Continuous-Time Bandpass (BP) sigma delta modulator (SDM)

Zoltan Nemes

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES

Spécialité : Nano Electronique et Nano Technologies

Arrêté ministériel : 25 mai 2016

Présentée par

Zoltan NEMES

Thèse dirigée par
Dominique MORCHE, Professeur, CEA,
et codirigée par
Maurits ORTMANNS, Professeur, Université de Ulm, et
Stéphane LE TUAL, Ingénieur, STMicroelectronics

préparée au sein du **CEA LETI au Laboratoire d'Architectures Intégrées Radiofréquence (LAIR) et STMicroelectronics et à l'Institut de Microélectronique de l'Université de Ulm**
dans **l'École Doctorale électronique, électrotechnique, automatique, traitement du signal (EEATS)**

Conception d'un récepteur à formation de faisceaux numérique pour petites cellules millimétriques 5G

Thèse soutenue publiquement le **30 juin 2021**,
devant le jury composé de :

Pr. Salvador MIR
Directeur de recherche laboratoire TIMA, Président
Pr. Hassan ABOUSHADY
Professeur associé à Sorbonne Université, Rapporteur
Pr. Boris MURMANN
Professeur à l'Université de Stanford, Rapporteur
Pr. Sven MATTISSON
Professeur assistant à l'Université de Lund, Membre
M. Cédric DEHOS
Ingénieur de recherche au CEA LETI, Invité

Acknowledgment

During any PhD, one goes through many hardships, and when comes the time to acknowledge the ones who allowed him to over pass these hardships, comes the fear to forget someone. I am no exception to this statement. Thankfully, I know that the people who kindly helped me will not feel offended if by mistake I do not mention them.

First, I would like to thank Stéphane Le Tual, my supervisor and college at ST. When I decided to undertake the challenge of a PhD, I asked him right away if he wanted to be part of the project. Thankfully he accepted, allowing me to wright these lines today. I must Also thank Olivier Rossetto for accepting me in the master program and then following me through to the end. It is during this master's degree that I met Dominique Morche who would become my director. He made me redo many, if not all, of my presentations multiple times, until it was deemed good enough. Thank you for believing in me when didn't myself, it was worse it. I was lucky to have as co-director Maurits Ortmanns, his mind speed and sharpness pushed me further than I could imagined. Thank you for your benevolent pressure.

I want to thank Sven Mattisson both for the exchanges we add early in my PhD and for being a member of my jury. Similarly, I want to thank Cedric Dehos for all the system discussions, for being a member of my CSI and finally for being part of the jury as well. I also thank Hassan Aboushady for his interest in my work during the CSI meetings and for being reporter jury member. I want to thank Boris Murmann for getting up that early on the defense day, and of course for being a reporter member of the jury. I also thank Mir Salvador for presiding the jury.

I want to thank Pierre-Olivier Jouffre for supporting this project all along as well as Yves Desserez for his precious help on the layout. I thank Michael Pelissier for our discussion on compress sensing and Alexandre Giry for teaching me the basics on power amplifiers.

I need to thank all my fellow PhD student at CEA/LETI and ST, for the technical discussions and all the rest.

I thank my family, in particular my mother my sister and my brother for supporting me until the end.

Last, I am proud to thank Aline for being the spark of this adventure and for joining me until its end.

Abstract:

The fifth-generation mobile network, or 5G, will become a standard for nearly all forms of wireless communication. In that purpose, it will use a larger part of the spectrum. The sub-6GHz 5G is currently being deployed. The millimeter wave spectrum exploitation will start in the coming years. This part of the spectrum is envisioned to provide enhanced Mobile Broad Band (eMBB) over small areas. These access points with limited coverage are called small-cells and are the focus of this manuscript. More specifically, the challenges of base station receivers for these small cells in the 28GHz band will be studied. This work is divided into three parts. First, the system is analyzed to establish the receiver requirements. Second, these requirements are used to propose an innovative receiver's architecture. Finally, an implementation of the proposed architecture is described and evaluated.

The system analysis starts from 5G's Key Performance Indicators (KPI). A system architecture in line with these KPI is established and becomes the basis to evaluate the receiver's requirements in a multiple operator scenario. One specific trait of the receiver is its beamforming ability using a large antenna array. While this approach has the potential to deliver the desired performances, it also increases the receiver's complexity. The system has many parameters (number of antennas, array topology, …), leading to many possible configurations. Finding the optimal configuration is very challenging. To alleviate this problem, as many parameters as possible were fixed, based on practical considerations. This significantly reduces the size of the problem and simplifies the analysis.

Beamforming consists in combining the signals from multiple antennas, to receive only the radio waves from a given direction, forming a beam in that direction. Prior to recombination, the signals must be delayed and weighted. The domain where these operations are performed defines the receiver architecture. When in the analog domain, it is analog beamforming. When in the digital domain, it is Digital Beamforming (DBF) and when in both domains, it is hybrid beamforming. DBF offers the best performances but has the most challenging RF front end implementation, requiring a full receiver per antenna. The proposed analysis shows that the performances required for these individual receivers are significantly relaxed, the challenge laying more on the digital side, due to the large amount of data to process in a short time.

Hence, receivers benefiting from relaxed requirements, while reducing the digital processing were investigated. Using band pass Sigma-Delta Modulators (SDM) for analog to digital conversion can reduce the digital processing, thanks to its oversampling and low-resolution output. The former provides a nearly free delay by just selecting the samples. The later provides a cheap multiplexer-based multiplication. To simplify the receiver as much as possible, RF sampling was investigated, where the receiver is reduced to the SDM. This was made possible by using a sub-sampling approach. Even though, the sampling frequency remains high, and closing the loop and compensating for Excess Loop Delay (ELD) are very challenging. One major result was to show that some sub-sampling SDM could be made ELD compensation free and provide more than one clock cycle to close the loop. This allows for a two-times interleaved quantizer and is a key enabler.

In addition to the interleaved quantizer, the proposed implementation features transformer-based resonators. The additional degree of freedom offered by the ratio between the primary and secondary inductances is very useful to improve power efficiency. While simulation results are below expectation, they are good enough for a proof of concept. A test chip integrating 8 parallel receivers was send for fabrication in a CMOS 28nm FDSOI process from STMicroelectronics and is yet to be tested.

Key words: 5G, Millimeter Wave, Digital Beamforming, Analog to Digital Converters, Band Pass Continuous Time Sigma Delta Modulators, Excess Loop Delay, Sub-sampling, RF Sampling, CMOS.

Résumé:

Le réseau mobile de cinquième génération, ou 5G, tend à devenir le standard pour l'ensemble des communications sans fil. Il utilisera une plus grande portion du spectre. Les déploiements actuels se concentre sur la bande sous 6GHz. L'exploitation du spectre millimétrique commencera elle dans les années à venir. Il servira à fournir un service mobile large band amélioré sur de petites surfaces. Ces points d'accès à couverture limitée s'appellent petite cellule et sont le sujet de ce travail de thèse. Le cœur de l'étude porte sur le récepteur de ces petites cellules, dans la bande à 28GHz. Elle se divise en trois parties. Une analyse système permettant d'établir les spécifications du récepteur, la proposition d'une architecture de récepteur innovante et une description et une évaluation de l'implémentation proposée.

L'analyse système se base sur les indicateurs de performance clés de la 5G. En partant d'une architecture en ligne avec ces indicateurs, on dérive les spécifications requises du récepteur dans un scénario multi-opérateur. Une caractéristique spécifique de ces récepteurs est leur capacité à former des faisceaux à l'aide de larges tableaux d'antennes. Bien que cette approche ait le potentiel pour satisfaire les objectifs de la 5G, elle est plus complexe. Les nombreux paramètres (nombre d'antennes, topologie du tableau, …) engendrent beaucoup de configurations possibles et trouver l'optimum devient difficile. Une solution est de fixe un maximum de paramètres sur la base de considérations pratiques, permettant une analyse simplifiée.

La formation de faisceaux se fait par la combinaison des signaux de plusieurs antennes pour recevoir les ondes provenant d'une direction privilégiée. Avant cette combinaison, les signaux sont retardés et pondérés. Le domaine dans lequel ces opérations sont faites défini l'architecture du récepteur. Si elles s'opèrent dans les domaines analogique, numérique ou une combinaison des deux, on parle de formation de faisceaux analogique, numérique ou hybride. L'approche numérique est la plus performante, mais la plus difficile à implémenter. Il faut une chaine complète de réception par antenne. L'analyse proposée montre que les performances requises pour ces récepteurs individuels sont relâchées, et que le défi se trouve dans la gestion en temps réel des données numériques.

Ainsi, les récepteurs permettant une réduction du traitement numérique furent investigués. L'utilisation de Modulateurs Sigma-Delta (MSD), pour la conversion analogique numérique, peut réduire le traitement numérique, grâce à leur sur-échantillonnage et leurs signaux de sortie de faible résolution. L'un permet la réalisation d'un retard presque gratuit en sélectionnant les échantillons. L'autre fourni une multiplication bas coût, à base de multiplexer. Pour simplifier le récepteur, l'échantillonnage direct du signal RF fut investiguée. Le récepteur est alors réduit au MSD. C'est rendu possible grâce au sous-échantillonnage. La fréquence d'échantillonnage reste élevée, et la fermeture de la boucle ainsi que la compensation du Retard de Boucle (RB) reste un défi. Un résultat majeur fut de montrer que certain MSD sous-échantillonnés pouvaient être réalisé sans compensation du RB et avec un temps de fermeture de boucle supérieure à une période d'horloge. Cela permet l'utilisation d'un quantificateur deux fois entrelacé en temps, et rend cette approche réalisable.

En plus du quantificateur entrelacé, l'implémentation proposée présent des résonateurs à base de transformateurs. Le degré de liberté offert par le rapport entre les inductances du primaire et du secondaire est très utile pour améliorer la consommation énergétique. Bien que les résultats de simulations soient moins bons qu'escompté, ils sont suffisamment bons pour établir une preuve de concept. Une puce de test intégrant 8 récepteurs en parallèle fut envoyée en fabrication dans un procédé CMOS 28nm FDSOI de STMicroelectronics et reste à être mesuré.

Mots clés : 5G, Ondes Millimétriques, Formation de Faisceaux Numérique, Convertisseurs Analogique Numérique, Modulateurs Sigma Delta Passe Bande à Temps Continu, Retard de Boucle, Sous-échantillonnage, Échantillonnage RF, CMOS.

List of Publications:

- Z. Nemes, D. Morche, S. L. Tual and M. Ortmanns, "A Feasibility Study of Digital Beamforming for 5G mmWave Massive MIMO Base Station Receivers," 2018 16th IEEE International New Circuits and Systems Conference (NEWCAS), Montreal, QC, Canada, 2018, pp. 1-5.

Table of Contents

List of Acronyms

AAP	Adaptive Array Processing
ABF	Analog Beamforming
AC	Alternating Current
ACLR	Adjacent Channel Leakage Ratio
ADC	Analog to Digital Converter
AF	Array Factor
ALTR	ALTernate channel leakage Ratio
AoA	Angle of Arrival
AWGN	Additive White Gaussian Noise
BB	Base Band
BBP	Base Band Processor
BG	Back Gate
BiCMOS	Bipolar-CMOS
BLER	BLock Error Rate
BPCTSDM	Band Pass Continuous Time Sigma-Delta Modulator
BPDTSDM	Band Pass Discrete Time Sigma-Delta Modulator
BS	Base Station
CDMA	Code Division Multiple Access
CDS	Correlated Double Sampling
CMOS	Complementary Metal Oxid Semiconductor
CMRR	Common Mode Rejection Ratio
CP	Charge Pump
CSI	Channel State Information
CT	Continuous Time
CTFBIR	Continuous Time Feedback Impulse Response
CTSDM	Continuous Time Sigma-Delta Modulators
CVDAC	Capacitively coupled Voltage DAC
DAC	Digital to Analog Converters
DBF	Digital Beamforming
dBm	decibels per milli-Watt
DC	Direct Current
DFT	Discrete Fourier Transform
DQAM	Differential QAM
DR	Dynamic Range
DS	Delay and Sum
DT	Discrete Time
DTFBIR	Discrete Time Feedback Impulse Response
DTSDM	Discrete Time Sigma-Delta Modulator
DWS	Delay Weight and Sum
ELD	Excess Loop Delay
eMBB	enhanced Mobile Broad Band
ESD	Electro-Static Discharge

EVM	Error Vector Modulation
FD	Frequency Divider
FDMA	Frequency Division Multiple Access
FDSOI	Fully Depleted Silicon on Insulator
FIR	Finite Impulse Response
GSG	Ground Signal Ground
GSM	Global System for Mobile communication
HBF	Hybrid Beamforming
HPBW	Half Power Beam Width
HRZ	Half delayed Return to Zero
IBI	In Band Interferers
ICI	Inter Carrier Interference
IDAC	Current DAC
IF	Intermediate Frequency
IIP3	third order Input Intercept Point
IL	Insertion Loss
IM3	third order Inter-Modulation
IM5	fifth order Inter-Modulation
IRR	Interferer Rejection Ration
ISF	Impulse Sensitivity Function
KPI	Key Performance Indicators
LDPC	Low Density Parity Check
LED	Light Emitting Diode
LF	Loop Filter
LNA	Low Noise Amplifier
LO	Local Oscillator
LoS	Line of Sight
LP	Low Pass
LPDTSDM	Low Pass Discrete Time Sigma-Delta Modulator
LPF	Low Pass Filter
LTI	Linear Time Invariant
LTV	Linear Time Varying
MACLR	Modified ACLR
M-BS	Macro-cell BS
MC-UCA	Multiple Concentric UCA
MIMO	Multiple Input Multiple Output
ML	Maximum Likelihood
MMSE	Minimum Mean-Square Error
mMTC	massive Machine Type Communication
MPW	Multi-Project Wafer
MRC	Maximum Ratio Combining
MU Massive MIMO	Multiple Users Massive MIMO
MU-MIMO	Multiple User MIMO

Near-ZIF	Near Zero Intermediate Frequency
NF	Noise Figure
NLoS	None Line of Sight
NOMA	Non-Orthogonal Multiple Access
NPR	Noise Power Ratio
NRZ	Non-Return to Zero
NTF	Noise Transfer Function
NTN	Non-Terrestrial Networks
NZ	Nyquist Zone
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OIP3	third order Output Intercept Point
OMA	Orthogonal Multiple Access
OoBI	Out of Band Interferers
OS	OverShoot
OSR	Over Sampling Ratio
PA	Power Amplifier
PAPR	Peak to Average Power Ratio
PCB	Printed Circuit Board
PFD	Phase Frequency Detector
PLL	Phase Lock Loop
PM	Phase Margin
PN	Phase Noise
PSD	Power Spectral Density
PSL	Peak Side Lobe
PSS	Phase Shift and Sum
PVT	Process and Temperature Variations
QAM	Quadrature Amplitude Modulation
RB	Resource Block
RE	Resource Element
RF	Radio Frequency
RMS	Root Mean Square
RRC	Root Raise Cosine
Rx	Receiver
RZ	Return to Zero
SAR	Successive Approximation Register
S-BS	Small-cell BS
SCS	Sub Carrier Spacing
SDM	Sigma-Delta Modulators
SF	Source Follower
SFDR	Spurious Free Dynamic Range
SH	Sample and Hold
SINR	Signal to Interferer plus Noise Ratio
SMS	Short Messaging System

SNR	Signal to Noise Ratio
SOI	Silicon On Insulator
SPI	Serial Parallel Interface
SQNR	Signal to Quantization Noise Ratio
SRx	Single Receiver
SSB	Single Side Band
STF	Signal Transfer Function
TDD	Time Division Duplexing
TDMA	Time Division Multiple Access
THD	Total Harmonic Distortion
TSPC	True Single-Phase Clock
Tx	Transmitter
UCA	Uniform Circular Arrays
UE	User Equipment
ULA	Uniform Linear Arrays
UPA	Uniform Planar Arrays
URLLC	Ultra-Reliable Low Latency Communication
VCO	Voltage Controlled Oscillator
ZF	Zero Forcing

List of Figures

List of Tables

1 CHAPTER I: INTRODUCTION

The history of wireless communication goes back to the end of the 19th century, with the pioneer work on wireless telegraphy from Guglielmo Marconi. This was made possible thanks to the understanding of the electromagnetism laws, theorized by James Clerk Maxwell in 1865, and experimentally confirmed by Heinrich Hertz in the 1880's. From that point on the field of wireless communication developed tirelessly until today and the current deployment of the 5th generation mobile network, going by the name of 5G.

The DNA of modern wireless communication is often tied to the year 1948. This year saw two major advances that are still the base for 5G. First is the publication by Claud Shannon of his funding article on information theory, which is the base to devise efficient and error free communication systems in a noisy environment. Second is the invention of the bipolar junction transistor by William Shockley, which allows to build more compact and power efficient transceivers.

While these two fields are still at the root of today's mobile systems, they have greatly evolved. Modern error correcting codes are nearly reaching Shannon's limit, and transistors' performances, as well as their integration, have improved by several orders of magnitude. It is now common to find chips with several billion transistors in a volume smaller than Shockley's original single transistor.

Like many fields, wireless communications were greatly developed under military impulse, before reaching our everyday life. They are now omnipresent in the modern society. This manuscript will focus on the use case consisting in providing a two-way communication link to a mobile user that can roam virtually anywhere. This has been achieved by deploying networks of fixed antennas, each of them covering an area called a cell. All the antennas form what is called a mobile network. Today, the fifth generation is being deployed. Each of them has one or more technical specificities which are often inherited by the following generations.

The first generation was based on an analog modulation of the signal and was designed to only carry the voice. Already, the covered areas were divided into cells, each of them covered by a single antenna, or Base Station (BS), connected to the landlines, thereby making a cellular network.

The second generation introduced a digital modulation of the signal, improving significantly on the communications' quality. It also allows for the transmission of small amount of other than voice data. This was used to create the Short Messaging System or SMS. This second generation, named Global System for Mobile communication (GSM), was the first mobile network to meet a commercial success.

The third-generation innovation was about handling users' multiple access. While previous generation were using a combination of Frequency Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA), 3G used Code Division Multiple Access (CDMA) which allows for multiple users to communicate over the same frequency at the same time. It also improved the network capability to transmit larger amount of data. Its commercial success was not on par with its predecessor. In its early deployment, its improved ability for data transfer was virtually unused since no device could actually use it efficiently at the time. Once the smartphone revolution came in, its data rate proved to be insufficient, rapidly calling for the deployment of the next generation.

The fourth-generation network answered the call for higher data rate, introducing multiple innovations. First is a combination of a new modulation, the Orthogonal Frequency Division Multiplexing (OFDM) with an increase of the channel bandwidth up to 20MHz. OFDM Allows for a better bandwidth efficiency as well as being more robust to fast fading channels, making it a good solution for wider bandwidth and allowing the channel width increase. Second is the possibility to use carrier aggregation up to three channels for a maximum cumulated bandwidth of 60MHz. Finally, it also introduced Multiple Input Multiple Output (MIMO) communications with up to eight antennas on each side. In the

latest release of the standard all these improvements combined allows for a theoretical downlink data rate up to 300Mb/s, and 150Mb/s in uplink. In practice users rarely get rates above 100Mb/s, and typical rates are generally in the range from 1Mb/s to 10Mb/s. This discrepancy between theoretical and experienced data rates has two major reasons. Channels with a quality allowing for peak rate are extremely rare, if not inexistent, and the network is often near saturation of its capacity. This means that, even if the communication channel between a user and a BS is good enough for a 100Mb/s downlink communication, because the BS must serve multiple users at the same time it may not be able to provide this data rate without reducing the service to other users. Hence the limitation comes more from the network total throughput than the user link itself. This limitation comes from, either the BS capacity itself or from the backhaul link establishing the connection between the BS and the core network.

One of 5G goals is to provide a solution for this limitation. It will be shown that it is in fact much more ambitious than that and aims at becoming a standard for nearly all forms of wireless communications.

1.1 5TH GENERATION MOBILE NETWORK OVERVIEW

5G has been marketed as a disruptive technology with a capacity increase of two orders of magnitude compared to 4G, with data rates up to 10Gb/s or even 20Gb/s delivered to a single user. While this is already an ambitious goal, it aims at providing much more than that. As already discuss one challenge coming along increased peak rate is increased network capacity. This is generally referred to as enhanced Mobile Broad Band (eMBB). Beyond that, 5G also aims at providing the framework for massive Machine Type Communication (mMTC), also called the Internet of Things (IoT), where a very large number of devices are connected to the network. The last goal for 5G is to provide an Ultra Reliable Low Latency Communication (URLLC) system for mission critical applications. There are also discussions about including Non-Terrestrial Networks (NTN) using satellites or any other kind of air born vehicles to create a network. It is not clear yet if it will or how it could be part of 5G, so it will not be discussed further but it shows the ambition of 5G to propose a common framework for as many wireless communication systems as possible.

1.1.1 The three aspects of 5G: mMTC, URLLC, and eMBB

The evolution from 4G to 5G is often represented by the Key Performance Indicators (KPI) diagram in Figure 1-1.

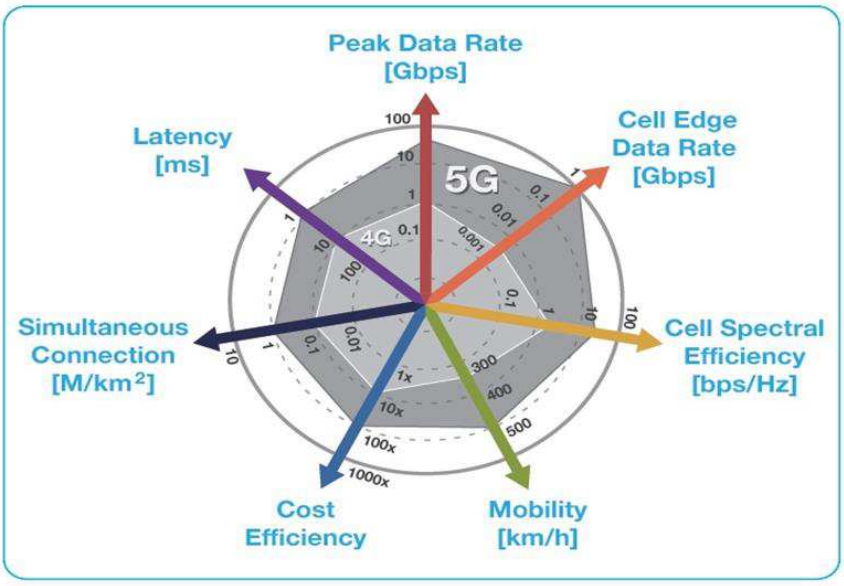

Figure 1-1: 4G versus 5G Key Performance Indicators

While the data it contains are accurate this diagram must be properly read. While 5G's target is to outperform 4G by one or two orders of magnitude in all aspects, it is not aiming at doing so all at the same time. This is simply and purely forbidden by the laws of physics as they are currently known.

For example, peak data rate, cell edge data rate and cell spectral efficiency will be the playground of eMBB but in these conditions the number of simultaneous connections will never even be close to a million connections per square kilometer. This level of connection density is only aimed for mMTC, while eMBB targets a user density of ten thousand users per square kilometer, each of them experiencing an average data rate of 100Mb/s, far from the ten or twenty gigabit per second peak rate. Similarly, the one millisecond latency only applies to URLLC. The target latency for eMBB is around 10 milliseconds, on par with current 4G performances.

One of the challenges is to provide a framework that is compatible with all three cases. An overview of mMTC and URLLC will be presented here. eMBB will be the focus of this manuscript and will be further detailed in the subsequent sections. First, a short description of the radio frame structure together with some vocabulary is required.

1.1.1.1 Radio frame structure

For a given cell, there is only a limited amount of time-frequency resource dedicated to 5G. The radio frame structure basically describes how this resource is organized, how it is divided into sub-pieces of resources. In general, it is desirable for all sub-resources to be orthogonal to each other. This means that a given piece of spectrum at a given time will belong to a single sub-resource. By doing so it is then possible to allocate these different sub-resources to different users. This gives rise to the broad category of Orthogonal Multiple Access (OMA). 5G uses an OFDM modulation, hence the specific multiple access technique used is Orthogonal Frequency Division Multiple Access (OFDMA).

A frame corresponds to a piece of spectrum, also called a channel, for a given duration. In 5G the frame duration is fix and equal to ten milliseconds, while the channel bandwidth is flexible.

In the time domain the frame is divided into ten sub-frames of one millisecond. Each sub-frame is divided into a flexible number of slots, each of them being made of 14 symbols. There is a case where a slot can be made of only 12 symbols but, for the sake of brevity, it will not be discussed here. The number of slots per sub-frame then depends on the symbol duration. This will be defined when looking at the structure in the frequency domain. Figure 1-2 gives a visual representation of this frame slicing for the case of one slot per sub-frame.

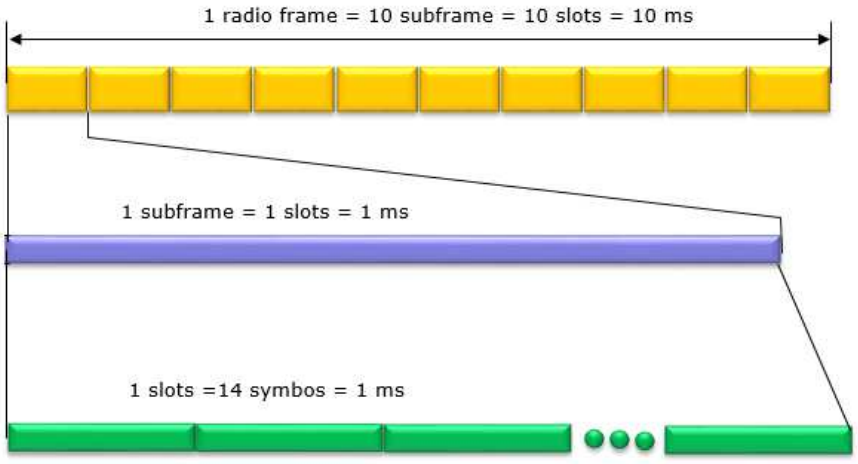

Figure 1-2: Radio frame division and sub-division in the time domain

In the frequency domain the channel bandwidth is divided into sub-channels. This is characterized by the Sub Carrier Spacing (SCS) which is flexible. It is the SCS Δf that also defines the symbol duration $T_{sym} = 1/\Delta f$. The smallest SCS Δf is 15kHz and can be increased by powers of two up to 240kHz. Doubling the SCS halves the symbol duration and doubles the number of slots per sub-frame.

It is common to represent one sub-frame by the resource grid from Figure 1-3. One slot of one sub-carrier is called a Resource Element (RE). The REs are assembled by groups of twelve consecutive sub-carriers over the same slot to form a Resource Block (RB). RBs are the smallest resource that can be allocate to a user. From this basic description of the radio frame structure, the challenges of mMTC and URLLC can be discussed.

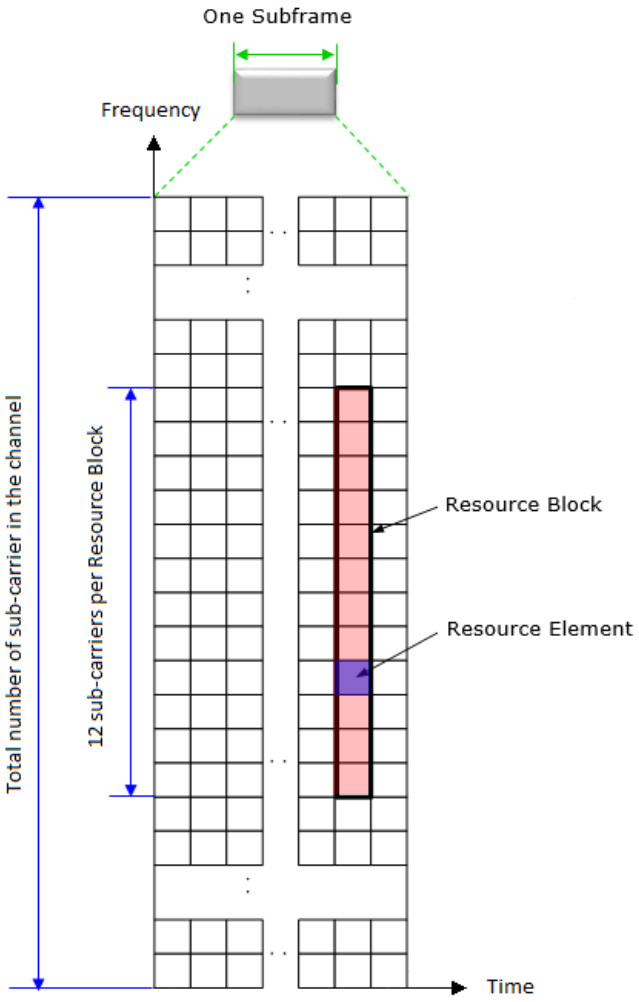

Figure 1-3: Representation of the resource grid

1.1.1.2 mMTC overview

The purpose of mMTC is to provide a long range ultra-low power low data rate wireless link with a simultaneous connection density of one million devices per square kilometers or more. These objectives pose two challenges, to connect all these devices at the same time, and to do it with a high level of power efficiency. To address these challenges several new technologies are envisioned. The focus here will be only on the two major ones.

The challenge on connection density is linked to the RB size. Based on the following note a simple back of the envelope calculation can be made:

A resource block always contains the same number of symbols, the 14 symbols of a slot over 12 sub-carriers for a total of 168 symbols. For larger subcarrier spacing, one RB will occupy a larger bandwidth over a smaller duration such that its surface in the time-frequency plane is constant. Hence the total number of available RBs is independent of the SCS, so only the case of the smallest SCS of 15kHz needs to be considered in the following argument.

Considering an RB bandwidth of 180kHz (12 sub-carriers of 15kHz) over a 1ms slot, then a 1GHz of bandwidth, for the duration of a radio frame, contains around 56000 RBs. This can cover an area of $0.056km^2$ with the target one million devices per square kilometer connection density. It is equivalent to a cell radius of about 135m. It is clear here that, if the minimum payload size remains an RB, the desired user density or the long range targeted by mMTC cannot be reached. It would also be very spectrally inefficient since, for applications such as remote sensor reading, it is expected to transmit only few bits of information at a time. The numbers provided here only give the order of magnitude, but show that 5G's radio frame structure, regardless of its high flexibility, cannot accommodate mMTC efficiently.

One intuitive solution would be to reduce the RB size in order to have more of them available. But even reducing the RB to a single symbol would only allow a cell of 1.7km and would leave no room for eMBB and URLLC.

The envisioned solution is to use Non-Orthogonal Multiple Access (NOMA) [1-1] which allows for the BS to connect multiple devices using the same RB. The details on how this is achieved will not be explained here but in practice it is equivalent to build artificial flexibility on the RB size, hence creating enough of them to address all the devices.

The challenge on power efficiency partly comes from the digital processing. To be able to communicate properly, the data must be encoded before transmission. It is said that redundancy is added. This allows to recover some eventual transmission errors, ensuring proper communication. This will be studied more deeply into the next chapter. Current coding technologies such as turbo codes or Low-Density Parity Check (LDPC) allow to reach near optimal channel capacity but at the cost of relatively high digital processing. This digital processing capacity is not expected to be present on devices that aims at ten years' lifetime over a small battery. This will be solved by using a new code family called polar coding [1-2]. Under certain circumstances compatible with mMTC link requirements, these codes allow to reach optimal channel capacity at a much lower digital processing cost.

These two main technologies, together with the flexible radio frame structure, are the principal enablers for mMTC as envisioned for 5G

1.1.1.3 URLLC

The purpose of URLLC is exactly as its name suggests, to provide an ultra-reliable low latency communication link with a moderate data rate. In more practical terms the target is to achieve a latency of one millisecond for the over the air interface with a BLock Error Rate (BLER) blow 10^{-6}. In comparison the target BLER in 4G is 0.1.

To achieve these performances, the system physical layer has been optimized and the radio frame structure high flexibility allows the possibility to prioritize URLLC traffic over mMTC and eMBB. As for mMTC the details on how it is done will not be discussed here.

1.1.2 eMBB: Foreseen technologies

The challenge for eMBB is to provide not only an improved peak rate but also an improved average rate and user density. These last two metrics can be aggregated into the area throughput expressed in bits per second per square meter. Because the user density target is much lower than that of mMTC, around ten thousand users per square kilometer, the flexibility of the radio frame structure is enough to

ensure the user density as long as the area throughput is up to the targeted specification. Hence, the focus can only be on peak rate and area throughput.

To reach the desired performances, the main bottle neck is the lake of available spectrum. Two solutions are envisioned to solve this problem. The first one is to allocate more spectrum for 5G and the second is to exploit a new source of diversity, the space dimension.

The additions of new bands are split into two categories, the sub-6GHz one and the millimeter wave one. The first one will lead to deployments that can be seen as an evolution of the 4G network. The second one will require a more disruptive solution since classical approach would be highly inefficient.

As its name suggests, the sub-6GHz category will englobe the bands below 6GHz. Strictly speaking millimeter waves correspond to wavelength between 1mm and 10mm. This corresponds, in the frequency domain, to the bands between 30GHz and 300GHz, assuming the speed of light in the vacuum is $3 \times 10^8 m/s$. In the context the of 5G, millimeter waves refer to frequencies between 24GHz and 100GHz. For the rest of this manuscript, unless explicitly said otherwise, the 5G definition of millimeter waves will be assumed.

Spatial diversity will be exploited in both frequency categories but with potentially different approaches. In the sub-6GHz range, it will be done through a technique called Multiple Users Massive MIMO (MU Massive MIMO), generally referred to as Massive MIMO. In the millimeter waves part of the spectrum, it will most likely be exploited through a different technique called beamforming, even though it is not entirely clear yet, as Massive MIMO could also be a solution. The discussion about the pros and cons of each technique will be a part of this manuscript.

1.1.2.1 Massive MIMO in sub-6GHz bands for large coverage

Sub-6GHz frequencies generally share the same good propagation property, which make them good candidates for wide area coverage. But they are not the best fit when it comes to peak rate since it simply requires a large amount of bandwidth which is not available at these lower frequencies.

Even though this type of base station is not targeting the peak rate, providing the average rate over a large area requires a lot of system throughput. The larger the area the more users must be served and the higher the system throughput must be. To cover large areas, using a similar technology as 4G, would require significantly more spectrum than available in the sub-6GHz spectrum.

To alleviate this limitation, 5G proposes to exploit the spatial dimension. When looking at the resource grid in Figure 1-3, this means add a third dimension to it. To some extend this is already partly used in previous generation by sectoring the cells, and in 4G by the use of MIMO. In the first case the cell is generally split in three or four sectors each of them covered by a separate antenna that have the appropriate level of directivity. This allows, in theory, to use the whole available spectrum in each sector. In practice there are some limitations mainly because the link between the base station and the user is not in Line of Sight (LoS) but rely on reflections and diffraction of the propagating waves. Even with the appropriate directivity, the radiated signals from one sector may pollute the neighboring ones, limiting the spectral reuse.

In 4G the possibility to use MIMO was introduced. In the general sense, the MIMO theory describes communication system with multiple inputs and outputs. In the case of 4G, it corresponds to the case where both the base station and the User Equipment (UE) have multiple antennas. For example, in down link, the base station will send four different signals using four different antennas. All four signals are simultaneous and at the same frequency, i.e. they use the same time-frequency resource. The UE on its side, also have four antennas and will receive a superposition of the four signals on each of its antennas. Depending on what is called the spatial richness of the Radio Frequency (RF) channel, each UE antenna will see a different superposition of the four original signals. These variations in the superposition

comes from the fact that each signal will reach each antenna though different paths, undergoing different delays and attenuations. If these different delays and attenuations, called the Channel State Information (CSI), are known, the original signals can be recovered by solving a four-equation system with four unknowns.

In theory it could allow to increase the data rate by a factor up to four. In practice channels that are not spatially rich enough will lead to under constrained equation systems and a data rate increase lower than four. 4G has provision to support MIMO communication up to 8×8 antennas but in practice the spatial richness rarely exceeds a data rate increase larger than three or four. This limitation comes from the antennas' vicinity on the UE and on the base station leading to highly correlated channels between them.

The MU massive MIMO envisioned for 5G is significantly different. It aims at alleviating the inter-sector pollution as well as the limited richness of the MIMO channel in two ways.

Taking the same example as above but where the four antennas on the UE actually belong to four different devices at four different locations in the cell. They are now physically separated offering a much better spatial diversity. But to reconstruct the signals each UE need all four antenna signals but can access only its own. The same receiving strategy cannot be used here but this can be corrected by looking at the problem from a different angle.

If knowing the CSI and having all four antenna signals allows to recover the original signals, maybe it is possible to reverse the process. If the base station knows the CSI it can emit a specific superposition on its own antennas such that the UEs only receive the desired signal. Indeed, this is possible and is known as Multiple User MIMO (MU-MIMO). Using this approach brings two benefits, it limits the inter-sector pollution, and it provides spatial rich MIMO channels for a much larger performance increase of the cell area throughput compared to 4G MIMO.

It also comes with new challenges. One of them is the processing complexity which increase exponentially with the number of users. That is where the "Massive" from Massive MIMO comes into play. In this context it means that one side has many more antennas than the other one. In 5G, the side with more antennas will be the base station, with potentially hundreds of them, forming an antenna array. When in that configuration it can be shown that near optimal performances can be achieved with simplified algorithm of linear complexity [1-3].

With this last piece, the basic aspects of MU Massive MIMO, as it is envisioned in 5G, has been covered. It is an efficient way to exploit spatial diversity and significantly improves the system performance for coverage and average rate. The peak rate and user density will be solved by the introduction of beamforming in the millimeter wave spectrum.

1.1.2.2 Millimeter Waves Beamforming small cells for locally ultra-dense areas

The sub-6GHz spectrum is very efficient in covering wide areas with moderate data rates but is not sufficient to provide the desired peak rate or user density. To cover these KPIs the idea is to use the millimeter wave spectrum. At these frequencies, the bandwidth available is much larger and would allow to reach the target peak rate.

Millimeter wave were generally considered unfit for mobile communication due to their poor propagation properties. The idea here is to exploit this limited propagation by building small cells of some tens of meters in the specific areas where high user density is needed, typically in dense urban areas. These small cells will overlap with the macro cells forming a heterogeneous network and providing everywhere connectivity as well as the desired user density and peak rate.

From afar millimeter waves base stations have a very similar architecture, compared to their sub-6GHz counter parts, using antenna arrays to spatially address multiple users. They different mainly by an

aspect brought by the smaller covered area. The users are almost always in a LoS configuration. The goal then becomes to form a beam pointing only the user at stake. This technique is in use since a long time in radar application and is called beamforming.

The purpose is to form a specific beam for each user, which can now benefit from the whole spectrum available in the beam. In other word, each beam can be seen as a parallel resource grid, multiplying the cell capacity by the number of beams.

It is through the combination these three aspects, wider bandwidth, small-cells, and multiple beamforming that 5G proposes to solve the challenge of user density and peak rate.

1.2 Focus of this work: Small cell receivers with large antenna arrays

This manuscript focuses on the challenges brought by multiple beamforming millimeter wave Small Base Stations (S-BS). In particular, it will discuss the constraints this approach imposes on the electronic hardware. Using large antenna arrays, with potentially hundreds of elements, requires very efficient transceivers, to be a viable option. This is an even higher challenge to do so in the millimeter wave domain and over very large bandwidth.

A transceiver is made of two parts, a transmitter and a receiver. Each of them has their own set of challenges. This work's focus only on the receiver end of it. The final goal is to propose a receiver that can unlock this technology and contribute to the delivery of 5G promises. To this end, the manuscript will be organized as follow:

In Chapter 2, a system analysis for a system targeting a bandwidth of 1GHz around a carrier frequency at 28GHz will be proposed. The outcome of this analysis will be two-fold. First, is to understand the basic interactions between all the parts of the system. Second, is to provide a coarse sizing of the system to ensure, if possible, user density and peak rate KPIs.

Chapter 3 will use chapter 2 results to evaluate the constraints imposed on the electronics. This will allow to derive a receiver specification. From that point, the feasibility of such a receiver will be established through an extensive analysis of the state of the art of receivers' building blocks with a classic architecture.

Chapter 4 will investigate further the receiver's architecture and propose an innovative approach centered around RF sampling sigma-delta modulators.

Finally, Chapter 5 will detail the proposed implementation and Chapter 6 will summarize and conclude this work.

1.3 References

[1-1] M. Vaezi, R. Schober, Z. Ding and H. V. Poor, "Non-Orthogonal Multiple Access: Common Myths and Critical Questions," in IEEE Wireless Communications, vol. 26, no. 5, pp. 174-180, October 2019, doi: 10.1109/MWC.2019.

[1-2] E. Arikan, "Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels," in IEEE Transactions on Information Theory, vol. 55, no. 7, pp. 3051-3073, July 2009.

[1-3] H. Q. Ngo, E. G. Larsson and T. L. Marzetta, "Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems," in IEEE Transactions on Communications, vol. 61, no. 4, pp. 1436-1449, April 2013.

2 CHAPTER II: SYSTEM ANALYSIS

Modern communication systems are of an extremely high level of complexity. With the addition of beamforming and millimeter wave carriers, 5G is increasing it even more. This creates the need for new system analysis approaches adjusted to these cases. In order to limit the complexity of the analysis, it is possible to adjust classical methods for a specific type of networks. One such method will be proposed in this chapter, that is adapted to mobile network using millimeter wave beamforming BS and deployed in a small cell arrangement.

The usual approach to system analysis is first builds a model with free parameters such as the transmitter (Tx) output power, receiver's (Rx) noise performances or the Tx to Rx distance. Then this model is used to study the system's sensitivity to each of these parameters. This gives insight on the system behavior and allow for performances optimization. Unfortunately, with the increasing system complexity, design parameters are generally not free, but interdependent. For example, as it will be seen, the antenna array number of elements affects the link budget and the spatial multiplexing capability of the system. It is often possible to make linear combinations of the primary interdependent parameters to create a set of secondary free parameters. But these are generally not as insightful as the primary ones in term of design parameters.

Instead, a more practical approach is proposed. Since the target performances are already set by 5G's KPI, the choice was made to adopt a reverse strategy. Instead of optimizing the system performances, a model is built based on design parameters and used to look at the different configurations providing the desired performances. Then, the one that has the lowest power consumption is selected. The power consumption will be mostly evaluated through two parameters, Power Amplifier (PA) output power and digital processing complexity. This method does not guaranty optimality but allows for a better understanding of the design parameters impacts.

The discussion on the network architecture will be based on two pillars, information theory and beamforming theory. These theories will be covered in the first two sections. Once the architecture is set, this will allow to establish the link budget that will be used to size the system such that it reaches 5G's KPI. Finally, the impact of such a system will be evaluated in the case of a multiple operator deployment, in particular on the characteristics of the interfering signals.

2.1 INFORMATION THEORY

Information theory is the basis for all modern digital communications. It was mostly laid down by Claud Shannon in his 1948 reference papers [2-1] and [2-2]. In the following lines, after introducing the fundamental results of this theory, they will be used in order to get some insight of the effects of several parameters on wireless links.

2.1.1 Noisy-channel coding theorem

The original article, "A Mathematical Theory of Communication" tackles a much wider range of challenges, namely information source modeling, data compression and channel coding. The focus will be here on the later.

Let us assume a binary symmetric noisy channel, i.e. a channel that can carry symbols of information which are bits and has a probability p_{flip} of flipping them when passing through it (Figure 2-1).

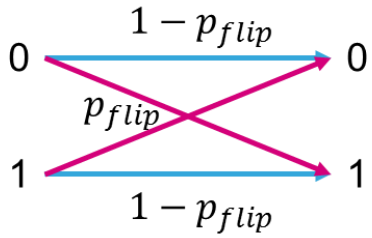

Figure 2-1: Binary symmetric noisy channel

If no precaution is taken it is obvious that the received data will be corrupted. It is possible to devise some simple methods to reduce the probability of error. For example, each symbol may be transmitted three times and a majority vote may be used for decoding. This scheme would clearly reduce the probability of error, at the cost of reducing the transmission rate. The new probability of error would be $p_{new} = p_{flip}^3 + 3 \times p_{flip}^2 \times (1 - p_{flip})$ corresponding to the cumulated probabilities of having two or more bits flipped out of three during the transmission. For the case where $p_{flip} = 0.1$ the new probability of error goes down to $p_{new} = 0.028$ while obviously the data rate is divided by three.

Intuition would suggest here that when the probability of error is driven down, the transmission rate would go to zero, which is certainly the case for this simple coding scheme. The main result from Shannon's publication is that for any noisy channel it exists a way of encoding and decoding the data, i.e. adding and removing redundancy, such that the error rate can be arbitrarily small while maintaining a strictly positive channel capacity. This channel capacity is generally referred to as Shannon's capacity or Shannon's limit of the channel and is denoted C.

In the case of the binary symmetric noisy channel this capacity would be:

$$C = 1 - \left(p_{flip} \times \log_2(p_{flip}) + (1 - p_{flip}) \times \log_2(1 - p_{flip})\right) \quad (2.1)$$

For the case where $p_{flip} = 0.1$ then $C = 0.53$ bits of information per channel use. This means it exist a way of encoding the bits to be transmitted such that the data rate is only about halved, and the error rate can be set arbitrarily close to zero. This is a far better result compared to simple coding scheme described earlier. While the origin of this formula and the existence proof of such a code is beyond the scope of this manuscript, the interested reader can go to [2-3] for in-depth explanations. As mentioned by Shannon himself in his original paper, this theorem suffers from the same drawback as most existence theorems. While it proves the existence of such codes, it does not provide a way of building them. Thankfully in the 70 years that have passed since his publication, codes such as Turbo-code [2-4] or Low Density Parity Check (LDPC) [2-5] nearly reaching Shannon's limit have been found. For simplicity it will be assumed that the system to be built will be close enough to Shannon's limit, so it can be used as a first order approximation for the channel capacity.

2.1.2 Shannon-Hartley theorem

The Noisy-channel coding theorem just presented is a general result. While the binary symmetric noisy channel was a good example for didactic purposes it is not especially useful to the case of modern wireless communication. Most systems will modulate the information around a carrier frequency within a definite bandwidth. It is this piece of spectrum that is called the RF channel. Its accurate modeling can be quite complex. To keep things simple, the simplest model known as the Additive White Gaussian Noise (AWGN) Channel model will be used. The impact of more accurate channel models will be discussed later, once the rest of the system is built up. The AWGN channel consists of a flat frequency response channel adding white noise to the signal and where the amplitude distribution of the noise follows a Gaussian distribution. In these conditions the Shannon-Hartley theorem [2-2] states that the

channel capacity only depends on the channel bandwidth B, the signal power S and the noise power N according to the following equation:

$$C = B \times \log_2\left(1 + \frac{S}{N}\right) \tag{2.2}$$

Where B is expressed in Hertz, S and N in Watts and C in bits per second. This is a very convenient formulation since it brings back the problem of capacity estimation to an evaluation of bandwidth and Signal to Noise Ratio (SNR) which are common figures of merit for RF devices.

2.1.3 Signal power versus Bandwidth

One of the major goals of 5G is to improve significantly the energy efficiency per bit in order to keep the network power consumption to a level comparable to the one of the 4G network while providing an increase in network performance of two orders of magnitude or more. One major contributor to the power consumption is the transmitter Power Amplifier (PA) consumption, that is closely linked with the required signal power at the receiver's end for proper communication. What the "receiver's end" means will be detailed later, during the receiver's architecture description. Reducing the required signal power would have a second positive effect by reducing the level of interfering signals in a multi-user multi-operator environment.

The purpose here, using the Shannon- Hartley theorem, is to gain insight on the interaction between bandwidth and signal power for a given data rate. Based on equation (2.2) the signal power S at the receiver's end as a function C, B, and N can be expressed as:

$$S = N \times \left(2^{\frac{C}{B}} - 1\right) \tag{2.3}$$

In the case of an AWGN channel the noise has a constant Power Spectral Density (PSD) within the channel bandwidth and its total in band power is simply the product of its PSD N_0 by the channel bandwidth B. Let us suppose the noise here is the antenna thermal noise at the temperature $T = 290K$, then $N_0 = k_b \times T$, where k_b is the Boltzmann constant. Equation (2.3) can then be rewritten as:

$$S = k_b \times T \times B \times \left(2^{\frac{C}{B}} - 1\right) \tag{2.4}$$

Figure 2-2 plot the received signal power in dBm and SNR in dB for a channel capacity of $1Gbit/s$ for a bandwidth from $100MHz$ to $2GHz$. Considering first the received signal power, represented by the blue curve, it can be seen that increasing the bandwidth decreases the signal power. But this decrease goes with diminishing return when increasing further the channel width. Starting from $100MHz$, doubling the bandwidth once reduce the signal power by $12.2dB$. Doubling it another time, for $B = 400MHz$, the reduction drops to $5.2dB$. Doubling it one more time, for $B = 800MHz$, the reduction falls to $2.3dB$. Due to this diminishing return and the potential reduction in electronics efficiency when handling wider bandwidth, widening the channel width is only desirable up to a certain point. Another drawback when increasing the bandwidth is that it also requires more available spectrum for the whole system, which is a scarce resource in the sub-6GHz range used by current mobile networks. This is one of the reasons a part of 5G is focusing on millimeter wave spectrum where more bandwidth is available.

Considering now the received SNR, represented by the red curve, one can note the same behavior. The same frequency doubling leads to an SNR reduction of $15.2dB$, $8.2dB$ and $5.3dB$ respectively. Again, wider bandwidth brings diminishing return. It can be seen here that it also has the potential to relax the requirements on noise performances on the receiver's hardware. It may be interesting to use larger bandwidth them what was first suggested by the signal power analysis alone.

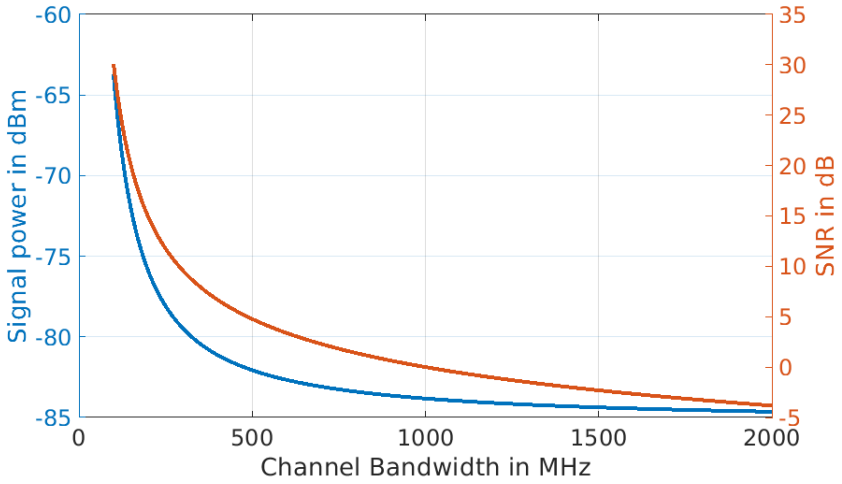

Figure 2-2: Signal power and SNR versus channel bandwidth for a constant channel capacity of 1Gbit/s

Reducing the signal power at the user end mean reducing the PA output power. Since the PA is one of the major power consumption contributors in RF systems, increasing the signal bandwidth can be seen as a way to reduce the system overall power consumption.

2.1.4 Conclusion

This initial analysis, based on the basic principle of Information Theory, allows to reach three main conclusions.

The first one is that Shannon's capacity can be used as a first order approximation of a channel capacity thanks to the advent of efficient codes such as Turbo-code or LDPC.

The second one is that the problem of reliable communication can be brought back to a problem of signal to noise ratio on an RF channel, allowing to translate a problem of stochastic nature into a design one.

The third one is that increasing the RF channel bandwidth has the potential to reduce the system overall power consumption by reducing the output power required for Power Amplifiers, a major contributor of current mobile network power consumption. This has also the potential to relax the required dynamic range and noise performances of the receiver's hardware thanks to reduced interference levels and required SNR. These benefits come at the cost of increased requirement on available spectrum and hardware bandwidth. While finding the optimal tradeoff on that matter is a question with no definite answer, the next sections and chapters will give hints on where it could lie.

2.2 BEAMFORMING

At its heart, beamforming consists in using multiple antennas in order to send or receive a given signal to or from the desired direction. Interestingly this idea is almost as old as radio itself, i.e. the use of electromagnetic wave for wireless communication. Electromagnetism laws were unified in 1865 by James Clerk Maxwell, and experimentally confirmed by Heinrich Hertz in the 1880s, followed by a series of publications in 1887. Practical implementations of radio transmission were soon realized by Guglielmo Marconi, among others, in the mid 1890'. The first reported use of antenna array was in 1901 [2-6], also by Marconi for the first Transatlantic transmission. Less than a decade after the first radio was built, beamforming was already seeing some practical use. The fact that this technology has stimulated intensive research throughout the 20th century and is still at the heart of future mobile networks gives an idea of how rich the subject is.

In this section, an overview of the main topics that beamforming is made of will be presented. It will go through the basic concepts and solve the most classic problems in order to build some intuition in the mechanics of beamforming. From there, based on a solid state of the art, the different implementations that have been proposed for millimeter wave beamforming systems will be discussed. There are three of them called Analog Beamforming (ABF), Hybrid Beamforming (HBF) and Digital Beamforming (DBF). Finally, these three implementations will be detailed. This will give significant incentives to investigate further one of these three solutions.

2.2.1 Basic principle

The main idea is to combine the signals of all antennas to increase the overall received power. Let us start with the simple two antennas system depicted in Figure 2-3-a. Let us consider the case where this system is used to receive signals. Since an antenna is a purely passive device, transmitting signals is a completely symmetrical problem, so all of the following results hold in the case of transmission.

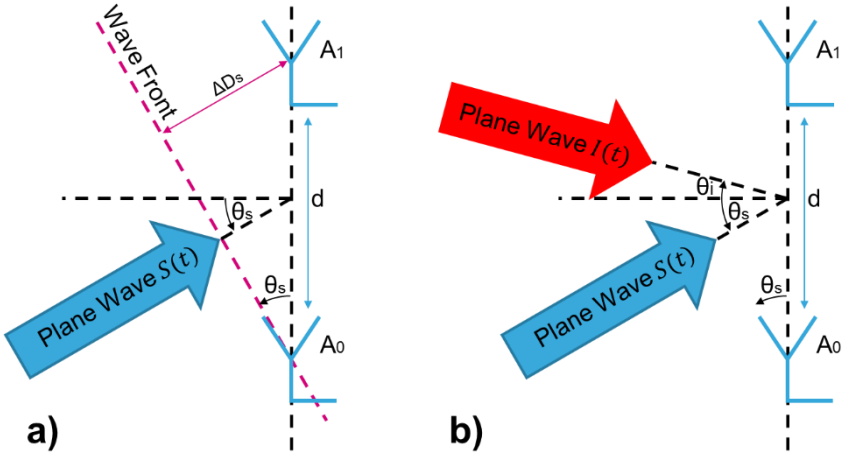

Figure 2-3: Two elements antenna array with a) One incoming plane wave b) Two incoming plane waves

Let us consider a plane wave incoming on the antenna array with an Angle of Arrival (AoA) θ_S, the angle formed by the incoming wave propagation direction and the normal to the antenna array. This wave will not reach the different antennas at the same time. The signal received by the antenna A_1 will be delayed compared to that of antenna A_0. The value of this delay Δt is coming from the plane wave traveling the extra distance ΔD_S at the speed of light c and depends on the AoA θ_S and the antenna spacing d:

$$\Delta t = \frac{\Delta D_S}{c} = \frac{d \times \sin(\theta_S)}{c} \tag{2.5}$$

To recombine the antenna signals in order to maximize the received power from the plane wave, this delay needs to be compensated for. Once they are properly aligned, they are summed up in a coherent manner, therefore, the amplitude of the recombined signal grows linearly with the number of antennas. As a consequence, the signal power grows in a quadratic way.

In the transmit case this means that a single antenna system, driven by a single PA, can be replaced by an N_{ant} antenna system where each antenna is driven by a PA whose output power is divided by N_{ant}^2. This allows to reduce significantly the total PA power consumption in a system using a large number of antennas. This may sound unphysical at first. The origin of this power reduction comes from the fact that an antenna array increases the directivity compared to a single antenna system, meaning that overall

radiated power is lower and more focused in a beam toward the desired direction, thus the naming of this technique: Beamforming.

In the receive case, as shown in section 2.1, the interest is more on the received SNR. Making the reasonable assumption that the thermal noise added by each antenna is uncorrelated, then the resulting noise power after beamforming will grow linearly with the number of antennas. Since the signal power grows in a quadratic way this means that the SNR grows linearly with the number of antennas:

$$SNR_{BF} = SNR_{ant} \times N_{ant} \Leftrightarrow SNR_{BF_{dB}} = SNR_{ant_{dB}} + 10 \times \log_{10}(N_{ant}) \quad (2.6)$$

Where SNR_{ant} is the SNR for a single antenna and SNR_{BF} is the resulting SNR after beamforming. It is important to remember the noise un-correlation condition that will be used as a design constraint later in order to ensure the system performances.

Let us now consider the case of Figure 2-3-b where a second incoming signal $I(t)$ with a different AoA θ_i is added. This signal will be an interferer for $S(t)$. Let us assume the antennas are tuned to receive signals around $f_c = 28GHz$, are isotropic, and have for example a spacing $d = \lambda_c/2$, where $\lambda_c = c/f_c$ is the carrier wavelength in the void. Let us take $S(t)$ and $I(t)$ to be sinewaves at frequencies $f_S = 26.6GHz$ and $f_I = 29.4GHz$ respectively, corresponding to Graph 1 and 2 in Figure 2-4. Graphs 3 and 4 represent the received signals $S_{A0}(t)$ and $S_{A1}(t)$ at the corresponding antennas when the AoA are $\theta_S = 45°$ and $\theta_i = -15°$. One can see large interferences between the two signals at each antenna. Graph 5 plots the original signal $S(t)$ and the beamformed signal when $S_{A0}(t)$ and $S_{A1}(t)$ are delayed and summed to maximize the power received from the AoA θ_S.

The first noticeable effect is the amplitude increase. As expected, it is equal to the number of antennas; two in this case. Second the interference caused by $I(t)$ has almost disappeared. This is called spatial filtering, i.e. the ability of a system to receive selectively signals from one direction while filtering out the interferers from other directions.

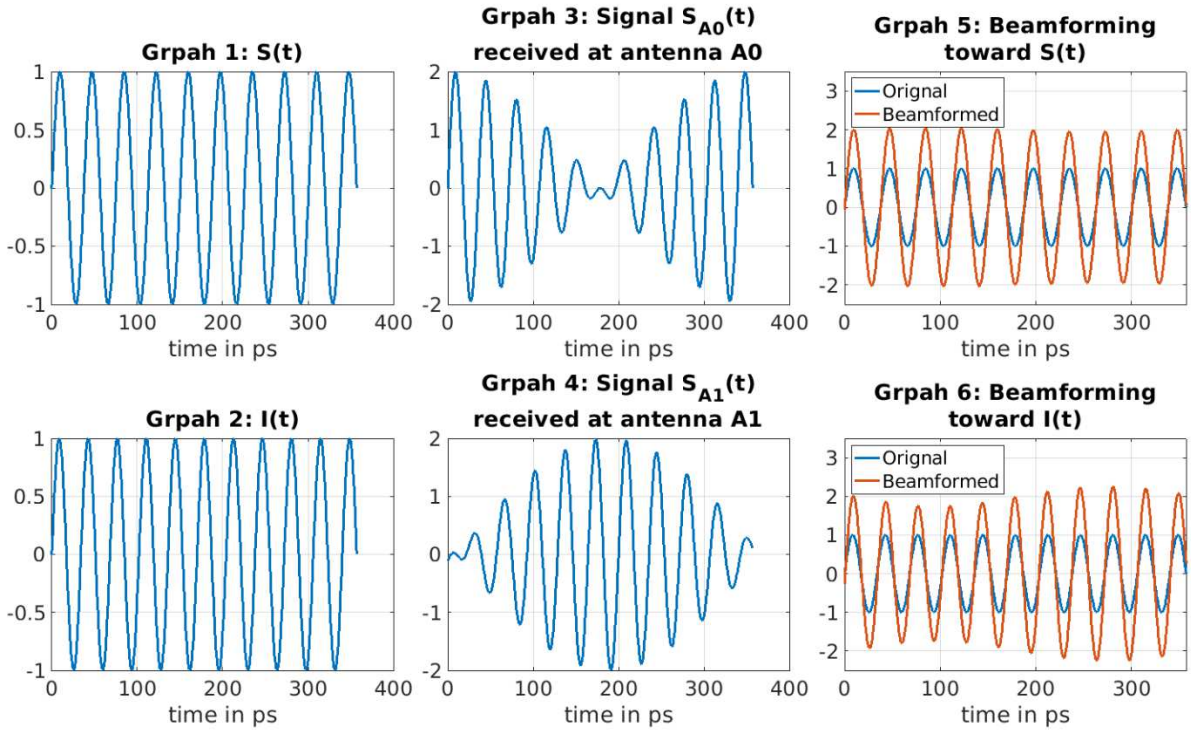

Figure 2-4: Plots of incoming plane waves, antenna received signals and beamformed signals

Finally, a different delay and sum recombination can be made in order to maximize the power received from $I(t)$ (Grpah 6). Again, the same expected amplitude increase is displayed, and a significant portion of the interference caused by $S(t)$ has been removed. This means that, if properly used, a single antenna array can receive multiple signals from different AoA. This is true even if the signals are using the same time-frequency resource. This example is finely tuned in order to show the desired effect but generally a two antennas array will not have such good spatial filtering capacity. While there are many more subtleties to it, spatial filtering will generally improve when more antennas are used.

To get more insight on the filtering capability of a multi antenna system, it is useful to look at its radiation pattern. Let us start with the case of aligned antennas which are called linear arrays. Assuming delays are applied on the antenna signals to receive plane wave from an AoA θ_{BF}, one can look how much signal coming from a different AoA θ would appear at the output. Plotting this for all AoA, using polar coordinates, will give the radiation pattern of the linear antenna array.

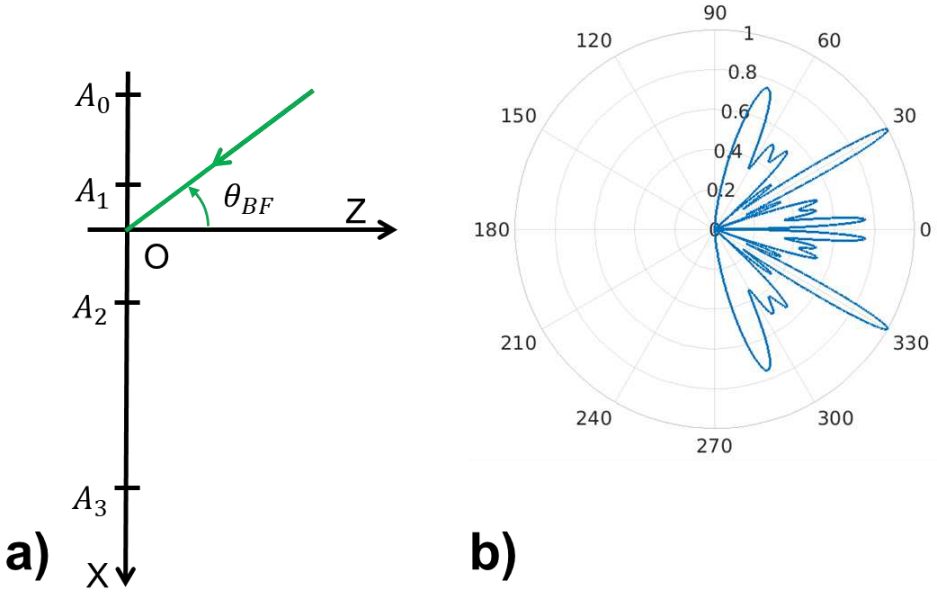

Figure 2-5: a) Linear antenna array b) Corresponding radiation pattern

Let us assume the setup of Figure 2-5-a where incoming signals are sinewaves at center frequency f_c. The positions of the antennas are $(-3\lambda_c\,;\,-\lambda_c\,;2\lambda_c\,;6\lambda_c)$. Expressing the antenna position in unit of $\lambda_c = c/f_c$ makes the results true for any frequency f_c and thus more general. Figure 2-5-b represent the normalized radiation pattern of the array for a beamforming angle $\theta_{BF} = \pi/6$. As expected, the output amplitude is maximum for waves with AoA θ_{BF}.

Also, it can be seen that it varies widely for other AoA. In particular, in this example, there is another AoA at $\theta = -\pi/6$ where the beamformed amplitude is maximum. This means that waves coming from that direction will not be filtered at all which is highly undesirable for the targeted application. One sufficient condition to prevent such a phenomenon is to have the spacing of adjacent antennas below $\lambda_c/2$ [2-7 Ch. 22.6]. The next section will focus on such antenna arrays.

2.2.2 Beamforming algorithm

In the section above, the simplest and the most natural way to perform beamforming was implicitly used, i.e. compensating the propagation delays to sum up coherently the antenna signals and receive waves coming from a given direction. More generally beamforming processing consist in delay, weight and sum operations. A beamforming algorithm is simply the method used to choose the different delays and weights to be applied to each antenna signal before summation.

There are two main approaches to this problem. The first one is to choose the delays and weights to physically shape the beam in the desired direction while minimizing the side lobes. This is what is typically used in radar systems where the beam scans multiple directions, searching for an object echo. This approach only requires knowing the AoA of arrival of the different users. This set of AoA will represent the Channel State Information (CSI). In this manuscript, this approach will be referred to as beam-shaping.

The second approach is Massive MIMO, where the beamforming coefficients to be applied are based on the evaluation of the channel propagation properties. In that case it is these propagation properties that forms the CSI. In the case of a Line of Sight (LoS) configuration the resulting radiation pattern will also result into a main beam toward the target user, hence it can also be seen as a beamforming algorithm.

In this section, a general description of these algorithms' principles will be made and the most common implementations they have seen will be given. For the sake of simplicity, beam-shaping will be studied using linear and evenly spaced antenna arrays called Uniform Linear Arrays (ULA). In Massive MIMO, the array topology is captured by the CSI and its mathematical formalism does not require to make any assumption on it.

One important aspect of these studies will be on the processing complexity of these algorithms. If only uplink is considered, the processing complexity only impact the system power efficiency and latency. The strongest constraint on processing complexity actually comes from downlink. The foreseen solution is to exploit the reciprocity property of the channel, i.e. it has the same propagation properties in both directions, to reuse the CSI estimated during uplink in downlink. This imposes for the uplink CSI estimation to be done before processing the downlink. This could be addressed by extending the uplink duration in Time Division Duplexing (TDD). Unfortunately, the communication is almost always downlink favored, which means there are little incentives to increase uplink time. Another constraint comes from the CSI validity duration. If the estimation takes too long, by the time the system is ready for downlink the CSI will not be valid anymore, making this approach simply unviable.

From these studies, the strong points and weaknesses of the different approaches will be emphasized. This will allow to get some insight on the tradeoffs between system complexity and performances. Based on this understanding, the foundations for three different beamforming receiver architectures will be laid down. These architectures will then be studied in greater details in subsequent sections.

For now, only the narrow band case where a delay is equivalent to a phase shift will be considered. It means that applying a real coefficient and a delay can be implemented by a single complex coefficient. The implication of using wide band signals in such systems will be studied later.

2.2.2.1 Uniform Linear Array (ULA) under Delay and Sum (DS) processing

As described earlier a ULA is an antenna array where antennas are uniformly spaced along a line. Here, only the case where this spacing is equal to half the wavelength λ_c in the air of the center frequency of interest will be considered. Let us assume a ULA with N_{ant} antennas, named A_0 to $A_{N_{ant}-1}$, and an incident cosine wave $S(\theta, f)$ incoming with angle θ and frequency f. Note that, for the case treated here, the signal frequency f does not necessarily equals the center frequency f_c. This gives a more general result. Let us place the antennas along the X axis, with A_0 at the origin and such that other antennas abscises are positives. The antenna signal S_{A_n} is delayed by Δt_n with respect to S_{A_0}. The delay Δt_n is then obtain using equation (2.5) with $d = \lambda_c/2$:

$$\Delta t_n = \frac{n \times \frac{\lambda_c}{2} \times \sin(\theta)}{c} = \frac{n \times \sin(\theta)}{2 \times f_c} \tag{2.7}$$

The signal received by antenna A_n can then be expressed by:

$$S_{A_n} = \cos\big(2 \times \pi \times f \times (t - \Delta t_n)\big) \tag{2.8}$$

To form a beam in the direction θ_{BF}, the delays Δt_{BF_n} for this angle must be compensated:

$$\Delta t_{BF_n} = \frac{n \times \sin(\theta_{BF})}{2 \times f_c} \tag{2.9}$$

One must not confuse the incoming wave AoA θ and the beamforming angle θ_{BF}, which is the angle where the beam is steered, and can be different from θ. Using (2.9) Leads to equation (2.10) for the output signal S_{BF} after beamforming:

$$\begin{aligned} S_{BF}(N_{ant}, \theta_{BF}, \theta, t, f) &= \sum_{n=0}^{N_{ant}-1} \cos\Big(2 \times \pi \times f \times \big(t - \Delta t_n + \Delta t_{BF_n}\big)\Big) \\ &= \sum_{n=0}^{N_{ant}-1} \cos\left(2 \times \pi \times f \times \left(t - n \times \frac{\sin(\theta) - \sin(\theta_{BF})}{2 \times f_c}\right)\right) \end{aligned} \tag{2.10}$$

After some more manipulation (see Annex 2.1) (2.10) can be reformulated as follow:

$$\begin{aligned} &S_{BF}(N_{ant}, \theta_{BF}, \theta, t, f) \\ &\quad = G_{sp}(N_{ant}, \theta_{BF}, \theta) \times \cos\Big(2 \times \pi \times f \times \big(t - \Delta \mathrm{t}(N_{ant}, \theta_{BF}, \theta)\big)\Big) \end{aligned} \tag{2.11}$$

With the delay

$$\Delta \mathrm{t}(N_{ant}, \theta_{BF}, \theta) = \frac{(N_{ant} - 1)}{4 \times f_c} \times (\sin(\theta) - \sin(\theta_{BF})) \tag{2.12}$$

And

$$G_{sp}(N_{ant}, \theta_{BF}, \theta, f) = \frac{\sin\left(N_{ant} \times \frac{\pi}{2} \times \frac{f}{f_c} \times (\sin(\theta) - \sin(\theta_{BF}))\right)}{\sin\left(\frac{\pi}{2} \times \frac{f}{f_c} \times (\sin(\theta) - \sin(\theta_{BF}))\right)} \tag{2.13}$$

Where G_{sp} is the spatial transfer function or Array Factor (AF). Plotting G_{sp} in polar coordinates versus θ, while N_{ant}, θ_{BF} and f are held constant, gives the array radiation pattern. Since G_{sp} and Δt are independent of time, it can be deducted from (2.11) that the output is a weighted and delayed version of the signal carried by the incoming wave. For a given AoA θ the delay Δt is independent of the signal frequency so it will not be the source of a frequency selective behavior of the system response. It is generally not the case for G_{sp} , depending explicitly on f.

For $\theta = \theta_{BF}$, $G_{sp}(N_{ant}, \theta_{BF}, \theta)$ is undefined, but its extension by continuity at this point is N_{ant}, which is the expected gain in the main beam. One can also see that for this AoA and this one only, the gain is independent of the input frequency, and will therefore have a flat frequency response in the main beam. Figure 2-6 plots the spatial transfer function for ULAs of 4, 8 and 32 antennas for an input signal at $f = f_c$, for $\theta_{BF} = -\pi/6, 0, \pi/4$.

All the spatial transfer functions plotted here present similar characteristics, a succession local extremums and zeros that goes to minus infinity on a logarithm scale. The part between two successive zeros is called a lobe. The one with the highest value is called the main lobe or main beam.

Two trends can be observed. The first one is that, when increasing the number of antennas, the main beam becomes narrower. The second one is that the number of zeros is nearly equal to the number of antennas. Increasing the array size, increases the number of zeros and improve the likelihood for an interferer coming from a random direction to fall near one of the zeros where the attenuation is high. In other words, an array with more antennas will generally have a better spatial filtering capability.

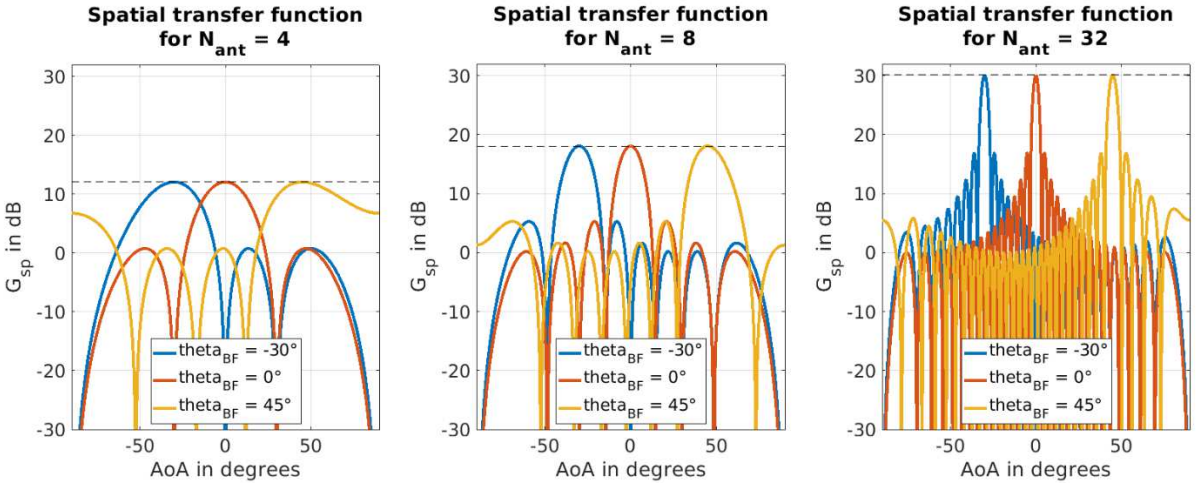

Figure 2-6: Spatial transfer function for ULAs with 4, 8 and 32 antennas for various beamforming angles

Spatial transfer functions are often evaluated by the use of two figures of merit, the Half Power Beam Width (HPBW) and the Peak Side Lobe (PSL). The HPBW is defined as the range of angles where the received power is greater than half the maximum power. It is analogous to the notion of bandwidth in the frequency domain. The PSL is the height of the highest lobe that is not the main one. Figure 2-7 plots the HPBW and the PSL as the number of antennas grows for a beamforming angle of $\theta_{BF} = 0$. The HPBW has a linear relationship on a log-log scale, i.e. it is divided by two when the number of antennas is doubled. The PSL, while improving with the number of antennas, reaches a plateau around -13.26dB.

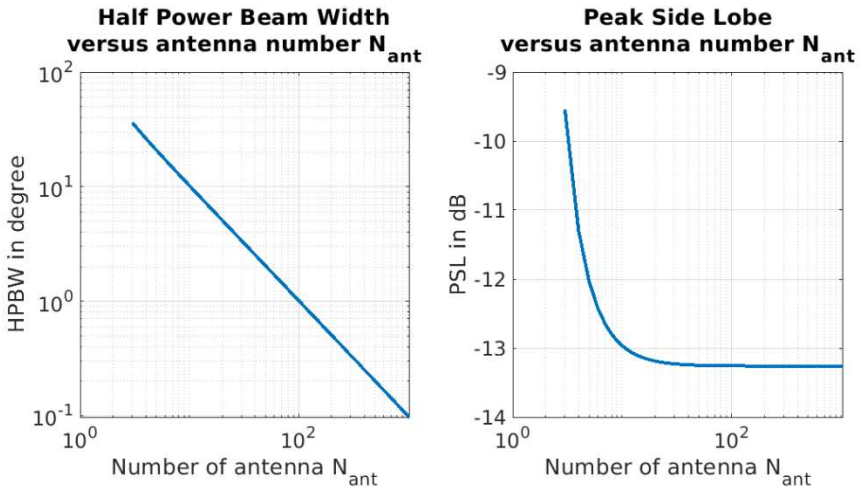

Figure 2-7: HPBW (left) and PSL (right) for ULAs versus the number of antennas N_{ant}

Figure 2-8 plots the HPBW and the PSL for the three ULA of Figure 2-6 for beamforming angles going from $-\pi/3$ to $\pi/3$. It can be seen that the HPBW not only depends on the number of antennas but also on the steering angle.

Also note that, when steering too much, the PSL start to degrade. Thankfully, the more antennas the later this happens. This means that when choosing the array size, the desired steering capability to ensure the target HPBW and PSL performances must be considered.

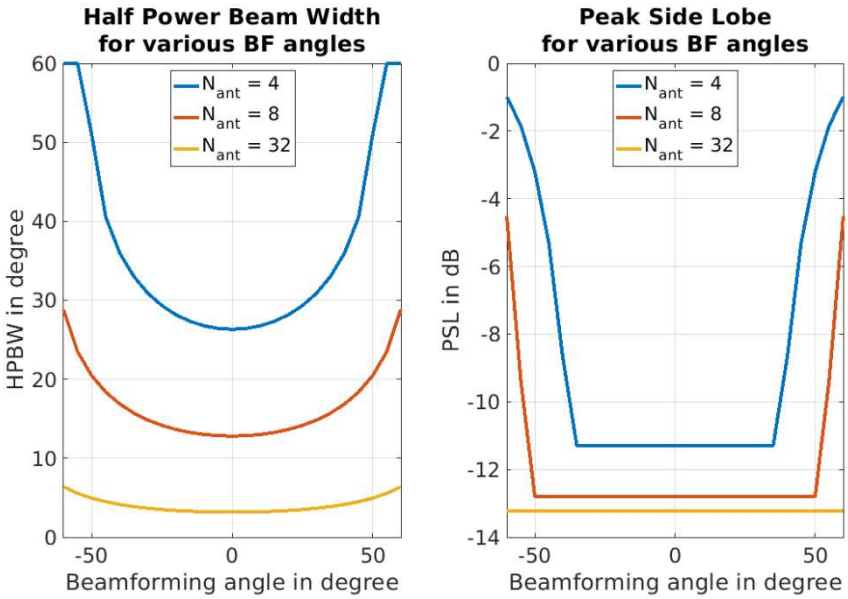

Figure 2-8: HPBW and PSL for ULA's with 4, 8 and 32 antennas for various beamforming angles

2.2.2.2 Beam-shaping

The purpose is to use non-unit weights on the antennas to modify the beamforming spatial transfer function. This is called the Delay Weight and Sum (DWS) processing, which englobes the delay and sum approach described earlier. To get some intuition on how it works let us use an example with the same ULA as described above.

A signal coming from an angle θ, while the delays are adjusted to receive a signal from AoA θ_{BF} is expressed by equation (2.10), recalled here for convenience:

$$\begin{aligned} S_{BF}(N_{ant}, \theta_{BF}, \theta) &= \sum_{n=0}^{N_{ant}-1} \cos\left(2 \times \pi \times f \times \left(t - \Delta t_n + \Delta t_{BF_n}\right)\right) \\ &= \sum_{n=0}^{N_{ant}-1} \cos\left(2 \times \pi \times f \times \left(t - n \times \frac{\sin(\theta) - \sin(\theta_{BF})}{2 \times f_c}\right)\right) \end{aligned} \tag{2.10}$$

Let us note $\delta t = \frac{\sin(\theta)-\sin(\theta_{BF})}{2\times f_c}$. The output signal S_{BF} is the sum of delayed versions of the incoming signal. This is exactly the same signal processing as a digital Finite Impulse Response (FIR) filter with a number of taps equal to the number of antennas N_{ant}, each tap being of unit coefficients and delay δt. The details of how equation (2.11) is derived from equation (2.10) is given in Annex 2.1. It is done by the introduction an imaginary part. This gives rise to a complex notation that is very common in digital signal processing. Using this complex notation and adding weighting factors a_n in (2.10) leads to (2.14):

$$S_{BF}(N_{ant},\theta_{BF},\theta) = \sum_{n=0}^{N_{ant}-1} a_n \times e^{i\times 2\times\pi\times f\times(t-n\times\delta t)} = e^{i\times 2\times\pi\times f\times t} \sum_{n=0}^{N_{ant}-1} a_n \times e^{-i\times 2\times\pi\times f\times n\times\delta t} \quad (2.14)$$

The complex notation allows to separate the signal from the array factor. To fully emphasis the parallel with FIR filters, noting $Z = e^{i\times 2\times\pi\times f\times\delta t}$, the array factor can be reformulated as:

$$G_{sp}(N_{ant},\theta_{BF},\theta) = \sum_{n=0}^{N_{ant}-1} a_n \times Z^{-n} \quad (2.15)$$

This is exactly the same form as the Z-Transform of a FIR filter. It is therefore possible to use the same mathematical tools to determine the coefficients to modify the array factor as desired ([2-8] [2-9]). Using for example the Chebyshev coefficients [2-8] on ULA achieves constant side lobes of any level and minimum beam width for that level of side lobe. This is called the Dolph-Chebyshev Array. Thanks to the reuse of digital signal processing techniques, many more alternative exist [2-7 Ch. 23].

In general, using non-unity coefficients, while having the potential to improve the spatial filtering transfer function, will broaden the main beam and reduce the amount of energy it contains, meaning a potential reduction in SNR for a receiver. It is a tradeoff between noise reduction and interferer rejection capability. It can be noted that this approach requires a LoS configuration between the user and the BS, you cannot point a beam at a user you cannot see. Also, in general this approach filters out the additional energy received from the multi-path propagations outside the LoS path, not fully exploiting the channel capacity.

Let us look at the processing complexity of beam-shaping. It can be divided into three functions, the CSI estimation, the processing of the beamforming coefficients and the beamforming processing itself, which applies the coefficients to the received signals.

In the context of beam-shaping, the CSI is reduced to the AoA of the different users. Its initial estimation can be processing intensive. Thankfully, a user displacement between two successive communications is small. It can then be assumed that the subsequent AoAs for the same user will be near the previous ones. This allow to reduce drastically the range of AoA to be scanned to evaluate future CSI. It is also possible to develop tracking algorithms with even lower complexities. This makes the CSI estimation processing negligible compared to the remaining processing.

Since the CSI has a very simple representation, and because the coefficients to be applied for a given user beam are independent of the other users, it is possible to preprocess the coefficients for a predetermined mesh of AoA and store them in a code book. One could even envision some synergy with the tracking algorithm where the coefficients for the potential future AoA with the highest likelihood are locally preloaded to cut on the memory access time. With these assumptions the processing of the beamforming coefficients also becomes negligible.

Only remains the beamforming processing which is incompressible. It is clear here that the main advantage of this approach comes from its simple algorithmic complexity.

2.2.2.3 Massive MIMO

The origin of Massive MIMO is rooted in 2 different domains, the Adaptive Array Processing (AAP) and the Multiple Input Multiple Output (MIMO) systems.

The goal in AAP is to adjust the beamforming to the received signals. Early works [2-10] [2-11] mostly focused on the beam self-alignment along the incoming signal AoA. It was generally done using analog feedback loop circuits aiming at maximizing the received power. Rapidly the ability to filter an interferer from a given direction was added [2-12]. A good introduction on the subject was done by William GABRIEL in [2-13]. He takes the Howells-Applebaum servo-loop example and use it to derive all the major matrix relationships in use today. These remain true in the context where each signal is coming from only one angle, i.e. there is no multi-path propagation for a given signal. The following years saw more investigations in the context of mobile network for cellphones with more complex channel models and wider bandwidth. This is nicely summed up in [2-14] where emphasis is put on both the spatial processing from the antenna array and the time processing to be used to combat channel multipath fading.

In the context of wireless communication, the MIMO theory is used to describe a system where N_t transmitters and N_r receivers are working at the same time on the same frequency in the same area. It is clear that such a configuration will lead to significant in band interference at the receivers between the transmitted signals. Through MIMO formalism it is possible to show that, in some environments, proper communication can still be achieved. The channel is then described by an $N_r \times N_t$ matrix $\boldsymbol{H}$ where the $h_{i,j}$ complex coefficients describe the interaction between the j^{th} transmitting antenna and the i^{th} receiving one. The modulus of the complex coefficient represents the attenuation and its argument the phase rotation. In that case $\boldsymbol{H}$ represent the Channel State Information (CSI).

The matrix $\boldsymbol{H}$ contains many pieces of information about the MIMO channel. Among them is the spatial diversity offered by the environment. The maximum number of data streams is equal to the rank of $\boldsymbol{H}$ and is bounded by $\min(N_t, N_r)$. What is called a data stream here is a signal that can recovered, after the MIMO processing, with the same SNR regardless of the presence other streams over the same MIMO channel. If more data streams where to be used than the rank of $\boldsymbol{H}$, they would interfere with each other in a way that the MIMO processing could note undo. This would degrade the signal's SNR and the achievable data rate.

The 4G standard already propose such capability for a single use case, meaning that the transmitters all belong to the user and the receivers to the BS for uplink. A user and a BS may have up to eight antennas each (8×8 MIMO), allowing at best to improve the data rate by a factor of eight. In Practice the spatial diversity offered by the environment rarely allows more than three of four data streams.

In the context of 5G, it is Multiple User MIMO (MU-MIMO) systems that are of interest. For uplink, this correspond, for the N_t transmitter, to be different users, and the N_r receivers, to be all the antennas available at the BS. A system is called Massive MIMO when the number of antennas at the BS is large compared to the number of users. The mathematical theory behind such systems has a lot in common with AAP and in many cases was inspired by it.

Let us consider N_U users, each of them having one antenna and transmitting signals at the same time, on the same frequency, in the same area, to the same BS equipped with an array of N_{ant} antennas. The channel is then described by a $N_{ant} \times N_U$ matrix $\boldsymbol{H}$. The signal received by the BS is expressed by:

$$\boldsymbol{y} = \boldsymbol{H} \times \boldsymbol{x} + \boldsymbol{n} \qquad (\text{'.16})$$

Where the component x_j of vector $\boldsymbol{x}$ is the transmitted signal by the j^{th} user, y_i the signal received by the i^{th} antenna of the BS array, and $\boldsymbol{n}$ is the noise vector of the BS antenna array.

Optimal performances can be achieved at the BS, assuming CSI are known, by using algorithms such as Maximum Likelihood (ML). It consists in trying all the possible combination of transmitted symbols $\boldsymbol{x}$. The estimated transmitted signal $\hat{\boldsymbol{x}}$ is the one that minimizes the distance between the actual received vector $\boldsymbol{y}$ and $\boldsymbol{H} \times \hat{\boldsymbol{x}}$, i.e. the input that maximize the likelihood of the observed signals. While

giving optimal performances this method has a processing complexity that is exponential with the number of users, the size of the symbol constellation and the number of antennas at the BS.

Linear processing algorithms do not suffer from such an exponential complexity, at the cost of sub-optimal performances. Thankfully, an important result from [2-15] states that, under the Massive MIMO approximation, i.e. the number of receiving antennas is much larger than the number of transmitting ones, linear processing algorithms become near optimal. It is for this specific reason that Massive MIMO is set as a center technology in sub-6GHz 5G. To know if it is also useable in the millimeter wave domain is one of the questions this manuscript is trying to answer.

Assuming the CSI represented by the matrix $\boldsymbol{H}$ is known, linear processing consists in applying a detection matrix $\boldsymbol{A^H}$ to the received signal $\boldsymbol{y}$, where $\boldsymbol{A}$ depends on $\boldsymbol{H}$ and $\boldsymbol{A^H}$ is the Hermitian transpose of $\boldsymbol{A}$, to produce $\boldsymbol{\hat{x}}$, the estimation of $\boldsymbol{x}$.

$$\boldsymbol{\hat{x} = A^H \times y = A^H \times H \times x + A^H \times n} \tag{2.17}$$

There are three common ways to build $\boldsymbol{A}$ respectively called Maximum Ratio Combining (MRC), Zero Forcing (ZF) and Minimum Mean-Square Error (MMSE) [2-16]:

$$\boldsymbol{A} = \begin{cases} \boldsymbol{H} \\ \boldsymbol{H \times \left(H^H \times H\right)^{-1}} \\ \boldsymbol{H \times \left(H^H \times H + \frac{1}{p_u} \times I\right)^{-1}} \end{cases} \tag{2.18}$$

Where p_u is the average received power from each user and $\boldsymbol{I}$ is the identity matrix. From (2.18) the processing complexity of the three different algorithms can be compared. As beam-shaping the complexity has three components. The CSI evaluation, the detection matrix evaluation, and its application to the received signals.

The CSI evaluation is the same for all three algorithms and is much more complex compared to beam-shaping, simply because the CSI is represented by many more parameters. It is usually estimated through the help of a pilot, a predefine sequence send by the user at the beginning of uplink. By comparing the actual received signal with the original signal it is possible to extract the CSI. It exists many different methods to perform this estimation. Giving an estimation of the problem's complexity is difficult. Nonetheless, from the survey in [2-36], it can be stated that, for a pilot-based estimation, the complexity is at least proportional to $N_t^2 \times N_r^2$. Since the target is systems with large number of antennas, this makes the channel estimation a significant contributor to the overall processing complexity.

The processing complexity of the detection matrix varies for the different algorithm. MRC is the simplest with a null complexity since it is directly the CSI matrix $\boldsymbol{H}$. On that point it is even simpler than beam-shaping. ZF is significantly higher with two matrix multiplication and mainly a matrix inversion which can be relatively processing intensive. MMSE have about the same number of operation but also requires the additional information of the average received power from each user. This information can be acquired from $\boldsymbol{H}$ at the cost of addition processing, making MMSE the algorithm with the highest complexity.

Finally applying the detection matrix is common to all three algorithms and is the same as the beamforming processing in beam-shaping.

Each of these algorithms have different properties. MRC tends to maximize the power received from each user, when ZF tends to minimize the inter-user interferences by steering zeros in the interferers' direction, and MMSE is a compromise in-between that aims at maximizing the SNR.

The author in [2-15] also studied the capacity of a mobile network with BS array of "infinite" size. The result is that it does not grow indefinitely but is limited by the pilots' contamination. This contamination limits the achievable accuracy on CSI acquisition and consequently limits the network capacity. This is a significant step in understanding the potential of such systems. This can also be used as a criterion to evaluate how many antennas is "infinite". It has been shown in [2-17] that, when the array is large enough, MRC is near optimal for communication using low SNRs while ZF is near optimal for communication with high SNRs. Finally, MMSE is near optimal in all SNR scenarios.

Overall, Massive MIMO has the capacity exploit the channel capacity in a near optimal way. This performance increase comes at the cost of significantly higher processing complexity compared to beam-shaping.

2.2.2.4 Conclusion

Two approaches to beamforming have been seen. One really focused on creating a beam toward a user and which is of a simpler algorithmic complexity but with sub-optimal performances. The second uses the Massive MIMO approach which has the potential for near optimal performances, but at the cost of significantly higher algorithmic complexity. From here three general architectures for millimeter wave small base station receivers can already be envisioned.

One would use a lower bandwidth and hence higher SNRs to maintain performances. In that case a Massive MIMO approach using ZF or MMSE would be a better algorithm choice. The higher complexity of such algorithms would drive the design toward minimizing the number of antennas. To maintain a high SNR that would require higher PA output power, also leading to larger interferers. This would lead to an architecture of lower complexity with higher requirements toward Dynamic Range (DR) and noise performances but with reduced needs in bandwidth.

The second would be using a higher bandwidth, large enough to ensure low SNRs. Here, a Massive MIMO approach using MRC would be a good fit. Its lower complexity, which is essentially the CSI estimation and the beamforming processing, would allow for more antennas. Lower SNR and higher number of antennas means lower PA output power and reduced interferer. This would relax the BS hardware constraint on DR and noise performances but increase the requirement for bandwidth.

The last one would be using the largest possible bandwidth reaching an even lower SNR. The algorithm choice would naturally go to the lowest complexity of DWS beam-shaping approach. This would allow for an even larger antenna array, significantly reducing the radiated powers and interferer levels, drastically reducing requirements on DR and noise performances and potentially compensating for its sub-optimal performances.

A tradeoff is appearing between the architectural and the algorithmic complexity. To find the optimal point on this tradeoff, if it exists, is a hard question. As is will be shown later, there are other parameters affecting this tradeoff and that can be used to choose to investigate an interesting part of the design space. One of these parameters in the array topology. Until now the main focused was on ULA. In the next section, the impact of different array topologies will be studied.

2.2.3 Antenna Array Topology

Beyond the Uniform Linear Arrays already studied, there are other array topologies in the literature ([2-18] [2-19] [2-20] [2-21]) which offer interesting properties. In this section, the focus will be on a subset of them, called $\lambda/2$ uniform array, where antennas are evenly spaced by a distance of $\lambda/2$. In particular, Uniform Planar Arrays (UPA) and Uniform Circular Arrays (UCA) will be studied. The performances of a given array topology can be evaluated through many figures of merit. Here, the same as before will be used, the HPBW and the PSL.

The purpose of this study is to gain understanding of the topology's impact on the array performances. It has been seen that the DWS approach, while it can reduce significantly the PSL, always does it at the cost of widening the HPBW and reducing the total useful received power. The question is: Is it possible to improve the PSL performances of the DS beamforming by acting on the array topology? That would allow to maximize the useful received power and reduce the HPBW without compromising the spatial filtering capability caused by degraded PSL.

2.2.3.1 Uniform Planar Array

One limitation of the ULA's is their ability to steer only along one direction. To overcome this limitation, one can extent the array topology to two dimensions. Antennas are placed in a rectangle shape and evenly spaced by $\lambda_c/2$ in both directions. To describe this problem, the coordinate system is set as per Figure 2-9, where antennas are located through their Cartesian coordinates (x_k, y_k) in the X-Y plan and the AoA through the angular spherical coordinates (θ, φ).

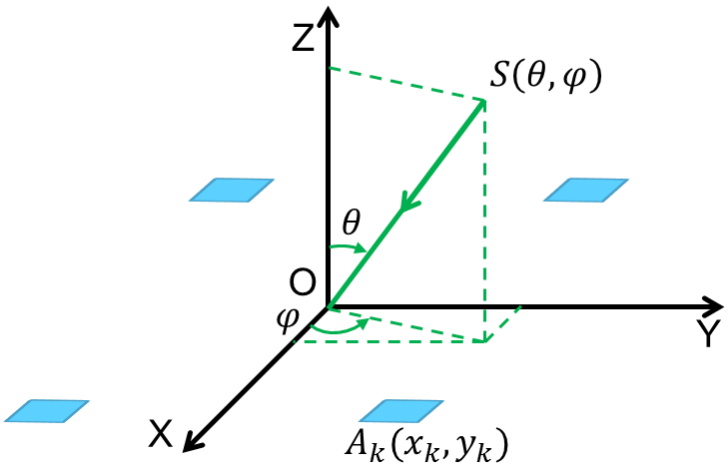

Figure 2-9: Used convention to describes UPAs

The method is the same as for ULAs, i.e. the delay a plane wave with AoA $(\varphi_{BF}\,;\theta_{BF})$ would have is processed for each antenna. These delays are then compensated to form a beam in that direction. Getting the spatial transfer function of the array requires to process, at the beamformer's output, the power coming from all directions. One way to look at it is as a ULA in the Y direction where the array elements are themselves ULAs in the X direction. From that perspective the spatial transfer function can easily be expressed as the combination of the two ULA as in equation (2.19).

$$G_{sp_{UPA}}(N_{ant}, \theta_{BF}, \varphi_{BF}, \theta, \varphi) = G_{sp_{ULA}}(N_{ant}, \theta_{BF}^{X}, \theta^{X}) \times G_{sp_{ULA}}(N_{ant}, \theta_{BF}^{Y}, \theta^{Y}) \qquad (2.19)$$

With $\theta_{BF}^{X} = arctan(tan(\theta_{BF}) * cos(\varphi_{BF}))\,, \theta_{BF}^{Y} = arctan(tan(\theta_{BF}) * sin(\varphi_{BF}))$ and $\theta^{X} = arctan(tan(\theta) * cos(\varphi))\,, \theta^{Y} = arctan(tan(\theta) * sin(\varphi))$.

