{bias} voltage is generated, which is used to control the common-mode voltage at the MDACs buffer's outputs.

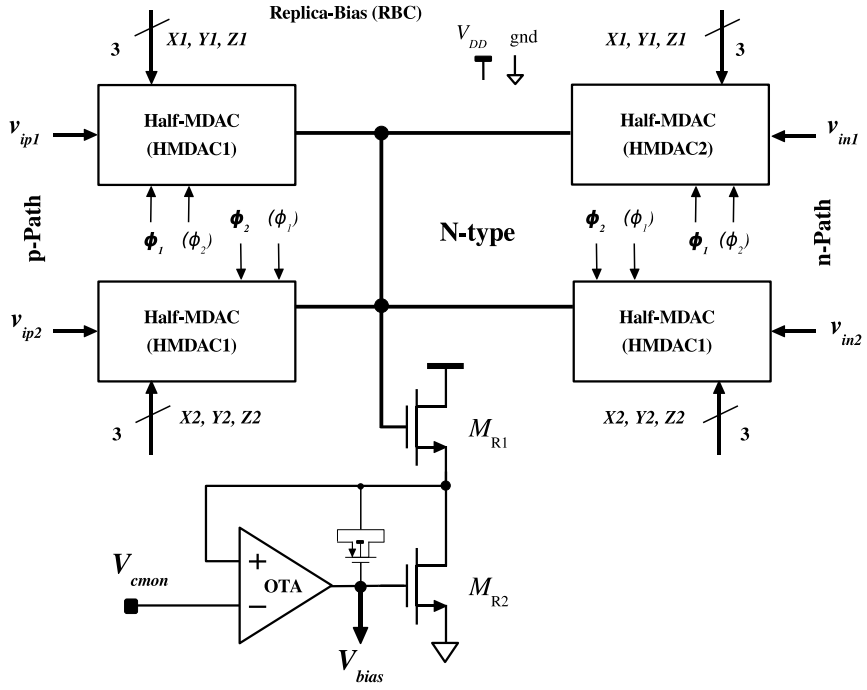


Figure 5.10: Schematic of the N-type RBC block.

The feedback loop used in the circuit of Figure 5.10 to set V_{bias}, uses a PMOS compensation capacitor to increase the stability around the loop formed by the OTA and transistor M_{R2}. The OTA used is depicted in Figure 5.11. It is an one stage single-ended cascode configuration OTA biased by an external current source, I_{bias}, of 2.5

μA. A current mirror is then employed to setup the differential pair bias current of M_6 to about 25 μA.

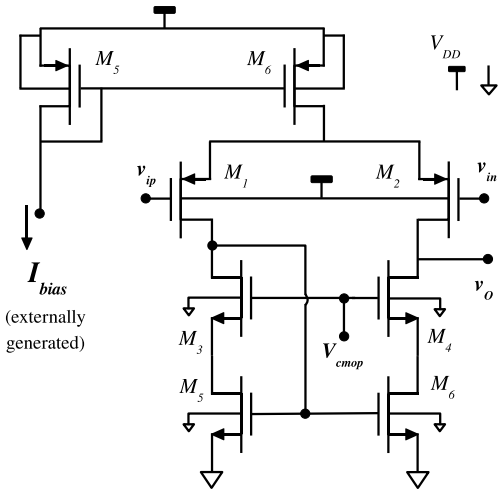


Figure 5.11: Schematic of the OTA used in the RBC block of N-type.

5.3.5 The front-end S/H circuit

The block diagram of the used front end S/H circuit is illustrated in Figure 5.12. It is simply derived from the P-type stage which is forced to operate always in Y segment of the MDAC's characteristic (sample, amplify by two and hold) with some circuitry simplifications. Those include:

- the addition of an output switch for each P-type HMDAC to allow output buffer sharing between the two channels (so only two buffers are required);
- simplification of the RBC block.

Since this front-end S/H has a gain of two, and that the differential input signal range for the pipelined stages is two times higher, this results in an input full-scale of the S/H of about 200 mV_{pp}.

The RBC block for the front-end S/H is depicted in Figure 5.13. This block is slightly different from the RBC used in each P-type stage since it does not use the input signal to generate the biasing reference voltage for the S/H. Instead V_{cmon} is sampled and amplified by the scaled-down P-type HMADC to emulate the operation of the original S/H HMDACs.

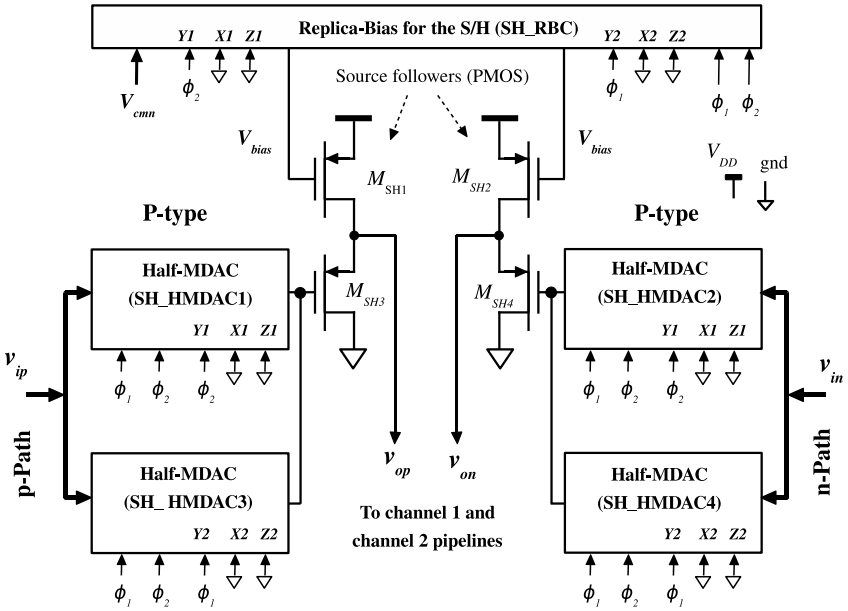


Figure 5.12: Block diagram of the input S/H with gain of two.

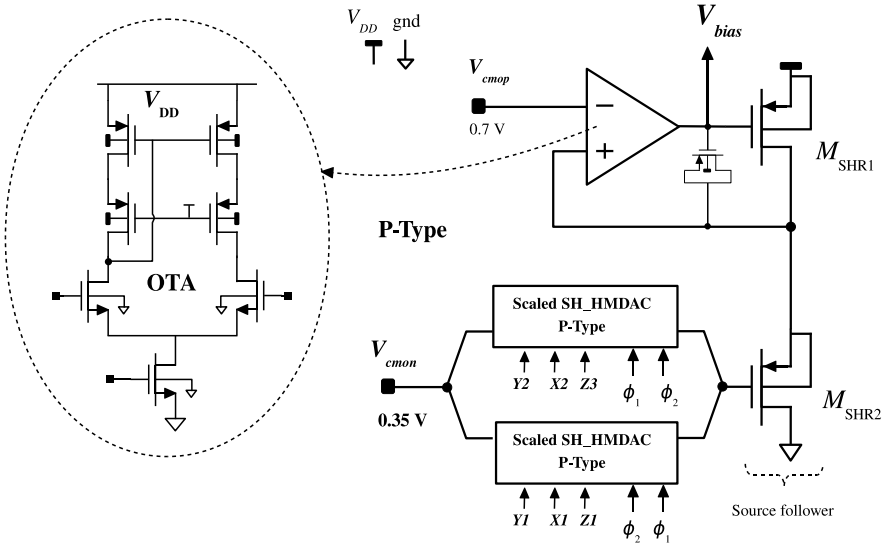


Figure 5.13: Input S/H replica bias circuit.

5.4 Digital logic and Clock Generation

5.4.1 Digital synchronization logic

To synchronize the quantized information related to a particular input sample that is resolved in a pipelined fashion, the circuit shown in Figure 5.14 is used, [127]. It is composed entirely by master-slave D flip-flops which respond on the falling edge of the clock signal. Signals $b_{0<i>}$ and $b_{q<i>}$ represent, respectively, the least and the most

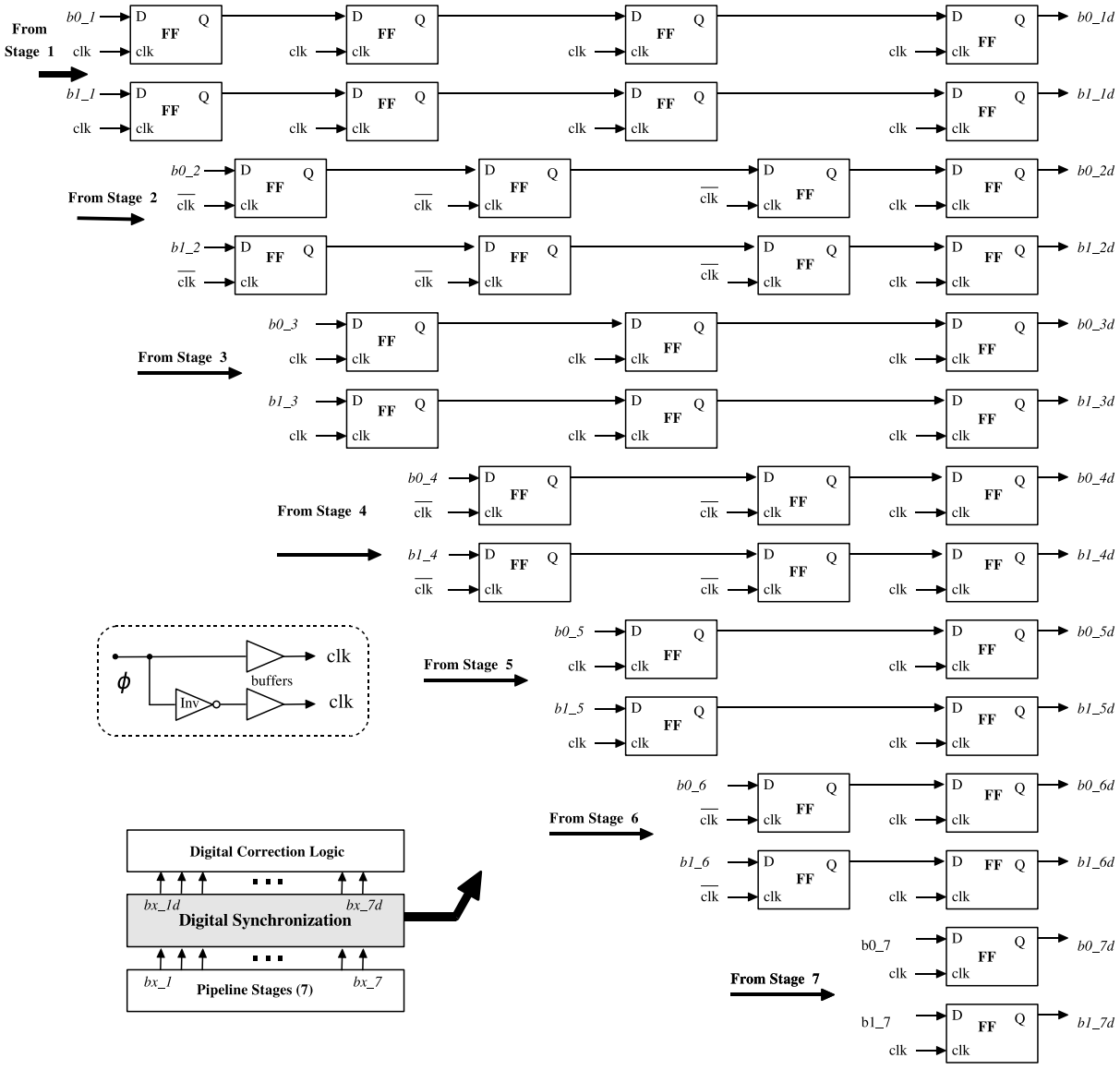


Figure 5.14: Single-channel digital synchronization logic.

significant bits in i-th pipelined stage, and $b_{0<i>d}$ and $b_{0<i>d}$ are the corresponding synchronized (delayed) versions.

5.4.2 Digital correction logic

To greatly reduce the accuracy requirement of the FQs the digital error correction technique is employed with an extra 0.5-bit *per* stage of redundancy, [127]. The error correction circuit for a 0.5-bit overlap between adjacent pipelined stages is depicted in Figure 5.15. The output least significant bit (LSB), b_0, is the delayed LSB of the last pipelined stage (2-bit flash ADC). The remaining bits (b_1, b_2, ... , b_7) result from regular additions as illustrated below. The digital output words coming from the

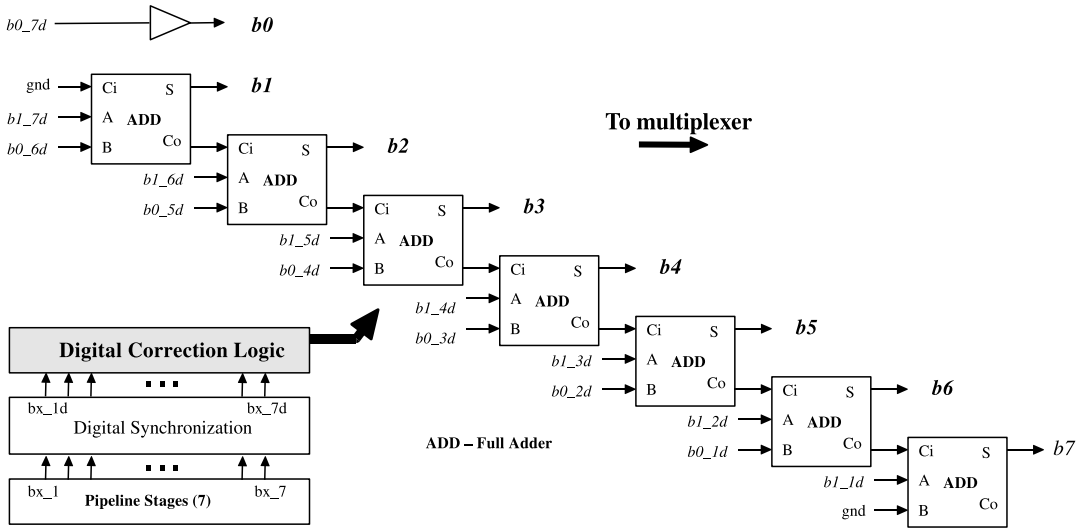


Figure 5.15: Digital error correction logic circuit for a single-channel ADC.

correction error circuit pass through an output register before being ready for external reading. The digital multiplexer, shown in Figure 5.1, is not implemented because of the prototyping nature of the design and also because of the need to have both digital output outside for testing purposes.

5.4.3 Clock Generation

The timing diagram of the clock signals is shown in Figure 5.16, and the clock-phases are generated by the circuit depicted in Figure 5.17 which comprises several differently sized inverters and two NAND gates. An external single-ended input clock, *clk* is converted into two non-overlapping clock-phases ϕ_1 and ϕ_2. The pair of cross-coupled NAND gates is used to generate the non-overlapping signals, inverter delay chains are used to provide the necessary "dead" time and also increase the driving capability.

The falling-edges of clock-phases ϕ_1 and ϕ_2 define the sampling instants of the switched circuits which, when operating in an interleaved fashion, have to be precisely phase-

shifted by 180^0. Since the inverter INV_1 represented in 5.17 adds an asymmetry to the circuit, it will be directly translated into a systematic time-skew error typically of the order of 50 to 100ps (depending on the adopted technology). In order to remove this systematic error an asymmetrical CMOS transmission gate (ATG) can be employed. This ATG is, therefore, capable of equalizing the delay due to inverter INV_1.

ϕ_{1D} and ϕ_{2D} are delayed versions of ϕ_1 and ϕ_2, respectively, and are used by the synchronization and output register circuits. Signals lat_1 and lat_2 are delayed versions of ϕ_1 and ϕ_2, respectively, and are employed by the comparators of the FQs. It is important to emphasize that the clock signals are re-buffered locally, in each pipelined stage.

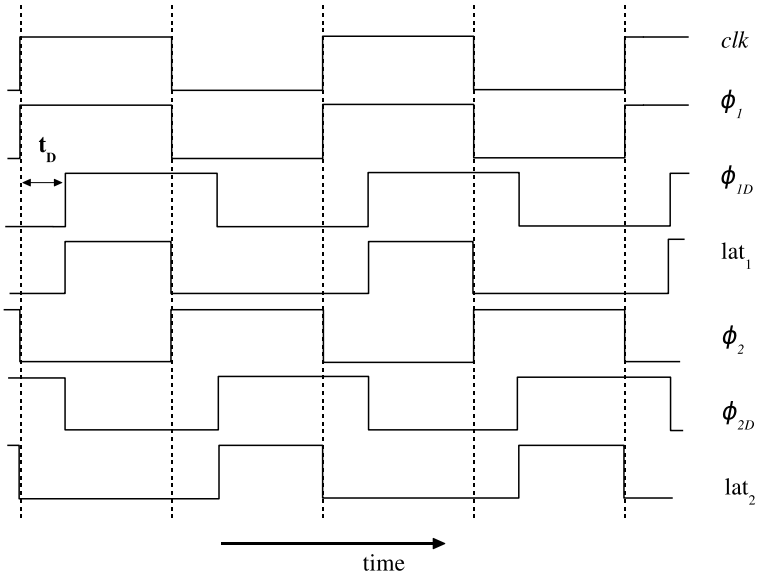


Figure 5.16: Clock phases timing diagram.

5.5 Noise analysis of the pipeline ADC

The pipeline architecture of the ADC dictates that the noise budget for each MDAC stage has to be weighted by the total gain of precedent stages, when referred to the converter's input. Although the Parametric Amplification is intrinsically noiseless, all the different kT/C components (where k is the Boltzmann constant and T is the absolute temperature) have to be taken into account. Additionally, other noise sources increase the total noise power at the SFs outputs during the amplification phase, including substrate noise and noise from the active devices used in the SF circuits. Since the digital noise-coupling contribution through the substrate can be reduced by using well known layout techniques it can be neglected in this analysis. The total ADC

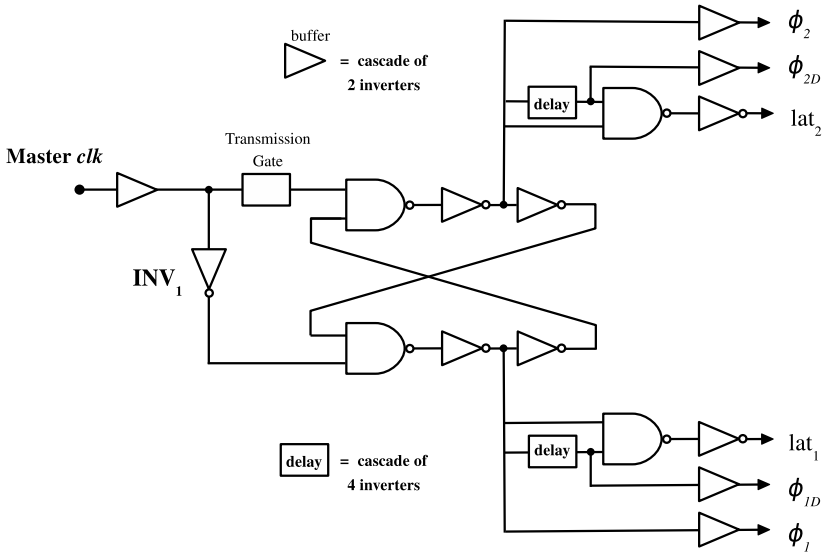


Figure 5.17: Schematic of the clock-phase generator used in the ADC.

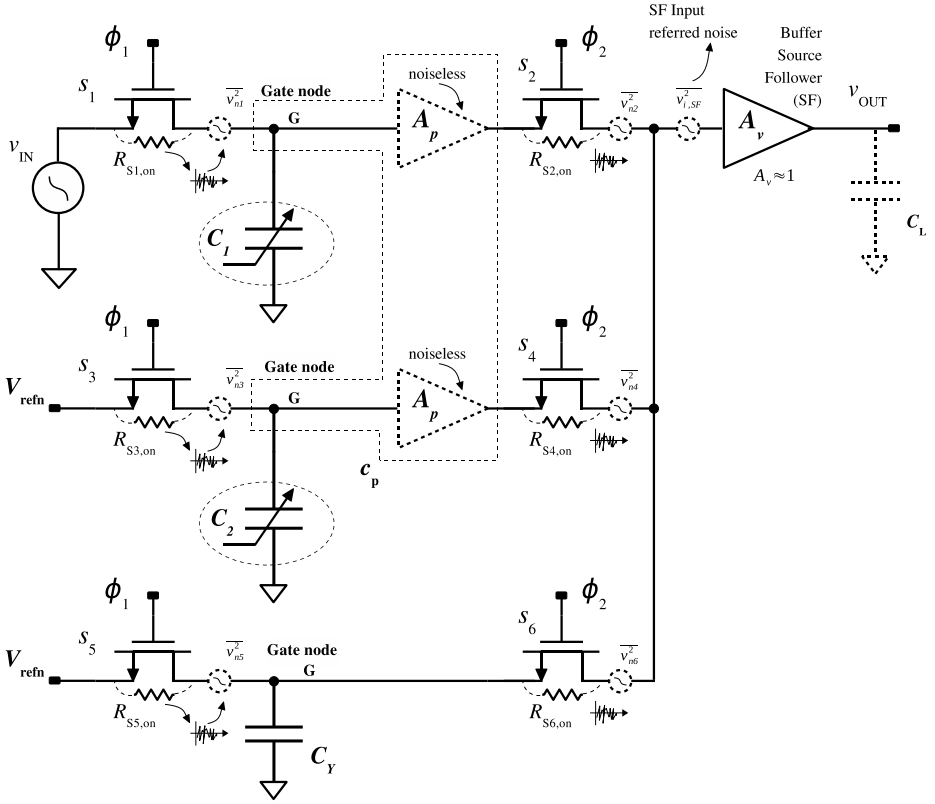


Figure 5.18: MDAC simplified model for noise analysis.

noise performance is mostly dependent on the front-end S/H and on the 1^{st} and 2^{nd} 1.5-bit MDACs. Hence, a simplified MDAC noise model, as the one shown in Figure 5.18, can be used for noise analysis since, as stated before, the S/H can be treated as an MDAC in Y configuration. The input referred noise of each MDAC can be described by

$$\begin{aligned}\overline{v_{i,n,MDAC}^2} = 2 \cdot \left(\frac{kT}{C_1} + \frac{kT}{C_2} + \frac{kT}{C_Y} \cdot \frac{1}{(G_{MPA,C1})^2} \right) \\ +2 \cdot \left(\left(\frac{kT \cdot \gamma \cdot E_{SF}}{C_L} \right) \cdot \left(\frac{1}{(G_{MPA,C1} \cdot G_{SF})^2} \right) \right).\end{aligned} \tag{5.5}$$

The first term represents the different kT/C noise contributions related to the parametric and constant MOS capacitances (biased in the inversion region). In this result, G_{MPA} represents the intrinsic noiseless gain of the parametric structure [99]. The remaining terms describe the noise from the SF at the output node which has to be input referred. The G_{SF} is the gain of the SF, which is slightly lower than unity mainly due to the body-effect, E_{SF} is the SF excess thermal noise factor ($E_{SF} \approx 2$) and γ is the transistor's excess noise factor ($\gamma \approx 1$). Since each stage is connected in a differential schematic, a factor of two is also affecting equation 5.5.

Considering that all MDAC have similar input referred noise level, the overall ADC input noise power is approximately given by

$$\overline{v_{i,n,ADC}^2} \approx \frac{\overline{v_{i,n,MDAC}^2}}{1} + \sum_{j=1}^{6} \frac{\overline{v_{i,n,MDAC}^2}}{4^j}. \tag{5.6}$$

Additional noise components, such as quantization noise, rms jitter noise and DNL "grass" noise, increase the total input referred noise power of the ADC. For an 8-bit ADC and considering a full-scale input signal, a noise power term due to 2 ps-rms jitter (corresponding to a 35.5 $\mu Vrms$ noise voltage) and a DNL "grass" at 1/3 LSB level (corresponding to 230 $\mu Vrms$), the expected SNR is about 40.5 dB (assuming that capacitance values $C_1 = 312$ fF $C_2 = 51$ fF $C_Y = 25$ fF and $C_L = 0.5$ pF are used).

5.6 Overall ADC simulation

The proposed and complete ADC implementation has been electrically simulated. The MOSFET model employed is BSIM3v3 from UMC 0.13 μm Logic 1P8M 1.2 V CMOS process technology, release V2.3P1. In the simulation setup, a differential sinusoidal signal is used to stimulate the ADC. This setup is adequate to analyze the ADC

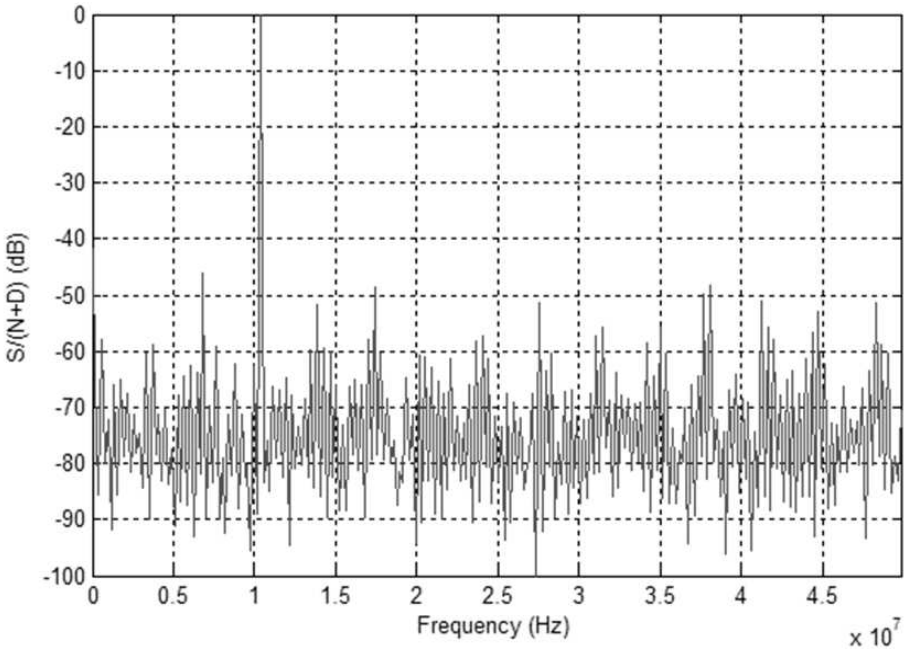


Figure 5.19: Simulated 512 bins FFT spectrum .

performance through the FFT analysis of its output. To avoid windowing, coherent sampling is used. Spectral result is depicted in Figure 5.19. In Figure 5.19 the input frequency (f_{in}) is near 10 MHz and the sampling frequency ($F_s = 2.f_{clk}$) is 100 MS/s. The total simulated rms power dissipation is 18.5 mW (9.4 mW: analog, 4.7 mW: mixed, 3.3 mW: digital and 1.1 mW: generale purpose). The simulated SNDR is 38.2 dB corresponding to an ENOB of 6.0 bits. Note that the dynamic performance parameters are further optimized during layout and post-layout (extracted) simulations, which will be addressed in Chapter 6. This is the main reason why a single FFT is shown. Moreover. since the pipelined stages do not employ any amplifier, the only critical performance parameters that have to be verifies in post-layout simulations are the gain error of the MDACs and the corresponding speeds in all X, Y and Z modes.

5.7 Summary

In this Chapter, the design of a time-interleaved pipeline ADC and the obtained simulated results are used to demonstrate the applicability of discrete-time amplification on such type of data converters.

The Chapter begins by presenting the adopted architecture while the remaining sections describe, in detail, all the major building blocks. It is shown that the parametric amplification is used in the flash comparator and also in the MDAC circuit.

To stabilize DC biasing of the source followers outputs, a RBC block has to be used. However, due to power and area savings this block is shared between the two channels operating in parallel.

The full ADC has been designed and optimized in a 130 nm CMOS technology. The performance of the ADC was checked through electrical simulations and exhaustive FFT analysis over PVT corners. The simulation results show that a resolution of 6-bit is possible to achieve without additional calibration scheme.

Chapter 6

Integrated prototype and experimental evaluation

6.1 Introduction

The design strategy explained in the previous Chapters is applied for building up a pipeline ADC that is fully based on a DT-MPA open loop technique. As already mentioned in Chapter 5, the 8-bit 120 MS/s ADC employs a time-interleaved structure and is fabricated in an 130 nm single-poly eight-metal (1P8M) CMOS process. Powered by a 1.2 V power supply, the prototyped ADC has been integrated only using MOS devices, i.e, the ADC does not contain any resistor, inductors or MIM capacitors, making it suitable for full integration in a digital CMOS technology.

Although the complete design process was presented in Chapter 5, in the next Section the integrated chip is analyzed from the layout perspective, which is complemented by the description of the chip floorplanning. Special emphasis is given in the layout of the blocks where the MOS parametric amplification is embedded.

Last Sections are dedicated to the prototype testing. In first place, a description is given about the design of the Printed-Circuit-Board (PCB) as well as the testing setup. Secondly, the measurements obtained for three chips samples are presented and analyzed. These results includes common ADC performance parameters drawn from FFT and INL/DNL experimental data.

6.2 Integrated prototype

After proper architecture selection, system level simulation, schematic design and electrical simulations, the natural following step in the design flow of an analog integrated circuit is the layout task. The latter is essentially divided in layout planning, layout drawing and, very important, parasitic extraction (both, with floating and ground connected parasitic components) and subsequent post-layout simulation. As stressed before, the MOS parametric amplification level is largely determined by the surrounding overlap, coupling and ground parasitic capacitances, including those intrinsic to the MOS device itself but also those associated to the interconnection lines. Since there is no intention to use any analog calibration scheme in the ADC, the intensive post layout simulations over PVT corners enabled the fine tunning of the size and layout of all MPA cells.

6.2.1 Full chip

Figures 6.1 and 6.2 display, respectively, the chip microphotograph and floor plan of the MPA based pipelined ADC[1]. The plot layout has been overlaid to reinforce the position of the ADC. In order to improve flexibility during the prototype testing

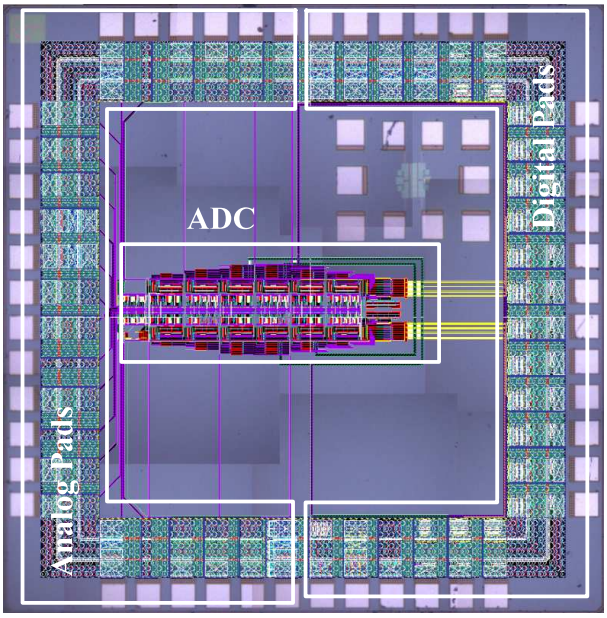


Figure 6.1: Die photo of the test chip with overlaid layout plot.

phase, all the voltage references and current bias (for the different RBC blocks) have

[1]The ADC is not visible since the chip has been automatically filled with dummy metals in order to meet the minimum metal density requirements.

been considered to be provided externally and connected to the chip through dedicated analog I/O Pads. Additionally, the digital multiplexing operation between the output bits of both channels of the time-interleaved ADC has not been included in the chip. This permits that all (8+8) bits, can be fully available outside, which is important for the prototype evaluation of each individual channel. Adding to these outputs (I/O Pads), all the V_{DD} and V_{SS} (ground) connections, makes the chip to be limited by the number of Pads that can be put together in the 1.5 mm x 1.5 mm available area. In this case, the maximum number of pads is 48. These 48 pads are distributed along a ring in which the digital pads are separated from the analog ones. The digital pad region includes the digital circuitry that delivers the digital outputs and inputs into and from the chip. The analog mixed-mode pad region delivers the input signal, supply voltages, current and voltages references to the overall ADC. A breaker pad cell is used between these two areas. The main objective of this cutting cell (gray pad in Figure 6.2) is to separate the V_{DD} and V_{SS} lines from digital and analog sections, thus minimizing the injection of digital noise through the supply bus into the analog circuitry.

The time-interleaved operation of the integrated ADC justifies a careful layout planning in order to reduce the performance degradation due to the mismatches between channels. Therefore, some layout rules are used to minimize these mismatches and they are listed below:

- to contribute for the reduction of clock skew effects, the clock distribution network for both channels is made as symmetrical as possible;
- the complete time-interleaved stage is laid-out altogether using a common-centroid approach;
- the two input signal wires connected to the inputs of the first S/H are made as symmetrical as possible.

In addition to those rules, the MPA cell structure and its sensitivity to the substrate resistance are also reflected in the layout. One of the consequences is the extensive use of substrate contacts rings for each individual MPA cell. This facilitates the reduction of the substrate resistance and consequently substrate noise.

All ground lines are distributed along the overall circuit and connections between devices to ground are made as short as possible. In other to minimize coupling effects from digital and clock signals to analog signal lines, it has been preferred to locate the formers at peripheral region of the ADC while the analog ones are located and center region. Therefore the clock-phase distribution is located at the boundaries of the ADC.

Local clock regeneration buffers are used wherever needed. Due to jitter reasons, the clock generator is placed close to the front-end S/H block. Standard layout procedures are followed, namely, dummy diodes are added to prevent antenna effects, wherever needed.

The total ADC effective area, excluding references generation which are feed externally, is approximately 0.12 mm^2.

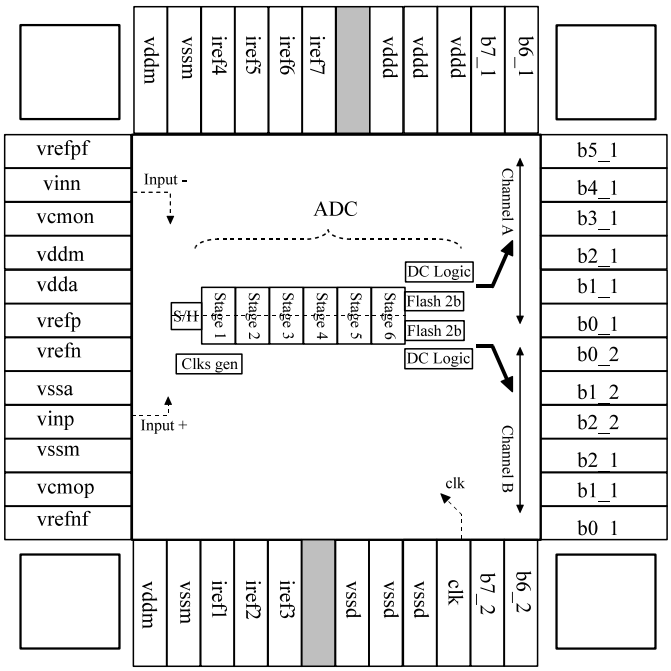


Figure 6.2: Chip floorplan.

Table 6.1 describes the pinout represented in the floor plan diagram of Figure 6.2.

6.2.2 Layout of the different building blocks

As indicated before, both pipeline stages from each channel of the time-interleaved ADC are merged altogether forming a compact layout of a full interleaved stage. Two layouts are implemented respectively for the N-type stage and P-type stage, which are shown in Figure 6.3. In the latter, it can be identified the full MDAC (formed by the two MDACs of both channels) circuit, the RBC block at left and two 1.5-bit FQs on the top and on the bottom. The digital part (encoders, ROM, digital correction) are placed at the top and at the bottom in order to be as far as possible from the main analog signal processing path, which is mainly concentrated at the center of the ADC. The layout of the two stages types is configured to facilitate the connection between two consecutive stages.

Pin Name	Value	Description
vdda	1.2 V	analog positive power supply voltage
vddm	1.2 V	mixed mode positive power supply voltage
vddd	3.3 V	digital positive power supply voltage
vssa	0 V	ground reference
vssm	0 V	ground reference
vssd	0 V	ground reference
vrefp	1.0 V	positive reference voltage for the HMDACs
vrefn	0 V	negative reference voltage for the HMDACs
vrefpf	0.7 V	positive reference voltage for the flash quantizers
vrefnf	0.35 V	negative reference voltage for the flash quantizers
vcmop	0.7 V	P-type output common-mode voltage
vcmon	0.35 V	N-type output common-mode voltage
iref1...iref7	2.5 μA	bias current for the replicas bias blocks
vinp	100 mVpp	positive input signal
vinn	100 mVpp	negative input signal
b7_1 ... b0_1	digital	binary output for channel #1. MSB...LSB.
b7_2 ... b0_2	digital	binary output for channel #2. MSB...LSB.
clk	digital	clock input

Table 6.1: List of all Pads used in the ADC prototype.

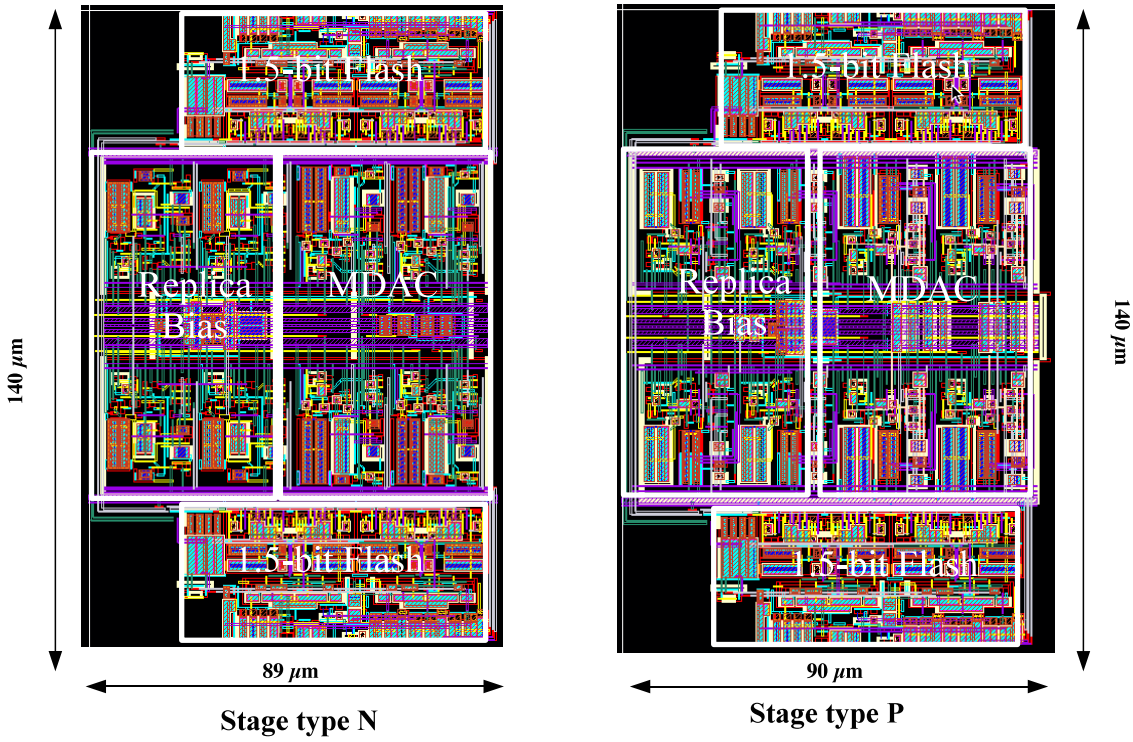


Figure 6.3: Layout of the time-interleaved pipeline stages (type N and P).

Each 1.5-bit FQ included in each pipelined stage comprises two comparators, SR latches and digital encoders. Figure 6.4 shows that the digital processing part is separated from the analog processing part in order to reduce digital noise coupling. To improve this isolation a substrate connection ring is used around the two regions. Also shown in Figure 6.4 is the 2-bit flash used in the last pipelined ADC stage. This is slightly different from the latter since it needs one extra comparator but, at same time, does not need the XYZ encoding logic. However, the same layout precautions are followed.

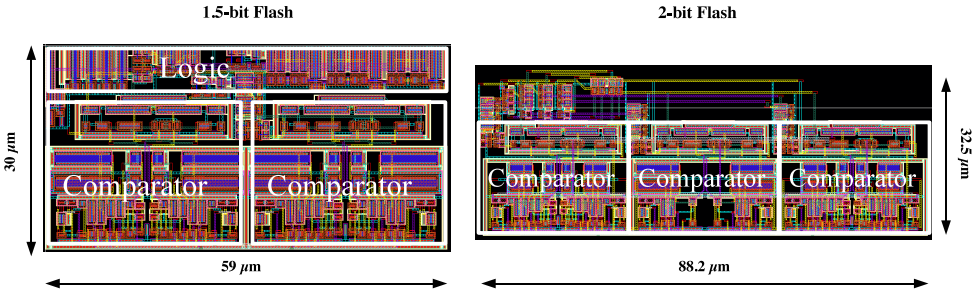


Figure 6.4: Layout plot of the 1.5-bit and 2-bit FQ.

Figure 6.5, shows a layout detailed view of each comparator used in the 1.5-bit FQ. It is indicated in the Figure, the areas where the MPA devices are laid out. It can also be observed that these MPA devices are strongly isolated, through the use of substrate contacts rings, from the switching noise. Once again, the objective is to reduce, as much as possible, the substrate noise coupling effects. The presented 50 mV comparator occupies an area of 23 μm by 29.4 μm.