As a consequence, most of the ULAs results are also true on UPAs. The behavior is similar except that the spatial transfer function now needs to be seen as a surface in a 3D space.

Let us plot this surface for a square antenna array with 8×8 elements and a steering angle $(\varphi_{BF} = -\pi/4\,; \theta_{BF} = \pi/6)$ (Figure 2-10-a). While having some esthetic properties it is not practical to get a good understanding. Figure 2-10-b plots the profile of this radiation pattern when sliced by a plan passing by the main beam center, the peak side lobe center and the origin. Let us call this the profile along the PSL. One can note that this profile is very similar to ULAs. The last item to be studied is the HPBW. It is now a solid angle. The contour of this solid angle is plotted on Figure 2-10-c, and the corresponding solid angle is the area inside this contour when projected on the unit sphere. It is

measured in steradian. Assuming MKSA unit system, the area covered by the HPBW at a distance $D = 50m$ from the antenna array is $A_{50m} = HPBW \times D^2 = 0.0451 \times 50^2 \approx 113m^2$. Such a surface could cover a large number of users that would therefore need to share the spectrum inside the beam.

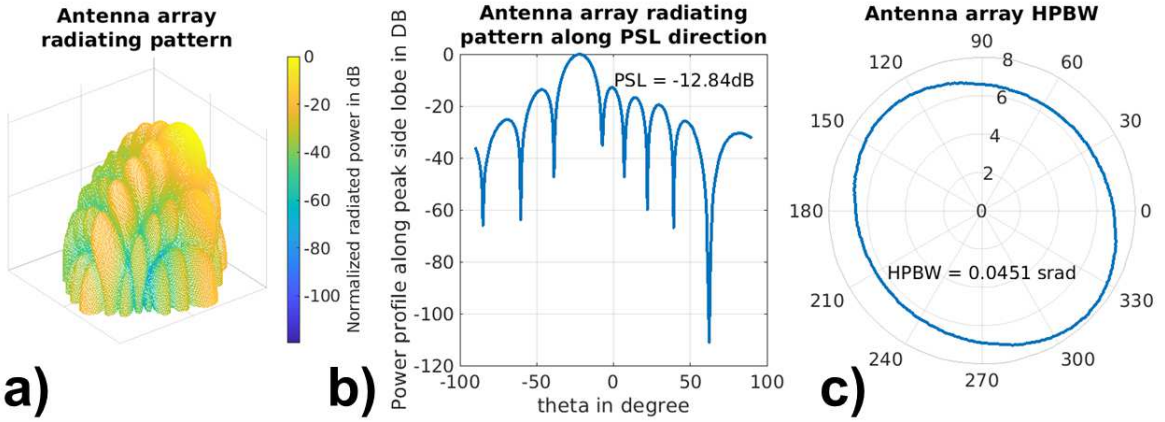

Figure 2-10: 64 element UPA a) Radiation Pattern b) Profile along PSL c) HPBW contour

Overall UPAs behaves in a very similar way compared to ULA. Figure 2-11 plots the HPBW and the PSL versus the number of antennas for $(\varphi_{BF} = 0\,;\theta_{BF} = 0)$. On can observe the same linear relationship on a log-log scale for the HPBW and the same limit for the PSL around $-13.26dB$.

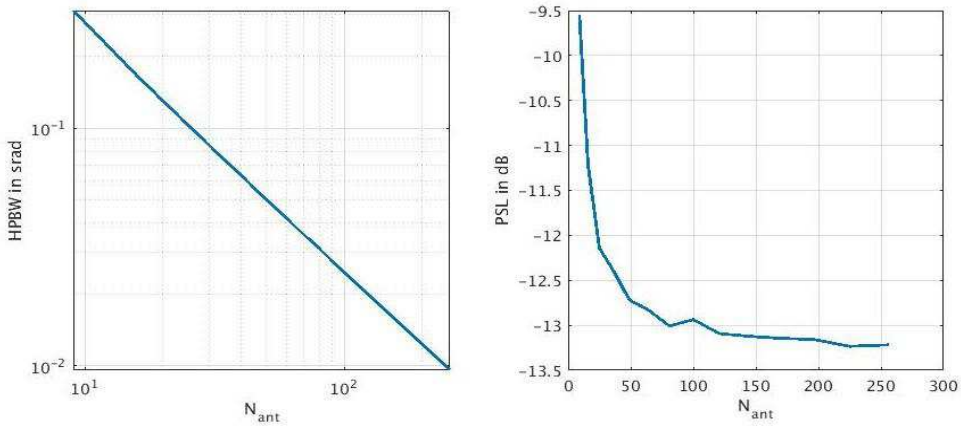

Figure 2-11: HPBW (left) and PSL (right) for square UPAs versus the number of antennas N_{ant}

While the best PSL is achieved with a reasonable number of elements, the same conclusion cannot be reached for the HPBW. Let us estimate the required number of antennas to have a HPBW of $1m^2$ at a $50m$ distance from the antenna array. This is the right order of magnitude for a 5G millimeter wave small cell. Thanks to Figure 2-11-b, the result from a 64 antenna array can be extrapolated. Since it was covering a surface of $113m^2$, 113 times more antennas would be needed to have a beam covering $1m^2$, which is about 7000 antennas. This is one of the major drawbacks of UPAs, while they offer good performances in terms of Peak Side Lobe, they require a very large number of antennas to achieve narrow beams.

2.2.3.2 Uniform Circular Array (UCA)

Using the same convention as with UPAs, UCAs will only differ by how the antennas are placed in the X-Y plan. They will be along a circle centered at the origin and evenly spaced on it. The radius needs to be adjusted to fit the desired number of antennas with the desired spacing $\lambda_c/2$. Then, the same evaluation as for the UPA is processed, as shown on Figure 2-12 and Figure 2-13.

The performances of UCAs vary drastically differently from UPA as a function of the number of antennas. While they require much less antennas to achieve narrow beams, their PSL is much less interesting, being around -8dB and independent of the number of antennas. The 64 element UCA of Figure 2-12 has a HPBW covering $11m^2$ at $50m$ distance, about ten times smaller compared to an equivalent UPA.

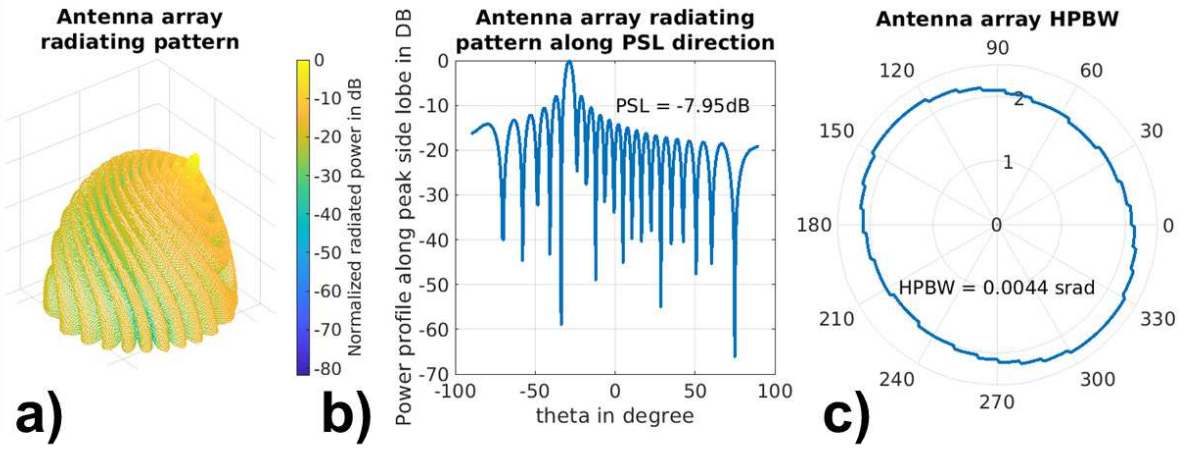

Figure 2-12: 64 element UCA a) Radiation Pattern b) Profile along PSL c) HPBW contour

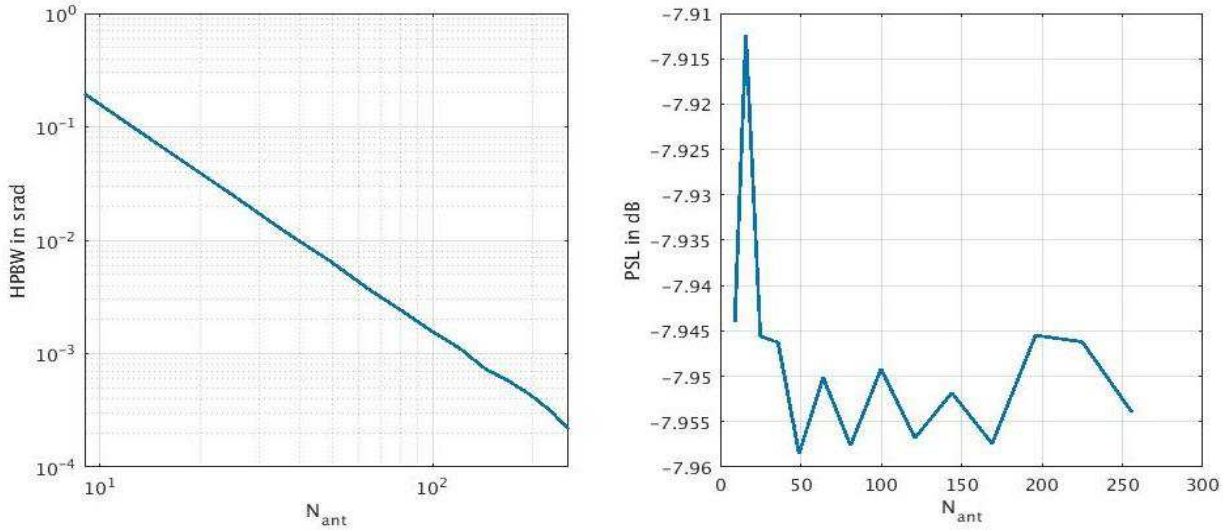

Figure 2-13: HPBW and PSL for UCAs versus the number of antennas N_{ant}

It is possible to achieve performances in between these two topologies in many different ways. For example, it is possible to use nonuniform antenna spacing [2-18]. One way of doing it, is to remove some of the antennas in a way that minimize the degradation of the radiation pattern. This is called thinned array. In [2-19] they use a genetic algorithm to optimize thinned ULAs and UPAs yielding interesting results.

Another possibility is to use Multiple Concentric UCAs (MC-UCA) [2-20] with different numbers of antennas on each ring. Figure 2-14 displays the performances of such an antenna array where the number of antennas on each circle is given by the following piece of the Fibonacci's sequence $[8\,;13\,;21\,;34\,;55\,;89\,;144]$, for a total number of 364 antennas. The choice of this sequence has no other justification than providing good performances in HPBW and PSL. The PSL for this array is $-16.52dB$, and its HPBW at $50m$ is $4.5m^2$, i.e. a circle of about $2.4m$ in diameter. In

[2-21] they develop MC-UCA with non-uniform spacing and with uniform excitation. This approach allows to reduce the side lobes below $-20dB$ with a narrow beam width and a reduced number of antennas. For later considerations it will be assumed that it is possible to design a 256 elements antenna array with PSL below -20dB and HPBW at $50m$ of less than $2.5m$ across.

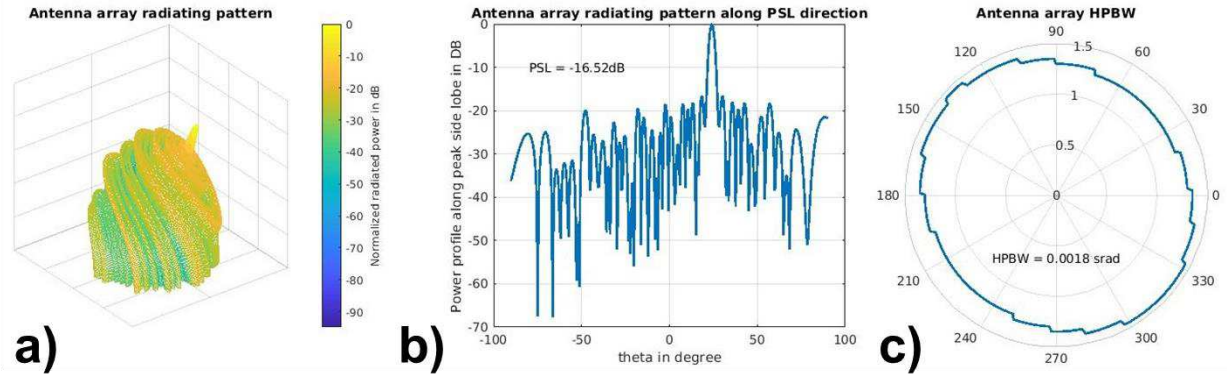

Figure 2-14: 364 elements MC-UCA a) Radiation Pattern b) Profile along PSL c) HPBW contour

2.2.3.3 Conclusion

It may seem that MC-UCAs outperforms other topologies in all aspects, but this is only due to the incompleteness of the used metrics. As an example, one could look at the amount of the total energy that is focused in the main beam, or equivalently how much energy is lost to the side lobes. In that regard, UPAs would perform better. Another way to compare antenna array topologies could be to evaluate the probability for an interfering signal coming from an angle outside of the main beam to fall on a side lobe. Here again, UPAs would most likely perform better. These remarks are here to recall that the proposed study is by no mean complete. The relevant figure of merit may vary with the application, the chosen system architecture and even on some implementation details. Nevertheless, some preliminary conclusions can be made. In the context of 5G, the antenna array needs to have 3 properties:

- The first one is to be able to control the beam toward any user, wherever they may be in the cell. This requires the ability to steer in two directions and implies a planar array.
- The second is for one beam to be narrow enough to cover the minimum number of users. Regardless of the topology, increasing the number of antennas reduces the beam width, but some topologies such as Uniform Circular Arrays perform better.
- The third is to filter out signals outside the main beam. If the goal is to minimize the number of antennas, this must be compromised with the second properties.

Finally, it was shown that a Multi Concentric Uniform Circular Array with some hundreds of elements can provide a narrow beam as well as low side lobes. This provides an idea of how many antennas could be required on such systems. It is also interesting to note that while this study was done with a beam-shaping approach in mind the array topology may also be of some importance for a Massive MIMO approach [2-22]. Different topologies may help in increasing the MIMO spatial richness for example. This is an interesting idea that is left for future work.

2.2.4 Beamforming receiver architectures

A beamforming receiver is made of two main components. The first one, described in the previous section, is the antenna array. The second is the receiver. Only the part of the receiver that acquire and process the antennas signals to deliver the appropriate signals to the demodulator will be considered here. On top of the functions of a classical receiver (amplifying, filtering, down mixing, digitization …) a beamforming receiver has to process and recombine the antenna signals to effectively form the beams.

To perform this task, there are three general architectures. They differ by the domain, analog, digital or a mix of both, in which the beamforming processing is implemented. They are respectively called Analog Beamforming (ABF), Digital Beamforming (DBF) and Hybrid Beamforming (HBF). Each of them has their pros and cons and will be detailed in this section.

2.2.4.1 Analog Beamforming

The operations required to perform beamforming, as described earlier, are of three kinds, phase-shifting or delay, weighting and summing. A typical block diagram of an ABF receiver is depicted in Figure 2-15-a

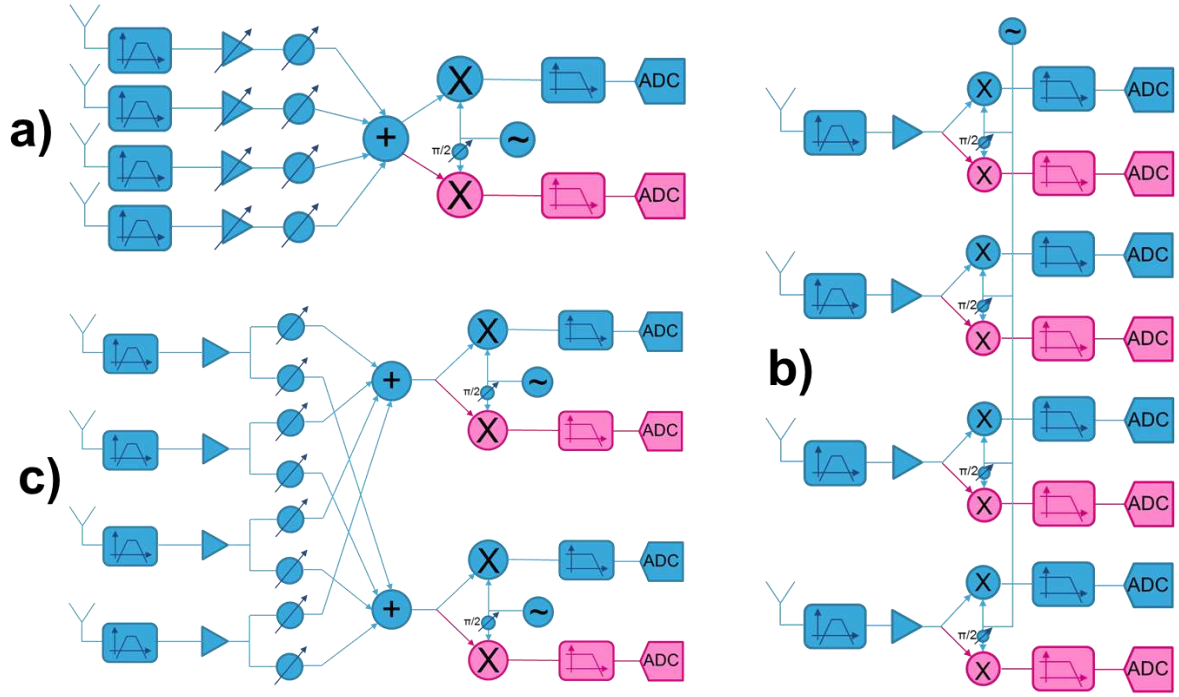

Figure 2-15: Beamforming Receiver block diagram: a) ABF b) DBF c) HBF

With this architecture the beamforming is implemented in the RF domain. It is also possible to perform it in the analog base band domain. An interesting trick for phase shifting can be done by phase shifting the LO instead of the signal, removing the phase shifter from the signal path.

The main advantages of ABF are its low power consumption and the Dynamic Range (DR) relaxation for the components coming after the beamforming. This is because interferers from other AoA have already been filtered.

This architecture also has some disadvantages. Most practical implementations use phase shifters instead of time delays. It will be seen later that this has an impact on the system ability to process large bandwidth, but its main limitation is its inability to create multiple separate beams since it has only one base band stream. Since this is necessary for BS it makes it unsuited, and it will not be considered any further in this analysis. Note that it might still be of interest for the UE.

2.2.4.2 Digital Beamforming

The block diagram of DBF is depicted in Figure 2-15-b. Each antenna comes with its own receiver up to the Analog to Digital Converter (ADC). The beamforming processing is done in the digital domain (not represented in Figure 2-15-b for the sake of clarity).

The main advantages of DBF are that the analog part of the receiver have relaxed noise requirements. It can create multiple beams without any modification of the analog part of the receiver. It only needs

to create a copy of the digital signals and apply a different beamforming to it. It has also the possibility to use time delays. Finally, it can benefit from the flexibility of digital processing.

The disadvantages are, on the analog side, the requirement of a full receiver per antenna and with higher constrains on DR compared to ABF. This is because no spatial filtering has happened yet, hence the receiver need enough DR to process the full interferers without saturating. This architecture generates a massive amount of data and the management of all the digital processing required is a huge challenge. In particular, the implementation of time delays. A classic method is to perform a band limited interpolation. This requires processing the convolution product between the signal and a cardinal sinus, which is a processing intensive operation and must be performed for each antenna.

2.2.4.3 Hybrid Beamforming

As its name suggests, this architecture performs the beamforming partly in the analog domain and partly in the digital domain. A fully connected architecture with two baseband streams is depicted in Figure 2-15-c. HBF refers more to a class of receivers than to a precis architecture because of the partitioning flexibility between the analog and the digital domains. Conclusions cannot be drawn without specifying more its architecture. Several classical HBF exists and most of them assumes only phase shifting and no weighting in the analog domain. The two most common are the fully connected and the sub-array architecture.

- Fully connected architectures see a combination of all the antenna signals at the input of each of the baseband processing unit (Figure 2-15-c). An interesting result from [2-23] is that, with several baseband processing units equal to twice the number of beams to be formed, a fully connected architecture has equivalent performances as a DBF one. It is also possible to maintain performances close to DBF with the same number of baseband processing units as the number of beams. For example, a fully connected HBF receiver with four baseband processing units can form two beams and achieve the same SNR as a DBF receiver, or up to four beams with a slight SNR degradation.
  The main advantage of fully connected HBF is that it significantly reduces the amount of digital data to be processed while maintaining the same level of performance.
  Its disadvantage is the complexity of the analog side. In particular, the number of phase shifters per antenna is equal to the number of baseband processing units, which is at best equal to the number of beams. This means the signal of each antenna must be split, i.e. its power is divided by the number of splits (if done in the RF domain. If done in the base band domain, it increases the number of analog components).
- The sub-array architecture is the aggregation of side-by-side ABF with an additional layer of processing on the digital signals of the ABFs output that forms the base band streams. While it does not suffer from the same analog complexity compared to the fully connected approach, it cannot fully exploit all the antennas when forming multiple separated beams.

Regardless of the HBF type, this architecture scales poorly with the number of beams. Since this is a mandatory feature for a BS, it makes it a less interesting candidate compared to DBF for future developments.

2.2.5 Implementation considerations

So far, only ideal components were considered, i.e. ideal antennas, receivers, (…). Obviously, a real implementation will introduce some non-idealities. It is important to have an idea of their impact on the performances, and if possible, to study the sensitivity of the different architectures to such imperfections. Here, a mostly empirical approach will be taken, based on simulations to remain concise.

2.2.5.1 Individual antenna radiation pattern

The first non-ideality to be looked at, is related to the antenna radiation pattern. Until now, the assumption of perfect isotropic elements was made. This is obviously not the case in reality. Planar arrays may use different antennas, but the most widely used is the patch antenna. It is composed of a patch of metal, often a square, on top of a ground plane separated by an insulator as shown in Figure 2-16.

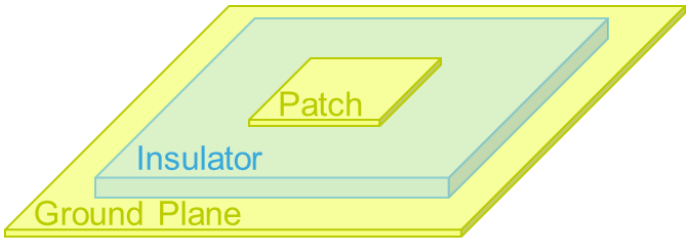

Figure 2-16: Basic architecture of a patch antenna

This kind of antenna is widely used since it can be made from cheap and well controlled Printed Circuit Board (PCB) manufacturing processes. It is especially convenient for antenna array since multiple patches can be put on the same ground plane, providing a simple solution for planar arrays. The first difference with the ideal case is that it radiates energy only on the top hemisphere. In that hemisphere, radiation is also not isotropic. In general patch antennas have wide HPBW often over $100°$ ([2-24]-[2-28]). Thankfully, since an antenna is a linear device, assuming they are all identical, the radiation pattern of an ideal array can be simply multiplied with the one of a single antenna to get the overall radiation pattern.

There are two major consequences. The first one is that it is not possible to steer efficiently outside of the single antenna HPBW. This means that to cover the $360°$ of a cell, four arrays will be required. The second one is that the side lobes outside of the single antenna HPBW will be attenuated compared to the ideal case. This could be exploited when optimizing the array topology.

2.2.5.2 Time delay versus phase shift

In RF, the implementation of phase shift is more convenient than time delay. However, this has some consequences. In particular, it affects the frequency response in the main beam and the radiation pattern. The left part of Figure 2-17 plots the frequency response over $1GHz$ bandwidth around a $28GHz$ center frequency, for various steering angles for both time delay and phase shift beamforming. As expected from equation (2.11) the time delay approach has a flat frequency response in the main beam regardless of the steering angle (all Time Delay curves are overlapped in Figure 2-17 left graph).

When using phase shift, the in-beam response is also flat for a steering angle of $0°$. This is intuitive since, in this situation, a plane wave reaches all the antennas at the same time, as a consequence, no phase shift or time delay is needed. As the steering angle increases, the response is less and less flat. The right graph of Figure 2-17 shows the frequency response for a $45°$ steering angle for ULAs with 32, 64 and 128 elements. It is clear that the array size has a significant impact. For the ULA with 128 antennas the attenuation at the edge of the band is greater than 12dB. Since, to achieve a narrow beam width a large array is necessary, the use of time delay is highly desirable for wide band operation.

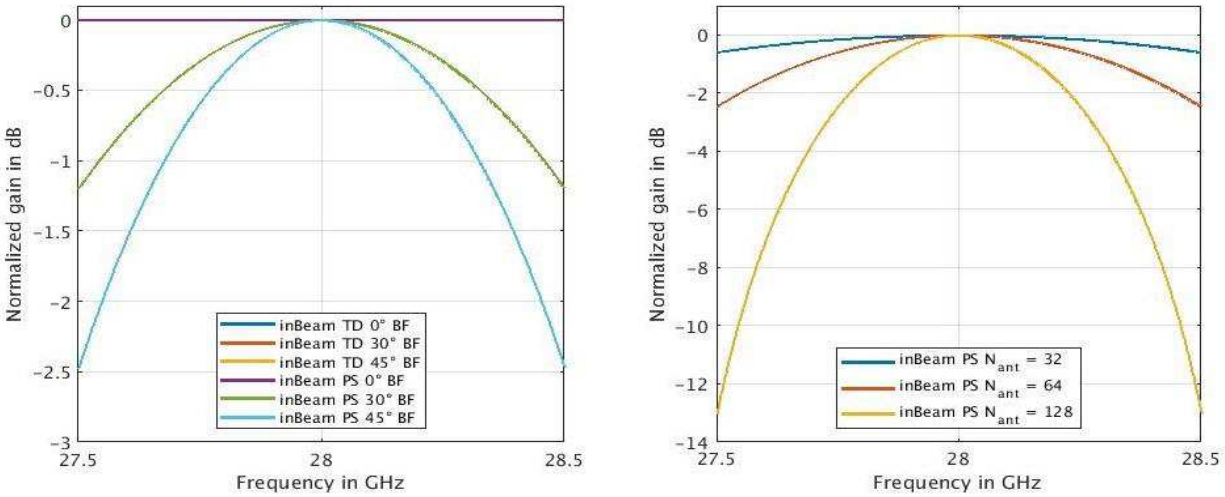

Figure 2-17: In Beam frequency response of time delay versus phase shift beamforming. Left: ULA with 64 elements for steering angles of 0°, 30° and 45°. Right: Steering angle of 45° using phase shift beamforming for ULAs with 32, 64 and 128 elements

Figure 2-18 plots the Array Factor for a 128 elements ULA for a single tone and for a wide band signal, using time delay and phase shift beamforming. The wide band signal is approximated by an evenly spaced 1001 tones signal over $1GHz$ bandwidth around $28GHz$ center frequency.

Both time delay and phase shift beamforming suffer from reduced null depth, but the effect is much more sever when phase shifting is applied. This is more severe for nulls near the main beam. One consequence is that a null-steering approach, such as ZF, will be ineffective for interferers near the main beam. One can also notice that the main beam broadens significantly, reducing the pointing ability of the system.

Overall, a time delay approach is very desirable since the adverse effects of phase shifting increase for wider bandwidth and larger arrays, while both of them are required to achieve 5G targeted performances.

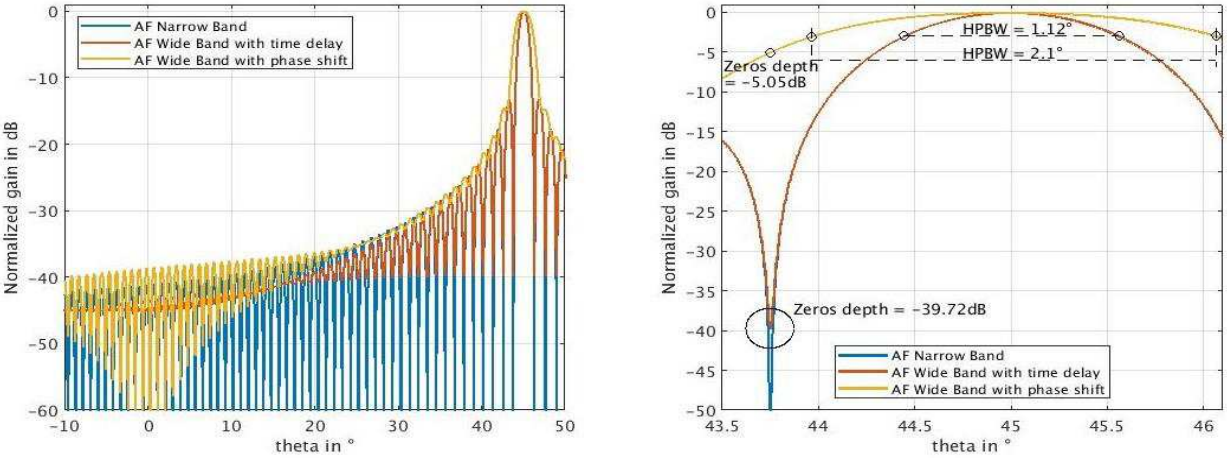

Figure 2-18: Array factor for a 128-element ULA for narrow and wide band signal with time delay and phase shift beamforming

2.2.5.3 Delay and Gain errors

Here, the impact of antenna signals inaccurate time delays and gains on the Array Factor will be studied, regardless of the inaccuracy source, whether it is CSI limited accuracy or hardware imperfections. In both cases, the inaccuracy is modeled as an additive white Gaussian error with zero mean. Then, the Root Mean Square (RMS) error increases is monitors to evaluate the impact. This analysis will first focus on delay and gain errors separately, and then jointly for narrow and wide band signals.

Figure 2-19 plots the Array Factor averaged over 200 runs for various RMS errors in delay and gain. It is clear from left graphs that both gain and delay errors have a similar effect of creating a floor of maximum filtering capability. It is not all that surprising since the math are very similar to the addition of a white Gaussian Noise on a signal. The difference being that, here, the spatial spectrum is observed, not the frequency one. On the right-side graphs, it can be seen that the main beam is nearly unaffected. As it will be detailed later, this possibility of always accurate pointing can be exploited to provide calibration signals.

Once the spatial filtering capacity required for the system is known, this result can be used to evaluate the required precision. A $-33dB$ floor corresponds to a 40% RMS gain error, while it is only $2ps$ RMS error on the time delay. A gain accuracy of 40% is very relaxed but $2ps$ is a much harder design challenge.

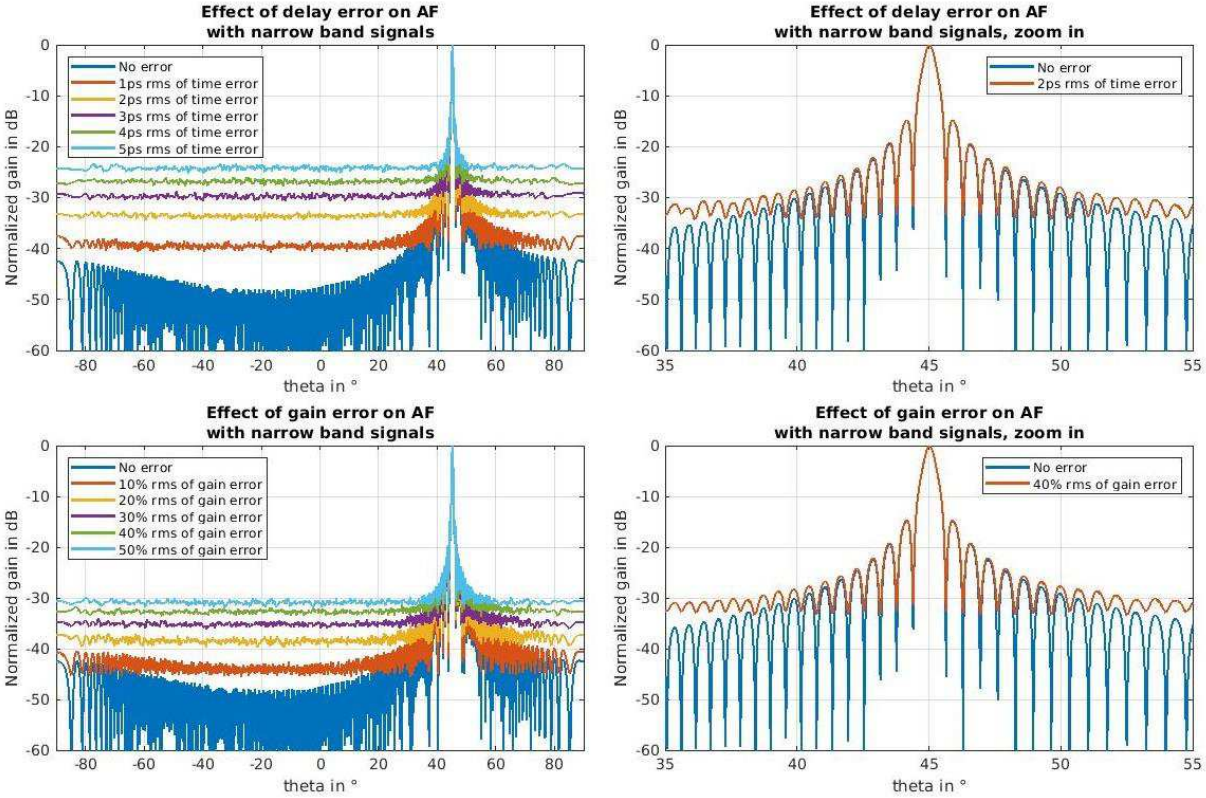

Figure 2-19: Top graphs: Average impact of time delay error on Array Factor. Bottom graphs: Average impact of gain error on Array Factor

Delay and gain errors adds many local effects that attenuated when averaging multiple runs. While general conclusions cannot be made when looking at a single run, it is still interesting to look at the different, from the same run, between narrow and wideband signals. Figure 2-20 plots the array factor for one such single run of cumulated gain and delay error with respective RMS values of 40% and $2ps$. This run was note particularly chosen and present similar characteristics compared to other runs that have been observed. The top graphs are for a narrow band signal and the bottom ones for a wide band

one. The wide band signal is modeled by 41 tons evenly spaced over $1GHz$ bandwidth at $28GHz$ center frequency.

Locally the gain and delay errors can give rise to significant side lobes. As for the averaged AF the main beam is unaffected. There is also little difference between the narrow and the wide band signal in the vicinity of the main beam. Further away from the main beam the wide band signal sees some significant smoothing compared to the narrow band. This is beneficial in general since it reduces the likelihood of a high side lobe.

On one hand the main beam is largely unaffected by the introduction of errors. In particular, its pointing accuracy is unaffected. The AF away from the main beam, on the other hand, sees strong limitations in its spatial filtering ability. Null-steering based beamforming would probably require stringent calibration constraints that might render them impractical even with a time delay implementation. Even for MRC or beam-shaping approaches, the timing accuracy is likely to be a challenge.

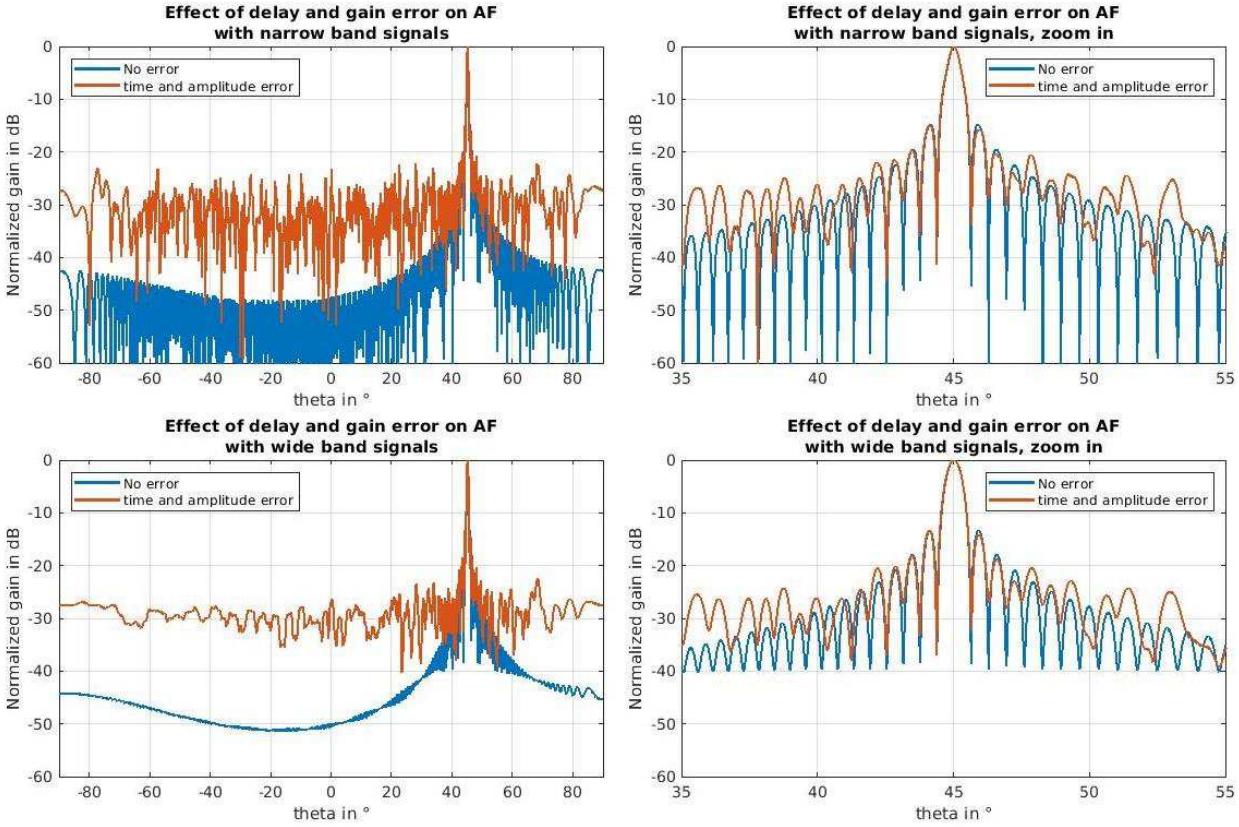

Figure 2-20: Top graphs: Impact of delay and gain error on Array Factor for narrow band signals Bottom graphs: Impact of delay and gain error on Array Factor for wide band signals

2.2.5.4 Beamforming with a wide band fast fading channel

In section 2.2.2 the different algorithms were studied under the narrowband approximation, i.e. when delays can be approximated by phase shifts. In section 2.2.5.2 it was shown that this approximation is not valid for the kind of systems studied in this manuscript; large antenna arrays using wideband channels. Here, a second aspect will be looked at, where this approximation induces impairments in the output. It is the channel frequency response. Until now, only the case of an AWGN channel was considered, where the frequency response is flat. For this to be true it would require for the wave traveling from the UE to the BS to go through a single path, ideally, the LoS one. Unfortunately, this is not the case. The RF channel's properties at 28GHz in urban areas have been studied in large cities such as Manhattan [2-29]. It has been shown that in average there are four different paths. The rays following

these paths, since traveling different distances, will reach the receiver at different times and with different amplitudes. This leads, through constructive and destructive recombination, to a frequency dependent channel response. This effect is called fast fading and cannot be overlooked by the significance of its impact on the channel response.

The frequency dependence impact on beamforming performances will be studied. To make some evaluation on that matter, a fast-fading channel model will first be built, and then used with different beamforming algorithms to compare them.

The model is based on a ray tracing multipath approach, to be used on the 28GHz channel model proposed in [2-29]. Each path is a cluster of sub-paths as depicted on the left side of Figure 2-21. The user is at the edge of a quarter cell of radius 50m making an elevation AoA of about $\theta = 45°$ and hold his UE at $1.5m$ height. The antenna array is the MC-UCA studied in section 2.2.3.2 and is located at 10m height. Based on measurements from [2-29], the model has four path clusters on top of the LoS and three sub-paths per cluster. The clusters of scatterers are uniformly distributed in a $50m$ cylinder above the ground, centered on the UE topping at $11.5m$ ($10m$ above the UE). They have an additional attenuation uniformly distributed between $0dB$ and $-10dB$ to account for reflection and diffraction losses. The scatterers within each cluster are normally distributed around the cluster center with a standard deviation of $50 \times \lambda_c$. This aims at reflecting the intra-cluster delay distribution reported in [2-29]. This model is rather simple but precis enough since it is only to evaluate the beamforming performances. The top graph in Figure 2-21 plots the channel frequency response for the two farthest left adjacent antennas and the farthest right.

The response is strongly frequency dependent. It is not necessarily an issue since OFDM modulations are naturally robust to such dependencies. The really important metric is the SNR over the whole band. The channel response seems to be uncorrelated, even for adjacent antennas (Figure 2-21 top-right graph). This means that for linear processing, the CSI must be acquired for each antenna. This is the reason why the channel must be represented by the $N_t \times N_r$ matrix $\boldsymbol{H}$ in section 2.2.2.3. On the contrary, in the beam-shaping approach, only the AoA needs to be evaluated.

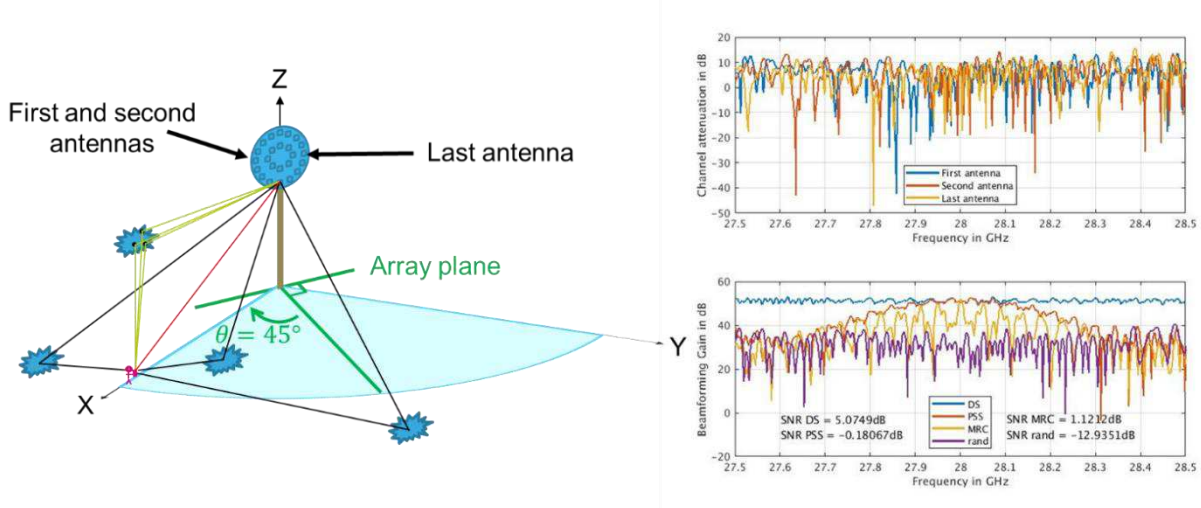

Figure 2-21: Left: multipath channel model. Top Right: Channel response for the two adjacent most left antennas and the most right one. Bottom Right: Beamforming gain for various beamforming methods

Three kind of beamforming will be examined, Delay and Sum (DS), Phase Shift and Sum (PSS), and Maximum Ratio Combining (MRC). In the following results, there is no interfering signals in the environment, such that the channel impact can be specifically observed. In this situation MRC will be

near optimal, and it can be assumed that ZF and MMSE will have the same behavior. As a sanity check, a beamforming using a random matrix with unit gains is also added, like in DS and PSS, and random phases. This will be used as a reference. To consider a beamforming algorithm is effective, it must perform significantly better than this random recombination, although it will only be used qualitatively. The beamforming gains and SNRs are reported in the bottom-right graph of Figure 2-21 for one realization of the channel. Figure 2-22 plots the SNRs for each beamforming for 100 runs.

The signal and antenna noises powers are adjusted such that the LoS beam would have a 5dB SNR under DS in an AWGN channel. Once fast fading is added to the channel model, the DS frequency response remains nearly flat with the expected $5dB$ SNR. It provides the best signal gain and SNR performances. It is important to note that the flatness, in this case, is due to the absence of scatterers on the LoS path in the model. Otherwise, the sub-path these scatterers would have created, would have change the frequency response of the LoS path. In general, a channel can only be as flat at its LoS path. DS seems to be the best here, but one must remember that the processing complexity of a time delay is very high.

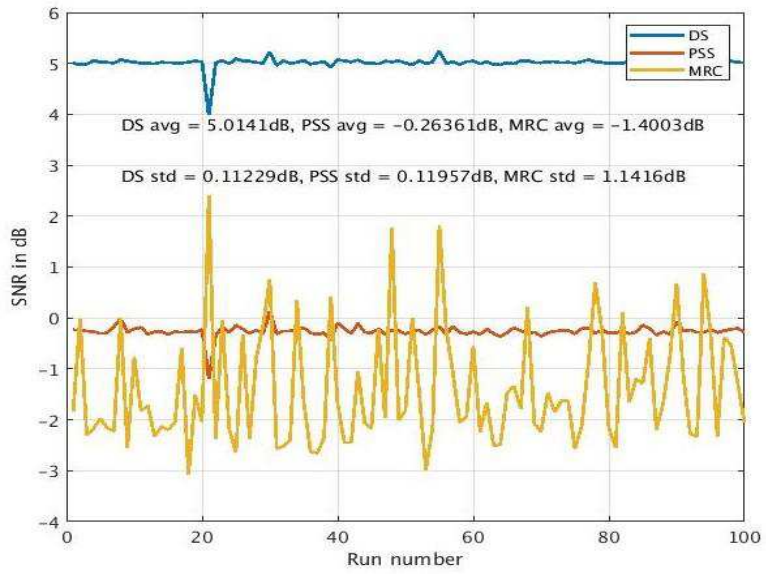

Figure 2-22: SNR for 100 channel realizations for DS, PSS and MRC beamforming

The PSS beamforming performs well in the middle of the band, but as expected from section 2.2.5.2, it suffers severely on the band edges. Interestingly the losses on the edges are limited by the performances of a random matrix, which limits the performance degradation.

Over an AWGN channel in LoS configuration, MRC should perform exactly as PSS since it results in exactly the same processing. When fast fading is added to the channel model, it can be seen that it performs worse than PSS and is the least consistent beamforming algorithm. It performs barely better than random beamforming and is clearly unusable as is.

From Figure 2-21, one can infer that the performance loss for PSS and MRC are mostly due to the frequency dependence of the phase shift required for accurate beamforming. In practice this problem is solved by splitting the channel bandwidth in sub-bands in which the narrowband approximation holds. The sub-band width is called the coherence band. In Annex 2.2, this coherence band was evaluated for PSS and MRC using the following definition. It is the largest sub-band width required to achieve an SNR within one decibel of the one delivered by DS.

To reach this level of performance PSS only needs to split the band in four sub-bands. The results for MRC vary widely with the channel characteristics and range from few sub-bands to hundreds of them.

What is interesting is that, when the sub-band width becomes small enough, MRC starts to output perform DS. While, in the example, DS always deliver a 5dB SNR, MRC can go up to 15dB SNR. This is thanks to its ability to exploit the spatial richness of the channel.

Unfortunately, this comes at the massive cost of performing CSI estimation, detection matrix processing and beamforming processing for each sub-band. This directly multiply the processing complexity by the number of sub-bands. It is likely that the performance increase brought by linear programming approaches, MRC, ZF, or MMSE, does not outweigh the processing cost. One last point going against these approaches the time for which the CSI are valid and is the subject of the next section.

2.2.5.5 Channel State Information acquisition

CSI acquisition is a complex topic. Whether it is the estimation of the AoA for beam-shaping or the channel matrix for Massive-MIMO, the task is difficult. One must scan a large solid angle to find and identify the desired users. The other need to evaluate the channel gain and phase between each user and each BS antennas, potentially for multiple bands. Evaluating which one has the higher processing complexity is hard and it will not be addressed here.

One thing that is easy to estimate is the required refresh rate of the CSI. For Beam-shaping it depends on 2 parameters, the user speed and the HPBW. While the first one cannot be adjusted by design there could be some more leeway on the second one. Assuming the CSI estimation gets the user in the middle of the beam, it is the time it takes for the user to exit the HPBW. Taking a user one meter away from the BS, moving at $50km/h$ as an extreme case, and assuming the HPBW of section 2.2.3.2 MC-UCA, the user would remain about $1.7ms$ within the beam. This time is proportional to the distance from the user to the BS. If this distance is known, it is even possible to adjust the CSI refresh rate to minimize overhead. Also, while the initial AoA estimation may be complicated, once it is acquired it may be possible to track the users with a potential reduction in processing since it is known that the future AoA will be near the previous one.

CSI acquisition refresh rate for Massive MIMO is given by the channel coherence time. It depends on the user speed and the carrier wavelength, none of which are adjustable by design. A classical estimation is $t_c = \frac{\lambda_c}{v_{UE}} \times \sqrt{\frac{9}{16\times\pi}} \sim 300us$ for a $28GHz$ carrier and a user at $50km/h$. Intuitively it corresponds to the time it takes for the user to move by about a wavelength. This CSI validity duration become a massive challenge for using a Massive MIMO approach in 5G millimeter wave small cells. As mentioned before, the purpose is to exploit the channel reciprocity by reusing the CSI acquired in uplink for downlink. What this shows here, is that this must be done in the space of few hundred microseconds. With the processing complexity exhibited in the previous section, this makes Massive MIMO a much less attractive solution compared to beam-shaping.

2.2.6 Conclusion

Beamforming is seen as an enabling technology for 5G thanks to its ability to increase the received power and SNR, to perform spatial filtering and to offer full spectrum reuse in each individual beams. Because multi-beam is mandatory only hybrid or fully digital architectures are possible. Performances depend on many intricate factors such as number of antennas, array topology, beamforming algorithm, hardware imperfections or RF channel properties. Because of the frequency dependence of the channel, the complexity of linear processing is likely to be too high for a practical implementation.

The choice is then to be made between a beam-shaping approach using time delays and phase shifts over four sub-bands. If the choice must be made on processing complexity the second option is probably a better solution. Fortunately, it will be shown that the architecture of the receiver, and of the ADC in particular for a DBF approach, can provide a solution combining the performances of time delay and the low processing complexity of phase shifting.

2.3 NETWORK ARCHITECTURE

The network is composed of all the base stations, the cells they are covering and the connection with the core network through backhaul links. Millimeter wave small-cells have the potential to provide high throughput but have strong limitations in terms of coverage. First, the 5G deployment strategy to overcome this limitation will be looked at. Second, the proposed solution to answer the challenge of backhaul connection will be described. Finally, the main characteristics of a small cell will be drafted out.

2.3.1 Heterogeneous Network

To overcome the coverage limitation of small cells, 5G envision the coexistence of sub-6GHz macro-cells and millimeter wave small-cells in a heterogeneous deployment to provide both high throughput and coverage. This is depicted in Figure 2-23 where the macro-cell radius d_{MC} is much larger than the small-cell radius d_{SC}.

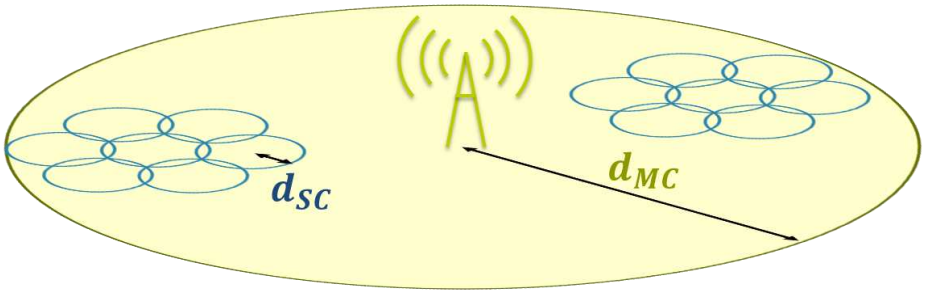

Figure 2-23: 5G Network deployment strategy

The idea is that the higher throughput allowed by millimeter wave small-cells is only required in the densest part of the cities. It is necessary to deploy small-cells only in these specific areas while the wider areas with lower user density can be covered using sub-6GHz macro-cells. It is likely that the small-cells will not be fully standalone but slaves to the nearest Macro-cell and that a significant portion of the scheduling will be done at the Macro-cell BS (M-BS). The Small-cell BS (S-BS) could be installed on streetlights, where power is readily available, to reduce deployment costs. This could go along with a campaign of replacing streetlamps with LEDs, allowing for more available power for the S-BS. The limiting factor would be their connection to the core network, the backhaul.

2.3.2 Wireless Backhauling

Because of the dense deployment of S-BS, a wired connection of each of them to the core network would have a prohibitive cost. One way to solve this problem is to use wireless backhaul. In particular using the same band as for user access at $28GHz$, since it is already available, is very interesting. Data could them be relayed to and from an Anchor BS (A-BS) with higher backhaul capabilities such as millimeter wave E-band wireless or wired connections.

This is possible because the inter-site distance is small, and each S-BS being surrounded by 6 neighbors in average, the likelihood of LoS configurations with one or more of its neighbors is high. Also, because S-BS are fixed it is also possible to position them and adjust the environment to favor these LoS conditions. A potential deployment scenario is depicted in Figure 2-24.

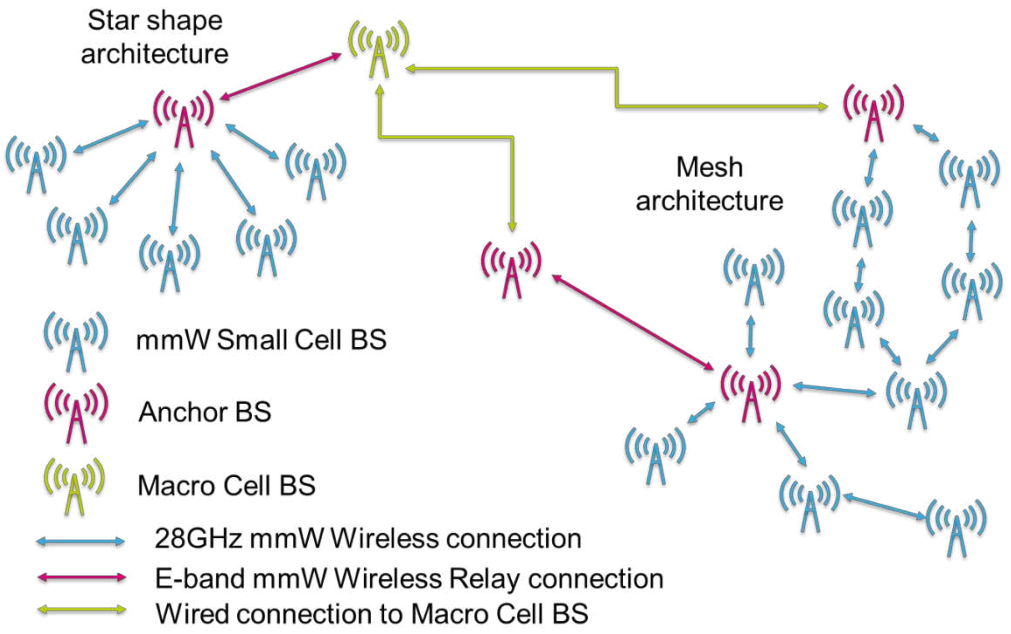

Figure 2-24: Potential deployment scenario for heterogeneous network and backhaul

The backhaul is achieved through multiple hops from S-BS to S-BS until reaching an A-BS or the M-BS forming a backhaul network. One last interesting possibility is that is can be made reconfigurable by using the beamforming capabilities of the S-BS. This would increase the network robustness to the failing of one S-BS or to absorb local and short traffic excess through multiple parallel backhaul routes to the M-BS.

2.3.3 Small cell architecture

Here the analysis will go deeper into the system and start looking at the different parameters of a small cell itself, meaning its radius, the number of sectors, and some characteristics of one sector. The focus will be on small cell operating at 28GHz, making no difference between the A-BS and the S-BS. The additional backhaul capacities of the A-BS, through E-band or wired connections, is out of the scope of this manuscript.

2.3.3.1 Small cell radius

There are two things that can constrain the cell radius. The first one is the link budget. This will be studied in the next section, proving not to be the limiting factor. The second one is the LoS probability. As shown before, for the beam-shaping proposed approach a LoS configuration is highly desirable. It is not a showstopper per say, receiving the strongest None Line of Sight (NLoS) path could be possible but it would have a significant toll on the link budget and potentially temper with the in-band frequency flatness after beamforming.

Several measurement campaigns were made throughout the world to characterize the 28GHz channel and in particular the LoS probability as a function of the distance from the user to the BS. Unfortunately, the amount of measurements for short distances, below $50m$, is very small. One such study is proposed by the authors of [2-29], where they performed measurements down to 20m for BS sitting at height of 7m. This is realistic if a S-BS deployment on streetlamps is considered. Since these results are based on a single measurement campaign, the numbers presented below are to be taken with a grain of salt. Nonetheless, it is enough to get a meaningful order of magnitude. Their proposed model for LoS probability $P_{LoS}(d)$ as a function of the distance d between the user and the BS will be used. It is given in equation (2.20):

$$P_{LoS}(d) = \min\left(\frac{d_1}{d}, 1\right) \times \left(1 - e^{-\frac{d}{d_2}}\right) + e^{-\frac{d}{d_2}} \tag{2.20}$$

Where d_1 and d_2 are fitting parameters extracted from measurements. Here, the parameters extracted from [2-29] measurements will be used, $d_1 = 24m$ and $d_2 = 45m$. The next thing to be known is the probability of LoS configuration for a user with a uniform probability of location in the cell as a function of the cell's radius d_{SC}.

$$P_{LoS_{Cell}}(d_{SC}) = \begin{cases} 1 & if\ d_{SC} < d_1 \\ \int_0^{2\times\pi} \int_0^{d_{SC}} P_{LoS}(d) \times P_{UE} \times d \times \partial d \times \partial\theta & if\ d_{SC} \geq d1 \end{cases} \tag{2.21}$$

The derivation of the case $d_{SC} \geq d1$ is provided in Annex 2.3 and the result is given bellow:

$$\begin{aligned} P_{LoS_{Cell}}(d_{SC}) = {} & \frac{d_1^2}{d_{SC}^2} \\ & \times \left(1 + 2 \times \left(\frac{d_{SC}}{d_1} - 1\right) \times \left(1 - \frac{d_2}{d_1} \times e^{-\frac{d_{SC}}{d_2}}\right)\right. \\ & \left. - 2 \times \left(\frac{d_2}{d_1}\right)^2 \times \left(e^{-\frac{d_{SC}}{d_2}} - e^{-\frac{d_1}{d_2}}\right)\right) \end{aligned} \tag{2.22}$$

It is not very insightful, so it is plotted in Figure 2-25 for d_{SC} going from $0m$ to $100m$ together with $P_{LoS}(d)$. As expected, both decrease as the distance or radius increase, but the cell LoS probability goes down slower. A high probability of LoS configuration is desired, ideally higher than 95%. Here, that would require a cell of about $35m$. That would be very small and require many small cells to cover large areas. Cells up to $50m$ radius will be considered. While keeping the LoS probability above 80% it requires twice as less cells to cover the same area. Also, the assumption of uniform distribution of the users is a worse case. In real life people are more likely to be at specific locations such as near a bus stop or a street bench, or along the sidewalk. It is then possible to adjust the S-BS position and orientation to account for this non-uniform distribution and optimize the LoS probability of the cell.

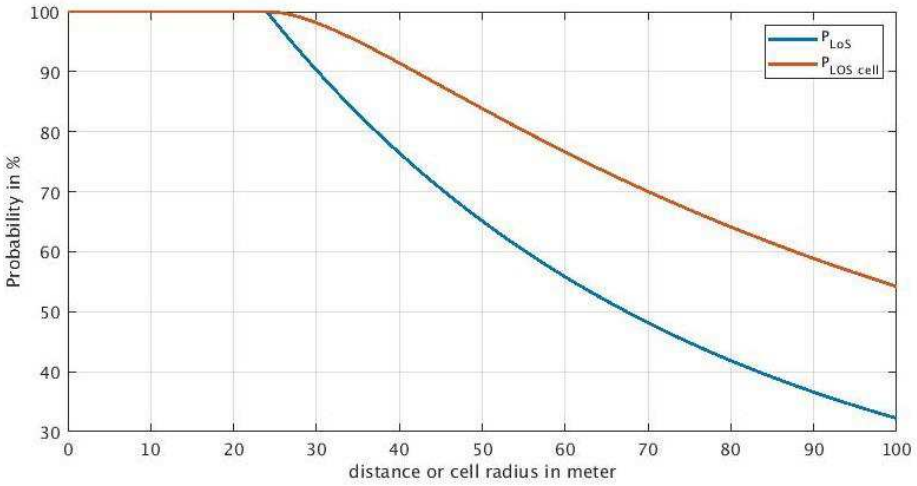

Figure 2-25: LoS probability for a user at distance d and average LoS probability in a cell of radius r

2.3.3.2 Small cell sectors

The antenna array considered are made of patch antennas on a ground plan. This means that it can radiate at best only in the hemisphere in front of it, the ground plan blocking any rear radiation. In these

conditions a circular cell needs to be split at least into two sectors. In practice, as it was shown in section 2.2.5.1, patch antennas have a HPBW of less than 180°. To have only three sectors it is required to have an HPBW of 120°. While some patch antennas can have such a wide view angle most of them are below 100°. A split in four sectors, requiring a HPBW of 90°, is a good compromise between limiting the number of sectors and acceptable requirements for the antennas.

Assuming a hexagonal deployment of S-BS, as described in Figure 2-26, one sector would be able to establish backhaul link with at most two neighboring S-BS. Under these assumptions, the number of beams dedicated to backhaul can be limited to two per sector with a total of six backhaul beams for the whole cell. Because those beams are of the same nature as the ones for the users, there is in practice no reason to have them dedicated to backhaul. The beam allocation can remain flexible. The total number of beams per sector should be adjusted such that the capacity of one beam times the number of beams matches 5G target area throughput. The next step is therefore to evaluate the capacity of a single beam. This is done through the analysis of the link budget.

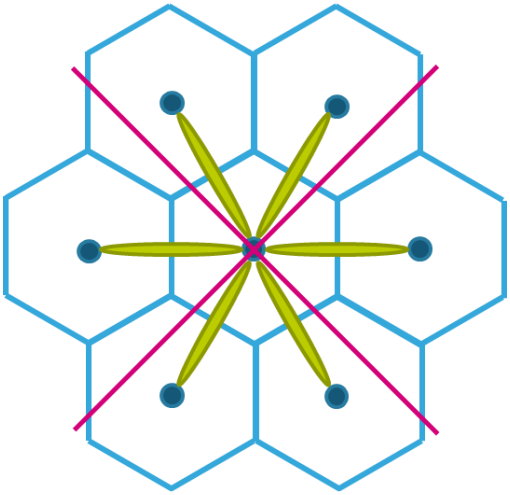

Figure 2-26: Distribution of backhaul links per sector for a cell division in four

2.4 LINK BUDGET

The link budget is an important tool in RF system design. It allows to put into perspective the first order impact of the design variables on the link performances. It summarizes the Transmitter (Tx) and Receiver (Rx) hardware contribution as well as the channel. As already shown, a beam-shaping DS approach give a flat frequency response and the expected beamforming gain of $10 \times \log_{10}(N_{ant})$. Thanks to that, the channel can be approximated using a simple AWGN model. In that case, the Friis equation becomes very handy. It is then possible to evaluate the interaction between the design variables.

2.4.1 Friis law

Friis law gives the simple relationship of (2.23) between the received power P_r and the transmit power P_t of a wireless link:

$$P_r = \left(\frac{\lambda}{4 \times \pi \times d}\right)^2 \times A_r \times A_t \times P_t \qquad (2.23)$$

With A_r and A_t the antenna gains of the receiver and the transmitter, d the Tx-Rx distance, and λ the carrier wavelength. This version is valid for single antenna receiver and transmitter and in the case where the peak antenna directivities are facing each other. In a beamforming system this is a reasonable assumption. This equation needs to be adjusted to the desired use case, i.e. a user with a near

omnidirectional antenna and a BS with an array of N_{ant} patch antennas. (2.23) can then be extended in (2.24) to account for this change:

$$P_r = \left(\frac{N_{ant} \times \lambda}{4 \times \pi \times d}\right)^2 \times A_r \times A_t \times P_t \tag{2.24}$$

Where A_r now represent the gain of a single receiving antenna. It is common to express this relationship in decibels per mW (dBm) as follow:

$$P_{r_{dBm}} = P_{t_{dBm}} + 10 \times \log_{10}\left(\left(\frac{\lambda}{4 \times \pi \times d}\right)^2 \times A_r \times A_t\right) + 20 \times \log_{10}(N_{ant}) \tag{2.25}$$

This equation can now be used to evaluate the impact of the design variables on the link budget.

2.4.2 Link budget sensitivity to design variables

The considered link budget is based on the received Signal to Interferer plus Noise Ratio (SINR). As it will be shown, these interferers are coming from the limited Interferer Rejection Ratio (IRR) of the beamformer. Additionally to the IRR, the SINR will be evaluated as a function of the Power Amplifier (PA) output power P_t, the Tx-Rx distance d, the thermal noise power N_{th} at the antenna level, the receiver's Noise Figure NF, the number of receiving antennas N_{ant} and the channel bandwidth B. Equation (2.25) already gives the relationship between all these variables except for N_{th}, NF and IRR. First, the noise related variables will be studied, then the equation for SINR will be derived.

2.4.2.1 Antenna thermal noise power

The thermal noise power of a resistor is given by (2.26):

$$N_{th} = \frac{v_n^2}{R} = \frac{4 \times k_b \times T \times R}{R} \times B = 4 \times k_b \times T \times B \tag{2.26}$$

With v_n^2 the mean square noise voltage, k_b the Boltzmann constant, T the temperature of the resistor, R the value of the resistor and B the bandwidth of interest. This formula holds for an antenna except that the temperature is not the one of the antennas itself but rather the average temperature of where it is looking. In some system such as satellite communications, this can have an impact but for the terrestrial mobile network considered, it is a reasonable assumption to use an ambient temperature $T = 290K$.

Assuming this antenna is loaded by a noise-less receiver with the matched input impedance R, the effective noise voltage seen by the load goes through a voltage divider:

$$v_l = \frac{\sqrt{v_n^2} \times R}{R + R} = \frac{v_n}{2} \tag{2.27}$$

The effective noise power delivered to the receiver is then:

$$N_{th_{eff}} = \frac{v_l^2}{R} = \frac{4 \times k_b \times T \times R}{4 \times R} \times B = k_b \times T \times B \tag{2.28}$$

It is important to remember that this classic result is true only under the assumption of a power matching between the antenna and the load. It is a common practice, called noise matching, to optimize the load impedance to minimize the noise instead of maximizing the power transfer. For simplicity, the hypothesis of power matching will be kept for this analysis.

2.4.2.2 Noise Figure

NF is a figure of merit that measures the degradation of the source thermal noise power, the receiving antenna in the present case, by the receiver's intrinsic noise. By definition, it is the ratio of the total

input referred noise power $N_{tot_{IR}} = N_{th_{eff}} + N_{Rx_{IR}}$ to the effective antenna noise power $N_{th_{eff}}$ at the reference temperature $T_0 = 290K$, under the hypothesis of power matching. The NF can then be expressed as:

$$\begin{aligned} NF &= 10 \times \log_{10}\left(\frac{N_{tot_{IR}}}{N_{th_{eff}}}\right) = 10 \times \log_{10}\left(\frac{k_b \times T_0 \times B + N_{Rx_{IR}}}{k_b \times T_0 \times B}\right) \\ &= 10 \times \log_{10}\left(1 + \frac{N_{Rx_{IR}}}{k_b \times T_0 \times B}\right) \end{aligned} \tag{2.29}$$

A noiseless receiver will have a $0dB$ NF and if the input matching is done using a real resistor the noise contribution of the receiver is the same as the antenna and the NF is $\sim 3dB$. While this definition is useful to understand the NF physical meaning it is not very convenient to use with multiple antenna systems. For that purpose, the formulation from equation (2.30) will be used.

$$\begin{aligned} NF &= SNR_{IN_{dB}} - SNR_{OUT_{dB}} \\ &= SNR_{SRx\,in_{dB}} + 10 \times \log_{10}(N_{ant}) \\ &\quad - \left(SNR_{SRx\,out_{dB}} + 10 \times \log_{10}(N_{ant})\right) \\ &= SNR_{SRx\,in_{dB}} - SNR_{SRx\,out_{dB}} \end{aligned} \tag{2.30}$$

It is the difference between the input and the output SNR. Equation (2.6) provides the SNR of an ideal beamforming receiver, meaning it adds no noise and performs ideal beamforming. This can be seen as the multiple antenna system input SNR. The output SNR is almost the same, the difference being that the SRx output SNR is degraded by the SRx internal noise. Equation (2.30) assumes the SRxs internal noises are uncorrelated. In that case the number of antennas of the array has the same effect on the input and output SNR, and the whole receiver's NF is the same as the single receivers' NF.

2.4.2.3 Signal to Interferer plus Noise Ratio

To evaluate the SINR, the interferers of interest are the in-band ones. In the context of multi-user Massive-MIMO, where a given band is locally operated by a single operator, these interfering signals, for a given user, are the other users remaining signals, after they underwent the spatial filtering from beamforming. Hence, the Interferer Rejection Ratio (IRR) of a beamformer is defined as the minimum attenuation it applies on signals outside the main beam.

In a first step, SNR will be evaluated. The received power is available from (2.25) and the total noise power at the receiver can be found by reversing (2.29):

$$\begin{aligned} SNR_{dB} &= 10 \times \log_{10}\left(\frac{\left(\frac{N_{ant} \times \lambda}{4 \times \pi \times d}\right)^2 \times A_r \times A_t \times P_t}{N_{ant} \times k_b \times T_0 \times B \times 10^{\frac{NF}{10}}}\right) \\ &= \left(P_{t_{dBm}} + Att_{Ch}\right) - \left(N_{th_{dBm}} + NF\right) + 10 \times \log_{10}(N_{ant}) \end{aligned} \tag{2.31}$$

With $Att_{Ch} = 10 \times \log_{10}\left(\frac{\lambda^2 \times A_r \times A_t}{(4 \times \pi \times d)^2}\right)$ the channel attenuation and $N_{th_{dBm}} = 10 \times \log_{10}\left(\frac{k_b \times T_0 \times B}{1mW}\right)$ the antenna thermal noise power in decibels per mW (dBm).

The first parenthesis is the received signal power, the second the total noise of a single receiver, with the difference of the two giving the SNR of a single receiver. The last terms correspond to the dependence of the total SNR to the number of antennas at the receiver, which is as expected.

It is now necessary to take the interferers into account and evaluate the Signal to Interferer plus Noise Ratio (SINR). There are two sources of interferences the $N_{beam_{UE}} - 1$ other user beams and the $N_{beam_{BH}}$ backhaul beams. Here, it is assumed that the user beams all use the same SNR_{UE} and the

backhaul beams the SNR_{BH}. Similarly, both beam types are assumed to have different Interferer Rejection Ration (IRR) respectively IRR_{UE} and IRR_{BH}. This is because the backhaul beams come from a fix known location. This allows for pre-processed null-steering hence better IRR. The user beams $SINR_{UE}$ can be expressed as:

$$SINR_{UE} = \frac{P_{r_{UE}}}{N_{Rx} + \left(N_{beam_{UE}} - 1\right) \times P_{r_{UE}} \times IRR_{UE} + N_{beam_{BH}} \times P_{r_{BH}} \times IRR_{BH}} \tag{2.32}$$

Where $N_{Rx} = N_{ant} \times k_b \times T_0 \times B \times 10^{\frac{NF}{10}}$ is the total Rx input referred noise power. In theory this input referred noise is different for each beam. But, as long as the beamforming coefficients have an amplitude close to one in average, it will actually be fairly independent from the beam considered. For the sake of simplicity, all beams will be assumed to have the same input referred noise power N_{Rx}.

Dividing the top and bottom by N_{Rx} allow to express the $SINR_{UE}$ as a function of the different SNR.

$$SINR_{UE} = \frac{SNR_{UE}}{1 + \left(N_{beam_{UE}} - 1\right) \times SNR_{UE} \times IRR_{UE} + N_{beam_{BH}} \times SNR_{BH} \times IRR_{BH}} \tag{2.33}$$

Similarly, the backhaul beam $SINR_{BH}$ can be expressed as:

$$SINR_{BH} = \frac{SNR_{BH}}{1 + N_{beam_{UE}} \times SNR_{UE} \times IRR_{UE} + \left(N_{beam_{BH}} - 1\right) \times SNR_{BH} \times IRR_{BH}} \tag{2.34}$$

Re-arranging (2.34) SNR_{BH} can be expressed as function of $SINR_{BH}$ and SNR_{UE}:

$$SNR_{BH} = \frac{\left(1 + N_{beam_{UE}} \times SNR_{UE} \times IRR_{UE}\right) \times SINR_{BH}}{1 - \left(N_{beam_{BH}} - 1\right) \times SINR_{BH} \times IRR_{BH}} \tag{2.35}$$

Injecting (2.35) in (2.33) and re-arranging the SNR_{UE} can be expressed as a function of both $SINR$ (see Annex 2.4).

$$SNR_{UE} = \frac{SINR_{UE} \times K_{BH}}{1 - \left(N_{beam_{UE}} \times K_{BH} - 1\right) \times SINR_{UE} \times IRR_{UE}} \tag{2.36}$$

With:

$$K_{BH} = \frac{1 + SINR_{BH} \times IRR_{BH}}{1 - \left(N_{beam_{BH}} - 1\right) \times SINR_{BH} \times IRR_{BH}} \tag{2.37}$$

The expression for the backhaul beam SNR_{BH} is very similar:

$$SNR_{BH} = \frac{SINR_{BH} \times K_{UE}}{1 - \left(N_{beam_{BH}} \times K_{UE} - 1\right) \times SINR_{BH} \times IRR_{BH}} \tag{2.38}$$

With:

$$K_{UE} = \frac{1 + SINR_{UE} \times IRR_{UE}}{1 - \left(N_{beam_{UE}} - 1\right) \times SINR_{UE} \times IRR_{UE}} \tag{2.39}$$

This now provides all the tools required to evaluate the system performances. The next step is to set the design parameters to achieve the desired performances.

2.5 SYSTEM SIZING AND CAPACITY

The first step is to size the system to achieve the area throughput of 5G KPI. Then, it will be evaluated how it complies with the other KPIs, specifically user density, average and peak rate for uplink and downlink. Last, the question of wireless backhaul will be studied.

2.5.1 System sizing

Let us estimate the requirement in SINR. The target area throughput for eMBB is $10Mb/s/m^2$ for downlink, and half of that for uplink. Assuming a $50m$ small cell radius, the total small cell throughput is around $80Gb/s$ for downlink and $40Gb/s$ for uplink. Using TDD with equal uplink and downlink time and allocating half of the uplink time for pilot transmission, the instantaneous cell throughput must be $160Gb/s$ or $40Gb/s$ for one sector. It is also required to estimate the number of beams per antenna array. For uplink, to remain within the Massive-MIMO approximation, the number of transmitting antennas must be small compared to the receiving antenna array number of elements. In the case of multi-user Massive-MIMO, where users are equipped with a single antenna, the number of beams is equal to the number of transmitting antennas. In section 2.2 it was shown that the S-BS antenna array will be made of few hundreds of elements. To guaranty an order of magnitude between the number of transmitting and receiving antennas, the number of beams dedicated to users will be fixed to $N_{beam_{UE}} = 10$ per sector. This gives a data rate of $4Gb/s$ per beam. Using (2.3), an $SINR_{UE}$ of $11.8dB$ is obtained.

The system sizing, namely fix the number of S-BS antenna, will be based on the following hypothesis. From [2-31] it is known that an NF of $10dB$ for the full single antenna receiver is a reasonable assumption. It was also assumed the array PSL is below $-20dB$. From section 2.2.5.3, it is known that a rejection much better than $30dB$ cannot be expected. Three assumptions are made here: First, the average interferer rejection ratio is $IRR_{UE} = -25dB$. Second, all individual antennas are the same and have a gain of $4dBi$. Third, at cell edge, the HPBW is large enough to cover multiple users. To serve multiple users within one beam, the $1GHz$ band is split in ten channels of $B_{Ch} = 100MHz$ each, in order to use Frequency Division Multiple Access (FDMA). In that configuration one cell can serve ten users per beam times ten beam times four sectors for a total of $NB_{UE} = 400$. The number of backhaul beams per sector is $N_{beam_{BH}} = 2$. The last parameter to be evaluated is the achievable IRR_{BH}. The evaluation is based on the following argument:

- The backhaul beams are different from users' beams in two points. First, they are produced by a large antenna array, therefore the radiated signal is highly directional and the multi-path other than the LoS one will have negligible level of power. This means to improve IRR_{BH}, one needs only to null-steer one zero in the LoS path. The second difference is that the backhaul beams come from fix and known directions. This reduces the amount of processing to be done to perform the required null-steering. It may also allow for some pre-processing.
- It was shown, in section 2.2.5.3, that such a null-steering approach would require a stringent calibration of the system, in particular the timing accuracy. It was also noted that the main beam was largely unaffected by timing and gain errors. Because the backhaul link comes from another S-BS having the same intrinsic quality of main beam, it can be used to provide a relatively strong reference signal. This known strong signal coming from a fix far away location can be used as a reference plane wave to calibrate the antenna array, and the timing in particular. This reference signal could be sent regularly at a slow rate, maybe once per second or per minute depending on what is needed. This can provide a continuous calibration tracking for power supply and temperature variations, with minimal overhead while guarantying performances at all times.
- Ultimately what limits the IRR_{BH} is the bandwidth. In section 2.2.5.2, on the right graph of Figure 2-18 the effect of wide band signals on the spatial transfer function can be seen. One of the observations is that, even with an ideal time delay approach, for a bandwidth of $1GHz$, the null depth is limited to about $-40dB$. This will be assumed to be the best achievable IRR_{BH}.

Using (2.31) and (2.36) the required user PA output power is plotted for a $50m$ range, as a function of the number of antennas at the S-BS to reach an $SINR_{UE}$ of $11.8dB$ for various $SINR_{BH}$. (Figure 2-27). Modern PA's at $28GHz$ often provide output powers beyond $16dBm$ even in Complementary Metal Oxide Semiconductor (CMOS) technologies [2-32]-[2-34]. It is common to use PA's at 6dB back off for linearity reasons. In order to satisfy the link budget, with a PA output power of $10dBm$, for an $SINR_{BH} \leq 33dB$, less than 15 antennas are necessary. For $SINR_{BH}$ of $35.62dB$ and $35.74dB$ the requirement is respectively of 256 and 2048 antennas. Clearly there is a point beyond which increasing the $SINR_{BH}$ is highly detrimental for the system.

It will be further assumed $SINR_{BH} = 33dB$. The previous argument neglects that, for backhaul, the signal sees interferences both at transmitting and receiving ends. In the worst case this leads to a degradation of $3dB$, providing an $SINR_{BH_{min}} = 30dB$. For $SINR_{UE}$ calculations, the $SINR_{BH} = 33dB$ must be used and, for backhaul data rate it is the $SINR_{BH_{min}} = 30dB$. The backhaul rate is then $5Gb/s$ per beam. For six beams that is a total of $30Gb/s$ of backhaul maximum capacity in one direction, uplink or downlink. The surrounding S-BS being at fix location there is no need for pilot. This makes both directions perfectly symmetrical.

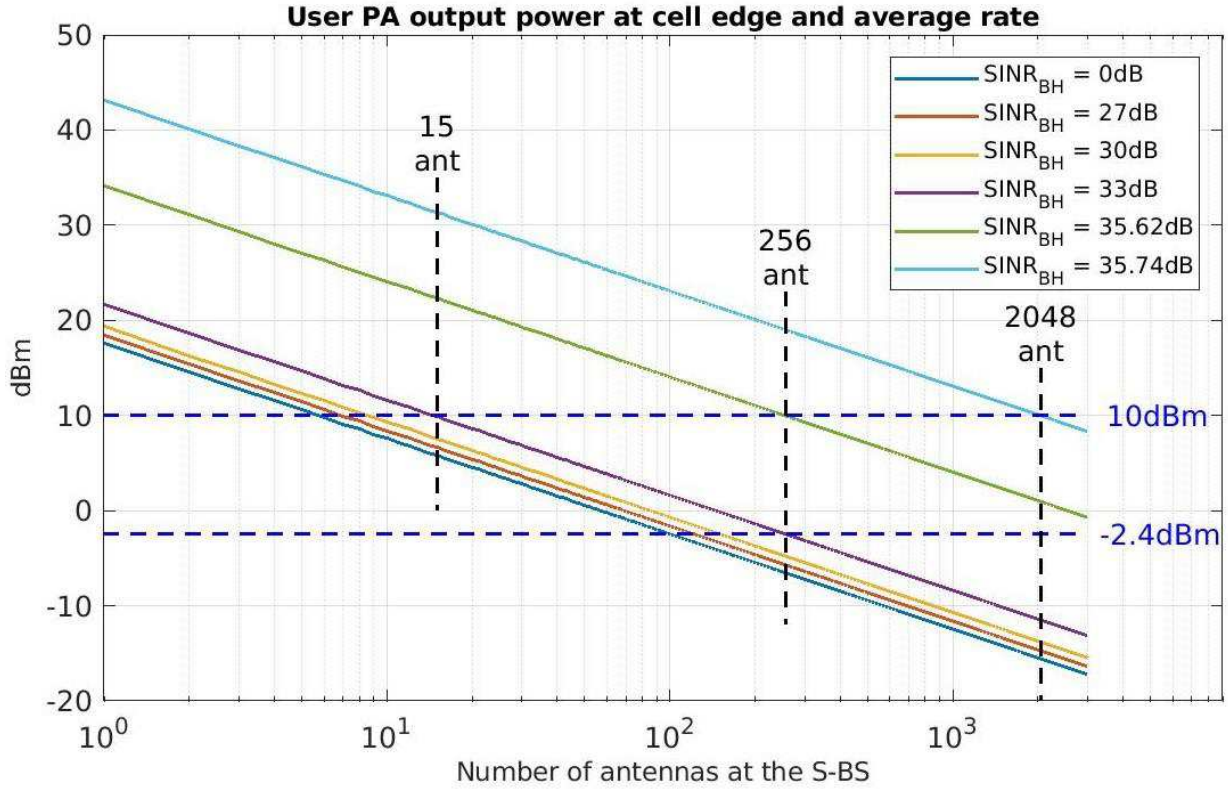

Figure 2-27: Users' PA output power at cell edge and average rate as a function of the number of receiving antennas for various $SINR_{BH}$

In these conditions only 15 antennas are required at the S-BS arrays to satisfy the user link budget. Because many more antennas are needed to have a thin enough beam, the number of antennas is primarily constrained by the HPBW and PSL. From the conclusion in section 2.2.3.2, a system with 256 antennas will be considered. The required PA output power is then $-2.4dBm$ providing more than $12dB$ of margin on the user's PA output power.

2.5.2 Compliance to 5G KPIs

All the design parameters of the system have now been fixed. In a hexagonal deployment of $50m$ radius small cells, each of them is divided in four sectors equipped with an antenna array with 256 elements. Each array can produce ten beams for user communication and two beams for backhaul. The next step is to consider the capacities of this system and compare them with the KPIs in [2-35].

1. User density

By splitting the $1GHz$ band in ten channels of $100MHz$ each to use FDMA, the maximum number of users per cell is $NB_{UE} = 400$. This corresponds to a user density of $50\,000\ user/km^2$ for a $50m$ radius cell. This is above the high user density scenario of $10\,000\ user/km^2$.

2. Uplink Average rate

5G's target average rate is $50Mb/s$ in uplink. The proposed system is design to provide the same performances across the cell. The average rate can be assimilated to the minimum rate when the S-BS is at maximum user density, i.e. the capacity of one $100MHz$ sub-channel. Such a channel has a capacity of $100Mb/s$ in uplink. This is twice the average rate targeted for 5G.

3. Downlink average rate

The previous analysis is valid only for uplink. The beamforming process in downlink is a different problem. In uplink the S-BS sees users' beams at the same power level thanks to power control. This means that each beam sees a similar level of interference from the other beams. In comparison with downlink the S-BS must radiate different power levels in each beam in order for the users to receive the same SINR, regardless of their location in the cell. A user close to the S-BS will see a large interference from a user at cell edge and that user at cell edge will not experience any significant interference from the close user.

This difference will probably lead to choose a different way of beamforming involving some null-steering. In contrast with uplink there is no "easy way" of calibration which shows that these approaches are challenging. For the same reasons as for backhaul, the downlink channel will have only a LoS path. One important consequence is that the channel response will be flat.

ZF would require splitting the 1GHz band in sub-bands. Thankfully only a limited number of sub-bands are necessary. There are two reasons for this. First, the LoS configuration with beamforming at the TX side ensure a flat frequency response of the channel. Second, the frequency response in the main beam varies slowly when using phase shift beamforming (Figure 2-17 right). The use of sub-bands is necessary only to alleviate this slow variation since the channel itself is flat. This limits the processing complexity increase. One natural choice would be ten $100MHz$ sub-channel. The processing remains acceptable, in particular since the validity of the CSI, i.e. the LoS direction of each user, is long lived compared to a fast fading channel. This allows for more time to do the beamforming processing.

Even with the challenge of calibration and increase complexity, the assumption is made that it is possible to achieve the same SINR performances in downlink as in uplink and the same equation will be used as a first order approximation. 5G's target average rate is $100Mb/s$ in downlink. The proposed system is design to provide the same performances across the cell. The average rate can be assimilated to the minimum rate when the S-BS is at maximum user density, i.e. the capacity of one $100MHz$ sub-channel. Such a channel has a capacity of $200Mb/s$ in downlink. This is twice the average rate targeted for 5G.

4. Uplink peak rate

The peak rate KPI for 5G is $10Gb/s$ in uplink. The peak-rate of the proposed system is evaluated assuming the following scenario. There is only one user in each sector, meaning no in band interferences other than the one from backhaul beams. Each user is using the whole $1GHz$ available bandwidth in his sector. Users are $25m$ away from the S-BS, and the UE is equipped with a $16dBm$ saturation power (P_{sat}) PA working at $6dB$ back off.

The available SNR is then $25dB$ giving an instantaneous peak rate of $8Gb/s$. Assuming, as before, a TDD system with equal uplink and downlink time and half of the uplink time dedicated to pilot transmission, the effective uplink falls to $2Gb/s$. This is well below the target of $10Gb/s$ of peak rate for uplink. In fact, in this TDD configuration, the instantaneous data rate of the beam must be $40Gb/s$ to reach target peak rate. Using $1GHz$ of bandwidth, it would require an SINR of $120dB$. This is neither near any current nor future technology. Thankfully 5G will be a heterogeneous network, all the performances are not to be provided only through the 28GHz band. Achieving peak rate will require a significant contribution from higher millimeter wave bands where even more spectrum is available.

5. Downlink peak rate

The downlink peak rate target is $20Gb/s$. It corresponds to the same $40Gb/s$ instantaneous rate since, in this scenario, the full uplink time slot is used for data transmission. This also requires the unreachable 120dB SNR. But the problem is not fully symmetrical. In uplink the achievable $SINR_{UE}$ is mostly limited by the maximum PA output power. In the downlink the limitation comes from the backhaul beam interferences. With an infinite PA output power, the downlink $SINR_{UE}$ is limited to about 43dB in the presence of two backhaul beams of $SINR_{BH} = 33dB$ and an $IRR_{UE} = IRR_{BH} = -40dB$. This IRR_{UE} assumption is reasonable since in this configuration the user beam is the same as the backhaul beams. For a PA total output power of 10dBm, accounting for user and backhaul beams, the $SINR_{UE}$ drops only 1dB at 42dB and the corresponding peak rate is $7Gb/s$. It is important to note that such a high $SINR_{UE}$ will be challenging to deal with at the UE side, in particular in terms of Dynamic Range (DR).

Table 2-1: 5G KPIs and System performances summary

	5G KPI	Proposed system				
		Performances	SINR	PA inst power	Cell throughput	Cell PA avg output power per bit per second
User density	$10\ ku/km^2$	$50\ ku/km^2$	N/A	N/A	N/A	N/A
Average rate UL	$50Mb/s$	$100Mb/s$	$11.8dB$	$-2.4dBm^*$	$40Gb/s$	$1.4\ pW \cdot b^{-1} \cdot s^{**}$
Average rate DL	$100Mb/s$	$200Mb/s$	$11.8dB$	$-7.1dBm^{**}$	$80Gb/s$	$0.7\ pW \cdot b^{-1} \cdot s^{**}$
Peak rate UL	$10Gb/s$	$2Gb/s$	$25dB$	$10dBm$	$8Gb/s$	$2.4\ pW \cdot b^{-1} \cdot s$
Peak rate DL	$20Gb/s$	$7Gb/s$	$42dB$	$10dBm$	$28Gb/s$	$177\ pW \cdot b^{-1} \cdot s$
Backhaul	N/A	$5Gb/s$	$30dB$	N/A	$30Gb/s$	N/A

* At cell edge. ** At maximum user density with users uniformly distributed.

Table 2-1 summarizes 5G's KPIs and the system performances (details in Annex 2.4). Only the target peak rates cannot be reached. The analysis shows that larger bandwidths are necessary and that it will require the contribution of higher millimeter wave bands. In terms of user access, it can provide the desired area throughput, but the wireless backhaul is not able to relay the data if the cell is at maximum

capacity. This means that the area throughput can be achieved only locally, on an area smaller than a small-cell.

The interesting point is that providing the average rate to the maximum user density turn out to be very power efficient and requires PA output power below $0dBm$. As already mentioned, an S-BS would be highly power constrained, within few hundred Watts. Focusing on the downlink, if $10dBm$ PAs are used, making the simplistic assumption of a 10% drain efficiency, the power consumption of the PAs alone will be around $100W$ for the four sectors of one S-BS. Using 0dBm PAs would divide this consumption by ten, assuming the PAs are properly optimized. In that scenario only the peak rate in downlink would be affected, and it would only be reduced by about 11%.

This example shows that there is some leeway in the S-BS power budget allocation increasing the feasibility likelihood of such a system.

2.5.3 Conclusion

The proposed system provides performances answering a significant portion of 5G KPIs when not covering them entirely. It is very efficient to provide lower data rate to a large number of users, fitting the use case of dense urban areas. The major outcome is that the hardest challenge is about backhaul when targeting 5G's area throughput. Due to the high number of small-cells, using wires to connect them to the core network would be very expensive. In band wireless backhaul appears like a natural and cheap solution but it was shown that it limits the system area throughput at a larger scale. More generally this enlighten the fact that the area throughput KPI is the most ambitious 5G objective.

The purpose of this manuscript is to study the S-BS receiver. S-BS and UE PA output power were only considered to ensure global feasibility. The conclusion is that the proposed system does not see any technological limitations from the PA performance requirements. The most challenging part will likely be on input DR for the UE, and on the S-BS transmitter calibration as well as on the required digital processing power. While these challenges can be technologically addressed today, they might be unreachable while staying within the power budget limit. Thankfully all these challenges are linked to high $SINR$. It was shown that there is leeway in easing them with acceptable performance reduction.

2.6 MULTIPLE OPERATOR SCENARIO

So far, the analysis was assuming a single operator scenario and was taking only In Band Interferers (IBI) into account. To be realistic, this scenario would require a cooperation of the operators to deploy a seemingly "Single Operator Network". This would benefit to the network in many ways. The deployment costs would be shared and therefore reduced. This would also benefit to the deployment speed and the network coverage. Finally, as it will be seen, it would also significantly reduce the constraints on hardware.

Despite all these economical and technical benefits this kind of cooperation has not been seen until today, probably because it makes the biggest operators loosing significant competitive advantage on the smaller ones. Also, it requires a cooperation of all the operators. This could probably be achieved only through regulation, but again it would require some kind of cooperation of the governments for this to enter an international standard such as 5G.

For these reasons 5G's deployment will assumed to be non-cooperative multiple operators. To perform the analysis of such a network, the worst case for a two-operator scenario will be first established. The consequences from that scenario will be derived to get an estimation of the Out of Band Interferers (OoBI).

2.6.1 Worst case scenario in a two-operator environment

Thanks to the high beamforming capability of the S-BS the radiated power will be only focused on the desired users. It will only cause OoBI when two users from two different operators are co-located. Even in that case, both users will receive a similar level of power. This means the UE does not require input dynamic range improvement because of the interferer. Only the S-BS input dynamic range in uplink is affected by the OoBI. This is a major difference compared to non-beamforming systems.

Let us assume two operators sharing a 1GHz band around 28GHz. Operator 1 (Op1) uses the lower half of the band, from 27.5GHz to 28GHz, and Operator 2 (Op2) the upper half from 28GHz to 28.5GHz. All links SINR are required to be 12dB to guaranty the communication average rate on a sub-channel. Using (2.36) this gives an 18dB SNR. The two S-BS dictate power control such that they receive their own users with powers on the same level. Said in a different way, an operator S-BS will always see a nearly constant power spectral density in its own band. According to the previously made hypothesis, users are equipped with a single isotropic antenna.

The worst-case scenario happens when users of different operators are co-located at the footstep of one of the operator's S-BS while being at cell edge of the other operator's one (Figure 2-28). This is because when co-located, the users cannot be spatially separated so the magnitude of the interferer is unaffected by the beamforming process. In Figure 2-28, when at the footstep of Op1's S-BS and at cell edge of Op2's S-BS, UE2 must radiate its maximum output power in order to satisfy its link budget. This creates a strong OoBI for Op1's S-BS which is trying to receive UE1 at a much lower radiated power.