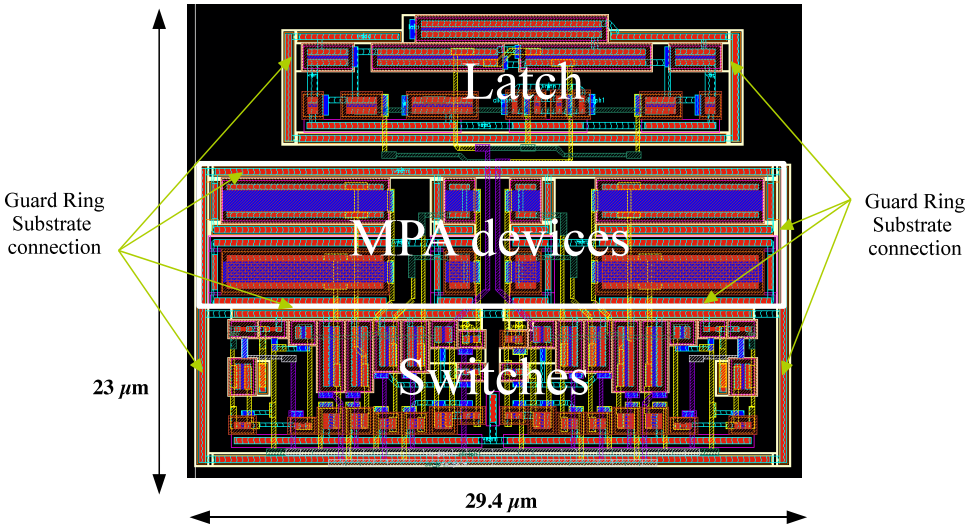


Figure 6.5: Layout plot of one 50 mV comparator.

The full stage MDAC is shown in Figure 6.6. It comprises two channel MDAC (which

constituted by two HMDACs) due to the differential structure employed, as explained in Section 5.3.2. Their layouts are implemented close to each other, in order to improve matching. Also visible in the Figure is a detailed view of one HMDAC. Similarly to the comparator explained previously, the MPA devices are isolated from the remaining part of the circuit by means of a substrate contact guard ring. The total area of the N-type full-stage MDAC is 74.5 μm x 42 μm while for the P-type the area is slightly higher, i.e., 74.5 μm by 45.7 μm.

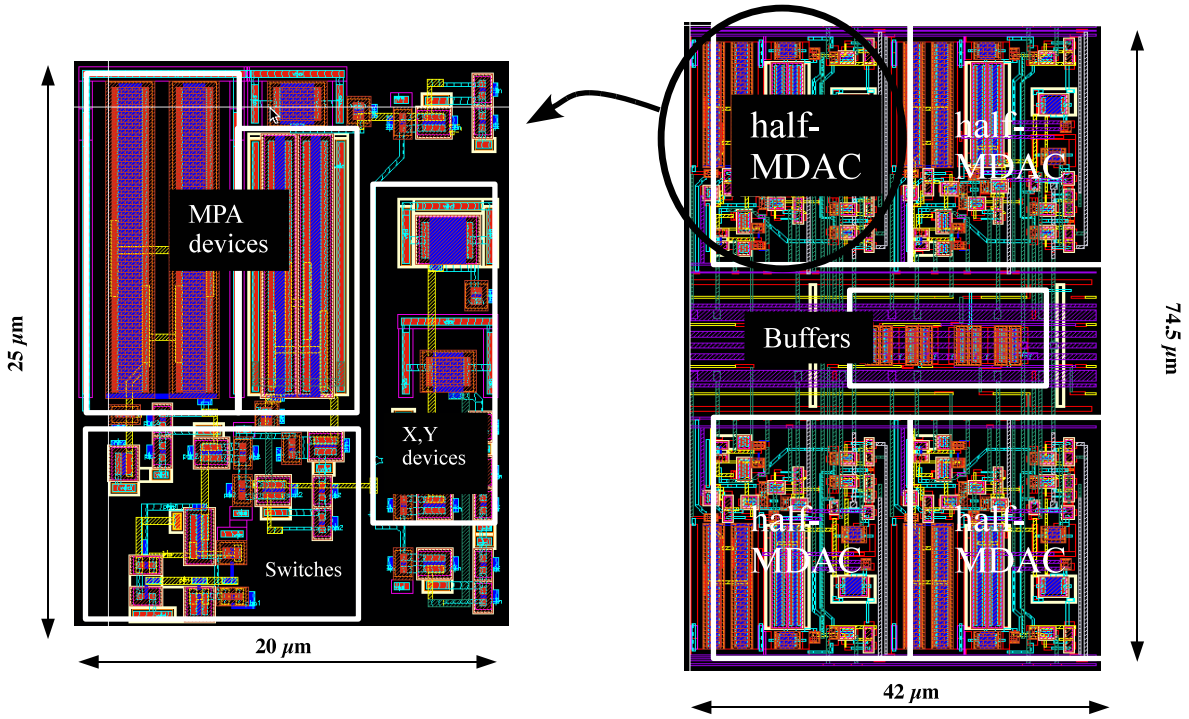


Figure 6.6: Layout plot of a time-interleaved MDAC of a N-type stage.

The complete time-interleaved stage also includes a RBC block. For the N-type case, the RBC layout is depicted in Figure 6.7. It is a replica of the main MDAC block, without the FQ, which is connected to a scaled SF. The auxiliary OTA connected to the SF sets the biasing voltage. All of these circuit are highlighted in the layout diagram. The RBC for the N-type occupies a similar area of the full-stage MDAC. The same happens for the P-type stage.

Finally, the layout of the input S/H is displayed in Figure 6.8. It has a similar layout structure of the standard full MDAC, but with a lower area as a result of a design scaling down approach (due to power savings reasons). Since this block has to achieve the overall resolution of the ADC, special care is taken during layout to implement a fully symmetrical arrangement and to protect all MPA devices with individual substrate guard rings.

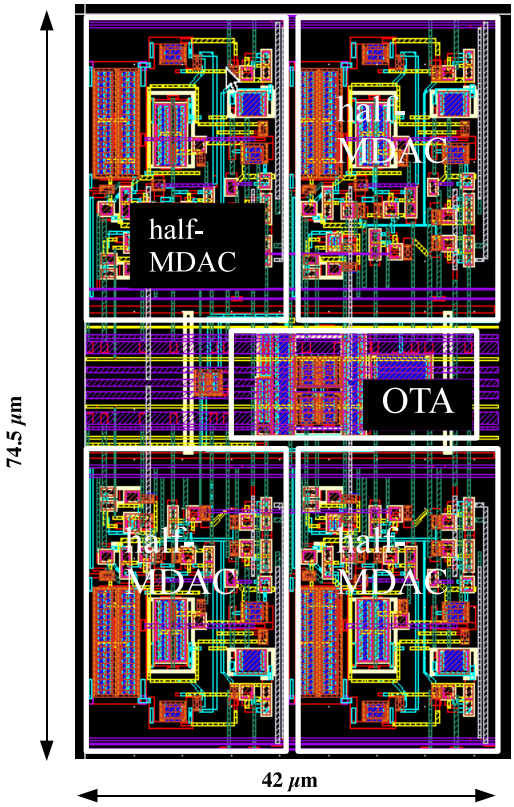


Figure 6.7: Layout plot of a RBC block used in a N-type stage.

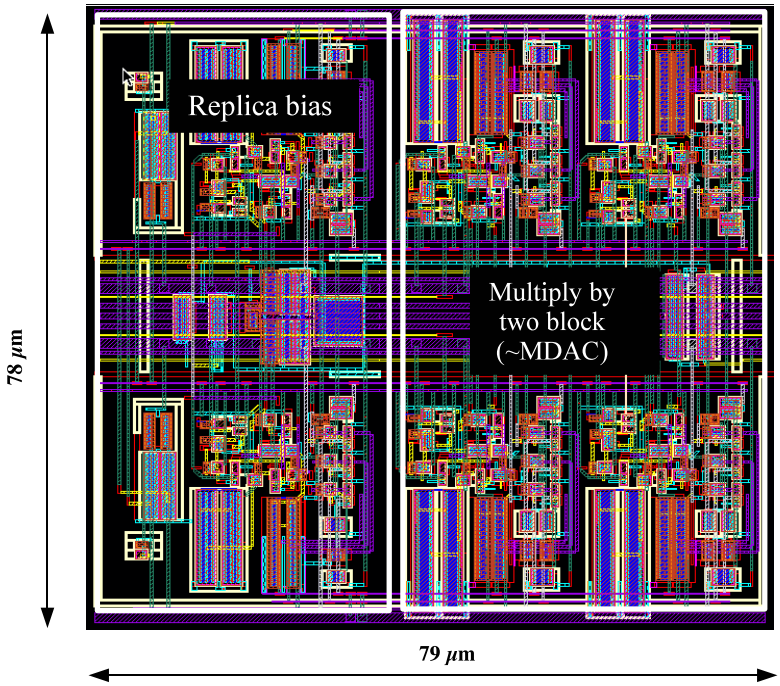


Figure 6.8: Layout plot of the front-end S/H.

6.2.3 Overall ADC post layout simulations

After validating the performance of the ADC at schematic level the layout process of the complete ADC targets the minimum possible area of implantation. However since the parametric amplification principle is very sensitive to the load conditions, the gain accuracy of the MPA cell used inside the P-type and N-type stages have to be optimized through exhaustive extracted layout simulations. All parasitic capacitances values above 0.1 fF have to be considered and extracted with the "coupled-mode" flag enabled. Furthermore, the power dissipation and speed of the SFs have been subjected of additional optimization.

Figure 6.9 shows an example of complete simulation of the extracted layout of the 8-bit ADC in which, the input signal frequency (f_{in}) is 10 MHz and the effective sampling frequency ($F_s = 2.f_{clk}$) is 100 MS/s. The total RMS power dissipation is 15.5 mW (5.52 mW: analog, 6.1 mW: mixed, 3.9 mW: digital). The simulated SNDR is 34.6 dB corresponding to an ENOB of 5.5 bits but, after adjusting the THD by optimizing the common-mode levels V_{cmop} and V_{cmon} voltages, the simulated SNDR improves to more than 41 dB, corresponding to an ENOB of 6.5 bits. Note that the common-mode voltages are set externally in this prototype version but an on-chip servo-loop mechanism also based on a replica circuit can be easily designed to automatically adjust their values to the optimum levels.

Two different figures-of-merit (FOM) for ADCs are usually employed. The first one, FOM_1, represents the used energy per conversion

$$\mathrm{FOM}_1 = \frac{P}{2^{\mathrm{ENOB}} \cdot F_s} \quad [\mathrm{pJ}] \tag{6.1}$$

where F_s is the sampling rate and P is the total power dissipation. A more complete figure-of-merit, FOM_2, includes the area, A (in mm^2)

$$\mathrm{FOM}_2 = \frac{P}{2^{\mathrm{ENOB}} \cdot F_s} \cdot A \quad \left[\mathrm{pJ} \cdot \mathrm{mm}^2\right] \tag{6.2}$$

State-of-the-art high-speed 8 bit ADCs reach FOM_1 of the order of 0.5 pJ and FOM_2 of the order of 0.3 to 0.5 $\mathrm{pJ.mm}^2$, [130]. This post-layout ADC simulation reveals a reasonable FOM_1 of 1.7 pJ but an expected FOM_2 of the order of 0.25 $\mathrm{pJ.mm}^2$. Moreover, the circuit is all made of transistors and does not require neither resistors, inductors, or MIM capacitors and all stages are made equal (no scaling-down was used).

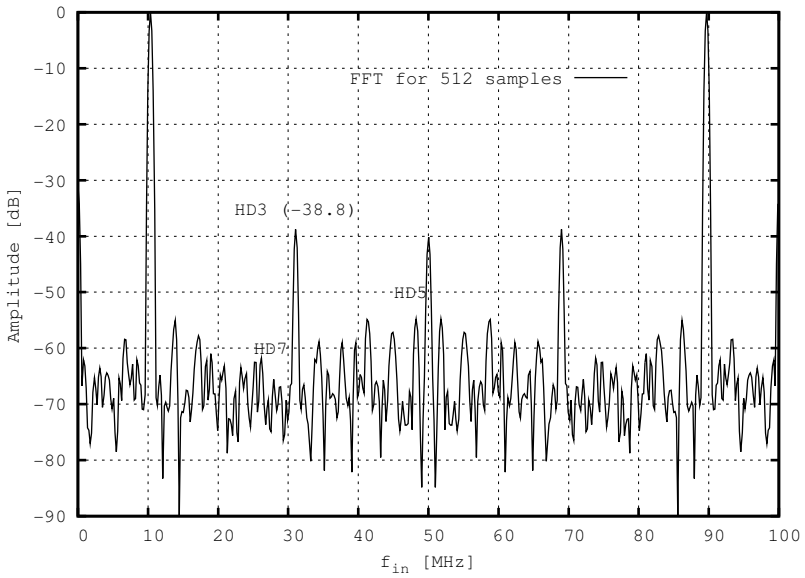


Figure 6.9: Example of a FFT obtained from post layout simulations.

6.3 Test bench

A direct bonding of the chip approach is followed in this design implied a dedicated PCB design. Additionally, this PCB also have to include several voltage and current references and biasing to allow the testing of the prototype ADC. Note that, in a final version of the circuit all of these bias and references can be included on-chip. Next, the PCB is presented as well as the test setup.

6.3.1 Printed circuit board used in testing

The designed PCB test board used to test the chip is shown in Figure 6.10. Note that, since 3 samples were tested, 3 of those PCB have been fully assembled. In Figure 6.10 it is outlined the position of the main PCB areas: voltages and current references, input signal and clock input, digital output, and the area where the chip is glued.

To facilitate the direct chip-on-board assembly, the PCB is coated with a 0.01 mil layer of chemical Gold to allow the Gold wire bonds to adhere to the PCB surface with low resistance contacts. To improve PCB performance, namely in terms of high frequency signal testing, a 4-layer PCB is designed where the two outer planes are assigned for routing and the inner planes are dedicated to the ground and power supply. In the latter case, the inner plane is split into two distinct and separated areas for the analog and digital voltage power supply, respectively, 1.2 V and 3.3 V. To ensure an adequate

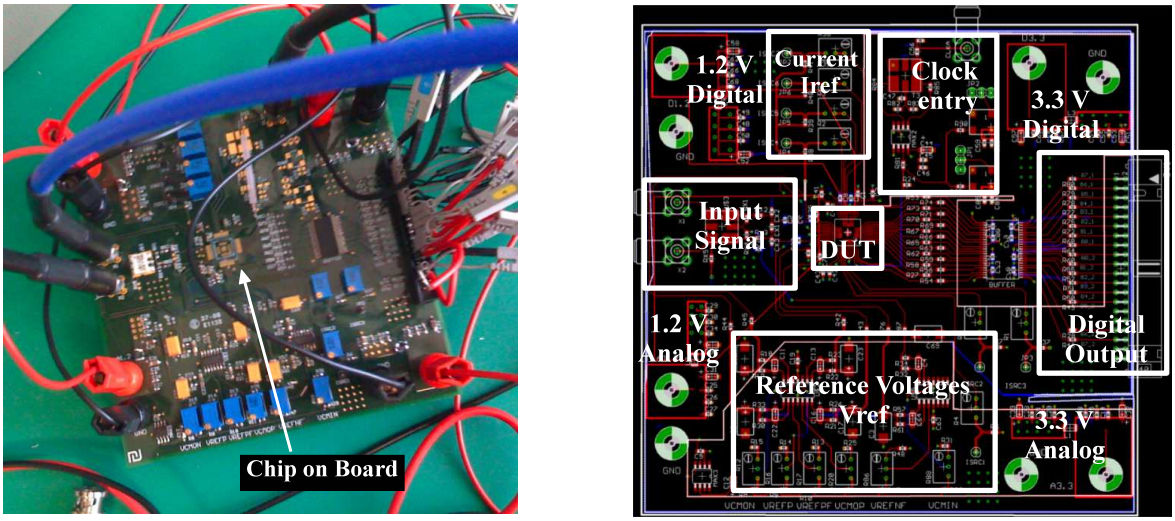


Figure 6.10: PCB used for testing the chip, which was directly wire-bonded to the board.

decoupling of all power supplies and all biasing DC inputs, set of 100 nF, 1 μF, and 10 μF SMD capacitors were placed not only at the input power supplies but also as close as possible to the chip pins.

The input differential signal is feed to the ADC through an RF transformer. A common mode voltage is added to the signal through the center tap of the transformer secondary winding. The differential ADC inputs are bypassed by a 100 pF capacitors to minimize S/H glitches and to improve frequency performance.

6.3.2 Test setup

The complete test setup is depicted in Figure 6.11. The power supply voltages are provided by a programmable power supply, Agilent HP 6624A, with current limiting. The input signal to the board is generated by a Marconi Signal Generator 2041. Due to the necessity of generating a low-jitter signal clock, a Stanford Clock generator CG635 with less than 2 ps jitter is used. The clock is directly connected to the chip board since this equipment support both 50 Ω or high impedance terminations and it also supports CMOS digital output voltage up to 3.3 V.

The signal and clock generators are programmed through the GPIB interface to allow coherent sampling of the input signal. The output bits from the digital output buffers are connected to a logic analyzer, Agilent 16702B, which collects the data coming from both channels. Through the GPIB connection the data is automatically transfered to a LABView program running on the PC. This program is responsible to set all

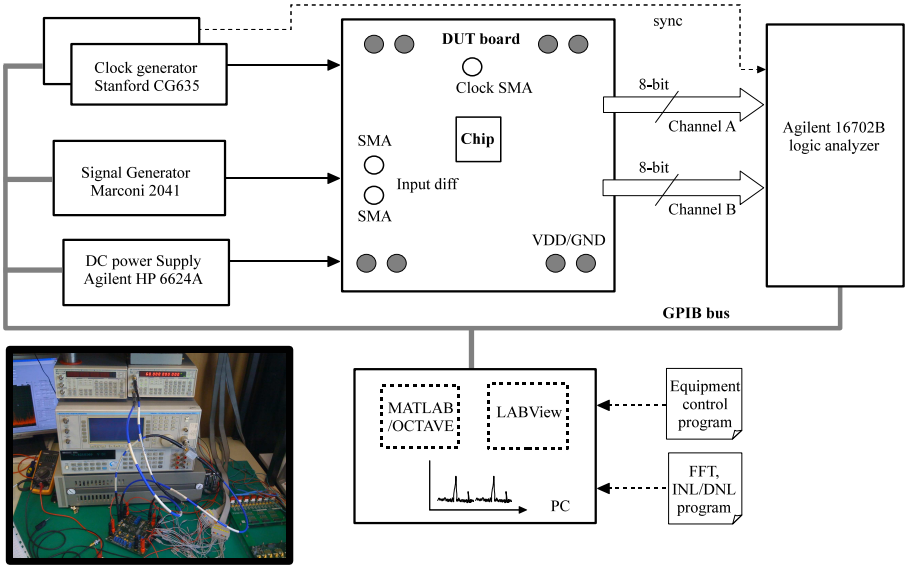


Figure 6.11: Block diagram of the experimental test environment.

equipments, including power supplies. It also presents and calculate dynamic ADC key performance parameters obtained from FFT calculations, such as SNR, THD, among others. Additional results as DNL and INL are calculated from MATLAB or OCTAVE dedicated programs, based on data delivered by the LABView environment.

6.4 Measurement Results

As stated in Section 6.2, the ADC features an active area of 0.12 mm^2 where only MOS devices are used. It dissipates less than 14.5 mW (average power) (at 120 MS/s and 1.2 V supply). This total power, which excludes the external band-gap references and reference buffers, is distributed through the digital correction (29%), the MPAs cells and source followers (34%), and the comparator's PFBL and clock buffers (37%). Notice that this ADC was not power optimized since all stages are equally sized to minimize layout effort. Since the SNR of a pipeline ADC is mainly determined by the front-end S/H, 1^{st} and 2^{nd} stages [127], by downscaling the others, power savings up to 50% could be reached in each MPA unit.

Figures 6.12 and 6.13 show the measured DNL and INL errors obtained with one chip sample running at 120 MS/s. They are within 0.8/+1.4 LSB and ± 2.0 LSB, respectively. For the same chip sample unit, Figure 6.14 displays a measured FFT for a 20 MHz input signal frequency (f_{in}) and 120 MS/s sampling frequency. This particular sample achieves a peak SNR of 39.7 dB, a SFDR of 49.3 dB and a peak THD of 47.5 dB, corresponding to an ENOB of 6.2 bits.

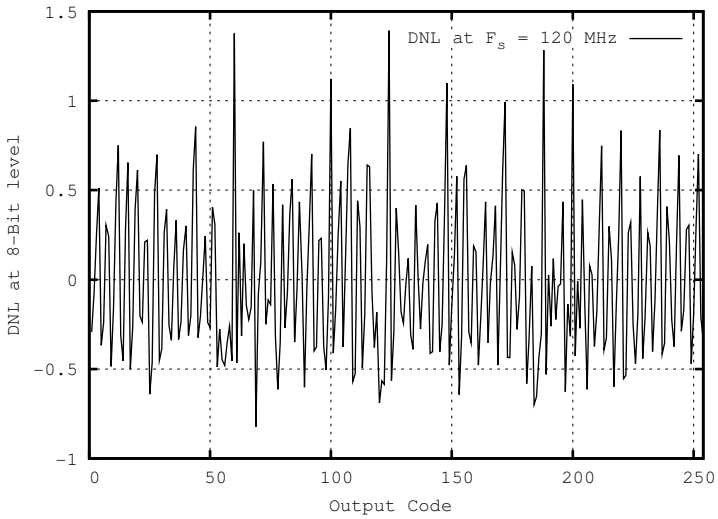


Figure 6.12: DNL experimental results from one sample.

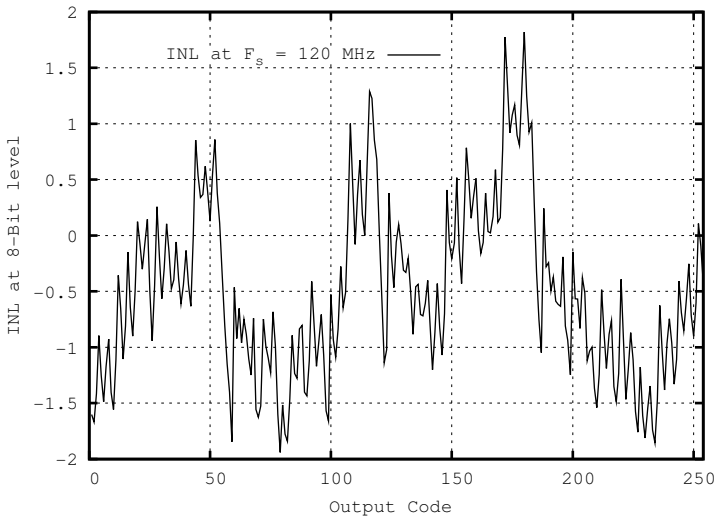


Figure 6.13: INL experimental results from one sample.

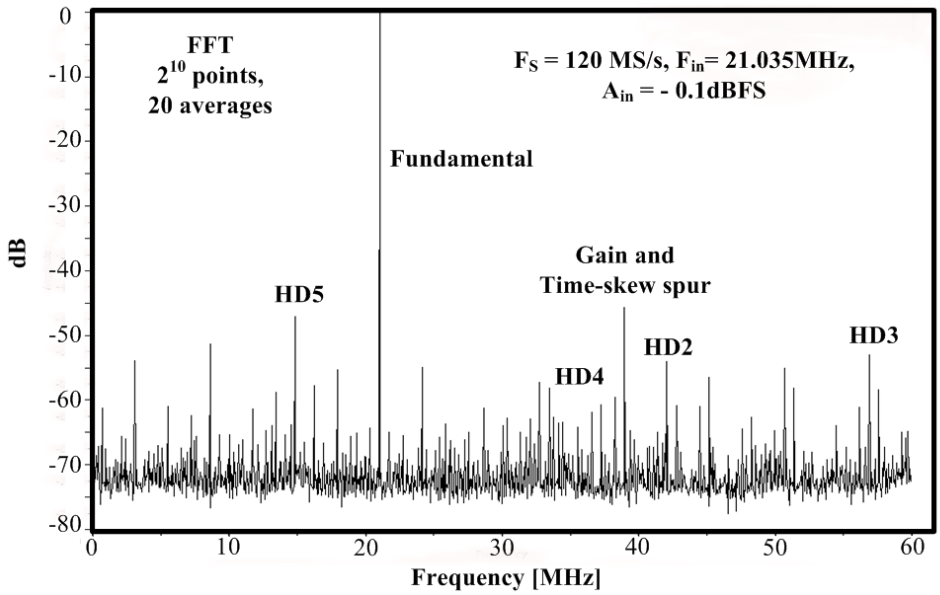


Figure 6.14: Calculated FFT from experimental data obtained from one sample.

The dynamic performance of this ADC was experimentally evaluated in 3 different samples (mounted using direct-bonding) by sweeping the input signal frequency at F_s = 120 MS/s and by sweeping the sampling rate for f_{in} = 20 MHz. Note that the exact value of the used frequency are determined by the coherent sampling testing approach. For F_s = 120 MS/s, the ADC exhibits a flat ENOB higher than 6 bit (with

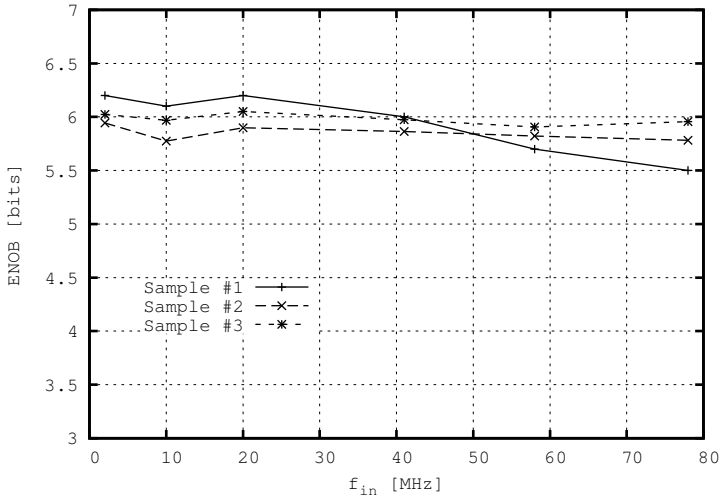


Figure 6.15: Measured ENOB versus f_{in} (-0.1 dBFS and F_s = 120 MS/s). Results obtained form 3 different chip samples.

a peak of 6.2 bit) up to f_{in} = 41 MHz, as shown in Figure 6.15. The same flatness behavior in the ENOB is achieved for sampling rates up to 130 MS/s (Figure 6.16). Due to the mismatch and time-skew between channels of the time-interleaved ADC,

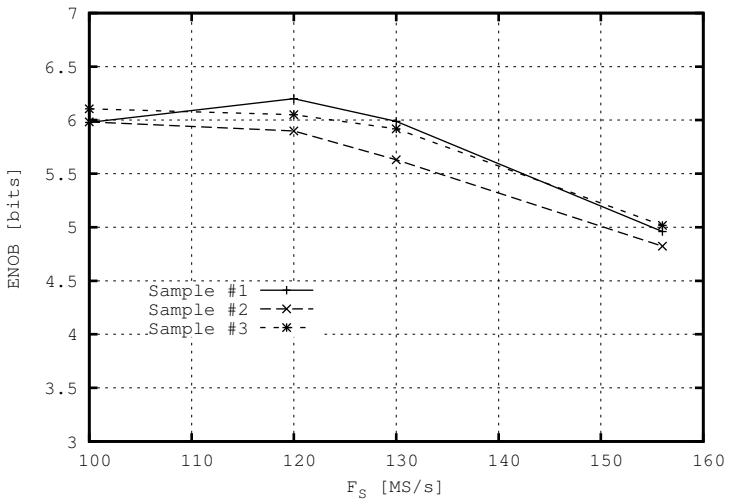


Figure 6.16: Measured ENOB versus F_s (-0.1 dBFS and f_{in} = 20 MHz). Results obtained form 3 different chip samples.

Sample	Total ENOB [bit]	ENOB difference between channels [bit]
#1	6.2	0.1
#2	5.89	0.4
#3	6.03	0.3

Table 6.2: Difference between ENOB from both interleaved channels.

the ENOB achieved by each individual channel presents variation between them that ranges from 0.1 bit for the best sample, to 0.4 bit for the worst one, as indicated in Table 6.2. For the same conditions as used in ENOB case, Figures 6.17 and 6.18, show the THD variations obtained with the three tested samples. For F_s = 120 MS/s, the ADC exhibits, approximately, a flat THD over the input signal frequency range. In opposite, the increase of the sampling frequency (higher than 130 MS/s) strongly degrades the total ADC THD, indicating that the MPA operation begins to be very affected. For the same conditions as used in previous case, Figures 6.19 and 6.20, show the SFDR variations obtained with the three tested samples. The SFDR is also an important dynamic performance metric since it can show the distortion that is not necessarily harmonically related to the input signal frequency. For F_s = 120 MS/s, the ADC exhibits, approximately, a flat SFDR over the input signal frequency range, except for chip sample 1, which presents a significant degradation for input frequencies higher than 30 MHz. It is also observed that the increase of the sampling frequency (higher than 130 MS/s) strongly degrades the total ADC SFDR. Figure 6.21 shows the experimental results of the total power dissipation obtained from the three test samples and as function of the conversion rate. As expected, the total power tends to increase slightly with F_s.

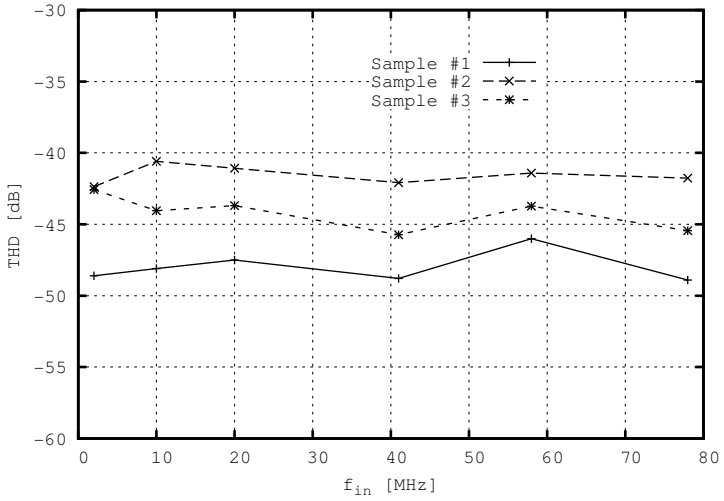


Figure 6.17: Measured THD versus f_{in} (-0.1 dBFS and F_s = 120 MS/s). Results obtained form 3 different chip samples.

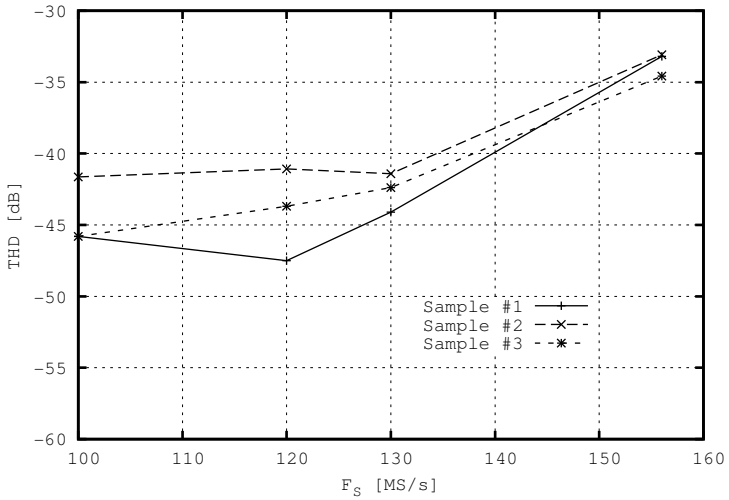


Figure 6.18: Measured THD versus F_s (-0.1 dBFS and $f_{in} = 20$ MHz). Results obtained form 3 different chip samples.

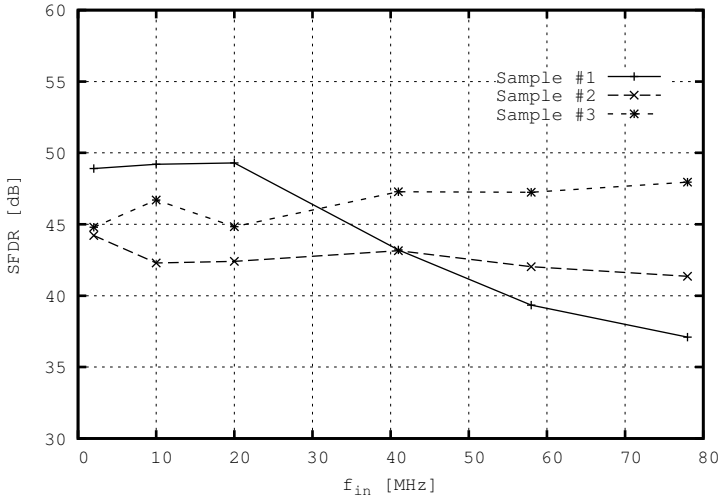


Figure 6.19: Measured SFDR versus f_{in} (-0.1 dBFS and F_s = 120 MS/s). Results obtained form 3 different chip samples.

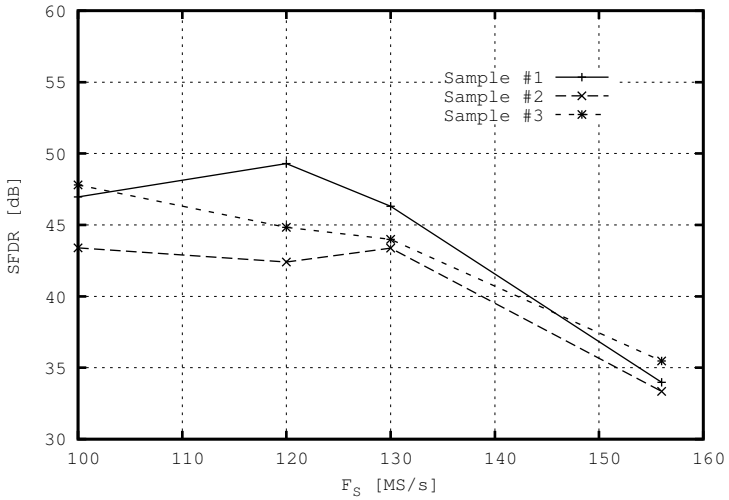


Figure 6.20: Measured SFDR versus F_s (-0.1 dBFS and f_{in} = 20 MHz). Results obtained form 3 different chip samples.

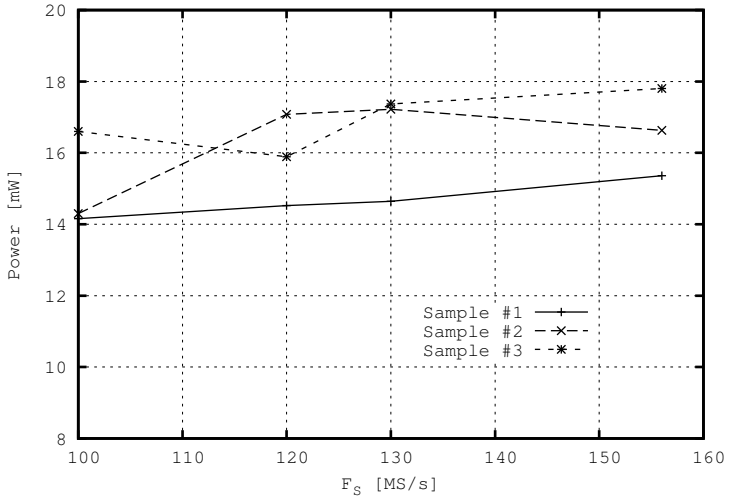


Figure 6.21: Measured Power consumption versus F_s (-0.1 dBFS and f_{in} = 20 MHz). Results obtained form 3 different chip samples.

Based on the above results the calculated FOM_2 is better than 191 [fJ.mm^2 *per* conversion] was experimentally achieved with this ADC. Key features and measurement results are summarized in Table 6.3.

6.5 Summary

In this Chapter, a practical implementation of a time-interleaved ADC fully based on discrete-time parametric amplification was described.

The ADC was integrated in a 130 nm CMOS technology only using MOS devices. The parametric amplification included in the design of the chip has proven that can reach moderate resolution performance without any calibration scheme. The full ADC exhibits more than 6 bit of ENOB and a good power and area efficiency.

Parameter	**Data and Measured Results**
Technology	CMOS 130 nm 1P8M
Die Area	0.12 mm^2
Resolution	8-bit
Conversion Rate	120 MS/s
Supply Voltage	1.2 V
Power Dissipation	14 mW - 17 mW
Analog Input	
Differential Input full-scale range	200 mVpp
Differential Internal signal swing	400 mVpp
Converter Key Performance Indicators ($F_s = 120$ MS/s , $f_{in} = 20$ MHz)	
DNL	-0.8 / +1.4 [LSB @ 8-bit level]
INL	-2.0 / +2.0 [LSB @ 8-bit level]
SNR	39.7 dB
THD	-47.4 dB
SNDR	39.1 dB
SFDR	49.3 dB
ENOB	6.2 bits
FOM	191 fJ*mm^2

Table 6.3: ADC key features and measured Results.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

Reactance based signal amplification and conversion, as well as their system integration targeting modern digital transceivers, have been the main subject of the exposed material in this work.

The presented system design space, carried out in Chapter 2, has two essential directions. One reflects the requirement of modern wireless systems that branches in two subareas: multi-standard, high data rate applications, and low data rate but ultra low power and cost solutions.

The other line of research pursued reflects the challenges imposed by nanoscale CMOS technologies, which difficults the integration of traditional analog circuitry, as opamps. However, the nanoscale era provides unprecedented digital processing capability. Consequently it is presented that an emerging response is the use a digital analog assisted circuit configurations to smooth the performance requirements of analog building blocks. Another indicated approach uses analog signal processing but in discrete-time domain. Those digital assisted techniques relax the requirements of the amplifiers which can be simplified towards open-loop configurations. It is concluded that the latter can be achieved by using parametric amplifier circuitry.

The reactance based amplifier relies on the ability of a nonlinear capacitor to transfer energy between circuit tanks when conveniently driven by both an input and a pump signals. Furthermore, it was highlighted in Chapter 3 that this amplification is intrinsically noiseless, since it is a reactance and not a resistance that is involved in the process. Chapter 3 ends by showing that, the continuous time parametric amplification

is possible to implement in modern CMOS technologies. It included a description of a mixer with parametric based amplification gain.

High quality bandpass filters are needed in the former amplifier for proper operation. For lower frequency, these LC type filters tends to occupy large areas thus pushing the implementation of the amplifier into deep GHz band. The alternative presented in Chapter 4 overcomes the need to use those filters, by using the MOS varactor in a discrete-time mode of operation. The entire Chapter was dedicated to the analysis of the MPA cell and its applications. It has been shown that, the intrinsic maximum achievable gain obtained by the structure is dependent on the technology parameters (as the oxide thickness, among others) and not directly on the area of the device. It was also demonstrated that, the extrinsic parasitic capacitances play a major role on the cell performance degradation. Among all, it has been shown that, the effect of the overlap capacitance can be reduced by choosing higher rather than minimum length devices. However, the time response analysis revealed that this length should be not too high since the time needed to remove the channel, during amplification, is proportional to L^2.

A MPA structure was presented, which was obtained from the original cell by dividing it in two half-size devices, which are connected by their drain but leaving this node floating. This enables a reduction of the total effective overlap seen between the gate and D/S node. Additionally, the structure presents a better model for simulation since when connecting the drain to source in a single MOS device, the BSIM3v3 does not consider the existence of the channel resistance. Distortion and noise analysis presented in this Chapter, revealed that moderate resolution level can be achieved without calibration. The remaining of the Chapter was dedicated to the application of the MPA cell. It was demonstrated the feasibility of the design of a MBTA, a comparator and a merged DT Mixer/IIR filter with embedded parametric amplification.

Chapter 5 described the design of an 8-bit high-speed, 14 mW pipeline ADC fully based on parametric amplification. In this approach, it was proved the feasibility to build a complete MOS-only ADC. An architecture of two time-interleaved pipeline ADC has been presented. Each individual pipeline branch comprises a front-end S/H followed by 6 pipelined stages with minimum resolution (1.5-bit) and by a 2-bit FQ. The details concerning the design of each building block were given and electrical simulations have confirmed the operation the overall A/D conversion system.

The data obtained from experimental evaluation of the previous designed ADC was presented in the Chapter 6. Layout considerations as well as details about the design of the measurement setup were also described. Measured results obtained with three

different chip samples have proven the feasibility of an high speed pipeline ADC only using MOS-only approach and totally supported by parametric amplification.

7.2 Future Work

Among other solutions, the application of the MOS parametric amplification principle constitutes an interesting approach to overcome some of the problems which arise when implementing traditional analog circuits in CMOS digital nanoscale technologies. Most promising is the complementary use of both continuous-time and discrete-time domain parametric based signal processing. Despite the parametric approach had been more popular in the past and today is mostly used in the optical domain, its utilization in MOS technology is still an emerging technique that will need further work.

A direct extension of the presented work is to implement the ADC in lower technology nodes like 65 nm, 40 nm or 32 nm. This needs to be preceded by a complementary study around the performance of the MPA cell under this deep nanoscale environment. The MPA cell operation model has, therefore, to include Short channel and Quantum Mechanical effects [23], like gate direct tunneling, energy Quantization, threshold voltage shift or gate capacitance degradation. The latter can be modeled by incorporating the concept of an effective oxide thickness $t_{ox,eff}$, [23]. Also the timing and noise model have to be adapted accordingly.

At MPA cell level configuration additional improvements can be added to facilitate a gain digital control (for digital analog assisted design type) either by adjusting the load capacitance [101] or by employing a cascade of MPA cells.

Concerning the parametric based pipeline ADC, additional stages can be used to extend the digital correction with the objective of reaching higher resolutions. Due to the modular nature of the pipeline operation, a power optimization tuning can also be applied by downscaling the size of the last pipeline stages. Parametric based low resolutions FQ can be used in the design of the Multi-bit Sigma-Delta ADC, using nonlinear DAC in the feedback loop as a way to improve the dynamic range.