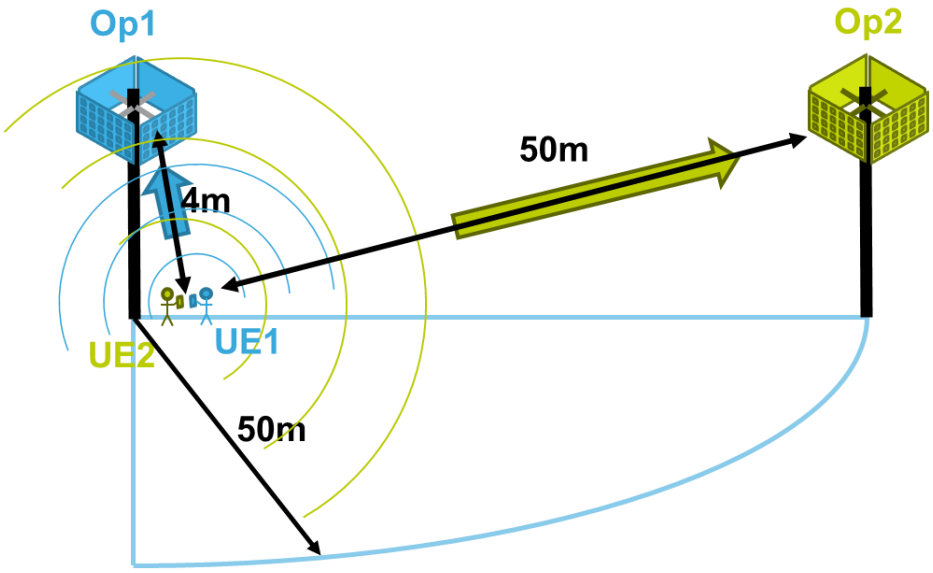

Figure 2-28: Worst case UE configuration

2.6.2 OoBI Power Spectral Characteristics

To determine the OoBI spectral characteristics, two pieces of information are required: Its power spectral density and its frequency location. The first one can be evaluated as a function of the amplitude of the Near-Far effect and the link SINR. The second one can be evaluated with some practical consideration on UE Adjacent Channel Leakage Ratio (ACLR) and ALTernate channel leakage Ratio (ALTR).

2.6.2.1 The Near-Far effect

The power difference of the signal received by Op1 from UE1 and UE2 is called the Near-Far effect and is expressed as:

$$R_{NF} = 20 \times \log_{10}\left(\frac{d_{UE2,S-BS2}}{d_{UE1,S-BS1}}\right) \quad (2.40)$$

The larger this ratio, the worse it is. Its maximum is reached when UE1 and UE2 are the closest to Op1 S-BS and the farthest from Op2 S-BS. In this case, the S-BS is assumed to be installed on streetlight poles at 5m height with UE being in average 1m high. The smallest distance to the S-BS happens when the UE is at its footstep and is $d_{min} = 4m$. The longest distance is at cell-edge, i.e. $d_{max} = 50m$. In these conditions the $R_{NF_{max}} = 22dB$. In practice margin needs to be taken to account for potential beam miss-alignment, power control accuracy, users' antenna anisotropy and so on. This value is rounded up to 30dB to account for these sources of degradation, corresponding to 8dB of margin on UE2 link budget.

2.6.2.2 ACLR and ALTR effect on wide band signals

Many metrics are used to describe the non-linearity of a component. Most of them, such as the third or fifth order Inter-Modulation (IM3 or IM5) or the third order Input or Output Intercept Point (IIP3 or OIP3), describe the effect of the non-linearity on a two-tone signal.

The origin of these metrics come from the analytical approach used to model non-linear systems. The output of a Nth order system is described by the weighted sum of N plus one terms, where each one of them is the ith power of the input signal for i from zero to N. In general, the 0th order weight is zero, the first order weight characterizes the linear part of the system, and the higher order weights characterize the non-linear behavior of the system. Using a two-tone input signal, with the appropriate amplitude, allows to characterize individually the odd order weights of a narrow band system by measuring the in-band intermodulation products. While this approach allows for a lot of analytical insight, it is not very appropriate to evaluate simply the impact on the output signal when dealing with wide band input signals. For that reason, Adjacent Channel Leakage Ratio (ACLR) and ALTernate channel leakage Ratio (ALTR) are generally preferred in modern wireless communications.

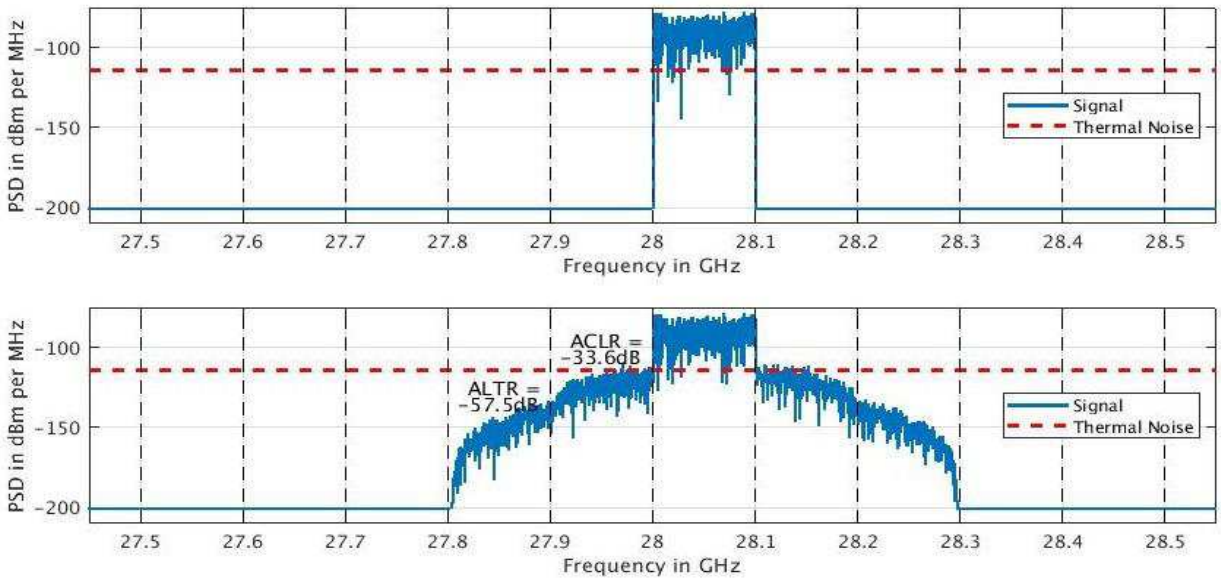

Figure 2-29: Impact of non-linearity on wide band signals. Top: original signal. Bottom: signal after undergoing non-linearity of third and fifth order

When a wide band signal undergoes a non-linear process, it leads to a phenomenon known as spectral regrowth. Figure 2-29 plots an example relevant to this scenario. One can see that the signal power seems to leak onto the adjacent channels. Obviously, this can be a problem if another user is using one of them. To limit such a detrimental effect, standards generally specify a maximum acceptable power

leakage in the adjacent channel and in the channel next to the adjacent one, also called the alternate channel, in terms of ACLR and ALTR. By definition they are the ratios, expressed in decibel, of the power in respectively the adjacent and the alternate channels to the power in the main channel. For this analysis, the values of -30dB for the ACLR and -40dB for the ALTR used in 4G [2-38] will be assumed.

2.6.2.3 Signal PSD profile

Based on the previous hypothesis, the profile of the signal received by a single antenna of Op1 S-BS will be sketched. With a target SNR of 18dB after beamforming, UE1 power must be received at:

$$P_{UE1} = N_{th} + NF_{SRx} + SNR - 10 \times \log_{10}(N_{ant}) = -94 + 10 + 18 - 24 = -90dBm \qquad (2.41)$$

It is interesting to note that this is below the noise power of a single receiver by about $6dB$. Then, from the Near-Far effect, the power received by S-BS$_1$ from UE2 can be evaluated to be $30dB$ above UE1. This gives $P_{UE2} = -60dBm$. Using Friis law from equation (2.23) UE2 PA output power is evaluated to be 5dBm. Two hypotheses will be made here: First, at this power level, the ACLR of UE2's PA is $-30dB$. Second, for lower level of PA output power, the leakage in the adjacent channels is dominated by third order non-linearity. This results in a $3dB$ reduction of the leakage for every $1dB$ reduction on the PA output power. Equivalently, it can be seen as a $2dB$ ACLR improvement for every $1db$ reduction on the PA output power.

Let us number the channels from one to ten, from lower to higher frequencies with the first channel centered at $27.55GHz$, and the following ones spaced by $100MHz$, as depicted in Figure 2-30. If UE1 is in the fifth channel and UE2 in the sixth one, using the first hypothesis, UE2 will leak $-90dBm$ of power in UE1's channel limiting its SINR to $0dB$ at best. This is of course unacceptable.

There are two possibilities to solve this problem. The first obvious one is to move UE2 to channel 8, 9 or 10, so there is no significant leakage in channel 5 and below. This is only possible if those channels are not already occupied by other co-located Op2 users.

The second is to reduce UE2 power and consequently its data rate. To estimate UE2 required power back off BO_{UE2}, it is first required to evaluate how much the leaked power must be reduced. Then, based on the second hypothesis, the amount of back off required for UE2 can be evaluated.

The goal is for the leaked power from UE2 on UE1's channel to have minimal effect on UE1 link. To that end the noise power in the fifth channel must not be degraded significantly. For the leaked power to be acceptable, the assumption is made that it must remain at least 10dB below the fifth channel noise power. Let us call this value $R_{L_{\text{lim}}} = -10dB$. The acceptable leaked power $P_{leak_{lim}}$ is then given by equation (2.42)

$$\begin{aligned} P_{leak_{lim}} &= N_{th} + NF_{SRx} - 10 \times \log_{10}(N_{ant}) + R_{L_{\text{lim}}} = -94 + 10 - 24 - 10 \\ &= -118dBm \end{aligned} \qquad (2.42)$$

Using the second hypothesis, the leaked power can be expressed as a function of UE2 initial power $P_{UE2_{init}} = -60dBm$, its $ACLR_{UE2} = -30dB$ and its back off BO_{UE2}:

$$P_{leak_{UE2}} = P_{UE2_{init}} + ACLR_{UE2} + 3 \times BO_{UE2} \qquad (2.43)$$

Equating (2.43) and (2.42) and rearranging, the required back off BO_{UE2} to keep UE2's leakage below the limit of acceptability can be expressed as:

$$BO_{UE2} = \frac{P_{leak_{lim}} - \left(P_{UE2_{init}} + ACLR_{UE2}\right)}{3} = \frac{-118 - (-60 - 30)}{3} = -9.33dB \qquad (2.44)$$

With this back off, the power leaked into the alternate channel, i.e. the fourth one, will be at least another 10dB lower. This is thanks to the better -40dB ALTR performances compared to the -30dB ACLR. In that case leaked power will have no noticeable impact on the fourth channel.

In practice, PAs, near their saturation output power, exhibit less than 2dB of ACLR reduction per 1dB output power reduction. One example of this is given in [2-37]. This is caused by interactions between the third and fifth order intermodulation product. This makes the proposed hypothesis the worst case in terms of interferer power level. The user in the sixth channel will need to reduce further its output power in order to keep its leakage in the fifth channel below an acceptable power, resulting in an overall lower power of the OoBI formed by the users in channels 6 to 10.

A similar line of reasoning can be held for the leakage in the alternate channel from a user in the seventh channel. This gives a power reduction of the seventh channel user by $4dB$.

Finally, the case when both the sixth and the seventh channels are used simultaneously must be considered. Assuming the leaked power by the sixth channel ACLR is the same as the leaked power by the seventh channel ALTR, the total leakage will increase by $3dB$. For this total leakage to remain 10dB below the fifth channel noise power, the sixth and seventh channels must apply some additional back off. To account for this and for the other approximations, it will be assumed that the sixth channel user must reduce its output power by $12dB$ and the seventh by $6dB$. Their respective SINR would be reduced to $0dB$ and $6dB$ corresponding to data rates of $25Mb/s$ and $50Mb/s$, which are still useful data rates.

Finally, the PSD profile of the Op1 S-BS can be drawn (Figure 2-30). This is the worst-case scenario where Op2 have five co-located users at Op1 S-BS footstep. The green signals represent the OoBI. Its total power is $-55dBm$ while the blue signal total power is only $-83dBm$.

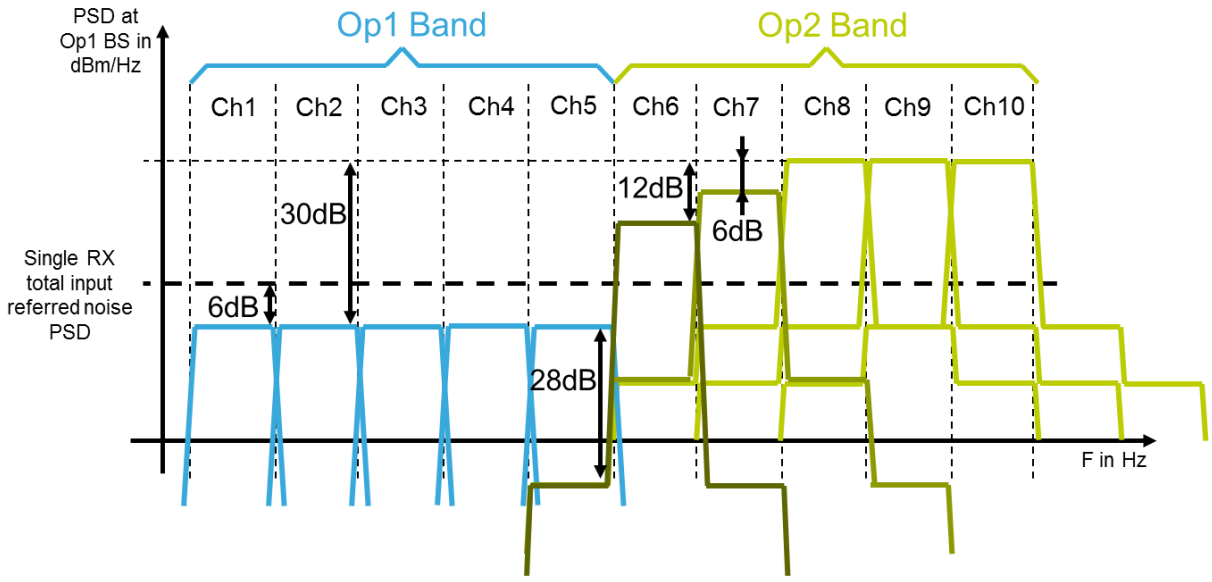

Figure 2-30: Op1 S-BS single receiver PSD profile

The PA output power of the users in channels 8 to 10 in that scenario is evaluated to be around $5dBm$. Because of linearity, it is assumed that the users in channel 6 and 7 will not be allowed to increase their output power. But it is likely for users in channel 8 to 10 to push their output power up to their maximum

of $10dBm$. The interferer can then be assimilated to the signals in channels 8 to 10, the power in channels 6 and 7 being negligible, leading to a total interferer power at Op1 S-BS of $-50dBm$.

2.6.3 Conclusion

After describing the worst-case scenario with two operators and making a proper analysis of the consequences, the power spectral density profile of the interfering signal was established. This signal is by far the most powerful one the system will have to deal with and will set the requirement in input dynamic range. Its detailed spectral characterization will allow for accurate specification of the S-BS receivers. This can potentially avoid over design and allow for proper power optimization.

2.7 CONCLUSION

Starting from the basis of information theory and beamforming, a methodology to analyze millimeter wave beamforming systems with a large number of antennas for wireless communication was developed. Using the outcomes of this analysis and 5G's KPIs as inputs, the dimensions of the proposed system were set. Finally, an analysis of the consequences of a multiple operator deployment scenario was proposed. The major results in this chapter are the following:

First, the peak rate cannot be achieved only with the $28GHz$ band. Much larger bandwidth needs to be used to achieve such rates at reasonable levels of power. Thankfully, such bandwidth is forecasted to be used at $60GHz$ band for example.

Second, the area throughput biggest limitation is not the user access but the backhauling of the data from and to the core network. The proposed wireless backhaul cannot handle the area throughput on wide areas, but it can still make a significant contribution to the problem. Nonetheless, backhaul will probably be the most challenging part in 5G millimeter wave deployment.

Third, except for the S-BS receiver that will be studied in the following chapters, it has been shown that the proposed system does not imply unreachable performances from the other part of the system, namely the S-BS transmitter and the UE transceivers, showing the feasibility of the proposed system.

Finally, all the required inputs needed for precisely specifying the S-BS receiver were derived. This will be the subject of the next chapter.

2.8 REFERENCES

[2-1] C. E. Shannon, "A mathematical theory of communication," in The Bell System Technical Journal, vol. 27, no. 3, pp. 379-423, July 1948.

[2-2] C. E. Shannon, "A mathematical theory of communication," in The Bell System Technical Journal, vol. 27, no. 4, pp. 623-656, Oct. 1948.

[2-3] D. Mackay, "Information Theory, Inference and Learning Algorithms". Available online: www.inference.org.uk/itprnn/book.pdf

[2-4] C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes. 1," Proceedings of ICC '93 - IEEE International Conference on Communications, Geneva, Switzerland, 1993, pp. 1064-1070 vol.2.

[2-5] D. J. C. MacKay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," in Electronics Letters, vol. 32, no. 18, pp. 1645-, 29 Aug. 1996.

[2-6] P. K. Bondyopadhyay, "The first application of array antenna," Proceedings 2000 IEEE International Conference on Phased Array Systems and Technology (Cat. No.00TH8510), Dana Point, CA, 2000, pp. 29-32.

[2-7] Sophocles J. Orfanidis, “Electromagnetic waves and antennas,” Rutgers University, August 1st 2016.

[2-8] C. L. Dolph, "A Current Distribution for Broadside Arrays Which Optimizes the Relationship between Beam Width and Side-Lobe Level," in Proceedings of the IRE, vol. 34, no. 6, pp. 335-348, June 1946.

[2-9] T. T. Taylor, "Design of line-source antennas for narrow beamwidth and low side lobes," in Transactions of the IRE Professional Group on Antennas and Propagation, vol. 3, no. 1, pp. 16-28, Jan. 1955.

[2-10] A. Gangi, "The active adaptive antenna array system," in IEEE Transactions on Antennas and Propagation, vol. 11, no. 4, pp. 405-414, July 1963.

[2-11] R. Ghose, "Electronically adaptive antenna systems," in IEEE Transactions on Antennas and Propagation, vol. 12, no. 2, pp. 161-169, March 1964.

[2-12] B. Widrow, P. E. Mantey, L. J. Griffiths and B. B. Goode, "Adaptive antenna systems," in Proceedings of the IEEE, vol. 55, no. 12, pp. 2143-2159, Dec. 1967.

[2-13] W. F. Gabriel, "Adaptive arrays - An introduction," in Proceedings of the IEEE, vol. 64, no. 2, pp. 239-272, Feb. 1976.

[2-14] A. J. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," in IEEE Signal Processing Magazine, vol. 14, no. 6, pp. 49-83, Nov. 1997.

[2-15] T. L. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," in IEEE Transactions on Wireless Communications, vol. 9, no. 11, pp. 3590-3600, November 2010.

[2-16] H. Q. Ngo, E. G. Larsson and T. L. Marzetta, "Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems," in IEEE Transactions on Communications, vol. 61, no. 4, pp. 1436-1449, April 2013.

[2-17] E. Björnson, M. Bengtsson and B. Ottersten, "Optimal Multiuser Transmit Beamforming: A Difficult Problem with a Simple Solution Structure [Lecture Notes]," in IEEE Signal Processing Magazine, vol. 31, no. 4, pp. 142-148, July 2014.

[2-18] R. Harrington, "Sidelobe reduction by nonuniform element spacing," in IRE Transactions on Antennas and Propagation, vol. 9, no. 2, pp. 187-192, March 1961.

[2-19] R. L. Haupt, "Thinned arrays using genetic algorithms," in IEEE Transactions on Antennas and Propagation, vol. 42, no. 7, pp. 993-999, July 1994.

[2-20] C. Stearns and A. Stewart, "An investigation of concentric ring antennas with low sidelobes," in IEEE Transactions on Antennas and Propagation, vol. 13, no. 6, pp. 856-863, November 1965.

[2-21] M. Alvarez-Folgueiras, J. A. Rodriguez-Gonzalez and F. Ares-Pena, "High-Performance Uniformly Excited Linear and Planar Arrays Based on Linear Semiarrays Composed of Subarrays With Different Uniform Spacings," in IEEE Transactions on Antennas and Propagation, vol. 57, no. 12, pp. 4002-4006, Dec. 2009.

[2-22] P. Wang, Y. Li, Y. Peng, S. C. Liew and B. Vucetic, "Non-uniform linear antenna array design for millimeter wave MIMO channels," 2015 9th International Conference on Signal Processing and Communication Systems (ICSPCS), Cairns, QLD, 2015, pp. 1-5.

[2-23] F. Sohrabi and W. Yu, "Hybrid Digital and Analog Beamforming Design for Large-Scale Antenna Arrays," in IEEE Journal of Selected Topics in Signal Processing, vol. 10, no. 3, pp. 501-513, April 2016.

[2-24] C. Liu, S. Xiao and Z. Zhang, "A low profile linearly polarized patch antenna with wide beam-width in E-plane," 2016 IEEE International Workshop on Electromagnetics: Applications and Student Innovation Competition (iWEM), Nanjing, 2016, pp. 1-3.

[2-25] J. Xu, H. Ke, Y. He and Y. Luo, "A Wideband U-Slot Microstrip Patch Antenna for Large-Angle Mmw Beam Scanning," 2018 IEEE International Conference on Computer and Communication Engineering Technology (CCET), Beijing, 2018, pp. 142-145.

[2-26] M. Wang, X. Xu and M. He, "The parametric analysis of H-shaped patch antenna," 2011 IEEE International Conference on Microwave Technology & Computational Electromagnetics, Beijing, 2011, pp. 250-252.

[2-27] C. -. Su, S. -. Huang and C. -. Lee, "CP microstrip antenna with wide beamwidth for GPS band application," in Electronics Letters, vol. 43, no. 20, pp. 1062-1063, 27 September 2007.

[2-28] Z. Pan, W. Lin and Q. Chu, "Compact Wide-Beam Circularly-Polarized Microstrip Antenna With a Parasitic Ring for CNSS Application," in IEEE Transactions on Antennas and Propagation, vol. 62, no. 5, pp. 2847-2850, May 2014.

[2-29] M. Samimi, T. Rappaport, "Characterization of the 28 GHz Millimeter-Wave Dense Urban Channel for Future 5G Mobile Cellular," NYU WIRELESS TR 2014-001Technical Report, June 24 2014.

[2-30] B. Wang, F. Gao, S. Jin, H. Lin and G. Y. Li, "Spatial- and Frequency-Wideband Effects in Millimeter-Wave Massive MIMO Systems," in IEEE Transactions on Signal Processing, vol. 66, no. 13, pp. 3393-3406, 1 July1, 2018.

[2-31] S. Mattisson, “Overview of 5G requirements and future wireless networks,” in ESSCIRC 2017 - 43rd IEEE European Solid State Circuits Conference, Sept 2017, pp. 1–6.

[2-32] B. Moret, V. Knopik and E. Kerherve, "A 28GHz self-contained power amplifier for 5G applications in 28nm FD-SOI CMOS," 2017 IEEE 8th Latin American Symposium on Circuits & Systems (LASCAS), Bariloche, 2017, pp. 1-4.

[2-33] Y. Zhang and P. Reynaert, "A high-efficiency linear power amplifier for 28GHz mobile communications in 40nm CMOS," 2017 IEEE Radio Frequency Integrated Circuits Symposium (RFIC), Honolulu, HI, 2017, pp. 33-36.

[2-34] D. Thomas, N. Rostomyan and P. Asbeck, "A 45 % PAE pMOS Power Amplifier for 28GHz Applications in 45nm SOI," 2018 IEEE 61st International Midwest Symposium on Circuits and Systems (MWSCAS), Windsor, ON, Canada, 2018, pp. 680-683.

[2-35] Deliverable D2.6 v0.2 “Final report on programme progress and KPI”, 5G-PPP, EURO-5G, October 2017.

[2-36] O. Elijah, C. Y. Leow, T. A. Rahman, S. Nunoo and S. Z. Iliya, "A Comprehensive Survey of Pilot Contamination in Massive MIMO—5G System," in IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 905-923, Secondquarter 2016.

[2-37] X. Ding and L. Zhang, "A High-Efficiency GaAs MMIC Power Amplifier for Multi-Standard System," in IEEE Microwave and Wireless Components Letters, vol. 26, no. 1, pp. 55-57, Jan. 2016.

[2-38] LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception (3GPP TS 36.101 version 16.8.0 Release 16)

2.9 ANNEX 2.1

In this annex, the derivation from equation (2.10) (recalled below) to equation (2.11) is provided.

$$S_{BF}(N_{ant},\theta_{BF},\theta)=\sum_{n=0}^{N_{ant}-1}\cos\left(2\times\pi\times f\times\left(t-\frac{n}{2\times f_c}(\sin(\theta)-\sin(\theta_{BF}))\right)\right) \tag{A.2.1}$$

The imaginary part to each cosine can be added to complete the complex exponential:

$$S_{BF_{complex}}(N_{ant},\theta_{BF},\theta)=\sum_{n=0}^{N_{ant}-1}\mathrm{e}^{j\times2\times\pi\times f\times\left(t-\frac{n}{2\times f_c}\times(\sin(\theta)-\sin(\theta_{BF}))\right)} \tag{A.2.2}$$

$$S_{BF_{complex}}(N_{ant},\theta_{BF},\theta)=\mathrm{e}^{j\times2\times\pi\times f\times t}\times\sum_{n=0}^{N_{ant}-1}\left(\mathrm{e}^{-j\times\pi\times\frac{f}{f_c}\times(\sin(\theta)-\sin(\theta_{BF}))}\right)^n \tag{A.2.3}$$

The summation part is the sum of a geometric sequence of common ratio $q=\mathrm{e}^{-j\times\pi\times\frac{f}{f_c}\times(\sin(\theta)-\sin(\theta_{BF}))}$ so its value is $\frac{1-q^{N_{ant}}}{1-q}$.

$$S_{BF_{complex}}(N_{ant},\theta_{BF},\theta)=\mathrm{e}^{j\times2\times\pi\times f\times t}\times\frac{1-\mathrm{e}^{-j\times N_{ant}\times\pi\times\frac{f}{f_c}\times(\sin(\theta)-\sin(\theta_{BF}))}}{1-\mathrm{e}^{-j\times\pi\times\frac{f}{f_c}\times(\sin(\theta)-\sin(\theta_{BF}))}} \tag{A.2.4}$$

Factorizing the top and bottom by the half angle gives:

$$\begin{aligned}&S_{BF_{complex}}(N_{ant},\theta_{BF},\theta)\\&\quad=\mathrm{e}^{j\times2\times\pi\times f\times t}\times\mathrm{e}^{-j\times(N_{ant}-1)\times\frac{\pi}{2}\times\frac{f}{f_c}\times(\sin(\theta)-\sin(\theta_{BF}))}\\&\quad\times\frac{\mathrm{e}^{j\times N_{ant}\times\frac{\pi}{2}\times\frac{f}{f_c}\times(\sin(\theta)-\sin(\theta_{BF}))}-\mathrm{e}^{-j\times N_{ant}\times\frac{\pi}{2}\times\frac{f}{f_c}\times(\sin(\theta)-\sin(\theta_{BF}))}}{\mathrm{e}^{j\times\frac{\pi}{2}\times\frac{f}{f_c}\times(\sin(\theta)-\sin(\theta_{BF}))}-\mathrm{e}^{-j\times\frac{\pi}{2}\times\frac{f}{f_c}\times(\sin(\theta)-\sin(\theta_{BF}))}}\end{aligned} \tag{A.2.5}$$

Using Euler's formula, the top and the bottom part of the quotient can be simplified:

$$\begin{aligned}&S_{BF_{complex}}(N_{ant},\theta_{BF},\theta)\\&\quad=\mathrm{e}^{j\times2\times\pi\times f\times t}\times\mathrm{e}^{-j\times(N_{ant}-1)\times\frac{\pi}{2}\times\frac{f}{f_c}\times(\sin(\theta)-\sin(\theta_{BF}))}\\&\quad\times\frac{2\times\mathrm{j}\times\sin\left(N_{ant}\times\frac{\pi}{2}\times\frac{f}{f_c}\times(\sin(\theta)-\sin(\theta_{BF}))\right)}{2\times\mathrm{j}\times\sin\left(\frac{\pi}{2}\times\frac{f}{f_c}\times(\sin(\theta)-\sin(\theta_{BF}))\right)}\end{aligned} \tag{A.2.6}$$

$$S_{BF_{complex}}(N_{ant}, \theta_{BF}, \theta) = e^{j\times 2\times\pi\times f\times\left(t-\frac{(N_{ant}-1)}{4\times f_c}\times(\sin(\theta)-\sin(\theta_{BF}))\right)} \times \frac{\sin\left(N_{ant}\times\frac{\pi}{2}\times\frac{f}{f_c}\times(\sin(\theta)-\sin(\theta_{BF}))\right)}{\sin\left(\frac{\pi}{2}\times\frac{f}{f_c}\times(\sin(\theta)-\sin(\theta_{BF}))\right)} \quad \text{(A.2.7)}$$

Taking the real part of it gives back the original signal:

$$S_{BF}(N_{ant}, \theta_{BF}, \theta) = \cos\left(2\times\pi\times f\times\left(t-\frac{(N_{ant}-1)}{4\times f_c}\times(\sin(\theta)-\sin(\theta_{BF}))\right)\right) \times \frac{\sin\left(N_{ant}\times\frac{\pi}{2}\times\frac{f}{f_c}\times(\sin(\theta)-\sin(\theta_{BF}))\right)}{\sin\left(\frac{\pi}{2}\times\frac{f}{f_c}\times(\sin(\theta)-\sin(\theta_{BF}))\right)} \quad \text{(A.2.8)}$$

2.10 ANNEX 2.2

In this annex, the effects of the number of sub-bands and the RF channel spatial richness on PSS and MRC beamforming will be studied. The conditions are the same as in section 2.2.5.4. Figure A.2.1 plots the SNR versus the number of sub-bands under PSS beamforming for RF channels of three different richness.

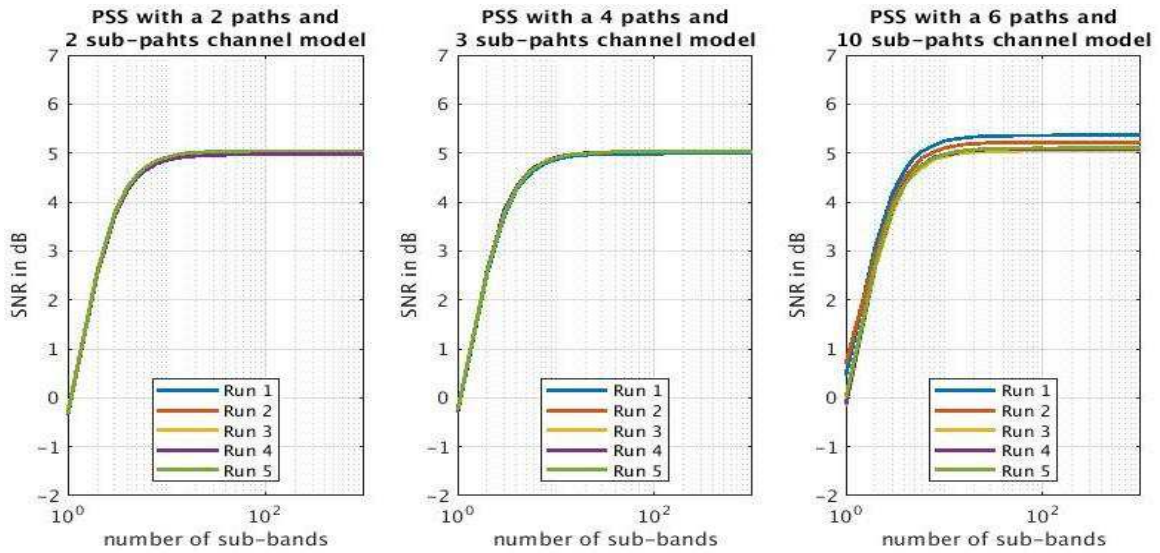

Figure A.2.1: SNR vs the number of sub-bands under PSS beamforming from 3 different channels

As the number of sub-bands increase performances tends to DS beamforming and is mostly unaffected by the richness of the channel. The interesting result here is that only a few numbers of sub-bands is required to approach DS performances. Four sub-bands always get you within $1dB$ of maximum performance. This means that the limitation of PSS can be overcome with a reasonable increase in complexity. In particular, it is not required to make any addition CSI acquisitions. The PSS weights can be processed just knowing the AoA regardless of the number of sub-bands.

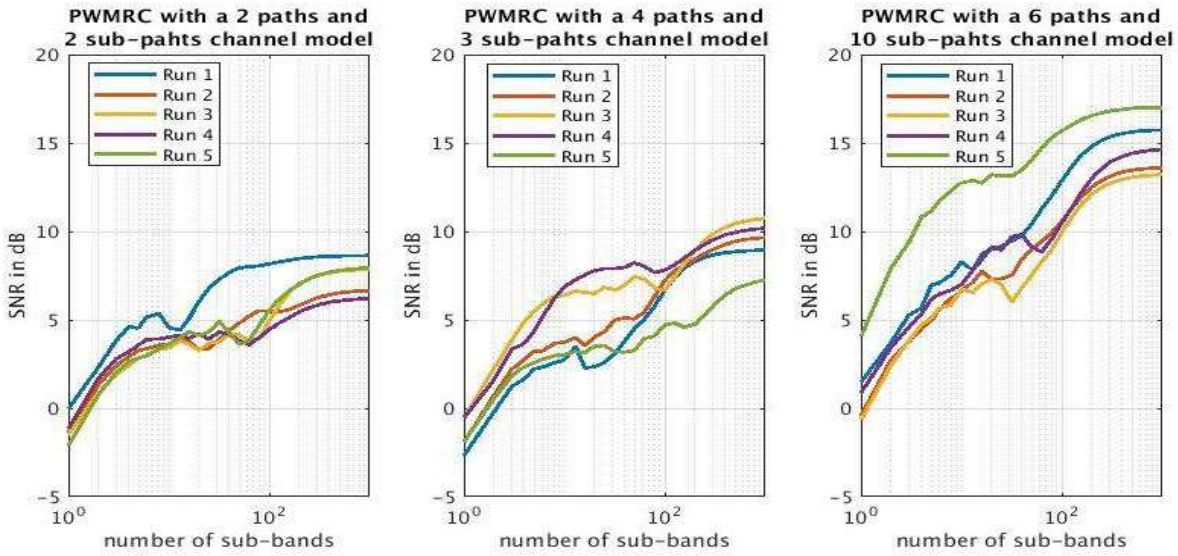

Figure A.2.2: SNR vs the number of sub-bands under MRC beamforming from 3 different channels

For MRC in Figure A.2.2, the results are significantly different. The maximum performances are always better than DS and is highly dependent on the channel richness but requires many more sub-bands, typically few hundreds. The performance increase ranges from about $2dB$ up to $10dB$ in the example. When the channel is rich it is possible to achieve better performances, or to achieve similar performances with fewer antennas, but because the required number of sub-bands is so high it is unlikely that can offset the processing complexity increase. On top of that this is interesting only if a rich channel is likely enough. Even though many measurement campaigns have been done, very few measurements have been done below a 50m TX-RX distance and therefore reliable channel statistics are not yet available.

One interesting investigation is proposed in [2-30]. They suggest exploiting the sparsity of the channel in the angular-delay domain, i.e. there is very few multipath arriving with different AoA and delays. The idea is to receive each path with an independent beam and then to realign them in the time domain before recombination. The complexity would be increased only by the number of paths, ~4 in average, and the CSI acquisition would also be reduced since only made of the AoA and relative delays of each path. It is also an interesting approach because it can be made modular since an independent beam can be used for any path of any user. The game would be, for a system capable of N_b independent beams, how to allocate them optimally to the N_u users to be served. But the same question remains, can the added performances offset the power cost of the extra processing?

2.11 ANNEX 2.3

This annex provides the step-by-step derivation of the $P_{LoS_{Cell}}(r)$. For recall $P_{LoS}(d)$ is:

$$P_{LoS}(d) = \min\left(\frac{d_1}{d}, 1\right) \times \left(1 - e^{-\frac{d}{d_2}}\right) + e^{-\frac{d}{d_2}} \tag{A.2.9}$$

The LoS probability for a cell of radius d_{SC}, expressed in the polar coordinates, with a uniform probability of location of the user $P_{UE} = \frac{1}{\pi \times d_{SC}^2}$ is expressed by:

$$P_{LoS_{Cell}}(d_{SC}) = \begin{cases} 1 & if\ d_{SC} < d_1 \\ \int_0^{2\times\pi} \int_0^{d_{SC}} P_{LoS}(d) \times P_{UE} \times d \times \partial d \times \partial\theta & if\ d_{SC} \geq d1 \end{cases} \tag{A.2.10}$$

Focusing on the case where $d_{SC} \geq d_1$, because $P_{LoS}(d)$ is independent of the polar angle θ and P_{UE} is independent from both integration variables, it can be reformulated as:

$$\begin{aligned} P_{LoS_{Cell}}(d_{SC}) &= P_{UE} \times \int_0^{2\times\pi} \partial\theta \times \int_0^{d_{SC}} P_{LoS}(d) \times d \times \partial d \\ &= \frac{2 \times \pi}{\pi \times d_{SC}^2} \times \int_0^{d_{SC}} P_{LoS}(d) \times d \times \partial d \end{aligned} \tag{A.2.11}$$

$$P_{LoS_{Cell}}(d_{SC}) = \frac{2}{d_{SC}^2} \times \int_0^{d_{SC}} P_{LoS}(d) \times d \times \partial d \tag{A.2.12}$$

To evaluate this integral it must be split using Chasles' relation at d_1 :

$$P_{LoS_{Cell}}(d_{SC}) = \frac{2}{d_{SC}^2} \times \left(\int_0^{d_1} d \times \partial d + \int_{d_1}^{d_{SC}} \left(\frac{d_1}{d} \times \left(1 - e^{-\frac{d}{d_2}} \right) + e^{-\frac{d}{d_2}} \right) \times d \times \partial d \right) \tag{A.2.13}$$

The first term is straight forward to integrate. Using the integral linearity property, the second term is split as follow:

$$P_{LoS_{Cell}}(d_{SC}) = \frac{2}{d_{SC}^2} \times \left(\frac{d_1^2}{2} + d_1 \times \int_{d_1}^{d_{SC}} \left(1 - e^{-\frac{d}{d_2}} \right) \times \partial d + \int_{d_1}^{d_{SC}} e^{-\frac{d}{d_2}} \times d \times \partial d \right) \tag{A.2.14}$$

The first integral is straight forward to integrate and an integration by part is used for the second one.

$$\begin{aligned} P_{LoS_{Cell}}(d_{SC}) = &\frac{2}{d_{SC}^2} \\ &\times \left(\frac{d_1^2}{2} + d_1 \times \left(d_{SC} - d_1 + d_2 \times \left(e^{-\frac{d_{SC}}{d_2}} - e^{-\frac{d_1}{d_2}} \right) \right) \right. \\ &\left. + \left[-d_2 \times e^{-\frac{d}{d_2}} \times d \right]_{d1}^{d_{SC}} - \int_{d_1}^{d_{SC}} -d_2 \times e^{-\frac{d}{d_2}} \times \partial d \right) \end{aligned} \tag{A.2.15}$$

The last integral is straight forward:

$$\begin{aligned} P_{LoS_{Cell}}(d_{SC}) = &\frac{2}{d_{SC}^2} \\ &\times \left(\frac{d_1^2}{2} + d_1 \times \left(d_{SC} - d_1 + d_2 \times \left(e^{-\frac{d_{SC}}{d_2}} - e^{-\frac{d_1}{d_2}} \right) \right) \right. \\ &\left. - d_2 \times \left(e^{-\frac{d_{SC}}{d_2}} \times d_{SC} - e^{-\frac{d_1}{d_2}} \times d_1 \right) - d_2^2 \times \left(e^{-\frac{d_{SC}}{d_2}} - e^{-\frac{d_1}{d_2}} \right) \right) \end{aligned} \tag{A.2.16}$$

Rearranging leads to:

$$P_{LoS_{Cell}}(d_{SC}) = \frac{d_1^2}{d_{SC}^2} \times \left(1 + 2 \times \left(\frac{d_{SC}}{d_1} - 1\right) \times \left(1 - \frac{d_2}{d_1} \times e^{-\frac{d_{SC}}{d_2}}\right) - 2 \times \left(\frac{d_2}{d_1}\right)^2 \times \left(e^{-\frac{d_{SC}}{d_2}} - e^{-\frac{d_1}{d_2}}\right)\right) \quad \text{(A.2.17)}$$

It can be confirmed that, when $d_{SC} = d_1$ then $P_{LoS_{Cell}}(d_1) = 1$ so continuity is ensured with the case $d < d_1$. When d_{SC} goes to infinity it gives the following asymptote:

$$P_{LoS_{Cell}}(d_{SC}) \xrightarrow[r \to +\infty]{} 2 \times \frac{d_1}{d_{SC}} \xrightarrow[r \to +\infty]{} 0 \quad \text{(A.2.18)}$$

Figure A.2.3 plot these different results with $d_1 = 24m$ and $d_2 = 45m$

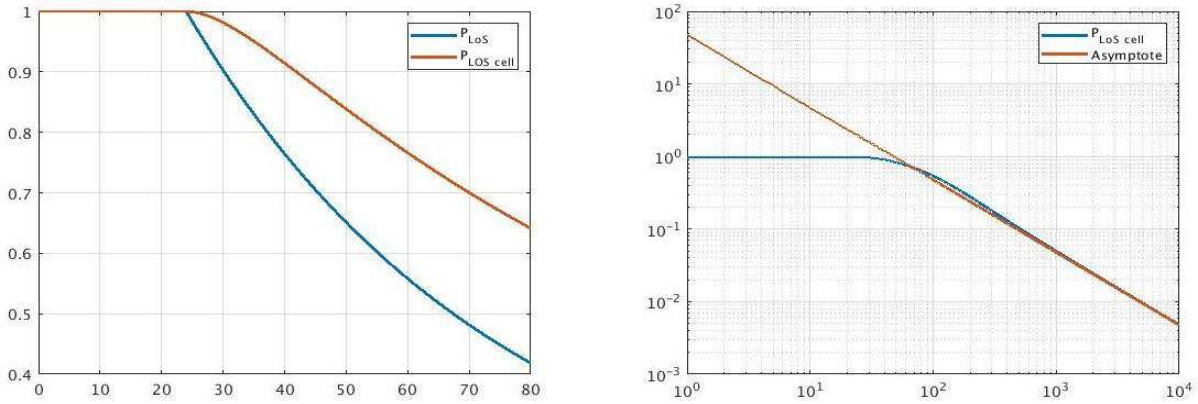

Figure A.2.3: Left $P_{LoS}(d)$ and $P_{LoS_{Cell}}(d_{SC})$. Right) $P_{LoS_{Cell}}(d_{SC})$ and its asymptote at infinity on a log-log plot

2.12 ANNEX 2.4

In this annex, the calculation details of Table 2-1, recalled here for convenience, is provided.

Table A.2.1: 5G KPIs and System performances summary

	5G KPI	Proposed system				
		Performances	SINR	PA inst power	Cell throughput	Cell PA avg output power per bit per second
User density	$10\ ku/km^2$	$50\ ku/km^2$	N/A	N/A	N/A	N/A
Average rate UL	$50Mb/s$	$100Mb/s$	$11.8dB$	$-2.4dBm^*$	$40Gb/s$	$1.4\ pW \cdot b^{-1} \cdot s^{**}$
Average rate DL	$100Mb/s$	$200Mb/s$	$11.8dB$	$-7.1dBm^{**}$	$80Gb/s$	$0.7\ pW \cdot b^{-1} \cdot s^{**}$
Peak rate UL	$10Gb/s$	$2Gb/s$	$25dB$	$10dBm$	$8Gb/s$	$2.4\ pW \cdot b^{-1} \cdot s$
Peak rate DL	$20Gb/s$	$7Gb/s$	$42dB$	$10dBm$	$28Gb/s$	$177\ pW \cdot b^{-1} \cdot s$
Backhaul	N/A	$5Gb/s$	$30dB$	N/A	$30Gb/s$	N/A

* At cell edge. ** At maximum user density with users uniformly distributed.

2.12.1 Derivation of equation (2.36)

First, let us setup the notations. All the signals considered have the same bandwidth B. S_{UE} is the signal power of interest, N_{th} is the receiver's thermal noise and NF its noise figure. There are two different sources of interference, the ones from the other users and the ones from the backhaul signals of power S_{BH}. There is a total of $N_{beam_{UE}}$ users, where $N_{beam_{UE}} - 1$ are interferers, and there is $N_{beam_{BH}}$, all of them being interferers. The users' rejection ratio is IRR_{UE} and the backhaul one is IRR_{BH}. The $SINR_{UE}$ can expressed as follow:

$$SINR_{UE} = \frac{S_{UE}}{N_{th} + \left(N_{beam_{UE}} - 1\right) \times S_{UE} \times IRR_{UE} + N_{beam_{BH}} \times S_{BH} \times IRR_{BH}} \quad \text{(A.2.19)}$$

Factorizing the top and the bottom by N_{th} gives:

$$SINR_{UE} = \frac{SNR_{UE}}{1 + \left(N_{beam_{UE}} - 1\right) \times SNR_{UE} \times IRR_{UE} + N_{beam_{BH}} \times SNR_{BH} \times IRR_{BH}} \quad \text{(A.2.20)}$$

With $SNR_{UE} = S_{UE}/N_{th}$ and $SNR_{BH} = S_{BH}/N_{th}$. Similarly, the backhaul $SINR_{BH}$ is expressed as:

$$SINR_{BH} = \frac{SNR_{BH}}{1 + N_{beam_{UE}} \times SNR_{UE} \times IRR_{UE} + \left(N_{beam_{BH}} - 1\right) \times SNR_{BH} \times IRR_{BH}} \quad \text{(A.2.21)}$$

(A.2.21) can be rearranged to express the SNR_{BH} as function of $SINR_{BH}$ and SNR_{UE}:

$$SNR_{BH} = \frac{SINR_{BH} \times \left(1 + N_{beam_{UE}} \times SNR_{UE} \times IRR_{UE}\right)}{1 - \left(N_{beam_{BH}} - 1\right) \times SINR_{BH} \times IRR_{BH}} \quad \text{(A.2.22)}$$

The following quantity is named D:

$$D = \frac{N_{beam_{BH}} \times SINR_{BH} \times IRR_{BH}}{1 - \left(N_{beam_{BH}} - 1\right) \times SINR_{BH} \times IRR_{BH}} \quad \text{(A.2.23)}$$

(A.2.22) is re-written:

$$SNR_{BH} = \frac{1 + N_{beam_{UE}} \times SNR_{UE} \times IRR_{UE}}{N_{beam_{BH}} \times IRR_{BH}} \times D \quad \text{(A.2.24)}$$

Injecting (A.2.24) into (A.2.20) and rearranging gives:

$$SINR_{UE} = \frac{SNR_{UE}}{(1 + D) + \left(N_{beam_{UE}} \times (1 + D) - 1\right) \times SNR_{UE} \times IRR_{UE}} \quad \text{(A.2.25)}$$

Reversing (A.2.25) provide an expression of SNR_{UE} without any dependence on SNR_{BH}:

$$SNR_{UE} = \frac{SINR_{UE} \times (1 + D)}{1 - \left(N_{beam_{UE}} \times (1 + D) - 1\right) \times SINR_{UE} \times IRR_{UE}} \quad \text{(A.2.26)}$$

Naming K_{BH} the quantity $(1 + D)$ gives back equation (2.36). The UE and backhaul play symmetrical roles so (2.38) is obtained by simply changing the UE subscripts by BH and vis-versa.

2.12.2 Uplink average rate

The target rate is $UL_{avg} = 100Mb/s$. The considered system uses a TDD duplexing with equal time for uplink and downlink, and half of the uplink time dedicated to pilot transmission. The uplink conveys actual data only for one fourth of the time. The average instantaneous rate must be four times higher to

compensate, i.e. $UL_{avg_{inst}} = 400Mb/s$. The cell is divided in $N_{sect} = 4$. Each sector is equipped with an antenna array with $N_{ant} = 256$ elements and can create up to $N_{beam_{UE}} = 10$ and $N_{beam_{BH}} = 2$. Each UE beam has a bandwidth of $B = 1GHz$ divided into 10 sub-channels of $B_{sub} = 100MHz$. This configuration can serve up to $N_{users} = 400$ per cell. The required $SINR_{UE}$ for each user to achieve $UL_{avg_{inst}}$ over the sub-channel bandwidth B_{sub} is given by:

$$SINR_{UE} = 10 \times \log_{10}\left(2^{\frac{UL_{avg_{inst}}}{B_{sub}}} - 1\right) = 11.8dB \tag{A.2.27}$$

And the required SNR is given by (2.36) recalled here for convenience:

$$SNR_{UE} = \frac{SINR_{UE} \times K_{BH}}{1 - \left(N_{beam_{UE}} \times K_{BH} - 1\right) \times SINR_{UE} \times IRR_{UE}} \tag{A.2.28}$$

With:

$$K_{BH} = \frac{1 + SINR_{BH} \times IRR_{BH}}{1 - \left(N_{beam_{BH}} - 1\right) \times SINR_{BH} \times IRR_{BH}} \tag{A.2.29}$$

Making the hypothesis of $IRR_{UE_{dB}} = -25dB, SINR_{BH} = 33dB$ and $IRR_{BH} = -40$ the required SNR_{UE} is $18.2dB$. Note that the $SINR_{BH}$ value to be used with this equation is the 33dB that ignores the interference from one side of the backhaul link. The effective $SINR_{BH_{eff}}$ is about $3dB$ lower at $30dB$. It is assumed that the receivers' noise figure is $NF = 10dB$. By reversing equation (2.31), recalled here for convenience, the required user PA output power $P_{t_{dBm}}$ can be evaluated as a function of its distance d_{UE} to the S-BS.

$$SNR = \left(P_{t_{dBm}} + Att_{Ch}\right) - \left(N_{th_{dBm}} + NF\right) + 10 \times \log_{10}(N_{ant}) \tag{A.2.30}$$

With $Att_{Ch} = 10 \times \log_{10}\left(\frac{\lambda^2 \times A_r \times A_t}{(4 \times \pi \times d_{UE})^2}\right)$ the channel attenuation and $N_{th_{dBm}} = 10 \times \log_{10}\left(\frac{k_b \times T_0 \times B_{sub}}{1mW}\right)$ the antenna thermal noise in dBm in the sub-channel bandwidth B_{sub}.

$$P_{t_{dBm}}(d_{UE}) = SNR - Att_{Ch}(d_{UE}) + \left(N_{th_{dBm}} + NF\right) - 10 \times \log_{10}(N_{ant}) \tag{A.2.31}$$

A user at cell edge, meaning $d_{UE} = 50m$, need to deliver a PA output power of $P_{t_{dBm}} = -2.4dBm$. To evaluate the total average PA output power P_{tot} of all the users in the cell, the assumption is made that they are uniformly distributed in the cell of radius R. First, the average PA output power for one user is processed. The probability ∂p_{UE}^2 of a user to be within the infinitesimal are $d_{UE} \times \partial d_{UE} \times \partial \theta_{UE}$ is the product of this area by the uniform probability density $\rho_{UE} = \frac{1}{\pi \times R^2}$ over the whole cell area, that is $\partial p_{UE}^2 = \frac{d_{UE} \times \partial d_{UE} \times \partial \theta_{UE}}{\pi \times R^2}$. The PA output power P_t expressed in linear for is:

$$P_t(d_{UE}) = \frac{10^{\frac{SNR}{10}} \times (4 \times \pi \times d_{UE})^2 \times k_b \times T \times B_{sub} \times 10^{\frac{NF}{10}}}{\lambda^2 \times A_r \times A_t \times N_{ant}} \tag{A.2.32}$$

The average user PA output power $P_{t_{avg}}(R)$ is the integral of $P_t(d_{UE})$ weighted by the probability ∂p_{UE}^2 over the cell area:

$$
\begin{aligned}
P_{t_{avg}}(R) &= \int_0^{2\pi}\int_0^{R} \frac{10^{\frac{SNR}{10}} \times (4\times\pi\times d_{UE})^2 \times k_b \times T \times B_{sub} \times 10^{\frac{NF}{10}}}{\lambda^2 \times A_r \times A_t \times N_{ant}} \\
&\quad \times \frac{d_{UE} \times \partial d_{UE} \times \partial\theta_{UE}}{\pi \times R^2} \\
&= \frac{10^{\frac{SNR}{10}} \times 16 \times \pi \times k_b \times T \times B_{sub} \times 10^{\frac{NF}{10}}}{\lambda^2 \times A_r \times A_t \times N_{ant} \times R^2} \times \int_0^{2\pi}\int_0^{R} d_{UE}^3 \times \partial d_{UE} \times \partial\theta_{UE} \\
&= \frac{10^{\frac{SNR}{10}} \times 16 \times \pi \times k_b \times T \times B_{sub} \times 10^{\frac{NF}{10}}}{\lambda^2 \times A_r \times A_t \times N_{ant} \times R^2} \times \frac{2\times\pi\times R^4}{4} \\
&= \frac{10^{\frac{SNR}{10}} \times (4\times\pi)^2 \times k_b \times T \times B_{sub} \times 10^{\frac{NF}{10}}}{\lambda^2 \times A_r \times A_t \times N_{ant}} \times \frac{R^2}{2}
\end{aligned}
\tag{A.2.33}
$$

This can conveniently be written in the logarithmic form as:

$$
P_{t_{dBm_{avg}}}(R) = P_{t_{dBm}}(d_{UE}=1) + 10\times\log_{10}\left(\frac{R^2}{2}\right) \xrightarrow[R=50]{} -5.4dBm \tag{A.2.34}
$$

The total average PA output power P_{tot} for all the users of the cell is the power contribution of the 400 users multiplied by the on-time ratio $R_{t_{on}} = 1/2$ of the PA in the uplink. Note here that while data are transmitted only during one quarter of the time, the PA must be on half of the time to transmit the pilot and the data. This is the power cost of the pilot.

$$
P_{tot_{dBm}}(R) = P_{t_{dBm_{avg}}}(R) + 10\times\log_{10}(N_{users}\times R_{t_{on}}) \xrightarrow[R=50]{} 17.6dBm \tag{A.2.35}
$$

Note that this is the PA output power. To have the PA power consumption one must consider the PA efficiency. This is not evaluated here because the goal is only to compare the power consumption of the different scenarios, assuming the same PA efficiency in all cases.

The product of the uplink average rate UL_{avg} by the number of users N_{users} gives the cell uplink throughput $UL_{avg_{thpt}} = 40Gb/s$. Finally, the average uplink efficient $UL_{avg_{eff}}$ characterized by the average PA output power per bit per second is:

$$
UL_{avg_{eff}} = \frac{0.001\times 10^{\frac{P_{tot_{dBm}}(R=50m)}{10}}}{UL_{avg_{thpt}}} = 1.4\,pW\cdot b^{-1}\cdot s \tag{A.2.36}
$$

This is homogenous to joules per bit, but since only the PA output power is considered, not its power consumption, the unit of Watt per bit per second is more relevant. It is mostly useful for relative comparison of different scenarios.

2.12.3 Uplink peak rate

The analysis of the peak rate is very similar and somewhat simpler. There is now only one user per sector giving $N_{users} = 4$. Each of them has a PA output power $P_{t_{dBm}} = 10dBm$. The total PA output power is:

$$
P_{tot_{dBm}} = P_{t_{dBm}} + 10\times\log_{10}(N_{users}\times R_{t_{on}}) = 13dBm \tag{A.2.37}
$$

They are located at $d_{UE} = 25m$ from the S-BS. Each of them has the whole bandwidth B available for themselves. The achievable $SINR_{UE}$ is only limited by the thermal noise of the receiver and the interference from the backhaul beams. First, the SNR_{UE} is evaluated with (2.31) where the thermal

noise on the whole band is now $N_{th_{dBm}} = 10 \times \log_{10}\left(\frac{k_b \times T_0 \times B}{1mW}\right)$. Then, the $SINR_{UE}$ is evaluated using (2.33). This gives an achievable $SINR_{UE}$ of $25dB$ and an instantaneous peak rate of $8Gb/s$. The average uplink peak rate is a fourth of that due to TDD duplexing and pilot transmission, giving $UL_{peak} = 2Gb/s$. In these conditions the cell throughput is $UL_{peak_{thpt}} = N_{users} \times UL_{peak} = 8Gb/s$. The uplink efficiency is then:

$$UL_{peak_{eff}} = \frac{0.001 \times 10^{\frac{P_{tot_{dBm}}}{10}}}{UL_{peak_{thpt}}} = 2.4\, pW \cdot b^{-1} \cdot s \tag{A.2.38}$$

2.12.4 Downlink average rate

The analysis of the downlink average rate is very similar to the uplink. There are three differences. The first one is that there is no pilot transmission, so the average rate is half of the instantaneous rate. Therefor $DL_{avg} = 200Mb/s$ and $DL_{avg_{inst}} = 400Mb/s$. This is the same instantaneous rate as for uplink leading to the same SINR. The second difference is that the PA output power expression now becomes:

$$P_{t_{dBm}}(d_{UE}) = SNR - Att_{Ch}(d_{UE}) + \left(N_{th_{dBm}} + NF\right) - 20 \times \log_{10}(N_{ant}) \tag{A.2.39}$$

Its dependence on N_{ant} is different because the antenna array is on the transmit side. The last difference has the same cause, i.e. the S-BS is the transmitter, meaning the number of PAs is the number of antennas time the number of sectors $N_{PAs} = N_{ant} \times N_{sect} = 1024$ and the total PA output power is:

$$P_{tot_{dBm}} = P_{t_{dBm_{avg}}} + 10 \times \log_{10}(N_{users} \times R_{t_{on}} \times N_{PAs}) = 17.6dBm \tag{A.2.40}$$

It may seem surprising that it is the same as for uplink, but it is no coincidence. In the uplink, the antenna array at the S-BS improves its sensitivity by $10 \times \log_{10}(N_{ant})$ allowing to reduce the user PA output power by the same amount. In the downlink the PA out power decreases as $20 \times \log_{10}(N_{ant})$, but because there is now one PA per antenna, there is in total N_{ant} times more PAs, increasing the power consumption as $10 \times \log_{10}(N_{ant})$. The result is a reduction in PA output power by only $10 \times \log_{10}(N_{ant})$, the same as in the uplink case.

Finally, the cell throughput is twice that of uplink at $DL_{avg_{thpt}} = 80Gb/s$ and the link efficiency is:

$$DL_{avg_{eff}} = \frac{0.001 \times 10^{\frac{P_{tot_{dBm}}}{10}}}{DL_{avg_{thpt}}} = 0.7\, pW \cdot b^{-1} \cdot s \tag{A.2.41}$$

The efficiency is twice as good as for uplink because there is no pilot transmission in the downlink.

2.12.5 Downlink peak rate

The achievable SINR is limited by the available PA output power, i.e. $10dBm$. There is a slight difference with uplink because here the PA must also provide the power for the backhaul beams, so not all of the 10dBm are available to build the $SINR_{UE}$. The setup is the same as the uplink peak rate except that it is now the S-BS sending data to the UE. The first step is to evaluate the PA output power required for the backhaul beam. Using (A4.4), recalled here, provides the SNR_{BH} as a function of SNR_{UE}.

$$SNR_{BH} = \frac{SINR_{BH} \times \left(1 + N_{beam_{UE}} \times SNR_{UE} \times IRR_{UE}\right)}{1 - \left(N_{beam_{BH}} - 1\right) \times SINR_{BH} \times IRR_{BH}} \tag{A.2.42}$$

The PA output power required to provide this SNR_{BH} is as follow:

$$P_{t_{BH}} = SNR_{BH} \times \frac{k_b \times T \times B \times 10^{\frac{NF}{10}}}{Att_{Ch_{lin}}(d_{BH}) \times N_{ant}^3} \tag{A.2.43}$$

Since there is an antenna array on both sides of the link the PA output power is reduced by N_{ant}^3. The available PA output power for the user link is then:

$$P_{t_{UE}} = 10^{\frac{10dBm}{10}} - N_{beam_{BH}} \times P_{t_{BH}} \tag{A.2.44}$$

And the SNR_{UE} is:

$$SNR_{UE} = \frac{P_{t_{UE}}}{k_b \times T \times B \times 10^{\frac{NF}{10}}} = SNR_{UE_{NI}} - N_{beam_{BH}} \times SNR_{BH} \times \frac{d_{BH}^2}{d_{UE}^2 \times N_{ant}} \tag{A.2.45}$$

With $SNR_{UE_{NI}}$ the available SNR if there is no backhaul interference.

$$SNR_{UE_{NI}} = 0.001 \times 10^{\frac{10dBm}{10}} \times \frac{Att_{Ch_{lin}}(d_{UE}) \times N_{ant}^2}{k_b \times T \times B \times 10^{\frac{NF}{10}}} \tag{A.2.46}$$

Plugging (A.2.22) into (A.2.46) gives:

$$SNR_{UE} = SNR_{UE_{NI}} - \frac{SINR_{BH} \times \left(1 + N_{beam_{UE}} \times SNR_{UE} \times IRR_{UE}\right)}{1 - \left(N_{beam_{BH}} - 1\right) \times SINR_{BH} \times IRR_{BH}} \times \frac{N_{beam_{BH}} \times d_{BH}^2}{d_{UE}^2 \times N_{ant}} \tag{A.2.47}$$

It can then be re-arranged to give the following expression of SNR_{UE}:

$$SNR_{UE} = \frac{SNR_{UE_{NI}} \times \left(1 - \left(N_{beam_{BH}} - 1\right) \times SINR_{BH} \times IRR_{BH}\right) - SINR_{BH} \times \frac{N_{beam_{BH}} \times d_{BH}^2}{d_{UE}^2 \times N_{ant}}}{1 - \left(\left(N_{beam_{BH}} - 1\right) \times IRR_{BH} - N_{beam_{UE}} \times IRR_{UE} \times \frac{N_{beam_{BH}} \times d_{BH}^2}{d_{UE}^2 \times N_{ant}}\right) \times SINR_{BH}} \tag{A.2.48}$$

The numerical evaluation gives $SNR_{UE} = 50.7dB$ and $SNR_{BH} = 45dB$. Using (A.2.20) gives $SINR_{UE} = 42dB$ providing a downlink peak rate of $DL_{peak} = 7Gb/s$. Looking at the difference between SNR and SINR for both user and backhaul links it is clear that the performances are limited by interference. The PA output power dedicated to the backhaul links and user links are:

$$P_{BH_{dBm}} = 10 \times \log_{10}\left(N_{beam_{BH}} \times P_{t_{BH}}\right) = -4.8dBm \tag{A.2.49}$$

$$P_{UE_{dBm}} = 10 \times \log_{10}\left(N_{beam_{UE}} \times P_{t_{UE}}\right) = 9.9dBm \tag{A.2.50}$$

Most of the PA output power is dedicated to the user links. Finally, the user link efficiency is expressed as:

$$DL_{peak_{eff}} = \frac{R_{t_{on}} \times N_{ant} \times 0.001 \times 10^{\frac{P_{UE_{dBm}}}{10}}}{DL_{peak}} = 177pW \cdot b^{-1} \cdot s \tag{A.2.51}$$

3 CHAPTER III: RECEIVER SPECIFICATIONS

Now that the proposed system is well described, a more detailed analysis of the S-BS receiver, the main focus of this manuscript, can be done. It has been shown that only a DBF or an HBF architecture is possible because of the requirements of the multiple beam capability. In this manuscript, the focus will be put on the DBF architecture since it is the most promising. To begin with in this chapter, a specification of such a receiver compatible with the system described in the previous chapter will be established. From that specification, a feasibility study of DBF receivers will be made. To do so, in a first step, the building block specifications of a given receiver architecture will be derived. Then, based on a bibliographic study of recent literature, a performance and power consumption evaluation will be performed.

3.1 HIGH LEVEL RECEIVER'S SPECIFICATIONS

From the previous chapter's analysis, the receiver's specifications will be derived. A Digital Beamforming Receiver (Rx) is composed of an antenna array where each of them is connected to a Single Receiver (SRx). Here, the interest is only in the analog part of the receiver, which starts at the antenna and ends right after the ADC. For each of the metrics of interest, the whole receiver and the single receivers will be specified when it make sense. A system with 256 antennas is assumed, with all the assumption made in the previous chapter.

3.1.1 Center Frequency, Bandwidth, Noise Figure and Sensitivity

The three first metrics are the initial working hypothesis. The center frequency is $28GHz$ and the band of interest is $1GHz$ around that center frequency. This means that the receiver must have at least $1GHz$ of input bandwidth and that the sampling rate of the output digital should be at least $2Gsps$. These specifications are the same for the Rx and the SRx.

The hypothesis that SRx have a Noise Figure of $NF_{SRx} = 10dB$ was also used. From equation (2.30), it is known that the whole receiver NF_{Rx} is the same the single receiver NF_{SRx}.

It is also common to specify the sensitivity of a receiver. By definition, it is the minimum received power allowing proper communication, i.e. the SNR at the Rx output is high enough to ensure the target data rate. In the present case this can be reformulated as the minimum amount of power the receiver needs to deliver an output signal with the desired SNR. This metric is defined only at the Rx level, since the SRxs by themselves are not foreseen to deliver a target data rate. If one sub-channel of $B_{ch} = 100MHz$ with a target $SNR = 18dB$ is considered, then the receiver's sensitivity is expressed as:

$$P_{Sensi} = 10 \times \log_{10}\left(\frac{k_b \times T \times B_{ch}}{1mW}\right) + NF_{Rx} + SNR - 10 \times \log_{10}(N_{ant}) = -90dBm \qquad (3.1)$$

Generally, a standard defines a reference sensitivity for receivers working in normal conditions. In the next sections, the value from (3.1) will be used as the reference sensitivity.

3.1.2 Linearity and Local Oscillator's Phase Noise

These two characteristics need to be specified together because they affect the receiver's performance under the same circumstances, i.e. in the presence of a strong out of band interferer. In general, a standard will allow for some amount of desensitization D of the receiver in these conditions. For 5G millimeter wave this amount is not yet defined, but classical values range from $3dB$ to $6dB$. Because millimeter wave circuit design is challenging, the later value of $D = 6dB$, which is more relaxed, will be assumed.

The impact of the receiver's non-linearity and Local Oscillator (LO) Phase Noise (PN) can be characterized by the amount of disturbing power P_d that will be added in channel 5 from the presence of the OoBI. From the desensitization definition, it is known that the cumulated added power should not degrade the receiver's sensitivity by more than $6dB$. It is equivalent to a degradation of the Noise Figure by the same amount. The degraded Noise Figure is then expressed as:

$$NF + D = 10 \times \log_{10}\left(\frac{N_{th} + N_{Rx} + P_d}{N_{th}}\right) \tag{3.2}$$

The acceptable disturbing power P_d is obtained by re-arranging (3.2) and expressing N_{Rx} as a function of $N_{th_{dBm}}$ and NF:

$$P_{d_{dBm}} = N_{th_{dBm}} + NF + 10 \times \log_{10}\left(10^{\frac{D}{10}} - 1\right) = -103dBm \tag{3.3}$$

Because there is no a priori on the relative contribution of the non-linearity and the LO phase noise, the assumption will be made that each of them can contribute up to half of P_d or $-106dBm$. The disturbing mechanisms can now be studied more deeply.

3.1.2.1 Linearity

The effect of non-linearity is the same as the one studied in section 2.6.2.2, namely spectral regrowth. The difference is that the input signal is the OoBI, and the relevant parameter is the power leaked in the fifth channel. The power leaked into the lower channels, one to four, is equally important, but, by nature, the non-linearity impacts reduce as the frequency shift from the OoBI increases. Hence the fifth channel will be the most affected one. If the degradations are acceptable for it, they will also be acceptable for the lower channels.

The proposed specification is that the leaked power should not exceed half of P_d in the presence of the OoBI. This is not a common way of specifying linearity, so it needs to be translated into a more standard specification such as the 3rd order Input Intercept Point (IIP3). There is a closed form expression derived in [3-1] that relates the IIP3 to the ACLR for a multi-tone input. Using this closed form is not straight forward since it depends on the number of tones used to model the input signal. This requires a way of choosing this number which is not detailed in [3-1].

Instead, a pure simulation approach is used. A choice must be made about the way non-linearity is modeled. A first solution would be to use a simple third order model. This has some limitations when the input signal becomes too large. After producing the initial desired gain compression behavior, the third order contribution starts dominating the fundamental signal, leading unrealistic gain regrowth. Modeling the non-linearity with a finite number of harmonic contributors will always lead to the same result at some point. Since the receiver is not expected to be used in a strongly non-linear region, this might not be an issue, but evaluating the limit where the model starts to deviate from a physical behavior is difficult.

To alleviate this issue, the non-linearity is modeled by an arctangent function and the input and output signals are scaled such that the third order non-linearity corresponds to a given IIP3. The output signal is scaled such that the linear gain is one. The arctangent being a function bounded between $[-\pi/2\,;\,\pi/2]$, monotonous and growing, it ensures a gain compression only behavior. Its bounded property even models gain saturation. Different functions with the same properties could have been used, but, since the exact non-linear behavior will depend on the currently unknown receiver's implementation details, it is difficult to assess which function is the best model. The choice on arctangent is, to some extent, arbitrary. One must recall that the best suited method is to evaluate the power leaked in the fifth channel using the non-linearity of the chosen implementation.

Figure 3-1 plots this leaked power in the fifth channel from spectral regrowth for various values of IIP3, using the arctangent non-linearity model described above, when the input signal is the OoBI at $P_I = -50dBm$. The arctangent linear gain is set to one to obtain directly the input referred distortion. To limit this leaked power below the targeted $-106dBm$, the IIP3 must be $-27dBm$ or better. When the receiver's non-linearity is caused by cascaded building blocks, for the same overall IIP3, the spectral regrowth is slightly worse since building blocks down the chain will also have intermodulation products between the signal and the spectral regrowth from the previous building blocks. For that reason, additional margin will be taken and the $IIP3$ specification will be set at $-25dBm$. In theory, non-linearity contributes to the receiver's desensitization in a second way which is gain compression. One can see from the top and bottom right graphs of Figure 3-1 that the power difference between the input and the output signal is $0.3dB$. This means that the level of linearity required for acceptable spectral regrowth renders gain compression negligible.

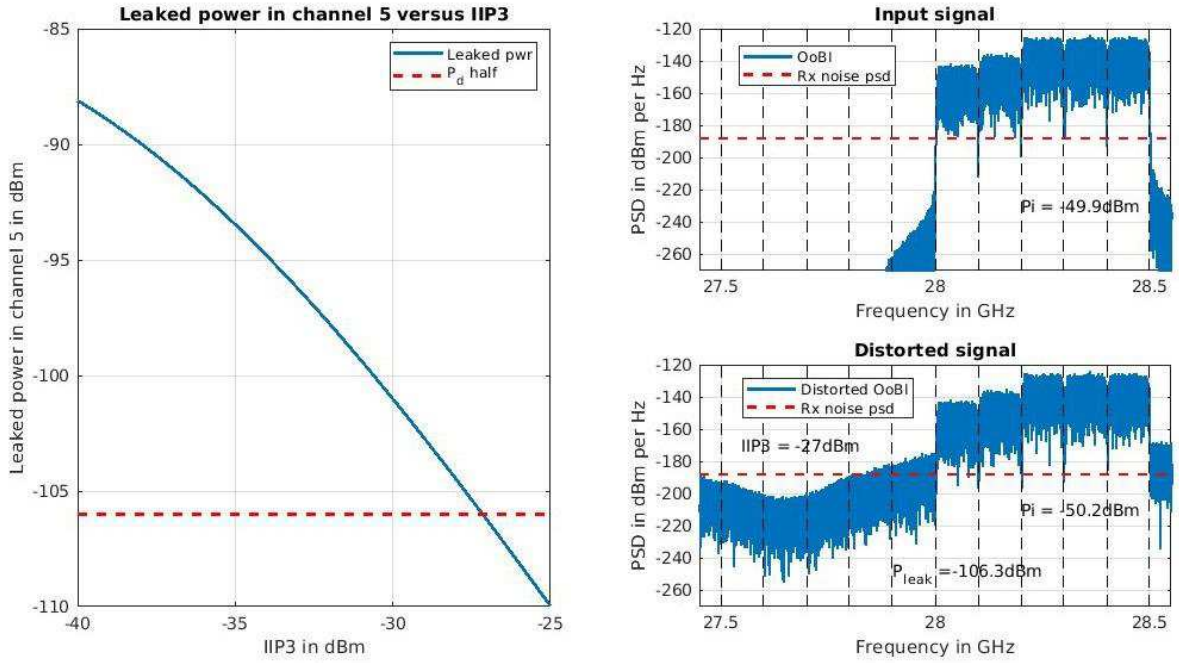

Figure 3-1: Left graph: Leaked power in channel 5 from non-linearity versus IIP3. Top right graph: input signal. Bottom right graph: Output distorted signal.

The specification derived above is for the whole receiver. It also holds for a SRx. The reason is that non-linearity produces the same effect when applied on the same signal, the OoBI in the current case. This means the disturbances will be largely correlated from one SRx to the next and will grow in a quadratic way with respect to the number of antennas. In other words, linearity does not benefit from the array factor as thermal noise does. This can potentially lead to stringent linearity requirement. Thankfully, the OoBI is known to be at some significant frequency offset from the channel of interest, allowing for more relaxed IIP3 requirement. This enlightens the benefit of the interfering signal precise description.

3.1.2.2 Local Oscillator (LO) Phase Noise

The local oscillator in a receiver is used to perform the frequency translation of the RF signal to an Intermediate Frequency (IF) or directly down to base band. Ideally, it is done through the mixing of the RF signal with a pure sinewave, the local oscillator output. Unfortunately, real oscillators do not produce ideal signals which introduces perturbations in the transmitted and received signal.

First, the nature of these non-idealities will be analyzed. This will be done through the study of oscillators in a first step, and their use in a Phase Lock Loop (PLL) in a second step. Then, their two main impacts on the received signal, random symbol rotation and reciprocal mixing, will be studied.

3.1.2.2.1 Oscillators

An oscillator can be described as a feedback system made voluntarily unstable in such way that it oscillates at the desired frequency. The block diagram of such a system is depicted on Figure 3-2-a.

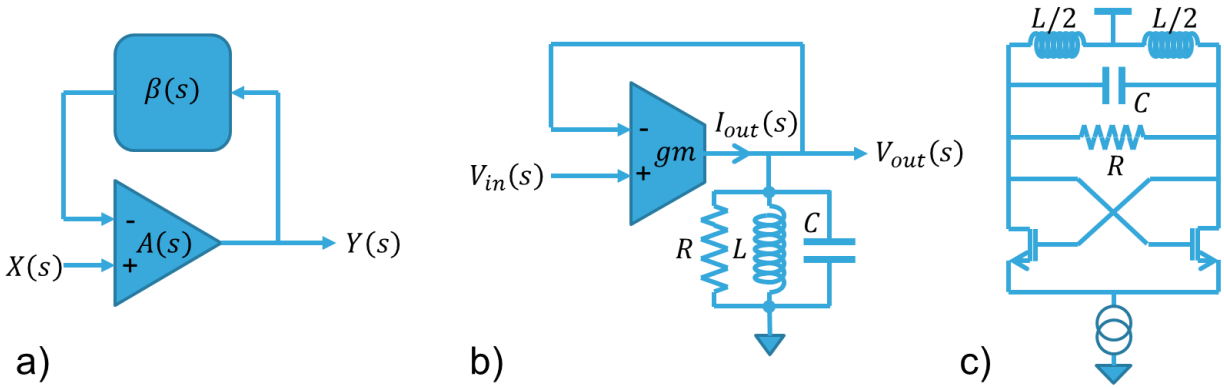

Figure 3-2: a) Linear Time Invariant (LTI) model of an oscillator b) Block diagram for an RLC resonator c) Differential implementation using a cross-coupled pair for negative trans-conductance

Equation (3.4) gives its output Laplace transform. The transfer function (3.5) is obtained after some re-arrangements.

$$Y(s) = \big(X(s) - \beta(s) \times Y(s)\big) \times A(s) \tag{3.4}$$

$$H(s) = \frac{A(s)}{1 + A(s) \times \beta(s)} \tag{3.5}$$

For this system to be unstable, its transfer function must have at least one pole on the right half part of the complex plan. For real systems, the poles come by complex conjugate pairs. The particular configuration having a single pair of poles sitting on the imaginary axis at $\pm j \times \omega_0$ describes a harmonic oscillator. This translates into the condition expressed by (3.6):

$$A(j \times \omega_0) \times \beta(j \times \omega_0) = A(-j \times \omega_0) \times \beta(-j \times \omega_0) = -1 \tag{3.6}$$

This is equivalent for the open loop gain $A(s) \times \beta(s)$ to provide 180° of phase shift at its unity gain frequency ω_0. The negative feedback is turned into a positive one leading to the desired instability.

The block diagram of a parallel RLC oscillator is depicted on Figure 3-2-b. A trans-conductance pushes current into an RLC resonant load. The load output voltage is then negatively fed back to the trans-conductance input. The transfer function of this system is given by (3.7):

$$H_{RLC}(s) = \frac{gm \times Z_{RLC}(s)}{1 + gm \times Z_{RLC}(s)} \tag{3.7}$$

With:

$$Z_{RLC}(s) = \frac{\frac{s}{C}}{s^2 + \frac{1}{R \times C} \times s + \frac{1}{L \times C}} = \frac{\frac{s}{C}}{s^2 + \frac{\omega_0}{Q} \times s + \omega_0^2} \tag{3.8}$$

And $\omega_0 = \frac{1}{\sqrt{L \times C}}$ and $Q = R \times \sqrt{\frac{C}{L}}$. Making the correspondence with the linear model gives $A(s) = gm \times Z_{RLC}(s)$ and $\beta(s) = 1$. At ω_0 the open loop gain is:

$$A(j \times \omega_0) \times \beta(\mathrm{j} \times \omega_0) = \frac{gm \times Q}{\omega_0 \times C} = gm \times R \tag{3.9}$$

Oscillation condition gives the intuitive condition on $gm = -\frac{1}{R}$. It must compensate for the resistive losses in the parallel RLC tank.

A classical way of achieving this negative trans-conductance is the use of a cross-coupled pair in a differential structure, as depicted in Figure 3-2-c. Such practical implementation cannot guarantee the exact value of gm. In the model, if $|gm|$ is too small there is no oscillation and if it is too large the oscillation amplitude diverges toward infinity. In practice $|gm|$ is set sufficiently above the ideal value to ensure oscillation. A real trans-conductance will experience saturation on its output voltage preventing the oscillation amplitude to diverge.

In the ideal model, the oscillator needs an input signal to start. Once it has started, the input signal can be set to 0. In practice the intrinsic noise of the oscillator components will trigger the oscillation start and the input can be completely removed. It is said that an oscillator is an autonomous system. When it is running without any control it is called a free running oscillator. The output wave form can be modeled by a $2 \times \pi$-periodic function $p(t)$ and characterized by its amplitude A_0 and frequency f_0:

$$v_{out}(t) = A_0 \times p(2 \times \pi \times f_0 \times t) \tag{3.10}$$

While the intrinsic noise is useful to trigger the oscillation start, it has detrimental effects on the output signal quality. In the next section, how exactly oscillators are affected by this noise will be studied.

3.1.2.2.2 Phase noise in free running oscillators

Device noises introduce perturbations in the oscillator's output signal in the shape of amplitude and phase errors and can be modeled as per (3.11):

$$v_{out}(t) = A_0 \times \big(1 + n(t)\big) \times p\big(2 \times \pi \times f_0 \times t + \phi(t)\big) \tag{3.11}$$

Because of the saturation happening in the oscillator the contribution to the total output noise from the amplitude noise is small. This is even more so when the oscillator's output is buffered by a limiting amplifier in order to create a square wave. For these reasons only the phase noise is to be considered.

The mechanism that transforms currents and voltages perturbations into phase error have been studied for decades. The modern theory providing satisfying predictions and insight for design optimization was laid down by Ali Hajimiri in his article "A General Theory of Phase Noise in Electrical Oscillators" [3-2]. The key aspect is his formulation of the problem using a Linear Time Varying (LTV) system. This is because the same perturbation applied at different times in the cycle will have different effects on the resulting phase noise. In this theory the oscillator can then be characterized through its Impulse Sensitivity Function (ISF), a periodic function with the same period as the oscillator which describes the oscillator phase noise sensitivity at each point in time of one cycle.

Since the ISF is a periodic signal, it can be decomposed into Fourier series components. The end result is that the noise near each of these components is down converted near DC in the phase term ϕ from (3.11). This noise down conversion is weighted by a coefficient that is proportional to the noise

frequency squared invers. In other words, the noise around the higher Fourier components will have less effect.

The next step is to understand how this phase noise affects the oscillator's output. If a sinewave is used as the periodic function in (3.11) and the amplitude noise is neglected for the reason previously mentioned, the following result is obtained:

$$\begin{aligned} v_{out}(t) &= A_0 \times sin\big(2 \times \pi \times f_0 \times t + \phi(t)\big) \\ &= A_0 \\ &\times \big(\sin(2 \times \pi \times f_0 \times t) \times \cos\big(\phi(t)\big) + \cos(2 \times \pi \times f_0 \times t) \times \sin\big(\phi(t)\big)\big) \end{aligned} \tag{3.12}$$

In the case of small $\phi(t)$ this can be approximated by:

$$v_{out}(t) \cong A_0 \times \big(\sin(2 \times \pi \times f_0 \times t) + \cos(2 \times \pi \times f_0 \times t) \times \phi(t)\big) \tag{3.13}$$

The phase noise is up converted around the carrier. For this reason, it is common to look at the Single Side Band (SSB) PSD of the oscillator. It is the noise PSD relative to the carrier power on one side of the carrier frequency expressed in decibel carrier per Hz (dBc/Hz). This gives a good estimation of the phase noise in the case of small variations. It is worth noting that this approximation is not always true, in particular for low frequency offsets. Nonetheless, it remains a useful measure that will be used later.

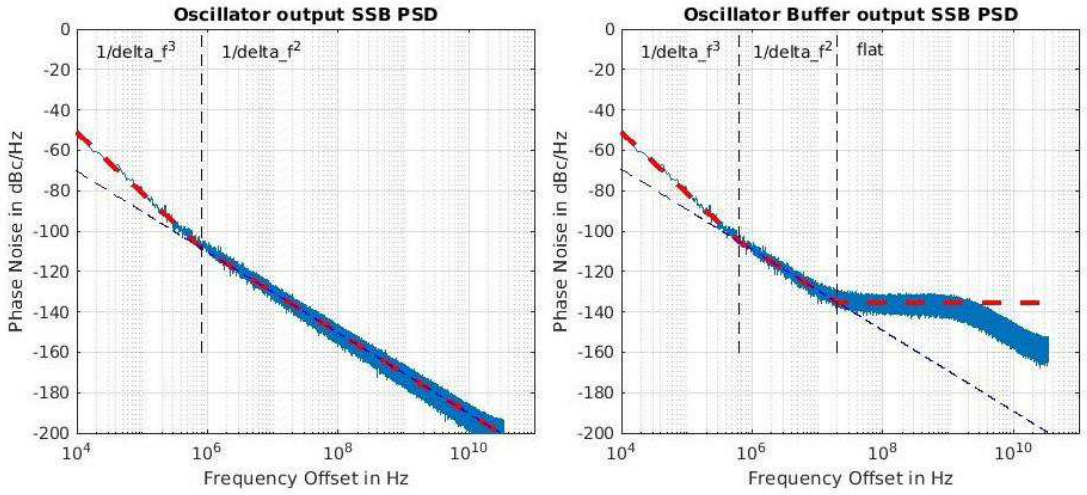

Figure 3-3: Typical Single Side Band PSD of a free running oscillator on the left and of its limiting amplifier output on the right

Figure 3-3 plots the typical case of a free running oscillator and its buffered output. The part of the spectrum with a $1/f^3$ slope correspond to the up conversion of devices Flicker noise near DC, the zeroth harmonic of the ISF. The $1/f^2$ zone correspond to the folding of the devices white noise around all the other ISF harmonics. To make use of the oscillators signal it must be buffered. This buffer will be outside the oscillator's loop; hence its noise will not be affected by the oscillator. It will simply add with it. Assuming the buffer noise is from thermal origin, it will white Gaussian noise. This buffer also has a limited bandwidth that will filter out its own noise past its cutoff frequency. An example of a buffered oscillator SSB PSD is depicted on the right graph of Figure 3-3.

Oscillators can be characterized by their two different corner frequencies and few points of their SSB PSD. It will be seen later that the colored noise and the white noise affect the system in different ways. The very low frequency phase noise can reach high power and can be regarded as frequency instability. This is highly undesirable because the receivers LO frequency must be the same as the transmitter's one for proper demodulation. Integrated oscillators are known to be relatively bad regarding that point, and often not useable as is. In the next section, how this can be mitigated will be looked at.

3.1.2.2.3 Phase Lock Loop

To overcome this issue, oscillators are often used in feedback system that aligns the oscillator's phase to that of an external reference. This is called a Phase Lock Loop (PLL). If the reference and the oscillator were running at the same frequency, one could directly use the reference signal and the PLL would be of no use. For this reason, PLLs are always used with a multiplying factor between the reference frequency and the oscillator's one. To be used in such a loop the oscillator frequency must be controllable, often through a control voltage. It is then called a Voltage Controlled Oscillator (VCO).

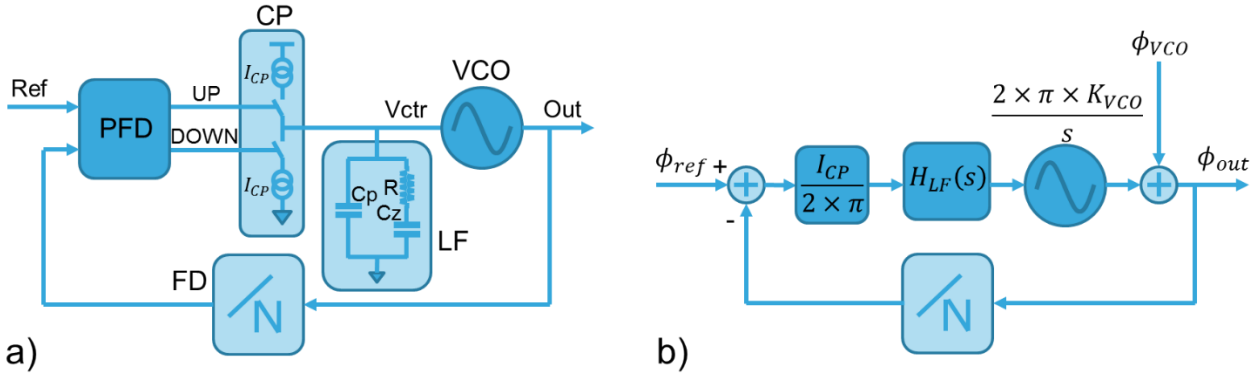

Figure 3-4: a) PLL Classical implementation b) Equivalent Linear Time Invariant (LTI) model

Figure 3-4-a depicts a classical implementation for PLLs. The VCO output goes through a Frequency Divider (FD) in the feedback part of the loop. It is then compared with the reference by a Phase Frequency Detector (PFD). This component measures the delay between the rising edges of its input signals. When the feedback signal edge is lagging behind the reference one, the PDF will issue an Up pulse with a width equal to the lag. When it is the reference signal edge that is lagging a Down pulse is issued, also with a width equal to the lag. These UP and DOWN pulses control a Charge Pump (CP). It will source or sink a constant current for the duration of the corresponding Up or Down pulses into the Loop Filter (LF). This generates the VCO control voltage. The LF allows to adjust the PLL bandwidth and ensure the loop stability.

Figure 3-4-b gives the block diagram of the equivalent Linear Time Invariant (LTI) model of the PLL. The PFD is modeled as continuous time linear subtractor and the CP by a linear gain equal to its current divided by $2 \times \pi$ to account for the transformation from time to phase of the PFD output. This linear approximation is acceptable as long as the LF is a Low Pass (LP) with a cutoff frequency much lower than the comparison rate of the PFD. The control voltage of the VCO adjusts its output frequency. The phase being the integral of the frequency, it is modeled by an integrator with a linear gain K_{VCO}. It represents the sensitivity of the VCO output frequency to the control voltage. The VCO's phase noise is modeled by an additive noise source at its output with the spectral characteristics of the buffered oscillator from Figure 3-3 right graph. The following equation (derived in Annex 3.1) allows to describe the output phase noise of the PLL:

$$\phi_{out}(s) = H_{\phi_{VCO}}(s) \times \phi_{VCO}(s) + H_{\phi_{ref}}(s) \times \phi_{ref}(s) \tag{3.14}$$

With:

$$H_{\phi_{ref}}(s) = N \times \frac{H_{OL}(s)}{1 + H_{OL}(s)} \tag{3.15}$$

$$H_{\phi_{VCO}}(s) = \frac{1}{1 + H_{OL}(s)} \tag{3.16}$$

Where ϕ_{out} is the total phase noise at the PLL output, $H_{\phi_{VCO}}$ is the close loop transfer function from the VCO noise input to the PLL output, $\phi_{VCO}(s)$ is the VCO's phase noise, $H_{\phi_{ref}}$ is the close loop transfer function from the reference input to the PLL output, ϕ_{ref} is the reference clock's phase noise and H_{OL} is the PLL open loop gain from the reference input to the divider by N output.

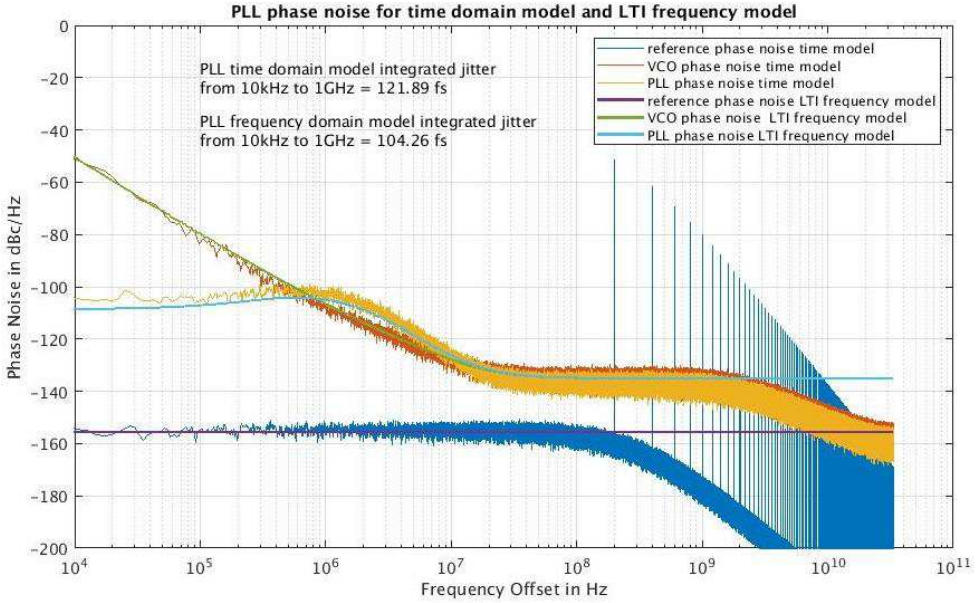

Figure 3-5: Phase Noise of a matlab time domain implementation of the PLL versus the LTI frequency model with a center frequency of 22.4GHz

The open loop gain $H_{OL}(s)$ has a low pass characteristic with infinite gain at DC. It is close to a double integrator. As a consequence, $H_{\phi_{ref}}(s)$ also has a low pass characteristic with a gain of N in the band. $H_{\phi_{VCO}}(s)$ is high pass with unit gain in its band. They have about the same cut off frequency. From (3.14) it can be deduced that the PLL output phase noise will be:

- N times that of the reference at low frequency
- the VCO phase noise at high frequency
- In between both for a transition zone

Figure 3-5 plots the SSB of a PLL using the VCO from Figure 3-3:

The PLL output phase has the expected characteristics. Now that the LO non-idealities have been well characterized and explained, their effects on the received signal can be studied. There are two of them and they will be looked at individually in the next two sections.

3.1.2.2.4 Random symbol rotation

Random symbol rotation is a major consequence of LO imperfections. To understand it, it is first necessary to describe the general idea of complex modulations. Then it will be seen how phase noise leads to symbol rotation.

Most of the modern modulations use complex constellations. This means that the bits to be transmitted are gathered in small groups of N to form a symbol. For one symbol there are 2^N possible configurations. Each of these configurations are mapped to a point of the complex plan in an orderly fashion. This is called the constellation. Figure 3-6 gives a classical constellation mapping for two-bit and four-bit symbols.

The main advantage of this mapping is that each symbol has only one-bit difference from its closest neighbors. When recovering the transmitted symbols, if the wrong symbol is estimated, it is more likely that this wrong symbol will be one of its closest neighbors. If this happens then there is only one bit of error instead of potentially four if the mapping was poorly chosen.

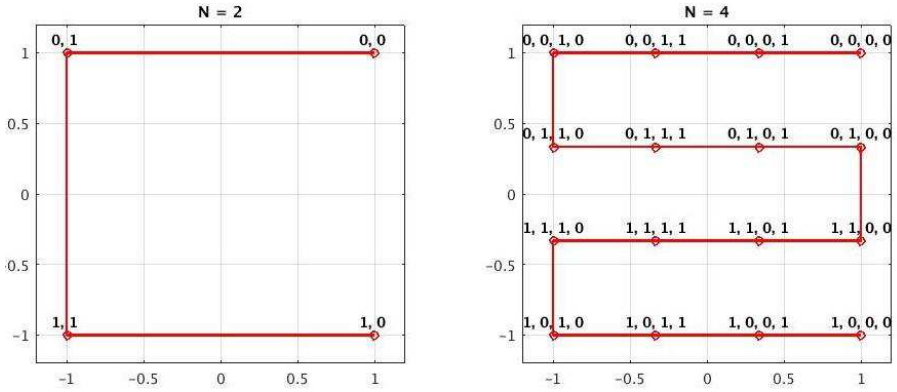

Figure 3-6: Symbol mapping for 2-bit symbols on the left and 4-bit symbols on the right

Once the symbols are mapped to their respective constellation points, their real parts are used to perform amplitude modulation on cosine wave. This forms the in-phase signal. Their imaginary parts are used for amplitude modulation on a sine wave. This forms the quadrature signal. Because in the periodic function space sine and cosine are orthogonal, the in phase and quadrature phase signals can be summed up and send over the same channel simultaneously while remaining separable at the receiver. Sine and cosine being in quadrature, meaning there are the same waveform with a 90° phase shift, this kind of modulation is called Quadrature Amplitude Modulation (QAM). This allows a straight doubling of the channel capacity for the same bandwidth occupation. For this reason, it is a must in nearly all modern wireless systems when high data rate and spectral efficiency are significant objectives.

Let us now look at the impact of PN on a QAM modulation. Let us assume the wireless communication chain from Figure 3-7. The Tx Base Band Processor (BBP) takes the input bit stream and generate the I and Q signals. They are then fed to Digital to Analog Converters (DAC) and up mixed in quadrature. The signal then travels through the RF channel made of the PA, the Tx antenna, over the air, the Rx antenna and finally the Low Noise Amplifier (LNA). It is then down mixed in quadrature to Base Band (BB). As it will be seen, there is an additional step of low pass filtering to remove the high frequency image produced during down mixing. Finally, the I and Q signals are digitized by ADCs and the transmitted symbols are estimated by the BBP to reproduce the transmitted bit stream.