At application level, the use of MOS parametric amplification can be extended over the complete receiver by using an optimal building blocks partition. In fact by using a heterodyne architecture, a first RF stage formed by a MOS continuous time parametric downconverter mixer can be used to bring the desired modulated carrier to a lower IF frequency. Not only the signal is downconverted but at the same time, it is amplified with low noise injection (due to the nature of the parametric operation).

This IF frequency is such that a DT parametric based stage can already process the signal and feeding the ADC, which by its turn also use discrete-time MOS parametric amplifier. With this approach the number of integrated inductors and capacitors can be substantially reduced, since the most the receiver is built with MOS devices, with clear benefits in area and, consequently, costs.

Appendix A

One approach for RF front-end receiver design

A.1 Introduction

While the cost per minimum size transistor decreases with CMOS scaling, the price per wafer unit area reaches a higher value when compared with previous technology node, at least during the initial time frame of a given generation. Among all the available passive on-chip components, the inductor does not scale well with CMOS evolution since for a given inductance value it remains approximately with the same area. This corresponds to a relative increase of the cost associated with the analog RF front-en. Therefore, research efforts have been made in order reduce to minimum the number of inductors inside the chip, by means of improved circuit techniques. Moreover, the necessity to reach wideband transceivers also makes the narrowband LC thanks less attractive.

Another issue that arises from the lowering of the power supply voltage is the difficulties to design high-gain operational amplifier without degrading the bandwidth and/or power consumption. Instead of maintaining the high-gain objective for this block, one can use alternative designs that explore simpler amplifier circuits based, for example, in open-loop configurations, as seen in Chapter 2. Remarkably, the error due to lower gain can be compensated by means of clever system architecture and self-calibration techniques (both analog or digital).

A.2 Low Noise Amplifier

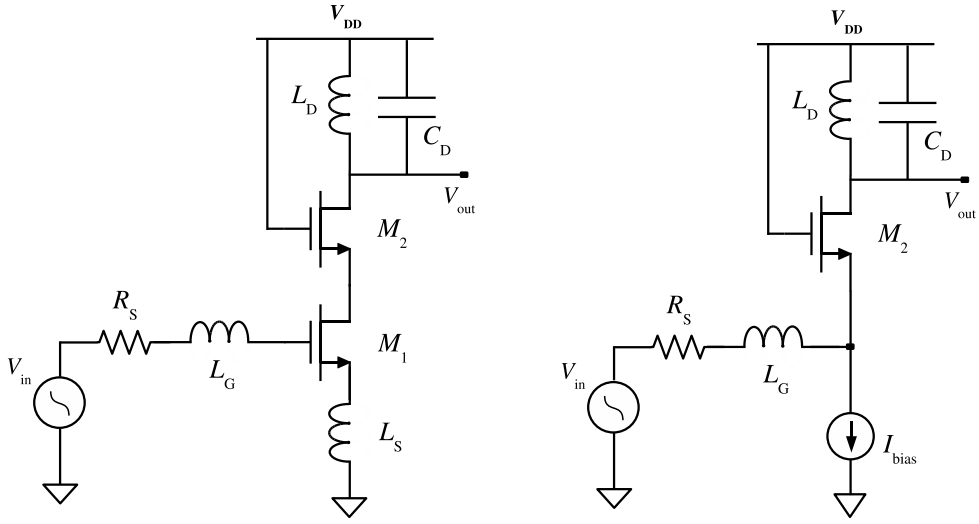
Signal amplification is a fundamental operation which is split over the receiver chain, ranging from RF to baseband domains. The distribution of the gain along this signal path depends on the type of the wireless system application, but the traditional approach reinforce the received weak signal just after the antenna stage and before demodulating it to baseband. In this case, the accuracy of the absolute amplification value is not the main issue but rather the noise and distortion contributions, which degrades the overall receiver SNDR. This first amplifier stage is usually a Low Noise Amplifier (LNA).

In a classical narrowband communication scenario, the LNA is usually designed under a RF tunned amplifier approach, in which the noise figure (NF) is optimized. This tunned approach is usually supported in high Q matching networks (traditionally LC tanks) sized to cope NF requirements and/or input and output impedance matching [30]. Since those narrowband filtering networks rely on high Q inductors, their integration in a pure digital CMOS technology is limited.

By making available the setting of each individual device geometry, the integrated MOS transistor has introduced new degrees of freedom to size an LNA. In fact, the design process has evolved from a classical approach in which the noise contribution from the active device is minimized through the optimization of a converted source impedance, Z_s, to a more flexible one in which the width and the length of the MOS transistor are both used to optimize noise, gain and linearity performance. Furthermore, with the integration of the LNA and following RF front-end blocks inside the same chip, the 50 Ω matching requirements between them are removed thus reducing power consumption.

Being the first active amplifying block after the receiving antenna, the LNA design space is delimited by the need to reduce noise injection during amplification while achieving a high frequency response, under narrowband conditions. To achieve these major requirements, the LNA is usually built around single stage topologies, as depicted in Figure A.1. The common-source (CS) with a cascode device connected to the drain and an inductive source degeneration, A.1(a) has been extensively used and analyzed in the past, [131, 30]. In comparison with the common gate (CG) topology, represented in Figure A.1(b), the common source has a superior performance in terms of minimum Noise Figure (NF) and voltage gain (A_v), in the GHz range,[14]. In fact, from [132], the common gate voltage gain is given approximately by

$$A_{v,CG} = G_{m,eff,CG} \cdot Z_L \approx \frac{1}{2R_S} \cdot Z_L \tag{A.1}$$



(a) Common Source Configuration (b) Common Gate Configuration

Figure A.1: Narrowband Low Noise Amplifiers.

whereas the gain of the common source is determined by, [131],

$$A_{v,CS} = -G_{m,eff,CS} \cdot Z_L = g_{m1} Q_{in} \cdot Z_L = -\frac{g_{m1}}{\omega_0 C_{gs} \left(R_S + \omega_T L_S\right)} \cdot Z_L \approx -\frac{\omega_T}{\omega_0} \cdot \frac{1}{2R_S} \cdot Z_L \tag{A.2}$$

at ω_0 resonance frequency and where R_S counts for the input signal source impedance. ω_T represents the transition frequency. Both equations are valid within narrowband configuration, where it is assumed that the input network are either at resonance state and under impedance matching conditions. These results shows that the CS stage tends to reach higher voltage gain values due to the increase of the transition frequency associated with the CMOS scaling. However, for a given technology node, the common gate configuration maintains the gain approximately constant up to higher frequencies making this stage an important option for microwave CMOS LNA, [132].

The increase of f_T in deep submicron MOS transistor contributes to lower the NF of the CS type LNA while the impact in CG is less visible, as can be checked respectively in A.4 and (A.3). Moreover, it has been demonstrated, [30], that the CG has minimum of 2.2 dB, when the input is under matching conditions, while the CS configuration can reach lower values.

$$\mathrm{NF_{CS}} \approx 1 + \frac{R_g}{R_S} + \gamma g_{d0} R_S \left(\frac{\omega_0}{\omega_T}\right)^2 \tag{A.3}$$

$$\mathrm{NF_{CG}} \approx 1 + \frac{\gamma}{\alpha} \frac{1}{g_{m2} R_S} \tag{A.4}$$

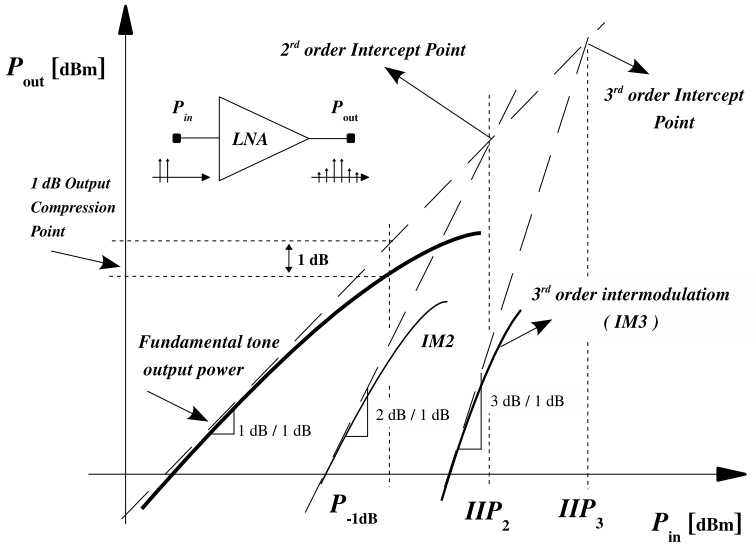


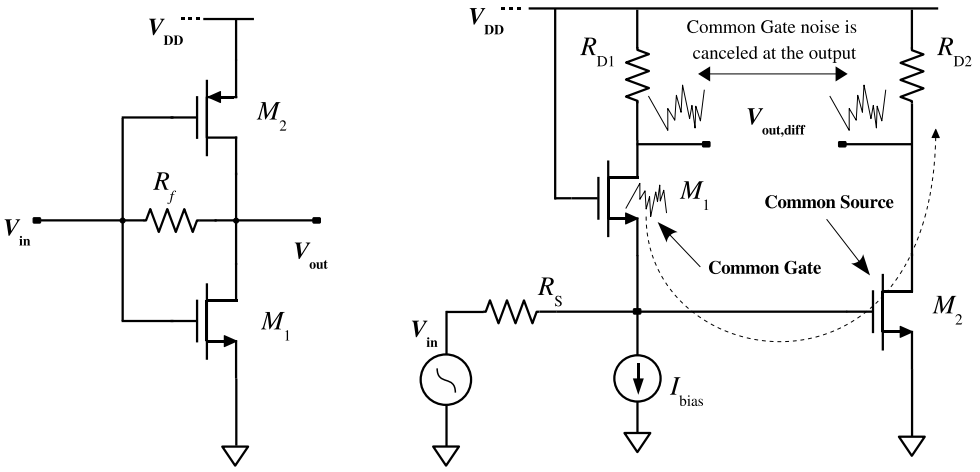
Figure A.2: Definition of the 1 dB compression point, IIP_2 and IIP_3.

where γ is the channel thermal noise coefficient and α is equal to g_{m1}/g_{d0}. The g_{d0} parameter is the channel conductance when V_{ds} is zero and g_m is the device transconductance. R_S and R_g counts for the input signal source impedance and input LNA resistance connected to the gate, respectively.

To improve the output to input isolation around the LNA transconductance transistor, by reducing the effective coupling effect between the drain to gate due to the overlap parasitic capacitance, a common gate device can be cascoded on the top of the common source stage, as depicted in Figure A.1.

Another important aspect is the total distortion of the overall receiver which is much dependent on the linearity of the LNA. As for the complete receiver, the linearity characterization of the LNA is also described by a 1 dB compression point, second and third order interception points, IIP_2 and IIP_3. All of them are illustrated in Figure A.2. The objective of these metrics is to define the level of the input power signal above which the nonlinearities effects have to be taken into account. Besides the effects of third order intermodulation products, the second order ones have a serious negative impact when in presence of close interferers, specially in DCR receivers where they fall directly into baseband. This increases the energy around DC and potentially reducing the detectable dynamic range due to saturation effects. To reduce their effect and as a complement to an optimization of the LNA DC biasing, one can use a differential structure which is known by its ability to attenuate significantly second order nonlinearities and parasitics effects caused by substrate coupling, [133].

Following the emerging of SDR and multi-band receiver, there is a growing interest



(a) Self Biased Inverter LNA (b) LNA with a noise and distortion cancellation

Figure A.3: Wideband LNAs.

on the design of wideband LNA, covering frequency range up to 6 GHz (SDR) or 10 GHz (UWB),[134]. Instead of using multiple narrowband LNA that are concurrently activated for each band, a better option in terms of cost and area, consists of replacing the inductors by resistors in the LNA circuits. These inductorless LNA expands its frequency response to a wideband type, specially in deep-submicron CMOS. In fact, removing the LC tanks facilitates the full integration of the RF front-end making it more flexible and efficient in terms of area. But in the other hand, the design gets more exposed to noise and non linear effects resulting in higher 2^{nd} and 3^{rd} order intermodulation products, thus degrading the signal to noise plus distortion ratio. As referred before, the use of a differential structure can reduce the 2^{nd} order intermodulation effects, but at the expanse of requiring an input balun and also duplicating the current consumption. One technique that can be adopted to lower the power consumption is to reuse the current by stacking a PMOS on top of an NMOS device [135], in a self-biased inverter type configuration as shown in Figure A.3. The overall transconductance G_m increases to $g_{mn} + g_{mp}$ requiring only on half of the current needed by the previous topology.

Another interesting approach presented in [38], proposes a noise and distortion canceling technique that simultaneously implements a single to differential conversion, which is of major importance since the antenna output is single-ended. By integrating the single-ended to differential conversion function, the internal or external inductor based balun is no longer needed. Moreover, if the antenna is close enough to the LNA input (to neglect standing wave effects) even the 50 Ω input match can be dropped, reducing even further the number of inductors, [38].

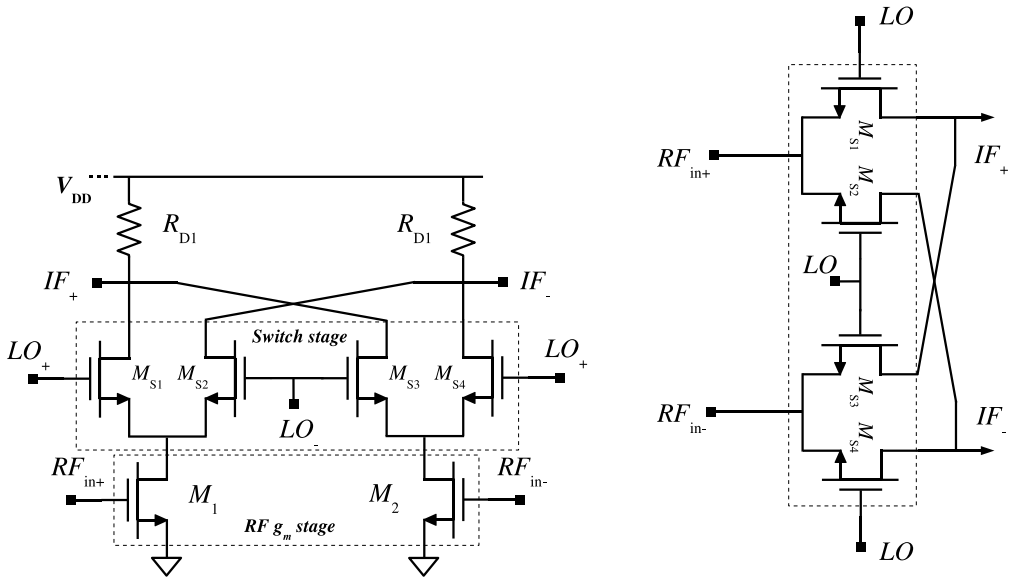
The schematic of the circuit represented in Figure A.3 is a balun based LNA consisting of a CG stage simultaneously connected to a CS stage. The CG stage implements a wideband input impedance matching and gain, while the CS stage generates an inverted output signal. This inversion permits the cancellation of the CG noise at the differential output. Besides this operation, this circuit can simultaneously implement noise and distortion ($IM2$) canceling, and output balancing as analyzed in [38, 134].

A.3 Frequency translation

After the amplification done by the LNA, the signal is downconverted to IF or directly to baseband by means of a frequency translation performed by a circuit mixer. This task can be accomplished with either nonlinear devices, like diodes and transistors, or by time-varying elements, like switches or varactors (this subject is analyzed in Chapter 3). Several mixer topologies have been proposed in the past, but one can roughly group them in active or passive mixers. More recently, also a emerging type of discrete-time sampling mixer has also gained some interest. The required type of mixer is much dependent on the target system requirements.

In comparison to the active topologies, the passive mixer do not provide conversion gain (e.g., they are lossy) but tends to achieve better linearity. As usual, single-ended and differential structures are available. The latter structures are the preferred when it is important to reduce second order non-linearity effects. As expected, gain, noise and linearity are fundamental specifications for mixers, which differs from the LNA by the fact that they are calculated between signals with different frequencies. Figure A.4 shows the common Gilbert cell based active mixer and also the corresponding passive version. A closer look on both schematics reveal that they differ in the way the switching quad operates. In the former case, the mixer commutates signal current (from the transconductance RF stage) whereas in the passive configuration it commutates directly the RF signal voltage. In both cases, the local oscillator drives the switching stage with sufficiently high amplitude while in the active version, a biasing currents flows through the transconductance branches.

The conversion gain of the commutating stage is obtained by averaging the gain over one period of the LO signal, [136, 137]. For a 50 % duty cycle ideal square-wave LO signal, this conversion gain approches $2/\pi$, thus corresponding to a loss. However, in the active mixer, this loss can be compensated by the transconductance gain, resulting



(a) Active (b) Passive

Figure A.4: Active and passive mixer.

in a total conversion voltage given by

$$A_{v,ActiveMixer} = \frac{2}{\pi} \cdot g_m R_{D1}. \tag{A.5}$$

In the passive mixer the transconductance stage is not present and therefore the total conversion gain is equal to $A_{v,PassiveMixer} = 20.\log(2/\pi) = -3.9$ dB, which is effectively a loss. However, since there is no DC current, it usually consumes less power than the active counterpart, [138]. Additionally, due to the unexistence of DC current, the passive mixer is characterized by a substantially lower level of flicker noise, [117], which is very important for low-IF and direct conversion receiver. Nevertheless, a mixer have usually a poor noise performance, since the transfer of white noise from multiple frequency bands to the output have to be considered in the signal-to-noise ratio calculation.

Another class of mixers are based on track and hold (T/H) circuits (see appendix B). A T/H circuit performs downconversion and sampling functions thus generating a low frequency discrete-time output signal. The previous referred mixer topologies do not present this discrete time characteristic. The sampling mixers can be divided into sub-sampling type, direct sampling type or over-sampling one. The sub-sampling mixer presented in [139] achieved a NF of 21 dB with a -6 dBm IIP_3, for a 2.4 GHz RF band and a 100 MHz sampling frequency. For the family of direct sampling mixer, two strategies can be followed: voltage sampling and charge sampling. In [140], a voltage sampling mixer achieved a NF of 25 dB with 22 dBm IIP_3 for a 1.6 GHz RF carrier

and a 1.55 GHz sampling frequency. Regarding the preceding case, in the topology proposed in [123], the voltage signal is firstly converted to current/charge and only after the T/H operation is performed.

A recent approach presented [42], propose to perform the mixing operation in discrete-time domain. To support this operation, the architecture need that the input signal be oversampled by a T/H running at 8 times faster than the carrier frequency.

In Chapter 3, an alternative way of achieving mixing operation is presented, which is based on variable reactance instead of conductance based approaches.

A.4 Co-design and merging strategies

The implementation of a direct conversion transceiver in a digital CMOS technology, must by supported on a optimized design strategy to fully integrate the RF front-end.

One of the approaches is to co-design together the LNA, Mixer and Local Oscillator (LO). This is better achieved by avoiding the 50 Ω matching buffers and networks and AC coupling capacitors between these three blocks. Also to reduce area the number of inductors are reduced to minimum, and wherever possible replaced by simple resistances. An example of one co-design method can be found in [141, 142, 143]. The design process begins by maximizing the LNA voltage gain for a given input match criteria (50 Ω in this case). Then, the LO is designed to maximize the signal output swing voltage and improve the In-phase (I) and Quadrature-phase (Q) signals accuracy for a given power. With the obtained DC components at the output of these blocks, the mixer is then optimized to reach a reasonable conversion gain and NF. In order to reduce the total area, the design minimizes, as much as possible, the use of inductors and AC coupling capacitors.

The number of integrated inductors is further reduced by using an RC based oscillator. The oscillator circuit shown in Figure A.5 comprises two RC integrators [39]. Each one is realized by a differential pair (transistors M_1) and a capacitor (C_1). As expected, current I_{tune} and the capacitor sets the oscillator frequency. There is an additional differential pair (transistors M_L), with the output cross-coupled to the inputs, which performs two related functions:

- compensation of the losses due to R to make the oscillation possible (a negative resistance is created in parallel with C_1);
- amplitude stabilization, due to the non-linearity (the current source I_{level} controls

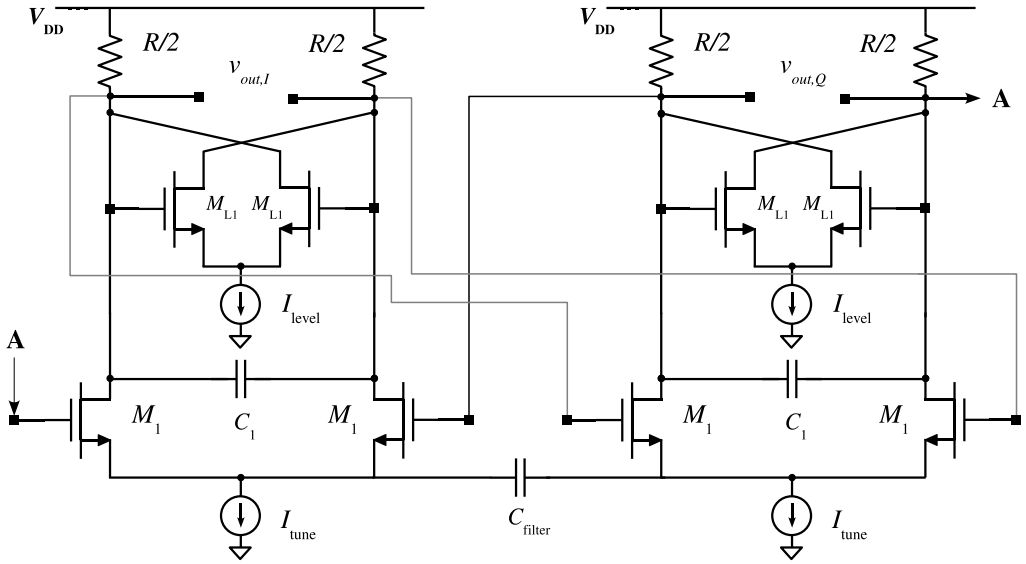


Figure A.5: RC Two-Integrator Oscillator circuit with quadrature outputs, based on [39].

the amplitude).

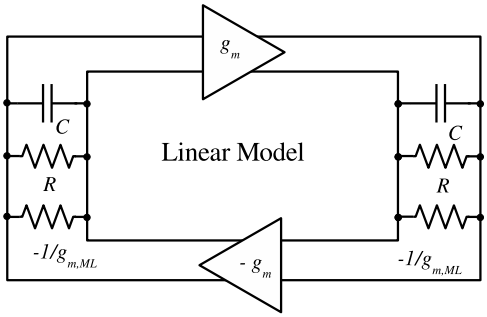


Figure A.6: Linear model for the RC Two-Integrator Oscillator, based on [39].

The oscillator frequency in a quasi-linear conditions is obtained from the linear model depicted in Figure A.6, where the negative resistance is realized by the cross-coupled differential pair (M_L), and R represents the integrator losses due to the pairs of resistances $R/2$. From [39] and considering that the negative resistance compensates the losses of the integrators, the oscillation frequency is approximately given by

$$-\frac{1}{g_{m,ML}} = R \qquad \Rightarrow \qquad \omega_0 = \frac{g_m}{C} \tag{A.6}$$

and the output amplitude can be approximated as

$$v_{out} \approx I_{level} R \ . \tag{A.7}$$

If a low distortion output is required for the quadrature output oscillator, a filtering technique can be set just by adding an extra capacitor C_{filter} between the two tuning current sources. The introduction of this element reduce both the harmonics at this point and the oscillator phase-noise. Finally, to start the oscillations the condition $g_m > 1/R$ must be met.

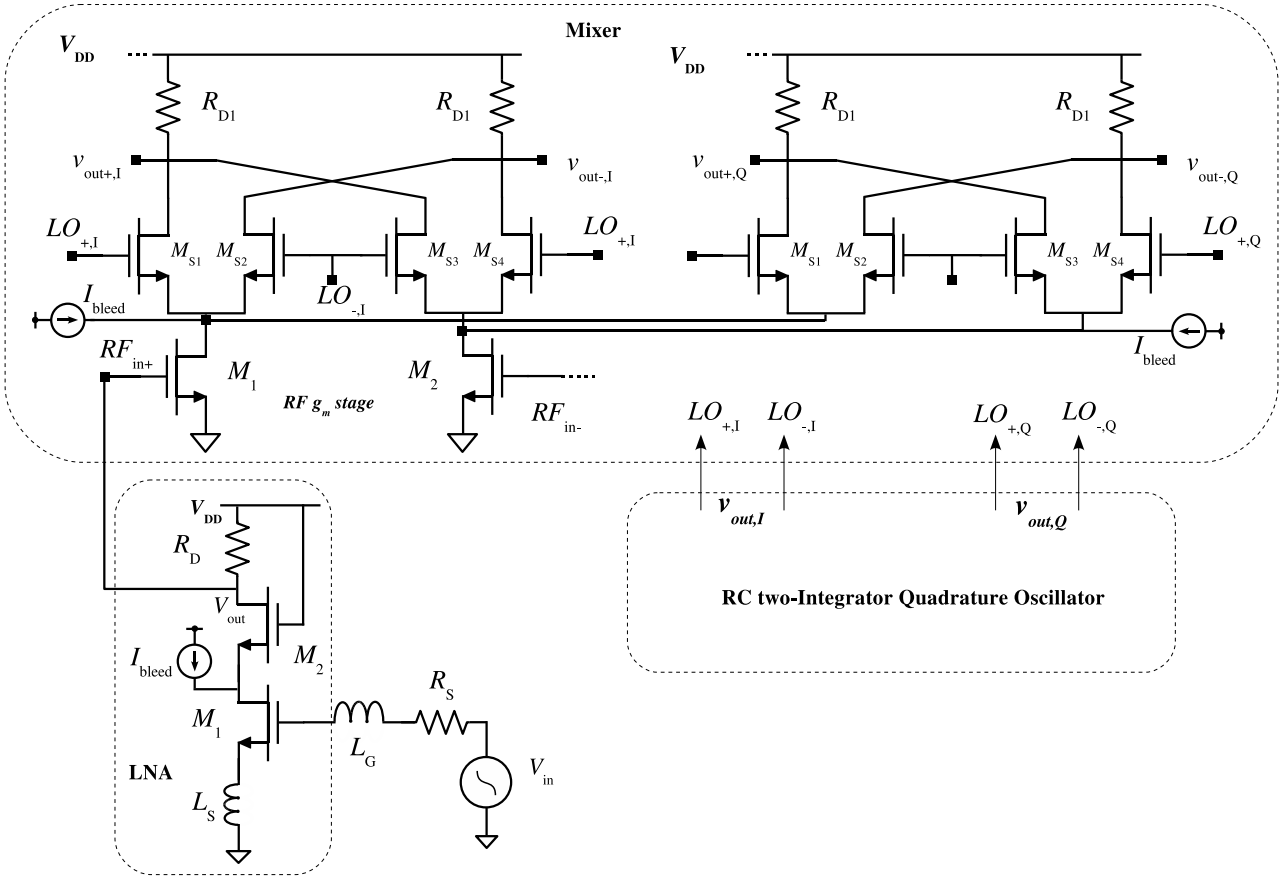
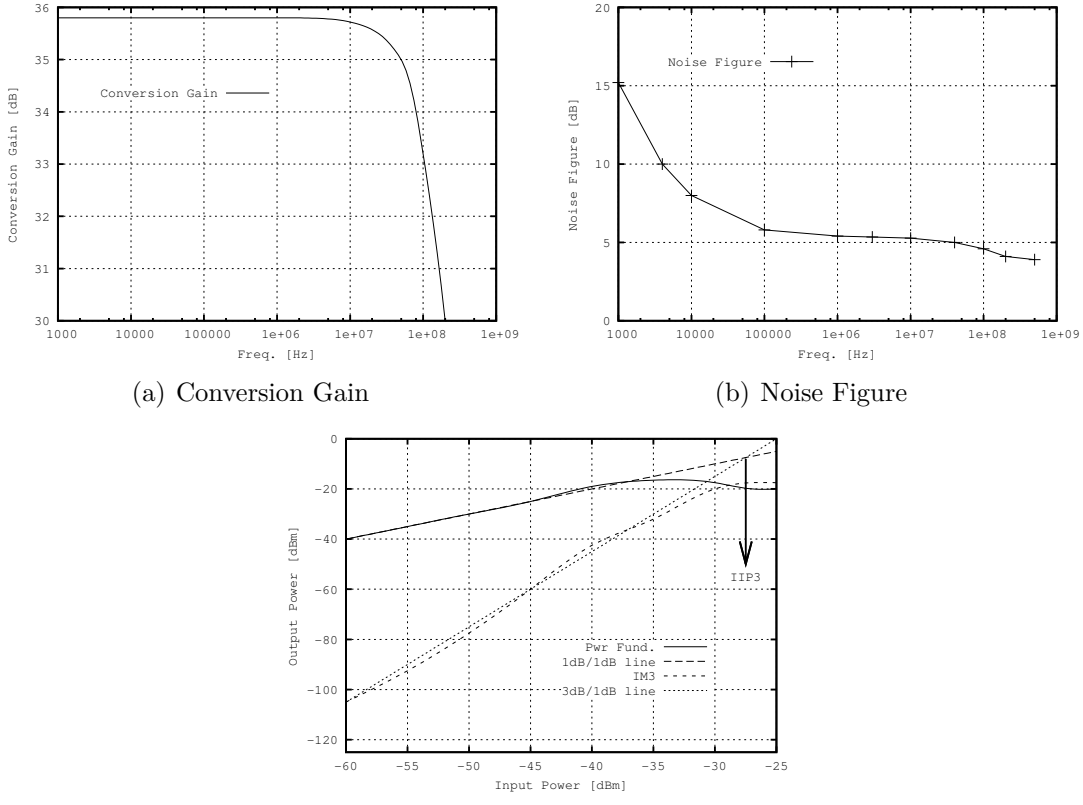


Figure A.7: LNA, Mixer and LO circuit diagram.

The low-IF RF front-end illustrated in Figure A.7 has a reduced number of integrated inductors, and coupling capacitors and interstage matching networks are not used. In this case, the LNA output DC component is used to bias the mixer transconductance stage, thus controlling the mixer conversion gain. For the transconductance part, minimum lengths transistors are used to maximize speed and gain, and the width is adjusted according to the DC voltage at the LNA output node. An additional current is injected into the mixer transconductance to improve linearity and to set the conversion gain and NF. By controlling this current, the DC output level from the LNA will have less impact on the mixer output voltage. The oscillator amplitude needs to be maximized to properly drive the mixer switching transistors and reduce the $1/f$ output noise.

The front-end has been designed in a 130 nm CMOS technology and simulated with SpectreRF by means of a Periodic Steady-State (PSS) analysis, and using BSIM3v3 models. For a target frequency of 900 MHz and a 50 Ω antenna input matching, the transconductance of the LNA input transistor is set to 60 mS. To optimize the quadrature output voltages of the LO while maintaining a reasonable value of the phase noise, the integrator load resistor is set to 314 Ω. The remaining transistors sizes are determined to be $W/L = 15$ μm /0.25 μm and $W/L = 10.8$ μm /0.25 μm, respectively, for the transconductance stage and for the cross-coupled pair. Accounting these results, the RF transconductance stage of the mixer is then determined with transistors of size equal to $W/L = 30$ μm /0.13 μm, considering a bleeding current of 4 mA. The simulated results obtained for the RF front-end conversion gain, NF and IIP_3 are shown in Figure A.8 and summarized in Table A.1. In order to improve



(a) Conversion Gain

(b) Noise Figure

(c) IIP_3

Figure A.8: SpectreRF PSS simulation results for the co-design of LNA, mixer and LO.

linearity and better adjust to receiving signal strength, the LNA gain can be reduced by reducing the bleeding current and adjusting accordingly the load resistance to meet the same DC output (thus, keeping the mixer and LO unchanged). To evaluate the impact of the LNA reduced gain, the LNA load resistance is reduced from 700 Ω to 72

Parameter @ 10 MHz	LNA high gain	LNA Low gain
Voltage Conversion Gain	35.7 dB	25.5 dB
NF	5.28 dB	6.16 dB
P_{1dB}	-37.16 dBm	-28.95 dBm
IIP_3	-27.16 dBm	-17.33 dBm

Table A.1: Simulation results for the co-design.

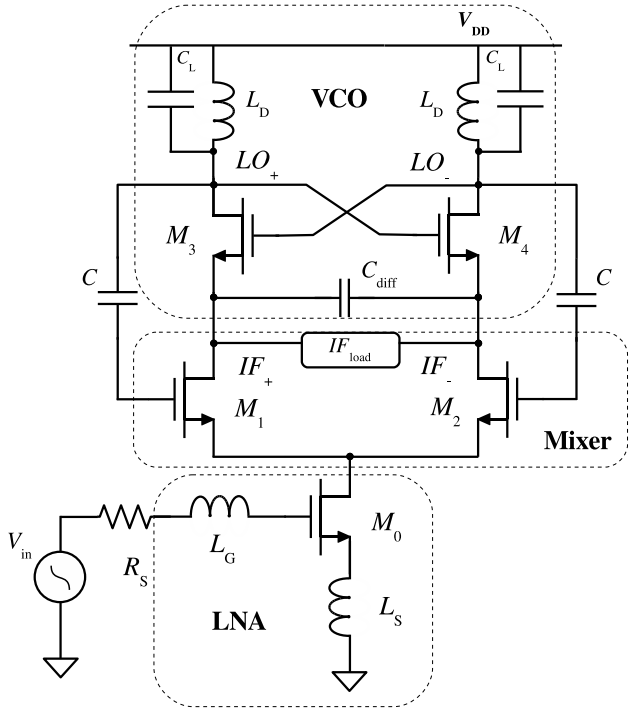


Figure A.9: An LNA, mixer and Oscillator merged cell, LMV, [144].

Ω and the bleeding current is also removed. The results from the PSS simulations are shown in Table A.1.

A slightly different approach is achieved by merging all of these three block into a single one. A good example of this approach is the LNA-Mixer-VCO (LMV) presented in [144], which is formed by stacking an LNA, a mixer and a VCO over the same bias current. LC-tank VCOs and mixers share a common principle which is represented by a pair of switching devices. The former does not offer high impedance at low frequencies and, therefore, an additional stage has to be used to provide isolation between the LO output and the down-converted signal. Analyzing the LMV cell proposed in [144] and reproduced in Figure A.9, one may conclude that :

- M_0 is transconductance stage to implement a narrowband LNA at RF, but also supply the bias current to the VCO;
- M_3 and M_4, provide the negative resistance to induce oscillations;

- M_1 and M_2 operates as switches to implement the mixing function. The mixer part and the output of the VCO are connected through capacitor C;
- although three transistor are stacked, the minimum supply voltage required is equal to only one threshold plus three overdrive voltages and, therefore, maintaining the compatibility with low power supply voltages, [144].

The results presented in [144], reveals that the implementation of the LMV in a 130 nm CMOS technology targeting a 2.2 GHz application, reached a conversion gain of 36 dB, a 4.8 dB NF and an IIP_3 of -19 dBm, which a measured power consumption of 11 mW. One of the major drawbacks is the use of integrated inductors.

Appendix B

Sampling and Switches

The signal amplitude quantization and digitalization is usually and closely associated with discretization in time domain, [94]. Therefore, a signal sampling process has to occur before or at the input of the ADC. Usually this sampling is periodically performed at a known rate (sampling rate F_s), thus implying the need for clock signals. A simple track and hold circuit is illustrated in Figure B.1 which is formed by one sampling switch and one storage capacitor. During the sampling phase the input signal is simply tracked since the switch is closed and therefore the output signal is ideally equal to the input one. At the end of the sampling phase the switch opens and the signal value at this instant is stored in the capacitor and maintained during the hold phase. The ADC will process this stored sampled and not the changing signal during tacking phase as it could generate aperture errors, [94]. Therefore, from the ADC perspective it is the sample and hold (S/H) operation that matters.

It is well known that the sampling process changes the signal spectrum by adding repeated signal replicas at multiples of the sampling frequency F_s and one must follow to the Nyquist criteria to be able to preserve original signal information, [94]. Additionally, the frequency response of S/H follows a *Sinc* type envelope with zeros at integer multiples of F_s and low pass characteristic.

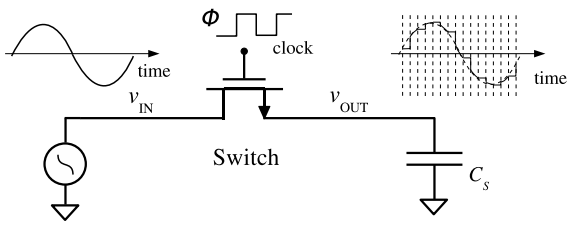


Figure B.1: Basic Track and Hold circuit.

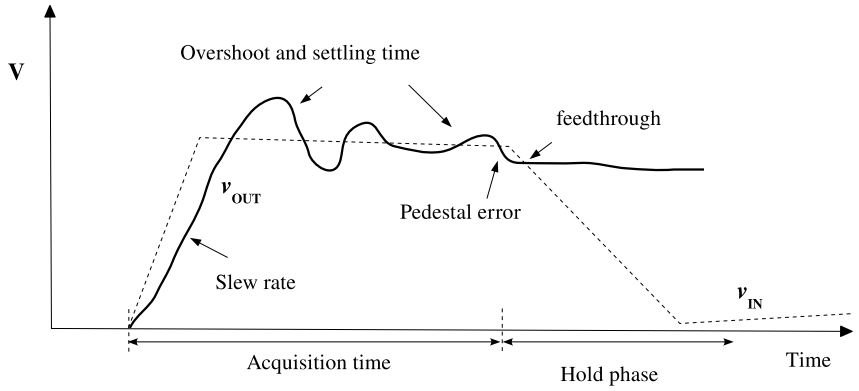


Figure B.2: S/H errors.

S/H errors	Origin	Impact
Pedestal error	Clock feedthrough, Charge injection	Higher THD
Leakage	Gate leakage	Higher THD, Limits minimum f_s
Aperture	Clock jitter	Limits dynamic range and frequency input
Acquisition time	Finite RC bandwidth and finite slew rate	Limits tracking accuracy, limits f_s
Nonlinear and Input-dependent sampling	Input-dependent v_{GS}	Lower THD

Table B.1: S/H circuit performance issues.

In general terms the front-end S/H stage used in RF sampling receivers or at the input of the ADC affects the achieved linearity. This is the result of two main error contributions coming from input-dependent charge injection and tracking nonlinearity due to the ON-resistance modulation of the sampling switches. As consequence the digitized sample signal differs from the ideal one reflecting errors during both tracking and the hold mode, as shown in Figure B.2 and summarized in Table B.1, [94, 116].

Most of the nonidealities of the S/H increase the total harmonic distortion. It depends on the amplitude of the input signal, input frequency, rise and fall times of the sampling clock and the amplitude of the clock signal. In case of sampling at RF, since the signal amplitude is usually small, a simple switch properly sized is usually enough. On the other extreme, the use of S/H at the input of high resolution ADC implies that linearization recipes (e.g., bottom-plate sampling, bootstrapping) have to be applied in the design of the most critical signal switches, [24]. In these cases lowering the switch resistance by just increasing the switch size does not improve automatically the linearity performance. In fact, the increase of the transistor size also increase the nonlinear parasitic capacitances, resulting in higher charge injection. Charge injection creates offset and nonlinear errors which are dependent of the input signal, at some extent.

In MOS technology, the most basic switches shown in Figure B.3 are implemented using a NMOS, a PMOS transistor or alternatively a compound of both, which is designated by a CMOS switch. With respect to the first basic configuration, the CMOS structure

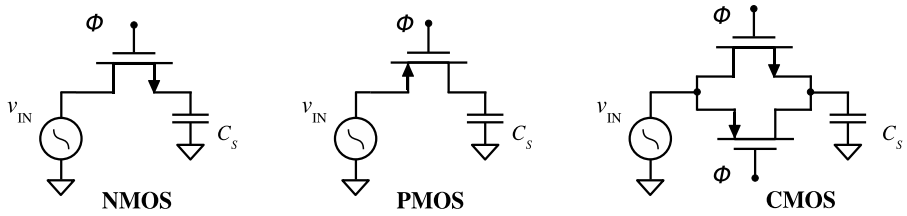
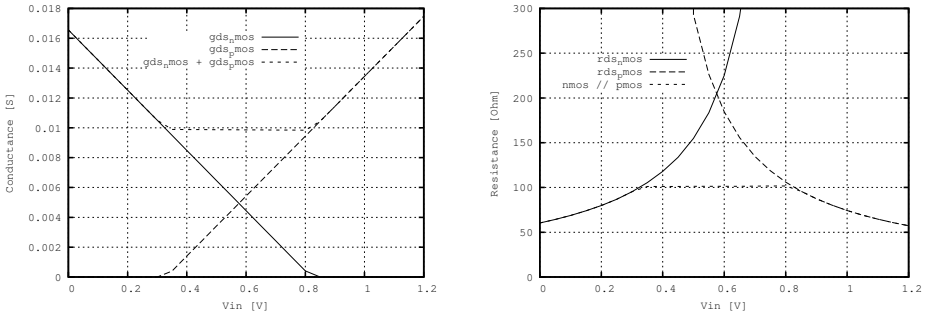


Figure B.3: Basic MOS switches.



(a) Total conductance

(b) Total Resistance

Figure B.4: Total conductance and resistance of the CMOS switch.

(ATG) supports a higher input voltage swing. Additionally, the CMOS switch, if correctly sized, can present a ON-resistance approximately flat within certain range, improving the linearity level. In fact, assuming that the transistors of the CMOS switch always settles towards the linear region, when in conducting state, one can obtain the total conductance which is then given by B.1.

$$g_{ds,cmos} = \begin{cases} g_{ds.nmos} = K_n \frac{W_n}{L_n} (v_{Gn} - v_{IN} - V_{tn}) & , \quad , v_{IN} < V_{tp} \\ g_{ds,nmos} + g_{ds,pmos} = K_n \frac{W_n}{L_n} (V_{DD} - V_{tn} - V_{tp}) & , V_{tp} < v_{IN} < V_{DD} - V_{tn} \\ g_{ds,pmos} = K_p \frac{W_p}{L_p} (v_{IN} - v_{Gp} - V_{tp}) & , v_{IN} > V_{DD} - V_{tn} \end{cases} \tag{B.1}$$

To obtain the flat conductance region in B.1, one must size the PMOS transistor width to be $W_p = m.W_n$, where m is the relation between the transistors mobility factor, i.e., $m = K_n/K_p$, assuming the same value for their length. Figure B.4 show the total theoretical output conductance of a CMOS switch, for a 130 nm CMOS technology. It is interesting to verify that the regions only dominated by the NMOS or the PMOS device, the resistance of the CMOS is nonlinear.