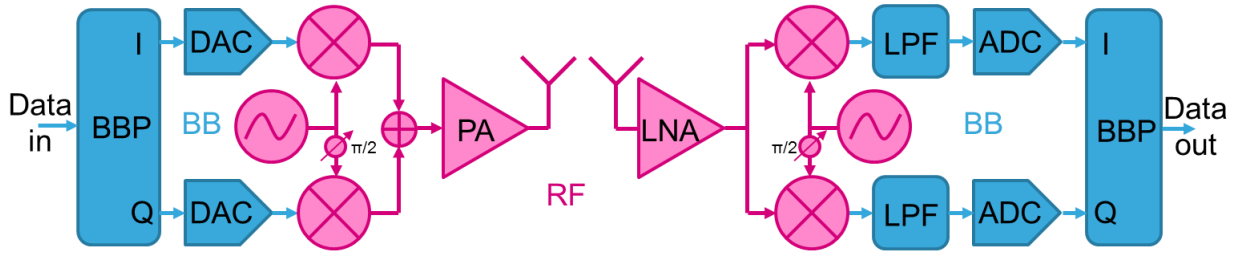

Figure 3-7: Block diagram of a wireless transmission chain with homodyne transmitter and receiver

Only the LOs effect are of interest here. To simplify the analysis, it will be assumed that all the blocks have unit gain and are ideal except for the local oscillators. The mixers perform a perfect multiplication between their two inputs, the DACs and ADCs have infinite resolution, and the Low Pass Filter (LPF) are perfect brick wall filters with unit gain and linear phase in the band and zero gain outside. Both Tx and Rx LOs have the same frequency and a random but fix phase difference φ. Without loss of

generality, it can be assumed that the phase of the Tx LO is zero at the origin of time. Finally, both of them are tinted with their respective phase noise ϕ_{TX} and ϕ_{RX}.

The RF signal $S_{RF_{TX}}$ sent by the transmitter can then be expressed as follow:

$$S_{RF_{TX}}(t) = I(t) \times \cos(\omega_{LO} \times t + \phi_{TX}) + Q(t) \times \sin(\omega_{LO} \times t + \phi_{TX}) \tag{3.17}$$

If the base band I/Q signals are expressed as a complex value, in terms of its modulus ρ and argument θ, as in (3.18) the transmitted RF signal $S_{RF_{TX}}$ can be rewritten in the more compact way of (3.19).

$$S_{BB_{TX}}(t) = I(t) + j \times Q(t) = \rho(t) \times e^{j \times \theta(t)} \tag{3.18}$$

$$\begin{aligned} S_{RF_{TX}}(t) &= Re\left[S_{BB_{TX}}(t) \times e^{-j \times \omega_{LO} \times t}\right] = Re\left[\rho(t) \times e^{-j \times (\omega_{LO} \times t + \phi_{TX} - \theta(t))}\right] \\ &= \rho(t) \times \cos\left(\omega_{LO} \times t + \phi_{TX}(t) - \theta(t)\right) \end{aligned} \tag{3.19}$$

One can note from (3.19) that a QAM modulation is a special case of phase and amplitude modulation. As a consequence, most of the results that will be derived here are generalizable to any such single carrier modulations. The received in phase and quadrature phase signal $I_{RX}(t)$ and $Q_{RX}(t)$ at the ADCs' output in Fig. 3.8 are expressed by (3.20) and (3.21) where the LPF performs image filtering:

$$\begin{aligned} I_{RX}(t) &= LPF\left(\rho(t) \times \cos\left(\omega_{LO} \times t + \varphi + \phi_{RX}(t)\right) \times \cos\left(\omega_{LO} \times t + \phi_{TX}(t) - \theta(t)\right)\right) \\ &= \frac{\rho(t)}{2} \times \cos\left(\theta(t) + \varphi + \phi_{RX}(t) - \phi_{TX}(t)\right) \end{aligned} \tag{3.20}$$

$$\begin{aligned} Q_{RX}(t) &= LPF\left(\rho(t) \times \sin\left(\omega_{LO} \times t + \varphi + \phi_{RX}(t)\right) \times \cos\left(\omega_{LO} \times t + \phi_{TX}(t) - \theta(t)\right)\right) \\ &= \frac{\rho(t)}{2} \times \sin\left(\theta(t) + \varphi + \phi_{RX}(t) - \phi_{TX}(t)\right) \end{aligned} \tag{3.21}$$

Finally, the received complex base band signal $S_{BB_{RX}}(t)$ can be expressed as:

$$S_{BB_{RX}}(t) = \frac{\rho(t)}{2} \times e^{j \times (\theta(t) + \varphi + \phi_{RX}(t) - \phi_{TX}(t))} = \frac{1}{2} \times S_{BB_{TX}}(t) \times e^{j \times (\varphi + \phi_{RX}(t) - \phi_{TX}(t))} \tag{3.22}$$

From this expression it is easy to see that the received symbols undergo a fixed rotation of φ from the fixed phase difference between Tx and Rx LOs, and a random rotation from the cumulated Tx and Rx LO phase noise. As mentioned previously this result is independent of the values of $\rho(t)$ and $\theta(t)$ and holds for any single carrier amplitude and phase modulations.

Figure 3-8 plots the received symbols from a wireless link only impaired by Tx and Rx LO phase noise for a 16-QAM and a 256-QAM modulation. The Tx and Rx phase noise are uncorrelated and have the characteristics of Figure 3-5. As expected, one can see a fix rotation of the entire constellation from the fix phase difference between Tx and Rx LOs. This is not an issue since it can be easily evaluated during pilot reception and then compensated. Also, the symbol random rotation, that spreads in a circular shape the received symbols' location, can be seen. On one hand, for the 16-QAM the phase noise seems unlikely to lead to a symbol estimation error. On the other hand, all the outer constellation points of the 256-QAM are nearly merged and will most certainly lead to symbol estimation errors. There are two visible trends here. The first one is that constellation points further away from the center get more spread out. The second one is that denser constellations are more sensitive to this phenomenon.

This variable impact depending on the constellation locus makes it inadequate to specify phase noise just as an SNR loss. This is because increasing the signal power does not reduce the problem, as phase noise is injected during a multiplication type operation by the mixer. Another point to note is that it is more likely for a symbol to collide with a neighboring one that is not one of its closest neighbors. This leads to more bit errors for the same amount of symbol errors.

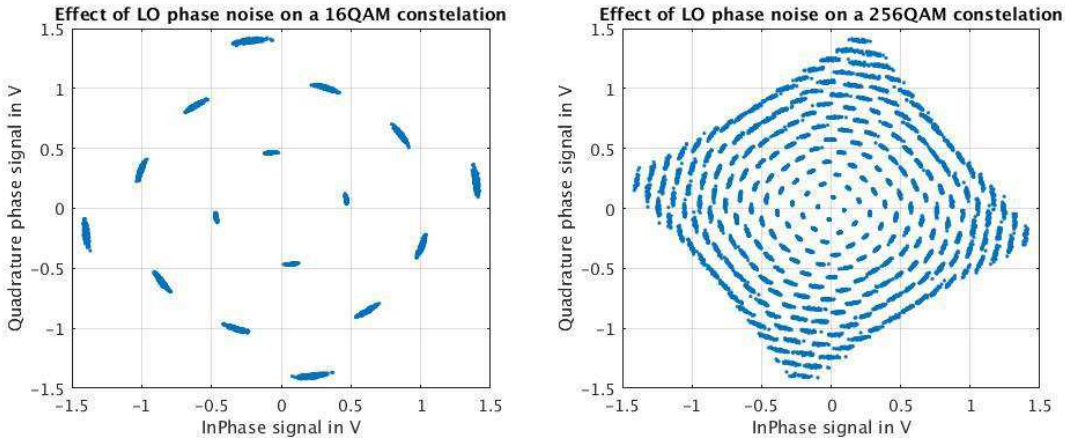

Figure 3-8: Impact of phase noise on received QAM symbols for 100us long symbol stream at 100Msymbols/s

A positive aspect is that this rotation is not impacted by the entire phase noise spectrum. In a TDD system, as the one considered here, the data are processed by the BBP by packets of one time slot, T_{TDD}~$100\mu s$ as it has been seen before. This means that the very low frequency phase noise will not have the time to develop and will only appear as a fix phase offset that can be evaluated with the pilot. In the present case this means that the phase noise below $f_{min} \approx \frac{1}{T_{TDD}} = 10kHz$ is not a big concern.

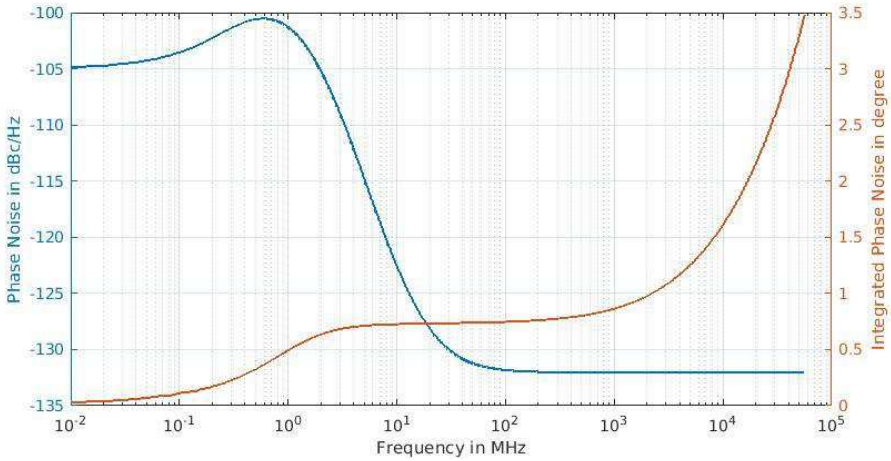

Figure 3-9: PLL phase noise and integrated phase noise as a function of frequency

Let us now consider the frequencies above $10kHz$. Fig. 3.10 plots the running integral of the PLL's phase noise described earlier.

It can be seen that, in the lower frequencies, there is a strong contribution around the PLL bandwidth. It is followed by a relatively flat portion up to about $1GHz$. The flat noise contribution below that frequency remains negligible compared to the colored noise contribution. Finally, when integrating the frequencies above 1GHz the flat noise starts to bring a significant contribution and seems to blow off toward infinity. In practice the PLL output buffer have a finite bandwidth such that the integrated phase will reach a plateau. But since this buffer also acts as a limiting amplifier to recreate a square wave, its bandwidth is typically in the range of three times the LO frequency. This means that, in the present case, the frequency offset is close to $50GHz$ away from the carrier, which is very high.

This very high buffer bandwidth may seem like an issue. To understand that it is not the case, it is necessary to look at the impact in the frequency domain of such phase noise on the modulated signal. As for the LO, high frequency phase noise will appear in the spectrum of the modulated signal as noise at high frequency offset from its center frequency. Any part outside the channel bandwidth can potentially be filtered. It is a legitimate question that, once a symbol underwent a rotation, will filtering undo this rotation? To know if this the case, let us run an experiment. Let us assume the system uses a 16-QAM modulation with Root Raise Cosine (RRC) match filters both at Tx and Rx. Figure 3-10 plots the received BB signal spectrum and the received symbols for three different cases. From left to right, the case without any phase noise, the case when phase noise is added after the Rx match filter and finally the case when it is added in between Tx and Rx match filters are displayed. The last case being the more realistic one.

As expected, the received symbols for the first case are un-rotated. As previously mentioned, the spectrum for the unfiltered case displays a wide band noise outside the channel bandwidth. Also, the received symbols are strongly rotated. Finally, the third case, where the phase noise outside the channel is filtered, only displays a small amount of symbol rotation.

From that experiment, the conclusion can be reached that, in order to evaluate the RMS phase noise, it is only necessary to worry about the frequencies up to the channel bandwidth or so. There is no need to be accurate on this higher bound since the contribution in that region is negligible. Finally, it can be seen that random symbol rotation is mostly affected by low frequency colored phase noise around the PLL bandwidth upper hand. This is often referred to as close in phase noise.

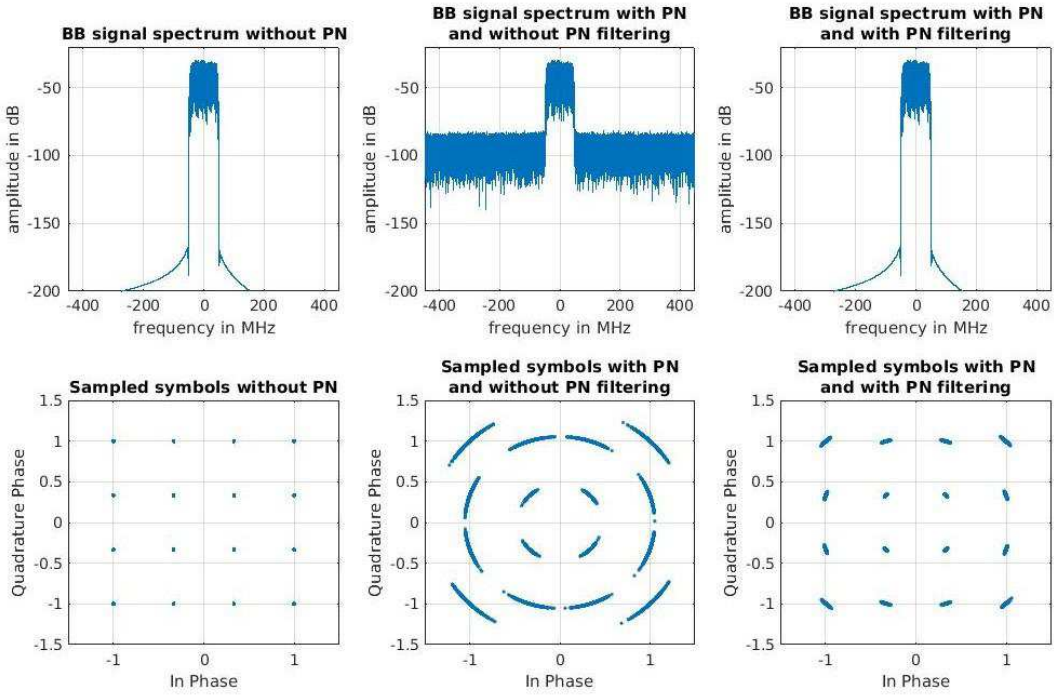

Figure 3-10: Received signal spectrum and received symbols for three different cases: left) No phase noise. center) Phase noise added after filtering. right) Phase noise added before filtering

Even if some effects of phase noise have been analyzed, some work is still needed to derive some specifications.

First the different impact of phase noise for different constellation points may have large detrimental effect on the performances of the error correcting algorithms. This analysis is well beyond the scope of this manuscript and prevent any formal evaluation of data rate loss.

Second, different modulations are impacted differently. For example, Differential QAM (DQAM) which encodes the transmitted symbols as the difference between any two successive constellation points is much less affected by low frequency phase noise. This is because it introduces some processing similar to Correlated Double Sampling (CDS). Multi carrier modulations such as OFDM will experience an additional perturbation which is the loss of sub-carrier orthogonality introducing Inter Carrier Interference (ICI). Since the modulation that will be used in millimeter wave 5G is not yet well defined, it is not yet possible to accurately specify the LO close in phase noise required to reach the desired data rate.

Third and last, the link performance is affected by both Tx and Rx phase noise. Since in the present case the available hardware will be highly asymmetrical, i.e. a user terminal on one hand and base station on the other, it is hard to know how to split the total acceptable close in phase noise between both ends of the link.

For all these reasons, the specifications of close in LO phase noise is left to future work.

3.1.2.2.5 Reciprocal mixing

Reciprocal mixing is the name given to the interaction between the out of band LO phase noise and a strong interferer during down mixing. Using the previously made analysis for a specific case, this interaction will be quantified. This will then allow to establish a specification for the receiver out of band LO phase noise.

As it has been seen, the mixing of the signal with a local oscillator high frequency phase noise give rise to a similar noise at the corresponding frequency offset and with a similar level of power compared to the mixed signal. When this signal is an interferer, this noise may fall into the band of the desired channel. Let us consider the Rx input signal S_{Rx} from (3.23):

$$S_{Rx}(t) = \rho_{ch}(t) \times \cos\left(\omega_{RF} \times t - \theta_{ch}(t)\right) + \rho_I(t) \times \cos\left(\left(\omega_{RF} + \Delta_{f_I}\right) \times t - \theta_I(t)\right) \tag{3.23}$$

It is composed of the desired signal and an interferer, namely users from different operators, operating in an adjacent band at a frequency offset Δ_{f_I}. When going through the mixer both get corrupted by the LO phase noise and the BB signal can be expressed as:

$$\begin{aligned} S_{BB_{RX}}(t) = \frac{1}{2} \times \big(&\rho_{ch}(t) \times e^{j\times(\theta_{ch}(t)+\phi_{RX}(t))} \\ &+ \rho_I(t) \times \mathrm{e}^{j\times\theta_I(t)} \times \mathrm{e}^{-j\times\Delta_{f_I}\times t} \times \mathrm{e}^{j\times\phi_{RX}(t)}\big) \end{aligned} \tag{3.24}$$

The interferer noise from phase noise mixing will appear on the desired channel as additive noise. What is necessary to evaluate is the amount of noise power falling into the desired channel. To do so, the interferer characteristics evaluated in the previous chapter will be used. In section 3.2.1, it was evaluated that the maximum acceptable noise power from reciprocal mixing is $P_{RM} = -106dBm$. It is now necessary to evaluate the maximum LO noise floor satisfying this condition. As the interferer frequency offset $\Delta_{f_I} \approx 300MHz$ is at a much larger frequency offset than the colored phase noise, it will not have any effect and can be ignored for that matter. Hence the interferer can be approximate by a single tone with the same total power P_I and a phase noise with a constant PSD $\phi_{LO_{PSD}}$ expressed in dBc. P_{RM} can then be expressed as:

$$P_{RM} = P_I + \phi_{LO_{PSD}} + 10 \times \log_{10}(B_{ch}) \tag{3.25}$$

And the desired phase noise floor as:

$$\phi_{LO_{PSD}} = P_{RM} - P_I - 10 \times \log_{10}(B_{ch}) = -136dBc \quad (3.26)$$

To confirm this value, a simulation where the interferer is made of modulated signals with a total power of -50dBm, mixed with the PLL output signal from the previous section, was run. Its phase noise floor is around -135dBc which suits the purpose. The results are plotted in Figure 3-11.

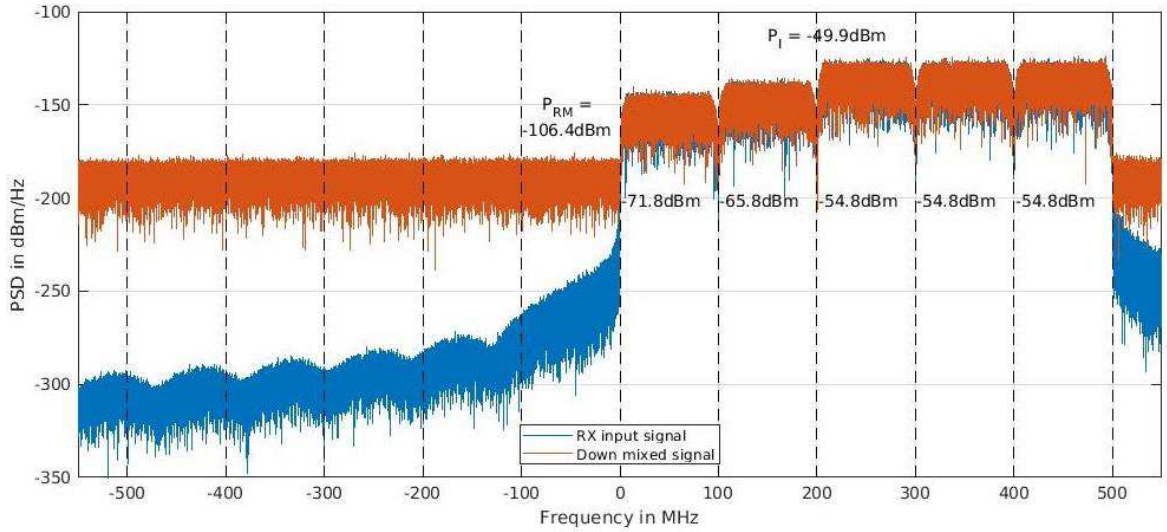

Figure 3-11: Evaluation of reciprocal mixing between a modulated interferer and a noisy LO

The blue signal is the received signal in the ideal case. The small amount of noise in channels one to five come from the truncation of the RRC pulse used for the match filter on the transmitter side. It remains well below the $-174dBm/Hz$ of thermal noise and has no noticeable effect. The red curve is the resulting signal from down mixing the RF signal with the noisy LO from Figure 3-5. With a flat phase noise of about $-135dBc/Hz$, a reciprocal mixing induced noise in channels one to five of about $-106dBm$ per channel is observed as predicted by (3.25). Note here that, compared to non-linearity it is not only channel five but all channels that are equally affected. With this simulation, the proposed LO flat phase noise specification of $-136dBc/Hz$ is confirmed.

Generally, a single LO is used for the whole receiver and, since the interfering signal is the same for all antennas, this specification is not relaxed by the array gain. If multiple LOs were to be used, as long as their noise are uncorrelated, this specification could be relaxed by $10 \times \log_{10}(N_{LO})$ where N_{LO} is the number of LO used, which can be different from the number of antennas. While this is always true for flat noise, since it is coming from the LO output buffer, it might not always be true for close in phase noise, if the PLLs are fed with the same reference. The optimization of the number of LOs and references is a topic in itself which is left to future work. The worst-case assumption of a single LO feeding all the SRx with a $-136dBc/Hz$ specification for the flat phase noise will be used.

3.1.3 Conclusion

In this section, the receiver's specifications for center frequency, bandwidth and sampling rate, Noise Figure and Sensitivity, Desensitization, linearity and LO flat phase noise were derived. It was done using the conclusions of the previous analysis. Even though the numerical values are based on hypothesis that might change, the methodology remains valid and will allow for a quick adjustment of the specifications if necessary. The problem of specifying the close in phase noise could not be analytically solved. Although a matlab simulation environment was set up and could allow for empirical specification, too many parameters, such as Tx/Rx budget splitting or modulation type, are missing for a proper evaluation. Also, the system optimization with respect to the number of LOs to be used is left to future work. Finally, while the current results are not entirely satisfactory, they are sufficient to

provide the order of magnitude of required performances. They will be used as a basis for the receiver's building blocks specifications.

3.2 FEASIBILITY OF DBF

This section will allow to reach the final conclusion from the system analysis, the demonstration that receivers dedicated to large antenna array digital beamforming millimeter wave are feasible. This final argumentation will go as follow. First, the Rx architecture will be described, and in particular its building blocks. Then, from the previous section's Rx specifications, the building block specifications will be derived. Finally, a bibliographical study will be performed for each of them. This last step has two purposes. First to determine if the specifications are simply reachable. Then to make a power consumption assessment to evaluate the attractiveness of the DBF approach for millimeter wave 5G.

3.2.1 Single Receiver architecture description

For the sake of clarity, a homodyne receiver architecture (Figure 3-12-a) has been assumed so far, i.e. the Rx LO frequency is the same as the RF signal center frequency. In practice this architecture is sometime avoided because it suffers from LO self-mixing. Due to the finite isolation between the mixer's inputs, the LO signal leaks into the RF one and is then mixed with itself. This results in a strong DC component in the down mixed base band signal which have multiple detrimental effects. In particular, until this DC component can be removed, generally in the digital domain, the receiver's dynamic range must be increased.

One solution to this problem is to down convert the signal not to base band but to a low intermediate frequency, just high enough such that DC falls outside the signal band. This approach is often called Near Zero Intermediate Frequency or Near-ZIF. The DC from self-mixing can then simply be removed by a DC block. Ideally this initial mixing does not need to be a quadrature one. Also, with a low enough intermediate frequency, the IF signal can be directly digitized with a single ADC with a bandwidth close to that of the signal (Figure 3-12-b).

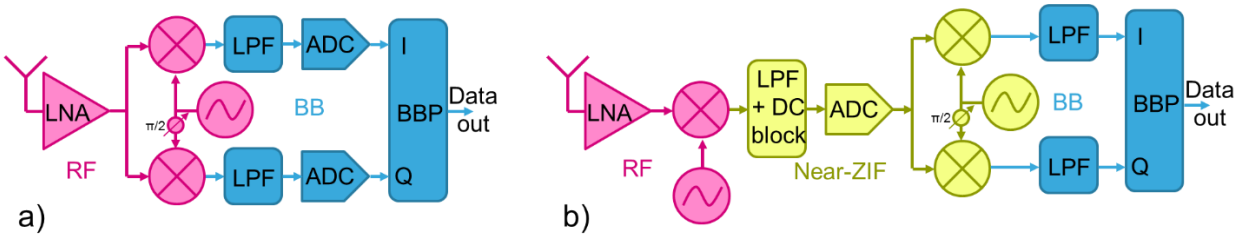

Figure 3-12: a) Rx homodyne architecture b) Rx Near-ZIF naive architecture

But this naive implementation would suffer from a nearly non-existing image frequency rejection. This is detailed in Figure 3-13. If the desired signal is contained in a 1GHz band around 28GHz and acceptable IF would be for example at 600MHz. The lower end of the signal band would sit at 100MHz, far enough from DC such that self-mixing can be removed easily, yet the upper end of the signal band would be at 1.1GHz, limiting the excess bandwidth requirement on the ADC compared to the signal bandwidth.

In that configuration the desired signal sits at $f_c = f_{LO} + f_{IF}$. The issue is that the image frequency $f_I = f_{LO} - f_{IF}$ will also be down converted at the IF and interfere with the desired signal. This is not to be confused with the high frequency signal image produced at $f_c + f_{LO}$ by the IF down-mixing. Here the discussion is about a different signal, an interferer, that would be sensed by the antenna, and which would sit at the image frequency $f_I = f_{LO} - f_{IF}$.

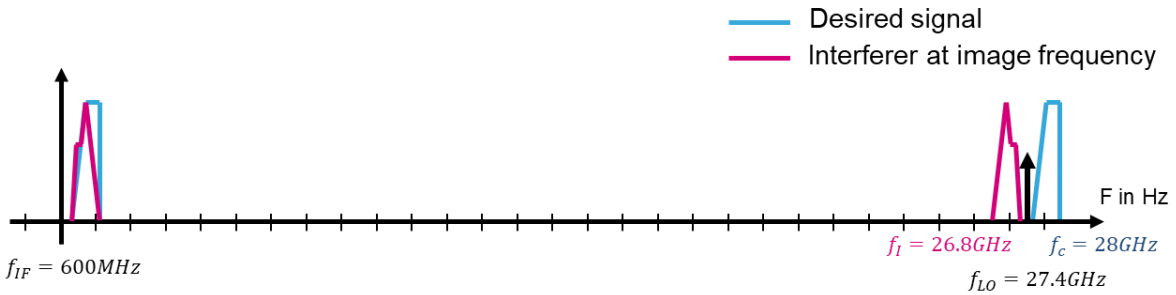

Figure 3-13: Impact of Image frequency on a Near-ZIF naive implementation

In a classic super-heterodyne receiver with a higher IF, the RF signal is bandpass filtered to remove this image frequency signal before down-mixing. In a Near-ZIF configuration the image frequency is too close to the RF to be properly filtered. The image frequency must be dealt with differently.

There are two common technics to do so, the Hartley (Figure 3-14) and the Weaver approaches (Figure 3-15). The details on how these two approaches work is given in Annex xx. They both require an RF quadrature down-mixing.

The Hartley solution applies a 90° phase shift on the quadrature-phase IF signal and recombines it with the in-phase IF signal to cancel the image signal. These phase-shift and recombination operations can be performed in the analog domain (Figure 3-14-a) or the digital one (Figure 3-14-b).

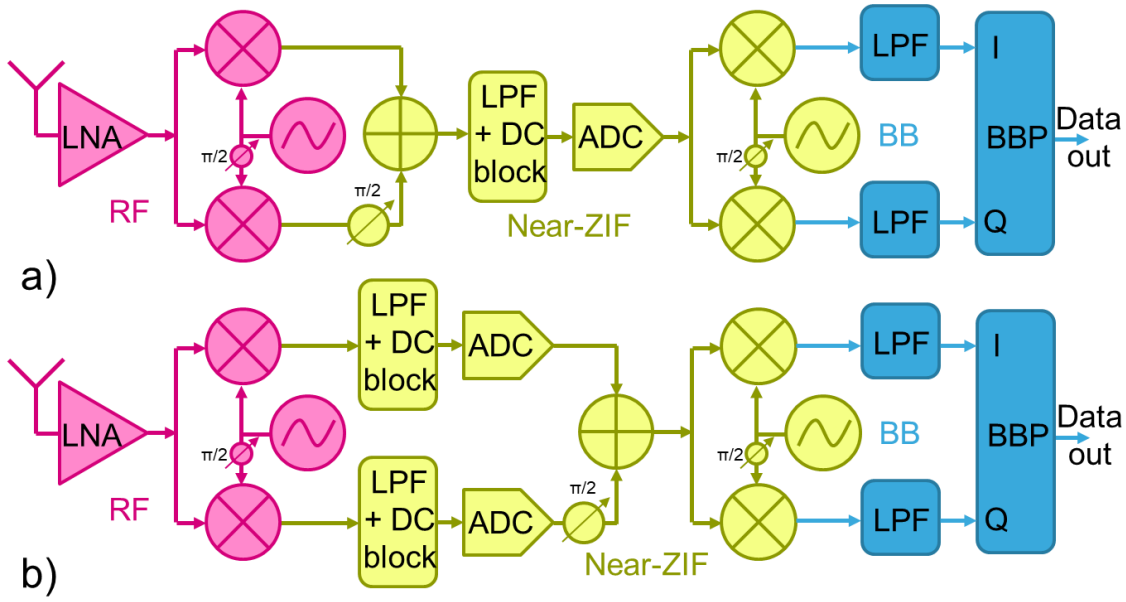

Figure 3-14: Near-ZIF receiver using the Hartley image cancelling approach a) with an analog implementation, b) with a digital implementation

The Weaver solution performs a quadrature down-mixing to base band for both the in-phase and the quadrature-phase IF signals. These four base band signals are then recombined appropriately to form the desired complex signal while cancelling the image signal. As for the Hartley approach, this can be implemented in the analog or the digital domain. Figure 3-15 displays a digital implementation of the Weaver solution.

Both approaches show similar performances in terms of image rejection ratio in the range of 30dB to 40dB. The limitation comes from the quadrature down-mixing accuracy from RF to IF, in the range of 2° or 3°, and the gain mismatch between the in-phase and quadrature-phase paths in the range of 0.2dB

or 0.3dB. The interfering signal at the image frequency can be of two kinds, the receiver's internal noise (including the antenna noise) or an external interferer.

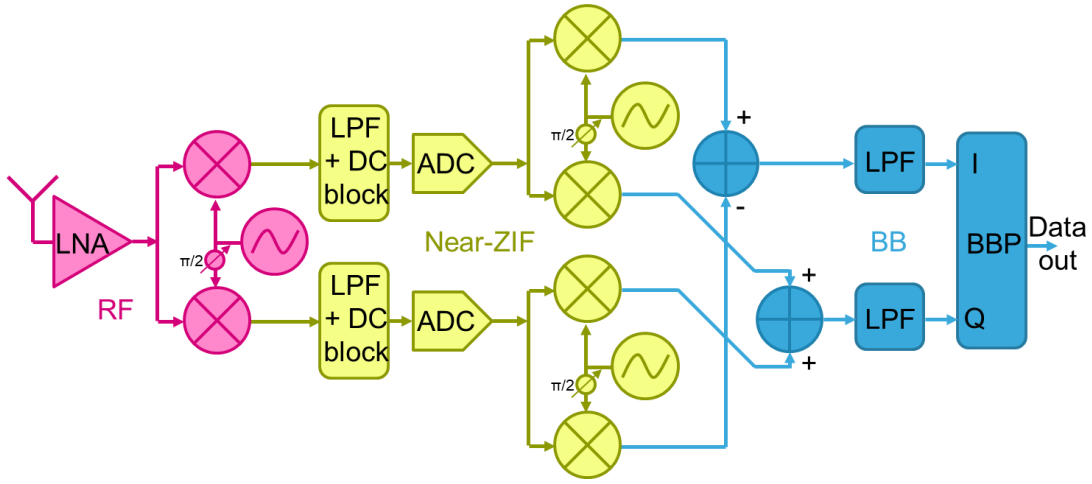

Figure 3-15: Near-ZIF receiver using the Weaver image cancelling approach with a digital implementation

If it is noise, 30dB or image rejection ratio would be enough since the image frequency signals of the different single receivers would be uncorrelated, and their power would evolve in the same way as the in-band noise during beamforming.

If it is an external interferer, then again there are two cases, if the interferer comes from the same direction as the desired signal or not. If it is coming from a different direction, then its power must remain low enough to avoid saturating the receiver. Then it will undergo the image rejection attenuation and the spatial filtering from beamforming. This is likely to be enough so its contribution to SINR degradation can be ignored. If it is coming from the same direction as the desired signal, then it must be low enough such that it does not degrade significantly the received SINR. Using the PSL hypothesis of -20dB from section 2.2.3, the image signal power would require to be at least that much below the acceptable level for the case when it comes from a different direction.

Regardless of the image cancelling approach, the receiver's analog part, from the antenna to the ADCs outputs, is the same in the case of a digital implementation and is essentially the same as a homodyne receiver. The main difference lays with the requirements on the ADC. In the homodyne case it needs a bandwidth half that of the signal but with a higher dynamic range to handle the DC from self-mixing, while in the Near-ZIF case it needs a bandwidth equal to that of the signal but with a lower dynamic range.

It is not straight forward to know beforehand which approach will be the most power efficient, but the ADC survey from [3-36] helps getting some understanding on the matter. It provides multiple figures of merits over all the ADCs published for more than two decades at ISSCC and VLSI. Here the focus will be on the two major ones, the Walden and the Schreier Figure of Merits (FoM). Their formulas are given in equations (3.27) and (3.28) respectively.

$$FoM_W = \frac{P}{f_S \times 2^{ENOB}} \tag{3.27}$$

$$FoM_S = SNR + 10 \times \log_{10}\left(\frac{f_S/2}{P}\right) \tag{3.28}$$

With P the ADC power consumption, f_S its sampling frequency, SNR its Signal to Noise Ratio and $ENOB = (SNR - 1.76)/6.02$ its Effective Number Of Bits.

On one hand, these two FoM describe the same tradeoff between power and sampling speed. In both cases, doubling the clock speed doubles the power. On the other hand, they describe a different tradeoff between power and resolution. From the Walden FoM to preserve a constant value, adding one bit requires to double the power. From the Schreier FoM, adding on bit requires four times the power. Of course, both FoM cannot be right at the same time.

The Walden FoM is more meaningful when the power is dominated by the charging and discharging of parasitic capacitances. In that case it is said to be process limited since a finer lithography would lead to smaller capacitances and hence lower power. In general, this FoM is applicable for low resolution ADCs.

The Schreier FoM is more meaningful when the SNR is limited by thermal noise. In that case one more bit of resolution means an SNR $6.02dB$ higher, which means a signal power four times higher for the same noise power. In general, this FoM is applicable for high resolution ADCs.

As processes shrink down the limit between low and high resolution goes down. Currently it is assumed to be around 50dB of SNR. If the specification for the ADC resolution falls below that value, then the breakeven point between the homodyne and Near-ZIF receiver will be if the additional SNR required to handle the DC from self-mixing is roughly 6dB. If it is more that, then a Near-ZIF is likely to be more efficient. If the ADC specification is above 50dB then the breakeven point falls at 3dB.

To complete this analysis the sampling jitter must be considered since both solutions have different sampling frequencies. Figure 3-16 and Figure 3-17 are the plots of the Walden and the Schreier FoM extracted from [3-36]:

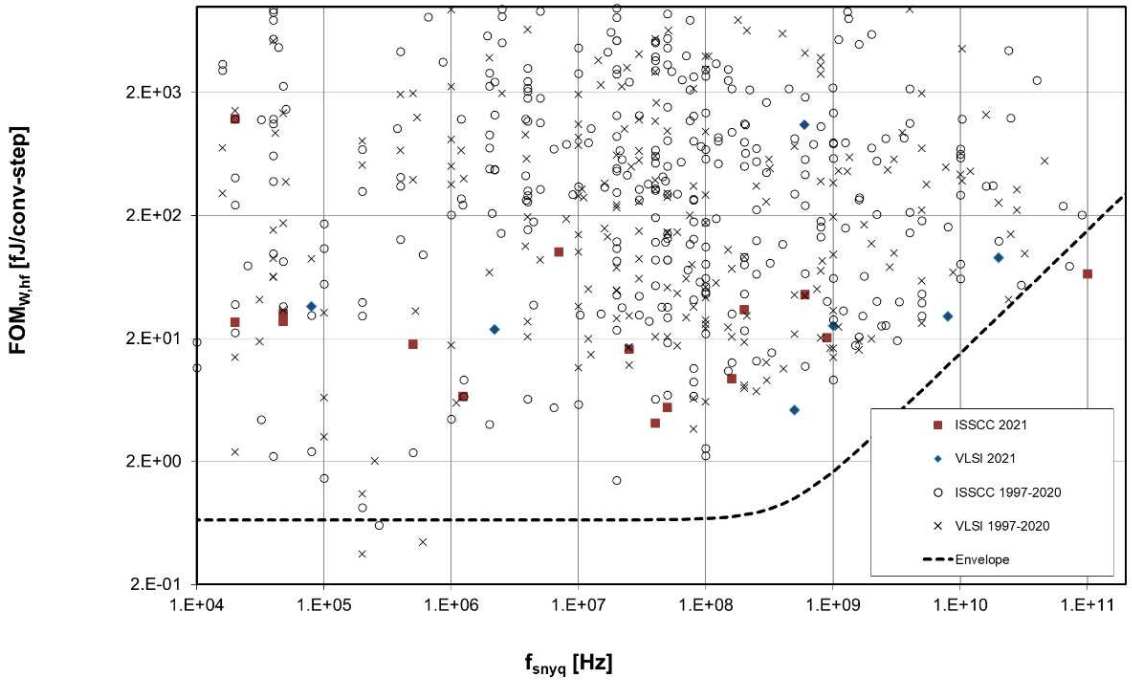

Figure 3-16: Walden Figure of Merit versus sampling frequency

In both cases the envelope is flat at low frequency and degrades at high frequency. This degradation is caused by the sampling jitter. For the Walden FoM, the corner frequency happens a little below 1GHz. This is of the same order of frequency as required in the present application, meaning some additional but limited SNR degradation are to be expected. For the Schreier FoM, the corner frequency happens almost a decade earlier and a greater SNR degradation can be expected.

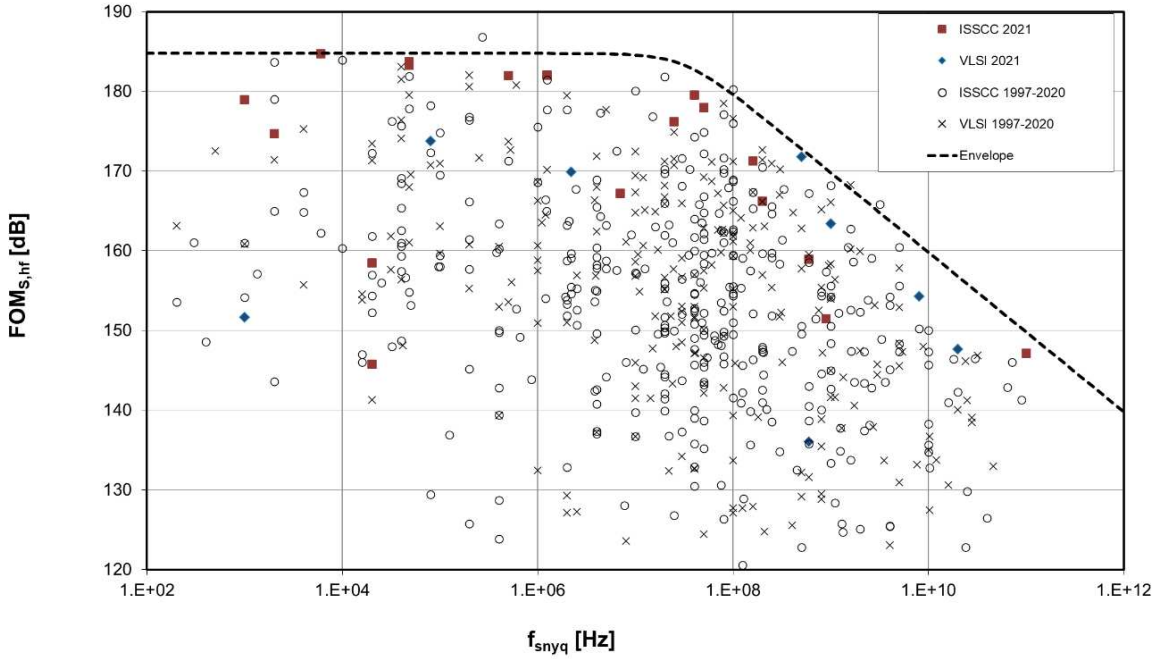

Figure 3-17: Schreier Figure of Merit versus sampling frequency

While clear conclusions cannot be reached from this discussion alone, it gives a better understanding of the cases where a homodyne or a Near-ZIF receiver would be more appropriate. For the following specification a Near-ZIF architecture will be assumed since it requires to do one less hypothesis on the DC level from self-mixing.

Since the focus is on DBF receivers, it is reasonable to suppose that the implementation will be made using a digital process such as deep sub-micron CMOS. While in general such technologies provide lower analog performances, they can deal with relatively high bandwidth thanks to the low parasitic capacitance of the deep sub-micron etching. Obviously, they also provide good digital processing performances. The implementation proposed in the second part of this manuscript is done in STMicroelectronics 28nm Fully Depleted Silicon On Insulator (FDSOI) technology. Whenever technological parameters are needed, this process will be used.

The primary interest is to specify the analog part of the receiver. The building block analysis will be limited to the LNA, the mixer, the LPF and the ADC. The LO has already been specified as best as possible. Also, only one need to be implemented and then be shared by all the SRxs, and so will be its power consumption. Its distribution to each SRx on the other hand should be considered for the power consumption evaluation. All the other building blocks will be specified in terms of center frequency, bandwidth, gain, NF and linearity. Finally, the ADC sampling rate, DR and resolution will be specified.

3.2.2 Center frequency and bandwidth

In terms of center frequencies, three of them need to be specified, the RF, the LO and the IF. The RF, as previously mentioned is one of the working hypotheses, is $f_c = 28GHz$. This will be the center frequency of the LNA and the mixer RF input.

This architecture is based on a low IF. The minimum possible IF allowing for the whole IF signal band to be in the positive frequencies is half of that band $f_{IF} = B/2 = 500MHz$. Because it is desirable to filter the DC signal from the mixer self-mixing, it is necessary to shift up a little bit this frequency.

Here, it will be shifted by a sub-channel width of $B_{ch} = 100MHz$. This gives a $600MHz$ IF frequency (f_{IF}).

For the mixer to translate the RF signal at this IF, the LO frequency can take two values $f_{LO} = f_c \pm f_{IF}$. Since f_{IF} is small, both of these frequencies are close to each other and there is no significant advantage in choosing one or the other. The lower one is arbitrarily chosen, that is $f_{LO} = f_c - f_{IF} = 27.4GHz$.

The LNA and mixer bandwidth are both equal to $1GHz$ (ten sub-channels). In the IF domain, the filter and the ADC need to accommodate an additional $100MHz$ of band (due to the DC shift). This makes the total bandwidth equal to $1.1GHz$.

3.2.3 Gain specification

It is interesting for a receiver to provide some gain, mostly for two reasons. First, to reduce the noise contribution of the subsequent links in the chain. This will be discussed in the next section. Second, to deliver the signal with a proper amplitude to the ADC to minimize its requirements. It is the second one that will be used to specify the SRx gain. Then, it will be allocated to the different building blocks.

As an initial step, it is necessary to evaluate the maximum peak-to-peak amplitude $S_{in_{pp}}$ of the input signal. As previously, the input signal power can be approximated to be the OoBI power since it is several orders of magnitude above any other signal. To evaluate S_{pp}, the Peak to Average Power Ratio (PAPR) of the signal can be used.

The PAPR of a modulated signal depends on the modulation type. OFDM is notoriously known to produce high PAPR such that TX BBP use crest reduction algorithms to relax the PA linearity constraint. From [3-3], the probability of a native OFDM PAPR to go above 15dB is below 10^{-6} and can be considered as negligible. Also, PAPR reduction techniques generally improve PAPR by about 3dB. Hence it will be assumed that the OoBI PAPR is 12dB. The peak amplitude at the LNA input, assuming an input impedance of $R_{in} = 50\Omega$, is:

$$S_{in_{pp}} = 2 \times \sqrt{R_{in} \times 0.001 \times 10^{\frac{P_{in_{dBm}} + PAPR}{10}}} \approx 6mV \tag{3.29}$$

In STMicroelectronics 28FDSOI technology, the nominal power supply is $1V$. Therefore, a good compromise between noise and linearity is for the signal to evolve in a voltage range of about $300mV$. Assuming a differential structure, the maximum ADC peak-to-peak input can be up to $600mV$. The SRx gain must then be around a hundred or $40dB$. This gain now needs to be split between the LNA, the mixer and the filter. A typical gain for an LNA is $20dB$. The choice is made to split equally the remaining $20dB$ between the mixer and the filter. It must be kept in mind that the proposed gain distribution is only an educated guess and that its specification should remain somewhat flexible when reviewing the literature.

3.2.4 NF specification

An NF target specification of $10dB$ was established and an acceptable contribution for each of the building blocks now needs to be allocate.

One important point, which has not been discussed yet, is the Insertion Loss (IL) of the connection between the antenna and the LNA. This path is, at the minimum, made of an RF filter and some PCB routing, from the antenna up to the package input pin, the package routing and finally the on-chip routing and matching network to the LNA input. In most cases, the antenna will also be used by a transmitter, adding a switch on the path. The author in [2-31] evaluate the overall IL to $3.9dB$. Since the present analysis so far did not considered this loss, it can be assumed to be as straight contribution to NF.

In the same study, an NF contribution evaluation of the other building blocks, is also provided. The LNA contribution is $2.7dB$, and the subsequent block contributions are $0.3dB$. This leads to a total $NF = 7.5dB$. The proposition here is to allocate the $2.5dB$ margin by raising the acceptable IL and LNA contributions to $4.5dB$ each, and that of the mixer to $0.4dB$. The reason for this allocation to the early building blocks is because, in general, the dominant contributions come from the elements before the first gain element, the LNA in the present case. Once the signal is amplified, the same noise power will contribute to a much smaller SNR degradation. The total NF of two consecutive gain stages is expressed in (3.30):

$$NF = 10 \times \log_{10}\left(10^{\frac{NF_1}{10}} + 10^{\frac{NF_2 - G_1}{10}}\right) \tag{3.30}$$

The NF contribution of the second stage is reduced by the gain G_1 of the first one. Applying this formula recursively for each stage together with the gain specification from the previous section leads to the NF specification of Table 3-1:

Table 3-1: NF specifications of the SRx building blocks

	Gain	NF contribution	Individual NF
FE	-4.5dB	4.5dB	4.5dB
LNA	20dB	4.5dB	4.5dB
Mixer	10dB	0.4dB	14.5dB
Filter	10dB	0.3dB	23dB
ADC	0dB	0.3dB	33dB

One can see that, despite the NF contribution of the mixer, filter and ADC being very small, their actual NF specifications are much more relax thanks to gain of the previous stages. This justifies, on one hand, the specification of a large gain for the LNA and on the other hand, the attribution of NF contribution mostly to the IL and the LNA.

3.2.5 Linearity specification

The same simulation-based approach will be used for the building blocks as the one that was used for the SRx specification. The total acceptable parasitic power in the fifth channel was specified from non-linearity. As the different stages may provide different gains, this power level needs to be scaled accordingly. To avoid gain dependent metrics, the fifth channel power is measured relative to the OoBI. This measure becomes very similar to an Adjacent Channel Leakage Ratio (ACLR). Because ACLR is used to characterize non-linearity in the presence of wide band signals, the proposed measure will be called Modified ACLR (MACLR). This specification for the whole receiver is then given by:

$$MACLR_{SRx} = P_{d_{dBm}} - 10 \times \log_{10}(2) - P_I = -106 - (-50) = -56dB \tag{3.31}$$

There are four stages that may contribute to non-linearity. The choice was made that each of them may contribute equally to the overall non-linearity. Assuming the worst case, i.e. non-linearity products are perfectly correlated, the contribution of the two first stages will be $6dB$ higher than the contribution of a single stage. Same goes for the two last stages. The total contribution of the four stages is then another 6dB higher for a total intermodulation power increase of 12dB. The individual MACLR specification of the building blocks must then be 12dB below that of the SRx to achieve the overall specification. This gives a building block $MACLR = -68dBm$. By simulation, the IIP3 corresponding to that specification is monitored. The results are plotted on Figure 3-18.

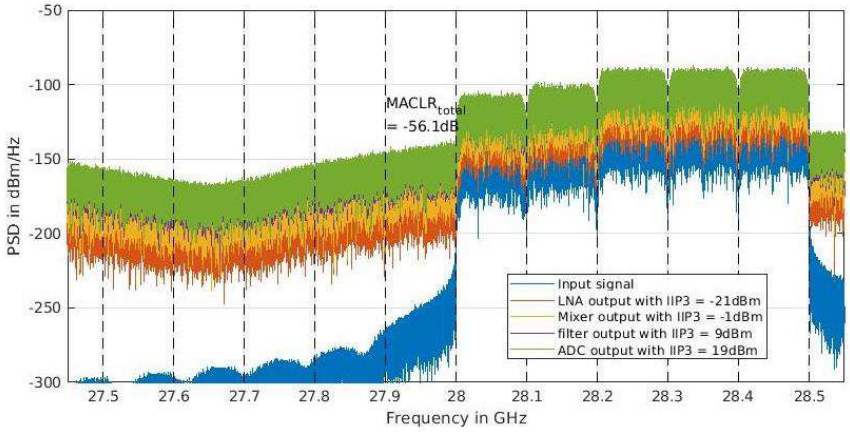

Figure 3-18: Output spectrum of the different building blocks for IIP3 compatible with MACLR specifications

The different specifications of the building blocks are summarized in Table 3-2.

Table 3-2: Linearity specifications of the SRx building blocks

	Gain	MACLR	IIP_3
LNA	20dB	-68dB	-21dBm
Mixer	10dB	-68dB	-1dBm
Filter	10dB	-68dB	9dBm
ADC	0dB	-68dB	19dBm
SRx	40dB	-56dB	-25dBm

ADC non-linearity is rarely specified in terms of IIP3. More often the Total Harmonic Distortion (THD) or the Spurious Free Dynamic Range (SFDR) are used. Both metrics are evaluated using a single tone input at near full-scale amplitude, generally $-1dB_{FS}$. The THD is the ratio of the Harmonics total power to the fundamental, expressed in decibel. The SFDR is the power ratio between the fundamental and the strongest spurious tone expressed in decibel. This spurious tone may not be a harmonic.

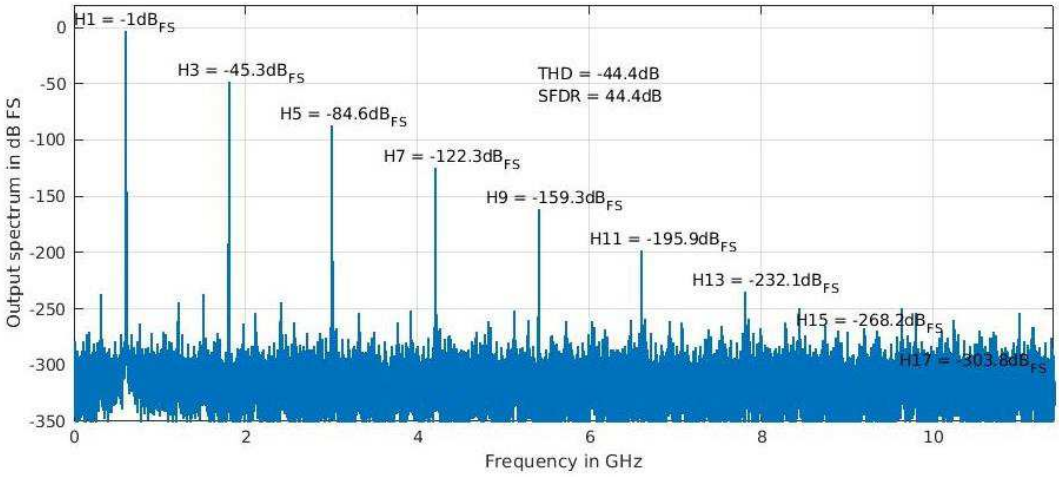

Figure 3-19: ADC output spectrum for a single tone input at $-1dB_{FS}$

For weakly non-linear systems, when SFDR is limited by the third harmonic, THD and SFDR have similar values with a sign difference. Figure 3-19 plots the ADC output for a single tone input using an

arctangent to model the non-linearity. While this model may not be very realistic, it still gives a good order of magnitude for the desired THD or SFDR.

As for NF, the optimal budget splitting is generally not an equal splitting. It is generally harder to have a good linearity when dealing with signals of greater amplitude. It would therefor make sense to allocate more budget later in the chain once the signal gets amplified. The difference between an optimized splitting and an equal one is not as radical for linearity as it is for NF. It will also be seen later that the specifications proposed in Table 3-2 are well within reach of the current state of the art. For this reason, this simple equal splitting of the linearity budget will be kept unchanged.

3.2.6 Image rejection and anti-aliasing filter

The purpose of this filter is double. First, it must remove the high frequency image from the mixer output. Second, it must remove, as much as possible, any parasitic signals from the spectrum outside of the ADC Nyquist zone of interest, regardless of the nature of the parasitic signals, noise or interferer. While some evaluation of the filtering required for image rejection can be made, it is not as easy for anti-aliasing. The reason being that no information is available about the potential parasitic signals outside the ADC bandwidth (The OoBI defined earlier sits within the ADC bandwidth and does not provide useful information on this matter). The proposed specification will therefore be based on the optimistic case where only image rejection and wide band thermal noise filtering are considered and may need revision in the future.

3.2.6.1 Image rejection

The high frequency image will introduce perturbations in the band when aliased by the ADC sampling operation. These perturbations are likely to be correlated between all the SRx, such that this specification does not benefit from the array factor. It is desirable for this aliased power to have negligible impact and be at least 20dB below the noise level of the input signal. For the sake of simplicity, all powers are Rx input referred. The noise power in one sub-channel in the presence of the OoBI, i.e. under the 6dB desensitization, is given by (3.32):

$$N_D = N_{th_{dBm}} + NR_{SRx} + D - 10\log_{10}(N_{ant}) = -102dBm \tag{3.32}$$

The acceptable remaining image power must be below $-122dBm$. The power in a 100MHz interfering channel is $-54.7dBm$. The minimum required attenuation is therefore $67.3dB$.

Let us now locate this high frequency image in the spectrum. With the input signal center frequency at $f_c = 28GHz$ and the LO frequency at $f_{LO} = 27.4GHz$, the high frequency image will sit at $f_{img} = f_c + f_{LO} = 55.4GHz$. The desired low frequency image is at $f_{IF} = 600MHz$ and the signal bandwidth is $B = 1GHz$. The filter cutoff frequency can be at best at $f_{flt} = f_{IF} + B/2 = 1.1GHz$. The high frequency image signal is about 1.7 decades away from the cutoff frequency. A first order filter would provide a 20dB per decade slope, giving about $34dB$ attenuation. This is clearly not enough. A second order filter would provide $68dB$ attenuation. This would be just enough. For a simple Butterworth filter, its adequate order would be three. It is preferred to have some margin. This allows for a slightly higher cutoff frequency, improving the in-band flatness response at the band upper edge.

3.2.6.2 Anti-Aliasing

The purpose of the anti-aliasing filter is to remove any signal outside the Nyquist zone of interest. As mentioned earlier, the only out of band signal that will be considered here is the receiver's thermal noise. The characteristics of the anti-aliasing filter are related to the ADC sampling frequency. It is usual to have a sampling frequency slightly higher than the minimum required by the signal bandwidth, typically by a factor ~1.5. This minimum sampling frequency, in the present case, would be twice the highest frequency of the input signal. It is $1.1GHz$ which gives a minimum sampling rate of $2.2Gsps$. A realistic sampling rate for such a system would be $3Gsps$ leading to a Nyquist frequency of $1.5GHz$.

Ideally the filter's cutoff frequency must be as low as possible, in order to remove as much noise as possible. It was shown in the previous section that it is also interesting for this cutoff frequency to be slightly higher than the input signal highest frequency of interest for improved in band flatness. A tradeoff between these two constraints needs to be done. One interesting question is: How high the cut off frequency can be without compromising the anti-aliasing characteristic of the filter? To answer this question, the notion of filter equivalent noise bandwidth will be used.

For classical filters such as the Butterworth filter, when a white noise is injected at its input, the total output power is bounded. The equivalent noise bandwidth is defined as the bandwidth a brick wall filter should have, such that, for the same input white noise, it produces the same total output power. This is assuming the filter has no noise contribution to the output.

A good practice for the anti-aliasing filter is for its equivalent noise bandwidth to be below the Nyquist frequency. The authors from [3-4] provide this value as a ratio compared to the 3dB cutoff frequency for multiple filters and orders. This ratio, for a third order Butterworth filter, is 1.047. This means that the filter cutoff frequency must remain below $f_{Nyquist}/1.047 \cong 1.43GHz$. It has also been seen previously that it must be above $1.1GHz$. Here, the anti-aliasing filter cutoff frequency specification is set in between at $1.3GHz$, keeping some margin on both sides.

3.2.6.3 Summary

The proposed filter is a third order Butterworth filter with a $1.3GHz$ cutoff frequency. The spacing with the high frequency image is then 1.63 decades and the attenuation is $97.7dB$, well above the $67.3dB$ required. Its noise equivalent bandwidth is $1.36GHz$. This is below the defined Nyquist frequency of $1.5GHz$. With no additional information about potential interfering signals outside the channel, it can be concluded that this filter is satisfactory for both image rejection and anti-aliasing.

3.2.7 Analog to Digital Converter specifications

During the specification of the previous building blocks, the ADC was already partially specified. In particular, it was established that the input full scale must be $600mV$, the $THD = -44.4dB$, and the sampling frequency $3Gsps$. The last specifications to be done are the SNR and SNDR. To do that, the allocated ADC NF contribution of $0.3dB$ and the THD will be used.

Starting with SNR, equation (3.33) evaluates the noise power the ADC may add over the band $B = 1GHz$ to limit the NF degradation to $0.3dB$.

$$N_{ADC_{dBm}} = N_{th_{dBm}} + NF + G_{SRx} + 10 \times \log_{10}\left(1 - 10^{\frac{-NF_{ADC_{cont}}}{10}}\right) = -45.7dBm \qquad (3.33)$$

Assuming flat noise, this gives a maximum noise PSD of $-135.7dBm/Hz$. The total ADC noise power, up to the specified 1.5GHz Nyquist frequency, is $-44dBm$.

It is now necessary to relate this value to a full-scale tone. Assuming the SRx input impedance is 50Ω and knowing that the gain defined earlier are voltage gain, this noise power is $-57dB_{Vrms}$. A full-scale sinewave will have a $0.3V$ amplitude and an RMS value of $-13.5dB_{Vrms}$. The required ADC SNR is the difference between the two. This gives an $SNR_{ADC} = 43.5dB$.

The THD has a value similar to the SNR, giving a good balance of the constraints split between noise and linearity. The SNDR is evaluated by summing the ADC's total output noise and its total harmonic distortion power. This gives a $41dB$ $SNDR$.

This result benefits slightly from the chosen sampling rate of 3Gsps, 1.5 times higher than the minimum 2Gsps Nyquist rate required for a 1GHz signal bandwidth. This is called the Over Sampling Ratio (OSR). In the present case, it spreads the ADC total noise onto a wider frequency range than the signal

bandwidth. This reduces the amount of in band noise, improving the signal SNR, for the same total ADC output noise power. Without this OSR, the required $SNDR$ would only be $1dB$ higher. One can see here that this over sampling does not bring much advantage on the final required $SNDR$. Its main purpose is to relax the anti-aliasing filter.

3.2.8 State of the art on building blocks

The specifications derived above are summarized in Table 3-3:

Table 3-3: Building Blocks specifications summary

	BW	Gain	NF	MACLR	IIP3
FE	1GHz	-4.5dB	4.5dB	N/A	N/A
LNA	1GHz	20dB	4.5dB	-68dBc	-21dBm
Mixer	1GHz	10dB	14.5dB	-68dBc	-1dBm
Filter	1GHz	10dB	23dB	-68dBc	9dBm
ADC	1.5GHz	0dB	33dB	-68dBc	19dBm
RX	1GHz	40dB	10dB	-56dBc	-25dBm

Additionally, a $-136dBc/Hz$ LO flat phase noise was specified, a third order Butterworth filter with a $1.3GHz$ cutoff frequency and an ADC with a $3Gsps$ sampling rate, an $SNR = 43.5dB$, a $THD = -44.5dB$, and an $SNDR = 41dB$. The next step is to establish the state of the art for the LO, the LNA, the mixer, the filter and the ADC with two goals. The first one is to evaluate the feasibility of such a receiver. The second one is to make a power consumption evaluation.

3.2.8.1 Local Oscillator

The specification of a wide band phase noise PSD below $-136dBc/Hz$, typically beyond 300MHz offset, is somewhat uncommon and generally not reported in the literature. It can even be questioned if proper attention has been given to that matter in most designs. Often, the phase noise measurements do not even look at such high frequency offsets. Thankfully, it is not the case for all publications. The numbers for wide band phase noise PSD provided below in Table 3-4 are all graphically estimated.

Presumably for practical measurement reasons, the reported phase noise spectrums are often those of a frequency divided signal. It is then difficult to really know how this affects the measured performances or if the reported spectrums are compensated for it. Nonetheless, the values from Table 3-4 satisfy the specification. One can note that [*3-35*] reports the best performances and is implemented in advance digital node. Finally, it is worth mentioning that the proposed specification is based on a single LO for the 256 receivers. In practice, it is unlikely the case and this specification can most probably be relaxed. Also, the power consumption will be divided between multiple receivers and will unlikely be significant and will be ignore for now. Consequently, there is no limitation coming from the LO generation performances.

Table 3-4: PLL state of the art performance summary

Ref	Frequency	Wide band PN PSD	Power	Techno
[*3-32*] 2016	25.2GHz-30.4GHz	-147dBc/Hz*	87mW @???	CMOS 65nm
[*3-33*] 2017	26.2GHz-32.4GHz	-136dBc/Hz**	26.9mW @1V	CMOS 65nm
[*3-34*] 2017	27.4GHz-30.8GHz	-140dBc/Hz***	24.3mW @???	CMOS 65nm
[*3-35*] 2018	23.3GHz-30.2GHz	-147dBc/Hz****	31mW @1.2V	CMOS FDSOI 28nm

*Measured at the 3.5GHz divided output. **On the output divided by 4. ***On the output divided by 3. ****At 300MHz offset

3.2.8.2 LNA

Table 3-5 below summarizes the performances of recently published LNA relevant for this use case.

Table 3-5: LNA state of the art performance summary

Ref	BW	Gain	NF	IIP3	Power	Techno
[*3-7*] 2019	12GHz @ 28GHz	18.2dB	4.1dB	-5.4dBm*	9.8mW @1V	CMOS 65nm
[*3-8*] 2018	9.3GHz @ 27.8GHz	18.4dB	3.4dB-4.4dB	-4.9dBm	21.5mW @ 1.1V	CMOS 40nm
[*3-9*] 2018	From 24GHz to 33GHz	24dB	4dB	-13.4dBm*	18.5mW @1.1V	CMOS SOI 45nm
[*3-10*] 2019	From 24GHz to 43GHz	23dB	3.7dB	-15dBm	20.5mW @1V & 1.6V	CMOS FDSOI 22nm
[*3-11*] 2018	4.4GHz @ 33GHz	24.5dB	4dB	-15.9dBm*	27.5mW @2V	CMOS 28nm

*Estimated from IP1dB by IIP3 = IP1dB + 9.6dB

With the 5G approach, several 28GHz LNA have been published in the last couple of years. It can be seen from Table 3-5 that none of the derived specifications may be a blocking point. All the LNA were build using a CMOS process, but some, like 45nm CMOS Silicon On Insulator (SOI), are more RF oriented tahn others, like 22nm FDSOI which is a more digital technology. The power consumption ranges from 9.8mW to 27.5mW. For power estimation, the average value of 20mW will be used.

3.2.8.3 Mixer

Table 3-6 summarizes the performances of recently published mixers relevant for this use case. There are not so many publications on $28GHz$ mixers. From Table 3-6, it seems that there is no showstopper in terms of performances. However, none of the mixers reported here use a deep sub-micron digital technology. Reaching the desired performances in such a process for a fully integrated receiver may be challenging. Passive mixers generally have lower power consumption but lower gain, below 0dB. In the present case, this lower gain will have to be compensated somewhere else and will add power consumption. For this reason, the power estimation on an active architecture with a 20mW power consumption will be used.

Table 3-6: Mixer state of the art performance summary

Ref	RF BW @ IF	CG	NF	IIP3	LO power	Power	Techno
[*3-5*] 2013	20GHz-26GHz @300MHz	9.15dB	3.61dB	-12.6dBm**	0dBm	2.6mW (mixer) + 20.6mW (IF buffer) @1.5V	CMOS 0.13µm
[*3-6*] 2015	24.5GHz-36.5GHz @12GHz	6.4dB	6dB	15dBm*	0dBm	21.2mW @1.5V	CMOS 45nm SOI
[*3-12*] 2019	23GHz-25GHz @100MHz	26.1dB	7.7dB	-8.2dBm**	-3dBm	16.8mW @1.5V	CMOS 0.13µm
[*3-13*] 2018	23GHz-30GHz @???	-3dB	???	21dBm**	1dBm	10mW @ 1.2V	CMOS 90nm

*IIP3 achieved using nonlinearity cancellation by tweaking a back gate by hand. IIP3 = 2dBm when no tweaking applied. **Estimated from OP1dB by IIP3 = OP1dB – CG + 9.6dB or IP1dB + 9.6dB

3.2.8.4 Filter

Table 3-7 summarizes the performances of published filters relevant for this use case.

Table 3-7: Filter state of the art performance summary

Ref	Cut off Freq	Gain	NF	IIP3	Order	Power	Techno
[*3-14*] 2006	250MHz or 1GHz	0dB or 14dB @1GHz BW	N/A	N/A	3	3.2mW @1.8V	CMOS 0.18µm
[*3-15*] 2009	240MHz	From 0dB to 40dB	16.6dB @40dB Gain	-35.2dBm* @ 40dB Gain	6	2.9mW @1.2V	CMOS 90nm
[*3-16*] 2010	250MHz	From -9dB to 73dB	14dB @73dB Gain	-71dBV in-band -6dBV Out of band	6	56.4mW @1.2V	CMOS 0.13µm
[*3-17*] 2012	255MHz	From -2.57dB to 39.02dB	22.7dB	14dBm In-band @ 0dB Gain	6	2.3mW @1.2V	CMOS 90nm

*Estimated from OP1dB by IIP3 = OP1dB – Gain + 9.6dB or IP1dB + 9.6dB

There are very few publications with performances corresponding to these needs. Except for [3-14] the cutoff frequency is significantly lower than the specification. This is because these published works were targeting a UWB scenario with an RF bandwidth of ~500MHz and a BB bandwidth of ~250MHz for the I and Q signals. This means there is no technical impossibility.

One can also observe that all of them are third order or higher and can often provide much more gain than required. Linearity may prove to be more challenging but not out of reach. CMOS Process are used but not modern digital ones. This is because these publications are relatively old and such technologies simply did not exist or were not easily accessible at the time.

Since no specification for out of band blockers is available, it is rather hazardous to make any conclusion on the filter power consumption. A budget of 10mW of power consumption will be allocated, but this number needs to be taken lightly.

3.2.8.5 ADC

Here, the literature is rich in devices in the vicinity of the desired requirements (Table 3-8). Few things are interesting. First, in average the technologies used are advance digital ones. This shows the benefit of smaller technologies for high-speed ADCs. In particular, it facilitates heavily digitally assisted architectures and allow co-integration with digital function such as demodulation.

Table 3-8: ADC state of the art performance summary

Ref	Fs Nyquist	Full scale	SNDR	Power	Techno
[3-18] 2012	3Gsps	500mVpp diff	36.2dB	11mW @1.1V	CMOS 40nm
[3-19] 2013	5Gsps	300mVpp diff	30.9dB	8.5mW @0.85V	CMOS 32nm SOI
[3-20] 2015	5Gsps	1Vpp Single	30.25dB	5.5mW @1V	CMOS 65nm
[3-21] 2017	2.4Gsps	0.9Vpp diff	40.05dB	5mW @0.9V	CMOS 28nm

[3-22] 2017	1.5Gsps	650mVpp diff	50.1dB	6.92mW @0.95V	CMOS FF 14nm
[3-23] 2019	5Gsps	???	48.5dB	29mW @1/0.85V	CMOS 28nm
[3-24] 2019	6Gsps	960mVpp diff	39.9dB	41.1mW @0.9V/0.72V	CMOS FF 16nm
[3-25] 2019	3.6Gsps	500mVpp diff	31.8dB	2.6mW @1V	CMOS 28nm
[3-26] 2019	1.6Gsps	???	54.2dB	12.2+7.6*mW @0.9V/0.8V	CMOS 28nm
[3-27] 2019	1Gsps	???	46.65dB	2.1mW @1.1V	CMOS FDSOI 28nm
[3-28] 2019	1Gsps	1.2Vpp diff	60.02dB	7.6mW @1V	CMOS 28nm
[3-29] 2019	2.4Gsps	???	49.02dB	9.8mW @0.9V	CMOS FDSOI 28nm
[3-30] 2019	10Gsps	800mVpp diff	36.9dB	21mW @1V	CMOS 28nm
[3-31] 2019	4Gsps	???	39.9dBC	11.7mW @1V	CMOS 28nm

* Digital calibration power consumption estimated based on gate count

While none of the above ADC has the exact required performances, it is reasonable to assume that this is achievable within a power budget of 10mW. For example, the ADC from [3-21] only falls short on sampling rate, compared to the specification, and consumes only 5mW. Using a two-time interleaved architecture could double its sampling rate, achieving 4.8Gsps. This would consume slightly over twice the original power, around 10mW. In this case, only a 3Gsps sampling rate needs to be reached. This is certainly achievable within a 10mW power budget.

3.2.9 Conclusion on feasibility

To conclude on DBF feasibility let us first summarize the state-of-the-art evaluation. It was shown that none of the required specifications are out of reach, even in CMOS technologies. The mixer and the filter have no implementation example in very advance digital node, but it is hard to conclude if this is for fundamental reasons or just because of the low number of publications in these areas.

From a power consumption standpoint, the SRx can be evaluated to $100mW$. Let us put this number into perspective. The idea behind small cell deployment is to have S-BS on light polls. It would be a good opportunity to merge this deployment with the replacement of streetlights with Light Emitting Diodes (LED) based lamps. This would free up some power for the S-BS. While the available power on light polls may vary from place to place, it is reasonable to assume that it is around a couple of hundreds of Watts since incandescent light bulb can easily reach this level of power. If an S-BS with 4 sectors and 256 elements per sectors is assumed, the SRx total power consumption sums up to about 100W. That leaves a significant amount of power available for all the digital processing that needs to be done. Since this is a TDD system, the Rx consumption should be, at least partially, available for the Tx as well.

Within the digital processing, the power consumption of the digital portion of DBF must be considered as it would be more significant than for HBF. This part is much harder to evaluate for many reasons but is actually unlikely to be a real limitation. As it was seen in section 2.2.5.5, the biggest challenge in terms of S-BS processing is the CSI acquisition. This portion of the digital processing must be done for both DBF and HBF. Hence, if this is a showstopper, it is not a limitation caused by DBF, but by the very nature of beamforming itself.

Finally, this demonstrates the feasibility of digital beamforming both from a performance and a power consumption point of view.

3.3 CONCLUSION

Starting from the conclusions of the previous chapter, an in-depth analysis of a digital beamforming receiver was carried out. The required performances for the whole receiver as well as for one dedicated to a single antenna of the S-BS array were evaluated. A conclusion on the close in phase noise required on the local oscillator could not be reached. Except from that a fairly relevant set of specifications was established.

Based on a super heterodyne receiver architecture with a low intermediate frequency, the building block specifications were derived. A full specification of the image rejection and anti-aliasing filter could not be established since the nature of the potential interfering signals is unknown. The remaining building blocks were properly specified with success.

Finally, through a pretty extensive state of the art survey, the feasibility of the proposed digital beamforming receiver was established, and its power consumption evaluated. This power consumption, when put into perspective of the expected available power for an S-BS, is perfectly acceptable. This final remark makes the case for the proposed approach.

3.4 REFERENCE

[3-1] N.B.de Carvalho, Compact formulas to relate ACPR and NPR to two tone IMR and IP3, Microwave Journal december 1999

[3-2] A. Hajimiri and T. H. Lee, "A general theory of phase noise in electrical oscillators," in IEEE Journal of Solid-State Circuits, vol. 33, no. 2, pp. 179-194, Feb. 1998.

[3-3] T. Jiang and Y. Wu, "An Overview: Peak-to-Average Power Ratio Reduction Techniques for OFDM Signals," in IEEE Transactions on Broadcasting, vol. 54, no. 2, pp. 257-268, June 2008.

[3-4] W. Stanley and S. Peterson, "Equivalent Statistical Bandwidths of Conventional Low-Pass Filters," in IEEE Transactions on Communications, vol. 27, no. 10, pp. 1633-1634, October 1979.

[3-5] S. Kong, C. Y. Kim and S. Hong, "A K-Band UWB Low-Noise CMOS Mixer With Bleeding Path Gm-Boosting Technique," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 60, no. 3, pp. 117-121, March 2013.

[3-6] C. L. Wu, C. Yu and K. K. O, "Amplification of Nonlinearity in Multiple Gate Transistor Millimeter Wave Mixer for Improvement of Linearity and Noise Figure," in IEEE Microwave and Wireless Components Letters, vol. 25, no. 5, pp. 310-312, May 2015.

[3-7] S. N. Ali, M. Aminul Hoque, S. Gopal, M. Chahardori, M. A. Mokri and D. Heo, "A Continually-Stepped Variable-Gain LNA in 65-nm CMOS Enabled by a Tunable-Transformer for mm-Wave 5G Communications," 2019 IEEE MTT-S International Microwave Symposium (IMS), Boston, MA, USA, 2019, pp. 926-929.

[3-8] M. Elkholy, S. Shakib, J. Dunworth, V. Aparin and K. Entesari, "A Wideband Variable Gain LNA With High OIP3 for 5G Using 40-nm Bulk CMOS," in IEEE Microwave and Wireless Components Letters, vol. 28, no. 1, pp. 64-66, Jan. 2018.

[3-9] V. Chauhan and B. Floyd, "A 24–44 GHz UWB LNA for 5G Cellular Frequency Bands," 2018 11th Global Symposium on Millimeter Waves (GSMM), Boulder, CO, USA, 2018, pp. 1-3.

[3-10] L. Gao and G. M. Rebeiz, "A 24-43 GHz LNA with 3.1-3.7 dB Noise Figure and Embedded 3-Pole Elliptic High-Pass Response for 5G Applications in 22 nm FDSOI," 2019 IEEE Radio Frequency Integrated Circuits Symposium (RFIC), Boston, MA, USA, 2019, pp. 239-242.

[3-11] M. Keshavarz Hedayati, A. Abdipour, R. Sarraf Shirazi, C. Cetintepe and R. B. Staszewski, "A 33-GHz LNA for 5G Wireless Systems in 28-nm Bulk CMOS," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 65, no. 10, pp. 1460-1464, Oct. 2018.

[3-12] Y. Peng et al., "A ${K}$ -Band High-Gain and Low-Noise Folded CMOS Mixer Using Current-Reuse and Cross-Coupled Techniques," in IEEE Access, vol. 7, pp. 133218-133226, 2019.

[3-13] F. Chen, Y. Wang, J. Lin, Z. Tsai and H. Wang, "A 24-GHz High Linearity Down-conversion Mixer in 90-nm CMOS," 2018 IEEE International Symposium on Radio-Frequency Integration Technology (RFIT), Melbourne, VIC, 2018, pp. 1-3.

[3-14] S. D'Amico, J. Ryckaert and A. Baschirotto, "An up-to-1GHz low-power baseband chain for UWB receivers," 2006 Proceedings of the 32nd European Solid-State Circuits Conference, Montreux, 2006, pp. 263-266.

[3-15] S. D'Amico, A. Baschirotto, K. Philips, O. Rousseaux and B. Gyselinckx, "A 240MHz programmable gain amplifier & filter for ultra-low power low-rate UWB receivers," 2009 Proceedings of ESSCIRC, Athens, 2009, pp. 260-263.

[3-16] H. Y. Shih, C. N. Kuo, W. H. Chen, T. Y. Yang and K. C. Juang, "A 250 MHz 14 dB-NF 73 dB-Gain 82 dB-DR Analog Baseband Chain With Digital-Assisted DC-Offset Calibration for Ultra-Wideband," in IEEE Journal of Solid-State Circuits, vol. 45, no. 2, pp. 338-350, Feb. 2010.

[3-17] S. D'Amico, M. De Blasi, M. De Matteis and A. Baschirotto, "A 255 MHz Programmable Gain Amplifier and Low-Pass Filter for Ultra Low Power Impulse-Radio UWB Receivers," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 59, no. 2, pp. 337-345, Feb. 2012.

[3-18] Y. S. Shu, "A 6b 3GS/s 11mW fully dynamic flash ADC in 40nm CMOS with reduced number of comparators," 2012 Symposium on VLSI Circuits (VLSIC), Honolulu, HI, 2012, pp. 26-27.

[3-19] V. H. C. Chen and L. Pileggi, "An 8.5mW 5GS/s 6b flash ADC with dynamic offset calibration in 32nm CMOS SOI," 2013 Symposium on VLSI Circuits, Kyoto, 2013, pp. C264-C265.

[3-20] C. H. Chan, Y. Zhu, S. W. Sin, U. Seng-Pan and R. P. Martins, "26.5 A 5.5mW 6b 5GS/S 4×-lnterleaved 3b/cycle SAR ADC in 65nm CMOS," 2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers, San Francisco, CA, 2015, pp. 1-3.

[3-21] C. H. Chan, Y. Zhu, I. M. Ho, W. H. Zhang, S. P. U and R. P. Martins, "16.4 A 5mW 7b 2.4GS/s 1-then-2b/cycle SAR ADC with background offset calibration," 2017 IEEE International Solid-State Circuits Conference (ISSCC), San Francisco, CA, 2017, pp. 282-283.

[3-22] L. Kull et al., "28.5 A 10b 1.5GS/s pipelined-SAR ADC with background second-stage common-mode regulation and offset calibration in 14nm CMOS FinFET," 2017 IEEE International Solid-State Circuits Conference (ISSCC), San Francisco, CA, 2017, pp. 474-475.

[3-23] M. Guo, J. Mao, S. Sin, H. Wei and R. P. Martins, "A 29mW 5GS/s Time-interleaved SAR ADC achieving 48.5dB SNDR With Fully-Digital Timing-Skew Calibration Based on Digital-Mixing," 2019 Symposium on VLSI Circuits, Kyoto, Japan, 2019, pp. C76-C77.

[3-24] J. Han et al., "A Generated 7GS/s 8b Time-Interleaved SAR ADC with 38.2dB SNDR at Nyquist in 16nm CMOS FinFET," 2019 IEEE Custom Integrated Circuits Conference (CICC), Austin, TX, USA, 2019, pp. 1-4.

[3-25] C. Yang and T. Kuo, "A 3mW 6b 4GS/s Subranging ADC with Adaptive Offset-Adjustable Comparators," 2019 IEEE Custom Integrated Circuits Conference (CICC), Austin, TX, USA, 2019, pp. 1-4.

[3-26] M. Guo, J. Mao, S. Sin, H. Wei and R. P. Martins, "A 1.6-GS/s 12.2-mW Seven-/Eight-Way Split Time-Interleaved SAR ADC Achieving 54.2-dB SNDR With Digital Background Timing Mismatch Calibration," in IEEE Journal of Solid-State Circuits.

[3-27] Q. Fan and J. Chen, "A 1-GS/s 8-Bit 12.01-fJ/conv.-step Two-Step SAR ADC in 28-nm FDSOI Technology," in IEEE Solid-State Circuits Letters, vol. 2, no. 9, pp. 99-102, Sept. 2019.

[3-28] W. Jiang, Y. Zhu, M. Zhang, C. Chan and R. P. Martins, "A Temperature-Stabilized Single-Channel 1-GS/s 60-dB SNDR SAR-Assisted Pipelined ADC With Dynamic Gm-R-Based Amplifier," in IEEE Journal of Solid-State Circuits.

[3-29] Q. Fan and J. Chen, "A 2.4 GS/s 10-Bit Time-Interleaved SAR ADC with a Bypass Window and Opportunistic Offset Calibration," ESSCIRC 2019 - IEEE 45th European Solid State Circuits Conference (ESSCIRC), Cracow, Poland, 2019, pp. 301-304.

[3-30] E. Swindlehurst et al., "An 8-bit 10-GHz 21-mW Time-Interleaved SAR ADC With Grouped DAC Capacitors and Dual-Path Bootstrapped Switch," in IEEE Solid-State Circuits Letters, vol. 2, no. 9, pp. 83-86, Sept. 2019.

[3-31] Y. Lyu and F. Tavernier, "A 4-GS/s 39.9-dB SNDR 11.7-mW Hybrid Voltage–Time Two-Step ADC With Feed-Forward Ring Oscillator-Based TDCs," in IEEE Solid-State Circuits Letters, vol. 2, no. 9, pp. 163-166, Sept. 2019.

[3-32] A. Agrawal and A. Natarajan, "2.2 A scalable 28GHz coupled-PLL in 65nm CMOS with single-wire synchronization for large-scale 5G mm-wave arrays," 2016 IEEE International Solid-State Circuits Conference (ISSCC), San Francisco, CA, 2016, pp. 38-39.

[3-33] W. El-Halwagy, A. Nag, P. Hisayasu, F. Aryanfar, P. Mousavi and M. Hossain, "A 28-GHz Quadrature Fractional-N Frequency Synthesizer for 5G Transceivers With Less Than 100-fs Jitter Based on Cascaded PLL Architecture," in IEEE Transactions on Microwave Theory and Techniques, vol. 65, no. 2, pp. 396-413, Feb. 2017.

[3-34] S. Yoo, S. Choi, J. Kim, H. Yoon, Y. Lee and J. Choi, "19.2 A PVT-robust −39dBc 1kHz-to-100MHz integrated-phase-noise 29GHz injection-locked frequency multiplier with a 600μW frequency-tracking loop using the averages of phase deviations for mm-band 5G transceivers," 2017 IEEE International Solid-State Circuits Conference (ISSCC), San Francisco, CA, 2017, pp. 324-325.

[3-35] S. Ek et al., "A 28-nm FD-SOI 115-fs Jitter PLL-Based LO System for 24–30-GHz Sliding-IF 5G Transceivers," in IEEE Journal of Solid-State Circuits, vol. 53, no. 7, pp. 1988-2000, July 2018.

[3-36] B. Murmann, "ADC Performance Survey 1997-2021," [Online]. Available: http://web.stanford.edu/~murmann/adcsurvey.html.

3.5 ANNEX 3.1

In this annex is provided the equation derivation of the classic PLL architecture of Figure A.3.1.a

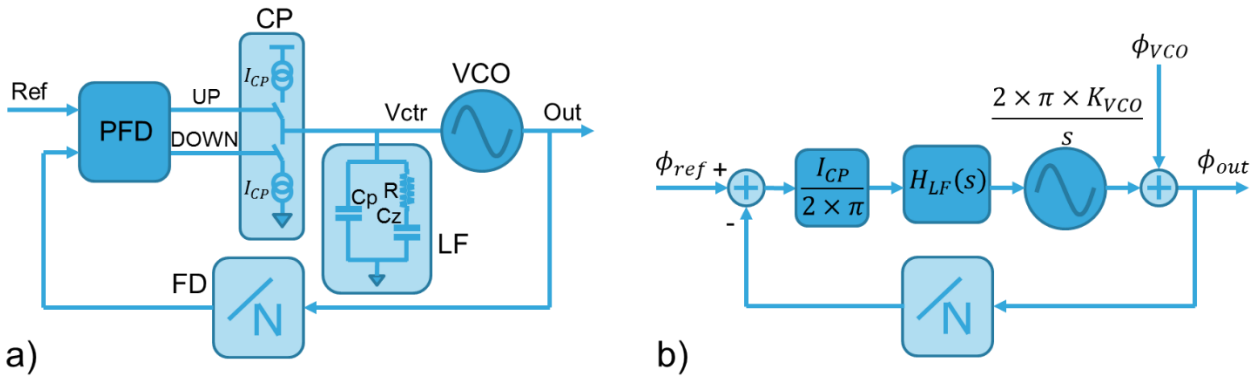

Figure A.3.1: a) PLL Classical implementation b) Equivalent LTI model

Figure A.3.1.b proposes an LTI model of the PLL. The PFD is modeled as continuous time linear comparator and the CP by a linear gain equal to its current divided by $2 \times \pi$ to account for the transformation from time to phase of the PFD output. This linear approximation is acceptable as long as the LF is a Low Pass (LP) with a cutoff frequency much lower than the comparison rate of the PFD. The control voltage of the VCO adjusts its frequency, the phase being the integral of the frequency it is modeled by an integrator with a linear gain K_{VCO}. It represents the sensitivity of the VCO output frequency to the control voltage variations. The VCO's phase noise is modeled by an additive noise source at its output with the spectral characteristics of Figure 3-3.

The following equations allow to describe the output phase noise of the PLL:

$$\phi_{out}(s) = H_{\phi_{VCO}}(s) \times \phi_{VCO}(s) + H_{\phi_{ref}}(s) \times \phi_{ref}(s) \tag{A.3.1}$$

With:

$$H_{\phi_{ref}}(s) = N \times \frac{H_{OL}(s)}{1 + H_{OL}(s)} \tag{A.3.2}$$

$$H_{\phi_{VCO}}(s) = \frac{1}{1 + H_{OL}(s)} \tag{A.3.3}$$

And:

$$H_{OL}(s) = \frac{I_{CP} \times H_{LF}(s) \times K_{VCO}}{s \times N} \tag{A.3.4}$$

The loop filter cannot be a simple integrator for stability reasons. A simple way to understand it is to study the open loop transfer function. One stability criterion is that it must intersect the $0dB$ line with a 20dB/decade slope. If $H_{LF}(s) \propto \frac{1}{s}$ then $H_{OL}(s) \propto \frac{1}{s^2}$ and the 0dB line is crossed with a $40dB/decade$ slope, leading to instability. One way to avoid this problem is to add a zero in the loop filter such that the $0dB$ line is crossed with the appropriate $20dB/decade$ slope. The most classical loop filter is depicted in Figure A.3.1.a. It also introduces one additional pole at higher frequency.

$$H_{LF}(s) = \frac{V_{ctr}(s)}{I_{in}(s)} = \frac{1 + s \times R \times C_z}{s \times (C_p + C_z) \times \left(1 + s \times R \times \frac{C_p \times C_z}{C_p + C_z}\right)} \tag{A.3.5}$$

$$s_p = -\frac{1}{R \times \frac{C_p \times C_z}{C_p + C_z}} \; and \; s_z = -\frac{1}{R \times C_z} \tag{A.3.6}$$

$$H_{LF}(s) = \frac{1}{s \times \left(C_p + C_z\right)} \times \frac{1 - \frac{s}{s_z}}{1 - \frac{s}{s_p}} = \frac{R \times C_z}{C_p + C_z} \times \frac{1 - \frac{s_z}{s}}{1 - \frac{s}{s_p}} \tag{A.3.7}$$

$$H_{OL}(s) = \frac{K_{VCO} \times I_{CP} \times R \times C_z}{N \times s \times \left(C_p + C_z\right)} \times \frac{1 + \frac{1}{s \times R \times C_z}}{1 + s \times R \times \frac{C_p \times C_z}{C_p + C_z}} = \frac{\omega_u}{s} \times \frac{1 - \frac{s_z}{s}}{1 - \frac{s}{s_p}} \tag{A.3.8}$$

With:

$$\omega_u = \frac{K_{VCO} \times I_{CP} \times R \times C_z}{N \times \left(C_p + C_z\right)} \tag{A.3.9}$$

There are three interesting point to look at, when $s = j \times \omega$ goes to zero, ω_u and infinity:

$$H_{OL}(j \times \omega) = \begin{cases} \overset{\omega \to 0}{\Longrightarrow} \dfrac{\omega_u \times s_z}{\omega^2} \\ \overset{\omega \to \omega_u}{\Longrightarrow} \dfrac{\omega_u}{\omega} \times \left| \dfrac{1 + j \times \frac{s_z}{\omega_u}}{1 - j \times \frac{\omega_u}{s_p}} \right| \times e^{j \times \left(\frac{3 \times \pi}{2} + \arctan\left(\frac{s_z}{\omega_u}\right) + \arctan\left(\frac{\omega_u}{s_p}\right) \right)} \\ \overset{\omega \to +\infty}{\Longrightarrow} \dfrac{\omega_u \times s_p}{\omega^2} \end{cases} \tag{A.3.10}$$

The zero s_z and pole s_p, and subsequently R, C_z and C_p, must be chosen to provide the desired Phase Margin (PM) at the unity gain angular frequency to ensure system stability. One easy way to do that is first to set $\omega_u = \sqrt{s_z \times s_p}$. The open loop gain at ω_u then becomes:

$$\begin{aligned} H_{OL}\left(j \times \sqrt{s_z \times s_p}\right) &= \left| \frac{1 - j \times \sqrt{\left|\frac{s_z}{s_p}\right|}}{1 + j \times \sqrt{\left|\frac{s_z}{s_p}\right|}} \right| \times e^{j \times \left(\frac{3 \times \pi}{2} - 2 \times \arctan\left(\sqrt{\left|\frac{s_z}{s_p}\right|} \right) \right)} \\ &= e^{j \times \left(\frac{\pi}{2} + 2 \times \arctan\left(\sqrt{\left|\frac{s_p}{s_z}\right|} \right) \right)} \end{aligned} \tag{A.3.11}$$

This way, ω_u effectively becomes the unit gain angular frequency and the PM can be expressed as:

$$PM = \pi - \arg\left(H_{OL}\left(j \times \sqrt{s_z \times s_p}\right)\right) = \frac{\pi}{2} - 2 \times \arctan\left(\sqrt{\left|\frac{s_p}{s_z}\right|} \right) < 0 \tag{A.3.12}$$

Note that the PM is negative which is unusual but not a problem. The stability criterion aims at measuring the distance to 180°, at which point the feedback becomes positive and the system unstable. There is no difference in moving the open loop phase shift one way or the other to reach stability.

The PM equation can be reversed to find the required pole/zero frequency ratio R_{PZ} and the filter element values.

$$R_{PZ} = \frac{s_p}{s_z} = \frac{C_p + C_z}{C_p} = \tan^2\left(\frac{\frac{\pi}{2} - PM}{2}\right) \xrightarrow{PM=-\frac{\pi}{3}} \tan^2\left(\frac{5 \times \pi}{12}\right) \approx 13.93 \quad \text{(A.3.13)}$$

$$|s_p| = \sqrt{R_{PZ}} \times \omega_u \; and \; |s_z| = \frac{\omega_u}{\sqrt{R_{PZ}}} \quad \text{(A.3.14)}$$

$$C_p + C_z = \frac{K_{VCO} \times I_{CP} \times \sqrt{R_{PZ}}}{N \times \omega_u^2} \quad \text{(A.3.15)}$$

$$C_p = \frac{C_p + C_z}{R_{PZ}} = \frac{K_{VCO} \times I_{CP}}{N \times \omega_u^2 \times \sqrt{R_{PZ}}} \quad \text{(A.3.16)}$$

$$C_z = \frac{K_{VCO} \times I_{CP}}{N \times \omega_u^2} \times \frac{R_{PZ} - 1}{\sqrt{R_{PZ}}} \quad \text{(A.3.17)}$$

$$R = \frac{1}{|s_z| \times C_z} = \frac{N \times \omega_u}{K_{VCO} \times I_{CP}} \times \frac{R_{PZ}}{R_{PZ} - 1} \quad \text{(A.3.18)}$$

The PLL close loop bandwidth is approximately ω_u. The phase noise is dominated by N times the reference clock phase noise inside this band, and by the VCO's one outside, with a transition zone in between. Figure A.3.2 plots the phase noise of time domain simulation implemented in matlab against the prediction of the LTI frequency model just described.

The center frequency is $22.4GHz$. The reference noise floor is $-155dBc/Hz$. A typical reference will exhibit a colored phase noise at low frequency. Here the simulation only goes down to $10kHz$ frequency offset where this behavior is not yet visible. The free running oscillator phase noise is $-52dBc/Hz$ at $10kHz$, $-80dBc/Hz$ at $100kHz$, $-105dBc/Hz$ at $1MHz$ and has a noise floor of $-135dBc/Hz$. The open loop unity gain frequency is $1.29MHz$ and the phase margin is 60°.

The times domain simulation and the frequency model are in good agreement. The main difference is that the time domain implementation uses a discrete time feedback divider, the FD, and phase comparator, the PFD. For this reason, the reference clock high frequency noise must be filtered to avoid detrimental aliasing. In practice the reference clock buffer has a limited bandwidth and act as a filter. It is less of an issue for the feedback clock since the sampling happens at the input of the frequency divider, i.e. at a much higher sampling frequency. Nonetheless, the oscillator buffer bandwidth must be considered during design. The other components of the time domain model are LTI and behave as expected. The jitter performance of the time domain model is slightly worth than what the LTI frequency domain model predicts probably due to some noise aliasing remaining.

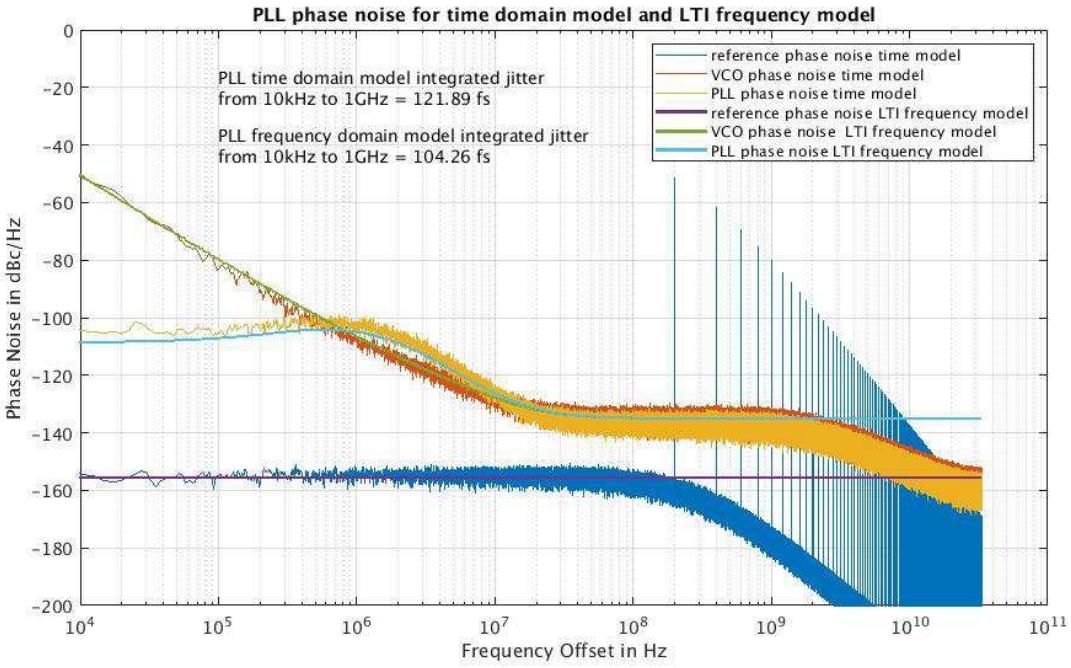

Figure A.3.2: Phase Noise of a matlab time domain implementation of the PLL versus the LTI frequency model with a center frequency of 22.4GHz

The unity gain frequency was chosen to minimize the integrated jitter from $10kHz$ to $1GHz$. This was done using the LTI frequency model since it is in good agreement with the time domain model and is much faster to simulate. The result of this optimization is depicted on the left graph of Figure A.3.3.

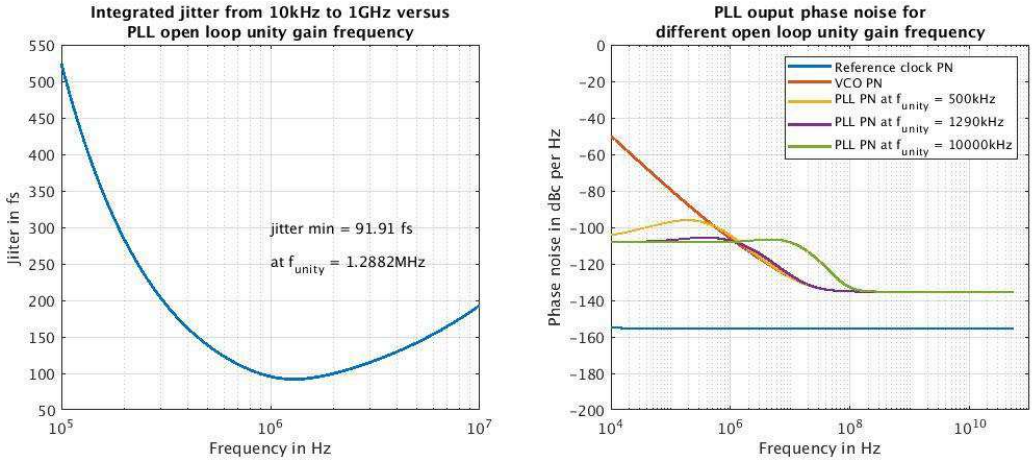

Figure A.3.3: Left) Integrated jitter from 10 kHz to 1 GHz versus PLL open loop unity gain frequency. Right) PLL output phase noise for different open loop unity gain frequency

Figure A.3.3's right graph plots the output phase of the reference clock, the free running oscillator and the PLL for open loop unity gain frequencies of $500kHz$, $1.29MHz$ and $10MHz$. When the bandwidth is too low the output phase noise is degraded by the colored noise of the VCO. When it is too high the reference clock phase noise gained by the feedback division ratio becomes more powerful than the VCO noise floor, degrading the overall phase noise performances. The left graph shows the importance of the unity gain frequency optimization allowing the integrated jitter to go down below $100fs$, while it can be well above $150fs$ for other bandwidths.

The time domain model gives a higher jitter. This may be caused by some aliasing remaining and some numerical error in particular in the integration process since at low frequency, where the PN is the highest, the frequency steps are relatively coarse.

3.6 ANNEX 3.2

In this annex are provided the equation derivations for the Hartley and Weaver image rejection techniques discussed in section 3.2.1. The Hartley approach, with an example of a digital implementation in Figure A.3.4, will be described first.

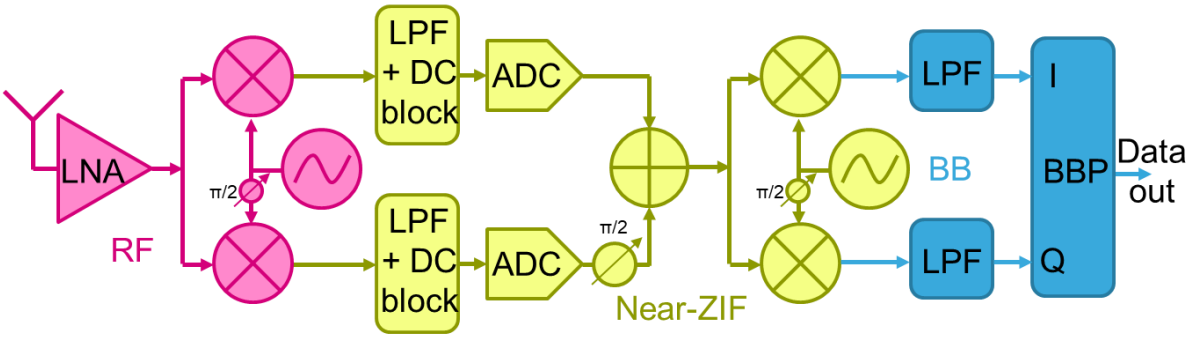

Figure A.3.4: Near-ZIF Receiver architecture with a digital implementation of the Hartley image rejection technic

The RF signal S_{RF} is first amplified by the LNA and quadrature down mixed to a low intermediate frequency ω_{IF}, providing $S_{I_{IF}}$ and $S_{Q_{IF}}$. These signals are the digitized. For the purpose of this argumentation the LPF is assumed to be a perfect brick wall filter and the ADC to have infinite resolution. The RF signal is composed of the desired signal at the center frequency ω_c and a signal at the image frequency ω_I as per equation (A.3.19).

$$S_{RF}(t) = \rho_S(t) \times \cos\big(\omega_c \times t - \theta_S(t)\big) + \rho_I(t) \times \cos\big(\omega_I \times t - \theta_I(t)\big) \tag{A.3.19}$$

The LO frequency ω_{LO} can be set at $\omega_c - \omega_{IF}$ or $\omega_c + \omega_{IF}$. In the present derivation only the case $\omega_{LO} = \omega_c - \omega_{IF}$ is treated. The second case can be directly obtained by setting $\omega_{IF_1} = -\omega_{IF_2}$ in the first case equations. For the first case the relationships between ω_c, ω_I and ω_{LO} are given in equations (A.3.20) and (A.3.21):

$$\omega_c = \omega_{LO} + \omega_{IF} \tag{A.3.20}$$

$$\omega_I = \omega_{LO} - \omega_{IF} \tag{A.3.21}$$

The RF signal can then be re-written in terms of ω_{LO} and ω_{IF}:

$$\begin{aligned} S_{RF}(t) = {} & \rho_S(t) \times \cos\big((\omega_{LO} + \omega_{IF}) \times t - \theta_S(t)\big) \\ & + \rho_I(t) \times \cos\big((\omega_{LO} - \omega_{IF}) \times t - \theta_I(t)\big) \end{aligned} \tag{A.3.22}$$

The quadrature down mixing of this RF signal gives the following two signals:

$$\begin{aligned} S_{I_{IF}}(t) = {} & \frac{\rho_S(t)}{2} \times \big(\cos\big((2 \times \omega_{LO} + \omega_{IF}) \times t - \theta_S(t)\big) + \cos\big(\omega_{IF} \times t - \theta_S(t)\big)\big) \\ & + \frac{\rho_I(t)}{2} \times \big(\cos\big((2 \times \omega_{LO} - \omega_{IF}) \times t - \theta_I(t)\big) + \cos\big(\omega_{IF} \times t + \theta_I(t)\big)\big) \end{aligned} \tag{A.3.23}$$

$$S_{Q_{IF}}(t) = \frac{\rho_S(t)}{2} \times \left(\sin\left((2 \times \omega_{LO} + \omega_{IF}) \times t - \theta_S(t)\right) - \sin\left(\omega_{IF} \times t - \theta_S(t)\right)\right) + \frac{\rho_I(t)}{2} \times \left(\sin\left((2 \times \omega_{LO} - \omega_{IF}) \times t - \theta_I(t)\right) + \sin\left(\omega_{IF} \times t + \theta_I(t)\right)\right) \quad \text{(A.3.24)}$$

The high frequency image is then low pass filtered giving:

$$LPF\left(S_{I_{IF}}(t)\right) = \frac{\rho_S(t)}{2} \times \cos\left(\omega_{IF} \times t - \theta_S(t)\right) + \frac{\rho_I(t)}{2} \times \cos\left(\omega_{IF} \times t + \theta_I(t)\right) \quad \text{(A.3.25)}$$

$$LPF\left(S_{Q_{IF}}(t)\right) = -\frac{\rho_S(t)}{2} \times \sin\left(\omega_{IF} \times t - \theta_S(t)\right) + \frac{\rho_I(t)}{2} \times \sin\left(\omega_{IF} \times t + \theta_I(t)\right) \quad \text{(A.3.26)}$$

Using the identity $\sin(a) = \cos\left(a - \pi/2\right)$, (A.*3.26*) can re-written as (A.*3.27*):

$$LPF\left(S_{Q_{IF}}(t)\right) = -\frac{\rho_S(t)}{2} \times \cos\left(\omega_{IF} \times t - \theta_S(t) - \frac{\pi}{2}\right) + \frac{\rho_I(t)}{2} \times \cos\left(\omega_{IF} \times t + \theta_I(t) - \frac{\pi}{2}\right) \quad \text{(A.3.27)}$$

The idea of the Hartley image cancellation technic is to note that the in-phase and quadrature phase IF signals both contain the RF and the image frequency signals carried by a cosine wave but with a different polarity and a 90° phase shift of the IF carrier. Phase shifting the quadrature phase IF signal by an additional 90°, as in (A.*3.28*), and summing the result with the in-phase IF signal will lead to a cancellation of the image frequency signal.

$$PS_{\frac{\pi}{2}}\left(LPF\left(S_{Q_{IF}}(t)\right)\right) = \frac{\rho_S(t)}{2} \times \cos\left(\omega_{IF} \times t - \theta_S(t)\right) - \frac{\rho_I(t)}{2} \times \cos\left(\omega_{IF} \times t + \theta_I(t)\right) \quad \text{(A.3.28)}$$

The image cancellation performances will be limited by the gain matching the in-phase and quadrature phase chains, and on the total accuracy of the quadrature down mixing and the 90° phase shift. In general, it is this second constraint that limits the overall cancellation performances of the Hartley approach.

As for the Hartley approach, the RF signal is quadrature down mixed to IF.

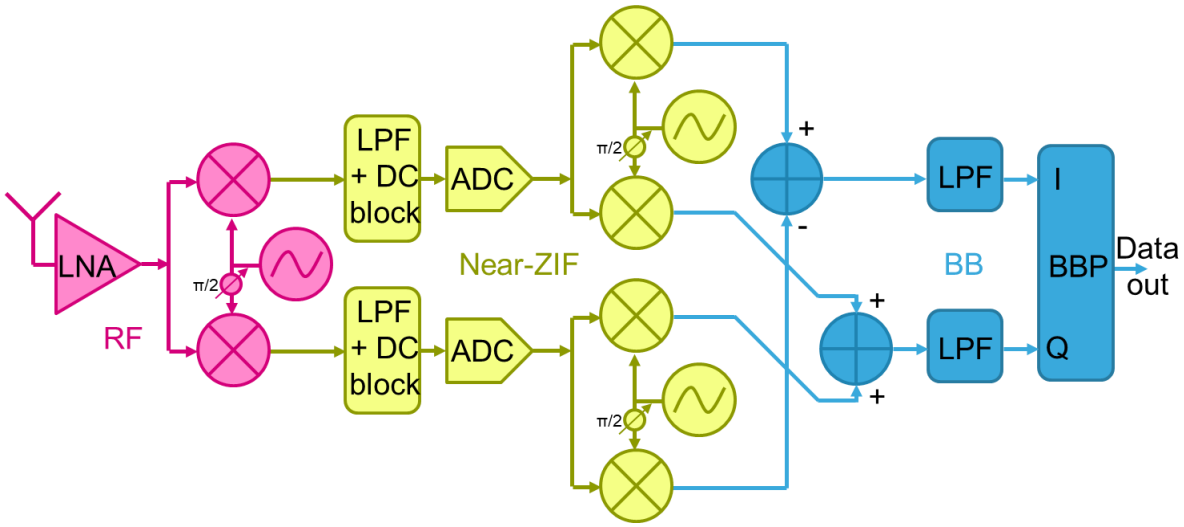

Figure A. 3.5: Near-ZIF Receiver architecture with a digital implementation of the Weaver image rejection technic

The low pass filtered IF signals are then themselves quadrature down mixed to base band producing the four signals in (A.*3.29*) (A.*3.30*) (A.*3.31*) and (A.*3.32*):

$$S_{I_{I_{BB}}}(t) = \frac{\rho_S(t)}{4} \times \cos\big(\theta_S(t)\big) + \frac{\rho_I(t)}{4} \times \cos\big(\theta_I(t)\big) \tag{A.3.29}$$

$$S_{I_{Q_{BB}}}(t) = \frac{\rho_S(t)}{4} \times \sin\big(\theta_S(t)\big) - \frac{\rho_I(t)}{4} \times \sin\big(\theta_I(t)\big) \tag{A.3.30}$$

$$S_{Q_{I_{BB}}}(t) = \frac{\rho_S(t)}{4} \times \sin\big(\theta_S(t)\big) + \frac{\rho_I(t)}{4} \times \sin\big(\theta_I(t)\big) \tag{A.3.31}$$

$$S_{Q_{Q_{BB}}}(t) = -\frac{\rho_S(t)}{4} \times \cos\big(\theta_S(t)\big) + \frac{\rho_I(t)}{4} \times \cos\big(\theta_I(t)\big) \tag{A.3.32}$$

For the sake of clarity, only the base band image is considered. Since all the base band processing in the digital domain is linear, the high frequency image can be filtered before or after the cancellation processing.

Subtracting (A.*3.32*) from (A.*3.29*) allows to cancel the image frequency signal from the in-phase base band signal, and summing (A.*3.30*)with (A.*3.31*) allows to do the same for the quadrature phase base band signal.

The image cancellation performances, as for the Hartley approach, will be limited by the gain matching of the in-phase and quadrature phase paths, and on the total phase accuracy of the three quadrature down mixing. In general, it is this second constraint that limits the overall cancellation performances of the Weaver approach.

4 CHAPTER IV: RECEIVER'S ARCHITECTURE

The feasibility of DBF for the Near-ZIF architecture was established, or at least for its analog part. But this does not resolve the challenge of the massive amount of digital processing required downstream. An attractive solution to this problem was originally proposed for ultrasound systems in [4-16], and was recently reused in the context of beamforming for 5G by the authors in [4-1]. They use band pass Sigma-Delta Modulators (SDM) to digitize a low IF. They use them in a particular configuration, allowing to optimize the digital processing regarding the following aspects:

- Digital down mixing
- True time delay
- Symbol rotation

In the first part, how SDM can bring these improvements will be studied. The result of this study will become the major motivation to investigate further SDM in the context of DBF receivers.

In the second part, starting from the basics of sigma-delta modulators, their characteristics will be studied. In particular, the focus will be put on how they can be used to reduce other building blocks constraints and even simplify the receiver's architecture.

Finally, in the third part, the proposed architecture will be derived: a direct RF sampling receiver based on a band pass continuous time sigma-delta modulator and using sub-sampling. This approach simplifies the receiver's architecture to the maximum, a sigma-delta modulator.

4.1 DIGITAL PROCESSING EFFICIENT IMPLEMENTATION

While the feasibility of the analog portion of a DBF Near-ZIF receiver was demonstrated, the question of the feasibility of its digital portion is still open. In this section, the potential complexity reduction a sigma-delta based DBF receiver can offer compared to a receiver using a Nyquist ADC will be reviewed. Ultimately, this will lead to the final reformulation of the question this manuscript is trying to answer, which is currently: Is digital beamforming for millimeter wave 5G system possible and how?

Starting from the assumption that an adequate SDM can be build, it will be shown how this can significantly simplify the beamforming digital processing. These simplifications revolve around three key points, the digital down mixing, the true time delay and the symbol rotation. An additional discussion will be carried out about decimation filters, which are necessary for SDM. Each of these points will be addressed separately. It is these four points that makes SDM so appealing for this use case.

4.1.1 Digital down-mixing

When digitizing an IF the final down-mixing needs to be done in the digital world. One significant benefit from digital down-mixing is that it can be done with no self-mixing at all. This requires generating a digital quadrature LO. A classic trick is to digitize the IF with a sampling frequency four time higher. The required LO frequency is then a fourth of the sampling frequency. The sine and cosine waves are then simply successions of ones, zeros and minus ones as depicted in Figure 4-1.

The down-mixing operation is then reduced to either do nothing, perform a sign change or set the sample to zero. Any multiplication is removed from the process. Only very low complexity operations remain. Another consequence is that the I and Q BB signals have every other samples equal to zero. These zeros are shifted by one sample between I and Q such that at each time step only one of the two may have a non-zero sample. This characteristic will be used later to further reduce the digital processing complexity.

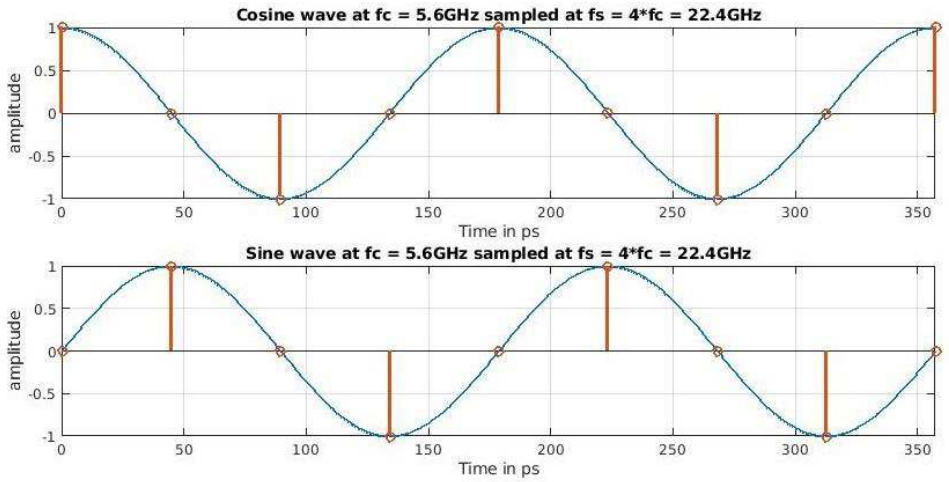

Figure 4-1: Cosine (top) and sine (bottom) waves sampled at $f_s = 4 \times f_c$

4.1.2 True time delay

The delay between the different antennas has two effects. To understand that let us reuse the equations derived in the previous chapter in the phase noise analysis. Ignoring the phase noise terms gives the following set of equations for the Tx BB and RF signals:

$$S_{BB_{Tx}}(t) = \rho(t) \times e^{j \times \theta(t)} \tag{4.1}$$

$$S_{RF_{Tx}}(t) = \rho(t) \times \cos\big(\omega_{LO} \times t - \theta(t)\big) \tag{4.2}$$

When the receiver has an antenna array, the antenna A_i of the array will receive the transmitted signal with a delay τ_i. This delay depends on the angle of arrival and the antenna position as detailed in chapter 2. The received signal $S_{RF_{SRx_i}}(t)$ of given antenna A_i is then expressed as:

$$\begin{aligned} S_{RF_{SRx_i}}(t) = S_{RF_{Tx}}(t - \tau_i) &= \rho(t - \tau_i) \times \cos\big(\omega_{LO} \times (t - \tau_i) - \theta(t - \tau_i)\big) \\ &= \rho(t - \tau_i) \times \cos\Big(\omega_{LO} \times t - \big(\omega_{LO} \times \tau_i + \theta(t - \tau_i)\big)\Big) \end{aligned} \tag{4.3}$$

Once the BB signal is recovered through quadrature down-mixing, equation (4.4) is obtained. Here it is assumed for clarity that the Rx LO has the same phase as the Tx one.

$$S_{BB_{SRx_i}}(t) = \frac{1}{2} \times S_{BB_{TX}}(t - \tau_i) \times e^{j \times \omega_{LO} \times \tau_i} \tag{4.4}$$

Two effects can be observed. First, the symbols are delayed and second, they are rotated. Ideal beamforming consists in compensating both effects before recombining all the signals. When beamforming is implemented through phase shifting, it assumes the symbol delay is negligible compared to the symbol duration and compensates only the symbol rotation. This is called the narrow band approximation.

This approximation is less and less true as the array size and the data rate increase. More antennas mean greater maximum distance between the antennas and therefor greater maximum delay between any two antennas. Higher data rate means smaller symbol duration. Both tends to go toward a weakening of the narrow band approximation. Figure 4-2 plots the maximum delay between any two antennas as a function of the AoA for the 364 elements MC-UCA of section 2.2.3.2 with an outer diameter of $24cm$.

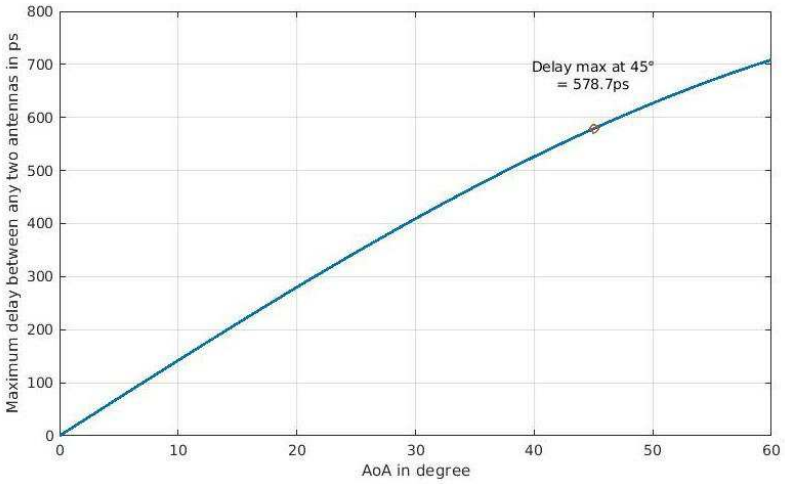

Figure 4-2: Maximum delay between any two antennas versus the angle of arrival for an MC-UCA of 364 elements with an outer diameter of 24cm

As one could expect this delay increases with the AoA going farther away from the normal angle. At the required limit angle of 45°, this delay is above $550ps$. For the considered bandwidth of $1GHz$, the symbol duration can go down to $1ns$. The maximum delay is clearly not negligible compared to the symbol duration, meaning the narrow band approximation is no more valid. Proper beamforming requires to add symbol delay compensation.

In sigma-delta modulators the signal is over sampled compared to the Nyquist rate imposed by the signal bandwidth. This is called the Over-Sampling Ratio (OSR). In the present case it means that each symbol is sampled many times with a short time step. The symbol delay compensation can then be partially done by simply selecting the right sample. If the sampling period is small compared to the symbol duration, the narrow band approximation can simply be ignored, and the delay compensation can be limited to a processing efficient sample selection. Thanks to SDM OSR it is possible to significantly reduce the amount of required digital processing required to implement digital true time delay. Only remains the symbol rotation processing. This is addressed in the next section.

4.1.3 Efficient symbol rotation

Symbol rotation is easily done by the complex multiplication of the symbol with a complex coefficient of unit modulus and argument of the desired rotation ϕ.

$$\begin{aligned} S_{rot} = S \times e^{j\times\phi} &= (I + j \times Q) \times (\cos(\phi) + j \times \sin(\phi)) \\ &= I \times \cos(\phi) - Q \times \sin(\phi) + j \times (I \times \sin(\phi) + Q \times \cos(\phi)) \end{aligned} \tag{4.5}$$

When complex numbers are expressed as real and imaginary parts, this operation requires four multiplications by two fixed coefficients and 2 additions. In section 4.1.1, it was mentioned that, out of the BB I and Q signals, only one of them was non-zero at a time. Hence, two of the four multiplications and both additions can be removed, and only two multiplications remain. This is a first processing reduction, but additional improvements can be done.

Together with over-sampling, sigma-delta modulators offer a second particularity which is to code the information in the frequency domain. This point will be clarified later but one consequence is that the sample's amplitude can be coded on a low number of bits, even down to a single one. Here, the value of this bit can simply be interpreted as the sample's polarity, meaning that it belongs to the set of values $\{-1; 1\}$. Therefore, the multiplication can be implemented with a simple multiplexer as depicted in Figure 4-3.

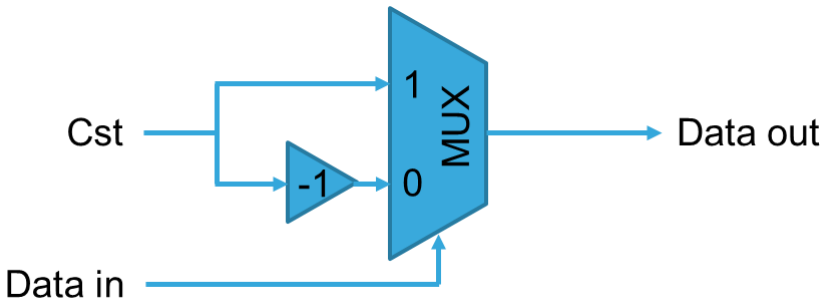

Figure 4-3: Multiplexer based multiplier

This allows for a nearly free multiplication where complexity is considered to be one. This reduced complexity comes at the cost of performing more multiplication due to the modulator's OSR. This will be detailed later in this chapter, but the OSR that will be used is around ten. It means the effective complexity for one multiplication is around this OSR value of ten.

To get a sense on if this is an improvement, it is necessary to compare it with a receiver using an ADC working at the Nyquist rate. In that case, its output would have the resolution evaluated in the previous chapter, i.e. about eight bits. Assuming the multiplication complexity of the Toom-Cook algorithm [4-2], a complexity of about 21 is reached. Hence the use of sigma-delta modulator has the potential to halve the cost of multiplications.

This multiplexer-based multiplier can also be used with a three level quantizer, and the authors in [4-1] use it with a five level quantizer. Fundamentally, this technique can be used with any number of quantization levels. As long as one of the input number is a constant it is technically possible to use it. In practice the multiplexer complexity grows exponentially and renders it interesting only for signals coded on a low number of levels. This will become one of the constraints when elaborating the receiver's architecture.

4.1.4 Decimation filter

While sigma-delta modulators allow for all these simplifications in the digital processing, they require one more step before the beamformed BB signal can be sent to the digital receiver. This is the filtering of the shaped quantization noise and the signal decimation to get rid of the unnecessary over sampling. These operations effectively convert the frequency coding of the information into the conventional amplitude coding used by synthetized digital circuits. They are generally implemented together by a so-called decimation filter. Again, this will be explained in more detail later. What is important to understand here is that, when looking at the sigma-delta output in the frequency domain, the quantization noise appears outside the band of interest. It is said to be shaped.

The purpose of the filtering is to remove this shaped noise. As a consequence, the requirement for the filter will depend on this noise's power. The question is then: When is it best to perform this filtering, before or after the beamforming operation, or a mix of both? There are two reasons allowing for a straightforward answer. First, all the benefits listed above are only applicable to the data before the decimation filter. Second, if performed after, only one filter needs to be implemented against one per antenna if performed before.

These reasons alone are so strong that it may seems strange to even think about it. But it is actually interesting since there is one additional benefit in having the decimation filter after beamforming. To explain it, a hypothesis that will be demonstrate later on will be used. This hypothesis is the un-correlation of the shaped noise between the SRx of the receiver. This means that the relative power of the out of band shaped noise, compared to the in-band signal, reduces after beamforming, since recombining incoherently. The direct consequence is the reduction of the filter requirement and an

additional reduction of the digital processing complexity. It could even be argued that further synergy could be found between this decimation filter and the digital demodulator, but this is left to future work.

4.1.5 Conclusion

After the demonstration of the feasibility of the analog part of a DBF Near-ZIF receiver, the question of the feasibility of its digital part was still open. All together the different techniques described here allow for a very low complexity true time delay digital beamforming where feasibility is hardly questionable since it essentially reduces to the complexity of one addition per antenna and per beam for recombination and one relaxed decimation filter per beam, all other operations having a negligible complexity in comparison. Hence, the question addressed in this manuscript can now be reformulated as: Is it possible for the analog portion of a sigma-delta based DBF receiver to achieve similar performances as a Near-ZIF receiver? A positive answer to this question would definitely make a strong case for digital beamforming in the context of millimeter wave 5G.

4.2 SIGMA-DELTA MODULATORS

Sigma-delta modulators were first introduced in 1962 [4-3] as an evolution of the delta modulator. This architecture rapidly proved its potential and was derived in multiple flavors [4-4]. There are today two large class of modulators implementation, the discrete and the continuous time ones, but both of them rely on the same principles. Starting from the basics, the underlying theory of sigma-delta modulators will be gradually unfolded to reach a deep understanding of the target architecture, the Band Pass Continuous Time Sigma-Delta Modulator (BPCTSDM). Based on this understanding, a new method of Excess Loop Delay (ELD) optimization will be developed for improved robustness to process, temperature and power supply variations.

4.2.1 Basic concepts

All the details of SDM modulators will not be reviewed here. The interested reader may go to one of the many references on the subject, such as the book from Richard Schreier and Gabor C. Temes, “Understanding Delta-Sigma Data Converters” [4-5]. It has been re-edited and augmented in 2017 and saw additions from Shanthi Pavan and is fairly complete and up to date. Here only some of the major results will be exemplified to get an intuitive understanding of SDMs.

The basic concept of sigma-delta modulators is the same for Continuous Time (CT) and Discrete Time (DT) modulators. The wide bandwidth targeted imposes the use of CT modulators. Regardless, this review will start with DT ones. First, because they are simpler to model and simulate, and second, because CT modulators can be brought back to a DT equivalent which is a classic way to study and design CTSDM.

Figure 4-4-a depicts the basic schematic of a DT modulator. In its simplest form it is a closed loop system composed of a loop filter $H(Z)$, a low resolution quantizer and a feedback DAC of the same resolution. The DAC output is negatively fed back to the input to close the loop. Here the input signal is discrete in time. To use such a circuit with a continuous time input, one must add a Sample and Hold (SH) circuit at the input.

To study this circuit, the equivalent LTI model from Figure 4-4-b is used. The quantizer is simply modeled as an additive noise source on the un-quantized signal. Its output is no more quantized, only noisy, so the DAC can simply be replaced by a delay. This delay ensures there is no delay free loop. Otherwise, the system could become un-causal, i.e. the current output could depend on its current value.

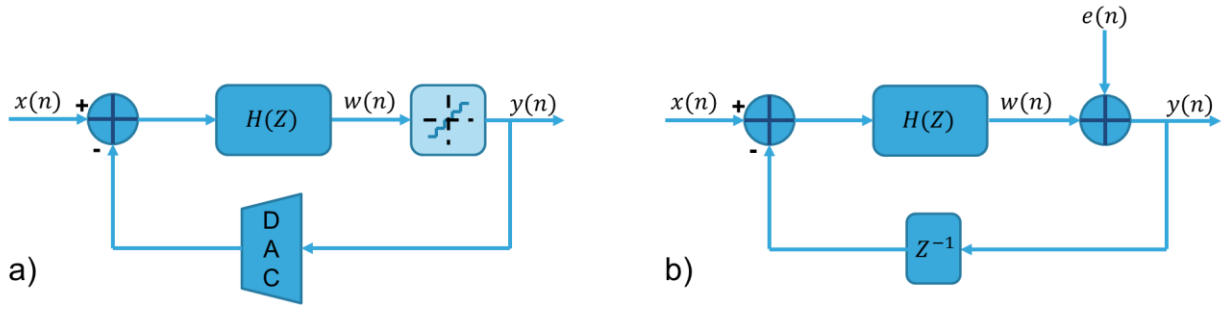

Figure 4-4: a) Basic schematic of a DT sigma-delta modulator, b) LTI equivalent model

There are two transfer functions of interest. The one from the input $x(n)$ to the output $y(n)$, called the Signal Transfer Function (STF), and the one from the quantizer noise input $e(n)$ to the output $y(n)$, called the Noise Transfer Function (NTF). The output is then expressed in the Z domain as:

$$Y(Z) = STF(Z) \times X(Z) + NTF(Z) \times E(Z) \tag{4.6}$$

With the following expression for the STF and the NTF:

$$STF(Z) = \frac{H(Z)}{1 + Z^{-1} \times H(Z)} \tag{4.7}$$

$$NTF(Z) = \frac{1}{1 + Z^{-1} \times H(Z)} \tag{4.8}$$

One can note that this is a similar form as the PLL studied in section 3.1.2.2.3. Therefore, the same conclusion can be reached. For the range of Z values where $|H(Z)| \gg 1$, the STF has unit gain and the NTF tends to zero. This means that the signals in this range of Z values are unaltered while the quantizer noise is attenuated. Note here that this effect is not limited to the quantization noise, but affects any signals added on the quantizer output. This means that all quantizer imperfections such as thermal noise, non-linearity and so on, will also be attenuated. This generally relax the quantizer requirements.

The loop filter $H(Z)$ may have two kind of characteristics, low pass or band pass. Each of them gives rise to their counter-part Low Pass Discrete Time Sigma-Delta Modulators (LPDTSDM) or Band Pass Discrete Time Sigma-Delta Modulators (BPDTSDM). Let us first start with the low pass modulators

4.2.1.1 LPDTSDM

The simplest configuration for these modulators is when the loop filter is a simple integrator with the transfer function of (4.9).

$$H(Z) = \frac{1}{1 - Z^{-1}} \tag{4.9}$$

The STF and the NTF then become:

$$STF(Z) = 1 \tag{4.10}$$

$$NTF(Z) = 1 - Z^{-1} \tag{4.11}$$

On one hand, the STF has a unit response independent of Z. On the other hand, the noise is differentiated. Only the noise variations from one sample to the next remains. The NTF has therefore a high pass characteristic. Noise variations that are much slower than the sampling rate bring nearly the same perturbation to two consecutive samples and get cancelled by this differentiation. Intuitively, this explains that, for a given bandwidth, increasing the sampling rate will allow to improve this noise

cancelation in the band. Indeed, as it will be seen, the Over Sampling Ratio (OSR) defined as per (4.12) will play a significant role in the performances of SDM.

$$OSR = \frac{f_s}{2 \times f_{max}} \tag{4.12}$$

Another interpretation is that the NTF has a zero at $\omega = 0$ with $Z = e^{j \times \omega}$. More generally what happens is that the poles from the loop filter $H(Z)$ become the zeros of the NTF. The modulator's order can then be defined as the number of poles of its loop filter or equivalently as the number of zeros of its NTF. The initial example is therefore a first order modulator.

This modulator can be represented using the model from Figure 4-5. With a single bit quantizer, it is very easy to implement in matlab (see ANNEX 4.1). The input can range from -1 to 1. This is called the full scale of the modulator and the input signal power is often referred to the power of a sinewave with a peak-to-peak amplitude equal to this full scale. Hence Input signals are often measured in decibel Full Scale (dB_{FS}), a classic unit for ADCs.

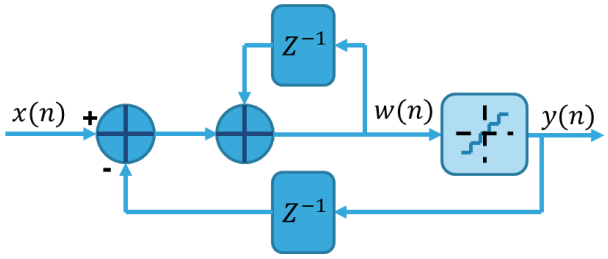

Figure 4-5: Simulation model of a first order LPDTSDM

The simplicity of high-level simulations of DT modulators makes it easier to experimentally explore their design space. This will significantly impact the way CT modulators are investigated.

Let us look at the main features of this modulator. Figure 4-6 plots different characteristics of the SDM. The left graph displays the evolution of the output SNR as the input signal power is increased. It presents a fairly linear characteristic up to an input power of nearly $0dB_{FS}$. It then drops abruptly. Intuitively this can be understood as follow. The high gain feedback loop tries to minimize the input signal. The DAC output being limited by its full scale; any larger signal cannot be efficiently minimized. Hence the loop is rendered dysfunctional beyond the full scale.

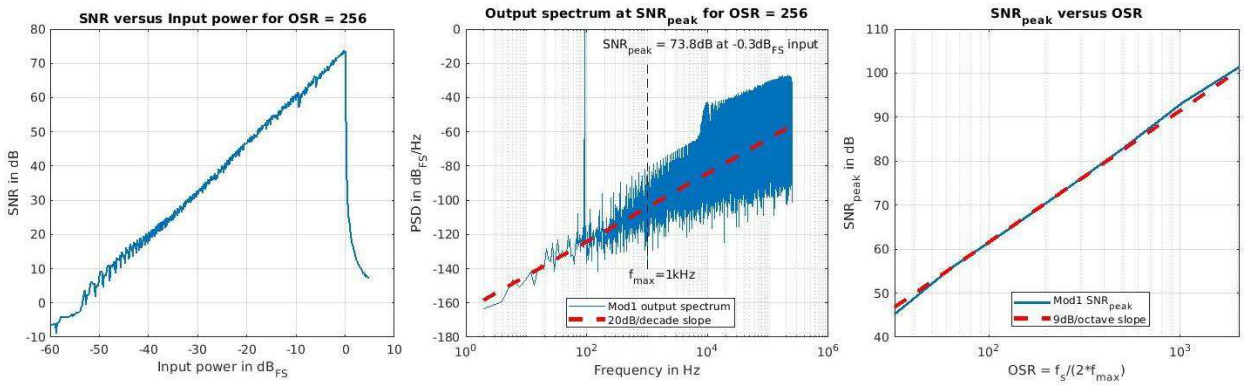

Figure 4-6: Simulation of a first order LPDTSDM. Left: SNR versus input power. Middle: Output spectrum at SNR_{peak} . Left: SNR_{peak} versus OSR

The middle graph in Figure 4-6 plots the output spectrum for the peak SNR and an OSR of 256. Even with a single bit quantizer the output SNR is above $73dB$ for a bandwidth of $1kHz$. One noticeable

characteristic is the $20dB/decade$ slope of the noise. It is shaped by the NTF. More generally the slope of the noise is $20dB/decade$ time the order of the modulator. The higher the order, the more attenuated the in-band noise is. This means that higher order modulators deliver better SNRs.

If one were to look at the output in the time domain, it would appear as a nearly random succession of plus ones and minus ones. It is only by looking at it in the frequency domain that the content of the signal can be understood. Hence it is said to be coded in the frequency domain.

Finally, the right graph of Figure 4-6 plots the peak SNR for different values of the over sampling ratio. The observed trend is that each doubling of the OSR adds about $9dB$ of SNR. The more general theoretical SNR formula for a modulator order L with a quantizer resolution N as a function of OSR is given by (4.13).

$$\begin{aligned} SNR_{peak}(dB) &= 10 \times \log_{10}\left(\frac{3}{2}\right) + 20 \times \log_{10}(2^N - 1) \\ +10 \times \log_{10}&\left(\frac{(2 \times L + 1)}{\pi^{2 \times L}}\right) + (2 \times L + 1) \times 10 \times \log_{10}(OSR) \end{aligned} \tag{4.13}$$

The two first terms correspond to the SNR of a classic Nyquist rate ADC. The two last ones correspond the SNR improvement brought by the modulator as a function of its order and OSR.

When applied to the modulator, this formula gives $SNR = 68.7dB$ for a full-scale input. The simulation gives an $SNR_{max} = 73.8fB$ for a $-0.3dB_{FS}$ input. The difference of $\sim 4dB$ is probably coming from the assumption that the quantization noise is white and uncorrelated with the input signal. It is known that this is generally not the case for low resolution quantizers. Nonetheless, together with the $20dB$ per decade noise slope and the $9dB$ per OSR octave SNR slope, this simulation is in good agreement with the theory.

The main conclusion here is that DTSDM benefit from a well-established theory and an efficient way to perform simulations. Only the surface of the low pass modulators was brushed but it is enough to get a somewhat intuitive understanding of such data converters.

4.2.1.2 BPDTSDM

To understand band, pass modulators, a more design-oriented approach will be used. The goal will be to look for an architecture that allows the attenuation of the quantization noise in a band centered on a frequency away from DC. A powerful tool to do that is to look at the position of the NTF zeros in the Z-domain complex plan.

Before making any arguments let us clear some mathematical notations. The Discrete Fourier Transform (DFT) is equal to the Z-Transform for the values of Z on the unit circle. That is $Z = e^{j \times \Omega}$. The DFT being $2 \times \pi$-periodic, only a limited range of Ω values can be considered. The choice was made to consider only the values in the interval $[-\pi; \pi[$. $\Omega = \frac{\omega}{f_s}$ is often called the normalized angular frequency.

In signal processing it is usual to use this normalized frequency. In most cases, signal processing systems are purely digital. Normalizing the raw data frequency once at the input allow then to use only normalized transfer functions avoiding any subsequent normalizing errors. While normalized variables are used for amplitudes and power, the use of the real frequency variable will be kept for the three following reasons:

- The purpose here is to study a mixed signal system where the input signal is a real analog signal, and its frequency cannot be physically normalized. Using the real frequency variable avoid

repeated normalization and de-normalization when going back and forth between the analog and the digital parts during the analysis

- People with a background in analog design, such as myself, generally have an intuition based on real frequencies. Normalized frequencies can sometimes be confusing.
- In general, normalizing allows to simplify some constants leading to simpler equations. Unfortunately, this also removes their associated units. This makes the equations homogeneity sometimes difficult to check. Since homogeneity is a powerful tool for error checking when handling unfamiliar equations, it is sometimes preferable to use un-normalized variables.

The unit circles in Figure 4-7 are labeled with fraction of the sampling frequency f_s. While this is mathematically incorrect, this notation will be kept remaining on a more analog intuition.

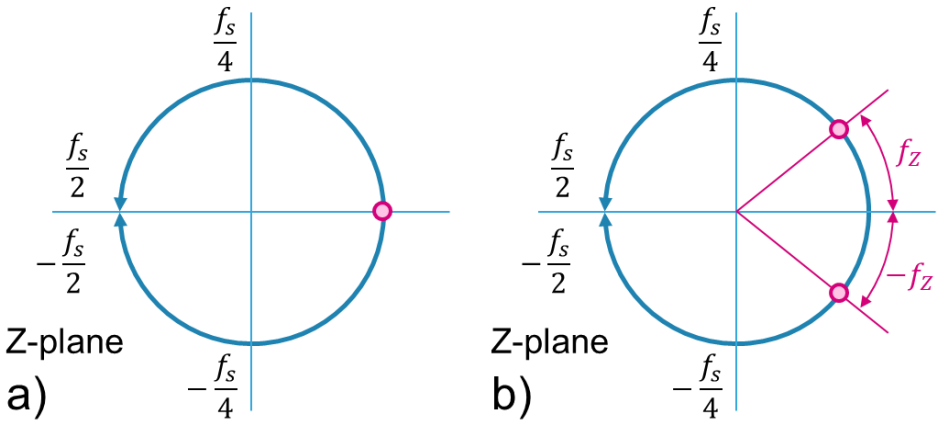

Figure 4-7: a) Zeros' location of a first order LPDTSDM. b) Zero locations of second order BPDTSDM

Figure 4-7-a gives the NTF zero location of the first order LPDTSDM seen just before. As already said, it is located at the angular frequency $\omega = 0$ corresponding to the point $Z = 1$ in the complex plan. To have a band pass modulator requires to move this zero at $\omega_Z = 2 \times \pi \times f_Z$ such that the noise attenuation now happens around f_Z. Since signals dealt with are real, their spectrums have symmetrical components at f and $-f$ with opposite phase. Hence, a second zero around $-f_Z$ must be added.

In general, it is not required to have it exactly at $-f_Z$, but that require the ability to create single complex poles. Modulators using such poles exist and are called quadrature modulators and are the complex extension of real modulators. In the current case, the investigation will be limited to real modulators using only real components. In this context complex poles can only be created by complex conjugate pairs. Therefore, band pass modulators are necessarily of an even order. The NTF from Figure 4-7-b can then be expressed by:

$$NTF(Z) = (1 - Z_Z \times Z^{-1}) \times (1 - \overline{Z_Z} \times Z^{-1}) = 1 - 2 \times Re[Z_Z] \times Z^{-1} + |Z_Z|^2 \times Z^{-2} \qquad (4.14)$$

Taking $Z_Z = e^{j \times \omega_Z \times T_S}$, (4.14) can be written as:

$$NTF(Z) = 1 - 2 \times \cos(\omega_Z \times T_S) \times Z^{-1} + Z^{-2} \qquad (4.15)$$

For a modulator with unit response STF for all frequencies, and injecting (4.15) into (4.6) gives:

$$Y(Z) = X(Z) + (1 - 2 \times \cos(\omega_Z \times T_S) \times Z^{-1} + Z^{-2}) \times E(Z) \qquad (4.16)$$

Adding $(-2 \times \cos(\omega_Z \times T_S) + Z^{-1}) \times Z^{-1} \times Y(Z)$ on both sides and dividing by the NTF gives:

$$Y(Z) = \frac{X(Z) - (2 \times \cos(\omega_Z \times T_S) - Z^{-1}) \times Z^{-1} \times Y(Z)}{1 - 2 \times \cos(\omega_Z \times T_S) \times Z^{-1} + Z^{-2}} + E(Z) \tag{4.17}$$

Equation (4.17) can be mapped to the schematic in Figure 4-8 with $K = 2 \times \cos(\omega_Z \times T_S)$.

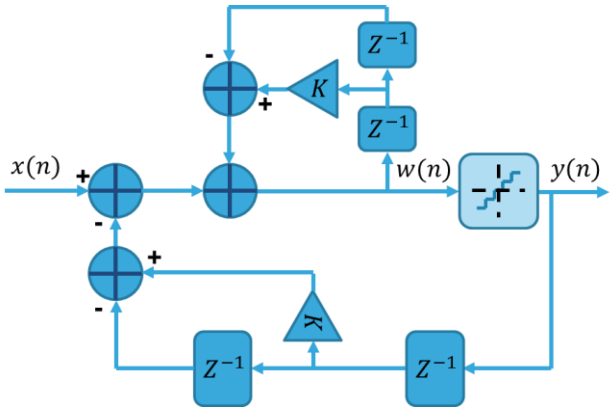

Figure 4-8: Simulation model of a second order BPDTSDM

The performances of this modulator are plotted in Figure 4-9 for a zero frequency $f_Z = f_s/7 \cong 73kHz$, a bandwidth of $1kHz$ and an $OSR = 256$.

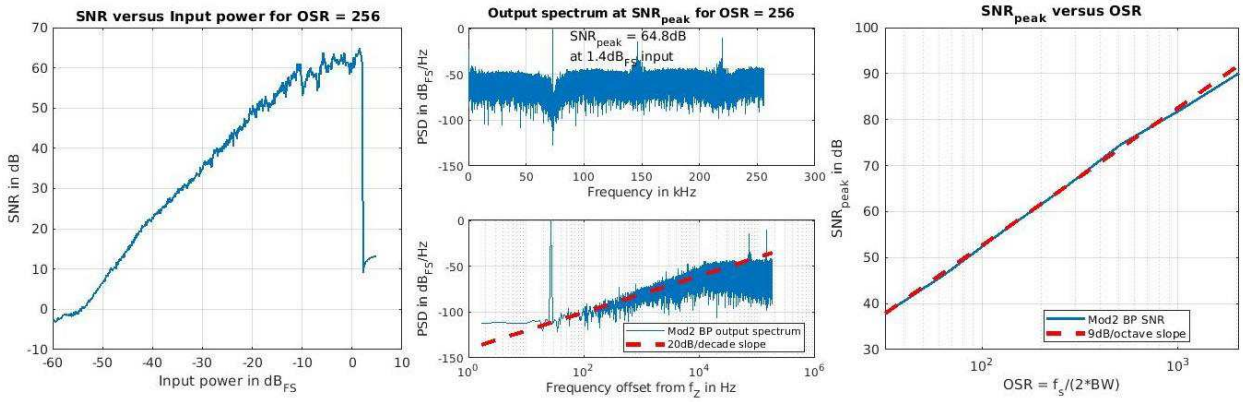

Figure 4-9: Simulation of a second order BPDTSDM. Left: SNR versus input power. Middle: Output spectrum at SNR_{peak}. Right: SNR_{peak} versus OSR

The overall behavior is similar to the first order low pass modulator. On the left graph, the SNR grows linearly with the input power for low power inputs. The curve gets an inflexion for high input powers but interestingly it breaks slightly beyond the full scale, around $1.4dB_{FS}$. The output spectrum at SNR_{max} in the middle graphs is plotted twice. First, using a linear frequency scale starting from DC on top. Here it can clearly be seen that the notch in the noise is at the desired frequency f_Z. Second, on a logarithmic frequency scale as frequency offset from the zero frequency f_Z. On this plot, the $20dB$ per decade noise slope is clearly visible, the same as the first order low pass modulator. Finally, on the right, the same SNR improvement of $9dB$ per octave of OSR is observed.

The conclusion is that band pass modulators have a noise shaping characteristic similar to a low pass modulator with half the order. Equation (4.13) is then adjusted to become (4.18).

$$SNR_{peak}(dB) = 10 \times \log_{10}\left(\frac{3}{2}\right) + 20 \times \log_{10}(2^N - 1) \\ +10 \times \log_{10}\left(\frac{L+1}{\pi^L}\right) + (L+1) \times 10 \times \log_{10}(OSR) \quad (4.18)$$

The band pass modulator simulation has a larger discrepancy compared to the low pass simulation. The peak SNR simulated is $64.8dB$ while (4.18) predicts 74.8dB. From the left graph of Figure 4-9 it is clear that the modulator's behavior becomes shaky for inputs above $-10dB_{FS}$, but if the curve were to be extended from data points below that, it would reach a value somewhere above $70d$B, close to equation (4.18) prediction.

The reasons for this discrepancy will be studied here. Only how it is affected by few factors, namely, the quantizer resolution and the nature of the input signal, will be looked at.

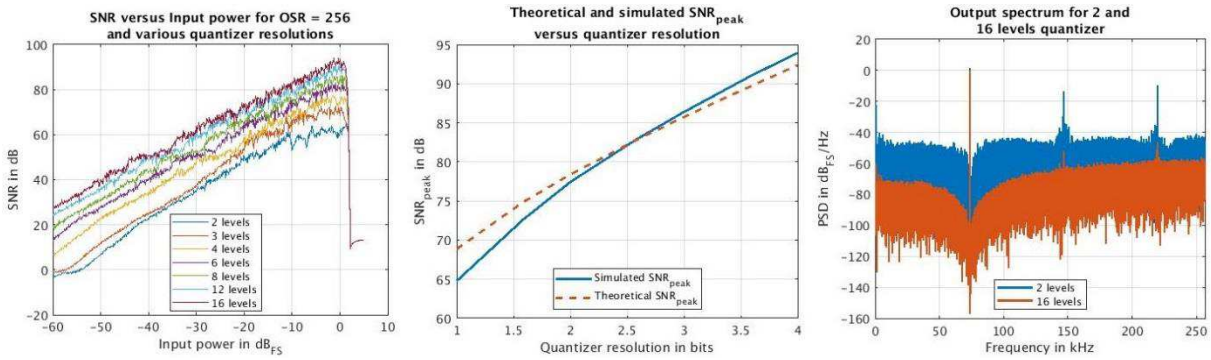

Figure 4-10: Simulation of a second order BPDTSDM for various quantizer resolution. Left: SNR versus input power. Middle: Theoretical and simulated peak SNR. Right: Output spectrum for a 2 and a 16 level quantizer

Let us start with the impact of the quantizer resolution. It is common in SDM to use a quantizer with a non-integer number of bits. It is then more convenient to describe them in term of the number of levels the output signal is coded on. The number of bits is then defined as the base two logarithm of the number of levels.

To see the impact of the quantizer resolution, simulations for various number of levels are performed. The results are plotted in Figure 4-10. On the left, the output SNR as a function of the input power is plotted. The more levels, the better the behavior, i.e. the SNR grows nearly linearly with the input power. When the simulated peak SNR and the theoretical one are compared (middle graph), the same behavior is observed. The right graph plots the output spectrum of a 2 and 16 level quantizer simulation. The increased resolution of the quantizer does not only reduce the noise in band but also out of band.

Intuitively, the performance loss is caused by an overloading of the quantizer by the shaped quantization noise itself. The overall noise power being reduced with more levels on the quantizer, this effect is reduced, and the ideal performances are restored.

While increasing the quantizer number of levels improves the modulator's behavior it is not compatible with the objective of efficient digital processing. The proposed study will be limited to a three level quantizer which already brings some benefits with respect to a two level quantizer. On top of that, it allows for a well-defined quantizer gain while maintaining a low complexity and the intrinsic linearity property of a two level quantizer.

So far, all the simulations provided were using a single tone input. Here, the modulator's behavior for multiple tones input will be studied. Figure 4-11 left graph plots the output SNR when the input signals are a one, two, six and ten tones signal. The X axis is the input power plus the input PAPR. This

corresponds to the input peak value. This way the modulator's performances break at the same point regardless of the input number of tons. While the peak SNR diminishes because of the higher PAPR of multiple tones inputs, the modulator's behavior is significantly improved, even with only two tones.

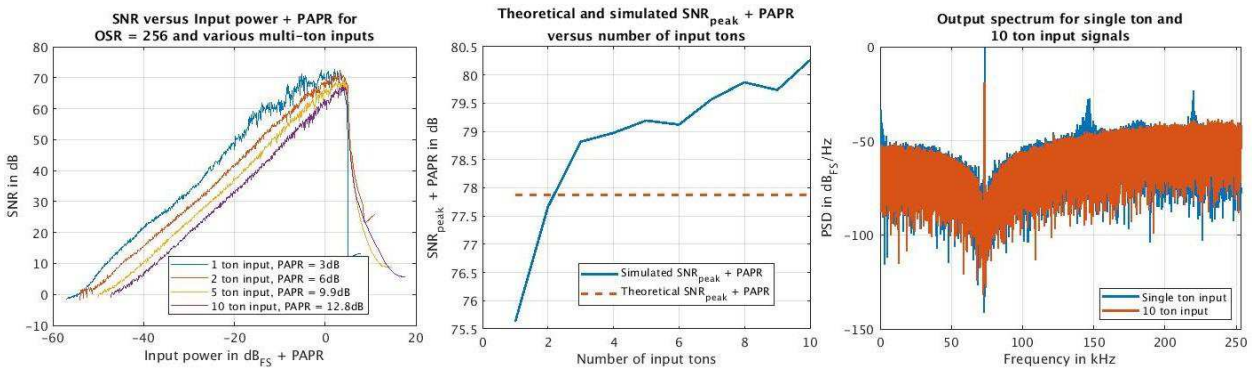

Figure 4-11: 3 level quantizer BPDTSDM. Left) SNR versus input power + PAPR for multiple tone inputs. Middle) Theoretical and simulated SNR +PAPR versus number of input tones. Right) SNR_{peak} spectrum for single and 10 tones inputs

The middle graph plots the theoretical and simulated peak SNR plus PAPR versus an increasing number of input tones. The theoretical peak SNR plus PAPR is constant since the input signal power is reduced by the PAPR such that the input peak value is at the modulator full scale. As the number of tones increases, the simulated peak SNR plus PAPR rapidly goes above the theoretical limit. This is possible because the used theory is too simple. It was assumed that the input signal could not go above the full scale of the ADC. While this is true at any time for Nyquist ADCs, for SDM, the signal can go above the full scale to some extent. This is due to their fundamentally different working principle. Intuitively the loop injects a feedback signal to the input node, trying to cancel the input signal. If the loop is successful, the signal information is contained into that feedback signal. If the input signal is too large, the limited amplitude of the feedback DAC output cannot cancel it, and the loop breaks down. But this does not happen instantly, if it is only for a short time and not too far out of the full scale, the loop may survive. This is exactly the case of high PAPR signals. While the PAPR of the multi-tones signal used here may differ from the one of an OFDM signal, it still gives the correct trend. Because the ADC SNR specification in section 3.2.7 was taking PAPR into account, it may be relaxed thanks to this feature.

Finally, the right graph displays the overlapped spectrums of a single tone and ten tones input. All the spurs in the shaped noise of the single tone input disappear for the multiple tone one. This reduces the overall power of the shaped quantization noise and prevent the premature overloading of the quantizer that degrades the performances. It is visible here that, for systems as the proposed one, intended for wide band signals, it is of prime importance to design the modulator using the appropriate input signals to avoid over designing, especially when target performances are already challenging.

4.2.1.3 Conclusion

While sigma-delta modulators can sometimes appear unintuitive, it has been shown that they can be properly described using the linear algebra of the Z-Transform. They also offer easy and efficient way of simulation. Finally, the first analysis carried out allows to reach some early conclusions. First, more level on the quantizer improves the modulator behavior. For the reasons discussed in section 4.1, a high number of levels cannot be used. A good compromise is to use a three level quantizer. Second SDM are inherently more robust to high PAPR signals. Third and last it is of prime importance to evaluate performances using input signals with the appropriate characteristics to avoid overdesign.

4.2.2 Continuous time modulators

The basic concept of Continuous Time Sigma-Delta Modulators (CTSDM) is very similar to their DT counterpart. It is a feedback system around a loop filter and a quantizer. The purpose is to have, in the band of interest, a unit Signal Transfer Function (STF) while attenuating the quantizer noise. Figure 4-12 depicts a Discrete Time Sigma-Delta Modulator (DTSDM) on the left and a CTSDM on the right. The major difference is that the input signal and the loop filter are continuous time entities. Hence a sample and hold is required ahead of the quantizer. Also, the DAC output must be pulse shaped to go back from the DT to the CT one.

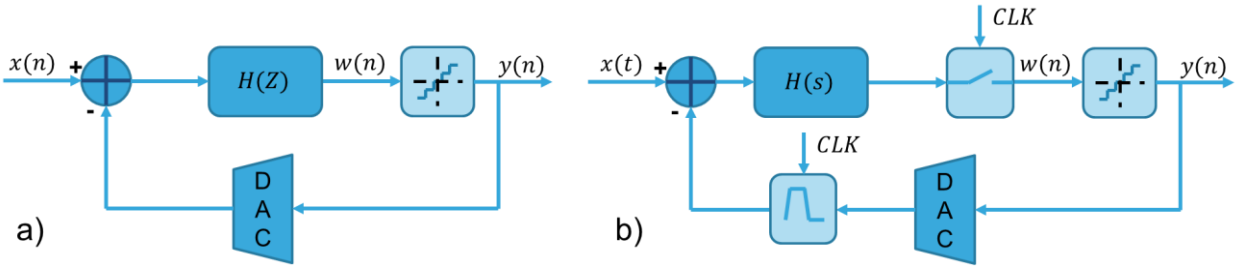

Figure 4-12: a) Discrete time sigma-delta modulator b) Continuous time sigma-delta modulator

Because of their mixed nature between discrete and continuous time, CTSDM are difficult to study directly. The most common technique, as it will be explained soon, is first to evaluate a discrete time equivalent of a CTSDM to study its NTF. Then, the model is completed with a CT signal path transfer function to study the modulators STF.

A classic design method is to start from a DTSDM and then to adjust the CT modulator such that its DT equivalent matches with the desired DTSDM. An efficient method to do so is called the "Impulse Invariant" design method.

One major difference in CTSDM is their sensitivity to Excess Loop Delay (ELD) in the feedback path. The quantizer and the DAC will require some time before they can settle their outputs. This added delay can have detrimental effects and need to be compensated.

Here these three points, modeling, design and ELD compensation, will be detailed.

4.2.2.1 CT modulator modeling

The full modeling of CTSDMs is obtained in two steps. First, a DT equivalent modulator is derived, allowing to study their NTF. Second, the modeling is completed with an additional continuous time signal path allowing the STF analysis.

4.2.2.1.1 DT equivalent modulator

The DT equivalent of CTSDM is obtain from a different, but mathematically equivalent, representation of the SDM. These alternate representations are given in Figure 4-13. They consist in moving the summing note at the input of the quantizer.

While these representations are mathematical equivalent to those of Figure 4-12, they are highly inadequate for implementation. Not only do they impose the duplication of the loop filter and the sample and hold but also, they expose the system to mismatch between these duplicates. These representations are only useful for analysis.

To have a discrete time equivalent requires two things. The first one, which is very obvious, is that the CT modulator is made into a DT system, i.e. a system having only DT inputs and outputs. The second is that this DT system behaves as the DT equivalent of interest. In the alternate representation of Figure 4-13-b, the part circled in pink has a discrete input and output. When looking at it as a black box, it is

indistinguishable from a discrete time system. This means it is a DT system. To find the discrete time equivalent only requires finding the DTSDM where its pink circle sub-part behaves in the same way.

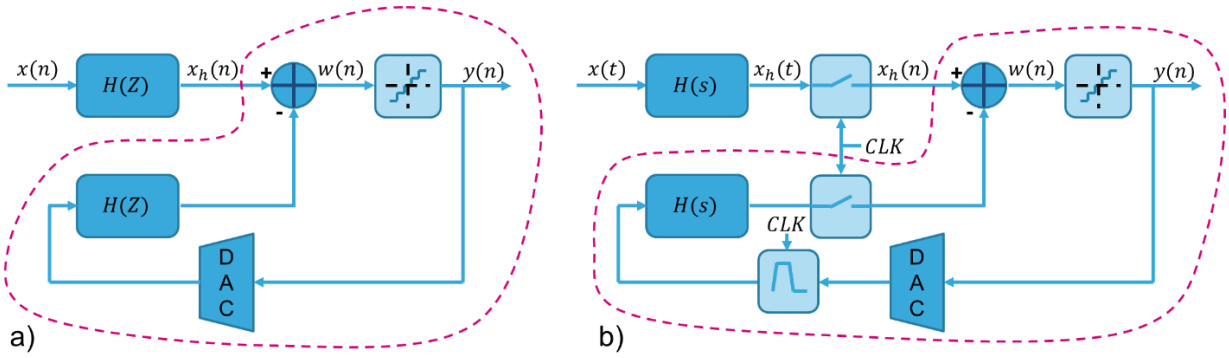

Figure 4-13: Alternate representation of a) Discrete time sigma-delta modulators b) Continuous time sigma-delta modulators

The important point to note here is that only the noise loop is fully included in this DT sub-part. Hence, the equivalent model will only be valid for the NTF, but not necessarily for the STF. In general, the STF will differ between a CTSDM and its DT equivalent. This will have an impact on the chosen design approach.

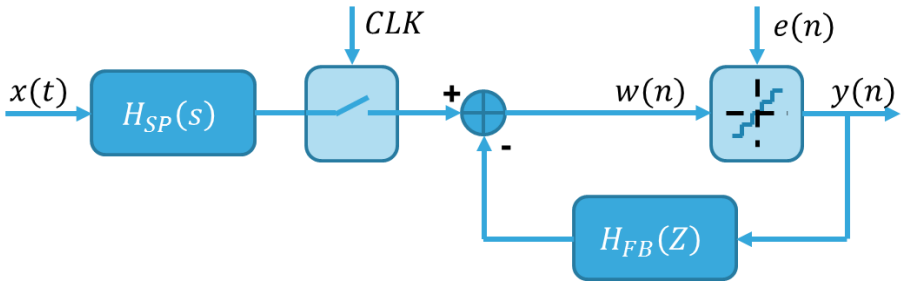

Figure 4-14: General representation of CTSDM

The analysis of the STF requires to consider the continuous time signal path. Using the DT equivalent a CTSDM can always be represented as a continuous time signal path $H_{SP}(s)$ and a discrete time feedback path $H_{FB}(Z)$ as described in Figure 4-14. The signal path is generally equal to the loop filter transfer function, but it can also include some additional signal conditioning. The feedback path corresponds to the DT equivalent modulator. The next step is to determine the NTF and STF.

4.2.2.1.2 NTF of a 1st order LPCTSDM

One method to determine $H_{FB}(Z)$ is to evaluate the CT impulse response from the quantizer output to the sample and hold input in Figure 4-13-b. Then to sample it and finally to apply the Z-transform to the sampled impulse response to obtain $H_{FB}(Z)$.

Let us treat the case of a first order low pass CTSDM. There are two things to determine, the pulse shape $p(t)$ coming after the DAC, and the loop filter $H(s)$. For the pulse shape, in this example, a Non-Return to Zero (NRZ) pulse, where the DAC output value is held constant between each sample, will be used. The pulse shaped output will then be a staircase function. For a sampling period T_S, $p(t)$ is described by:

$$p(t) = u(t) - u(t - T_S) \tag{4.19}$$

Where $u(t)$ is the heavy-side step function which has the value 0 for $t < 0$ and 1 for $t \geq 0$. Many other pulse shapes could have been used, the NRZ one is very common for low pass modulators and because it is simple, it will help making this example clearer.

The loop filter will be made of an ideal integrator, described by its Laplace transform in (4.20).

$$H(s) = \frac{1}{s} \tag{4.20}$$

The impulse response $h(t)$ of this integrator is the heavy-side step function as per (4.21):

$$h(t) = u(t) \tag{4.21}$$

The CT feedback impulse response $h_{FB}(t)$ is then given by the convolution of the two:

$$h_{FB}(t) = \int_{-\infty}^{+\infty} p(\tau) \times h(t - \tau) \times d\tau = \begin{cases} 0 & for\ t < 0 \\ t & for\ t \in [0, T_S[\\ T_S & for\ t > T_S \end{cases} \tag{4.22}$$

The next step is to sample $h_{FB}(t)$ with a sampling period T_S and to process its Z-Transform to get the equivalent $H(Z)$:

$$h_{FB}(n \times T_S) = T_S \times u\big((n-1) \times T_S\big) \tag{4.23}$$

$$H_{FB}(Z) = \sum_{n=-\infty}^{+\infty} h_{FB}(n \times T_S) \times Z^{-n} = T_S \times Z^{-1} \times \sum_{n=0}^{+\infty} Z^{-n} = T_S \times \frac{Z^{-1}}{1 - Z^{-1}} \tag{4.24}$$

Finally, the equivalent NTF can be processed:

$$NTF(Z) = \frac{1}{1 + H_{FB}(Z)} = \frac{1 - Z^{-1}}{1 - (1 - T_S) \times Z^{-1}} \tag{4.25}$$

Using a sampling period $T_S = 1$ gives back the first order LPDTSDM NTF from equation (4.11) in section 4.2.1.1. For a smaller sampling period the NTF will exhibit, on top of its zero, an integrating behavior. This gains up the noise in the band which is undesirable. The solution is to add a gain of $1/T_S$ in the continuous time feedback path. One way of doing it is by scaling the feedback pulse by $1/T_S$. This is equivalent to a sampling period normalization. (4.19) then becomes (4.26), (4.22) becomes (4.27), (4.23) becomes (4.28) and (4.24) becomes (4.29).

Then the NTF of the LPCTSDM is the same as the as the LPDTSDM previously studied, for any sampling period. One point to note here is that the original T_S factor comes from the sampling period, but it is compensated by a gain in the feedback loop. Because these two values have different physical origins, they will experience some mismatch. The impact of this mismatch can be predicted from (4.25). The NTF will experience some remnant integrating behavior limiting the noise attenuation in the band. Since this sampling period normalization does not exist in DT modulators, they cannot be impaired with such a phenomenon. This is a first difference between the DT and CT modulators. Other than that, the noise shaping ability of the CT will be exactly the same as the DT and need not to be re-simulated.

$$p(t) = \frac{1}{T_S} \times \big(u(t) - u(t - T_S)\big) \tag{4.26}$$

$$\begin{cases} 0 & for\ t < 0 \\ \dfrac{t}{T_S} & for\ t \in [0, T_S[\\ 1 & for\ t > T_S \end{cases} \tag{4.27}$$

$$h_{FB}(n) = \begin{cases} 0\ for\ n \leq 0 \\ 1\ for\ n > 0 \end{cases} \tag{4.28}$$

$$H(Z) = \frac{Z^{-1}}{1 - Z^{-1}} \tag{4.29}$$

4.2.2.1.3 STF of a 1st order LPCTSDM

The STF of a CTSDM has two main uses. The first one is to evaluate the in-band transfer function and second one is to evaluate the intrinsic anti-aliasing the CTSDM provides. To that purpose, the STF is usually represented over multiple Nyquist zones. While this is not fully rigorous since the output signal ends up folded on a single Nyquist zone, it gives the attenuation a given frequency will undergo before being folded.

From Figure 4-14, the input signal must go through $H_{SP}(s)$, then is sampled, and is finally fed to the quantizer, at which point it will experience the NTF. The effect of the signal path is entirely described by its transfer function $H_{SP}(s)$ and need no additional precision. Then comes the sampling operation. It has the effect of periodizing the signal spectrum with a frequency period equal to the sampling frequency $f_S = 1/T_S$. For an in-band signal on this low pass CTSDM, this is basically transparent and the STF is given by:

$$STF(f) = H_{SP}(s = j \times \omega) \times NTF\big(Z = e^{j \times \omega \times T_S}\big) = \frac{1 - e^{j \times \omega \times T_S}}{j \times \omega} \tag{4.30}$$

This equation is also valid for the improper STF use outside the first Nyquist zone for anti-aliasing evaluation. This can be done thanks to the periodic nature of the Z variable when going around the unit circle. When a signal in a higher Nyquist zone is at the modulator's input, first, it will be attenuated by the signal path integration, then folded on the first Nyquist zone by the sample and hold, and finally, it will be affected by the NTF. This is the valid interpretation of (4.30) for signals outside of the first Nyquist zone.

Figure 4-15 plots the main characteristics of the first order LPCTSDM with an OSR of 256 and a bandwidth of $1kHz$. The left top graph plots the gain of the STF in the band. The loop filter is an integrator with a $-20dB/Decade$ slope and the NTF is a differentiator with $20dB/Decade$ slope. The two of them compensate each other to give a flat STF.

This flat gain can be affected for low OSR values. Since the NTF is a discrete time transfer function, the $20dB/Decade$ slope is not held for all frequencies, it flattens in the vicinity of $f_s/2$, hence affecting the STF gain.

Figure 4-15 left bottom graph plots the STF in band phase. It is clearly linear, which is a very good property since it corresponds to a constant group delay in the band. This means that all the in-band frequencies will experience the same constant delay, hence preventing signal distortion.

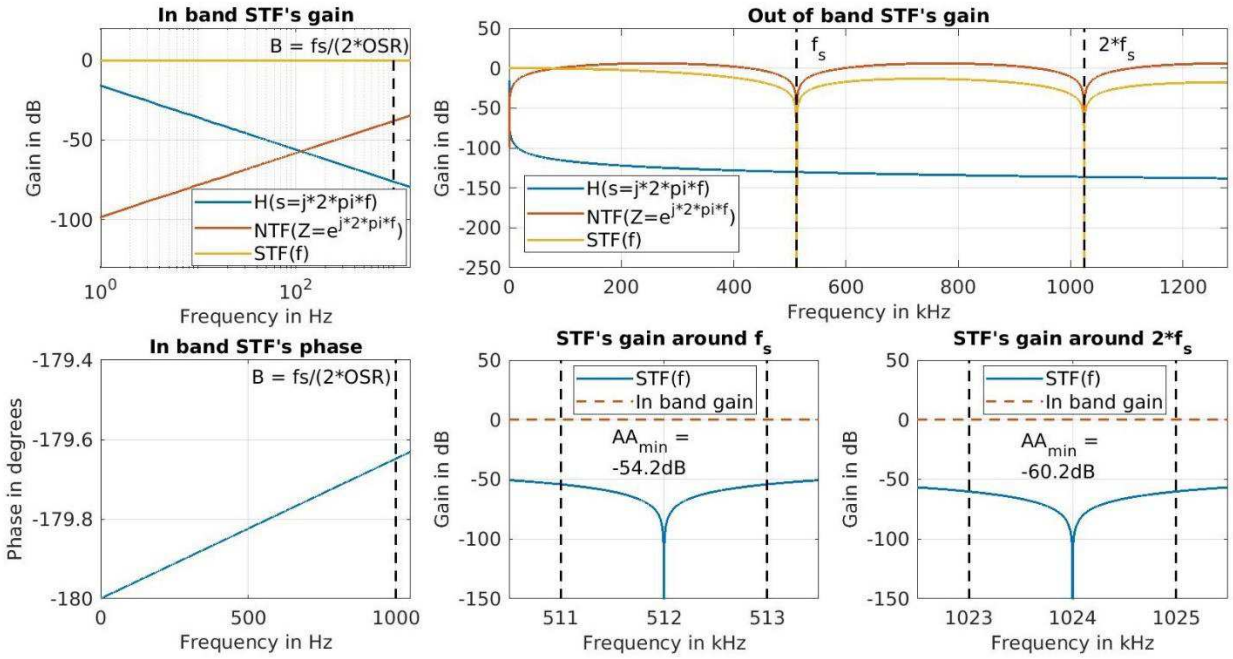

Figure 4-15: Signal Transfer Function characteristics of a first order LPCTSDM with an OSR of 256 and a bandwidth of $1kHz$

The top right graph plots the loop filter, the NTF and the STF over the five first Nyquist zones. The frequencies outside of the first Nyquist zone that will fold into the band are only the ones around integer multiple of the sampling frequency. While the loop filter gain was compensating the NTF zero in the first Nyquist zone, it only brings more attenuation in the following ones. Then the periodic zeros of the NTF translate into deep notches in the STF around integer multiples of the sampling frequency. The intrinsic anti-aliasing of CT modulators comes from the accumulations of these attenuations. The bottom middle and right graphs plot respectively a zoom of the STF around f_s and $2 \times f_s$. The lowest anti-aliasing filtering in the band are respectively $-54.2dB$ and $-60.2dB$. This $6dB$ difference correspond the $-6dB/Octave$ slope of the integrator going from f_s to $2 \times f_s$.

Here, two conclusions can be reached. Higher order modulators, since having a loop filter with a steeper slope, will have better anti-aliasing. Larger OSR leads to better anti-aliasing properties since the NTF periodic zeros will happen at higher frequencies where the loop filter has more attenuation.

4.2.2.1.4 Conclusion on CTSDM modeling

The overall mechanisms between DT and CT modulators are equivalent. In that regard, it is interesting to have a good understanding of DT modulators since they are easier to put into equation and to simulate. In particular, to any CTSDM, a DT equivalent can be associated that allows the analysis of the NTF. The complete modeling is obtained by adding a CT signal path. This allows to study CTSDM's STF. In particular, this allows to understand the well-known anti-aliasing property of CT modulators.

4.2.2.2 "Impulse invariant" design method

The design of CTSDM using the "Impulse invariant" method has been used for a very long time. Here, a version inspired by [4-6] will be presented. It is more specific to gmLC based BPCTSDM. It allows to set the NTF of the CTSDM to any NTF of the same order. One limitation is that is provides no design control of the STF. This design method consists in five steps:

1. Design of a discrete time modulator and evaluation of its feedback transfer function
2. Choose a CT modulator architecture with parametric feedback and/or feedforward coefficients
3. Extract the feedback impulse response and sample it

4. Process the Z-Transform of the sampled feedback impulse response
5. Adjust the feedback and/or feedforward coefficients to match the feedback transfer function of the DT modulator with the desired NTF.

Let us apply this method to the design of a second order Band Pass CTSDM.

1. *BPDTSDM design and impulse response evaluation*

Starting from the DT modulator from section 4.2.1.2, the feedback path transfer function can easily be found by setting $X(Z)$ and $E(Z)$ to zero in (4.17).

$$H_{FB}(Z) = \frac{2 \times \cos(\omega_Z \times T_S) \times Z^{-1} - Z^{-2}}{1 - 2 \times \cos(\omega_Z \times T_S) \times Z^{-1} + Z^{-2}} \tag{4.31}$$

2. *BPCTSDM architecture*

CTSDM architectures are very diverse. They can use different combinations of feedback and feedforward path, different nature of loop filters or DAC pulse shapes [4-5]. The target application requiring a high bandwidth, and therefore a high sampling frequency already imposes the CT nature of the SDM. Band pass modulators use resonators to build their loop filter. They can be of two natures, integrator based or gmLC based resonator. Again, the kind of bandwidth required imposes the use of gmLC based resonators. Feedforward path can be challenging to implement in integrated circuits when using LC resonators. The sheer physical size of the inductances implies long wire connections for the feedforward paths whose behavior cannot be overlooked. The strategy is to keep the architecture as simple as possible to ease the challenge of the very high bandwidth. Hence, the choice was made to use feedback only architecture. For reasons that will be made clear later, a Return to Zero (RZ) DAC pulse shape will be used, with the previous choice of a three level quantizer.

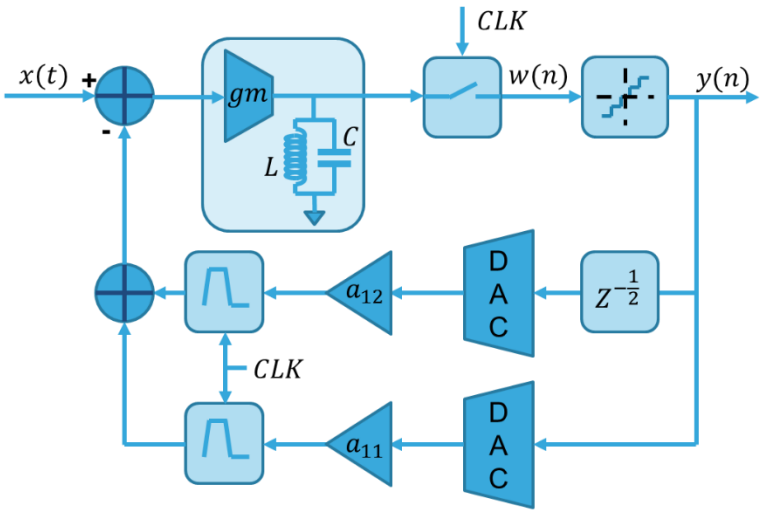

Figure 4-16: BPCTSDM parametric architecture

To control the DT feedback impulse response of a CTSDM it can be shown that it requires a number of free parameters equal to the modulator's order. In this architecture there is only one summing node since the second order is implemented using a single resonator. A single feedback, with coefficient a_{11}, on this node would not be enough. A solution to this problem was proposed by Shoaei and Snelgrove in [4-6]. It consists in adding a second path with a different DAC pulse shape and coefficient a_{12}. A classic alternate pulse shape is a Half delayed Return to Zero (HRZ) pulse. The final parametric architecture is depicted in Figure 4-16.

3. *Extraction of the sampled feedback impulse response*

Let us first determine the CT feedback impulse response. It will be made of the summation of the two feedback DAC pulses convolves with the impulse response of the gmLC resonator. The feedback pulses are described by (4.32) and (4.33).

$$p_{a11}(t) = a_{11} \times \left(u(t) - u\left(t - \frac{T_S}{2} \right) \right) \tag{4.32}$$

$$p_{a12}(t) = a_{12} \times \left(u\left(t - \frac{T_S}{2} \right) - u(t - T_S) \right) \tag{4.33}$$

The transfer function of the gmLC resonator is the product of the LC tank Z_{LC} impedance by gm. The impedance Z_{LC} can be obtained by setting $R = +\infty$ in the Z_{RLC} expression from ().

$$H_{gmLC}(s) = \frac{\frac{gm}{C} \times s}{s^2 + \frac{1}{LC}} = \frac{\frac{gm}{C} \times s}{s^2 + \omega_Z^2} \tag{4.34}$$

To ease the inverse Laplace transformation, it is re-written as a sum of quotients with first order polynomial denominators.

$$H_{gmLC}(s) = \frac{\frac{gm}{2 \times C}}{s + j \times \omega_Z} + \frac{\frac{gm}{2 \times C}}{s - j \times \omega_Z} \tag{4.35}$$

Using the known Laplace Transform $\mathcal{L}\{u(t) \times e^{a \times t}\} = 1/(s - a)$ and the linearity property of this transformation, the resonator's impulse response is obtained.

$$h_{gmLC}(t) = u(t) \times \frac{gm}{2 \times C} \times \left(e^{-j \times \omega_Z \times t} + e^{j \times \omega_Z \times t} \right) = u(t) \times \frac{gm}{C} \times \cos(\omega_Z \times t) \tag{4.36}$$

The last step is to convolve this impulse response with the DAC pulses.

$$\begin{aligned} h_{FB}(t) &= \int_{-\infty}^{+\infty} \left(p_{a11}(\tau) + p_{a12}(\tau) \right) \times h_{gmLC}(t - \tau) \times d\tau \\ &= \begin{cases} 0 & for\ t < 0 \\ \dfrac{gm}{\omega_Z \times C} \times a_{11} \times \sin(\omega_Z \times t) & for\ 0 \le t < \dfrac{T_S}{2} \\ \dfrac{gm}{\omega_Z \times C} \times \begin{pmatrix} a_{11} \times 2 \times \sin\left(\omega_Z \times \dfrac{T_S}{4} \right) \times \cos\left(\omega_Z \times \left(t - \dfrac{T_S}{4} \right) \right) \\ + a_{12} \times \sin\left(\omega_Z \times \left(t - \dfrac{T_S}{2} \right) \right) \end{pmatrix} & for\ \dfrac{T_S}{2} \le t < T_S \\ \dfrac{gm}{\omega_Z \times C} \times \begin{pmatrix} a_{11} \times 2 \times \sin\left(\omega_Z \times \dfrac{T_S}{4} \right) \times \cos\left(\omega_Z \times \left(t - \dfrac{T_S}{4} \right) \right) \\ + a_{12} \times 2 \times \sin\left(\omega_Z \times \dfrac{T_S}{4} \right) \times \cos\left(\omega_Z \times \left(t - 3 \times \dfrac{T_S}{4} \right) \right) \end{pmatrix} & for\ t > T_S \end{cases} \end{aligned} \tag{4.37}$$

Finally, to get the DT version of this feedback impulse response, it needs to be sampled, i.e. t is set to $n \times T_S$. For the sake of brevity, the continuous time and the discrete time functions notations will only be differentiated through their variables t or n; i.e. when writing $h(n)$ it formally means $h(n \times T_S)$.

$$h_{FB}(n) = \frac{gm}{\omega_Z \times C} \times \left(a_{11} \times h_{RZ}(n) + a_{12} \times h_{HRZ}(n)\right) \times u(n-1) \tag{4.38}$$

With:

$$h_{RZ}(n) = 2 \times \sin\left(\omega_Z \times \frac{T_S}{4}\right) \times \cos\left(\omega_Z \times \left(n - \frac{1}{4}\right) \times T_S\right) \times u(n-1) \tag{4.39}$$

$$h_{HRZ}(n) = 2 \times \sin\left(\omega_Z \times \frac{T_S}{4}\right) \times \cos\left(\omega_Z \times \left(n - \frac{3}{4}\right) \times T_S\right) \times u(n-1) \tag{4.40}$$

4. *Processing of the sampled feedback impulse response Z-Transform*

Because the Z-Transformation is linear, $h_{RZ}(n)$ and $h_{HRZ}(n)$ can be processed separately. Their transfer functions are given in (4.41) and (4.42).

$$H_{RZ}(Z) = \frac{\left(\sin(\omega_Z \times T_S) - \sin\left(\omega_Z \times \frac{T_S}{2}\right)\right) \times Z^{-1} - \sin\left(\omega_Z \times \frac{T_S}{2}\right) \times Z^{-2}}{1 - 2 \times \cos(\omega_Z \times T_S) \times Z^{-1} + Z^{-2}} \tag{4.41}$$

$$H_{HRZ}(Z) = \frac{\sin\left(\omega_Z \times \frac{T_S}{2}\right) \times Z^{-1} - \left(\sin(\omega_Z \times T_S) - \sin\left(\omega_Z \times \frac{T_S}{2}\right)\right) \times Z^{-2}}{1 - 2 \times \cos(\omega_Z \times T_S) \times Z^{-1} + Z^{-2}} \tag{4.42}$$

5. *Feedback coefficient evaluation*

Individually neither (4.41) nor (4.42) match the desired transfer function from (4.31) but this can be achieved with a linear combination of the two. What is needed it to find the values for a_{11} and a_{12} that provides the desired feedback transfer function. This translates into (4.43):

$$\frac{gm}{\omega_Z \times C} \times \left(a_{11} \times H_{RZ}(Z) + a_{12} \times H_{HRZ}(Z)\right) = \frac{2 \times \cos(\omega_Z \times T_S) \times Z^{-1} - Z^{-2}}{1 - 2 \times \cos(\omega_Z \times T_S) \times Z^{-1} + Z^{-2}} \tag{4.43}$$

Both sides have the same denominators. Only matching the numerators polynomials is need. This gives the equation system from (4.44):

$$\begin{cases} a_{11} \times \left(\sin(\omega_Z \times T_S) - \sin\left(\omega_Z \times \frac{T_S}{2}\right)\right) + a_{12} \times \sin\left(\omega_Z \times \frac{T_S}{2}\right) = \frac{\omega_Z \times C \times 2 \times \cos(\omega_Z \times T}{gm} \\ a_{11} \times \sin\left(\omega_Z \times \frac{T_S}{2}\right) + a_{12} \times \left(\sin(\omega_Z \times T_S) - \sin\left(\omega_Z \times \frac{T_S}{2}\right)\right) = \frac{\omega_Z \times C}{gm} \end{cases} \tag{4.44}$$

The solutions are given by (4.45) and (4.46).

$$a_{11} = \frac{\omega_Z \times C}{gm} \times \frac{\cos(\omega_Z \times T_S) \times \left(2 \times \cos\left(\omega_Z \times \frac{T_S}{2}\right) - 1\right)^2 - \cos\left(\omega_Z \times \frac{T_S}{2}\right) + \frac{1}{2}}{\sin(\omega_Z \times T_S) \times \left(2 \times \cos\left(\omega_Z \times \frac{T_S}{2}\right) - 1\right) \times \left(\cos\left(\omega_Z \times \frac{T_S}{2}\right) - 1\right)} \tag{4.45}$$

$$a_{12} = \frac{\omega_Z \times C}{gm} \times \frac{\cos\left(\omega_Z \times \frac{T_S}{2}\right) - \cos(\omega_Z \times T_S) - \frac{1}{2}}{\sin(\omega_Z \times T_S) \times \left(\cos\left(\omega_Z \times \frac{T_S}{2}\right) - 1\right)} \tag{4.46}$$

It is hard to interpret this result directly. Visually this corresponds to the fact that the DT impulse response of the target modulator will match the CT impulse response at the sampling instants. To observe it, it is first necessary to obtain the CT impulse response. This is done by plugging (4.45) and

(4.46) into (4.37). Then it is required to process the DT feedback impulse response. To do so, (4.31) is re-written into (4.47):

$$H_{FB}(Z) = -1 + \frac{A}{1 - e^{j\times\omega_Z\times T_S} \times Z^{-1}} - \frac{B}{1 - e^{-j\times\omega_Z\times T_S} \times Z^{-1}} \tag{4.47}$$

With A and B constant terms equal to:

$$A = \frac{e^{j\times\omega_Z\times T_S}}{2 \times j \times \sin(\omega_Z \times T_S)} \quad and \quad B = \frac{e^{-j\times\omega_Z\times T_S}}{2 \times j \times \sin(\omega_Z \times T_S)} \tag{4.48}$$

The impulse response is the sum of the individual impulse response of each of the three terms in (4.47). The first term is simply a negative Dirac pulse and the second and third are given by (4.49) and (4.50).

$$h_A(n) = \mathcal{Z}^{-1}\left\{\frac{A}{1 - e^{j\times\omega_Z\times T_S} \times Z^{-1}}\right\} = \frac{e^{j\times\omega_Z\times(n+1)\times T_S}}{2 \times j \times \sin(\omega_Z \times T_S)} \times u(n) \tag{4.49}$$

$$h_B(n) = \mathcal{Z}^{-1}\left\{-\frac{B}{1 - e^{-j\times\omega_Z\times T_S} \times Z^{-1}}\right\} = -\frac{e^{-j\times\omega_Z\times(n+1)\times T_S}}{2 \times j \times \sin(\omega_Z \times T_S)} \times u(n) \tag{4.50}$$

The impulse response sum of the two last terms is simplify in (4.51).

$$\begin{aligned} h_A(n) + h_B(n) &= \frac{e^{j\times\omega_Z\times(n+1)\times T_S} - e^{-j\times\omega_Z\times(n+1)\times T_S}}{2 \times j \times \sin(\omega_Z \times T_S)} \times u(n) \\ &= \frac{\sin(\omega_Z \times (n+1) \times T_S)}{\sin(\omega_Z \times T_S)} \times u(n) \end{aligned} \tag{4.51}$$

Finally, adding the negative Dirac pulse of the first term correspond to subtracting one at $n = 0$. Equation (4.51) evaluates to 1 for $n = 0$. Subtracting one sets the first term of the impulse response to zero. The feedback impulse response is then expressed by (4.52).

$$h_{FB_{DT}}(n) = \frac{\sin(\omega_Z \times (n+1) \times T_S)}{\sin(\omega_Z \times T_S)} \times u(n-1) \tag{4.52}$$

Note here that the first term, for $n = 0$, is zero. This is in agreement with the "no delay free loop" rule in discrete time systems.

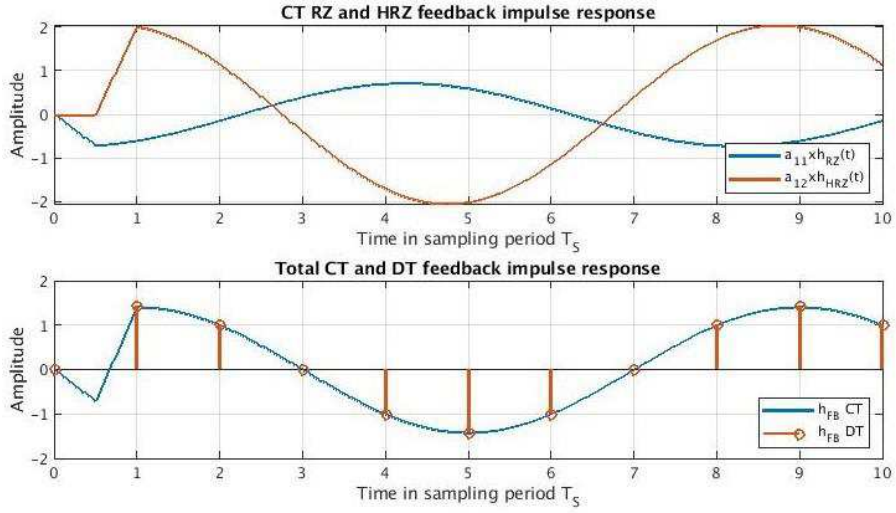

Figure 4-17: CT and DT feedback impulse response for $f_Z = f_s/8$

Figure 4-17 top graph plots the weighted impulse response of the RZ and HRZ feedback paths. The bottom graph plots the CT and DT total feedback impulse response. As expected, the CT impulse response matches the DT one at the sampling instants.

This method is very precise and provides the mathematical existence proof of a set of feedback coefficients allowing to match the desired NTF. It is also very involved. Here, the solution for the simplest BPCTSDM is derived, and the various analytical expressions are already too long for direct interpretation. In the following sections, a more numerical approach will be preferred where the feedback coefficients are obtained using some optimization algorithm, such as gradient descent, to match CT and DT impulse responses.

4.2.2.3 Excess Loop Delay (ELD)

One of the hypotheses in the previous analysis is overly idealistic. It is assumed that the quantizer and the DAC can deliver their outputs instantly. While this may be an acceptable assumption when dealing with low sampling frequencies, it is certainly invalid for the kind of frequencies investigated here. To understand the consequences of such a delay let us look at the impulse respond when it is present. Clearly delaying the feedback impulse response will lead to a sampling at a different point, altering the DT impulse response and the NTF. The question is to know if, for a given value of ELD, a different set of feedback coefficient can restore the desired NTF. One approach is to choose two points of the desired DT impulse response and adjust the feedback coefficients to match the CT impulse response on these 2 points. At the origin of time, for $t = 0$ and $n = 0$, the DT FB impulse response and the CT RZ and HRZ impulse response are all zero. This point will match for any value of the feedback coefficients. This means values of $n > 0$ must be used.

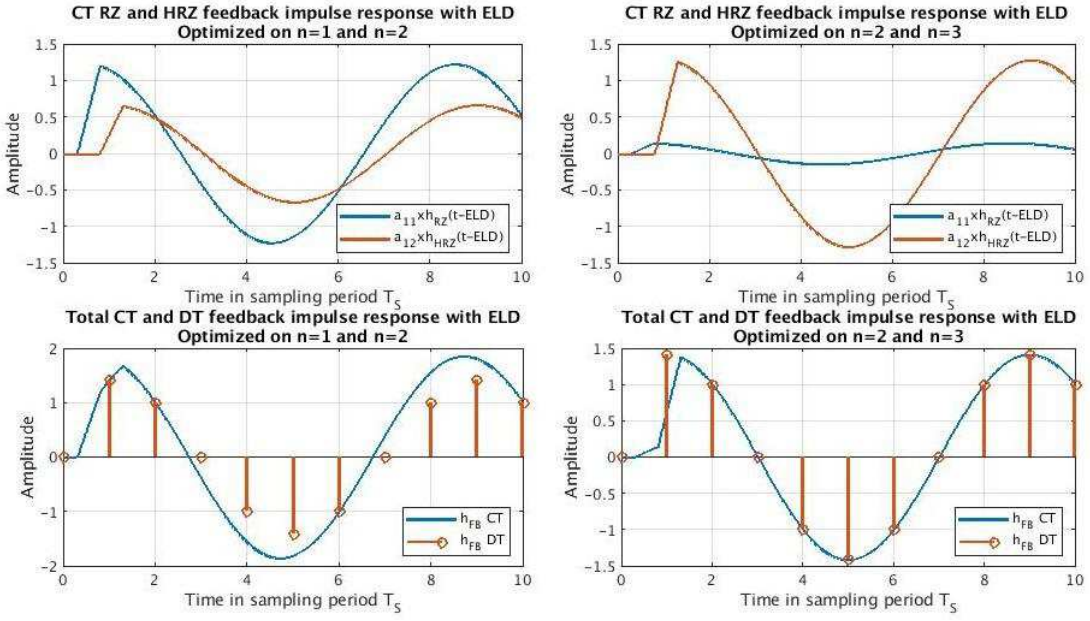

Figure 4-18: Impact of $ELD = 0.3 \times T_S$ on the CT impulse responses for $f_Z = f_s/8$. Left) Optimization done using $n = 1$ and $n = 2$. Right) Optimization done using $n = 2$ and $n = 3$.

Figure 4-18 plots the impulse responses for an ELD of $0.3 \times T_S$. The left plots are optimized using the DT point for $n = 1$ and $n = 2$, and the right plots are optimized using the DT point for $n = 2$ and $n = 3$. The first observation is that the feedback coefficients are widely different. In the first case, after the two optimized points, the DT and CT impulse response differ for most of the points. In the second case only the sample for $n = 1$ differs. This second option is much closer to the desired impulse response, but what is the impact of the remaining difference? The Z-transform of this impulse response can be obtained by a modification of (4.31).

$$H_{FB}(Z) = \frac{2 \times \cos(\omega_Z \times T_S) \times Z^{-1} - Z^{-2}}{1 - 2 \times \cos(\omega_Z \times T_S) \times Z^{-1} + Z^{-2}} + \left(h_{FB_{CT}}(T_S) - h_{FB_{DT}}(1)\right) \times Z^{-1} \quad (4.53)$$

Figure 4-19 plots the original NTF and the NTF affected by $0.3 \times T_S$ of ELD, optimized on $n = 2$ and $n = 3$ samples. One good point is that the zero is preserved. Unfortunately, the transfer function presents a huge gain around $125kHz$. Even though this is outside the band of interest this would have significant adverse effects, in particular quantizer overloading. In practice such an NTF is impractical and is considered unstable.