As previously mentioned, the S/H nonlinear degradation is mostly coming from the switch operation. Two of the nonideal effects occurs in the transition between the tracking phase to the hold phase, and are known as charge injection and clock feedthrough. When conducting, the MOS switch forms an inverted channel beneath the gate oxide

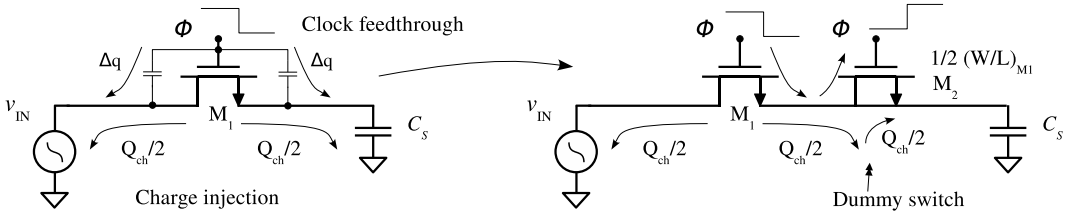


Figure B.5: Charge injection and Clock feedthrough in NMOS switch.

with a total charge of Q_{ch}. When the transistor tuns OFF (transition to hold mode), this channel charge is removed and, approximately, flows equally to the hold capacitor and to the input voltage source, v_{IN}, [29]. Because v_{IN} is assumed to be low-impedance, the injected charge has no effect on this node. However, on the hold capacitor side, this injected charge will corrupt the acquired sample, as illustrated in Figure B.5. Another charge error which is injected in the sampling capacitor is originated by the clock transition and the gate overlap parasitic capacitance. The overlap gate capacitance c_{ov} is equal by $C'_{ox}Wl_{ov}$ where l_{ov}, is the overlap length. A first order analysis results in a total voltage change at the sampling capacitor given by

$$\triangle v_{OUT} = \underbrace{-\frac{Q_{ch}(V_{DD}, v_{IN}, V_t, W, L)}{2C_S}}_{\text{Charge injection}} \underbrace{-\frac{C_{ov}}{C_{ov}+C_S}\cdot V_{DD}}_{\text{Clock feedthrough}} \tag{B.2}$$

where total channel charge is obtained from B.3.

$$Q_{ch} = C'_{ox}WL\cdot(V_{DD} - v_{IN} - V_{tn}) \tag{B.3}$$

$$V_{th} = V_{t0} + \gamma_n\left(\sqrt{|2\Phi_{Fn}| + v_{BS}} - \sqrt{2\Phi_{Fn}}\right) \tag{B.4}$$

The threshold voltage V_{th} depends on the bulk to source voltage v_{BS} which, for the NMOS switch is equal to the v_{IN}, adding an additional source of error.

The level of charge injection is strongly dependent on the input signal, and the error that appears at the output is not just a static offset, that could easily removed. In fact this input dependent charge injection is responsible for higher distortion. Deeper analysis can be found in [24].

In a P-type substrate MOS technology, the PMOS device is masked in a n-well area, which permits the individual access to the PMOS body terminal. This fact can be used to improve the linearity range of the PMOS switch by using bulk switching technique, represented in Figure B.6, [24]. When the PMOS switch is ON, the $S_{1,SB}$ switch connects the source and bulk terminals of the PMOS device thus reducing the v_{BS} to zero, thus eliminating the body effect. During the hold mode, $S_{2,SB}$ connects the bulk

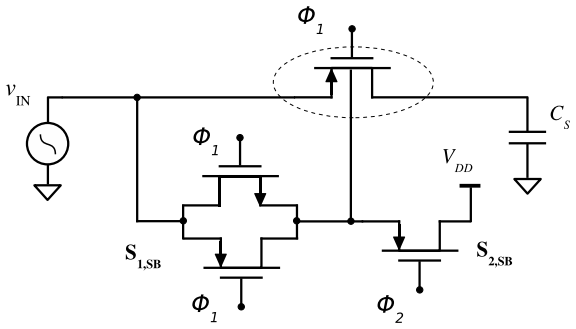


Figure B.6: Bulk switching configuration in a PMOS switch.

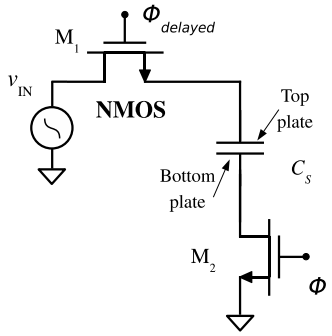


Figure B.7: Bottom plate sampling technique with an NMOS switch.

to the supply voltage to prevent positive source to bulk voltage.

An additional common method to minimize the input dependent charge injection is the bottom-plate sampling configuration, [116], represented in Figure B.7. In the circuit of Figure B.7, the M_2 transistor is turned off first to freeze the charge at the bottom plate of the sampling capacitor. Since this node is left floating, its charge can not change with the charge injection coming from M_1, when it turns off. The only charge injection is generated by M_2 which is essentially signal independent.

Despite the fact that the bulk-switching and bottom plate sampling makes the charge injection less dependent of input signal, it does not remove it. This can be achieved by inserting a dummy device. In Figure B.5, a half sized dummy transistor with shorted drain and source is added to the circuit. This dummy transistor is controlled by a phase-clock that is the complement of the signal controlling M_1. This technique relies on a opposite charge movement principle in which the charge injected by transistor M_1 is essentially matched by the charge induced by M_2, and the overall charge injection is canceled, [127].

During the tracking phase, the track and hold circuit can be analyzed as an RC circuit. However, as verified previously, the switch ON resistance is a nonlinear function of the

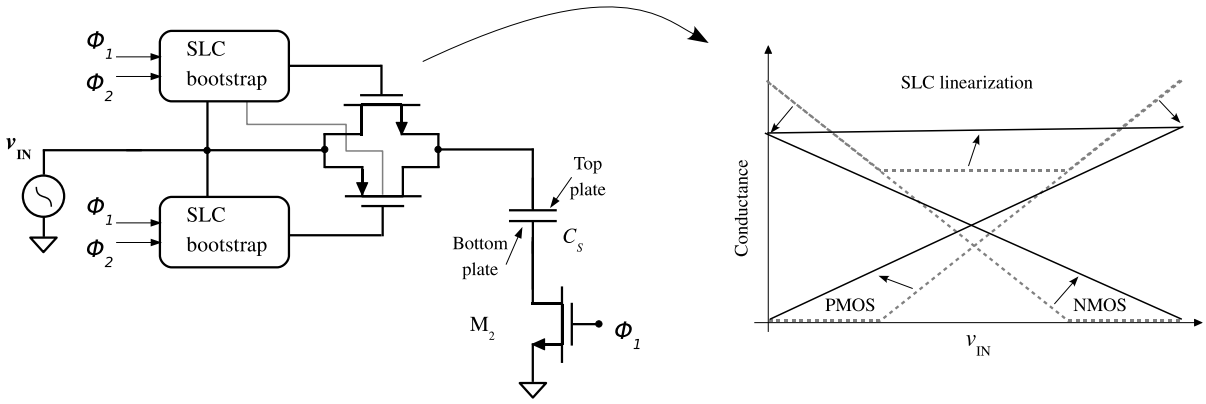


Figure B.8: SLC technique applied in CMOS switch.

the input voltage, generating signal degradation due to distortion. Instead of increasing the size of the switch, which can increase charge injection errors, one can modify the gate voltage accordingly to the input signal in order make the ON switch resistance less dependent of the input voltage. One common method is to use a bootstrap switch structure, [24], as depicted in Figure B.8. A detailed analysis can be found in [24].

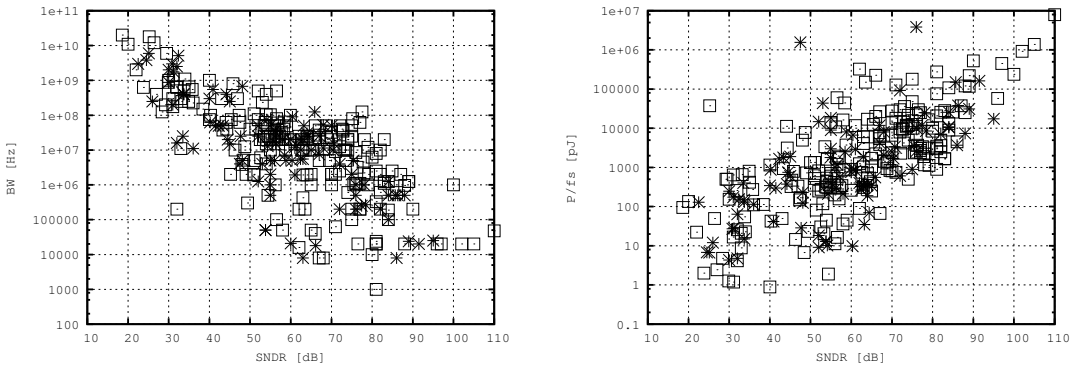
An important result for the S/H circuit is the effect of the noise coming from the transistor switch, in the sampling process, [116]. During the tracking phase, the circuit behaves like a RC network with a 3 dB bandwidth of $\omega_{3dB} = 1/R_{on}C_S$. Additionally, the spectral power density of the thermal noise associated with the ON resistance is the usual $4kTR_{on}$ [V^2/Hz] (where T is the temperature in Kelvin, k is Boltzaman's constant). The total noise that is sampled is then given by the well known

$$v_{noise,RMS} = \sqrt{\frac{kT}{C}}. \tag{B.5}$$

Intimately associated with the sampling process is the amplitude quantization and digital coding. This is a task performed by the ADC which is one the subjects treated in this work. Despite the diversity of specifications defined for this type of circuits, the most important ones are:

- Resolution;
- Frequency of operation;
- Input dynamic range;
- Power consumption.

To achieve target requirements for these specifications, several topologies of ADCs



(a) Bandwith (b) Energy

Figure B.9: ADC evolution trends in last decade, from [130].

have been developed. Each one have their one advantages and are able to achieve different performance specifications. For example, a Flash type ADC can achieve the highest operation speed but with low resolution. On the other extreme, recycling type of ADC can achieve high resolution, lower power but at a low frequency range. That is, depending on the application, the best ADC topology has to be chosen. Or alternatively, and considering the requirements for multisantdard transceivers, a reconfigurable ADC is a important target. Based on the data available from [130], Figure B.9 gives the ADC evolution trends in during last decade, with respect to energy *per* conversion and bandwidth.

Bibliography

[1] G. E. Moore, "Cramming more components onto integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp.114," *IEEE Solid-State Circuits Newsletter*, vol. 20, no. 3, pp. 33–35, Sep. 2006.

[2] S. Cherry, "Edholm's law of bandwidth," *2004. Digest of Papers. 2004 IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*, pp. 58–60, Jul. 2004.

[3] R. Metcalfe, "Metcalfe's law: A network becomes more valuable as it reaches more users," *Infoworld*, no. 17, 1995.

[4] Intel, "Transistor TimeLine," Intel Corporation, Tech. Rep., 2007. [Online]. Available: http://www.intel.com/technology/timeline.pdf

[5] SICAS. (2009) Semiconductor International Capacity Statistics: Status Report 2Q2009. [Online]. Available: http://www.sicas.info

[6] ITRS. (2009) ITRS report: 2008 edition. [Online]. Available: http://www.itrs.ney/reports.html

[7] J. Oliveira, J. Goes, B. Esperanca, N. Paulino, and J. Fernandes, "Low-Power CMOS Comparator with Embedded Amplification for Ultra-high-speed ADCs," in *IEEE International Symposium on Circuits and Systems, ISCAS 2007*, May 2007, pp. 3602–3605.

[8] J. Oliveira, J. Goes, N. Paulino, and J. Fernandes, "Improved Low-Power Low-Voltage CMOS Comparator for 4-Bit Flash ADCS for UWB Applications," in *14th International Conference on Mixed Design of Integrated Circuits and Systems, MIXDES '07*, 21-23 Jun. 2007, pp. 293–296.

[9] J. P. Oliveira, J. Goes, N. Paulino, J. Fernandes, and J. Paisana, "A multiplying-by-two CMOS amplifier for high-speed ADCs based on parametric amplification," in *15th International Conference on Mixed Design of Integrated Circuits and Systems, MIXDES '08*, 19-21 Jun. 2008, pp. 177 –180.

[10] J. Oliveira, J. Goes, N. Paulino, J. Fernandes, and J. Paisana, "New low-power 1.5-bit time-interleaved MDAC based on MOS capacitor amplification," in *15th IEEE International Conference on Electronics, Circuits and Systems, ICECS 2008*, Aug. 2008, pp. 251–254.

[11] J. Oliveira, J. Goes, M. Figueiredo, E. Santin, J. Fernandes, and J. Ferreira, "An 8-bit 120-MS/s Interleaved CMOS Pipeline ADC Based on MOS Parametric Amplification," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 57, no. 2, pp. 105–109, Feb. 2010.

[12] J. R. Custodio, J. Oliveira, L. Oliveira, J. Goes, and E. Brunn, "MOSFET-only Mixer/IIR filter with gain using parametric amplification," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2010, pp. 1209–1212.

[13] S. Ranganathan and Y. Tsividis, "A MOS capacitor-based discrete-time parametric amplifier with 1.2 V output swing and 3 μW power dissipation," in *IEEE International Solid-State Circuits Conference, ISSCC 2003, Digest of Technical Papers*, 2003, pp. 406–502.

[14] K. Iniewski, Ed., *Wireless Technologies: Circuits, Systems, and Devices.* NY, USA: CRC Press, 2007.

[15] OpenMoko. (2009) OpenMoko: 2009 edition. [Online]. Available: http://www.openmoko.org

[16] Android. (2009) Android 2.0. [Online]. Available: http://www.android.com

[17] C. A. Balanis, *Antenna Theory, Third Edition.* River Street, NJ, USA: John Wiley and Sons, 2005.

[18] I. D. O. Donnell and R. W. Brodersen, "An ultra-wideband transceiver architecture for low power, low rate, wireless systems," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 5, pp. 1623–1631, Sep. 2005.

[19] M. S. W. Chen and R. W. Brodersen, "A subsampling UWB radio architecture by analytic signaling," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, vol. 4, May 2004, pp. 533–536.

[20] IETF. (2010) Internet Engineering Task Force. [Online]. Available: http://www.ietf.org

[21] 3GPP. (2010) The 3rd Generation Partnership Project . [Online]. Available: http://www.3gpp.org/

[22] BTNode. (2008) BTNode: 2008 edition. [Online]. Available: http://www.btnode.ethz.ch/

[23] L. Wang, *Quantum Mechanical Effects on MOSFET Scaling Limit.* Sarrbrucken, Deutschland: VDM Verlag, 2009.

[24] A. J. G. Baptista, "Novel Techniques for the Design and Pratical Realization of Switched-Capacitor Circuits in Deep-Submicron CMOS technologies," Ph.D. dissertation, Faculdade de Ciências e Tecnologia - UNL, 2009.

[25] D. M. Binkley, *Tradeoffs and Optimization in Analog CMOS Design.* San Francisco, USA: John Wiley Sons Inc., 2008.

[26] J. Pekarik, D. Greenberg, B. Jagannathan, R. Groves, J. R. Jones, R. Singh, A. Chinthakindi, X. Wang, M. Breitwisch, D. Coolbaugh, P. Cottrell, J. Florkey, G. Freeman, and R. Krishnasamy, "RFCMOS technology from 0.25 μm to 65nm: the state of the art," *2004. Proceedings of the IEEE 2004 Custom Integrated Circuits Conference*, pp. 217–224, Oct. 2004.

[27] C. Enz, "An MOS transistor model for RF IC design valid in all regions of operation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 1, pp. 342 –359, Jan. 2002.

[28] H. Hassan, M. Anis, and M. Elmasry, "Impact of technology scaling on RF CMOS," in *SOC Conference, 2004. Proceedings. IEEE International*, Sep. 2004, pp. 97–101.

[29] Y. Tsividis, *Operation and Modeling of The MOS Transistor, Second Edition.* NY, USA: Oxford University Press, 1999.

[30] T. H. Lee, *The Design of CMOS radio-frequency integrated circuits, Second Edition.* Cambridge University Press, 2004.

[31] W. Liu, *MOSFET Models for SPICE Simulation including BSIM3v3 and BSIM4.* NY, USA: John Wiley and Sons, 2001.

[32] A. J. Annema, B. Nauta, R. van Langevelde, and H. Tuinhout, "Analog circuits in ultra-deep-submicron CMOS," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 1, pp. 132–143, Jan. 2005.

[33] Y. S. Lin, C. C. Wu, C. S. Chang, R. P. Yang, W. M. Chen, J. J. Liaw, and C. H. Diaz, "Leakage scaling in deep submicron CMOS for SoC," *IEEE Transactions on Electron Devices*, vol. 49, no. 6, pp. 1034–1041, Jun. 2002.

[34] M. J. M. Pelgrom, H. P. Tuinhout, and M. Vertregt, "Transistor matching in analog CMOS applications," in *International Electron Devices Meeting, IEDM '98 Technical Digest*, San Francisco, CA, USA, Dec. 1998, pp. 915–918.

[35] L. E. Larson, "Silicon technology tradeoffs for radio-frequency/mixed-signal "systems-on-a-chip"," *IEEE Transactions on Electron Devices*, vol. 50, no. 3, pp. 683–699, Mar. 2003.

[36] K. Lee, I. Nam, I. Kwon, J. Gil, K. Han, S. Park, and B.-I. Seo, "The impact of semiconductor technology scaling on CMOS RF and digital circuits for wireless application," *IEEE Transactions on Electron Devices*, vol. 52, no. 7, pp. 1415–1422, Jul. 2005.

[37] D. Leblebici and Y. Leblebici, *Fundamentals of High-Frequency CMOS Analog Integrated Circuits*. Cambridge University Press, 2009.

[38] S. C. Blaakmeer, E. A. M. Klumperink, D. M. W. Leenaerts, and B. Nauta, "Wideband Balun-LNA With Simultaneous Output Balancing, Noise-Canceling and Distortion-Canceling," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 6, pp. 1341–1350, Jun. 2008.

[39] L. B. Oliveira, J. R. Fernandes, I. M. Filanovsky, C. J. Verhoeven, and M. M. Silva, *Analysis and Design of Quadrature Oscillators*. Springer, 2008.

[40] J. Mitola, "The software radio architecture," *IEEE Communications Magazine*, vol. 33, no. 5, pp. 26–38, May 1995.

[41] A. A. Abidi, "The Path to the Software-Defined Radio Receiver," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 5, pp. 954–966, May 2007.

[42] Z. Ru, E. Klumperink, and B. Nauta, "A Discrete-Time Mixing Receiver Architecture with Wideband Harmonic Rejection," in *IEEE International Solid-State Circuits Conference, ISSCC 2008, Digest of Technical Papers*, 3-7 2008, pp. 322–616.

[43] R. G. Vaughan, N. L. Scott, and D. R. White, "The theory of bandpass sampling," *IEEE Transactions on Signal Processing*, vol. 39, no. 9, pp. 1973–1984, Sep. 1991.

[44] M. R. Yuce, A. Tekin, and W. Liu, "Design and performance of a wideband subsampling front-end for multi-standard radios," *AEU - International Journal of Electronics and Communications*, vol. 62, no. 1, pp. 41 – 48, 2008.

[45] B. Murmann, "Digitally Assisted Analog Circuits," *IEEE Micro*, vol. 26, no. 2, pp. 38–47, Mar.-Apr. 2006.

[46] V. Giannini, J. Craninckx, S. D'Amico, and A. Baschirotto, "Flexible Baseband Analog Circuits for Software-Defined Radio Front-Ends," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 7, pp. 1501–1512, Jul. 2007.

[47] B. Murmann, P. Nikaeen, D. J. Connelly, and R. W. Dutton, "Impact of Scaling on Analog Performance and Associated Modeling Needs," *IEEE Transactions on Electron Devices*, vol. 53, no. 9, pp. 2160–2167, Sep. 2006.

[48] J. Hu, N. Dolev, and B. Murmann, "A 9.4-bit, 50-MS/s, 1.44-mW pipelined ADC using dynamic residue amplification," in *IEEE Symposium on VLSI Circuits*, 18-20 2008, pp. 216–217.

[49] ——, "A 9.4-bit, 50-MS/s, 1.44-mW Pipelined ADC Using Dynamic Source Follower Residue Amplification," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 4, pp. 1057–1066, Apr 2009.

[50] M. E. Hines, "The Virtues of Nonlinearity–Detection, Frequency Conversion, Parametric Amplification and Harmonic Generation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 32, no. 9, pp. 1097–1104, Sep. 1984.