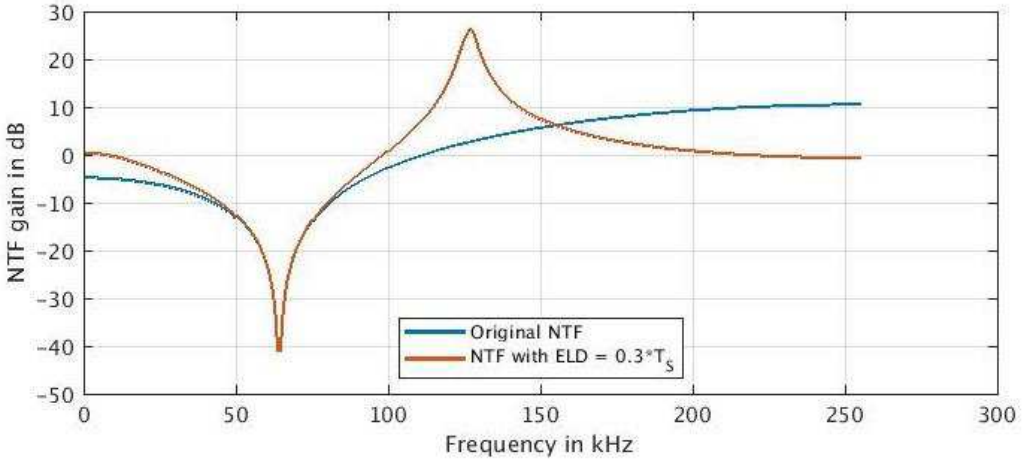

Figure 4-19: ELD impact of the NTF

The classic solution to this problem is to perform ELD compensation by the mean of an additional feedback path around the quantizer. The Architecture from Figure 4-16 must be modified as per Figure 4-20.

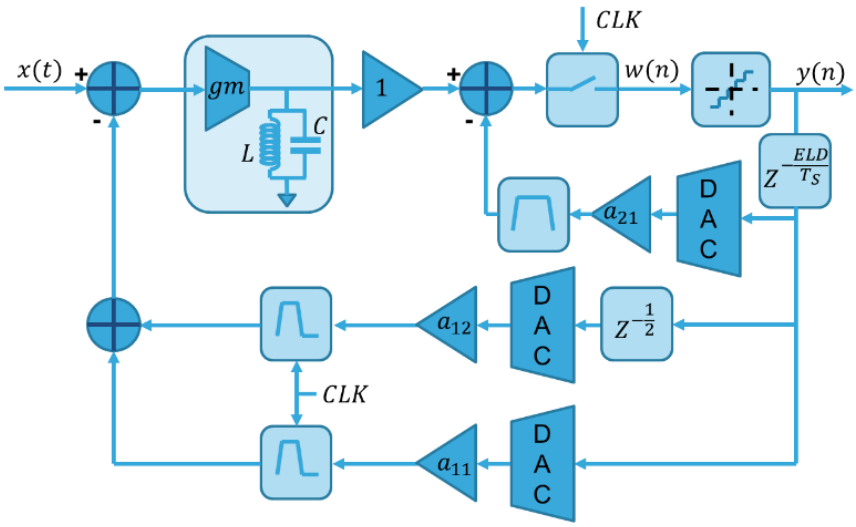

Figure 4-20: Modified BPCTSDM architecture for ELD compensation

Figure 4-21 plots the results with the ELD compensation. One can see the restoration of the DT impulse response. The price for this compensation is one additional feedback path and an additional buffer between the resonator's output and the ELD compensation feedback path. This buffer is required to isolate the resonator from the summing node. Otherwise, the compensation pulse would be affected by the resonator which would affect the remaining coefficients.

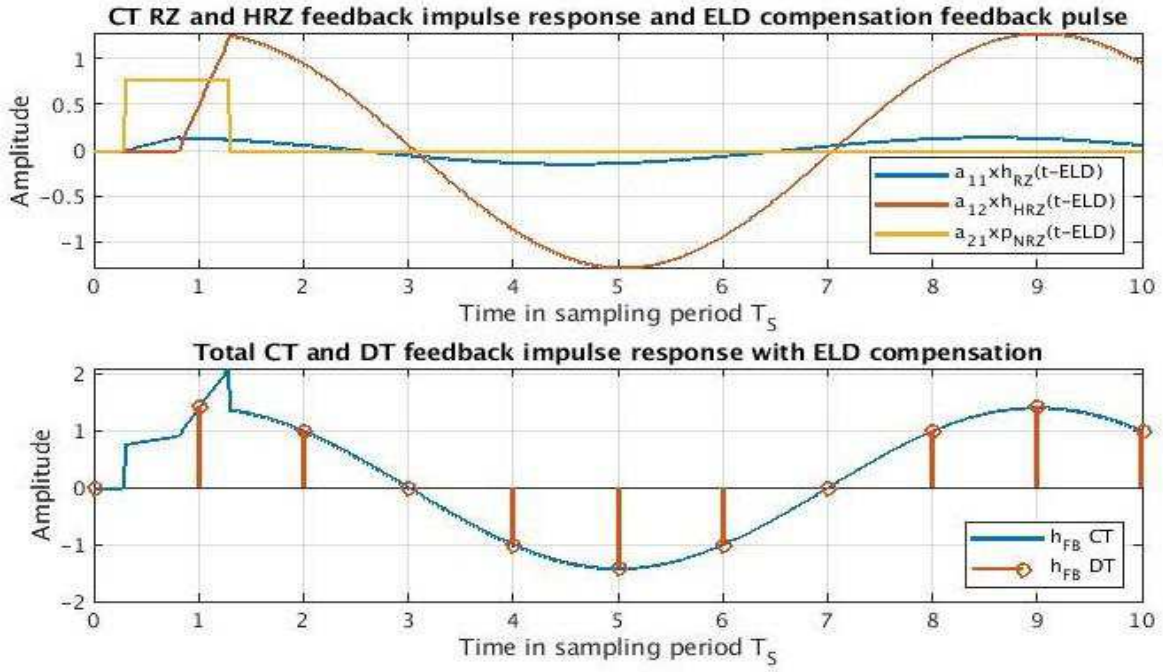

Figure 4-21: Feedback impulse response with ELD compensation

4.2.2.4 Conclusion

In this section, multiple aspects of continuous time sigma-delta modulators were covered. Going through their modeling and simulation, their intrinsic anti-aliasing was highlighted. Classic design and ELD effects were described and explained. Finally, an ELD compensation method was presented. It has the disadvantage of requiring addition components. In the next sections, it will be seen that for some specific configurations these extra components are not always necessary.

4.2.3 "$f_s/4$" Modulators

It was shown that for a sigma-delta loop to be functional the sampled continuous time feedback impulse response must match the desired discrete time one. One necessary condition on the feedback timing is that the loop must be closed before the first non-zero term of the DT impulse response. As the sampling frequency increases, this condition gets more and more challenging. In general, the first non-zero term is the second one for $n = 1$. In some configurations [4-7], when the center frequency is at one quarter of the sampling frequency, this term can be zero allowing for an additional clock cycle to close the loop.

$$H_{FB_{DT}}(Z) = \frac{2 \times \cos(\omega_Z \times T_S) \times Z^{-1} - Z^{-2}}{1 - 2 \times \cos(\omega_Z \times T_S) \times Z^{-1} + Z^{-2}} = \frac{-Z^{-2}}{1 + Z^{-2}} \tag{4.54}$$

Equation (4.54) gives the feedback transfer function of a second order band pass modulator at $f_s/4$. The numerator has only terms in Z^{-2}. This can be interpreted as a two-samples delay of the impulse response. Therefore, the first two terms for $n = 0$ and $n = 1$ will be zero.

For modulators of higher order, it is common to optimize the zeros' location in the band to maximize the SNR. Also, in practical implementation the quality factor of the resonators will not be infinite. This translates into a displacement of the zeros inward to the Z-plan's origin. Equation (4.55) gives the feedback transfer function of a fourth order band pass modulator.

$$H_{FB_{DT}}(Z) = \frac{\begin{pmatrix} 2\times\left(\left|Z_{Z_1}\right|\times\cos\left(\omega_{Z_1}\times T_S\right)+\left|Z_{Z_2}\right|\times\cos\left(\omega_{Z_2}\times T_S\right)\right)\times Z^{-1} \\ -2\times\left(1-2\times\left|Z_{Z_1}\right|\times\cos\left(\omega_{Z_1}\times T_S\right)\times\left|Z_{Z_2}\right|\times\cos\left(\omega_{Z_2}\times T_S\right)\right)\times Z^{-2} \\ 2\times\left(\left|Z_{Z_1}\right|\times\left|Z_{Z_2}\right|^2\times\cos\left(\omega_{Z_1}\times T_S\right)+\left|Z_{Z_2}\right|\times\left|Z_{Z_1}\right|^2\times\cos\left(\omega_{Z_2}\times T_S\right)\right)\times Z^{-3} \\ -\left|Z_{Z_1}\right|^2\times\left|Z_{Z_2}\right|^2\times Z^{-4} \end{pmatrix}}{\begin{matrix}\left(1-2\times\left|Z_{Z_1}\right|\times\cos\left(\omega_{Z_1}\times T_S\right)\times Z^{-1}+\left|Z_{Z_1}\right|^2\times Z^{-2}\right) \\ \times \\ \left(1-2\times\left|Z_{Z_2}\right|\times\cos\left(\omega_{Z_2}\times T_S\right)\times Z^{-1}+\left|Z_{Z_2}\right|^2\times Z^{-2}\right)\end{matrix}} \tag{4.55}$$

To benefit from an additional clock cycle, the term in Z^{-1} needs to be canceled. The numerator of (4.55) can then be factorized by Z^{-2} corresponding to a pure two clock cycle delay of the impulse response, setting its two first terms to zero. This can be translated into one of the two sufficient conditions of (4.56):

$$\begin{cases} \left|Z_{Z_1}\right| = \left|Z_{Z_2}\right| \\ \omega_{Z_1} = \dfrac{\omega_S}{4} + \Delta\omega \\ \omega_{Z_2} = \dfrac{\omega_S}{4} - \Delta\omega \end{cases} \quad or \quad \omega_{Z_1} = \omega_{Z\,2} = \frac{\omega_S}{4} \tag{4.56}$$

This means that the zeros must be either equally split around $f_S/4$ and of the same modulus, or at $f_S/4$ without restrictions on their modulus. One can note that this also cancel the terms in Z^{-3}.

For higher modulator's order, if it is an odd multiple of two, one of the complex conjugates zero pair must be at $f_S/4$ and the other ones must be paired by two and respect one of the two conditions in (4.56). When the modulator's order is an even multiple of two, all the complex conjugate zero pairs must be paired by two and respect one of the two above mentioned conditions.

This additional clock cycle gives more time to the quantizer, but it keeps the requirement on the sampling speed. To fully benefit from it, the quantizer can be time interleaved once, meaning two quantizer working at half rate with the second quantizer clock delayed by a half cycle.

As it will be seen all along the remaining of this manuscript, this additional clock cycle is one of the key enablers to answer the challenge of RF direct sampling. Additionally, this $f_S/4$ center frequency is one of the requirements for efficient digital down mixing. This provides some synergy between the analog and digital parts of the receiver.

4.2.4 Sub-sampling modulators

In the above section, modulators with an $f_S/4$ center frequency were discussed. In the aim of a direct RF sampling ADC that would require a sampling frequency of $4\times 28GHz = 112GHz$. In the literature there are few Nyquist ADCs with sampling frequencies approaching this value ([4-8] [4-9]) but they consume between 200mW and 800mW of power. This is well above the budget for the entire receiver. This wide band approach is inherently inefficient for this application where only $1GHz$ of bandwidth is required.

For sigma-delta modulators, the highest reported sampling frequencies are around $50GHz$ ([4-10] [4-11]). They use a very exotic superconducting process and their Josephson junctions. In a more conventional Silicon Germanium Bipolar-CMOS (BiCMOS) 130nm process, BPSDM were implemented with a $40GHz$ sampling frequency ([4-12] [4-13]). None of the reported SDM in CMOS processes exceeds $10GHz$ of sampling frequencies. All these sampling frequencies are clearly too low for the $f_S/4$ modulator.

The modulators considered so far were all receiving signals in the frequency range from 0 to $f_s/2$. This is called the first Nyquist Zone (NZ). The piece of spectrum between $(n-1) \times f_s/2$ and $n \times f_s/2$ is called the n[th] NZ. As already discussed, working in the first NZ with the target of RF sampling imposes an unprecedented sampling frequency. One solution is to use the spectrum folding properties of sampling. When the input signal is at a frequency above $f_s/2$, it will be folded back into the first NZ. Few examples of this approach can be found in the literature ([4-17],[4-18],[4-19]). The question is on how to modify the modulator to accommodate for this change. There are two things that are susceptible to be affected. The sampling frequency and corresponding OSR and the loop filter.

4.2.4.1 Sampling Frequency and OSR

On an "$f_s/4$" modulator the center frequency falls in the middle of the first NZ. For the input frequency to fold in the middle of the first NZ it requires to be in the middle of any NZ. To work in the second NZ the required sampling frequency falls at $37.33GHz$, in the third NZ it falls at $22.4GHz$, then $16GHz$ and so on. The frequency planning for the two last proposed sampling frequencies are depicted on Figure 4-22.

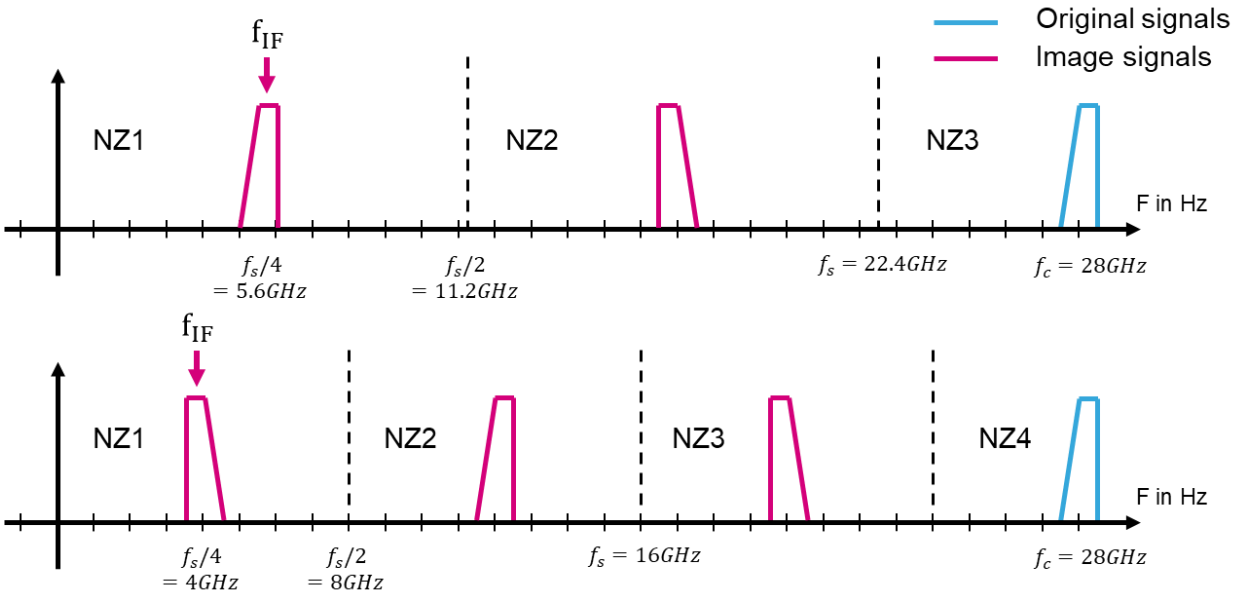

Figure 4-22: Frequency planning for a sampling frequency of $22.4GHz$ on top and $16GHz$ on the bottom

Working in a different NZ allows more acceptable sampling frequencies and gives back some control on the tradeoff between sampling frequency and OSR. As mentioned in section 4.2.3, the "$f_s/4$" architecture allows for time interleaving of the quantizer, halving the sampling rate of each quantizer. The different frequencies and OSR values for each NZ are summarized in Table 4-1.

Table 4-1: Frequency and OSR for inputs in NZ 1 to 5

	Intermediate Frequency	Effective sampling rate	OSR
NZ1	28GHz	56Gsps	56
NZ2	9.33GHz	18.67Gsps	18.67
NZ3	5.6GHz	11.2Gsps	11.2
NZ4	4GHz	8Gsps	8
NZ5	3.11GHz	6.22Gsps	6.22

The authors in [4-14], using a 28nm FDSOI CMOS process, achieve a sampling rate of $10Gsps$. From that it can be concluded that it should be possible to work down to the third NZ. Knowing that the target

is to use a three level quantizer, it is required to have enough OSR. In that regard the fifth NZ is unlikely to be usable. The reasonable choice is between the third and fourth NZ. It will be shown later that the third is actually the only choice. The next topic after OSR is to study how the loop filter is affected by sub-sampling.

4.2.4.2 Loop filter

The intuitive solution for the loop filter would be to keep it centered at the RF frequency while lowering down the sampling rate such that the loop filter center frequency sits in the middle of the desired NZ. To know if this is an acceptable solution one can look at the CT impulse response and find out if an adequate set of feedback coefficient exists that keeps its sampled version equal to the desired DT impulse response.

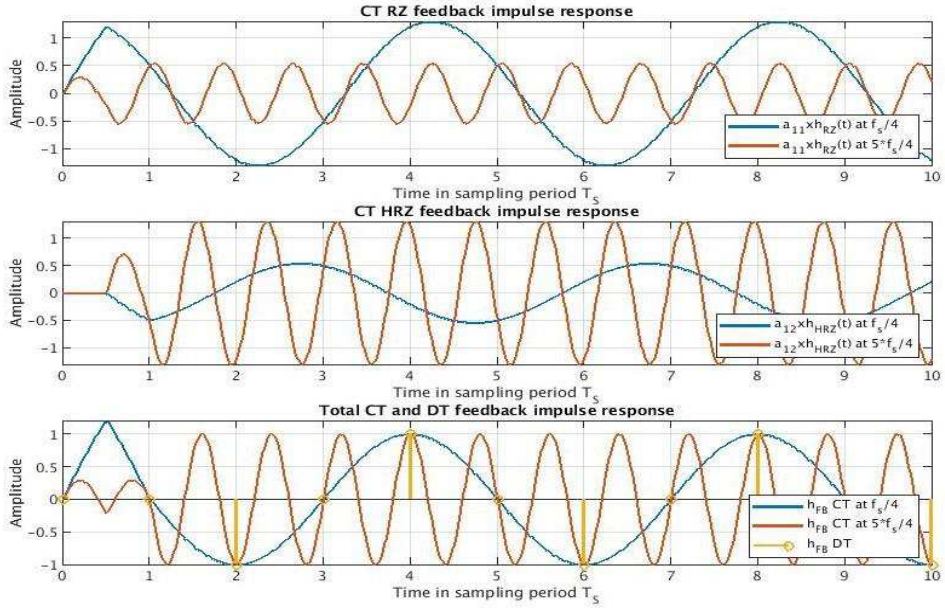

Figure 4-23: Impulse response comparison between $f_s/4$ and $5 \times f_s/4$ resonators

Indeed, such a solution exists for all NZ. The RZ, HRZ and total feedback impulse responses for a modulator coding the third NZ are plotted in Figure 4-23. The feedback coefficients must be adjusted differently but the sampled version of the total impulse response can be made equal to the DT one. Hence the NTF of the modulator is preserved while maintaining the additional clock cycle to close the loop.

4.2.4.3 Conclusion

A straight implementation of an "$f_s/4$" modulator would require an unprecedented sampling rate if a direct RF sampling architecture were to be targeted. To overcome this limitation, subsampling modulators, where the coded band sits in a higher NZ, were investigated. Based on some technological parameters, it was identified that the third or fourth NZ should be a good tradeoff between sampling rate and OSR. Finally, it was shown that the good property of an additional clock cycle to close the loop can be preserved when using a sub-sampling approach with the loop filter center frequency in the middle of any NZ.

4.2.5 ELD compensation in Sub-sampling modulators

At first glance the problem of ELD compensation may look identical to non-sub-sampling modulators. And indeed, the method described in section 4.2.2.3 would work in the same way for sub-sampling modulators. But there is a little more to it. Looking at Figure 4-23 one may note that the sub-sampling RZ and HRZ impulse responses cross zero multiple times between each sample. Intuitively this means

that it should exist a value of ELD where one of these zero crossing point coincides with a sampling point. Hence the value of this particular impulse response coefficient has no impact over the corresponding sampling point.

Taking the RZ impulse response of Figure 4-23, the first zero crossing point happens around $0.3 \times T_S$. If the ELD is set to $0.7 \times T_S$ this zero will fall on the second sampling point, for $n = 1$. The HRZ impulse response will only start to be non-zero after that sampling point. This means that the second sampling point will be zero regardless of the feedback coefficients' values. Hence for this particular value of ELD, both coefficients can be set to fit the rest of the impulse response without the need to compensate for the second sample.

In general, ELD is seen as an implementation impairment that need to be fixed. Its classic compensation is done by adding a feedback to fit the need of an additional degree of freedom in the design space. What will be shown here is that for sub-sampling architectures, ELD itself can be that extra degree of freedom. The approach here is to see ELD not only as an impairment, but also as a design variable.

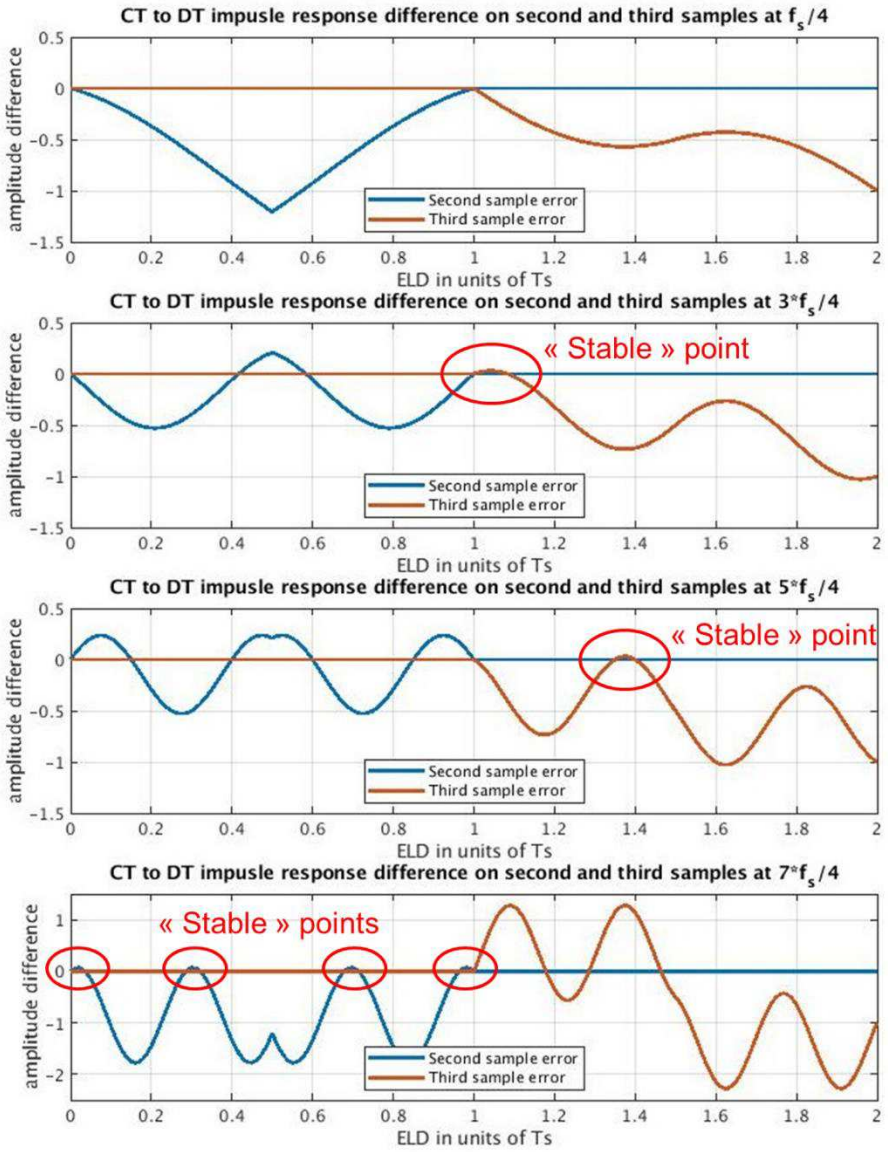

Figure 4-24: Second and third sample errors versus ELD for, from top to bottom, $f_s/4$; $3 \times f_s/4$; $5 \times f_s/4$; and $7 \times f_s/4$ modulators.

To investigate this matter, the impulse response error on the second and third samples will be looked at on uncompensated modulators working in the NZ1 to NZ4, and for ELD variations from 0 to $2 \times T_S$. The coefficients are obtained by fitting to the DT impulse response fourth and fifth samples. The results are plotted on Figure 4-24. For each modulator, the ELD values for which an additional feedback is not required is obtained when the error is zero for both the second and the third sample. In the first case, when working in the first NZ, there are only two such points, at 0 and T_S. The point at T_S correspond to the feature of having an additional clock cycle to close the loop, that was already exposed by the authors of [4-7].

When looking at the subsequent plots, one can see that the higher the NZ the more ELD feedback free points there are. Some of them have an interesting characteristic. They are near the top of a sine like curve. For significant range of ELD values around these points the error will remain low. One can hope that modulators with one of these ELD values will be more robust to ELD variations around that point.

In the sub-sampling modulators from Figure 4-24, only the ones working in the second and third NZ have one of these "stable" points for ELD greater than one clock cycle. The one working in the third NZ zone is very appealing since it is happening near $1.35 \times T_S$, providing an additional $0.35 \times T_S$ to close the loop. This is one of the reasons why the choice will be made to work in the third Nyquist zone. The discovery of this particular characteristic of sub-sampling SDM, having stable configurations in the presence of ELD and without the need for an additional fast feedback loop, is really one of the key enablers to increase significantly the sampling rate and one of the major contributions proposed by this manuscript.

4.2.6 Conclusion

Discrete time and continuous time sigma-delta modulators were studied. Through multiple examples, the impact of many design variables such as quantizer resolution and oversampling ratio were analyzed. How CTSDM are designed and studied was described, showing on the way their intrinsic anti-aliasing property. Finally, the problem of excess loop delay investigated, in general, and in the specific case of sub-sampling modulators. There, a new approach of dealing with ELD was proposed, that allows to get rid of the classically added feedback. A robustness analysis of the obtained modulators was also proposed. In the process it was identified that, for a direct RF sampling approach at $28GHz$, an "$f_s/4$" modulator working in the third Nyquist Zone is a strong candidate.

It is worth noting that the existence demonstration of ELD feedback free modulators is only for second order modulators. In the next sections, it will be demonstrated that this result holds for higher order modulators and is a key property for the proposed architecture to be viable.

4.3 PROPOSED ARCHITECTURE

The purpose here is to propose a sigma-delta based receiver architecture reaching the specifications evaluated in the previous chapter. First, a general architecture will be proposed. Then, using the understanding acquired thus far in this chapter, the design parameters will be gradually tuned until reaching the receiver's full picture. Then, the challenge of excess loop delay will be dealt with. The final touch will consist in improving the receiver's robustness to process variations. On the way, the tools to efficiently simulate and optimize SDM will be develop, allowing for proper characterization.

4.3.1 Architecture

The purpose is to propose an architecture capable to perform RF sampling. In the past this approach, using band pass sigma delta modulators, has been investigated many times in the sub-6GHz range ([4-17], [4-18], [4-19], [4-20], [4-21], [4-22]), and despite being seen as a promising technology it never made to the real world. One could ask why it would be better when moving to the mmWave part of the spectrum. Three reasons can be argued. First, most of the mentioned designs were aiming at providing

flexibility in terms of center frequency and bandwidth, adding complexity and making the resulting solution not competitive enough with the more classical solutions. Second, in the sub-6GHz wireless communication domain, the interferers induce the need for very large Dynamic Range. For example, GSM can require for the receiver to handle up to 90dB of DR. This makes the RF sampling approach extremely challenging, and to the best of my knowledge unsolved to date. In the present case the system analysis revealed no need for configurability and only limited DR. Last, most of the proposed design use integrated inductors which can be very area consuming in the sub-6GHz range and of limited quality factor. When going in the mmWave domain the passive size shrinks significantly, easing integration and the quality factor of inductors improves slightly. For these three reasons, despite the limited success of this approach in the past, the choice was made to go in that direction.

In the previous analysis, a number of architectural characteristics were already identified. A sub-sampling "$f_s/4$" architecture working in the third or fourth Nyquist Zone will be targeted. It will be using a three level quantizer and only feedback paths. Finally, it must be a continuous time modulator using gmLC based resonators. It has been seen that the resonators must satisfy one of two sets of conditions to benefit from the desired additional clock cycle to close the loop. The first set imposes conditions on the resonators center frequency as well as their quality factors. In a practical implementation the quality factor of an LC resonator is hard to control accurately. Therefore, only modulators satisfying the second set of conditions, which is to have their center frequency in the middle of the target NZ, will be considered.

What remains to be determined is the modulator order, the quality factor of the resonators and the working Nyquist zone. In the previous section, the design of higher order NTFs was not discussed. This matter is in fact very rich. Higher order NTFs must be optimized not only for their in-band noise suppression ability but also for their out of band noise gain, for stability reasons [4-5]. This is done by adjusting the zeros and poles of the NTF. This is not something is going to be discussed here. Instead, the web-based design-tool www.sigma-delta.de provided by the Institute of Microelectronics from Ulm University in Germany will be used for a fast performance evaluation of a given architecture. This tool offers multiple advantages. One of them is its ability to synthetized sub-sampling BPCTSDM up to the fourth NZ [4-15].

4.3.1.1 Modulator's order and working Nyquist Zone

First let us evaluate the potential of second, fourth and sixth order modulators using ideal resonators, when working in the fourth Nyquist Zone. The tool only converges for sixth order modulators. It is not clear if this non-convergence is due to a fundamental limit from the architecture or a limitation of the design tool. In any case, it will be considered that second and fourth order modulators are unfit for the target application while working in the fourth NZ. For the sixth order modulator the Signal to Quantization Noise Ratio (SQNR) remains below $35dB$, and the STF present a large peak with a maximum in band gain difference of about 25dB. Experimentally, the zeros' locations in the band need to be optimized to find an acceptable solution but it is preferable avoid this solution.

When working in the third Nyquist Zone, all three orders give solutions which properly behave, with different peak SQNR. The second, fourth and sixth order modulators provide respectively $37dB$, $42dB$ and $45dB$ of SQNR. Since adding more impairments will only degrade the performances it is reasonable to choose the sixth order modulator. The modulators architecture is then represented as in Figure 4-25.

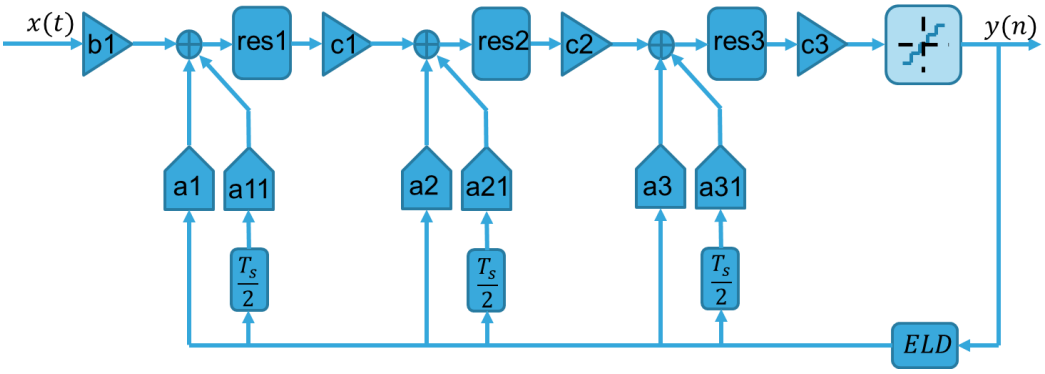

Figure 4-25: Modulator's architecture

4.3.1.2 Quality factor

The last parameters to determine are the quality factors of the three resonators. First, the quality factor of all resonators will be reduced equally and the point where performance start to degrade will be monitored. Surprisingly, the quality factor can be reduced significantly before any noticeable degradation happens. Until $Q = 30$ performances remain almost the same. The reason for this is the targeted wide band. When the Q-factor of a resonator increases the gain increases only around the resonant frequency, and only in this narrow band the noise will undergo more attenuation. The higher the Q, the narrower is the band where the gain is increased. In other words, very high Q-factors will reduce the noise only on a narrow band, and the total in band noise will be dominated by the noise in the rest of the band. Hence, until the Q factors drops significantly enough such that the gain start dropping on a wider band, its variation has a very limited impact on the resulting SQNR.

One important point to consider is the input matching. To connect this modulator directly to an antenna, one solution is to have a purely passive first stage, meaning only the resonator without a gm cell driving it. This allows to use this first resonator as a matching network. Using the quality factor expression from (3.8) and targeting an input impedance of 50Ω and a resonator quality factor of 30, the required values for the parallel LC network can be evaluated:

$$L = \frac{R}{Q \times \omega_0} = 9.5pH \quad and \quad C = \frac{1}{\omega_0^2 \times L} = 3.4pF \tag{4.57}$$

Both of these values are quite impractical at 28GHz. The inductance is too small and the capacitor too large. A reasonable inductance value would be at least $50pH$. The only solution is to give up on the resonator's quality factor. Reversing (4.57) gives the best quality factor that can be expected with a 50Ω matched input and a 50pH parallel inductance:

$$Q = \frac{R}{L \times \omega_0} = 5.68 \quad and \quad C = \frac{1}{\omega_0^2 \times L} = 646fF \tag{4.58}$$

Taking some margin for the implementation impairments, a quality factor of one will be assumed for the first resonator. The important point here is that, with the target architecture, the quality factor of the first resonator is imposed by matching. This means that techniques like Q-enhancement cannot be used since they would result in input mismatch.

On the contrary, this can be done on the subsequent resonators since inside the chip the physical distance between the resonators will be negligible compared to the wavelength at 28GHz, hence matching is unnecessary. A quality factor of 30 for the second and third resonators can then be targeted. During implementation, if the technology does not allow for this value with passive devices, a Q-enhancement circuitry could be added.

To summarize, a quality factor of one for the first resonator and 30 the two subsequent ones will be targeted. The next step will consist in characterizing the modulator in simulation.

4.3.2 Modulator simulation and characterization

The web-based design-tool only provides limited characterization of its performances. In particular, it only provides the STF from a limited frequency range and performs time domain simulations only using sinewave input signals. To overcome this limitation a matlab model was developed that reproduces the web-based design-tool results. This model is then used for an improved STF characterization and an adjustment of the input full scale.

4.3.2.1 SDM simulation model

Starting from the architecture of Figure 4-25, a Laplace transfer function can be associated to each block as in Figure 4-26. These transfer functions are the one used by the web-based design-tool. Using these block transfer function, a signal path transfer function $H_{SP}(s)$ and a feedback one $H_{FB}(s)$ can be processed. To process the STF and the NTF, the Z-Transform of the feedback path must be derived. This could be done analytically by processing the continuous time impulse response of $H_{FB}(s)$, sample it, and process its Z-Transform. This would be very heavy and would not allow to extend to real resonators where analytical expressions of their Laplace transform are unavailable.

Instead, a numerical approach is used where $H_{FB}(s)$ is convolved with a Dirac comb, the equivalent of sampling in the time domain. Since this is a numerical approach, this convolution cannot be done for all frequencies, the Dirac comb must be truncated. Figure 4-27 plot the STF for different convolving maximum frequencies up to a hundred time the sampling frequency and the STF provided by the web-based design-tool.

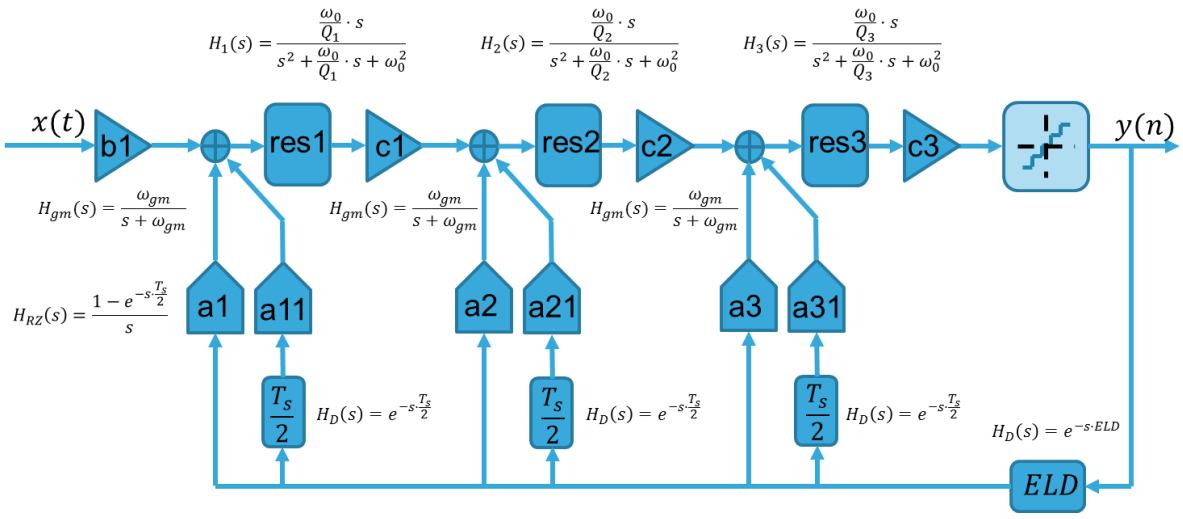

Figure 4-26: Modulator's architecture with its associated Laplace transforms

One can see that for a maximum frequency above fifty times the sampling frequency the STF is nearly aligned with the target STF. In the zoomed left graph of Figure 4-27 it can be seen that some small differences remains and that for $f_{max} = 100 \times f_s$ the error is about half that of $f_{max} = 50 \times f_s$. Going to higher maximum frequency improves further the results but starts to significantly increase the simulation time. For later simulations, a maximum convolving frequency of a hundred times f_s, which is a good compromise between simulation efficiency and accuracy, will be used.

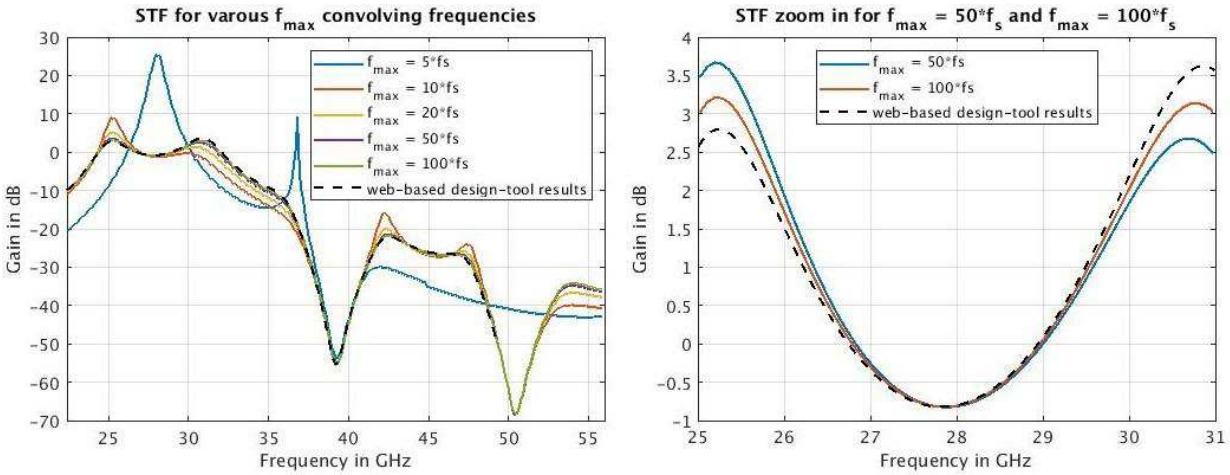

Figure 4-27: STF versus maximum convolving frequency

To perform time domain simulations as before requires knowing the location of the poles and zeros of $H_{FB}(Z)$ and rebuild the difference equation of the system. Since only a numerical evaluation of this transfer function is available this would require a numerical approach such as a gradient descent to find its local minimums and maximums over the complex plane. Instead, a different approach is used that exploits the relatively low quality factor of the resonators. While their impulse responses are infinite, they are also evanescent. Hence their effect can be well approximated by a convolution of the signal with a far enough truncated version of their impulse response.

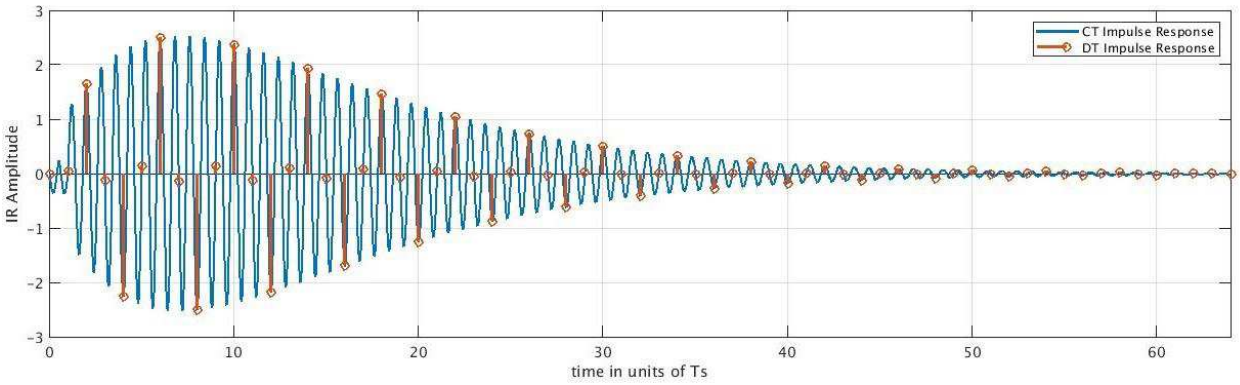

Figure 4-28: Continuous and Discrete Time Feedback Impulse Response over 64 sampling periodes

Figure 4-28 plots the DT and CT feedback impulse response over the 64 first samples. As expected, it decays quickly and is nearly zero by the 64th sample. To ensure simulation accuracy, the feedback impulse response will be truncated at the 256th sample. This will also allow to simulate modulators with higher quality factors. The modulator's feedback is then simply implemented by a 256-tap Finite Impulse Response (FIR) filter where the coefficients are the values of the truncated DT feedback impulse response. Figure 4-29 compares the results from both models and shows very similar behaviors. This allows the validation of the proposed model.

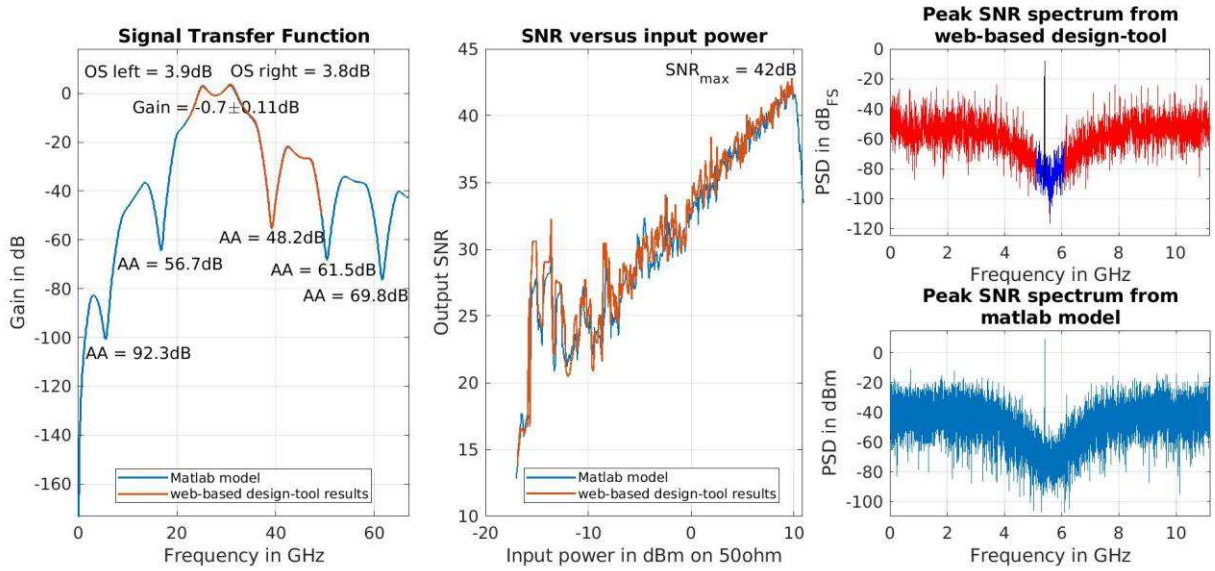

Figure 4-29: Comparison between web-based design-tool results and proposed matlab model results

4.3.2.2 STF characterization

This approach allows to plot the STF over more Nyquist zones (left graph from Figure 4-29) for a better characterization. The in-band gain flatness is characterized to be $\pm 0.11dB$ over $1GHz$ of bandwidth, which is very good. This shows another benefit of working directly at the RF frequency, it is easier to deal with wide bands.

Also, OverShoots (OS) around the band of interest are characterized. Here the overshoots are about $4dB$ while the in band average gain is about $-0.7dB$, giving a difference around $4.5dB$. The potential risk from these OS is on linearity performances. If a powerful interferer is located at one of them it will be gained up, potentially pushing the receiver into compression. Since the characteristics of a potential interferer outside the 1GHz band around 28GHz are not available, it is hard to know if this could be an issue.

Finally, this also allows to evaluate the anti-aliasing properties of the modulator. The first NZ falls at $5.6GHz$ that is in the sub-6GHz band of 5G, which could be problematic. There will be an important activity in that region of the spectrum. The minimum anti-aliasing there is $92.3dB$. Even though this is a high attenuation there could be an issue if the S-BS is co-located with a sub-6GHz BS outputting $50dBm$.

The subsequent NZ all fall in regions which may see activities from 5G, military or satellite applications. Considering the level of intrinsic anti-aliasing, they could prove to be problematic only in the case of co-located BSs. One potential solution would be to synchronize the TDD between the bands such that none of them actually emit while the other ones receive, but that is today only an option for 5G applications. Another point to consider is that for these bands the antennas generally have high and controllable directivity. Since the purpose is to communicate with a receiver away from the BS, it is unlikely for co-located millimeter wave BS to radiate large amount of power on each other's. Hence the proposed level of anti-aliasing is likely to be enough.

In the feasibility study from the previous chapter, the NF was budgeted with provisions for two filters, one in the RF front end, and one for anti-aliasing, before the ADC. If the S-BS is co-located with a sub-6GHz BS, it is not sure both filters can be removed, but the ant-aliasing provided by the modulator certainly allows to remove the anti-aliasing filter. One point that is missing here is the natural response of the antenna itself that will add more attenuation. With this additional attenuation it might be possible

to also remove the RF filter, but for the time being it will only be assumed that the anti-aliasing filter is removed.

4.3.2.3 Input full scale

Before determining the optimal input full scale, the NF specification from the previous chapter must be adjusted to the proposed architecture. In particular, the RF sampling approach allows to get rid of the mixer, and the intrinsic anti-aliasing property of CTSDM allows to get rid of the anti-aliasing filter. It is then necessary to re-allocate these budgets to other components. The NF budget taken for the external front end losses has no reason to be changed since they are unaffected by the proposed architecture. Then, only two budget remains, the Rx analog part and the quantization noise. Here, the choice was made to allocate the 0.4dB budget from the mixer to the Rx analog, and the 0.3dB budget from the anti-aliasing filter to the quantization noise, splitting nearly evenly between the two noise contributors. The updated NF budget is established in Table 4-2. The input full scale can now be evaluated.

Table 4-2: Updated NF budget

	FE	Rx Analog	Q-noise	Total
NF	4.5dB	4.9dB	0.6dB	10dB

It has been seen that SDMs can handle signals with PAPR going beyond their input full scale. This was not accounted for in the specification from the previous chapter. This could result in over-design. Here, a method is proposed to account for this specific trait of SDMs. The proper evaluation of the proposed modulator performances will be done by simulation. Since the highest input signal is very well known, i.e. it is the OoBI evaluated earlier, the input full scale is optimized when this signal is fed at the modulators input. The performance evaluation approach is inspired from Noise Power Ratio (NPR) measurements which are generally preferred over SNR measurement for wide band systems. The OoBI is generated and the thermal noise corresponding to a $9.4dB\ NF$ is added. This is the receiver's NF budget minus the allocated budget to quantization noise. Finally, a notch in the noise of the channel under test is created to finalize the input test signal.

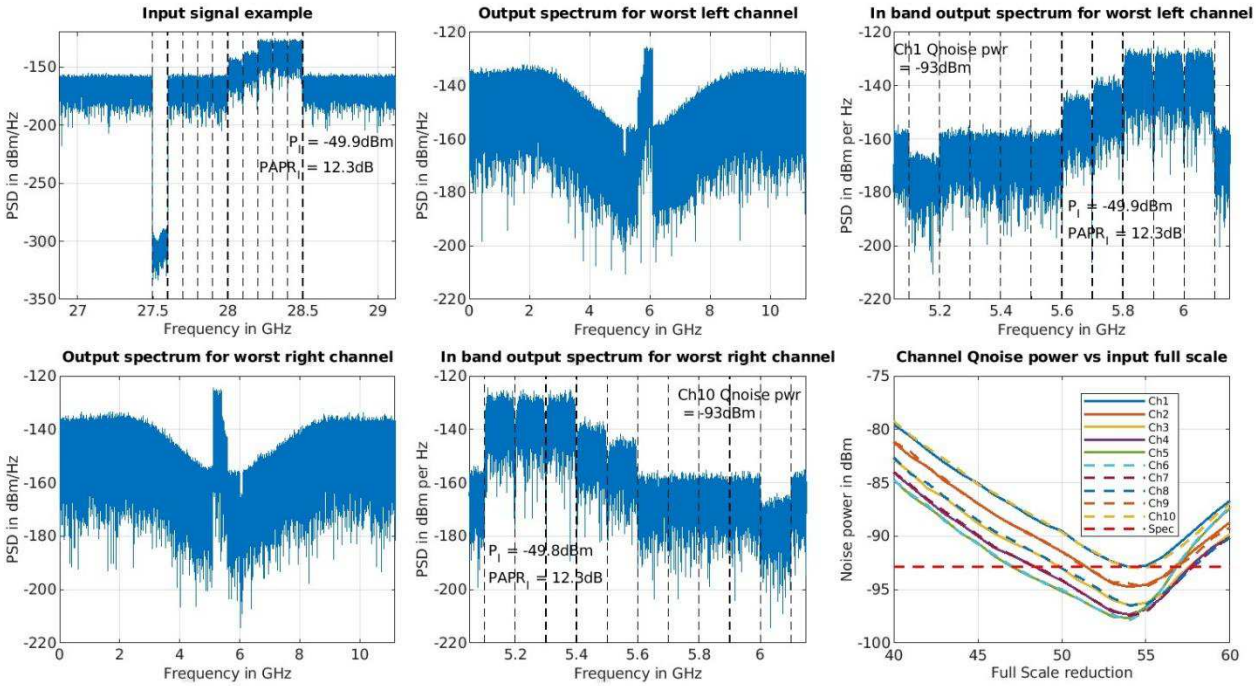

Figure 4-30: Input full scale optimization results

An example of such a signal is given on the top left graph of Figure 4-30 to test the first channel. The OoBI has the desired $-50dBm$ power and the specified $12dB$ of PAPR. These test signals are then injected at the modulator's input. For example, when testing the quantization noise power in the first channel, the output is as per the top middle graph from Figure 4-30, with an in-band zoom on the top right graph. A symmetrical OoIB is used when testing channels six to ten, as in the bottom left graph and its in-band zoom in the bottom middle graph.

To find the optimum input full scale, this test signal is injected while varying the input full scale. This variation is done through the application of an input full scale reduction factor on the reference modulator. Then, the quantization noise power in the channel under test is evaluated. This process is repeated for each channel, adjusting the noise notch location accordingly. The results are plotted on the bottom right graph of Figure 4-30 for each channel. The optimal full scale reduction factor is around $54dB$ and is the same for all ten channels. All channels are within the target specifications. The top middle and right graphs are the modulator output, when at its optimal input scale, while testing channel 1, i.e. the worst case in the band lower half. The bottom left and middle graphs display channel 10 performances in the same conditions. These worst channels are the most outer ones. This could be expected since the resonators are centered in the band, meaning the noise attenuation will be the strongest in the center and the weakest on the edges.

The noise power level in channel one and ten is just at the specification limit. The final version would require some margin with respect to that specification. It is interesting to wonder if the current architecture could be improved to achieve this margin. As already mentioned, increasing the resonators quality factors does not improve thing appreciably on the edges since it mostly reduces the noise in the middle of the band. Another approach could be to split the resonators' center frequencies trying to achieve a flat noise attenuation in the band. As already discussed, this is possible but would increase the complexity of the modulator's calibration. For the time being, as the purpose is only to provide a proof of concept, all resonator center frequencies will be kept at 28GHz, leaving this improvement to future work.

4.3.2.4 Conclusion

A model was built, allowing to reproduce the modulators synthesized by the web-based design-tool www.sigma-delta.de. From that model, the STF of a given solution, corresponding to the proposed architecture, was characterized. The proposed solution provides some appreciable ant-aliasing capability and allows to remove at least one of the filters of the conventional Near-ZIF receiver. Finally, the appropriate input full scale was determined, maximizing the performances for the targeted application.

4.3.3 Non-zero ELD modulators

During the analytical study of SDM, the possibility to have second order band pass modulators with more than one clock cycle of ELD was demonstrated. The purpose here will be to generalize this result to higher order modulators. Because the simulation model used here is different from the one used during the analytical study, it is first necessary to develop further this model for it to gain the ability to simulate SDM with ELD. For the sake of simplicity, this will be done for ELD below one clock cycle. Then, in a second step, the result from the analytic study on second order modulators will be generalized.

4.3.3.1 Modulators with ELD below one clock cycle

What is needed here is a method to design and simulate modulators with an ELD greater than zero. The chosen approach is to start from the reference zero ELD modulator from above and extract its feedback impulse response. It will be used as a reference. Then, a delay in the loop is added and a set of feedback coefficient that reproduces the reference impulse response is processed.

The proposed architecture has six feedback coefficients that can be tuned, or six free variables. Fitting the impulse responses on six samples leads to a system of six equations with six unknowns. In practice, this system is looked at as an optimization problem and solved numerically using a gradient descent. As the cost function, the square of the Euclidian distance of the current modulator DT impulse response to the reference one, on the six desired samples, is used.

The results of this method are plotted in Figure 4-31. Generally, it can be seen that the $0.4 \times T_S$ ELD modulator behaves almost exactly as the original one. The STF and NTF overlap almost exactly and the maximum SNRs are within $0.2dB$ of each other.

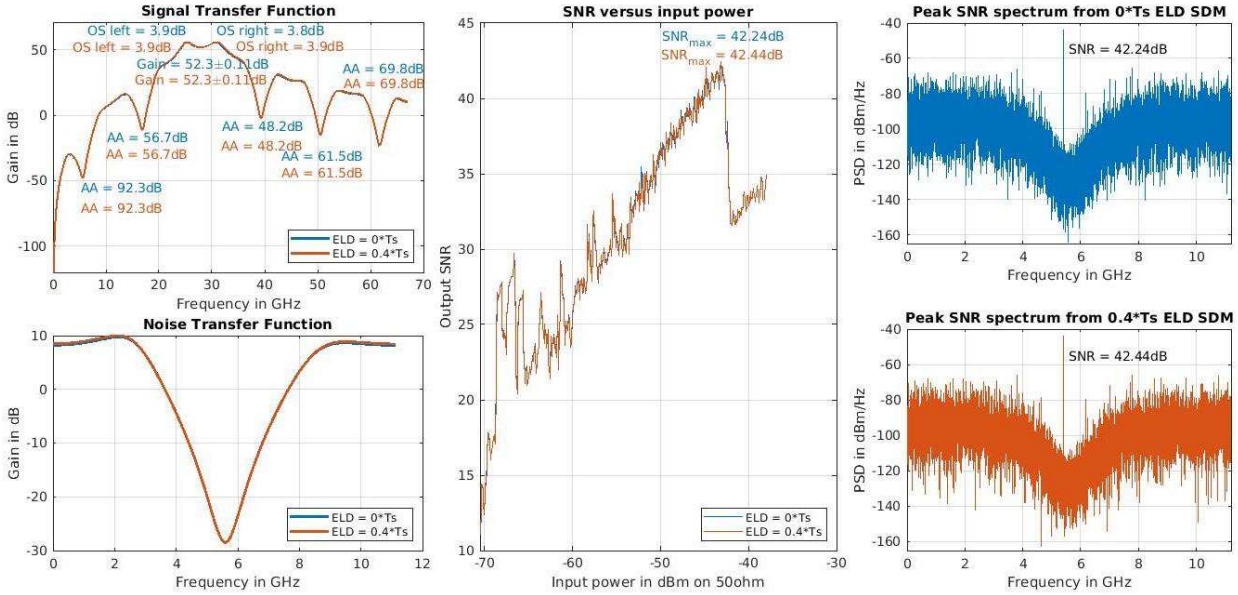

Figure 4-31: Comparison of modulators with respectively 0 *and* $0.4 \times T_S$ *ELD*

It is also interesting to look at feedback impulse response in Figure 4-32. The continuous time impulse responses differ in the beginning but are always the same at the sampling points. This means the discrete time impulse responses are the same, and so are the NTFs. Because modifications were only applied to the feedback path, the feed forward path has remained the same. Since the STF is the ratio between the feed forward TF and NTF, it was expected that it would remain unchanged.

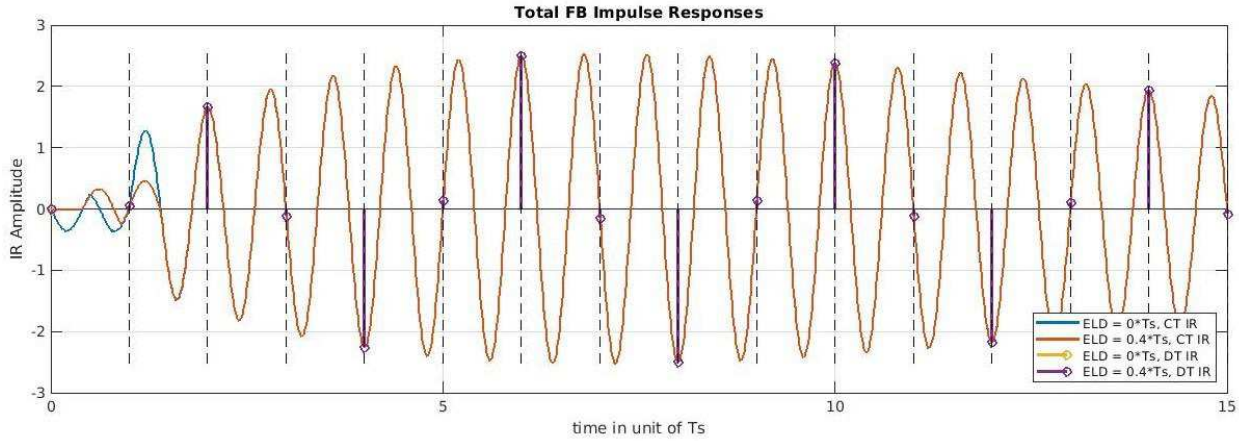

Figure 4-32: Comparison of modulators' feedback impulse responses for 0 *and* $0.4 \times T_S$ *ELD modulators*

What can be concluded here is that this method works at least to go from a zero ELD modulator to a $0.4 \times T_S$ ELD one. It has also been shown that, if the discrete time feedback impulse response of the modulator is preserved, then all its properties are also preserved.

From the analytical analysis, it is known that for some specific ELD values, second order modulators can be stable without the need for an addition feedback loop around the quantizer, for ELD compensation. To find out if these feedback free ELD point still exist, let us look at what happens when the ELD is varied from zero to T_S. The investigation goes as follows. First, the ELD is set to its new value. Then, the feedback coefficients are reset to unity, divided by the gain of each individual feedback path gains. The purpose here is for each coefficient to have a unity effect at the quantizer input, i.e. every feedback path has a similar effect in amplitude. This is used as the initial state before optimization. Finally, the optimization is run to fit the feedback impulse response to the desired reference one. This process is repeated for ELD values from 0 to T_S with a step of $0.01 \times T_S$.

In Figure 4-33 are plotted some of the results for this process. Left graphs plot the initial and final values of the cost function such that it can be evaluated how far the optimization went. On the top-right graph are the initial and final values expressed in dB. This helps to compare values when they are spread across multiple orders of magnitude. While it is uneven across the studied ELD range, the cost function final value is always better than the initial one by at least three orders of magnitude. Surprisingly, the fit is not perfect for zero ELD. A potential explanation is that the problem is most likely not convex, and that the gradient descent felt into a local minimum. It can be seen that the cost function final value can vary up to three orders of magnitude versus ELD. This will be studied shortly after.

The bottom-right graph plots the evolution of the normalized feedback coefficients. It can be described as piece wise continuous. The coefficients evolve continuously for a given range of ELD and abruptly change to different values in a subsequent ELD range. There are two notable facts to observe here. First, the coefficients a_{11}, a_{12}, a_{21} and a_{22} are generally large for most ELD values. Only punctually they are all small, like around $ELD = 0.4 \times T_S$. This will be investigated deeper in sections 4.3.4 and 4.3.5.

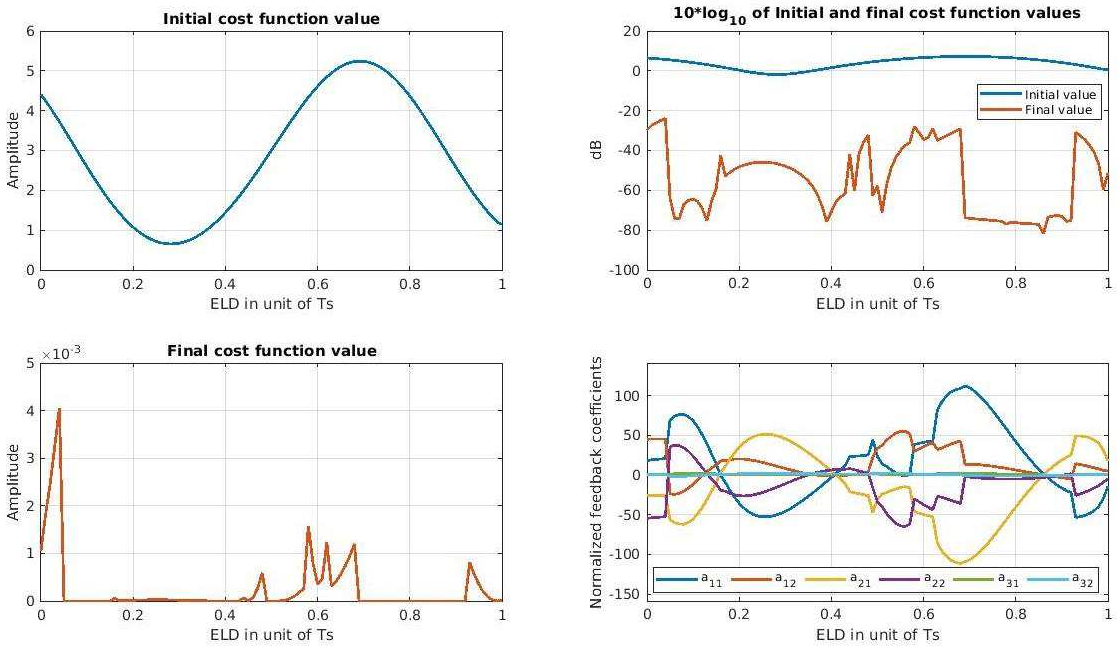

Figure 4-33: Optimization results for ELD values from 0 to T_S

Second the coefficients a_{31} and a_{32} remain small for the entire studied ELD range. This latter fact can be explained since these coefficients correspond to the most inner loop of the modulator. Its effect

happens first and is therefore the most effective to set the first DT feedback impulse response point. As a consequence, they have to remain of the same order of magnitude as this first point. Since in the current case this first point is nearly zero, these coefficients must remain small.

One important question that can be raised is: What is the impact of the optimization final value on the characteristics of the resulting modulator? Figure 4-34 left graphs compare the STF and NTF of the reference modulator with a modulator having an ELD equal to $0.04 \times T_S$. This was the ELD value resulting in the worst optimization final value. One can note that both the STF and the NTF are significantly altered. Surprisingly, the resulting spectrum and SNR in the middle graph remain acceptable. Hence it is more sensitive to compare the STFs and NTFs to evaluate the quality of the optimization.

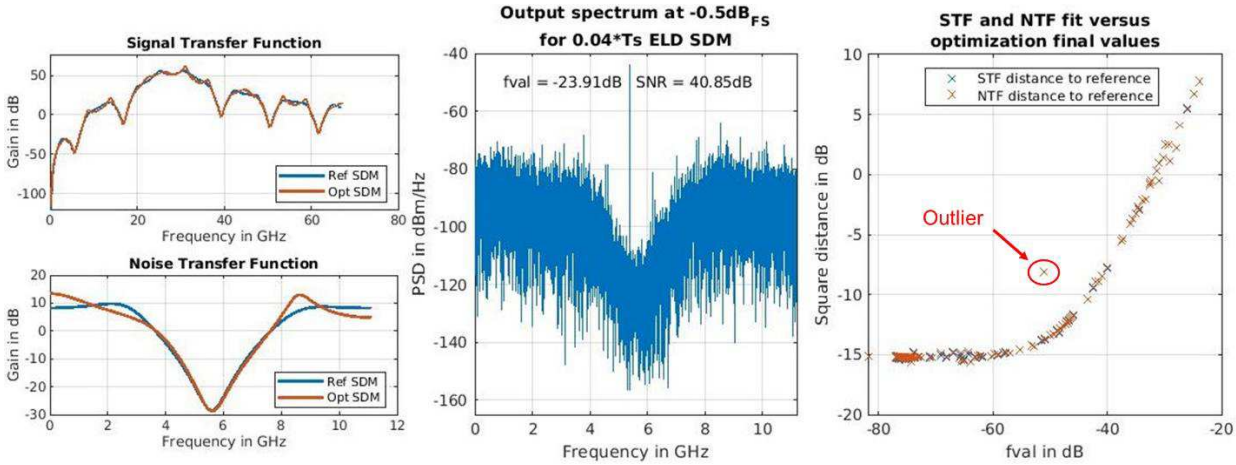

Figure 4-34: Optimization final value impact

The right graph plots the distance between the reference and the optimized STF and NTF curves in decibel as a function of the optimization final value. The distance between the reference and optimized STF and NTF is meant here only as a measure of how well the curves overlap. It is pretty clear that, for final values below $-55dB$, no significant improvement is observed anymore in the STF and NTF fits. Even though it is somewhat arbitrary, this limit value will be used for acceptable optimization outcome.

One may be surprised that the distance values between optimized and reference modulators are actually the same for STFs and NTFs. In fact, it is to be expected since the STF is the ratio between the feed forward TF and the NTF; when expressed in dB it becomes the difference. Since the feed forward TF remains unchanged, only the difference between the NTFs remains. This explains why the STF and NTF distances to reference are the same. This can be seen as another illustration that preserving the NTF will also preserve the STF.

One last note on this right graph can be made about the outlier around the point $(-50dB, -8dB)$. This point corresponds to the case where the ELD is exactly one clock cycle. This means that no matter what, the value of the impulse response at T_S will remain zero. This is because nothing has yet been feedback at that point in time. Since in the reference modulator this point is not exactly zero, the desired feedback impulse response can never be reached. Even though this first point is near zero (Figure 4-32) it can be seen that it has a significant impact on the resulting modulator. This aspect of the problem will be studied in the next section for modulators with ELD greater than one clock cycle.

As a conclusion, a method to generate modulators with ELD lower than one clock cycle was developed. The current method only provides good solutions for some ELD values in that range, but the tools needed to evaluate the acceptability of a solution was developed. The next step is to generate modulators with more than clock cycle of excess loop delay.

4.3.3.2 Modulators with ELD between one and two clock cycles

In section 4.2.3, it has been seen that the second sample $h_{FB_{dt}}(1)$ of the Discrete Time Feedback Impulse Response (DTFBIR) must be zero to benefit from an additional clock cycle. This is not the case for the current reference modulator, even though it is close. This was already noticed with the outlier point on Figure 4-34 right graph, for a one clock cycle ELD modulator. Going for ELD larger than that requires to modify the feedback impulse response, hence the NTF and the STF.

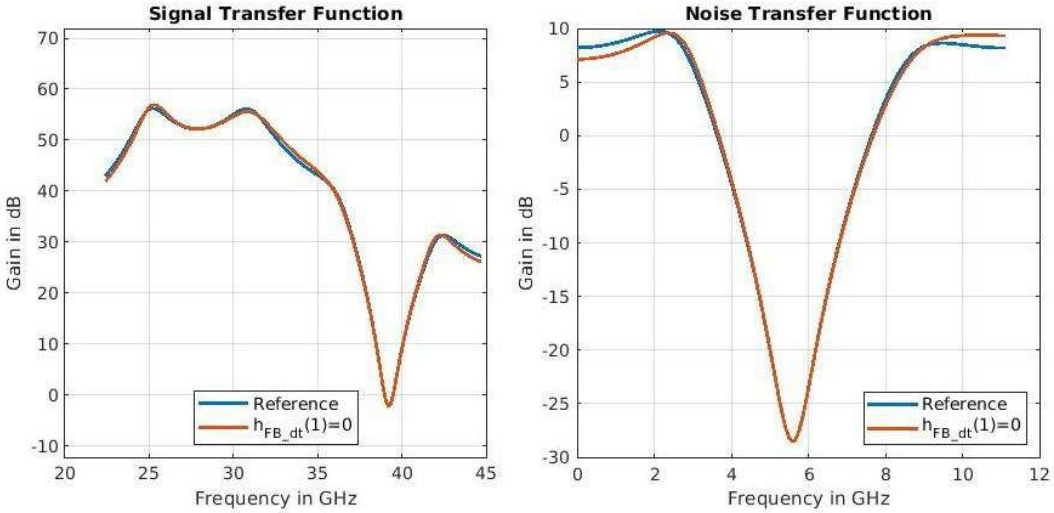

Figure 4-35: STF and NTF comparison when feedback impulse response second sample set to zero

Here, two ways of doing so are investigated. The first one is simply to set the second sample to zero. Figure 4-35 plots the comparison between the STF and the NTF with reference DTFBIR. The effect on the STF is very small. It is more pronounced on the NTF. In particular, one can note the loss of symmetry.

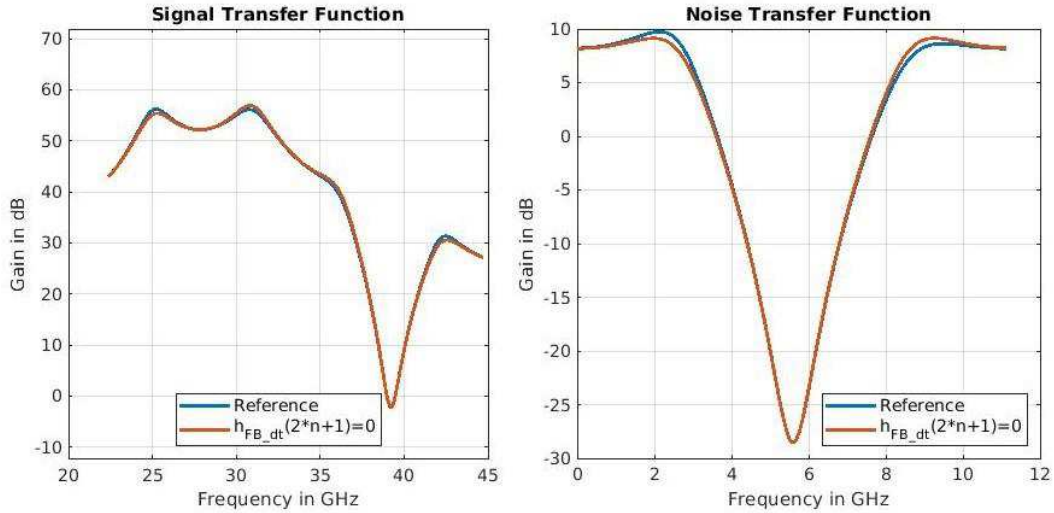

Figure 4-36: STF and NTF comparison when feedback impulse response odd samples are set to zero

For the second way of modifying the DTFBIR, one can note, from Figure 4-32, that samples with odd indexes are all close to zero. The modification consists in setting all these terms to zero. The resulting STF and NTF are plotted on Figure 4-36, together with the reference modulator. As before, the effect on the STF is very limited and it can be seen that the NTF symmetry is partially restored. This new NTF is even slightly better than the reference one. While the in-band noise rejection is the same, the out of

band gain is a little lower. In a physical implementation this can be better since it reduces the amplitude of the out of band quantization noise, hence driving the modulator in a more linear state of operation.

Going onward, this second option will be used as the new reference feedback impulse response. The range of ELD between one and two clock cycles can now be scanned with the same method as in the previous section. Overall, very similar conclusions are reached.

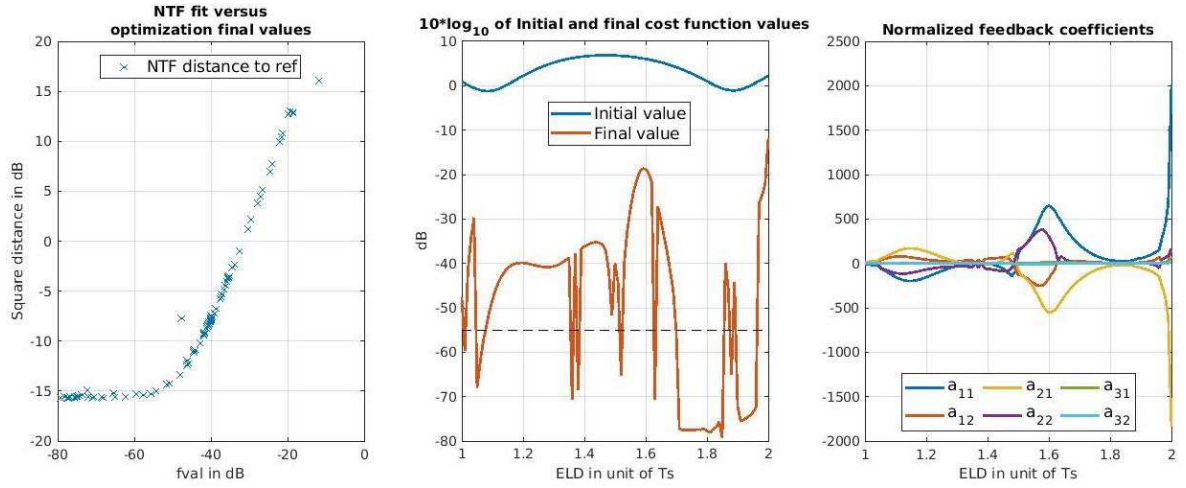

Figure 4-37: Optimization results for ELD values from 1 to 2 T_S

From Figure 4-37 left graph, it can be seen that the optimization final value's limit, for the resulting modulator to have the same characteristics as the reference one, remains around $-55dB$. From the middle graph, the same conclusion that the quality of the optimized modulator varies greatly with ELD can be reached. Finally, with the right graph, one can see that the evolution of the feedback coefficients is what is called piece wise continuous.

This last observation suggests that two ELD values close to each other are likely to lead to modulators with similar feedback coefficients. Hence, instead of running the optimization from a normalized initial point, it might be sensible to start from the solution of the previously processed ELD point. The results using this approach are plotted in Figure 4-38. Overall, they are very similar compared to the optimization from a normalized initialization. The main difference is that the optimization reaches better final values for more ELD values, especially for ELD between 1 and 1.5 clock cycle. Ultimately this shows that the optimization depends strongly on the initial point, hence proving that the problem is not convex, at least as it is formulated here. It is also very likely that the solution is not unique.

During the analytic study of a second order BPSDM, with a resonator of infinite quality factor, it was known that the system was defined and had one and only one solution. Hence the resolution method had little importance, as long as it was providing a solution, it was known it would be the best one. Based on this assumption, a gradient descent was chosen for its implementation simplicity. The difference is that, in the present case, the resonators do not have an infinite quality factor anymore. This leads to an under constrained system with potentially multiple solutions. This also gives a rough explanation for why solutions for large bands of ELD were found with only six feedback paths, i.e. six free variables, while solutions were expected only for specific ELD values, as suggested by the analytic study. The optimization process needs to be refined to reach better solutions.

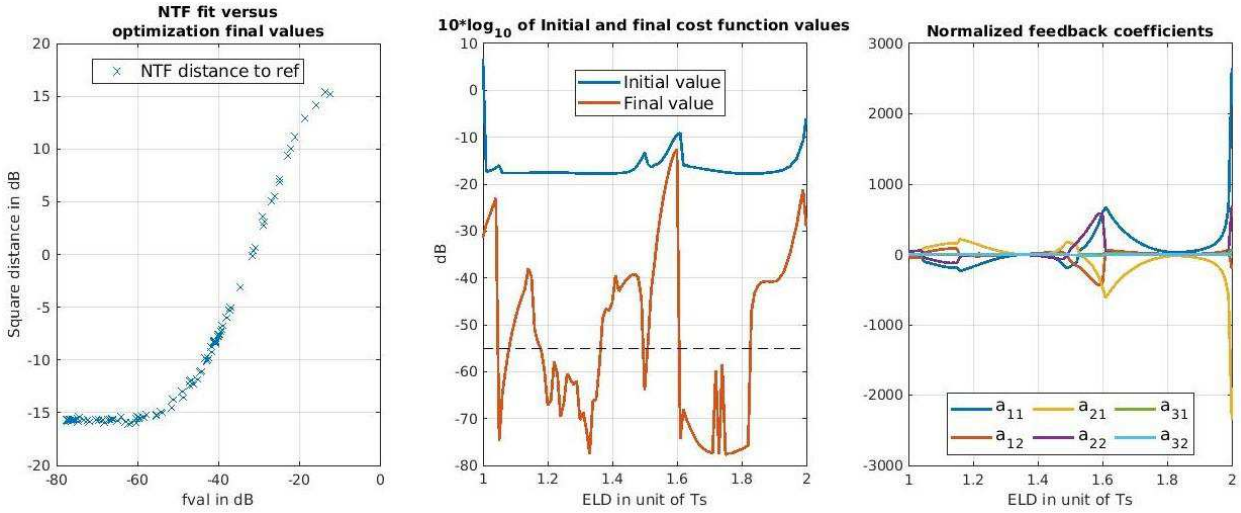

Figure 4-38: Optimization results for ELD values from 1 to 2 T_S using previous point for initialization

To improve the amount of design space that is covered by the optimization process, the proposed procedure is refined as follow. For each ELD value, multiple optimizations are run, each of them starting from a different initial point. The first two points are the ones just described, one starting from the previous ELD step solution and the other from the normalized state. The remaining optimizations are run from randomized initial states. They are obtained by multiplying the six feedback coefficients of the normalized state by a six-element random vector draw from a Gaussian distribution with zero mean, unity variance and zero correlation, i.e. the covariance matrix is the identity matrix.

Experimentally, it has been found that going past twenty runs only rarely brings new and better solutions and increases significantly the optimization time. Hence, the search will be limited to eighteen random runs plus the previous and normalized ones for a total of twenty runs.

The results of this refined procedure are gathered in Figure 4-39. The overall conclusions are the same. The main difference is that, now, the algorithm delivers acceptable solutions for a wider range of ELD values. This improves the algorithm ability to find acceptable solutions.

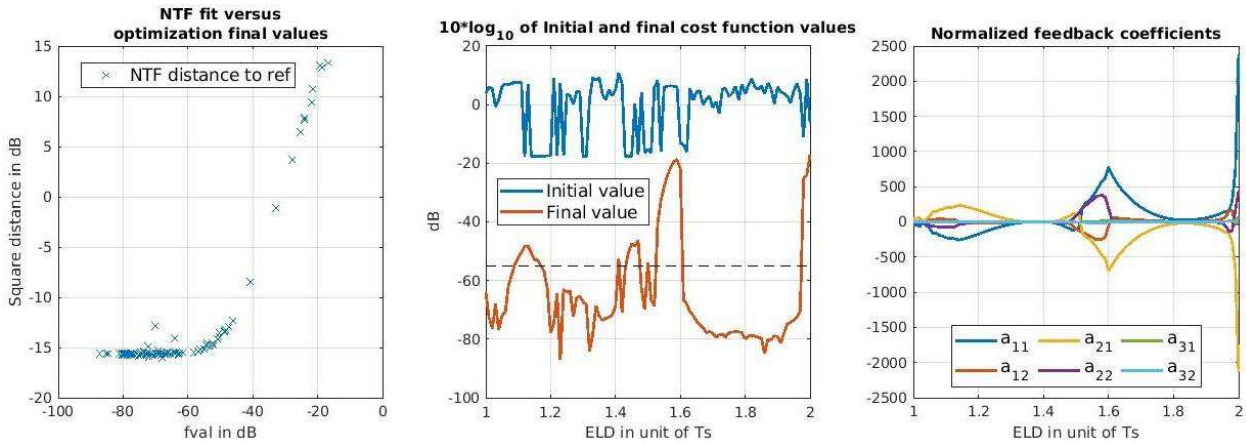

Figure 4-39: Optimization results for ELD values from 1 to 2 T_S using the refined procedure

Regardless of the optimization method it seems that no proper solution can be found for some range of ELD values. There are four such zone, one around $1.15 \times T_S$, the second around $1.45 \times T_S$, the third around $1.55 \times T_S$ and the last one just before $2 \times T_S$. A proper explanation for these zones is left for future work. For now, they will simply be considered as unusable.

4.3.4 Robustness optimization to feedback coefficient variations

The tools to generate modulators with the desired performances for almost any ELD values are now available. To be good candidates for a practical implementation, they also need to be robust to some level of random fluctuations in their design parameters (fluctuations induced by devices imperfections). Among all the design parameters of a modulator, two of them will be more specifically investigated, the feedback coefficients in this section and the ELD in the next one.

Here, in a first step, the sensitivity of a modulator to random variations of its feedback coefficients must be evaluated. More specifically, the goal is to investigate if some conditions on their values may be more robust to random fluctuations. Then, in a second step, these conditions will be added into the optimizer in order to produce more robust modulators.

4.3.4.1 Sensitivity to feedback coefficient variations

Let us first look at the contribution of each individual feedback paths to the Continuous Time Feedback Impulse Response (CTFBIR) in Figure 4-40. For each feedback path, there are the impulse responses of the none-delayed and half delayed path, as well as their sum. The last curve is the total CTFBIR overlapped with DTFBIR, showing their matching at the sampling points.

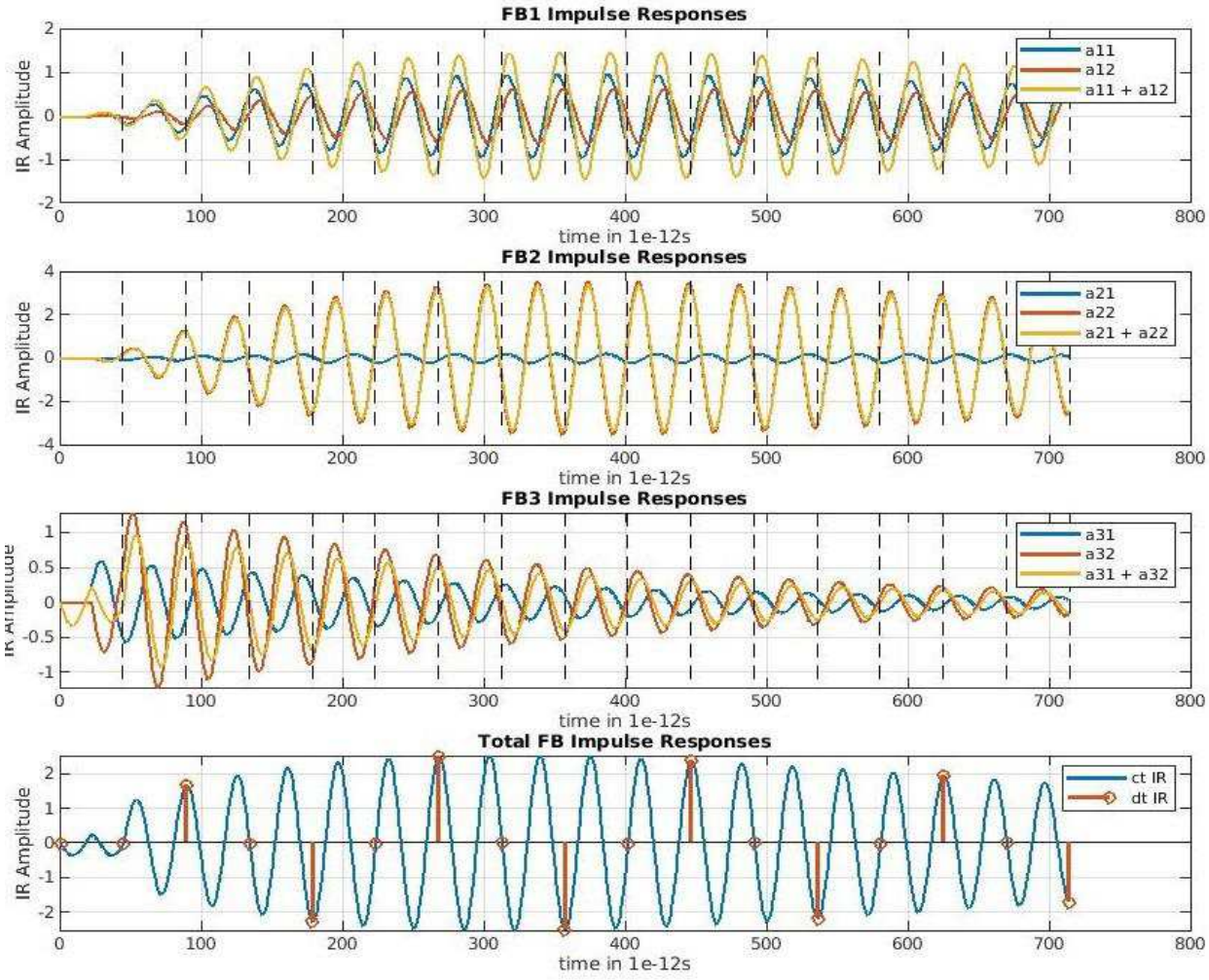

Figure 4-40: Individual continuous time impulse response of individual feedback paths of the reference modulator

Here, only relative fluctuations which can be induced by capacitance mismatch, for example, will be considered. When one coefficient varies by some percent, the impact on the total CTFBIR will depend

on the absolute value of that coefficient. If it is small, it will be small, if it is large, it will be large. Intuitively, one can conclude that a modulator with lower feedback coefficients will be less sensitive to their relative variations. Because they cannot be zero, otherwise they would be useless, it means that a good compromise must be, for each individual feedback path, to have a contribution in the same order of magnitude as the total CTFBIR.

With feedback path peak values below four, and a total CTFBIR peak value around two, the reference modulator respects this condition. To confirm this intuition, let us compare it with one with larger coefficients. From Figure 4-39 it can be seen that, for example, a modulator with an ELD of 1.2 clock cycle will exhibit good performances but has large coefficients. Its feedback path contributions are plotted in Figure 4-41. They are effectively between one and two orders of magnitude larger than the reference modulator.

To compare the robustness of these modulators, a random relative variation is first applied to the modulators' coefficients. Then, resulting performances are measured. The variations are emulated by adding a random vector to the vector formed by the feedback coefficients. This random vector comes from a Gaussian distribution with a diagonal covariance matrix and a variance of 10% for each coefficient. Finally, an input signal at $-1dB_{FS}$ is injected and the output SNR is measured.

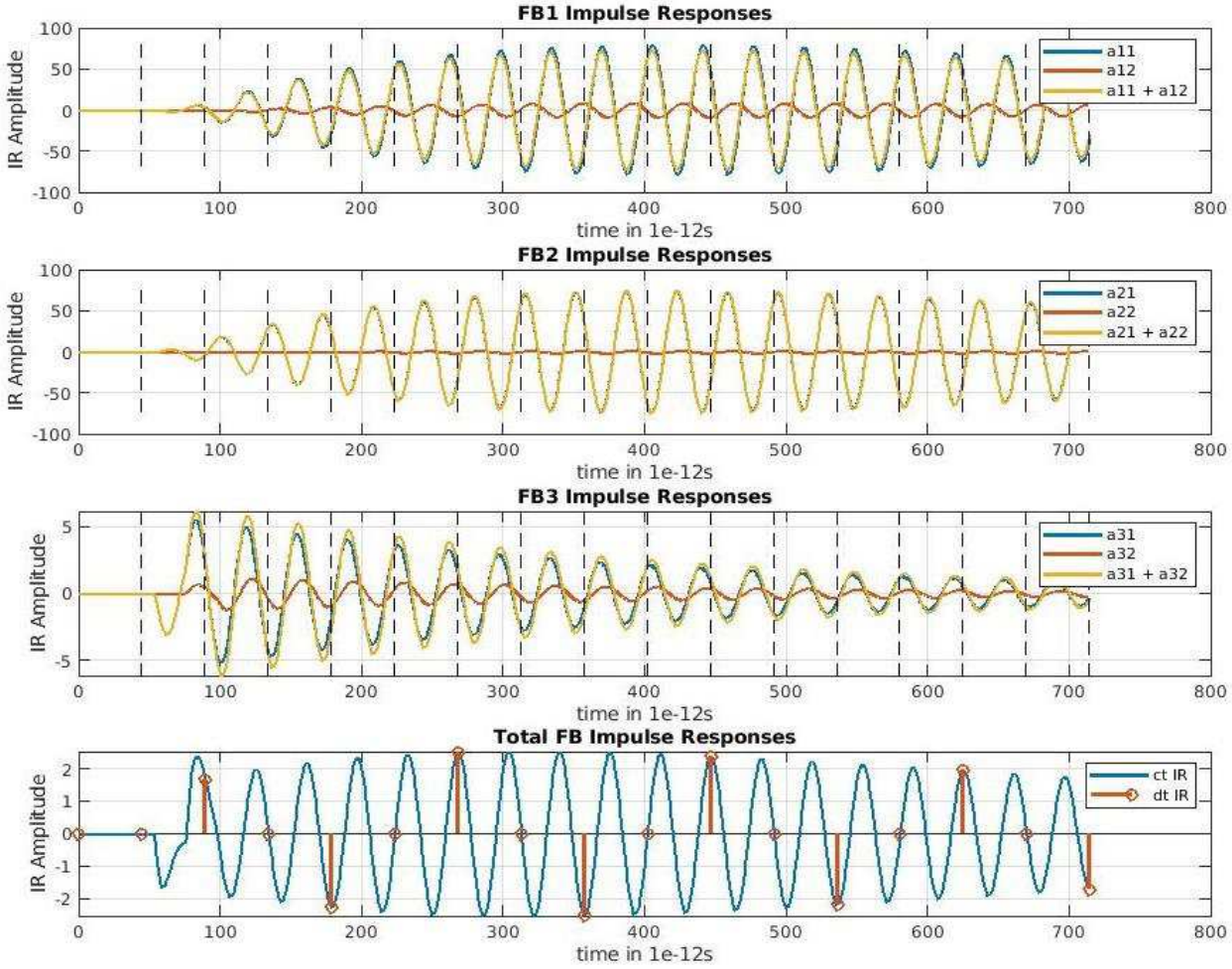

Figure 4-41: Individual continuous time impulse response of individual feedback paths of the 1.2 clock cycle ELD modulator

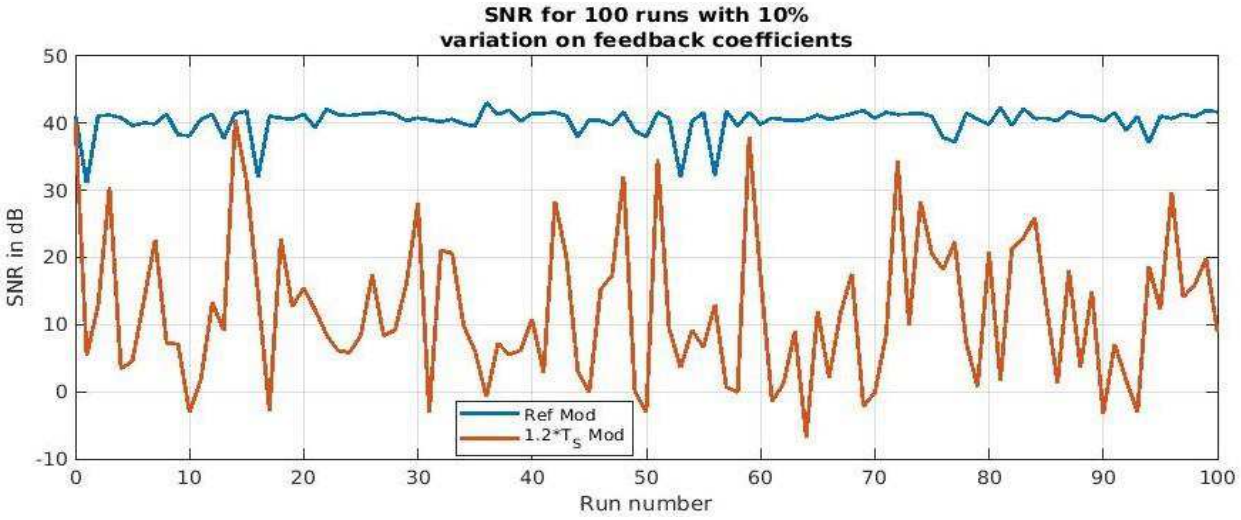

Figure 4-42: Reference and 1.2 clock cycle ELD modulators SQNR for Gaussian feedback coefficient varations over 100 runs

Figure 4-42 plots the results for both modulators for a hundred runs. In both cases the first point corresponds to the case with no variations. As expected, the modulator with larger coefficients is a lot more sensitive to their variations.

4.3.4.2 Robustness optimization to feedback coefficient variations

From the conclusions of section 4.3.4.1, in order to improve the optimization method, it may be desirable to add a second objective to the optimizer. As a measure of the feedback coefficients, the choice was made to use the square sum of their normalized values. Figure 4-43 plots the final value of the current optimizer as well as the feedback coefficients' square sum, both expressed in decibels for visualization convenience. One can note that the optimizer final value is often much smaller than the current target of $-55dB$. This gives some margin to find solutions with lower feedback coefficients.