[51] H. Heffner, "Solid-State Microwave Amplifiers," *IRE Transactions on Microwave and Techniques*, pp. 83–91, 1959.

[52] R. J. Duffin, "Impossible Behavior of Nonlinear Networks," *Journal of Applied Physics*, vol. 26, no. 5, pp. 603–605, 1955. [Online]. Available: http://link.aip.org/link/?JAP/26/603/1

[53] P. Johannessen, W. Ku, and J. Andersen, "Theory of nonlinear reactance amplifiers," *IEEE Transactions on Magnetics*, vol. 3, no. 3, pp. 376–380, Sep. 1967.

[54] J. F. B. da Silva, "Amplificação Paramétrica em Estruturas Periódicas," Ph.D. dissertation, Instituto Superior Técnico - UTL, 1968.

[55] A. V. D. Ziel, "On the Mixing Properties of Non-Linear Condensers," *Journal of Applied Physics*, vol. 19, no. 11, pp. 999–1006, 1948. [Online]. Available: http://link.aip.org/link/?JAP/19/999/1

[56] H. Heffner and G. Wade, "Gain, Band Width, and Noise Characteristics of the Variable-Parameter Amplifier," *Journal of Applied Physics*, vol. 29, no. 9, pp. 1321–1331, 1958. [Online]. Available: http://link.aip.org/link/?JAP/29/1321/1

[57] H. Suhl, "Proposal for a Ferromagnetic Amplifier in the Microwave Range," *Phys. Rev.*, vol. 106, no. 2, pp. 384–385, Apr 1957.

[58] W. W. Mumford, "Some Notes on the History of Parametric Transducers," in *Proceedings of the IRE*, vol. 48, no. 5, May 1960, pp. 848–853.

[59] C. S. Aitchison, R. Davies, and P. J. Gibson, "A Simple Diode Parametric Amplifier Design for Use at S, C, and X Band," *IEEE Transactions on Microwave Theory and Techniques*, vol. 15, no. 1, pp. 22–31, Jan. 1967.

[60] P. Lombardo, "Parametric Amplifiers and Masers 1969," *IEEE Spectrum Magazine*, no. 8, pp. 5–5, 1969.

[61] C. Bell and G. Wade, "Iterative Traveling-Wave Parametric Amplifiers," *IRE Transactions on Circuit Theory*, vol. 7, no. 1, pp. 4–11, Mar. 1960.

[62] P. Tien and H. Suhl, "A Traveling-Wave Ferromagnetic Amplifier," *Proceedings of the IRE*, vol. 46, no. 4, pp. 700 –706, Apr 1958.

[63] A. Cullen, "Theory of the travelling-wave parametric amplifier," *Proceedings of the IEE - Part B: Electronic and Communication Engineering*, vol. 107, no. 32, pp. 101–107, Mar 1960.

[64] R. Engelbrecht, "Nonlinear-reactance (parametric) traveling-wave amplifiers for UHF," in *Solid-State Circuits Conference. Digest of Technical Papers. 1959 IEEE International*, Feb. 1959, pp. 8–9.

[65] G. M. Roe and M. R. Boyd, "Parametric Energy Conversion in Distributed Systems," in *Proceedings of the IRE*, vol. 47, no. 7, Jul. 1959, pp. 1213–1218.

[66] J. Ranuarez, Y. Ramadass, and M. Jamal Deen, "CMOS distributed amplifiers," in *Devices, Circuits and Systems, 2004. Proceedings of the Fifth IEEE International Caracas Conference on*, vol. 1, 3-5 2004, pp. 29 – 36.

[67] B. Kleveland, C. Diaz, D. Vock, L. Madden, T. Lee, and S. Wong, "Monolithic CMOS distributed amplifier and oscillator," in *Solid-State Circuits Conference, 1999. Digest of Technical Papers. ISSCC. 1999 IEEE International*, 1999, pp. 70 –71.

[68] M. Egels, J. Gaubert, P. Pannier, and S. Bourdel, "A 52-GHz 8.5-dB Traveling-Wave Amplifier in 0.13 μm Standard CMOS Process," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 56, no. 5, pp. 1226 –1233, may 2008.

[69] R. E. Collin, *Foundations For Microwave Engineering, Second Edition.* NY, USA: McGraw Hill, 1992.

[70] I. Gutierrez, J. Melendez, and E. Hernandez, *Design and Characterization of Integrated Varactors for RF Applications.* John Wiley and Sons, 2006.

[71] J. Maget, M. Tiebout, and R. Kraus, "MOS varactors with n- and p-type gates and their influence on an LC-VCO in digital CMOS," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 7, pp. 1139–1147, Jul. 2003.

[72] E. Pedersen, "RF CMOS Varactors for 2 GHz Applications," *Analog Integrated Circuits and Signal Processing.*, vol. 26, no. 1, Jan. 2001.

[73] C.-Y. Wu and C.-Y. Yu, "A 0.8 V 5.9 GHz wide tuning range CMOS VCO using inversion-mode bandswitching varactors," in *Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on*, May 2005, pp. 5079–5082.

[74] P. Andreani and S. Mattisson, "A 2.4-GHz CMOS monolithic VCO based on an MOS varactor," *Proceedings of the IEEE International Symposium on Circuits and Systems, ISCAS '99*, vol. 2, pp. 557–560, Jul. 1999.

[75] ——, "On the use of MOS varactors in RF VCOs," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 6, pp. 905–910, Jun. 2000.

[76] K. A. Jenkins and H. Ainspan, "Characteristics of submicron MOS varactors," in *Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems*, Jan. 2006.

[77] Y.-J. Chan, C.-F. Huang, C.-C. Wu, C.-H. Chen, and C.-P. Chao, "Performance Consideration of MOS and Junction Diodes for Varactor Application," *IEEE Transactions on Electron Devices*, vol. 54, no. 9, pp. 2570–2573, Sep. 2007.

[78] J. M. Manley and H. E. Rowe, "Some General Properties of Nonlinear Elements-Part I. General Energy Relations," in *Proceedings of the IRE*, vol. 44, no. 7, Jul. 1956, pp. 904–913.

[79] H. E. Rowe, "Some General Properties of Nonlinear Elements. II. Small Signal Theory," in *Proceedings of the IRE*, vol. 46, no. 5, May 1958, pp. 850–860.

[80] W. Bennett, "Amplification in Nonlinear Reactive Networks," *IRE Transactions on Circuit Theory*, vol. 7, no. 4, pp. 440–446, Dec. 1960.

[81] R. H. Pantell, "General Power Relationships for Positive and Negative Nonlinear Resistive Elements," in *Proceedings of the IRE*, vol. 46, no. 12, Dec. 1958, pp. 1910–1913.

[82] S. A. Maas, *Nonlinear microwave and RF circuits-2nd ed.* Boston, USA: Artech House, 2003.

[83] S. Magierowski, J. Bousquet, Z. Zhao, and T. Zourntos, "RF CMOS Parametric Downconverters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, no. 3, pp. 518–528, Mar. 2010.

[84] P. J. Penfield and R. R. Rafuse, *Varactor Applications.* Massachusets, USA: MIT Press, 1962.

[85] S. Magierowski, H. Chan, and T. Zourntos, "Subharmonically pumped RF CMOS paramps," *IEEE Transactions on Electron Devices*, vol. 55, no. 2, pp. 601–608, Feb. 2008.

[86] S. Magierowski, T. Zourntos, J.-F. Bousquet, and Z. Zhao, "Compact parametric downconversion using MOS varactors," in *IEEE MTT-S International Microwave Symposium Digest*, 2009, pp. 1377 –1380.

[87] H. Chan, Z. Chen, S. Magierowski, and K. Iniewski, "Parametric conversion using custom MOS varactors," *EURASIP J. Wirel. Commun. Netw.*, vol. 2006, no. 2, pp. 20–20, 2006.

[88] Z. Zhao, J.-F. Bousquet, and S. Magierowski, "Coherent parametric RF down-conversion in CMOS," in *IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, 2010, pp. 377–380.

[89] A. Suarez and R. Melville, "Simulation-assisted design and analysis of varactor-based frequency multipliers and dividers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 3, pp. 1166–1179, Mar. 2006.

[90] B. Gray, J. S. Kenney, and R. Melville, "Behavioral modeling and simulation of a parametric power amplifier," in *IEEE MTT-S International Microwave Symposium Digest*, Boston, MA, Jun. 2009, pp. 1373–1376.

[91] T. E. Stern, *Theory on Nonlinear Networks and Systems.* Massachusetss, USA: Addison-Wesley, 1965.

[92] A. A. Kharkevich, *Nonlinear and Parametric Phenomena in Radio Engineering.* NY, USA: John F. Rider Publisher, 1962.

[93] W. E. Danielson, "Low Noise in Solid State Parametric Amplifiers at Microwave Frequencies," *Journal of Applied Physics*, vol. 30, no. 1, pp. 8–15, 1959. [Online]. Available: http://link.aip.org/link/?JAP/30/8/1

[94] R. J. Baker, *CMOS Mixed-Signal Circuit Design: Second Edition.* NY, USA: John Wiley and Sons, 2009.

[95] A. Abu-Labdeh and S. Al-Jaber, "Energy consideration from non-equilibrium to equilibrium state in the process of charging a capacitor," *Journal of Electrostatics*, vol. 66, no. 3-4, pp. 190–192, 2008.

[96] A. S. Porret, T. Melly, C. C. Enz, and E. A. Vittoz, "Design of high-Q varactors for low-power wireless applications using a standard CMOS process," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 3, pp. 337–345, Mar. 2000.

[97] R. E. Joynson, J. L. Mundy, J. F. Burgess, and C. Neugebauer, "Eliminating threshold losses in MOS circuits by bootstrapping using varactor coupling," *IEEE Journal of Solid-State Circuits*, vol. 7, no. 3, pp. 217–224, Jun. 1972.

[98] Y. Tsividis and K. S. Suyama, "Strange ways to use the MOSFET," in *Proceedings of 1997 IEEE International Symposium on Circuits and Systems, ISCAS '97*, vol. 1, Jun. 1997, pp. 449–452.

[99] S. Ranganathan and T. Tsividis, "Discrete-time parametric amplification based on a three-terminal MOS varactor: analysis and experimental results," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 12, pp. 2087–2093, Dec. 2003.

[100] R. S. Tsividis Yannis, "Mosfet parametric amplifier," October 2009. [Online]. Available: http://www.freepatentsonline.com/y2009/0251196.html

[101] S. Ranganathan, "Discrete-Time Parametric Amplification Using MOSFETs," Ph.D. dissertation, Columbia University, NY, USA, 2004.

[102] P. M. Figueiredo and J. C. Vital, "The MOS capacitor amplifier," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 51, no. 3, pp. 111–115, Mar. 2004.

[103] C. Enz and Y. Cheng, "MOS transistor modeling for RF IC design," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 2, pp. 186–201, Feb. 2000.

[104] R. Jindal, "Distributed substrate resistance noise in fine-line NMOS field-effect transistors," *IEEE Transactions on Electron Devices*, vol. 32, no. 11, pp. 2450–2453, Nov. 1985.

[105] B. Parvais, S. Hu, M. Dehan, A. Mercha, and S. Decoutere, "An Accurate Scalable Compact Model for the Substrate Resistance of RF MOSFETs," *IEEE Custom Integrated Circuits Conference, CICC '07*, pp. 503–506, Sep. 2007.

[106] R. Chang, M. T. Yang, P. Ho, Y. J. Wang, Y. T. Chia, B. K. Liew, C. Yue, and S. Wong, "Modeling and optimization of substrate resistance for RF-CMOS," *IEEE Transactions on Electron Devices*, vol. 51, no. 3, pp. 421–426, Mar. 2004.

[107] J.-J. Ou, X. Jin, I. Ma, C. Hu, and P. Gray, "CMOS RF modeling for GHz communication IC's," *Symposium on VLSI Technology, Digest of Technical Papers*, pp. 94–95, Jun. 1998.

[108] J. Han and H. Shin, "A scalable model for the substrate resistance in multi-finger RF MOSFETs," *IEEE MTT-S International Microwave Symposium Digest*, vol. 3, pp. 2105–2108, Jun. 2003.

[109] Y. Cheng and C. Hu, *MOSFET MODELING and BSIM3 Users Guide.* Dordrecht, The Netherlands: Kluwer Academic Publisher, 1999.

[110] R. T. Y. Mikami, "Large- and small-signal channel transit time delays in long-channel MOS transistors," *IEEE Transactions on Electron Devices*, vol. 24, no. 2, pp. 99–107, Feb 1977.

[111] W. E. Engeler, J. J. Tiemann, and R. D. Baertsch, "Surface charge Transport in Silicon," *Applied Physics Letters*, vol. 17, no. 11, pp. 469–472, Dec. 1970.

[112] J. Carnes, W. Kosonocky, and E. Ramberg, "Free charge transfer in charge-coupled devices," *IEEE Transactions on Electron Devices*, vol. 19, no. 6, pp. 798–808, Jun. 1972.

[113] R. Gregorian and G. C. Temes, *Analog MOS Integrated Circuits.* NY, USA: John Wiley and Sons, 1986.

[114] G. Xu and J. Yuan, "Accurate sample-and-hold circuit model," *Electronics Letters*, vol. 41, no. 9, pp. 520–522, Apr. 2005.

[115] P. Figueiredo and J. Vital, *Offset Reduction Techniques in High-Speed Analog-to-Digital Converters.* NY, USA: Springer, 2009.

[116] R. J. Baker, *CMOS Circuit Design, Layout, and Simulation: Second Edition.* NY, USA: IEEE Press - John Wiley and Sons, 2008.

[117] S. Chehrazi, R. Bagheri, and A. A. Abidi, "Noise in passive FET mixers: a simple physical model," *Proceedings of the IEEE 2004 Custom Integrated Circuits Conference*, pp. 375–378, Oct. 2004.

[118] S. Chehrazi, A. Mirzaei, and A. Abidi, "Noise in Current-Commutating Passive FET Mixers," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 2, pp. 332–344, Feb. 2010.

[119] J. Fernandes and M. Silva, "Asymmetric latches and nonideal performance of parallel ADCs," in *Proceedings of the 7th Mediterranean Electrotechnical Conference*, vol. 2, 12-14 1994, pp. 629–632.

[120] K. Kotani, T. Shibata, and T. Ohmi, "CMOS charge-transfer preamplifier for offset-fluctuation cancellation in low-power A/D converters," *Solid-State Circuits, IEEE Journal of*, vol. 33, no. 5, pp. 762 –769, may 1998.

[121] B. Razavi and B. Wooley, "Design techniques for high-speed, high-resolution comparators," *IEEE Journal of Solid-State Circuits*, vol. 27, no. 12, pp. 1916–1926, Dec. 1992.

[122] A. Yoshizawa and S. Lida, "A Gain-Boosted Discrete-Time Charge-Domain FIR LPF with Double-Complementary MOS Parametric Amplifiers," in *IEEE International Solid-State Circuits Conference, ISSCC 2008, Digest of Technical Papers*, San Francisco, CA, Feb. 2008, pp. 68–596.

[123] K. Muhammad and R. Staszewski, "Direct RF sampling mixer with recursive filtering in charge domain," in *Circuits and Systems, 2004. ISCAS '04. Proceedings of the 2004 International Symposium on*, vol. 1, 23-26 2004, pp. I – 577–80 Vol.1.

[124] K. Muhammad, D. Leipold, B. Staszewski, Y.-C. Ho, C. Hung, K. Maggio, C. Fernando, T. Jung, J. Wallberg, J.-S. Koh, S. John, I. Deng, O. Moreira, R. Staszewski, R. Katz, and O. Friedman, "A discrete-time Bluetooth receiver in a 0.13 μm digital CMOS process," *IEEE International Solid-State Circuits Conference, ISSCC 2004, Digest of Technical Papers*, vol. 1, pp. 268–527, Feb. 2004.

[125] R. B. Staszewski, K. Muhammad, D. Leipold, C.-M. Hung, Y.-C. Ho, J. L. Wallberg, C. Fernando, K. Maggio, R. Staszewski, T. Jung, J. Koh, S. John, I. Y. Deng, V. Sarda, O. Moreira-Tamayo, V. Mayega, R. Katz, O. Friedman, O. E.

Eliezer, E. de Obaldia, and P. T. Balsara, "All-digital TX frequency synthesizer and discrete-time receiver for Bluetooth radio in 130-nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 12, pp. 2278–2291, Dec. 2004.

[126] Z. Ru, E. Klumperink, and B. Nauta, "On the Suitability of Discrete-Time Receivers for Software-Defined Radio," in *IEEE International Symposium on Circuits and Systems, ISCAS 2007*, May 2007, pp. 2522–2525.

[127] João Carlos da Palma Goes, "Optimization of Self-Calibrated CMOS Pipelined Analogue-to-Digital Converters," Ph.D. dissertation, Instituto Superior Técnico - UTL, 2000.

[128] T. B. Cho, "Low-Power Low-Voltage Analog-to-Digital Conversion Techniques using Pipelined Architectures," Ph.D. dissertation, Berkeley University, Berkeley, USA, 1995.

[129] D. Camarero, *Time-Interleaved ADCs: A Mixed-Signal Solution for Clock-Skew Problems.* VDM Verlag, 2009.

[130] B. Murmann, "ADC Performance Survey 1997-2010," Tech. Rep., 2010. [Online]. Available: http://www.stanford.edu/murmann/adcsurvey.html

[131] D. K. Shaeffer and T. H. Lee, "A 1.5-V, 1.5-GHz CMOS low noise amplifier," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 5, pp. 745–759, May 1997.

[132] X. Guan and A. Hajimiri, "A 24-GHz CMOS front-end," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 2, pp. 368–373, Feb. 2004.

[133] D. Im, I. Nam, H.-T. Kim, and K. Lee, "A Wideband CMOS Low Noise Amplifier Employing Noise and IM2 Distortion Cancellation for a Digital TV Tuner," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 3, pp. 686–698, Mar. 2009.

[134] S. C. Blaakmeer, E. Klumperink, D. M. W. Leenaerts, and B. Nauta, "The Blixer, a Wideband Balun-LNA-I/Q-Mixer Topology," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 12, pp. 2706–2715, Dec. 2008.

[135] S. Wang, A. Niknejad, and R. Brodersen, "Design of a Sub-mW 960-MHz UWB CMOS LNA," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 11, pp. 2449–2456, Nov. 2006.

[136] J. Rudell, J.-J. Ou, T. Cho, G. Chien, F. Brianti, J. Weldon, and P. Gray, "A 1.9-GHz wide-band IF double conversion CMOS receiver for cordless telephone applications," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 12, pp. 2071–2088, Dec. 1997.

[137] M. Terrovitis and R. Meyer, "Noise in current-commutating CMOS mixers," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 6, pp. 772–783, Jun. 1999.

[138] R. Bagheri, A. Mirzaei, S. Chehrazi, M. E. Heidari, M. Lee, M. Mikhemar, W. Tang, and A. A. Abidi, "An 800-MHz–6-GHz Software-Defined Wireless Receiver in 90-nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 12, pp. 2860–2876, Dec. 2006.

[139] H. Pekau and J. W. Haslett, "A 2.4 GHz CMOS sub-sampling mixer with integrated filtering," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 11, pp. 2159–2166, Nov. 2005.

[140] D. Jakonis and C. Svensson, "A 1.6 GHz downconversion sampling mixer in CMOS," in *Proceedings of the International Symposium on Circuits and Systems, ISCAS '03*, vol. 1, 25-28 2003, pp. 725–728.

[141] J. Ferreira, I. Bastos, L. Oliveira, J. P. Oliveira, T. Michalak, P. Pankiewicz, B. Nowacki, P. Makosa, and A. Rybarczyk, "LNA, oscillator and mixer co-design for compact RF-CMOS ISM receivers," in *Mixed Design of Integrated Circuits and Systems, 2009. MIXDES '09. MIXDES-16th International Conference*, Lodz, Jun. 2009, pp. 291–295.

[142] J. P. Oliveira, L. Oliveira, J. Ferreira, I. Bastos, T. Michalak, P. Pankiewicz, B. Nowacki, P. Makosa, and A. Rybarczyk, "Co-Design Strategy Approach of LNA, Oscillator, and Mixer," *Intl. Journal of Electronics And Telecomunications Quartely*, vol. 55, no. 3, pp. 37–51, Dec. 2009.

[143] J. Ferreira, I. Bastos, L. Oliveira, J. P. Oliveira, T. Michalak, P. Pankiewicz, B. Nowacki, P. Makosa, and A. Rybarczyk, "Outstanding Paper Award presented to : LNA, oscillator and mixer co-design for compact RF-CMOS ISM receivers," in *Mixed Design of Integrated Circuits and Systems, 2009. MIXDES '09. MIXDES-16th International Conference*, Lodz, Jun. 2009, pp. 291–295.

[144] A. Liscidini, A. Mazzanti, R. Tonietto, L. Vandi, P. Andreani, and R. Castello, "Single-Stage Low-Power Quadrature RF Receiver Front-End: The LMV Cell," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 12, pp. 2832 –2841, Dec. 2006.

Index