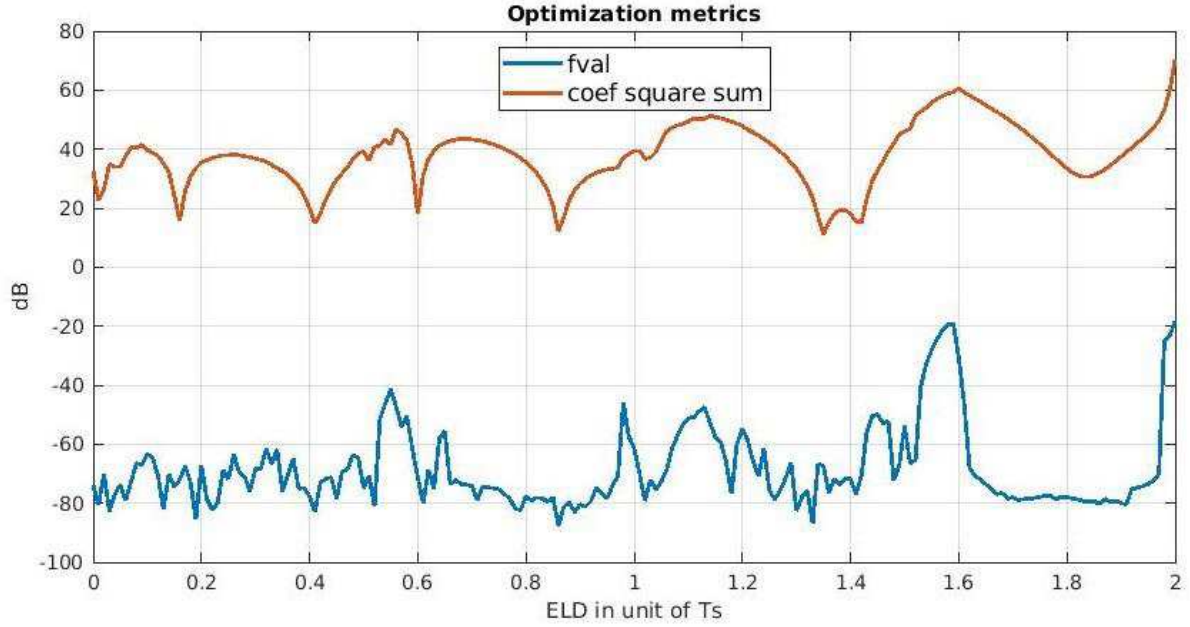

Figure 4-43: Optimizer final value and feedback coefficients' square sum

A new cost function that would account for both metrics is now needed. The simplest solution would be to use their sum. From Figure 4-43, it is clear that these metrics take very different values. A more sensible approach is to use a weighted sum. The reference modulator feedback coefficients' square sum is about $12dB$. Assuming this correspond to a near optimal solution, it is desirable to scale this value to the $-55dB$ required for the distance to the reference DTFBIR. That means an attenuation of $67dB$. The new threshold for acceptability is increase by three decibels, reaching $-52dB$.

Figure 4-44 plots the optimization results with this new cost function. Both metrics follow the same trend and are well optimized only for some specific ELD values. In fact, these ELD values are the ones predicted by the analysis of a second order BPCTSDM in section 4.2.5, as recalled in Figure 4-45. Proper optimization points align perfectly with the cases where there is no error on the second and third samples of the FBDTIR of a second order BPCTSDM. This extends the existence demonstration of ELD feedback free modulators to higher order modulators. Intuitively this can be explained by the fact that the outer loops, that increase the modulator's order, will go through multiple resonators, each time experiencing the corresponding group delay. Hence their effect is only significant on later samples of the DTFBIR, the first non-zero sample is mostly affected by the most inner loop. This is even more true as the ELD gets larger. It is therefore not so surprising that the result on second order modulators transferred to higher order ones.

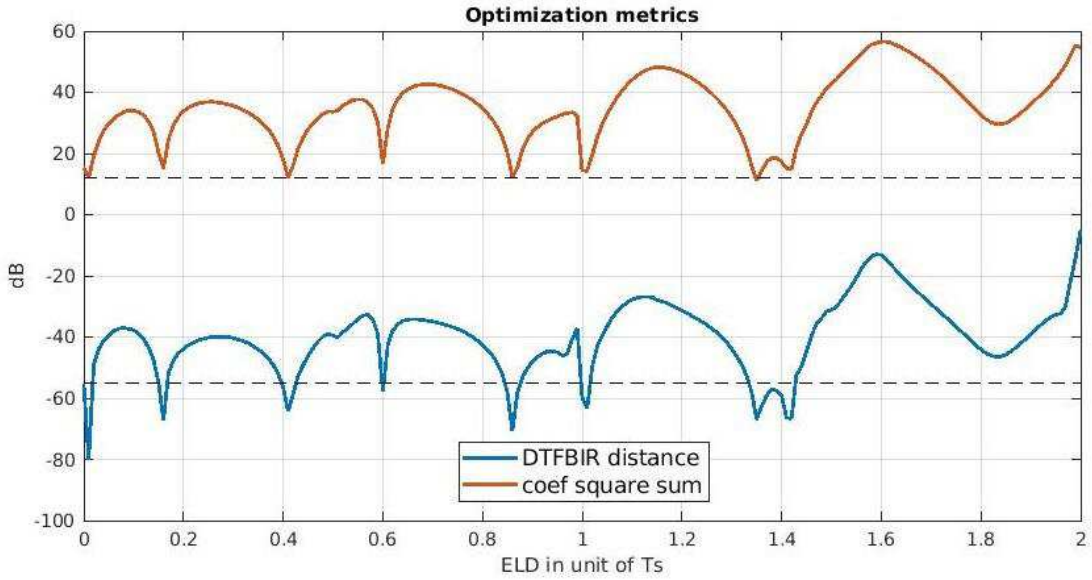

Figure 4-44: DTFBIR distance to reference modulator and feedback coefficients' square sum versus ELD

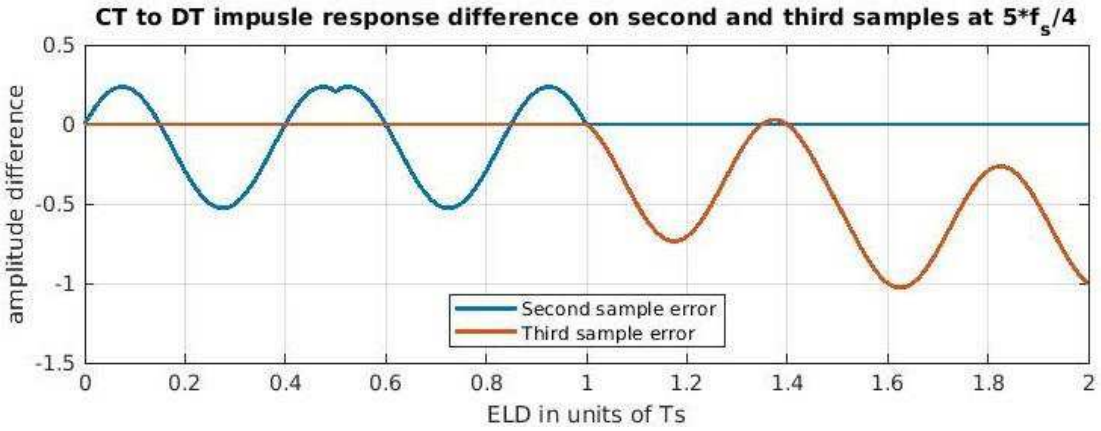

Figure 4-45: Second and third feedback impulse response sample errors versus ELD for a $5 \times f_s/4$ *second order BPCTSDM*

4.3.5 Sensitivity to ELD variations

Ideally, it would be best to use the same approach as in the previous section, i.e. first, evaluate the sensitivity and second tune the optimizer to reduce this sensitivity. Unfortunately, as it can be seen on Figure 4-44 red curve, the current optimizer gives out coefficient square sums just around the target threshold of $12dB$. This means that there is not much room left to optimize a third performance indicator without risking compromising the two others. Still, it is interesting to know the sensitivity of the modulators to ELD variations, to avoid picking a solution that has an unreasonable sensitivity to it.

First, the robustness of the reference modulator is evaluated. To do so, an ELD offset is added into the model and the output SNR for a $-1dB_{FS}$ input sinewave is measured. Note here that, for this new ELD value, the modulator is not re-optimized. The purpose is to evaluate the modulator's performances when the ELD is not as per design. This operation is repeated for various values of ELD offset such that a range wide enough around zero offset is covered. Finally, the range of ELD offset where the SNR remains better than the $40dB$ target is evaluated. This metric has no absolute relevance outside this specific use case but will be useful for relative comparison.

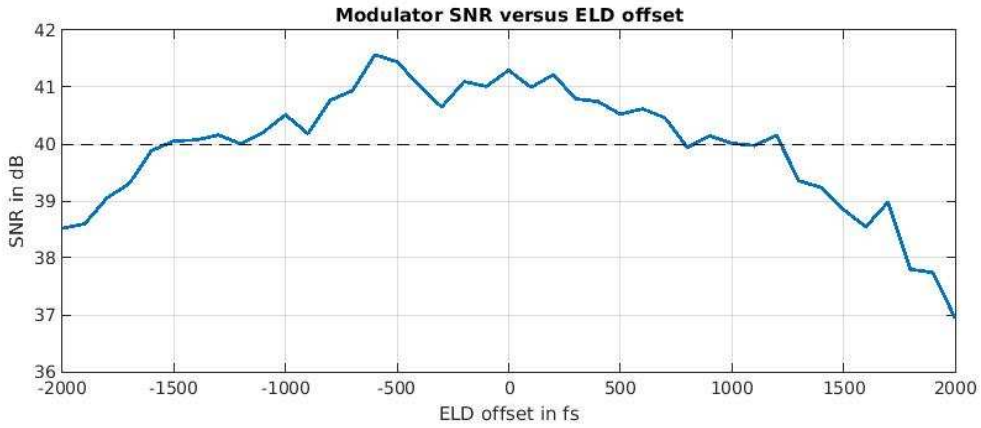

Figure 4-46: Reference modulator SNR versus ELD offset from $-2ps$ *to* $2ps$ *with a step of* $100fs$

Figure 4-46 plots the reference modulator's output SNR as a function of ELD offset. This offset is varied from $-2ps$ to $2ps$ with a $100fs$ step. The range of ELD offset for which the SNR remains better than $40dB$ is $2.7ps$. Let us call this the ELD band. It can be seen that the SNR is not a very accurate metric; hence the result of this analysis will be more qualitative than quantitative. One more note here is that, the reference modulator having zero ELD, the range with negative offset actually corresponds to a negative overall ELD. This is obviously unphysical and is only possible in a matlab model.

The same operation is now repeated for modulators with per design ELD between zero and two clock cycles. For each of them, their ELD band is measured. The results are plotted on Figure 4-47 top graph while the bottom graph plots the maximum SNR within the ELD band.

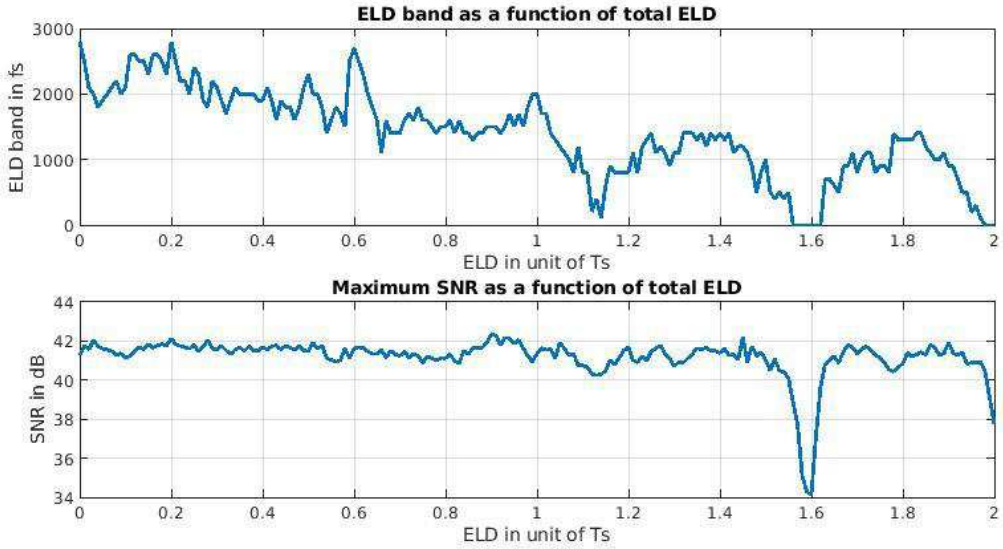

Figure 4-47: Top) ELD band for modulators with per design ELD between $0 \times T_S$ *and* $2 \times T_S$*. Bottom) Maximum SNR within the ELD band.*

There are two trends, one for ELD below one clock cycle and one for ELD above. In the first case, the larger is the ELD, the lower is the ELD band. The second case displays a maximum ELD band around $1.3ps$ with three regions where the robustness is especially poor, around 1.1, 1.6 and 2 clock cycles. These regions correspond to ELD values where the optimizer cannot find proper solutions.

From an ELD robustness point of view, it is desirable to have an ELD as low as possible and out of any of these “bad” regions. From an implementation point of view, the opposite is desired, an ELD as large as possible, to have additional time to close the loop. A large part of this chapter’s investigation turns around exploiting the additional clock cycle in ELD, allowed by "$f_S/4$" modulators. It is based on the assumption that, with a sampling clock running at $22.4Gsps$, a single clock cycle ELD would be impractical for implementation. Indeed, it will be seen in the next chapter that achieving an ELD below two clock cycle is already at the limit of what the technology can do with an acceptable amount of power. Hence, only ELD values above one clock cycle were considered. Only two sensible choices remain, to use an ELD around 1.35 or 1.8 clock cycle. It was shown in section 4.3.4 that modulators beyond 1.4 clock cycle of ELD display a high sensitivity to feedback coefficient variations. The only remaining option is to design for a modulator with an ELD around 1.35 clock cycle.

The SNR plot versus ELD offset is given in Figure 4-48. One can note that note only the ELD band is narrower but also the slopes before and after the ELD band are steeper compared to the reference modulator. This means that a wrong ELD value will be even less forgiving. While this value of ELD is the least bad option, it is clear that this will remain one of the toughest challenges for implementation.

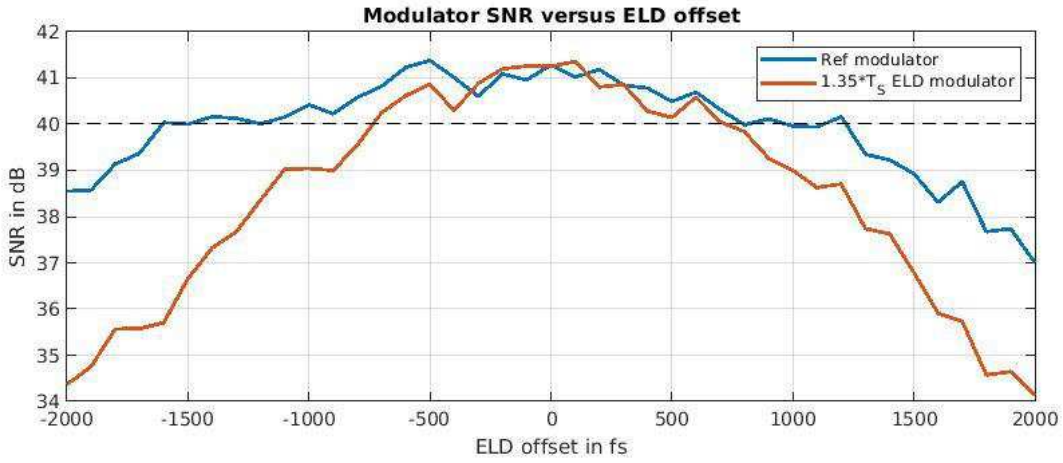

Figure 4-48: SNR versus ELD offset from $-2ps$ to $2ps$ with a step of $100fs$ for the reference and $1.35 \times T_S$ ELD modulators

4.3.6 Individual feedback path ELD optimization

Here, one last aspect of ELD will be investigated. Until now, all feedbacks were assumed to have the same ELD. Not only is this not easy to ensure, but it is even undesirable from a design perspective since this timing constraint is very challenging. Only one of the feedback loops needs to be fast enough to realize the first non-zero coefficient of the DTFBIR. When looking at the continuous time impulse response of the individual feedback path from Figure 4-41, it can be seen that the most inner one is the most efficient to perform this task. The value of the other feedback path at that sampling point being already very low, it is not compatible with keeping feedback coefficients as low as possible. Hence the time constrain can be relaxed on the two outer feedback paths. This is good from a design point of view since the most inner loop is closer to the quantizer, hence matching the timing requirement is easier, while the outer loops will physically bee farther away, which will consequently increase their respective ELD. The purpose here will be to evaluate what the individual ELD values can be for each loop.

Because the problem is now three dimensional, a nearly exhaustive search like before cannot be afford. Using a simple gradient descent was attempted but the cost function appears to be not only non-convex, but even sometimes discontinuous. To solve this issue, the problem was approached as follow: First, an exhaustive search only for the most outer loop ELD, let us call it ELD_1, was run while keeping the two other loops' ELD constant, respectively ELD_2 and ELD_3, at the original value. Then, ELD_1 is fixed to the largest value that gives an acceptable solution and an exhaustive search is run on ELD_2. Finally, an optimizer searching for the optimal values of ELD_1 ELD_2 and ELD_3 in a very close neighborhood is used.

The goal is to have as much ELD as possible to ease implementation, hence the ELD starting point is $ELD_2 = ELD_3 = 1.41 \times T_S$ corresponding to the last acceptable deep in Figure 4-44. Then, ELD_1 is scanned between $1.41 \times T_S$ and $2.5 \times T_S$. Ideally, the different ELD should have the following relationship: $ELD_1 > ELD_2 > ELD_3$. Hence, ELD_2 will be scanned across the same range. The results of these two scans are plotted in Figure 4-49. Here, the objective is for the ELDs to be as large as possible, hence the choice was made to fix $ELD_1 = 2.35 \times T_S$ in the last local minimum. For ELD_2, its value is fixed to $2.15 \times T_S$. Even though the optimization final values are not exactly below the threshold value they are still close enough and the subsequent local optimization will recover it.

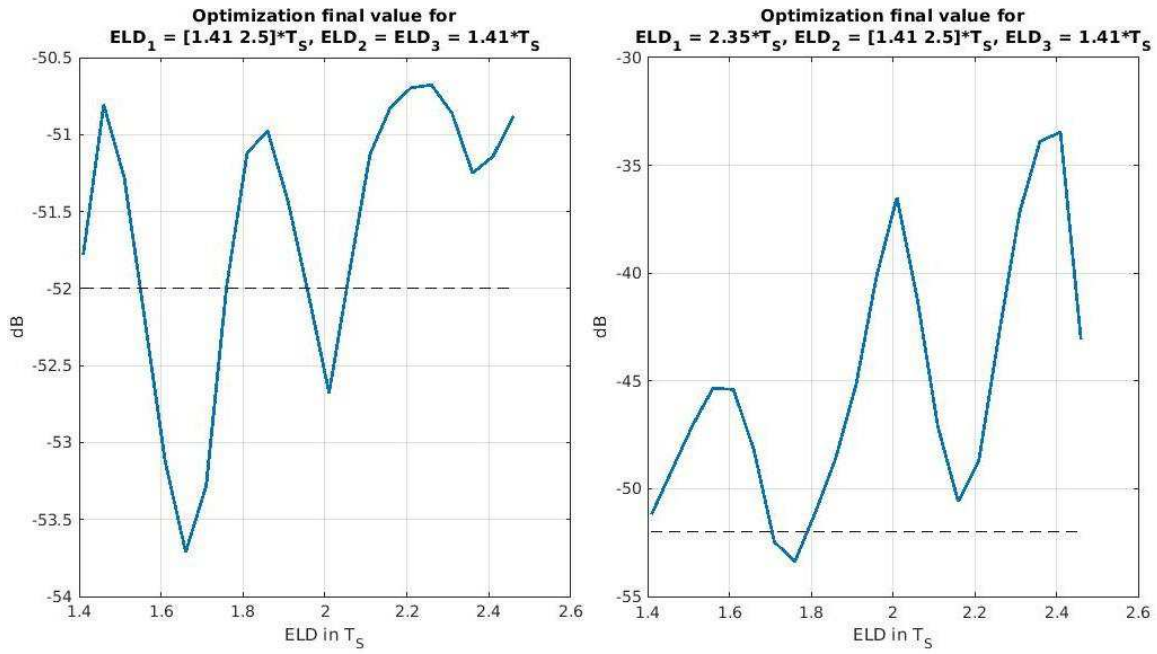

Figure 4-49: ELD_1 and ELD_2 scan for optimization's first step

An interesting outcome of the final local optimization is that even ELD_3 increases from $1.41 \times T_S$ to $1.4594 \times T_S$, which gives a little bit more time to close the most inner loop. ELD_1 and ELD_2 respectively settle at $2.3956 \times T_S$ and $2.1381 \times T_S$ for a cost function final value of $fval = -53dB$, just below the $-52dB$ threshold. This significantly relaxes the timing for these two loops giving some margin for implementation. Lastly, the level of robustness to ELD variation is preserved, as depicted in Figure 4-50.

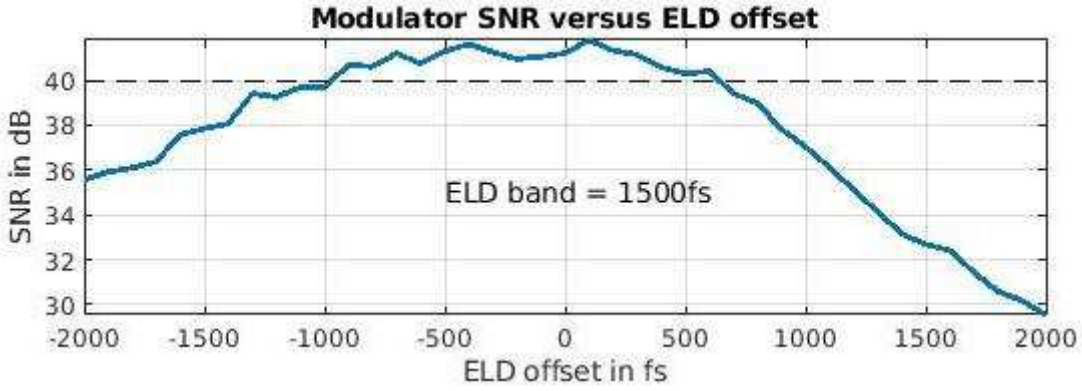

Figure 4-50: ELD robustness test on a modulator with individually optimized

As previously mentioned, the probable non-convexity of the cost function prevents this methodology to ensure optimality. Nonetheless, it provides a systematic method to obtain modulators with performances fitting the needs.

4.4 CONCLUSION

Starting from the observation that one of the major challenges in DBF is the amount of digital processing required, the proposed investigation focused on how this could be alleviated. The nature of SDMs output signals proved to have many interesting properties in that regard, motivating this investigation direction. It started by an analytic study of low order modulators, for the cases of low pass and band pass, as well as discrete and continuous time modulators. Based on this study, the architectural choices were made

toward an RF sub-sampling band pass continuous time sigma-delta modulator-based receiver. This architecture displays many advantages such as relaxed ELD requirement and intrinsic robustness to feedback coefficient random variations. While the ELD requirement is relaxed, it is still very stringent and will be one of the major challenges for implementation. Another challenge will be the timing accuracy, in the pico-second range, required to close the loop. These two points will be carefully addressed in the next chapter which deals with the implementation of the receiver.

4.5 REFERENCES

[4-1] S. Jang, R. Lu, J. Jeong and M. P. Flynn, "A 1-GHz 16-Element Four-Beam True-Time-Delay Digital Beamformer," in IEEE Journal of Solid-State Circuits, vol. 54, no. 5, pp. 1304-1314, May 2019.

[4-2] A. Zanoni, "Toom-Cook 8-way for Long Integers Multiplication," 2009 11th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, Timisoara, 2009, pp. 54-57.

[4-3] H. Inose, Y. Yasuda and J. Murakami, "A Telemetering System by Code Modulation - Δ-ΣModulation," in IRE Transactions on Space Electronics and Telemetry, vol. SET-8, no. 3, pp. 204-209, Sept. 1962.

[4-4] B. Razavi, "The Delta-Sigma Modulator [A Circuit for All Seasons]," in IEEE Solid-State Circuits Magazine, vol. 8, no. 2, pp. 10-15, Spring 2016.

[4-5] Shanthi Pavan; Richard Schreier; Gabor C. Temes, "Understanding Delta-Sigma Data Converters", IEEE, 2017.

[4-6] O. Shoaei and W. M. Snelgrove, "A multi-feedback design for LC bandpass delta-sigma modulators," 1995 IEEE International Symposium on Circuits and Systems (ISCAS), Seattle, WA, USA, 1995, pp. 171-174 vol.1.

[4-7] J. A. Cherry and W. M. Snelgrove, "Excess loop delay in continuous-time delta-sigma modulators," in IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, vol. 46, no. 4, pp. 376-389, April 1999.

[4-8] L. Kull et al., "22.1 A 90GS/s 8b 667mW 64× interleaved SAR ADC in 32nm digital SOI CMOS," 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), San Francisco, CA, 2014, pp. 378-379.

[4-9] L. Kull et al., "A 24-to-72GS/s 8b time-interleaved SAR ADC with 2.0-to-3.3pJ/conversion and >30dB SNDR at nyquist in 14nm CMOS FinFET," 2018 IEEE International Solid - State Circuits Conference - (ISSCC), San Francisco, CA, 2018, pp. 358-360.

[4-10] J. F. Bulzacchelli, Hae-Seung Lee, J. A. Misewich and M. B. Ketchen, "Superconducting bandpass /spl Delta//spl Sigma/ modulator with 2.23-GHz center frequency and 42.6-GHz sampling rate," in IEEE Journal of Solid-State Circuits, vol. 37, no. 12, pp. 1695-1702, Dec. 2002.

[4-11] A. Sekiya, K. Okada, Y. Nishido, A. Fujimaki and H. Hayakawa, "Demonstration of the multi-bit sigma-delta A/D converter with the decimation filter," in IEEE Transactions on Applied Superconductivity, vol. 15, no. 2, pp. 340-343, June 2005.

[4-12] T. Chalvatzis, E. Gagnon, M. Repeta and S. P. Voinigescu, "A Low-Noise 40-GS/s Continuous-Time Bandpass $\Delta\Sigma$ ADC Centered at 2 GHz for Direct Sampling Receivers," in IEEE Journal of Solid-State Circuits, vol. 42, no. 5, pp. 1065-1075, May 2007.

[4-13] T. Chalvatzis and S. P. Voinigescu, "A 4.5 GHz to 5.8 GHz tunable δσdigital receiver with Q enhancement," 2008 IEEE MTT-S International Microwave Symposium Digest, Atlanta, GA, USA, 2008, pp. 193-196.

[4-14] S. Le Tual, P. N. Singh, C. Curis and P. Dautriche, "22.3 A 20GHz-BW 6b 10GS/s 32mW time-interleaved SAR ADC with Master T&H in 28nm UTBB FDSOI technology," 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), San Francisco, CA, 2014, pp. 382-383.

[4-15] J. Wagner, F. Vogel, F. Kuhm and M. Ortmanns, "Automated Synthesis of Subsampling CT Bandpass ΣΔ Modulators with Non-Idealities," 2018 New Generation of CAS (NGCAS), Valletta, 2018, pp. 90-93.

[4-16] S. R. Freeman et al., "Delta-sigma oversampled ultrasound beamformer with dynamic delays," in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 46, no. 2, pp. 320-332, March 1999.

[4-17] N. Beilleau et al., "A 1.3 V, 26 mW 3.2 GS/s undersampled LC bandpass SD ADC for a SDR ISM band receiver in 130 nm CMOS," in IEEE Radio Freq. Integr. Circuits Symp. Dig.

[4-18] J. Ryckaert et al., "A 2.4 GHz Low-Power Sixth-Order RF Bandpass $\Delta\Sigma$ Converter in CMOS," in IEEE Journal of Solid-State Circuits, vol. 44, no. 11, pp. 2873-2880, Nov. 2009.

[4-19] S. Gupta, D. Gangopadhyay, H. Lakdawala, J. C. Rudell and D. J. Allstot, "A 0.8–2 GHz Fully-Integrated QPLL-Timed Direct-RF-Sampling Bandpass $\Sigma \Delta$ ADC in 0.13 μm CMOS," in IEEE Journal of Solid-State Circuits, vol. 47, no. 5, pp. 1141-1153, May 2012.

[4-20] E. Martens et al., "RF-to-Baseband Digitization in 40 nm CMOS With RF Bandpass $\Delta\Sigma$ Modulator and Polyphase Decimation Filter," in IEEE Journal of Solid-State Circuits, vol. 47, no. 4, pp. 990-1002, April 2012.

[4-21] A. Ashry and H. Aboushady, "A 4th Order 3.6 GS/s RF /spl Sigma//spl Delta/ ADC With a FoM of 1 pJ/bit," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 60, no. 10, pp. 2606-2617, Oct. 2013.

[4-22] A. Sayed, T. Badran, M. -M. Louërat and H. Aboushady, "A 1.5-to-3.0GHz Tunable RF Sigma-Delta ADC With a Fixed Set of Coefficients and a Programmable Loop Delay," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 67, no. 9, pp. 1559-1563, Sept. 2020.

4.6 ANNEX 4.1

In this annex is given a simple matlab LTI model of a first order discrete time low pass sigma-delta modulator with $1kHz$ of bandwidth and an OSR of 256. The modulator is technically implemented by the "for loop" at the end.

```
% Simulation parameters
NB_pts = 2^18;
% Modulator's parameters
A_FS = 1;
fmax = 1000;
OSR = 2^8;
fs = 2*fmax*OSR;
% Input signal parameters
fc = 91.7969; % The closest coherent frequency to 100Hz
A_dB_FS = -0.3;
```

```
A = A_FS*10.^(A_dB_FS/20);
% Generate time vector
ts = 1/fs;
t = (0:(NB_pts-1))*ts;
% Generate input signal
x = A*sin(2*pi*fc*t);
% Initialize modulator
w = 0;
y = ones(1,length(x));
% Run modulator
for n = 2:length(x)
    w = x(n) - y(n-1) + w;    % Process current value of w
    y(n) = A_FS*sign(w);      % Single bit quantizer
end
```

5 CHAPTER V: IMPLEMENTATION

Now that the receiver's architecture is established, work must be done toward its implementation. First, the required building blocks will be listed, and some of their implementation details will be provided, like the various possible circuit topologies or the reasons for the choices made in the proposed implementation.

Once the implementation of all the building blocks is clear, they need to be assembled to form the modulator. At this point, an additional round of modulator optimization using inputs from the electrical simulations will have to be done to account for the behavior difference between the ideal model and the implemented solution.

Finally, eight receivers were integrated into a test chip with an on-board memory and a digital interface to form an eight-channel digital beamformer. In the last part, this test chip top view and capabilities will be described as well as the top layout.

5.1 BUILDING BLOCKS TOPOLOGIES

This band pass continuous time sigma-delta modulator is composed of few building blocks: Resonators, feedback DACs and adders, weighting coefficients, a quantizer and a data and a clock distribution tree.

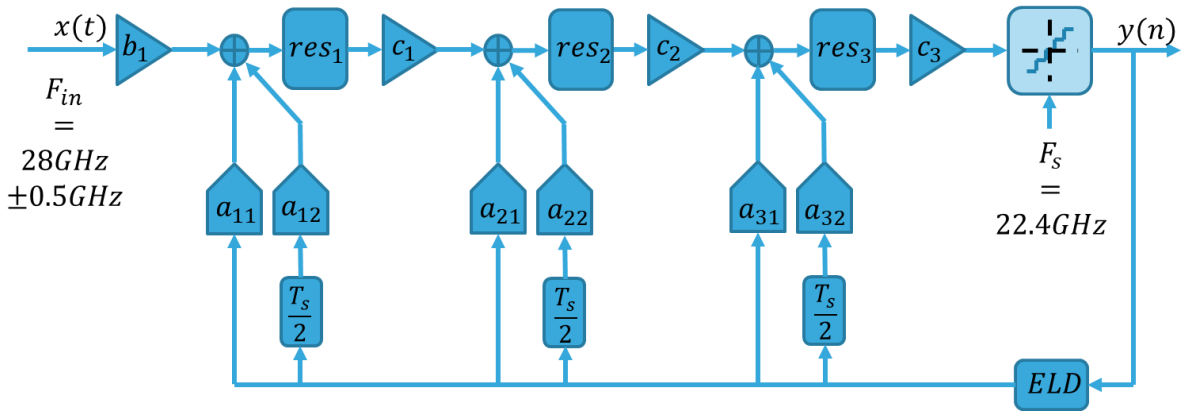

Figure 5-1: Modulators architecture

In the previous chapter, all the simulations were assuming these building blocks to be ideal. While the architecture was chosen to simplify as much as possible the implementation challenges, working with an input frequency F_{in} of $28GHz$ with $1GHz$ of bandwidth and sampling frequency F_s of $22.4GHz$ remains very ambitious and will require very careful design. Here, the blocks requiring lower complexity for their implementation will be treated first. The higher complexity blocks will be dealt with in a second time. In this section, the focus will be put on the different used topologies. The sizing methodology will be described in the next section. First, the feedforward weighting coefficients, b_1 and c_1 through c_3 will be discussed. Then, the analysis will focus on the feedback DACs, weighting coefficient, and the adders, which are all three implemented as one component. The following building block will be the resonators. To complete the analog path, the quantizer's implementation will be discussed. The last piece of this puzzle is the high-speed clock and data distribution tree required to drive the feedback DACs from the quantizer's output.

5.1.1 Feedforward weighting coefficients

Four different coefficients must be implemented, and it will be done in four different ways. The first coefficient is b_1. It will be produced by the input matching network. Coefficient c_1 and c_2 are realized in a similar manner, a gm-cell pushing current into the resonator. For input matching and noise reasons,

c_1 needs to be implemented with some differences and is a little bit more complex. Hence, between the two, the implementation of c_2 will be described first. Finally, c_3 simply corresponds to the quantizer's gain and will be discussed in the related section.

5.1.1.1 Coefficient b_1

The b_1 coefficient has no impact over the loop characteristics. For this reason, it is not necessary to control it accurately. In the present case, it will be resulting from the input matching network, so there is little margin to adjust it. Since it affects the input dynamic range, it must be properly evaluated to ensure the final receiver has the desired input dynamic range. If this metric were to be affected too much, the remaining parts of the modulator could be adjusted to recover the desired performances. Thankfully, this will not be necessary.

5.1.1.2 Coefficient c_2

Because it is simpler, the implementation of the c_2 coefficient will be discussed first. The principle schematic is given in Figure 5-2-a, and its transfer function is given in equation (5.1)

$$H(s) = \frac{gm}{C} \times \frac{s}{s^2 + \frac{1}{R \times C} \times s + \frac{1}{L \times C}} = c_2 \times f_s \times \frac{s}{s^2 + \frac{1}{R \times C} \times s + \frac{1}{L \times C}} \quad (5.1)$$

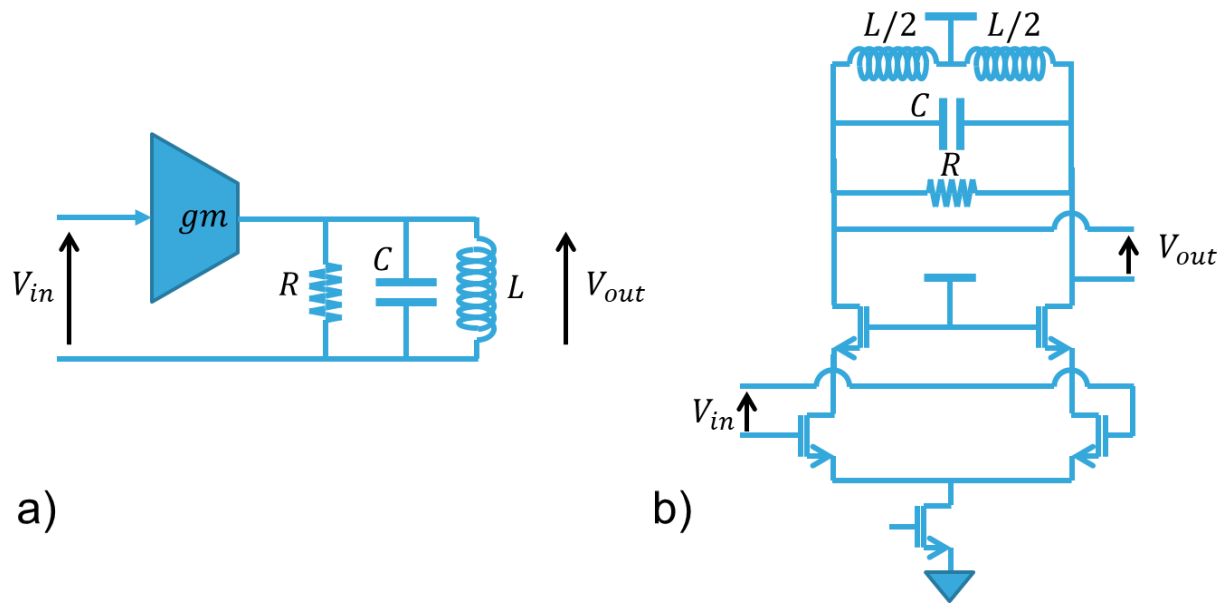

Figure 5-2: a) Implementation principle schematic of c_2. b) Transistor level schematic.

The coefficient is realized by the ratio between gm and C. To be precise, it would be the case if the sampling frequency were unity. To get the effective coefficient, this ratio must be normalized with the sampling frequency, $c_2 = \frac{gm}{C} \times \frac{1}{f_s}$. The sampling frequency being fix, the coefficient value can be adjusted through gm or C. One important note here is that more gm means more power. Hence to minimize power consumption, it desirable to have a smaller C. It will be seen that the feedback DAC will also impose some constraints on the value of C.

Figure 5-2-b provides the transistor level schematic. It is a classic differential pair. It is cascoded for two main reasons. First, it greatly limits the Miller effect, providing a more stable input impedance. Second, it increases the output impedance of the gm-cell, minimizing the resonator's quality factor degradation. For simplicity, the cascodes are biased directly with the power supply.

One challenge is to have a good control of c_2. The strategy is to be able to adjust it by controlling the amount of gm through the biasing current. Here, the target is to achieve a tuning range of about $\pm 30\%$ to compensate for process variation. Very often this tuning is obtained only by adjusting the common mode of the input voltage. The gain transistor then also plays the role of a current source. This has the merit of simplifying the circuit.

There are few issues with this approach. First, this leads to a poor Common Mode Rejection Ratio (CMRR), which could eventually make the loop unstable. Second, the requirements when sizing a transistor for RF performances are not compatible with the ones when sizing a current source. In the first case, as much gm as possible is needed while minimizing the parasitic capacitance. This translates into using transistors with small channel length. For a current mirror, a good output impedance and a large device for better matching are desired, which translate into transistors with large channel length. Since RF performances cannot be given up, it means the transistors' trans-conductance will be poorly controlled, leading to a poor control of c_2.

Therefore, the choice was made to control the operating point with a current source. Because of the low power supply of $1V$ in CMOS 28nm FDSOI, it is impractical to stack more than three transistors without risking depolarizing one of them. This could be done by using a higher voltage supply but handling two different supplies and ensuring safe operating area for all devices at all-time adds a complexity that is to be avoided if possible. Hence, the current source is implemented without a cascode. To improve its output impedance, it is made using a long channel device. Also, to improve the control of the bias current, the common mode of the input signal is adjusted such that the current source has the same V_{gs} and V_{ds} as the current mirror. To improves the robustness to Process and Temperature Variations (PVT), a main bias generator with a constant gm characteristic is used. The goal with this topology is to achieve the desired performances in terms of c_2 control and stability with the lowest possible complexity on the signal path.

5.1.1.3 Coefficient c_1

The principle schematic is the same as for c_2. The difference is in the implementation of the gm-cell. Since this is the first active stage, it will be part of the input matching equation. The gm-cell of c_2 has an input impedance that is essentially capacitive. Hence it will never present a real part close enough to 50Ω. The simplest solution is to add a 50Ω resistor in parallel of the input as well as an inductor to compensate the imaginary part brought by the input parasitic capacitor of the gm-cell. Unfortunately, this simple solution suffers from poor noise performances. The best achievable Noise Figure (NF) with such matching strategy is 3dB since the noise power of the matching resistor would be equal to the source noise power, hence doubling the overall input noise power.

RF designers have developed several Low Noise Amplifier (LNA) topologies allowing significantly better noise performances. The most widely used today is called inductively degenerated common source LNA. A differential implementation of this topology is depicted in Figure 5-3-a. As its names suggests, this circuit topology has its gain transistor connected as a common source and is degenerated by an inductor. This creates a feedback that leads to the appearance of a real part in the input impedance. Cancelling the remaining imaginary part is generally done by adding a series inductor on the input. This topology is very common because its real part is nearly frequency independent and has good noise and gain performances.

For the target application this topology has two major drawbacks. The first one is that the objective is to use the matching network also as the first resonator. For that purpose, something which behaves as a parallel RLC circuit, i.e. the gain reaches a maximum at the resonant frequency and decreases when away from that frequency, is required. The input impedance of an inductively degenerated common source LNA, when matching is done by a series inductor, behaves as series RLC network, i.e. the gain reaches a minimum at the resonant frequency and increases when away from that frequency. This is

incompatible with the resonator's characteristics needed for the modulator. A possible solution is to perform matching using a parallel element, but then the real part of the input impedance is not independent of frequency anymore. This makes this topology less attractive. The second drawback is that the feedback generated by the degeneration renders the effective trans-conductance GM nearly independent of the biasing current at the resonant frequency. This is incompatible with the current strategy for tuning c_1 through the biasing current. For these reasons it is necessary to use a different topology.

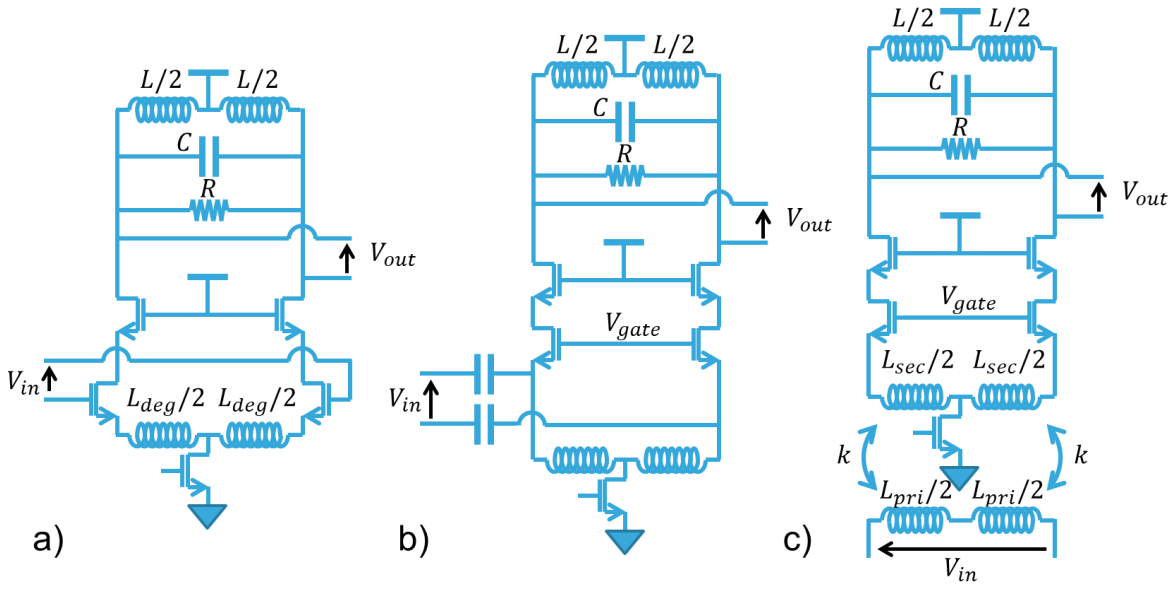

Figure 5-3: a) Inductively degenerated common source topology. b) Common gate topology with capacitive coupling input signal feed. c) Common gate topology with inductive coupling input signal feed.

Another possibility is to use a common gate circuit. Its input impedance is inversely proportional to gm, providing a nice controllable real part. One of the challenges with this topology is to feed the signal to the gain transistor source. The simplest solution is to use a coupling capacitance. Figure 5-3-b depicts a differential implementation of this circuit. With this approach the value of gm is set to provide the desired 50Ω input impedance. As a consequence, little margin would be left when adjusting c_1 using gm. An alternate solution is to feed the signal through a transformer as shown in Figure 5-3-c. With this approach, the transformation ratio can be used as an additional design parameter allowing for a wider range of acceptable gm values.

The transformer feed approach also brings few additional advantages. In the proposed system analysis, in the previous chapters, antenna arrays made of patch antennas were considered. These devices are intrinsically single ended. This input transformer is just the ideal place to put a balun to convert the single ended input into a differential one. A second advantage is that a transformer naturally provides some level of Electro-Static Discharge (ESD) protection. This means the amount of protection required is reduced, also reducing their parasitic capacitance. Last, Transformers are less susceptible to magnetic coupling with their surrounding environment. In a system where the goal is to integrate multiple receivers on the same die, that can only be a good feature to have.

These benefits are coming at the cost the reduced performances in noise and power efficiency of the common gate topology. In order to reduce the performance loss, it possible to use a gm-boosted common gate topology. The idea is to feed the signal on both the source and the gate of the transistor, with opposite phase. This way the gain transistors sees more swing, improving the gain of the stage. An early implementation [5-1] performed this opposite phase feeding using crossed coupled capacitor on a

differential common gate. This implementation allows at best to double the swing seen by the gain transistor. In [5-2] they propose an implementation using a transformer to perform this gate cross feeding, while the source remained on a capacitive feed. Using a transformer allows to use the transformation ratio to perform passive voltage gain, potentially improving noise and power performances.

A third implementation was proposed in [5-3] where both the source and the gate are fed through a three-way transformer. It is this last option that have been chosen and is depicted in Figure 5-4. As discussed before, the input transformer is also used as a balun. It is important to note that this input transformer plays multiple roles in the modulator. It is an important component of the matching network, it provides some ESD protection, performs the single end to differential conversion and will determine the b_1 coefficient. It must also provide some passive voltage gain to the input transistors gates. Finally, it will act as the modulator first resonator. Clearly, this is a key component on the proposed design and will require a very careful design.

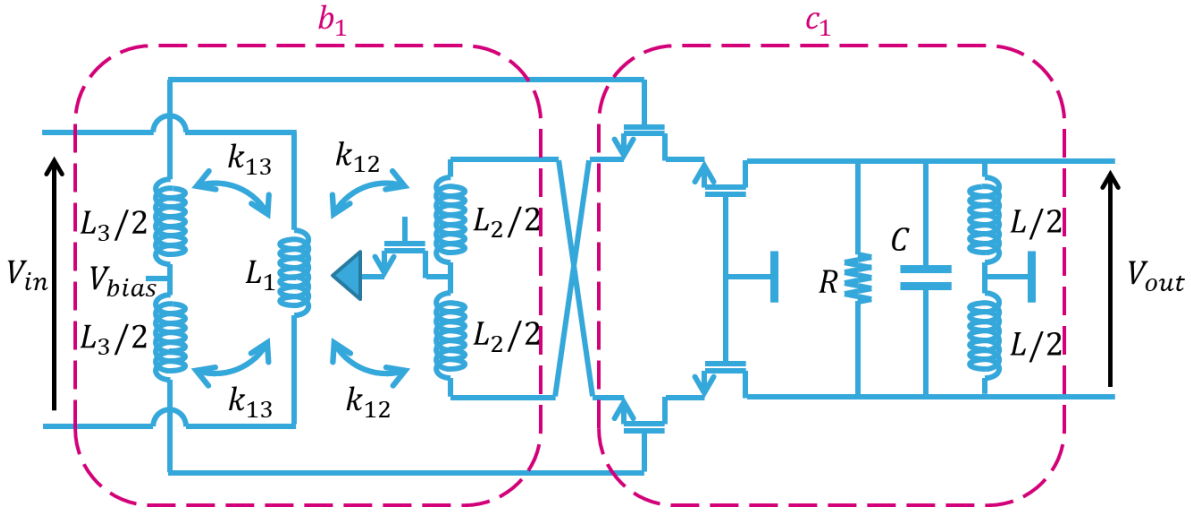

Figure 5-4: gm-boosted common gate LNA with tranformer feed

With this topology, the same biasing and tuning strategy as for c_2, tuning the trans-conductance through the biasing current, can be used. This allows to reuse many biasing blocks and simplify the design.

5.1.1.4 Coefficient c_3

The implementation of this coefficient is much simpler. It directly depends on the full scale of the quantizer. A 1.5-bit quantizer has two comparison levels. The c_3 coefficient can be adjusted by tuning these levels. This will be described in greater details during the quantizer description.

The implementation of all the feedforward coefficients have now been described. Even though, at a higher level they seem to be of a similar nature, their implementations are all very different. Next, the feedback coefficients will be discussed, together with the feedback DACs.

5.1.2 Feedback DACs

When considering the implementation of feedback DAC in SDM, one must consider two problems. The first one is how to implement the DAC itself and the second is how to sum its output on the signal path. The second problem will actually impose some constraints on the DAC implementation; hence it will be treated first.

5.1.2.1 Feedback summing method

Many implementations of feedback DACs were done using a current steering DAC ([4-1],[4-7]). It is well suited for high-speed operation and multi-bit modulators. In the case of a 1.5-bit quantizer, it might

be overdoing it. The complexity of the steering DAC circuitry in a low supply voltage deep sub-micron CMOS technology is somewhat high and its performances unnecessary. To understand why, let us look at its implementation in Figure 5-5.

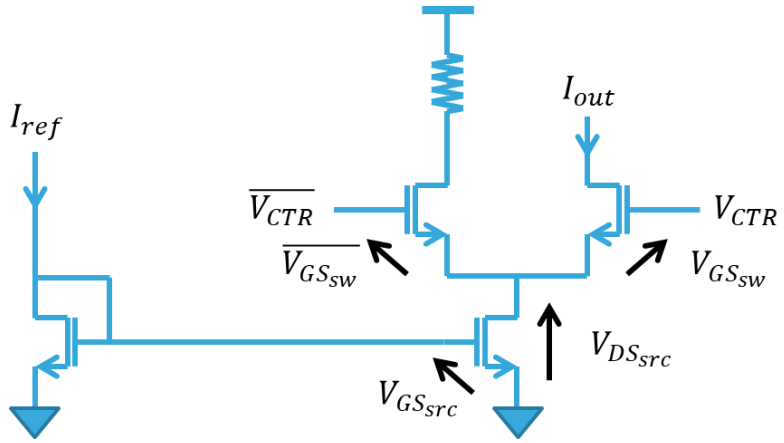

Figure 5-5: Simple implementation of a current steering DAC

It requires a current source and current switch to steer the source current in the feedback node or in a dummy load. Keeping the current flowing in the source when there is no current at the output allows for faster switching time, at the cost of higher power consumption. Stacking these two devices is by nature badly suited for low voltage processes. On one hand, the current source is implemented using a current mirror. To have a good matching on the copied current the transistors over drive voltage must be large. This reduces the sensitivity to transistor's threshold voltage variation. This large over drive voltage impose a large drain to source voltage ($V_{DS_{src}}$) to keep the transistor in saturation. On the other hand, the current switch is implemented by two transistors stacked on the current source and controlled by signals with opposite polarities. For these devices to be good switches, they need a large gate to source voltage ($V_{GS_{sw}}$). As a consequence, the gate control voltage $V_{CTR} = \mathrm{V_{DS_{src}}} + V_{GS_{sw}}$ must be even larger, and often require voltage beyond the typical power supply of low voltage technologies. One solution to this lack of voltage headroom is to use a second higher voltage supply. This has been done many times but adds a lot of complexity, in particular in ensuring the devices safe operating area at all times. As in section 5.1.1.2, this is to be avoided if possible.

Instead, the choice was made to use the Capacitively coupled Voltage DAC, or CVDAC, proposed by the authors in [5-4]. For a 1.5-bit resolution, it can be implemented by a circuit topology close to a CMOS digital gate. This is much simpler to implement in a digital process such as the targeted one.

The previous chapter analysis was assuming that a current DAC (IDAC) such as the one in Figure 5-6-a was used. Let us first process the transfer function and then compare it with the one of a CVDAC. First, the impedance of the resonator must be computed. The result is given in equation (5.2).

$$Z_{RLC}(s) = \frac{\frac{s}{C}}{s^2 + \frac{1}{R \times C} \times s + \frac{1}{L \times C}} = \frac{\frac{s}{C}}{s^2 + \frac{\omega_0}{Q} \times s + \omega_0^2} \tag{5.2}$$

The transfer function is then the product of this impedance with the IDAC current:

$$H_{IDAC}(s) = I_{DAC} \times Z_{RLC}(s) = \frac{I_{DAC}}{C} \times \frac{s}{s^2 + \frac{1}{R \times C} \times s + \frac{1}{L \times C}} \tag{5.3}$$

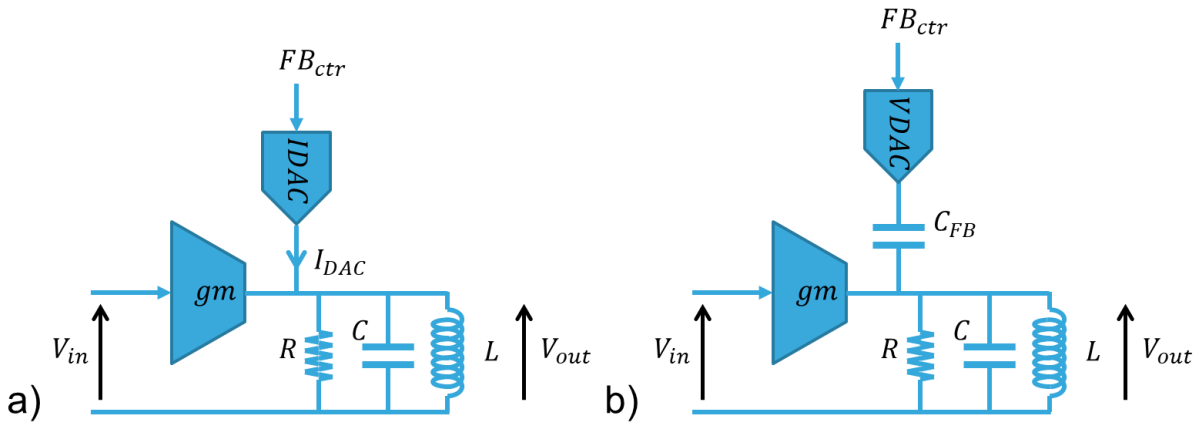

Figure 5-6: a) Classic current steering feedback DAC. b) Capacitively coupled Voltage feedback DAC

The second term is due to the resonator and the first one is the desired feedback coefficient, equal to the ratio between the DAC current and the resonator's capacitance.

The CVDAC from Figure 5-6-b can simply be seen as a voltage divider. The transfer function is then easily obtained in equation (5.4):

$$H_{CVDAC}(s) = \frac{Z_{RLC}(s)}{Z_{RLC}(s) + \frac{1}{s \times C_{FB}}} = \frac{s \times C_{FB}}{C_{FB} + C} \times \frac{s}{s^2 + \frac{1}{R \times (C_{FB} + C)} \times s + \frac{1}{L \times (C_{FB} + C)}} \tag{5.4}$$

As for the current DAC, this transfer function can be written as the product of two terms. As before the second one corresponds to the resonator. The DAC capacitance is now a part of the resonator's capacitance. The first term of (5.4) is the desired feedback coefficient. One can note here that this coefficient is no more frequency independent. It has a zero at the zero frequency. This means the feedback coefficient will be correct only for a single frequency. One could wonder if that affects the modulator's performances.

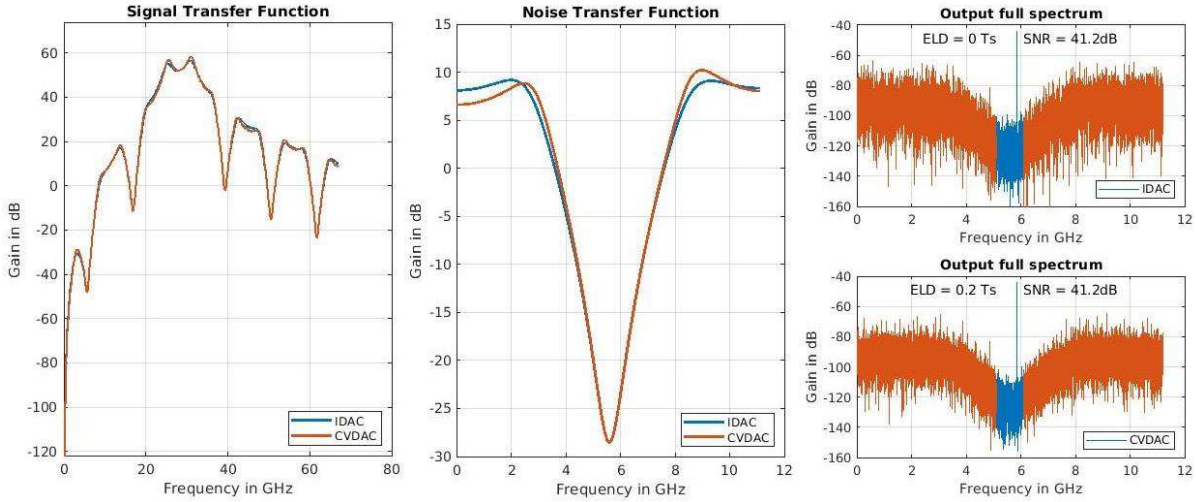

Figure 5-7: Comparison between an IDAC and CVDAC modulator

Another difference is that, for $s = j \times \omega$ the feedback coefficient is now imaginary. This means that the feedback signal will have a 90° phase lead. Here comes a very interesting feature of the CVDAC. To compensate this phase lead one can simply delay the feedback signal by a quarter of center frequency's period. When working in the third Nyquist Zone, this delay corresponds to $0.2 \times T_S$. Indeed, adding these two modifications in the high-level model, an additional zero in the feedback path and a delay of $0.2 \times T_S$, provides a functional modulator. The performance comparison is plotted in Figure 5-7. The STF and NTF are slightly different but still very acceptable and the SNR is identical.

If the optimizer is run the STF and the NTF can be completely recovered, as shown in Figure 5-8. With that answer, the viability of a capacitive feedback DAC is confirmed. It was mentioned in section 5.1.1.2 that the DAC would also impose a constraint on the resonator's capacitance. Since the feedback coefficient is proportional to the ratio of the feedback and the resonator's capacitances, implementing small coefficient means small feedback capacitance. Because there is a limit on how small it is reasonable to make a capacitor, it may be required to increase the resonators capacitance, for the feedback capacitor to be implementable. That would be detrimental for the power efficiency of coefficients c_1 and c_2 implementation. It will be seen in section 5.1.3, that this problem can be mitigated with the correct choice of resonator. The next step is to design the topology of the voltage DAC that will drive this feedback capacitor.

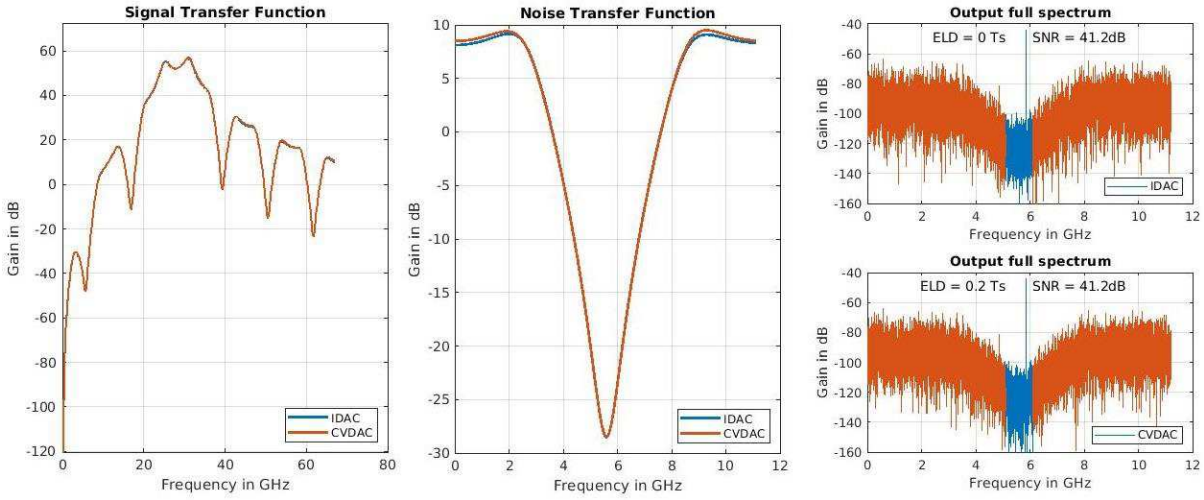

Figure 5-8: Comparison between an IDAC and CVDAC modulator after optimization

5.1.2.2 Return to Zero Voltage DAC topology

As mentioned before, to realize a 1.5-bit differential voltage DAC, a circuit topology close to a CMOS digital gate can be used. Two of them are needed to generate a differential signal. The biggest challenge is to accommodate for the $22.4Gsps$ sampling rate, especially with a return to zero pulse shape. This is at the edge of what the technology can do, and the DAC must be design very carefully to reach this level of performance. Its topology is given in Figure 5-9 and it behaves as follow: When the output (O_P, O_N) is the voltage pair $(VDD, 0)$, that correspond to a logical one. When it is $(0, VDD)$, that correspond to a logical minus one. To realize a 1.5-bit DAC, it is necessary to provide a third level corresponding to a logical zero. This is done by adding a switch between the outputs, to short them together. At that point, the other transistors must be off to avoid shorting the power supply. If the load on the output is properly balanced, the differential output should be zero with a common mode around $VDD/2$. To have this behavior, the appropriate logic to control the gates of the six transistors composing the DAC must be used. It is also through this logic that the Return to Zero (RZ) functionality is implemented. The control logic proposed in Figure 5-9 assumes that the input is given in thermometer code and that the control signals and their conjugates are available.

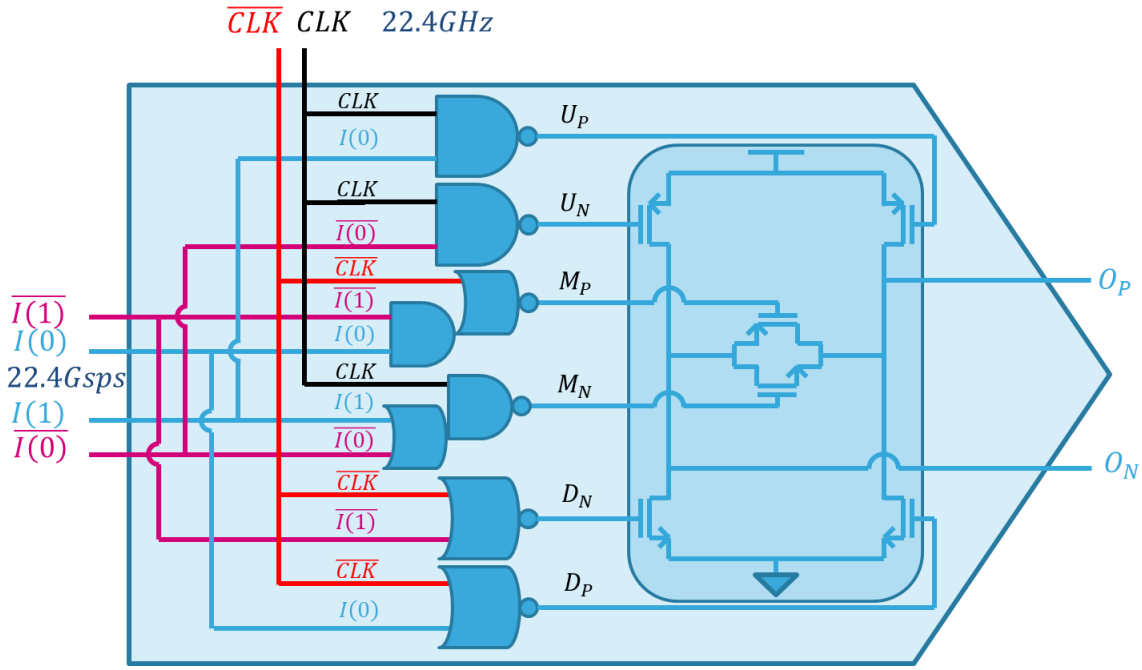

Figure 5-9: Feedback RZ Voltage DAC topology

The DAC's chronograms are given in Figure 5-10. When the clock is low, the output is set to a logical zero, and when the clock is high, the output is control by the input code. As long as the data is stable during the high pulse of the clock, the output signal's rising and falling edges will be driven by the clock's rising and falling edges. It is only necessary to ensure that the transition between two successive input data happens when the clock is low, but the stringent feedback timing constraint only applies to the clock, not to the data. In other words, the $22.3ps$ clock pulse must fit into the $44.6ps$ data window. This is very important for implementation. The data will come from a comparator. Its decision time may vary with the input signal. It is then difficult to ensure a precise timing for the data. While fitting the clock pulse within the data window remains a challenge at the considered frequencies, when this can be done, the proposed implementation output timing is set only by the clock's edges. This gives some level of tolerance to the data timing variation.

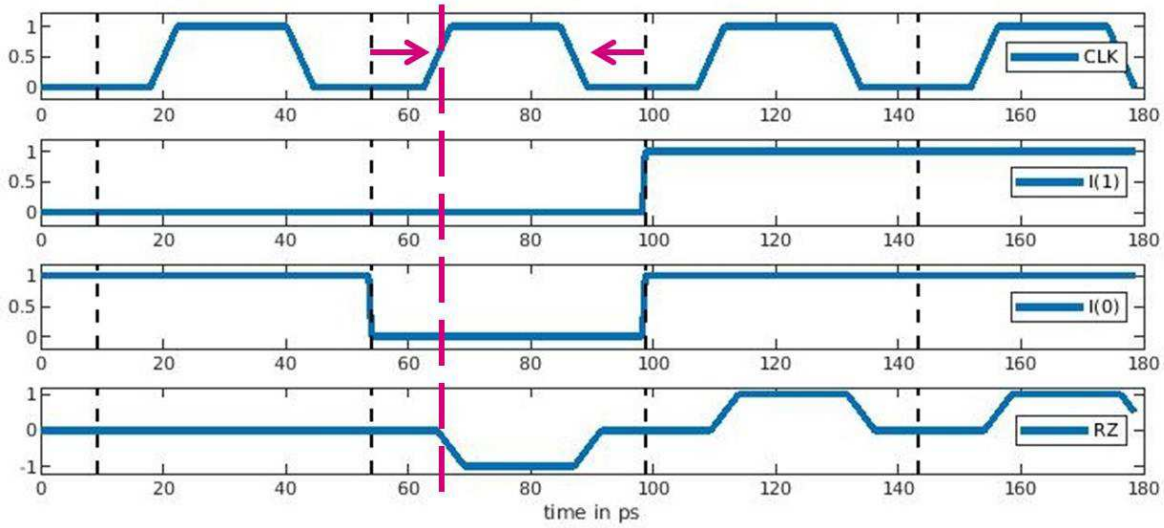

Figure 5-10: RZ Voltage DAC chronograms

The last point that needs to be addressed is the implementation of the Half delayed Return to Zero (HRZ) DAC. With this topology it is extremely simple, it is only required to cross the clock and delay the input data to ensure the clock high pulse happens when the data is stable. The assembly of the RZ and HRZ DACs is given in Figure 5-11. The delay of the data is simply realized with inverters. If the appropriate number of inverters to be used is odd, it is only required to cross the input data and its conjugate.

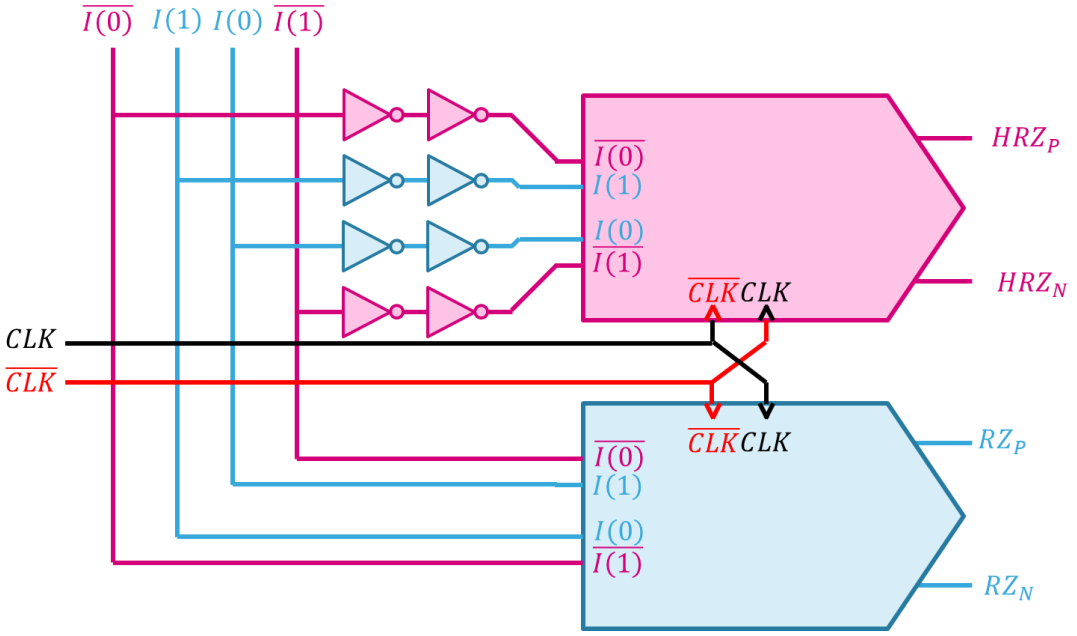

Figure 5-11: RZ and HRZ Voltage DAC assembly

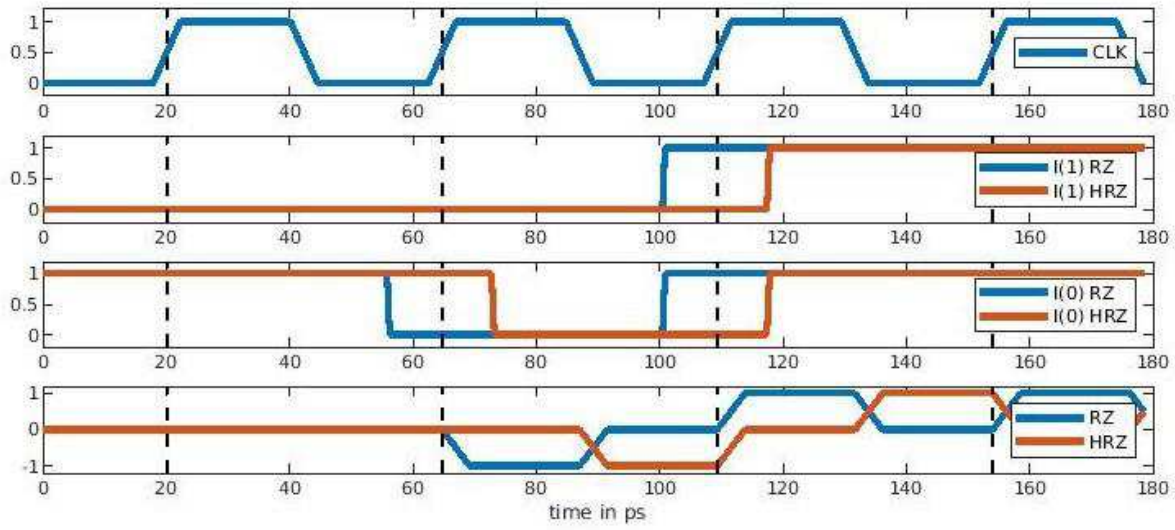

Figure 5-12: HRZ and RZ Voltage DAC assembly chronograms

The chronograms are given in Figure 5-12. As expected, the output of the HRZ DAC is delayed by half a clock cycle compared to the RZ DAC output. One limitation of this topology is its sensitivity to the clock duty cycle. If it is not 50%, the RZ and HRZ pulses will be of different width. This will have to be taken into account when designing the clock distribution network, in particular the CML to CMOS stage receiving the external $22.4GHz$ sinewave clock. Other than that, this topology is mainly made of

digital devices which are well suited for the target process. This will allow to operate at the desired clock rate.

5.1.3 Resonator topology

The resonator architecture has already been discussed in the previous chapter, and it was concluded that only a gm-LC based resonator could fit the frequency requirement. The gm part of it was already discussed, the next step is to take care of the LC part. The most common implementation is to use an inductor and tune the resonant frequency with a parallel capacitor to form an LC-tank. The output is the resulting voltage from the current pushed into the tank by the gm-cell.

In the present case, the choice was made to use transformers instead of inductors as depicted in Figure 5-13. The reasons to have a transformer on the first resonator were already given in section 5.1.1.3. Although impedance matching or single ended to differential conversion are not needed for the subsequent resonators, a transformer-based resonator would still benefit from the reduce magnetic coupling with its surrounding and a convenient middle point for biasing. It will be shown that the transformation ratio can provide an essential additional design parameter. It will allow to adjust conveniently the components' values to ease their physical implementation without compromising power efficiency.

To analyze the circuit, a detailed analysis of the resonator is needed. In a first step, the equation of the different transfer functions of such a resonator will be provided. Then, the behavior of the feedforward transfer function will be studied. Afterwards, the feedback transfer functions will be studied. Finally, the Q-enhancement circuit allowing to reach the desired quality factor will be discussed as well as how it can be tuned.

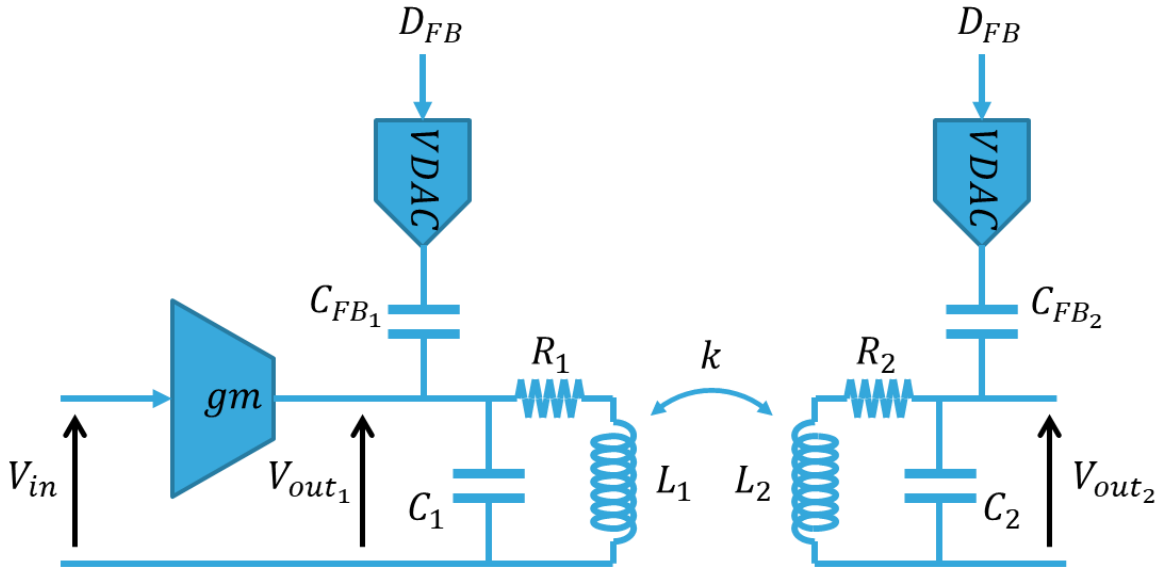

Figure 5-13: Proposed Transformer based gm-LC resonator

5.1.3.1 Resonator's transfer functions analytic equations

The chosen implementation is differential, but for the sake of simplicity, only the equations for a single ended equivalent will be provided. Four transfer functions will be processed, two from the input to the first ($Vout_1$) and second outputs ($Vout_2$), $H_{FF_1}(s)$ and $H_{FF_2}(s)$, one from the first feedback DAC to the first output, $H_{FB_1}(s)$, and the last one, from the second feedback DAC to the second output, $H_{FB_2}(s)$. To do so, the impedances at V_{out_1} and V_{out_2} are processed as well as the transformer voltage transfer function $H_{TF}(s)$ between V_{out_1} and V_{out_2}. These equations are derived under the approximation that the gm-cell has an infinite output impedance and that the voltage DAC has a null output impedance.

$$Z_{out_1}(s, C_{tot_1}, C_{tot_2}) = \frac{R_1 + Z_{L_1}(s, C_{tot_2})}{1 + s \times C_{tot_1} \times \left(R_1 + Z_{L_1}(s, C_{tot_2})\right)} \tag{5.5}$$

With:

$$Z_{L_1}(s, C_{tot_2}) = s \times L_1 \times (1 - k^2) \times \left(\frac{s^2 + \frac{R_2}{(1-k^2) \times L_2} \times s + \frac{1}{(1-k^2) \times L_2 \times C_{tot_2}}}{s^2 + \frac{R_2}{L_2} \times s + \frac{1}{L_2 \times C_{tot_2}}} \right)$$

$$Z_{out_2}(s, C_{tot_1}, C_{tot_2}) = \frac{R_2 + Z_{L_2}(s, C_{tot_1})}{1 + s \times C_{tot_2} \times \left(R_2 + Z_{L_2}(s, C_{tot_1})\right)} \tag{5.6}$$

With:

$$Z_{L_2}(s, C_{tot_1}) = s \times L_2 \times (1 - k^2) \times \left(\frac{s^2 + \frac{R_1}{(1-k^2) \times L_1} \times s + \frac{1}{(1-k^2) \times L_1 \times C_{tot_1}}}{s^2 + \frac{R_1}{L_1} \times s + \frac{1}{L_1 \times C_{tot_1}}} \right)$$

$$H_{TF}(s, C_{tot_2}) = \frac{V_{out_2}}{V_{out_1}} = \frac{\frac{s}{C_{tot_2}} \times k \times \sqrt{\frac{L_1}{L_2}} \times \left(1 + s \times R_2 \times C_{tot_2}\right)^2}{\left(s^2 + \frac{R_2}{L_2} \times s + \frac{1}{L_2 \times C_{tot_2}}\right) \times \left(R_1 + Z_{L_1}(s, C_{tot_2})\right)} \tag{5.7}$$

From these equations, the desired transfer functions can be computed:

$$H_{FF_1}(s) = gm \times Z_{out_1}(s, C_1 + C_{FB_1}, C_2 + C_{FB_2}) \tag{5.8}$$

$$H_{FF_2}(s) = gm \times Z_{out_1}(s, C_1 + C_{FB_1}, C_2 + C_{FB_2}) \times H_{TF}(s, C_2 + C_{FB_2}) \tag{5.9}$$

$$H_{FB_1}(s) = \frac{1}{1 + s \times Z_{out_1}(s, C_1, C_2 + C_{FB_2}) \times C_{FB_1}} \tag{5.10}$$

$$H_{FB_2}(s) = \frac{1}{1 + s \times Z_{out_2}(s, C_1 + C_{FB_1}, C_2) \times C_{FB_2}} \tag{5.11}$$

Even though this model is simplified, the equations are too complex for a direct analysis. The final model will come from an electro-magnetic extraction of the transformers layout, hence, the inductances and capacitances values of this model that would providing a good resonator are not directly necessary. The purpose here is to get the behavioral trends of transformer based resonators. Using the equations derived above, a numerical sensitivity analysis will be made by varying the different parameters individually, and studying the impact on the resulting transfer functions. The feedforward transfer functions will be studied first and then the feedback ones.

5.1.3.2 Feedforward transfer functions

Before studying the sensitivity to the different parameters, their initial values must be set. Let us start with a simple case as a reference, where both inductors and capacitors have the same values. The parameters are set as follow. First, gm is initialized with a realistic value based on the available knowledge of the technology and the power consumption budget. A value of $gm = 15mS$ will be used. Next are the inductors. It is necessary to initialize them with a useable value for integrated inductors at 28GHz. Here, the choice was $L_1 = L_2 = 400pH$. Integrated transformers at 28GHz can generally not

achieve coupling factors close to one. The maximum achievable value is around 0.8. It is set here at $k = 0.7$ such that, later on, it can be varied in both directions. Afterward, the inductors quality factors Q_1 and Q_2 need to be set. The resonator's target quality factor Q_{res} is 30, assuming the definition $Q_{res} = f_{res}/BW_{res}$, where f_{res} is the resonant frequency and BW_{res} is the three-decibel bandwidth. Q_1 and Q_2 are experimentally set such that $Q_{res} = 30$. Finally, the capacitances are adjusted such that the resonating frequency is $28GHz$. All the model's parameters are summarized in Table 5-1.

Table 5-1: Model parameters' values

gm	C_1	L_1	Q_1	R_1	k	L_2	Q_2	R_2	C_2
15mS	47.52fF	400pH	17.65	3.99Ω	0.7	400pH	17.65	3.99Ω	47.52fF

The resulting transfer functions are plotted in Figure 5-14. They are compared with the parallel RLC resonator used so far. On Figure 5-14 left graph, one can note that the transformer-based resonator exhibits a second resonance at higher frequencies. This is not an issue as long as it remains high enough in frequency. On the signal path there will be other components such as the gm-cells or the quantizer that will exhibit low pass characteristics, attenuating this second resonance.

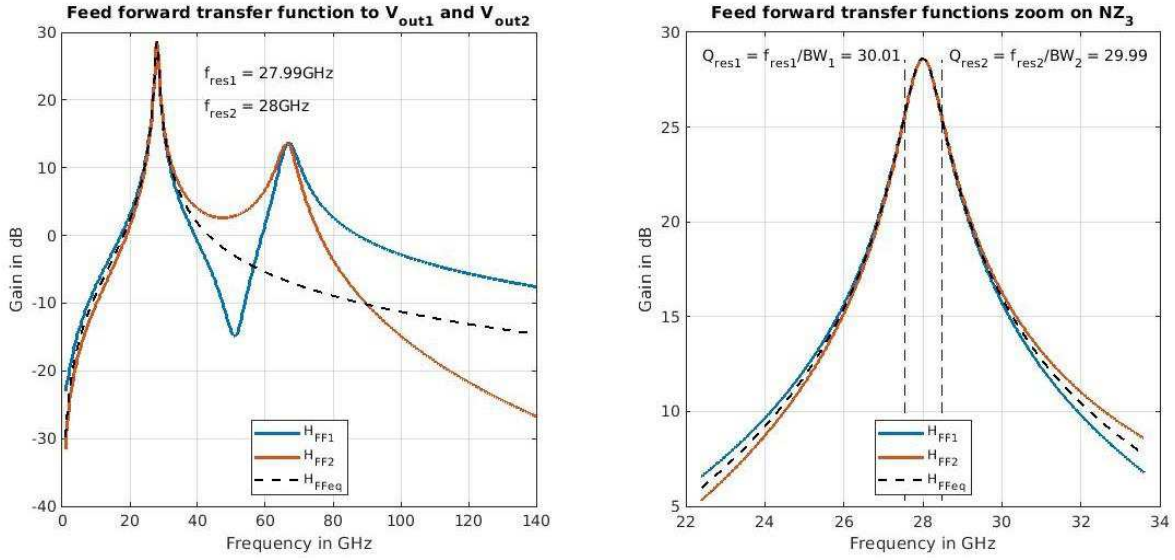

Figure 5-14: Feedforward transfer functions of a transformer-based resonator

It can also be noted that for H_{FF_2}, once beyond the second resonance, the decreasing slope seems steeper than the equivalent RLC resonator. The asymptotic analysis of (5.9) implies that the final slope will be the same as an RLC resonator. Experimentally, it is shown here that this slope will be shifted down significantly. This means that, far away from the resonant frequencies H_{FF_2} will provide better attenuation compared to an RLC resonator, although keeping the same asymptotic slope. While the second resonance may be detrimental from an anti-aliasing standpoint, this will provide better anti-aliasing properties for higher image frequencies. Even though it is rarely a problem in practice, it is still an interesting addition to the proposed transformer-based approach.

When zooming in around 28GHz shows that both transfer functions are very similar to the RLC resonator in the vicinity of the first resonant frequency. From that observation, it is likely that transformers are useable. This will be confirmed by simulations later. This parallel RLC resonator will be called the local equivalent. The equivalent inductance is $L_{eq} = 340pH$ with a quality factor of 30 and the equivalent capacitor is $C_{eq} = 95fF$. Here, yet another advantage of transformer-based resonators is highlighted. To achieve the same resonator quality factor, one needs inductors of a significantly lower quality factor when using a transformer. The used transformer model is too simple

to really quantify the improvement, but the effect is real and significant enough to be used. It has been used by the authors in [5-5], and many after them, to improve VCO's phase noise. This will be helpful to reduce the power consumption of the Q-Enhancement circuit.

Let us now study the impact of L_1, L_2, C_1 and C_2 variations. The results for $H_{FF_1}(s)$ are plotted in Figure 5-15, and for $H_{FF_2}(s)$ in Figure 5-16. When L_1 or L_2 are varied R_1 and R_2 are adjusted such that the inductors quality factors remain the same.

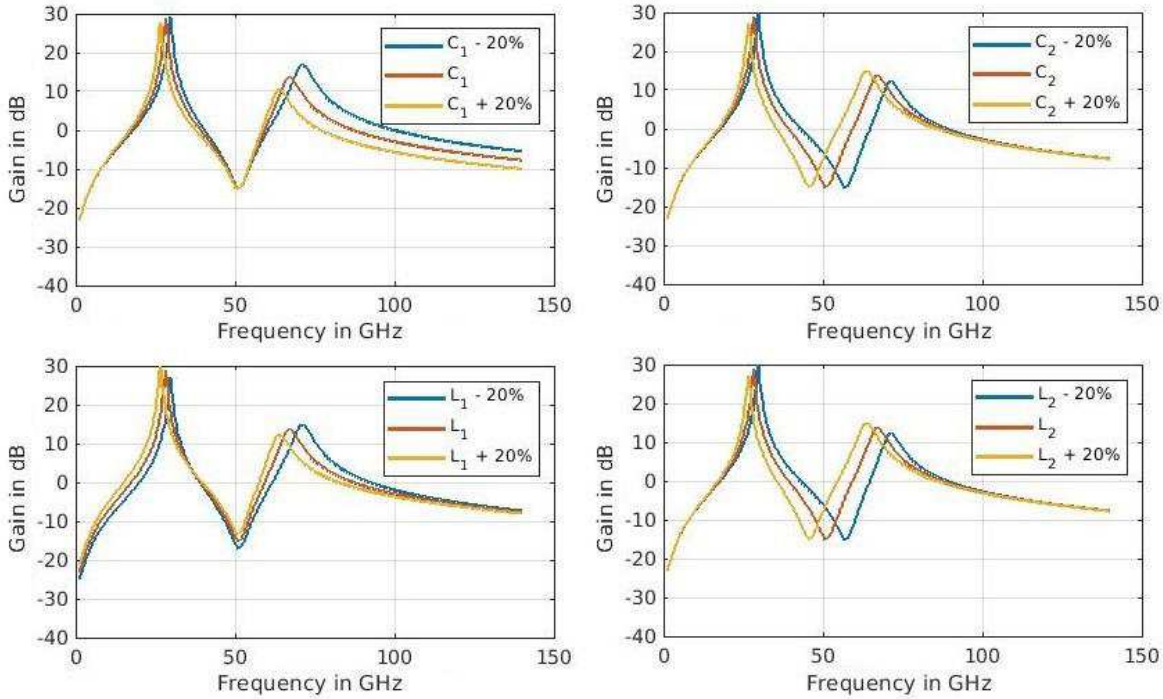

Figure 5-15: $H_{FF_1}(s)$ variations when sweeping C_1, C_2, L_1 and L_2.

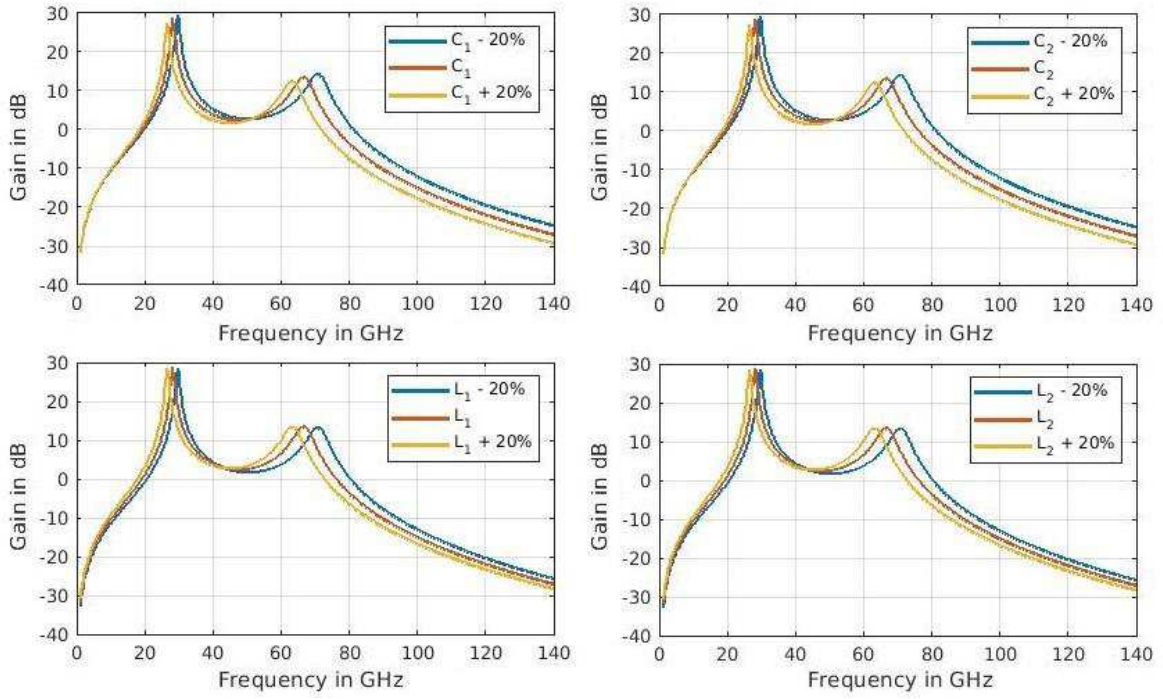

Figure 5-16: $H_{FF_2}(s)$ variations when sweeping C_1, C_2, L_1 and L_2.

The behavior is pretty much the same as an RLC resonator. More inductance or capacitor means lower resonating frequency and vice and versa. This is true for both resonances. A tunable capacitor can then be used to adjust the center frequency just like it would be done with an RLC resonator.

Next, the sensitivity to the coupling factor k variations is computed. The results are provided in Figure 5-17. This is very interesting, since it has a pole splitting effect: the larger k, the more spaced out the first and second resonance are. Since it is desirable for the second resonance to be as far as possible, it is necessary to seek transformers with the highest coupling factor possible. This will be a first constraint when sizing the devices.

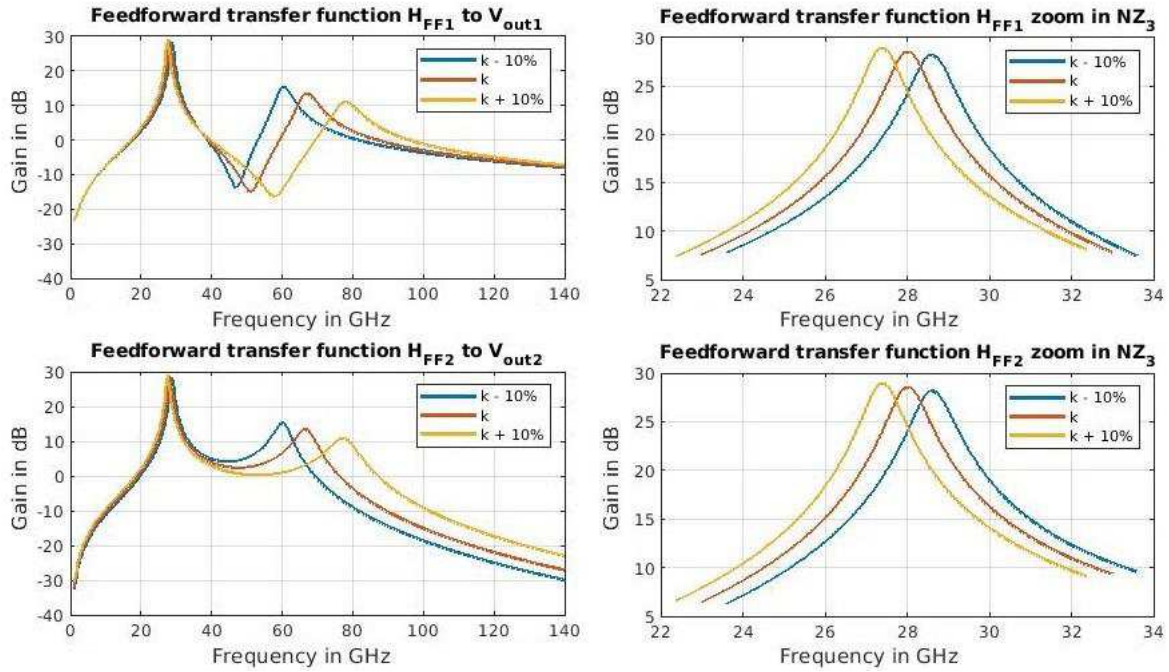

Figure 5-17: Feedforward transfer functions variations when sweeping k

Let us now study the impact of L_2/L_1 ratio. When doing so, not only R_1 and R_2 are re-adjusted for constant inductance quality factor, but also C_1 and C_2 are re-adjusted such that the resonating frequency remains $28GHz$ and $C_1/C_2 = L_2/L_1$. Figure 5-18 plots the results and interesting things can be observed.

While $H_{FF_2}(s)$ is nearly unaffected, $H_{FF_1}(s)$ sees a significant variation of its gain while maintaining the same quality factor. Another way to look at it is that $H_{FF_1}(s)$ and $H_{FF_2}(s)$ correspond to two different feedforward coefficients, but for the same amount of gm. Clearly, this can be exploited to reduce the amount of gm required and therefore the power consumption.

The next study case will be about using a different ratio between C_1 and C_2, with $L_1 = L_2 = 400pH$. C_1 is set to a different value and C_2 is adjusted to have the resonating frequency at $28GHz$. The results are plotted in Figure 5-19. Here, two distinct effects can be noted. The first one is the shifting of the second resonance toward the higher frequencies compared to the reference point. The second is a reduction in the resonator quality factor. While it could be good to push the second resonance to higher frequencies, if it is at the cost of quality factor degradation it is probably not desirable.

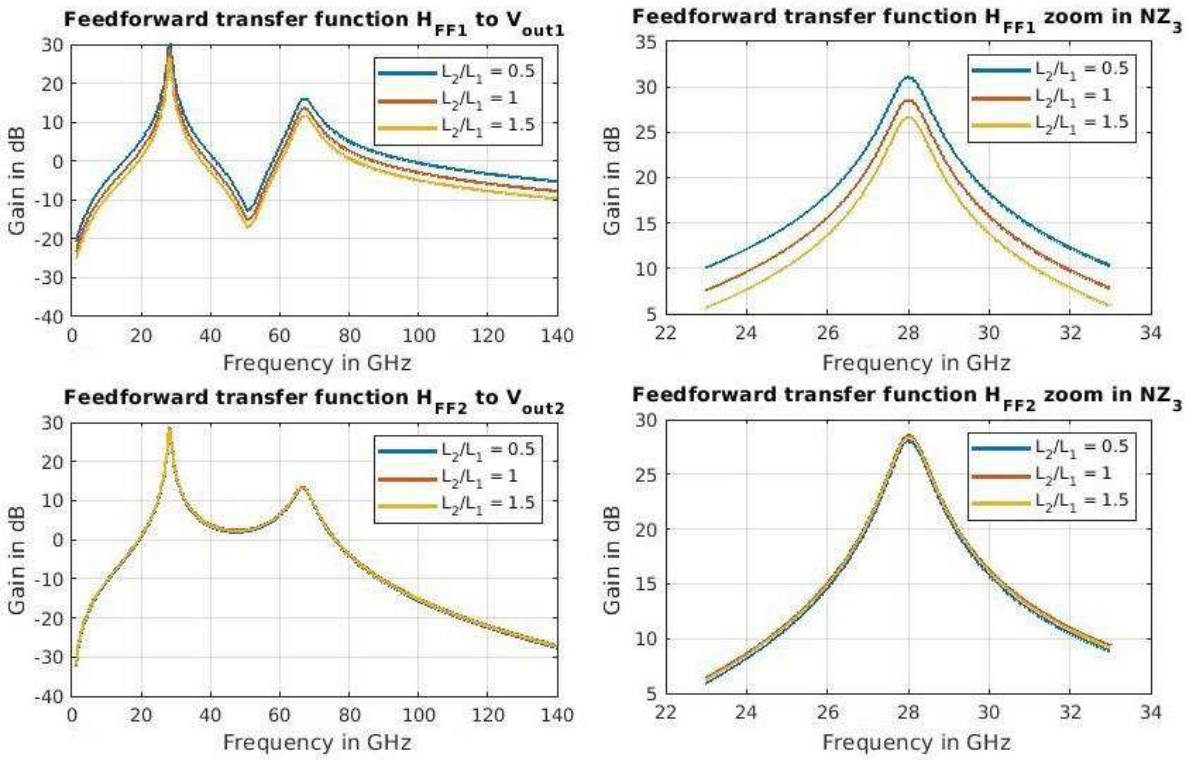

Figure 5-18: Feedforward transfer functions variations when sweeping the ratio L_2/L_1

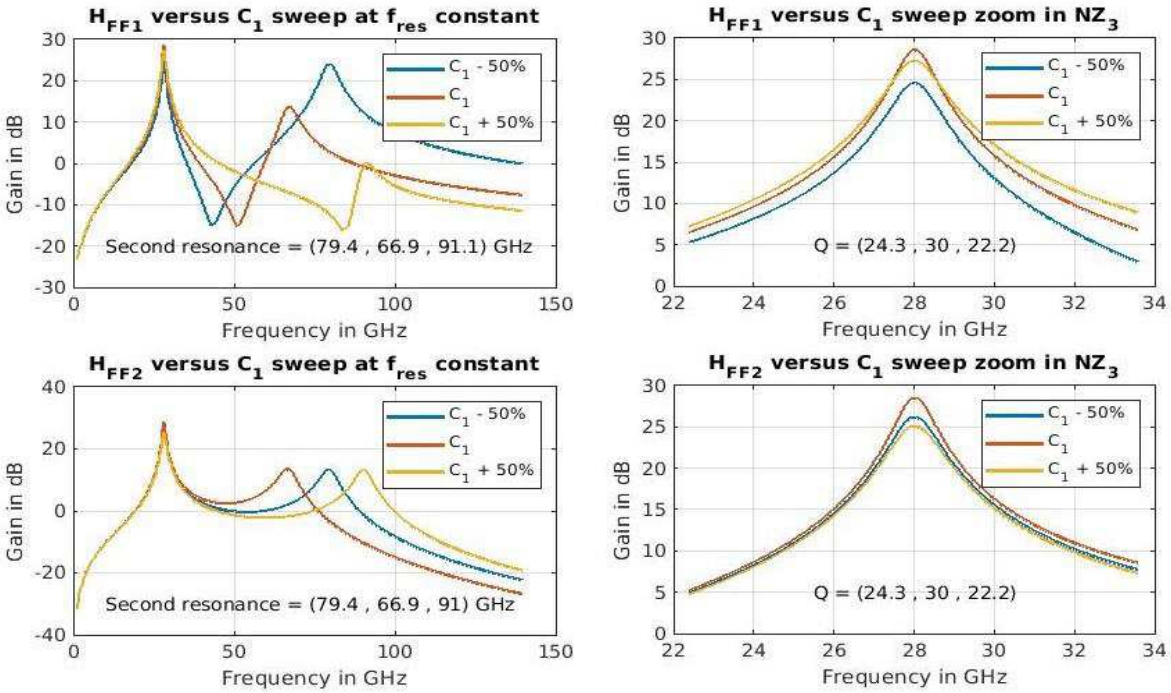

Figure 5-19: Feedforward transfer functions variations when sweeping C_1, with $L_1 = L_2 = 400pH$ and adjusting C_2 for a constant 28GHz resonating frequency.

To characterize this phenomenon the evolution of C_2, the quality factor Q and the frequency of the second resonance f_{SR} are monitored while C_1 is being varied (Figure 5-20).

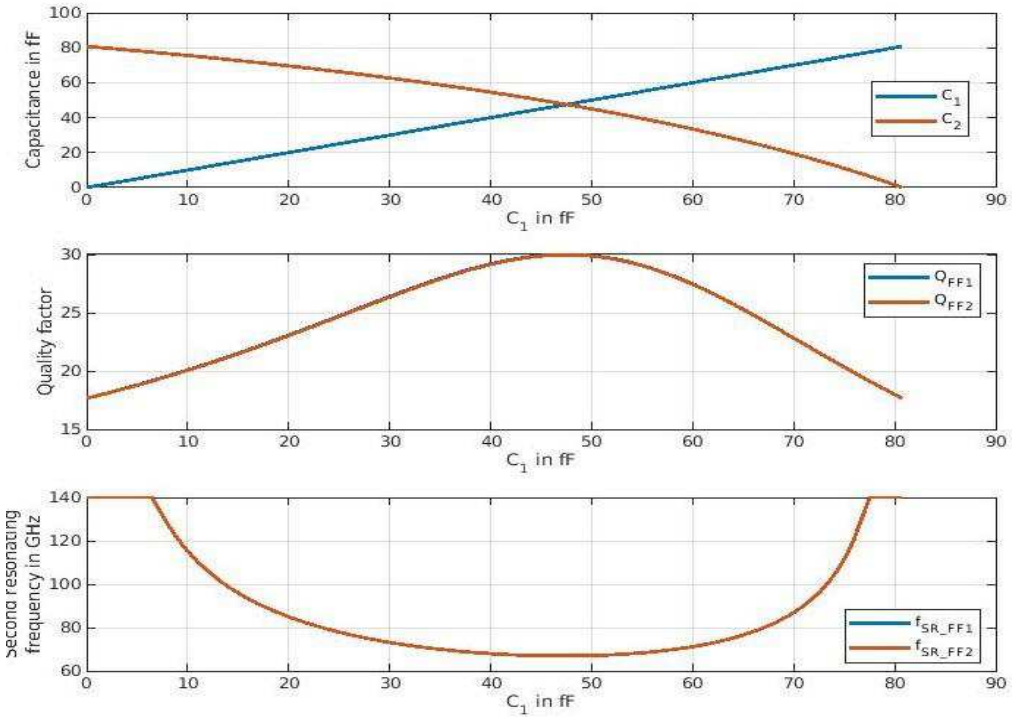

Figure 5-20: C_2, Q and f_{SR} as a function of C_1

The behaviors are exactly the same for $H_{FF1}(s)$ and $H_{FF2}(s)$. The quality factors pass by a maximum when $C_1 = C_2$. This is only the case because $L_1 = L_2$. More generally this maximum happens when $L_2/L_1 = C_1/C_2$. It also corresponds to the minimum frequency for the second resonance.

Clearly, it is not possible to get both a high Q and high second resonance. If the second resonance were to be pushed above $80GHz$ for example, that would degrade the quality factor from 30 down to 25. Here, the chosen strategy will go toward tuning C_1 and C_2 to maximize Q and maximize k to push the second resonance high enough. As shown in Figure 5-19 top left graph, when using $H_{FF1}(s)$, increasing C_1 from its optimal value not only pushes away the second resonance but also lowers its amplitude. If C_1 must be increased because of the second resonance, the configuration using V_{out_1} should be preferred. If C_1 must be decreased, it is then the configuration using V_{out_2} that should be preferred.

5.1.3.3 Feedback transfer functions

The impact of using a CVDAC instead of an IDAC was already studied. Here, the focus will be on how the feedforward and the feedback coefficients can be set in a judicious way. Noting that both feedback transfer functions from equation (5.10) and (5.11) are identical, only (5.11) will be studied. In particular it will be assumed that $C_{FB_1} = 0$. Figure 5-21 plots the comparison between the proposed transformer-based resonator feedforward and feedback transfer functions with an ideal parallel RLC resonator. The same conclusions as before can be reached. While the overall behavior can vary significantly compared to the ideal case, in the vicinity of the resonant frequency, it is almost identical.

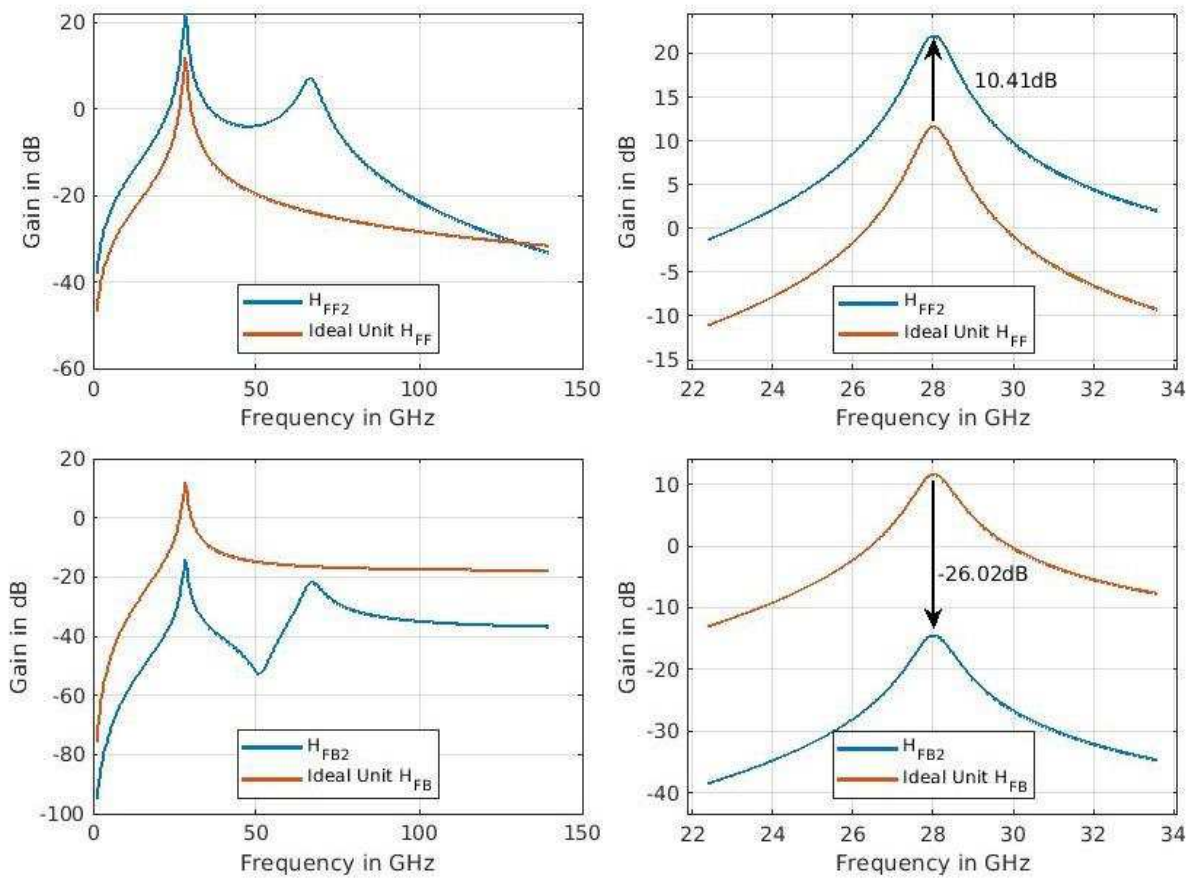

Figure 5-21: Transformer based resonator feedforward and feedback transfer functions to the second output V_{out2} and ideal parallel RLC resonator feedforward and feedback unit transfer functions.

The simulation parameters are summarized in Table 5-2. The ideal curves correspond to unit coefficients, allowing to evaluate the effective coefficients of the current proposal. The effective feedforward coefficient is $c_2 = 3.317$ and the feedback one is $a_2 = 0.05$.

Table 5-2: Model's parameters' values for Figure 5-21

gm	C_1	$L_1 + L_2$	L_2/L_1	Q_1	R_1	k	Q_2	R_2	C_2	C_{FB_2}
15mS	25.25fF	800pH	16	17.65	3.99Ω	0.7	17.65	0.47Ω	398.82fF	5.13fF

If the goal was to realize the same coefficients with a configuration using an ideal parallel RLC resonator and CVDAC while keeping C_{FB_2} the same, how much gm would be required? Equation (5.12) gives the relationship between the feedback coefficient in such a configuration.

$$a_2 = \frac{2 \times \pi \times f_c \times C_{FB_2}}{C_{FB_2} + C} \times \frac{1}{f_s} \tag{5.12}$$

This equation can be reversed to evaluate the required $C_{tot_2} = C_{FB_2} + C = 806.36fF$. Then, using equation (5.1) the amount of gm that would be required to realize the feedforward coefficient can be determined. It would be $59.91mS$. This is about four times more than with the transformer-based approach. The transformer effectively performs gm multiplication. This has the potential to save a lot of power. How much multiplication can be obtained depends on k and $\sqrt{L_2/L_1}$.

If the problem is taken from the other end, assuming a gm of $15mS$ is available then C_{tot_2} would be $201.88fF$ and C_{FB_2} $1.285fF$. Such a small capacitor can become problematic for implementation.

Overall, this technique can be very useful, when implementing small feedback coefficients, to keep the feedback capacitor to a feasible value without increasing the amount of gm, hence keeping a low power consumption.

There is another situation where the transformer can be used to simplify the implementation. It has to do with the case where the feedback coefficient is large. In that case, having a small equivalent capacitor is not necessarily an issue, and it is even good from a power consumption standpoint.

To understand how the transformer can help, another impairment that have been ignored so far can be considered. This is the shifting of the resonator's center frequency with process. Here, the plan is to use the classic approach of a digitally controllable parallel capacitance. A unit element of this tunable capacitance is given in Figure 5-22. Based on the on and off switch state equivalent circuits, the on and off state equivalent capacitance are $C_{eq_{on}} = C_u$ and $C_{eq_{off}} = C_u \times C_{off}/(2 \times C_u + C_{off})$. This gives the relative unit capacitance variation $\Delta C_u/C_u$ in equation (5.13).

$$\frac{\Delta C_u}{C_u} = \frac{1}{1 + \frac{C_{off}}{2 \times C_u}} \tag{5.13}$$

There is a tradeoff in the sizing of the switch. A large switch will have a low R_{on} allowing for small quality factor degradation in on state. But it also comes with a large C_{off} which reduces $\Delta C_u/C_u$. The total tank capacitance C comes mainly from two different sources, the parasitic capacitance C_{par}, and the tuning capacitance C_{tune}, which is an assembly of the unit element from Figure 5-22.

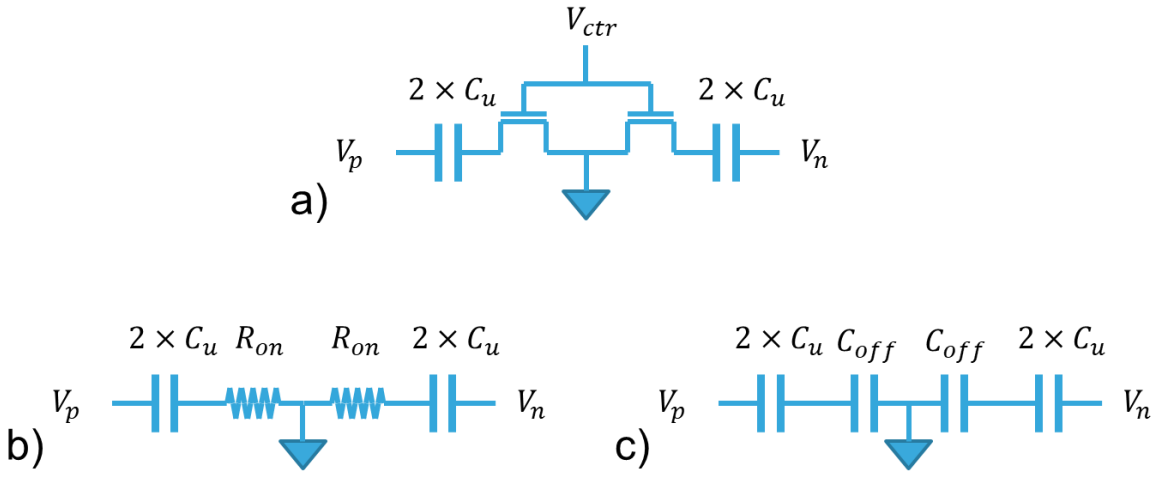

Figure 5-22: a) Digitally controllable unit capacitance. b) Equivalent circuit for on switch state. c) Equivalent circuit for off switch state

If necessary one can add some dead capacitance to tune the nominal center frequency, but for now it will be assumed there is none. The effective total capacitance variation is then given by (5.14):

$$\frac{\Delta C}{C} = 1 \pm \frac{C_{tune} \times \frac{\Delta C_u}{C_u}}{C_{par} + C_{tune}} \tag{5.14}$$

When C_{par} represent a larger portion of the total capacitance, $\Delta C_u/C_u$ must also increase to be able to cover the same range. This leads to an R_{on} degradation (to lower C_{off}) and a loss in quality factor. This phenomenon is amplified for smaller values of the total capacitance since some parasitics such as the gm-cell output capacitance, are fairly independent of the tank size. Using the transformer as before allows to have the tuning capacitor on the side that can tolerate more capacitance allowing for a better tradeoff on the switch sizing. It also allows for a larger tuning unit element that can potentially be easier to implement.

5.1.3.4 Q-enhancement circuit

So far it was simply assumed it was possible to realize a resonator with a quality factor of $Q = 30$, but how this could be possible was never discussed. Clearly, even with the improvement brought by the transformer-based resonator, this level of performance is unrealistic for a purely passive device in a digital process such as the targeted one. The choice was to use the classic cross-coupled differential pair from Figure 5-23 to implement the Q-enhancement circuit. This topology is very common in differential LC-VCO oscillators. The principle is fairly simple the cross-coupled differential pair provides some negative gm that compensate some of the parasitic resistance R_i. Unlike VCOs, only a portion of the losses must be compensated such that the circuit remains stable. This approach is also convenient since, with this solution, the negative gm can be controlled by the biasing current. This will allow to actually have a tunable Q-factor.

On Figure 5-23-a, it is displayed on a parallel RLC resonator. In the present case, the question is: Which end of the transformer is it better to put the Q-enhancement circuit? The quality factor of a transformer is the same on both sides; let us call it Q_{TF}. Also, for each side, a local equivalent parallel RLC resonator can be defined. Let us index the primary and secondary equivalents by 1 and 2 respectively.

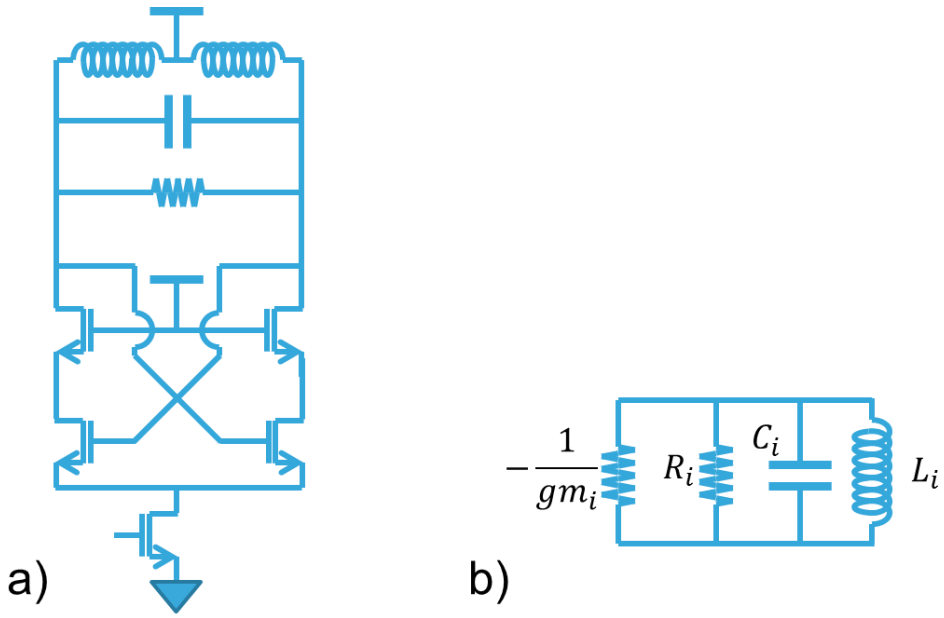

Figure 5-23: a) Negative-gm circuit topology. b) Principle single ended equivalent.

A parallel RLC resonator have the following relationship between R_i, Q_i and C_i:

$$R_i = \frac{Q_i}{2 \times \pi \times f_c \times C_i} \tag{5.15}$$

With this formula, the amount of negative gm required to improve on the transformer native quality factor Q_{TF} can be evaluated to reach the desired quality factor Q_i:

$$gm_i = 2 \times \pi \times f_c \times C_i \times \frac{Q_{TF} - Q_i}{Q_{TF} \times Q_i} \tag{5.16}$$

The first obvious observation is that this is a negative value, as one could expect, and the second is that gm_i is proportional to C_i. It will therefore be more power efficient to have the negative gm on the side with the lowest equivalent capacitance.

One last note here is that, if the cross-coupled pair is properly matched, this circuit also forces the signal to be differential. This can help correcting some imperfections of the input balun.

5.1.3.5 Conclusion

This was the final piece on the resonators' topology: All of them will be transformer based. The first one will be fully passive with no Q-enhancement and a targeted Q-factor of 1. It will have the task of performing the input matching, the single ended to differential conversion and to provide the two differential signals in opposite phase required for the gm-boosted common gate input stage. The two other resonators will be less constrained and will benefit as much as possible from the techniques exposed here to improve the power consumption. They will also benefit from a tunable Q-enhancement circuit that will allow them to have the desired Q-factor of 30. Finally, all three resonators will be tunable using a digitally controllable tuning capacitance. It was shown that the transformer-based approach, if properly used, can bring a significant improvement in power consumption compared to a traditional LC-Tank.

5.1.4 Quantizer implementation

The purposeful choice was made to have a sigma-delta architecture that allows a time interleaved quantizer thanks to the additional clock cycle available to close the loop. It allows only for interleaving once. This quantizer topology is often called "ping-pong". Its principle schematic is given in Figure 5-24.

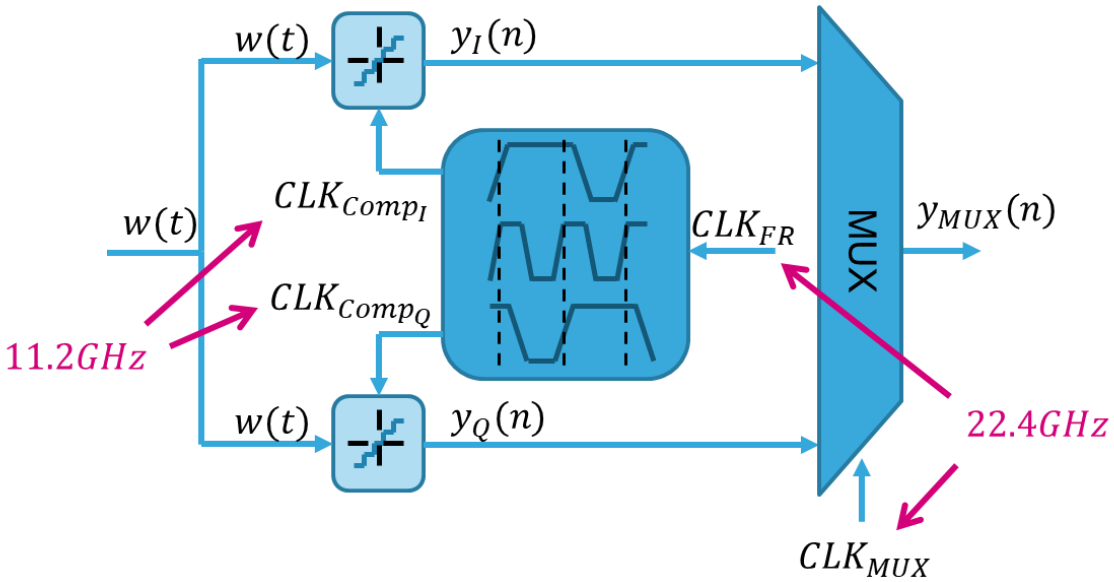

Figure 5-24: Time-interleaved quantizer principle schematic

It should be noted that the subscripts I and Q were added to the output names of the individual quantizers. These subscripts stand for In-phase and Quadrature-phase. It was mentioned in section 4.1.1 on down mixing that, for signals sampled at $f_S/4$, the quadrature down mixing is simply obtained through splitting the samples in two streams, I and Q, where even indexed samples go to the I stream and odd ones to the Q stream. Then for each stream, every other sample is inverted. This means that the

data from the top quantizer will end up in the I stream, the data from the bottom one in the Q stream, hence the naming. This convention will remain throughout the rest of this text.

The quantizer is made of two units in parallel running at half rate, or $11.2GHz$, and with opposite phase. The three building blocks that are the clock generator, the quantizers and the recombining output MUX will be detailed next.

5.1.4.1 Clock generator

The half rate clocks, CLK_{Comp_I} and CLK_{Comp_Q} are derived from the full rate clock $CLK_{FR} = 22.4GHz$. They are generated using two building blocks, the divider by two from Figure 5-25-a, and the pulse extender from Figure 5-26-c. Here, these two blocks will be described individually and then simulations of the overall Clock generator will be provided.

The divider by two must provide two half rate clocks in opposite phase. Halving the rate is done using a D-flip-flop with its output fed back to its input through an inverter. The opposite phase is obtained by having the dividers initialized in opposite states. The D-flip-flops are the dynamic True Single Phase Clock (TSPC) flip-flop from [5-6] depicted in Figure 5-25-b. To this classic TSPC flip-flop, a "set" or "reset" transistor was added, followed by an inverter to provide non-complementary output and a better drive. This solution allows for the needed very high-speed operation, as well as for the initial state control. The "set" and "reset" transistors are effective only if the clock or the input data is held low. For this purpose, the input clock is gated with a buffer with a low output when disabled.

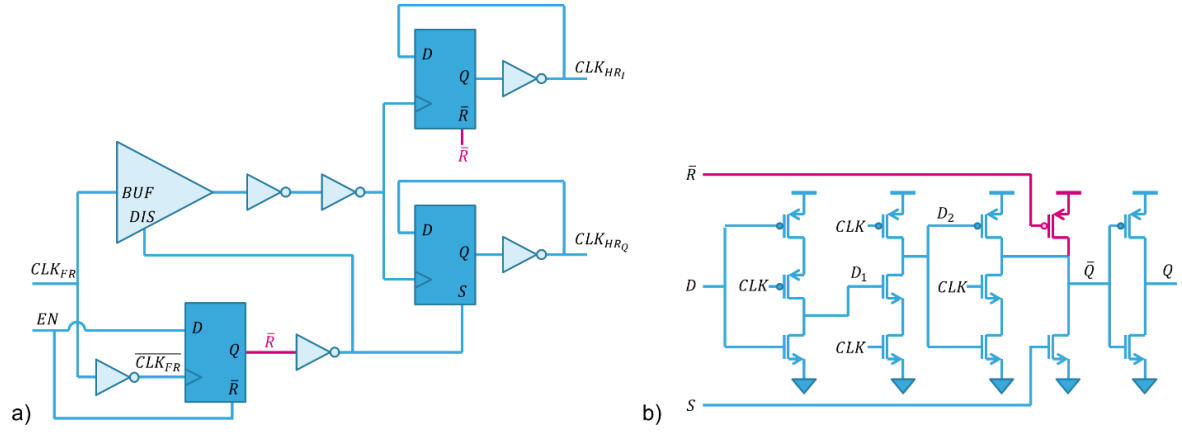

Figure 5-25: a) I and Q half rate clock generator b) True Single Phase Clock D-flip-flop topology

The buffer "disable", and flip-flop "set" and "reset" bits are locally retimed with $\overline{CLK}$ to ensure the proper start in opposite phase of the dividers by two. Between the buffer and the dividers by two, two additional inverters with increasing drive were inserted to provide a sharper edge.

When looking at Figure 5-25-a, the two elements used to re-time the "enable" bit, and inverter and a D-flip-flop, are directly connected to the input clock. Hence, if this clock is active, there will be some dynamic consumption left even when the block is disabled. Another thing that will have some consequences on power down consumption is that CLK_{HF_I} and CLK_{HR_Q} in Figure 5-25-a have opposite state in "disable" mode. This will be explained when discussing the comparators.

The second task to be performed on the half rate clocks is to extend their high pulses. This will give more regeneration time for the comparators and will be explained in more details shortly after. The classic method to extend a pulse is to feed a NAND gate with the clock and a delayed version of itself (Figure 5-26-a). Here, a slight variation of the NAND gate is proposed. The PMOS transistors on Figure 5-26-a, when activated, pull the output up to VDD. The chronograms from Figure 5-26-b shows that, by construction, CLK_{HR} is always ahead of CLK_{dly}. Hence the output pull-up is always performed by

the left PMOS driven by CLK_{HR}, and the right PMOS is never used and can be removed. This reduces the input and output parasitic capacitances, making the cell easier to drive and improving its own output drive. The counterpart is that for a short period, when CLK_{HR} is high and CLK_{dly} is low, the output becomes high impedance. If it were to stay in that state, leakage would corrupt the output after some time. Fortunately, the clock rate is more than fast enough to avoid this issue. In some sense, it becomes a dynamic pulse extender.

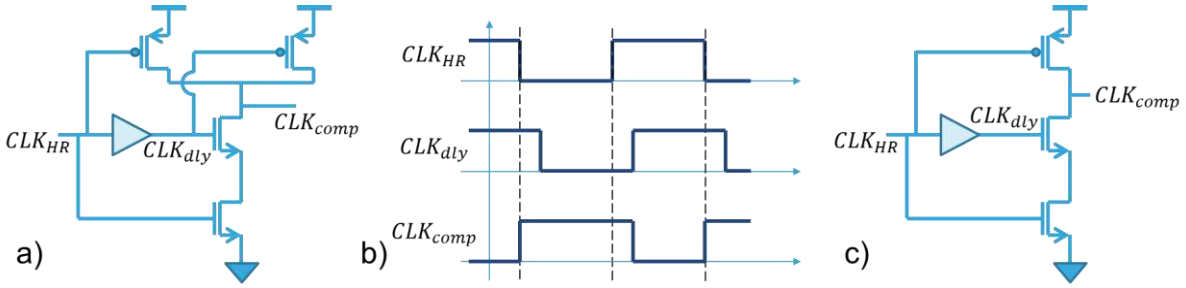

Figure 5-26: a) Classic pulse extender schematic. b) Pulse extender chronograms. c) Proposed pulse extender schematic

The simulation results of these two blocks forming the clock generator are given in Figure 5-27. It is done in typical conditions using a CC extraction of the layout. The overall behavior is the expected one with the final comparator's clock having a duty cycle of about 66%.

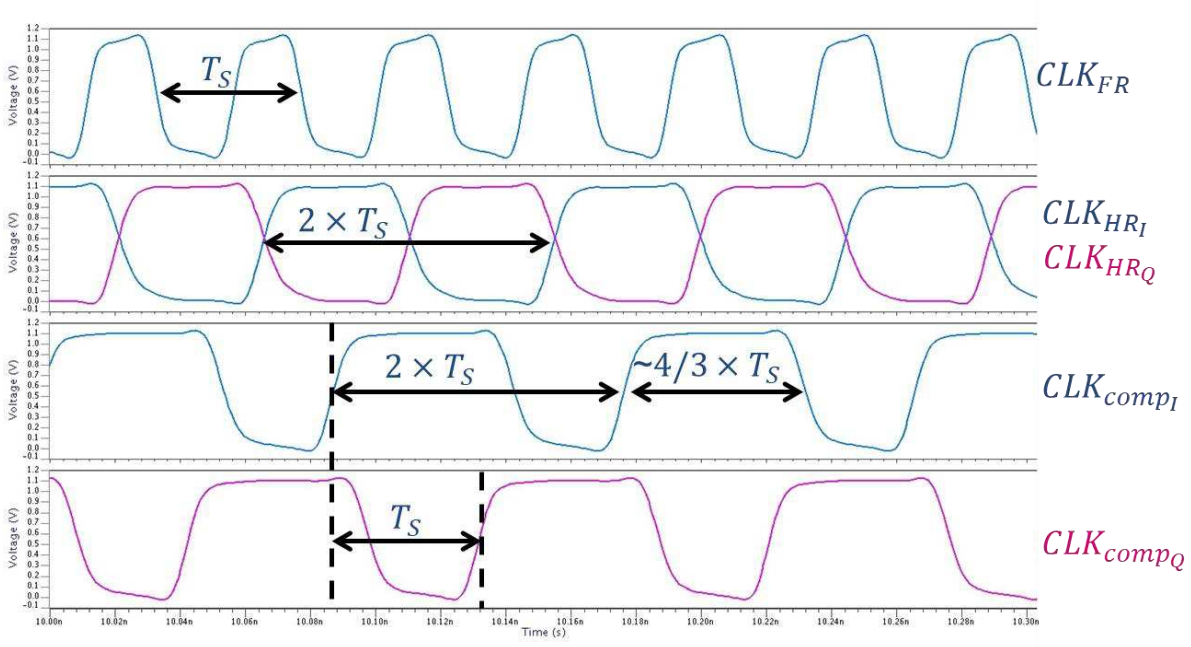

Figure 5-27: Clock generator simulation results

5.1.4.2 Single Quantizer

The quantizer is a 1.5-bit flash ADC, it has two comparison levels, one high and one low, symmetrical around zero. The thermometer coded output is obtained by using two comparators between the input signal and one of each of these references (Figure 5-28-a).

The comparator must fulfill several functionalities. First it needs to sample the input signal, then to perform the comparison against the reference and finally to hold the output data to be usable by the subsequent circuits. These three functionalities will be regrouped under the name latching comparator.

They are obtained by assembling a clocked comparator and a hold circuit (Figure 5-28-b). For the sake of clarity these two blocks will be described separately, starting with the clocked comparator.

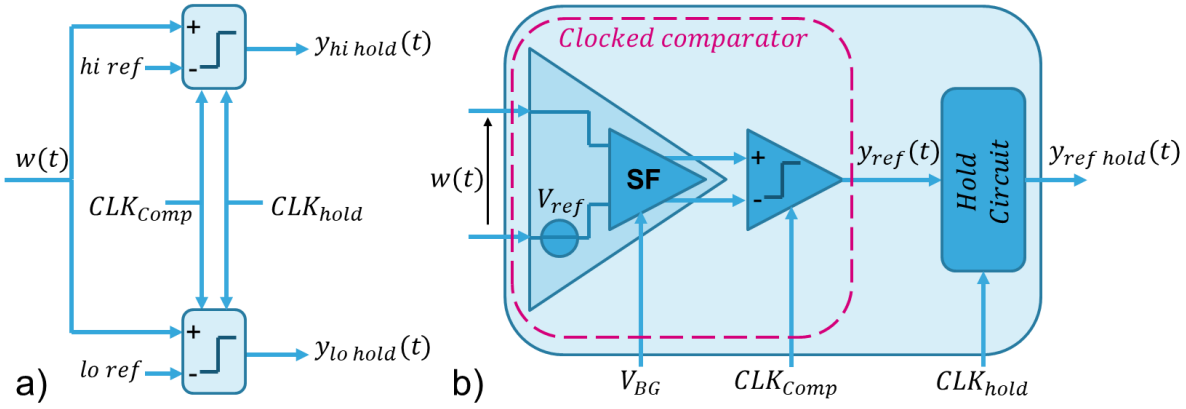

Figure 5-28: Quantizer and latching comparator principle schematics

The clocked comparators' topology is given in Figure 5-29. The fundamental principle is that of the double-tail latch-type voltage sense amplifier described in [5-7]. As its name suggests, this comparator only provides a latch functionality, and is not holding the output data during the reset phase, hence the need to add a subsequent hold circuit.

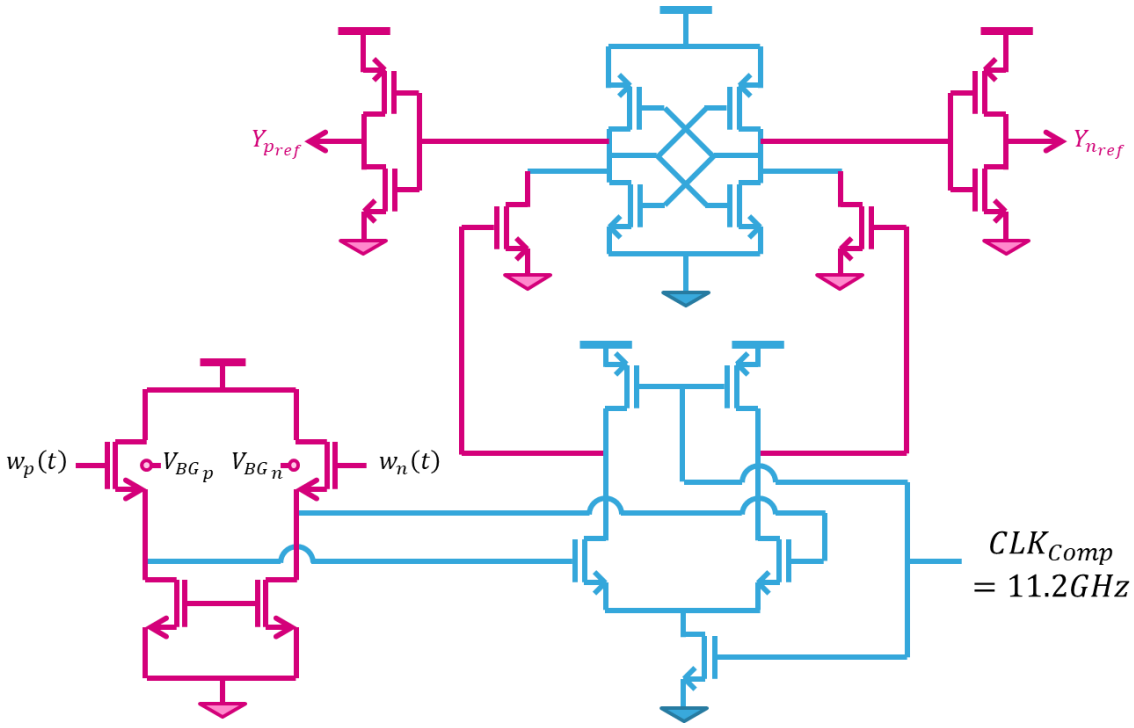

Figure 5-29: Clocked comparator's topology

This topology is effective when working at high speed and low supply voltage. It was successfully used in [5-8] as the core comparator in a 10-bit Successive Approximation Register (SAR) ADC working at 1MS/s. Later the authors in [5-9] proposed a slight variation where they added a gain stage at the front and removed the reset PMOS on the output stage. This last modification allows two things. First the comparator now operates with a single phase, and second the output regenerative latch draws current during the reset phase. While this is bad for power consumption, it improves the comparison time since the output latch does not need to wait for the current to flow to start regenerating. With this modification,

they successfully implemented a 6-bit Time Interleave (TI) SAR ADC working at 10GS/s. In that design, they had the comparator running at 10Gb/s. It is very close to the 11.2Gb/s required by the quantizer.

The proposed topology described in Figure 5-29 adds yet another variation. The input gain stage is replaced by a Source Follower as mentioned earlier. The main reason for this modification is because of the absence of a sample and hold circuit in front of this comparator. The SDM directly relies on the sampling capability of the clocked comparator itself. Hence, to get the input signal through, the input stage needs a bandwidth of at least 28.5GHz. To achieve such a large bandwidth with a gain stage, it would require a large amount of power. The first stage could have simply been removed, but the kick back of the regenerative latch revealed itself to be detrimental, making the presence of the input stage mandatory for improved reverse isolation. Hence a source follower was the simplest viable option from a power consumption standpoint.

One may wonder if the SF pseudo-differential nature could be a problem. This is not the case for two main reasons. First, the previous stage is truly differential, so its common mode gain is low, hence its output common mode is relatively stable. Second, it is a follower, i.e. it has unit gain which is obtain through a feedback mechanism. That makes the gain of each branch nearly equal and fairly independent of PVT, naturally limiting the CMRR degradation. In practice, some gain difference between the two branches can arise from bandwidth mismatch. To mitigate this issue, some margin is taken on the SF bandwidth, at the cost of some power consumption.

This additional stage was exploited to solve another problem. This comparator naturally only compares a differential signal with the "zero" level. To perform comparison with a different level, the Back Gate (BG) of the 28nm FDSOI process is exploited. Tuning the BG of a transistor performs a modulation of its threshold voltage V_{th}. In a source follower, this V_{th} modulation directly translates into output common mode variation. In the pseudo differential structure, this becomes an output offset that is naturally used by the following stage as a comparison level.

Both BG of the source follower input pair transistors are controlled by independent static DACs, such that the introduced offset can be positive or negative. Not only does this allow to implement both comparison levels, but it also makes them digitally controllable, and it can be used to perform offset calibration.

This is only possible because the feedforward coefficient c_3 needed is relatively large, at least 10. With the feedback DACs having a rail-to-rail output that means the comparison levels must be around $\pm VDD/(2 \times c_3) = 50mV$. The V_{th} modulation coefficient is about 80mV/V of BG tuning. That is enough to cover the need. In practice, the c_3 coefficient will be more around a hundred, leading to levels of $\pm 5mV$. The static DACs controlling the back gates have a 6-bit resolution and an output range of $[VDD/2\ \ VDD]$. For a 1V power supply this gives a quantization step of 7.9mV. With a V_{th} modulation coefficient of 80mV/V this leads to a comparison level quantization step of $635\mu V$ with a range of $\pm 40mV$. That gives enough accuracy to control the value of c_3 and enough tuning range to perform offset compensation.

The output chronogram is given in Figure 5-30. The behavior of this comparator is the same as the one described in [5-7] and will not be detailed here. One first note, is that the reset time is much shorter than the regeneration time. Also, it is the same for each cycle, while the regeneration time can vary widely with the input signal. This is the reason why the high pulse in the clock generator is extended. The duty cycle is moved from 50% to about 66%.

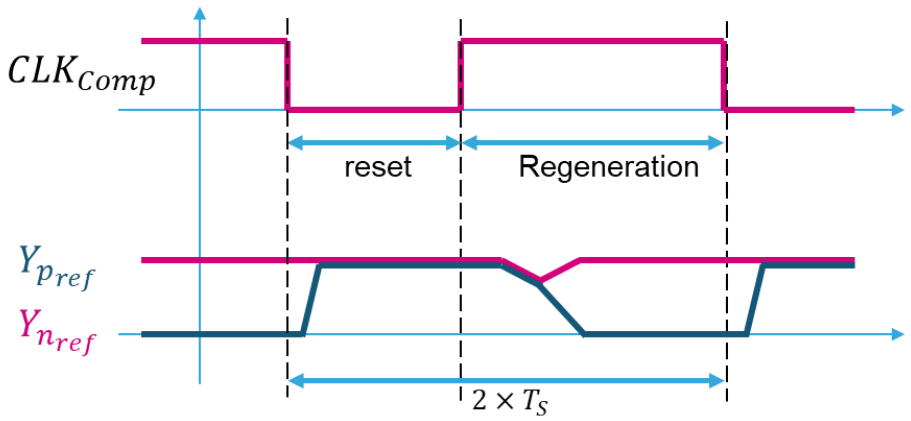

Figure 5-30: Clocked Comparator output chronogram

The implementation of the hold functionality of the comparator will now be described. It must hold its output during the reset time. A first idea would be to add a flip-flop similar to the one in Figure 5-25-a. But this can be optimized by exploiting one of the characteristics of the output signal from the chronogram in Figure 5-30: When in reset state, both outputs are high.

To understand how this can be used, let us first look at the chronograms of the TSPC D-flip-flop from Figure 5-31-b. When the clock is low, D_2 is held high which puts the output into a high impedance state. It will retain its value for some time, until the leakage current corrupts it. That is what makes this flip-flop a dynamic one. What is interesting here is that the comparator output behaves pretty much in the same way as D_2, i.e. it is high when the clock is low, and it follows the data when the clock is high. It is then possible to implement the hold function by just using the third stage of the flip-flop from Figure 5-31-a. This saves about two gate delays. It is important to optimize this delay because it is on the critical path to close the loop of the modulator, and as much delay as possible must be saved.

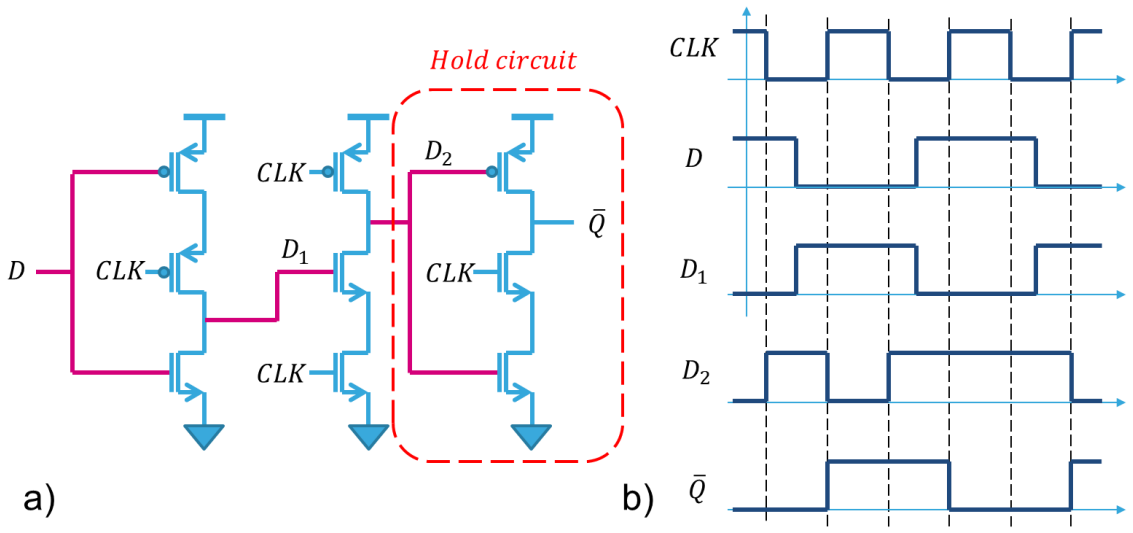

Figure 5-31: a) True Single Phase Clock D-flip-flop schematic.
b) True Single Phase Clock D-flip-flop chronograms

In practice, this output latch is not directly driven by CLK_{comp} but by a clock that has the same falling edge timing and a 50% duty cycle. This clock is called CLK_{hold} and allows to hold the data for a little longer compared to CLK_{comp}. Adding this hold function completes the set of functionalities the comparator needs, sampling, comparing with the reference and holding the result during the reset time.

Its simulation results are given in Figure 5-32 when the SF BGs are controlled to provide the high reference.

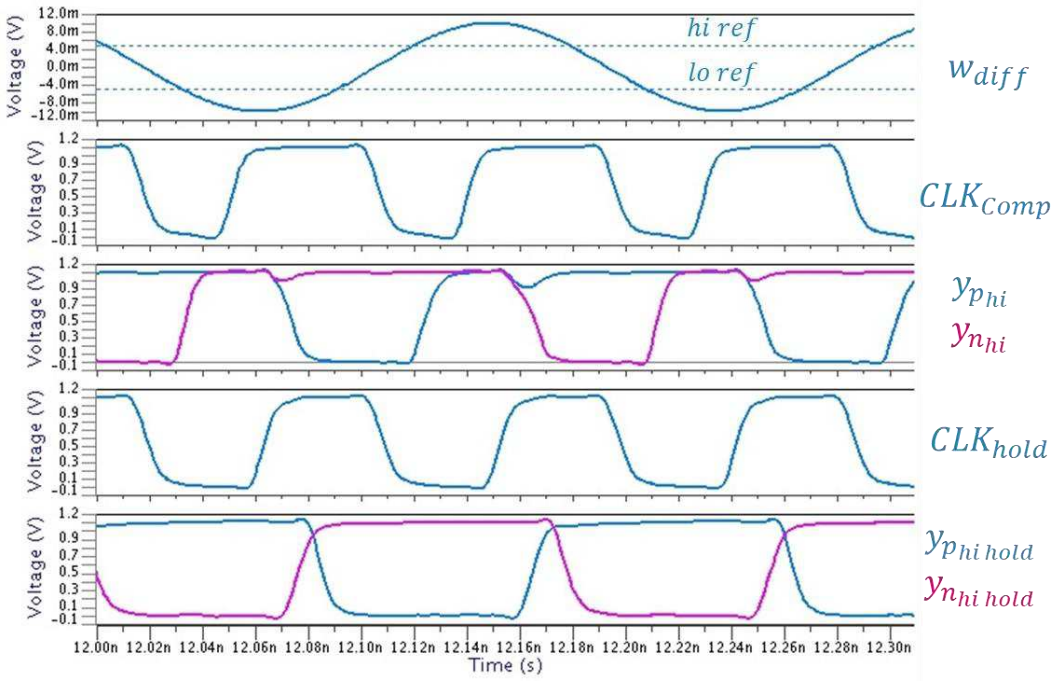

Figure 5-32: Simulation of the proposed latching comparator

From top to bottom, w_{diff} is the differential input signal, CLK_{comp} is the 66% duty cycle clock used for the comparator, $y_{p_{hi}}$ and $y_{n_{hi}}$ are its positive and negative outputs, before the holding circuit, CLK_{hold} is the 50% duty cycle clock driving the hold circuit, and finally, $y_{p_{hi\,hold}}$ and $y_{n_{hi\,hold}}$ are the positive and negative output of the hold circuit.

As expected, the clocked comparator samples and compares the input signal, and the held data have a well-defined value for a full clock cycle and can properly be used by the following stage. The output signals $y_{p_{hi\,hold}}$ and $y_{p_{hi\,hold}}$ are misaligned. This is because of the output high reset state, meaning one off the two signals always starts in its final state while the other one needs to make a transition. This makes this last signal slower, hence the misalignment.

This could have been mitigated by playing on CLK_{hold}'s duty cycle to hold the previous data until the regenerative latch has diverged. But that would delay the availability of the output data. Instead, a timing that was making the output data available as soon as possible was chosen, at the cost of this misalignment.

One can note a significant delay between CLK_{comp}'s rising edge and the time the regenerative latch starts its action. As a consequence, the sampling instant is not exactly the sampling clock rising edge but slightly later. Also, this delay has a dependency on the input signal which induces some signal dependent aperture jitter. While provision will be taken in the clock and data distribution tree to account for the unknown sampling instant, using tunable delay lines, the question on the signal dependent aperture jitter effect and potential mitigation techniques are left for future work. It will simply be assumed that this disturbance, since appearing at the quantizer level, will undergo the NTF, hence limiting its effect.

One last point to be discussed is the comparison duration. By nature, it will also be signal dependent and will lead to some variability in the data feedback timing. To some extent, this is mitigated by the

DAC. As explained in section 5.1.2.2, as long as the data is valid during the DAC clock high pulse, the DAC output timing only depends on its clock, not on the data. If the data is delayed too much it will arrive after the DAC clock rising edge and impact the feedback signal. In the present design no measure was taken when this happens. It is left to future work to add a non-divergence detection system that could for example feedback a zero in case of non-divergence.

5.1.4.3 Output merging MUX

The output data are recombined through a MUX, merging the two half rate input data streams into a single full rate output one y_{FR}. Its control input must be a half rate clock at $11.2GHz$ with the appropriate timing. Its schematic is given in Figure 5-33-a. It is basically two inverters with their output connected together. They have additional activation transistors such that activating one or the other inverter will allow the corresponding input to go through. When the sel bit is high, input a gets through otherwise it is b. To minimize the feed through of the blocked input, a and b are connected on the transistors further away from the output node.

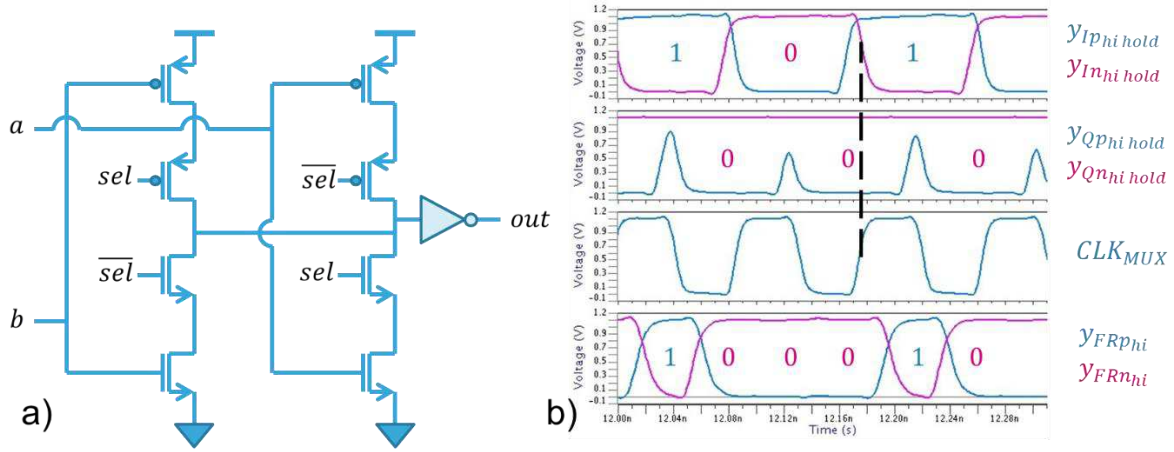

Figure 5-33: a) Output MUX topology. b) Simulation results

The simulation results are provided in Figure 5-33-b. It is a typical simulation using a Cc extracted view from the layout. Some bumps in $y_{Qp_{hi\,hold}}$ can be seen. They have the same origin as the misalignment between the positive and negative input signals. They correspond to the period when the latching comparator is transparent, but the regenerative latch has not diverged yet. Thanks to the CLK_{MUX} timing, these bumps are filtered out.

Also, the misalignment is reduced. When the MUX output is changing, its edge is driven either by the clock or the input, whichever comes last. From Figure 5-33-b it can be seen that, on one hand, CLK_{MUX} rising edge is delayed compared to the fastest input $y_{Ip_{hi\,hold}}$. The corresponding output edge of $y_{FRp_{hi}}$ is then driven by the clock. On the other hand, CLK_{MUX} is nearly aligned with the slowest input $y_{In_{hi\,hold}}$. The corresponding output edge of $y_{FRn_{hi}}$ is then also driven by the clock. Hence, both $y_{FRp_{hi}}$ and $y_{FRn_{hi}}$ are driven by the clock and are almost aligned. The remaining misalignment is caused by the MUX slew rate difference between the output rising and falling edges. Ideally, this would be compensated by increasing the PMOS transistor size. Unfortunately, this would increase too much its input capacitance for the previous stage to drive it properly. Since reducing the overall delay is a higher priority than having symmetrical complementary signals, the choice was made to have the MUX transistors to be slightly unbalanced to reduce the input capacitance.

In Figure 5-33-b, the input signals are always coming with the same timing relative to the MUX clock, but it was shown that it can vary, depending on the amplitude at the comparator's input. Occasionally, the delay will increase enough such that $y_{Ip_{hi\,hold}}$ and $y_{In_{hi\,hold}}$ will become slower than CLK_{MUX} and

the input misalignment will be transmitted to the MUX output. This could be mitigated by taking more margin on the delay applied to CLK_{MUX}, but again that would be at the cost of a higher overall delay. Instead, the choice was made to accept this occasional misalignment in favor of a better overall delay.

The challenge for this MUX is to provide the multiplexing functionality while adding the minimum delay. When it was designed, time was already running out and alternate solutions such as pass gates, could not be investigated to see if that would yield to a lower delay. This is one more optimization that is left to future work.

5.1.5 Clock and Data distribution trees

In the previous sections, all the elements of the loop were described. What remains to be done is closing the loop in a timely manner. Figure 5-34 provides a simplified schematic assembling all the building blocks described so far. Here, is also provided the chosen connections for the transformers. This will be discussed later. It also provides the architecture of the clock and data distribution trees.

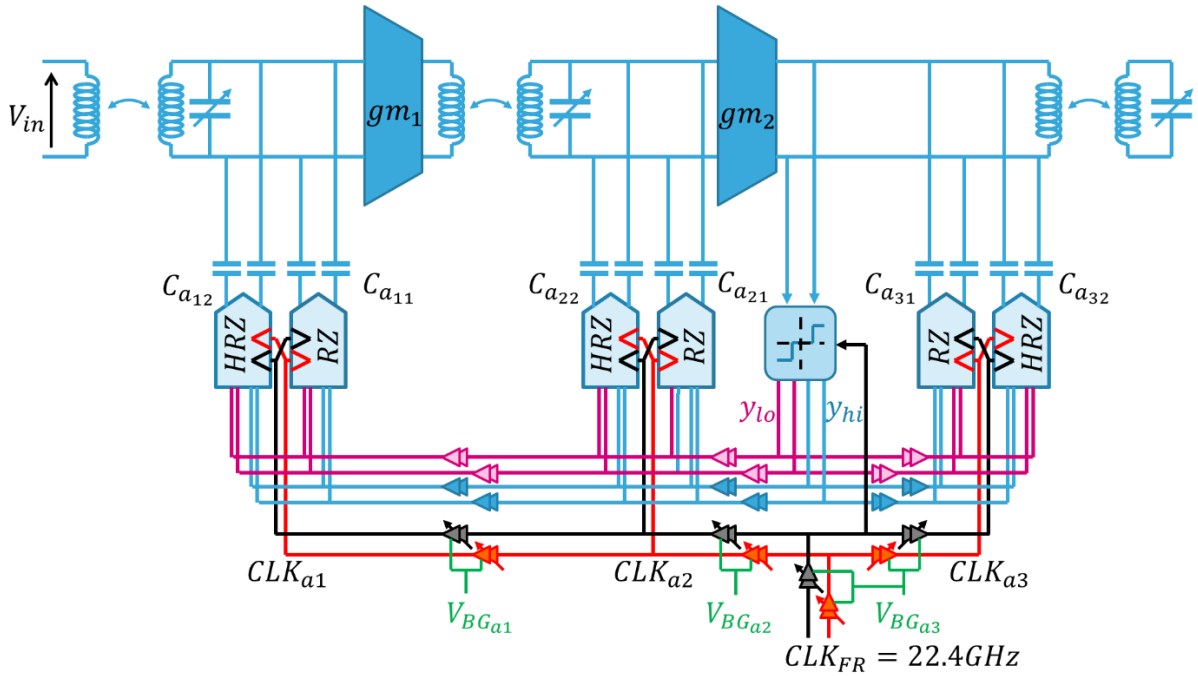

Figure 5-34: Modulator's simplified implementation schematic

In essence these trees are very simple. The data distribution tree is simply composed of buffers. Their roles are to drive the DAC's inputs while minimizing their delay such that the loop delay is equal or shorter than what is needed. In the next section, will see that the closing time constraints is limited to the most inner loop, and that solutions exist where the outer ones are more relaxed, allowing this architecture. As discussed in section 5.1.2.2 the data timing does not need to be very accurate: As long as the clock pulse fits into the data eye, the DAC output remains the same. The data buffers can then simply be some chains of inverters.

On the contrary, the timing of the DACs' clocks is very critical, but their absolute delays are not. Only their relative delay to the sampling clock matters. The clock tree will then be made of tunable delay elements to have the ability to precisely adjust these critical timings. Also, since this clock has a 50% duty cycle, inverting it equates to a half sampling period delay. This will be used to reduce the number of delay elements required to achieve the desired timing.

The tunable delay elements are simply achieved by tuning the back gate voltages of inverters as depicted in Figure 5-35-a. The back gate voltages are generating by two 4bits DACs with opposite controls.

Their output characteristics is given in Figure 5-35-b where all the input codes are swept in increasing order in a transient simulation. This dual control of the back gates aims at preserving the 50% duty cycle of the clock.

Each set of RZ and HRZ DAC has its dedicated tunable delay element. They have different numbers of elements to fit different needs in terms of tuning range and driving capability. In the architectural study it was shown that there is a range of about 1ps of ELD where the modulator's performances are within the target specifications. A delay step of about half a pico-second was chosen, meaning half a step or 250fs of feedback timing accuracy can be targeted.

The next constraint to take into account is the uncertainty about the sampling instant. The assumption is that the sampling happens at a point within the comparator's clock rising edge. Hence, the choice was made to cover that range or about 8ps. With a half pico-second step, this range can be achieved using a sixteen level DAC.

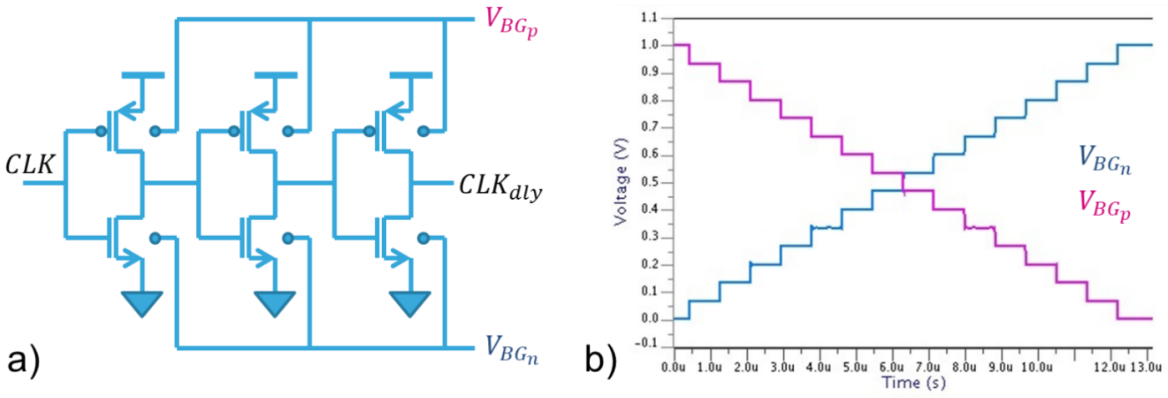

Figure 5-35: a) Tunable delay cell topology. b) Static control DAC characteristic.

The tuning range for each delay element is given in Figure 5-36. CLK_{a2} has a slightly shorter tuning range. However, since CLK_{a2} delay cell is fed by a clock coming for the middle of CLK_{a3} delay cell, it can also achieve 8ps of tuning range by delaying CLK_{a3} and CLK_{a2}. This reduces slightly the flexibility on the timing between the three clocks. This is not a problem since most of the tuning range is to cope with the unknown sampling point, which is common for all three clocks. It could even be argued that this clock tree could be simplified further. For the sake of testing, it was chosen to keep that flexibility. It also provides some margin to compensate for PVT variations.

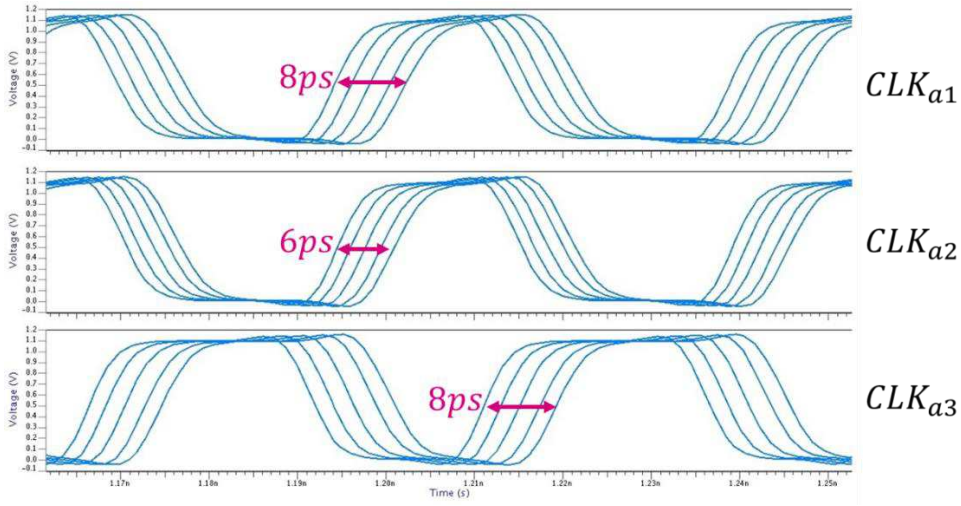

Figure 5-36: Delay elements tuning range

Finally, when the clock and data trees are connected to the DACs with a realistic load on them, the differential DAC output signal can be extracted from the simulation (Figure 5-37). One can see that the pulses are far from an ideal square shape. It will be seen in the next section how this was taken into account to perform the final optimization. Also, one can see bumps here and there in these feedback signals. They are caused by a global lake of accuracy between the clock and the data and the delays in the decoding logic that controls the DACs. When running at 22.4GHz, the process is really at its limit using CMOS gates. Moreover, this was designed the latest and time was missing to perform a more refined optimization. One obvious mistake was to not re-buffer the clocks after the delay cells to ensure constant and sharp rising edge, independent of the delay setting.

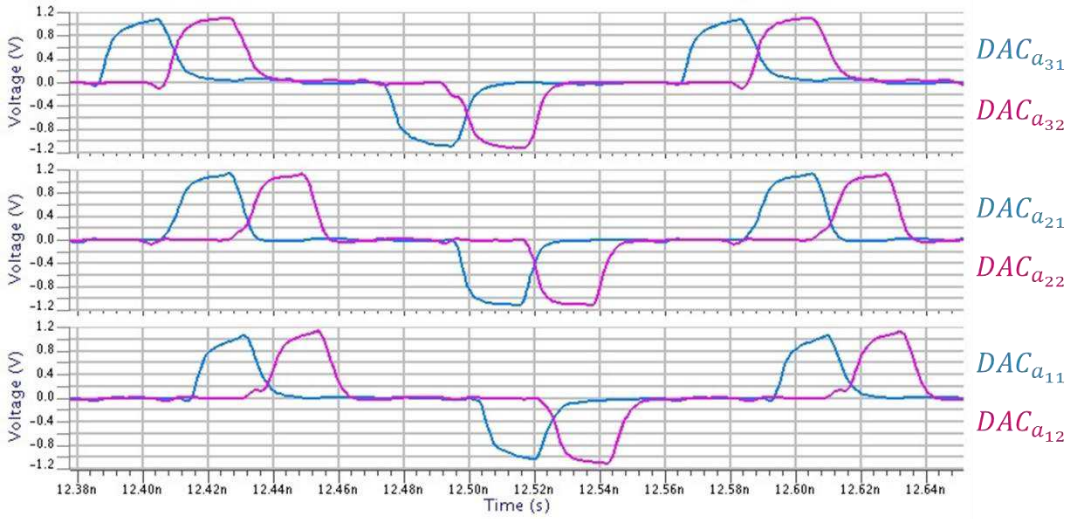

Figure 5-37: Feedback DACs' differential outputs

Similarly, one can see that the clocks duty cycles are visibly away from 50% or that the up and down pulses are not perfectly symmetrical. An attempt to account as much as possible for these non-idealities will be made in the final optimization, but it will not be possible to compensated for all of them. These imperfections will result in some performances degradation, explaining the difference between the system level model and the electrical simulation results.

5.1.6 Testing features and calibration procedure

For the sake of experimentation, a lot of configurability was added. In total there are 11 configurable parameters, the three feedforward coefficients c_1, c_2 and c_3, the three delays of the feedback paths, the three resonators center frequency, and the quality factors of the second and third resonators. In order for this configurability to be useful some testing features must be added such that the different parameters can be calibrated separately. The following two main testing features were added:

- Opening the feedback loop at the quantizer output and injecting a test signal in the feedback path. There are four different test signals internally generated, no signal, a sine wave at 5.6GHz, a cosine wave at 5.6GHz, and a single pulse repeated every 128 clock cycles.
- Enabling or disabling all feedback DACs individually

Using these testing features the following foreground calibration procedure can be envisioned:

- Comparator offset calibration
- Time interleaved quantizer gain calibration
- All resonators center frequency calibration
- Second and third resonators quality factor calibration
- Individual feedback path delay and feedforward coefficient calibration

- Close loop performance optimization

The comparator offset can be calibrated by opening the loop and not feeding back any signal. The quantizers inputs will then be only the feedforward path noise which has a zero mean. Averaging many measurements from the same comparator will give an image of its offset. The back gates of the comparator's source follower can then be tuned to minimize the offset.

The quantizer is two time interleaved, meaning it is made of two parallel quantizers running at half rate. While it is not possible to perform an absolute gain calibration, it is possible to perform a relative gain calibration of the two parallel quantizer such that they have the same gain. Opening the loop and injecting a sine wave will result in only one of the parallel quantizer to receive a non-zero signal. Through statistical measurements it is possible to get an image of its gain. Injecting a cosine wave allows to acquire an image of the second parallel quantizer. The comparators source follower back gates can then be tuned to set the two gains as close as possible.

The resonator center frequency can be calibrated by opening the loop and feeding back any of the 5.6GHz signal. Through the RZ and HRZ DACs this will lead to a strong tone at 28GHz at the resonator's inputs. Disabling the DACs from the two outer loops and making a statistical measurement at the quantizer provides an image of the third resonator's response at 28GHz. The calibration is then done by maximizing this response amplitude, meaning the resonant frequency is the desired 28GHz. Once this process is done for the third resonator it can be repeated for the second and finally the first resonator.

For the second and third resonators, it is possible to have an estimation of their quality factor by scanning points around the resonant frequency. This measurement then allows to perform the desired calibration on their quality factors.

The individual feedback delays and feedforward coefficient can be calibrated by opening the loop and feeding back the single pulse test signal. Starting with the most inner loop, disabling the two outer ones, and making statistical measurements, it is possible to evaluate the 128 first samples of its discrete time impulse response. The desired impulse response shape can then be obtained by tuning the corresponding clock delay, but it is not possible to calibrate its absolute amplitude. The same process can be repeated with the second loop. Once its impulse response has the desired shape, the c_2 feedforward coefficient can be adjusted such that the relative amplitude between the third and the second loop impulse responses is correct. Finally, the same process can be repeated for the last feedback loop.

This set of calibration should put the modulator in a functional state and close enough to its optimal configuration. The final calibration is done by closing the loop and looking for the near-by modulator's configuration that maximizes the SNR for a given external input test signal.

This calibration procedure is only a draft and needs to be refined once the test chip will be available for testing. Its main purpose is to reduce the calibration time by tuning the different parameters separately instead searching for the optimal configuration for all the parameters at once. This could be especially difficult since it is most likely a non-convex problem.

5.2 OPTIMIZATION METHODOLOGY

In the previous section, the topology of all the building blocks as well as their overall interconnections were described. To finalize the design, the transformers, feedback capacitors, gm-cells, and the feedback ELDs must be sized in order to reach the desired modulator. This will be done in two steps.

First, an LC based modulator will be generated, taking into accounts implementation constraints. For that modulator, the feedback and LC tank capacitors will be sized. This will give an idea on how the

different constraints are spread across the different elements. This will be used as the base line model to size the final implementation.

Second, the base line model will be adapted to a transformer-based design. The mechanics of transformers being significantly different from the inductors' one, it is not straight forward to derive analytically their characteristics. On top of that, the transformer model from section 5.1.3.2 is overly simplistic and many parasitic are missing. Instead, an empirical approach was chosen. Starting from the base line modulator, the different transfer functions of interest will be extracted, and the transformers will be iteratively tuned to get the physical implementation to match them. Once a result close enough is reached, the process will be reversed, injecting the transfer functions extracted from the electrical simulation into the model, to replace the ideal ones, and run a final optimization to retune the feedback capacitors.

This approach allows to account for two implementation non-idealities, namely the feedforward and feedback transfer functions and the feedback pulse shapes. These non-idealities can then be compensated by adjusting accordingly the feedback and feedforward coefficients.

During these two steps the goal will be to satisfy the following constraints as much as possible:

1. The feedback capacitors must be in the [1fF 5fF] range, such that it can easily be driven by the DACs and is not too small for implementation
2. The amount of gm on the gm-cells must be minimized to optimize power consumption
3. The input impedance must be 50Ω for proper input matching
4. The input referred noise of the receiver must be minimized

5.2.1 Initial LC based modulator sizing

For this initial modulator sizing it is necessary to first make a choice on the b_1, c_1, c_2 and c_3 feed forward coefficients. Then, the optimization process developed in the previous chapter to obtain the feedback coefficients can be performed. From there, in the following order, the feedback capacitance, the resonators total capacitances, the inductances and the amount of gm required on the gm-cells are derived.

From the targeted modulator generated by the online tool and the desired input dynamic range, it can be evaluated what the product of the four feed forward coefficients should be. In the present case, it should be equal to about 6000 or 75.5dB. In order to reach this gain, an initial guess for b_1, c_1 and c_2 will be made, and what remains is achieved using c_3, the quantizer gain. This way of proceeding was chosen since the quantizer gain can easily be tuned by simply adjusting its comparison levels.

To evaluate b_1, one point about the modeling must be clarified. The model used for optimization assumes an active input stage made of a gm-cell pushing current into the first resonator. In that case, b_1 is given by the ratio between gm and the LC resonator capacitance normalized by the sampling period. The voltage gain at the center frequency ω_c, between the gm-cell input and output is then given by equation (5.17):

$$H_{b_1}(s = j \times \omega_c) = b_1 \times f_S \times \frac{s}{s^2 + \frac{\omega_c}{Q} \times s + \omega_c} = b_1 \times f_S \times \frac{Q}{\omega_c} \tag{5.17}$$

For a unit b_1 coefficient and the target quality factor of 1 for the first resonator (c.f. section 4.3.1.2), that gives a gain of $-17.9dB$. In practice, it will be a passive stage, implemented by the input balun. Thanks to the gm-boosted architecture, a voltage gain around $3dB$ is expected. Hence, for the model to properly represent the proposed implementation, b_1 must be set to $3 + 17.9 = 20.9dB$. This does not mean that 20dB of passive gain is achieved, this is just a numerical consequence of the difference between the model and the actual implementation.

Replacing b_1 by c_1 or c_2 in (5.17) gives respectively the voltage gain across the first and second gm-cells. In both cases, as discussed in section 4.3.1.2, the targeted quality factor is $Q = 30$, which leads to a gain of $11.64dB$ for a unit feed forward coefficient. Here it will be assumed that 15dB of gain on each gm-cell can be achieved. The corresponding feedforward coefficient is $3.36dB$ or 1.47 in linear. It will be rounded up to 1.5 for both c_1 and c_2 or about $3.5dB$.

With that, the required value for c_3 is evaluated to $47.6dB$. This is about 240 and leads to a quantization step of $4.2mV$. It is a smaller quantization step than desired, but it remains achievable with the proposed comparator level tuning step of $635\mu V$.

The modulator parameters resulting from the optimization are given in Table 5-3. One can note that the coefficient pairs to each feedback node are of the same order of magnitude. This will allow to have both feedback capacitances in the same range, around the $5fF$ target. Another good property of this modulator is its ELD values. They are about $0.2 \times T_S$ larger than the equivalent modulator using IDACs. This shows that the result from section 5.1.2, on the compensation of the imaginary feedback coefficient, brought by the CDAC, done by adding a $0.2 \times T_S$ delay, is transferable to modulators with ELDs greater than one clock cycle. This gives additional pico-seconds to close the loop and increase the implementation viability.

Table 5-3: Initial LC based modulator coefficients, ELD and resonators Q-factor values

FF coefficients	$b_1 = 11.094$	$c_1 = 1.5$	$c_2 = 1.5$	$c_3 = 240.8$		
FB coefficients	$a_{11} = 6.973 \times 10^{-3}$	$a_{12} = -1.438 \times 10^{-2}$	$a_{21} = -1.405 \times 10^{-3}$	$a_{22} = 1.649 \times 10^{-3}$	$a_{31} = -2.842 \times 10^{-2}$	$a_{32} = -1.025 \times 10^{-2}$
ELD	$ELD_1 = 2.636 \times T_S$		$ELD_2 = 2.313 \times T_S$		$ELD_3 = 1.601 \times T_S$	
Resonator Q- factor	$Q_1 = 1$		$Q_2 = 30$		$Q_3 = 30$	

The performances of this modulator are plotted in Figure 5-38.

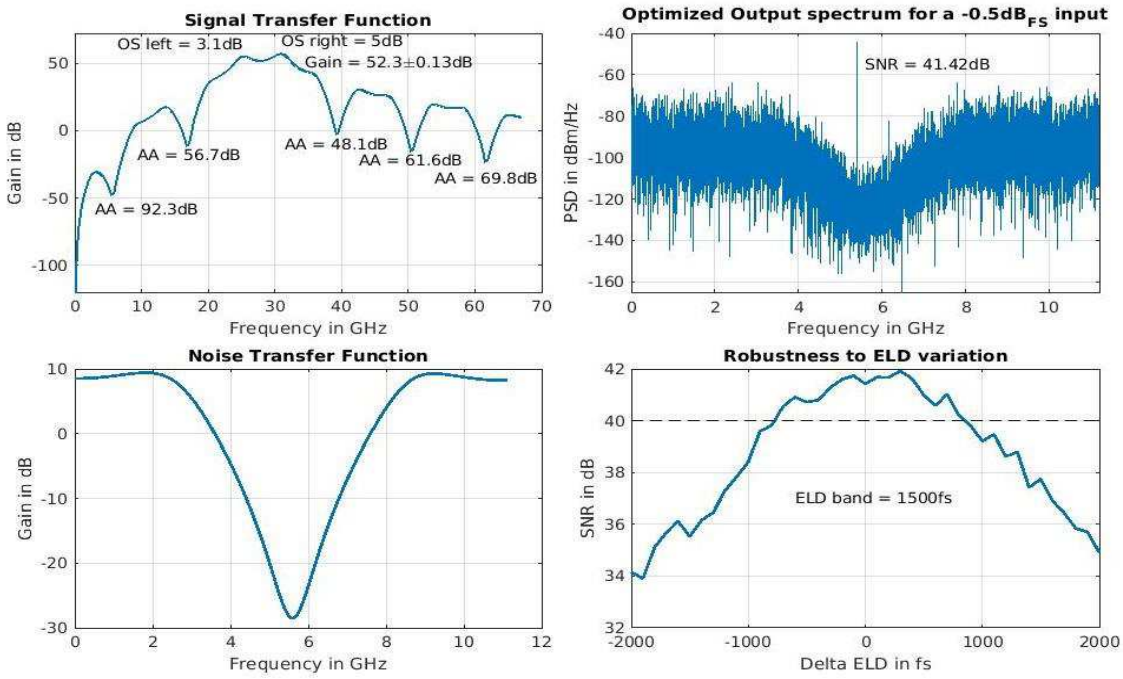

Figure 5-38: Initial LC based modulator perforamnces

The STF and NTF are preserved as well as the SNR. This comes as no surprise since these metrics are the direct objective of the optimization process. What is more interesting is the conservation of the robustness to ELD variation property. Ultimately this modulator has all the good properties targeted and is a very good base line for implementation.

The next step is to determine the different capacitor values. Equation (5.18) gives the value of a feedback coefficient as a function of the modulator center frequency f_c, the sampling frequency f_S, the feedback capacitance C_{FB}, and the LC tank total capacitance C_{total}. Using $5fF$ for C_{FB}, C_{total} can be evaluated.

$$a = 2 \times \pi \times \frac{f_c}{f_S} \times \frac{C_{FB}}{C_{total}} \tag{5.18}$$

Here, the target is 5fF for the two capacitors connecting the differential output of a voltage DAC to the resonator. Since they appear in series from the resonator standpoint, the corresponding coefficient will depend on half their value. Each resonator receives the output of two feedback DACs, corresponding to two different coefficients but referred to the same total resonator capacitance. Obviously both feedback capacitances cannot have the same value, so the total capacitance was chosen such that one of the feedback capacitances is 5fF and the second one is lower. The results are summarized in **Error! Not a valid bookmark self-reference.**.

Table 5-4: Initial Guess on resonators' total capacitance and feedback capcitances

$C_{total_1} = 1365\text{fF}$		$C_{total_2} = 11907\text{fF}$		$C_{total_3} = 690\text{fF}$	
$C_{11_{diff}} = 2.425\text{fF}$	$C_{12_{diff}} = 5\text{fF}$	$C_{21_{diff}} = 4.26\text{fF}$	$C_{22_{diff}} = 5\text{fF}$	$C_{31_{diff}} = 5\text{fF}$	$C_{32_{diff}} = 1.803\text{fF}$

On the total capacitance, one can see that the second resonator stands output with a very large value. As said already, this will require a large amount of gm to realize the desired c_1 feed forward coefficient. That is where the transformer approach will be effective, as it will be shown in the next section.

One note must be added on the capacitor's values from a matching point of view. First, since the modulator was optimized to be intrinsically robust to coefficients variations, their target accuracy can be as low as $\pm 10\%$. While this would be challenging to achieve in absolute value it is certainly within reach when matching two capacitors if their values are not too different. At first glance, it seems the feedback capacitances must be matched with their corresponding total capacitances in order to achieve the desired coefficients. The ratio between these two capacitors is typically between two and three orders of magnitudes, making the targeted matching accuracy challenging. Moreover, the total capacitance is made of many different types of capacitances such as gate and routing parasitic capacitances. Hence it will most likely vary differently with process compared to the feedback capacitance, making the problem even harder.

Thankfully, it is not necessary to match the feedback capacitance with the total one. To understand why let us first look at the most inner loop from Figure 5-1, recalled here for convenience. The feedback signals are weighted by a_{31} and a_{32}, summed up and finally weighted by c_3. One must remember that the only thing that matter is the effect of the feedback coefficient at the quantizer input. Hence, if a_{31} and a_{32} are multiplied by the same error factor e_3, this can be corrected by dividing c_3 by e_3. Since the feedback coefficients a_{31} and a_{32} are made by the ratio of C_{31} and C_{32} with the same capacitance C_{total_3}, if C_{31} and C_{32} are properly matched together, they will have the same mismatch with C_{total_3}. This results in both coefficients to be affected by the same error e_3 and can be compensated thanks to the configurability of c_3. Using the same method, the common errors e_1 and e_2 from the two remaining feedback loops can be compensated by using the configurability of c_1 and c_2.

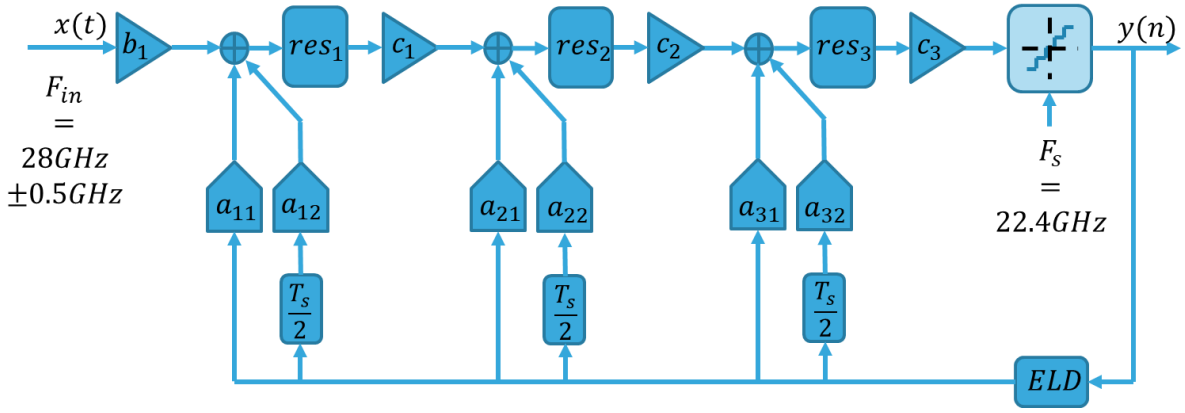

Figure 5-1: Modulators architecture

Thanks to this the feedback capacitances from a given loop only need to be matched between each other and not with their respective tank total capacitance. Since these capacitances are of a similar value and can be made of the same type of capacitor and close to each other in the layout, it is a lot easier to match them within few percent.

There is one drawback in using the feedforward coefficients to calibrate the feedback ones. The initial modulator's coefficients are chosen to achieve an SNQR of about 40dB, and such that the quantization noise is about 12dB below the receiver's thermal noise. Changing the feedforward coefficients as proposed would keep the SQNR performance but will change the quantization noise relative position to the thermal noise. If the quantization noise is higher, the receiver will have a better dynamic range but a worse noise figure. And it will be the opposite if the quantization noise is lower. This could be compensated by adding some configurability to b_1. In the proposed architecture, the passive input resonator is badly suited to achieve this kind of tunability. It is left to future work to adjust the receiver's architecture to add configurability to b_1.

5.2.2 Transformer based modulator sizing

To have this process running, it is first required to make the final fix to the topology. The feedback point on the transformers must be chosen. As seen just before, the first gm-cell will require a lot of trans-conductance. It is a good place to use the gm multiplication technique described earlier. This means the second feedback needs to be on the transformer secondary coil. For the two other transformers the choice is motivated by ELD. Since the timing constraints will be very aggressive, it is desirable to make things as close as possible in the layout. The most outer loop will then connect on the first transformer secondary and the most inner loop on the third transformer primary, making these two points as close as possible in the layout. This is how the final architecture from Figure 5-34 is obtained.

In practice, with the passive input resonator and gm-boosted common gate architecture, it is difficult to extract the transfer functions of the individual resonators from the electrical simulation. Instead, the extraction was made from the input and the feedback DAC outputs, all the way up to the quantizer input, to make the comparison with the theoretical model. For the forward path, the curves must be compared in an absolute way. This will allow to extract the required quantizer gain. For the feedback transfer functions, only the fact they have the proper shape is of interest, the gain will be adjusted last by tuning the feedback capacitances. Hence, only one of the RZ HRZ path will be compared for each feedback and their peak gain will be normalized with that of the model transfer function to make a relative comparison on their shapes.

For these extractions to be meaningful they need to take the layout into consideration. This was obviously an iterative process that is not described here. The final layout of the feed forward analog path is given in Figure 5-39. As already discussed, the input balun has a single ended input and two

differential outputs. The primary coil has two turns, the secondary one, connected to the input transistors' gates, has also two turns. Finally, the ternary coil, connected to the transistors' sources has only one turn.

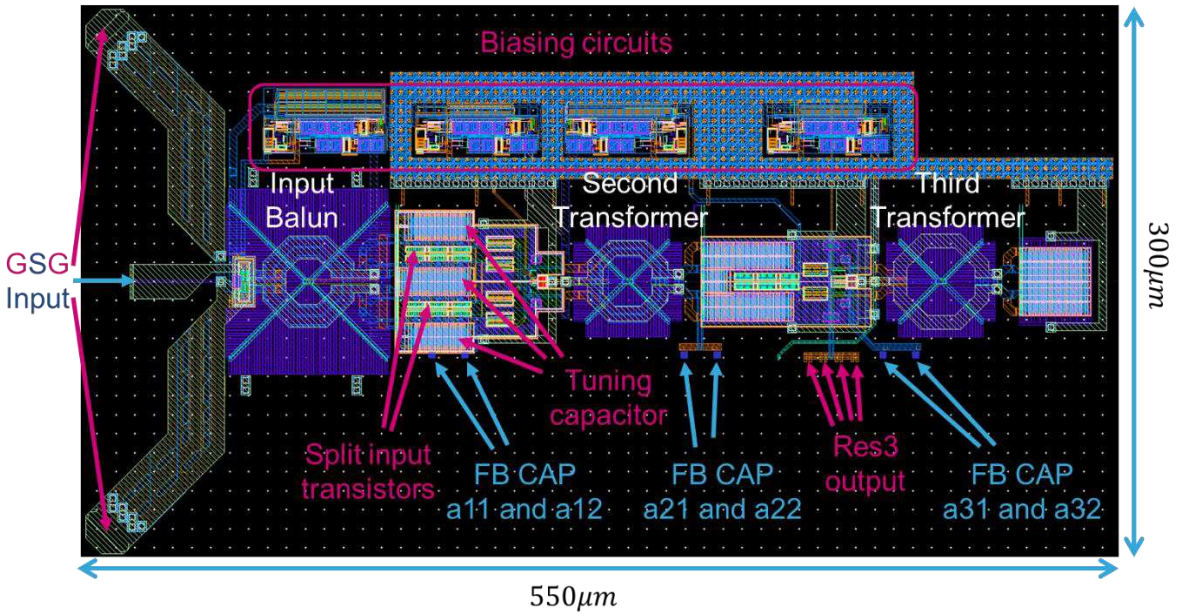

Figure 5-39: Feed forward analog path layout

This configuration was chosen in order to adjust the input matching while providing as much passive voltage gain to the gm-boosted common gate input stage. The S_{11} is given in Figure 5-40. It is lower than -10dB over a relatively large band from 25GHz to 40GHz, and it is lower than -14dB in the band of interest. This gives some margin to increase the first stage gm if necessary, without compromising the input matching.

The second transformer has four turns on its primary coil and one on its secondary one. As said already, the goal is to use the gm multiplication technique. A ratio of four was the best that could be achieved when considering layout constraints. To obtain this ratio requires to have four turns on the primary coil and one on the secondary. To achieve a higher ratio would require more turns on the primary coil. This means more inductance and more parasitic capacitance, driving the self-resonance frequency below the targeted 28GHz center frequency. Another reason for not going beyond this factor of four is that it is harder to achieve high coupling factors for transformer with a large difference on the number of turns between the primary and the secondary coil. In the present case, this would lower the second resonance to the point where it might become problematic. Hence, the gm multiplication technique was limited to a factor of four. Nonetheless, this is still a very nice power saving on the first active stage.

Despite this trick, about 30mS of gm is still needed on the first stage. This requires for the input transistors to be relatively large. To improve matching, it is common to use a common centroid layout for a differential pair. While this is good for matching, it makes the layout more complicated and increases the amount of parasitic capacitance. This is especially true in this case since both the source and the gate of the transistors must be accessed by the input signal. Instead, the transistors are placed along each other using relatively short multi-finger transistors so they can be close to each other for good matching.

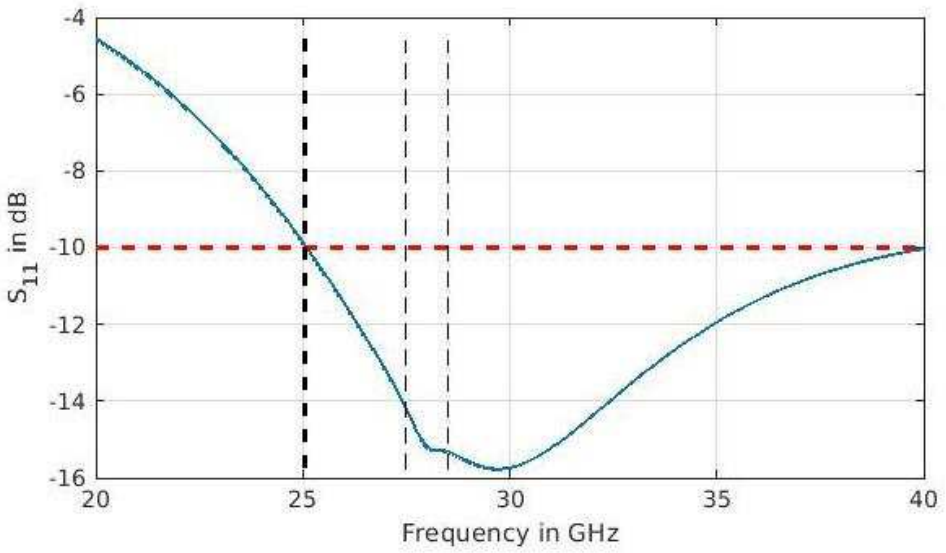

Figure 5-40: S_{11} parameter characterizing the input matching of the SDM

Using this layout approach, with the required transistor width, would lead to a very long layout. This would be bad for two reasons. First, the parasitic inductance of the routing would start to become non-negligible and would be difficult to account for because of its distributed nature along the transistors. Second, it would simply make the layout longer, meaning the feedback would have to travel a longer way, eating some of the timing margin. To alleviate these effects, the input transistors is split in two as shown in Figure 5-39. The access inductance to individual transistors is now halved and both splits are in parallel. The resulting parasitic inductance is then divided by four and the resulting layout is more compact reducing the distance for the feedback path.

The second stage requires about half the gm so the layout of one half of the first stage can be reused. In both cases, the tuning capacitances are compactly laid out around the transistors and the cascodes are deported on the right to avoid having parasitic capacitance between the stages inputs and outputs. Since there is a large voltage gain between these two points, this parasitic capacitance would lead to a large Miller effect and potential instabilities.

The third transformer has two turns on its primary coil and one on its secondary. The tuning capacitor is on the secondary coil. It allows to tolerate more parasitic capacitance, since the inductance seen from that side is lower and it avoids crowding the layout on the primary side. The third resonator's output is taken from the primary coil and split in four ways to feed the source followers of the individual comparators.

To have a full extraction, the source followers and the comparators are also added such that each stage is properly loaded. The extraction is done from the GSG input and the feedback capacitors all the way up to the comparator input.

Figure 5-41 plots the signal path transfer function. From DC to about 60GHz, the extracted TF matches very well the ideal LC one. After that, the downward slope of the extracted TF goes steeper than the ideal LC one. This difference is mostly due to the first and second resonators additional attenuation property that was ween in section 5.1.3.2, when transformers were studied individually, combined with the additional RC pole of the source follower stage. Quite surprisingly, there is not any clear second resonance. The exact explanation is not clear yet, but the hypothesis is that the second resonance of the first resonator, like its first one, is very weak and not visible. For the second and third resonator, maybe they have different coupling factors or and different ratios between input and output capacitors that makes the second resonances fall at different frequencies. In that case they would attenuate each other's

making them less visible. Finally, there is also a first order pole brought by the source follower that is designed to have a bandwidth slightly above 28GHz, and which would also attenuate these second resonances. Whatever the reason, it ends up providing a very good fit to the target signal path transfer function.

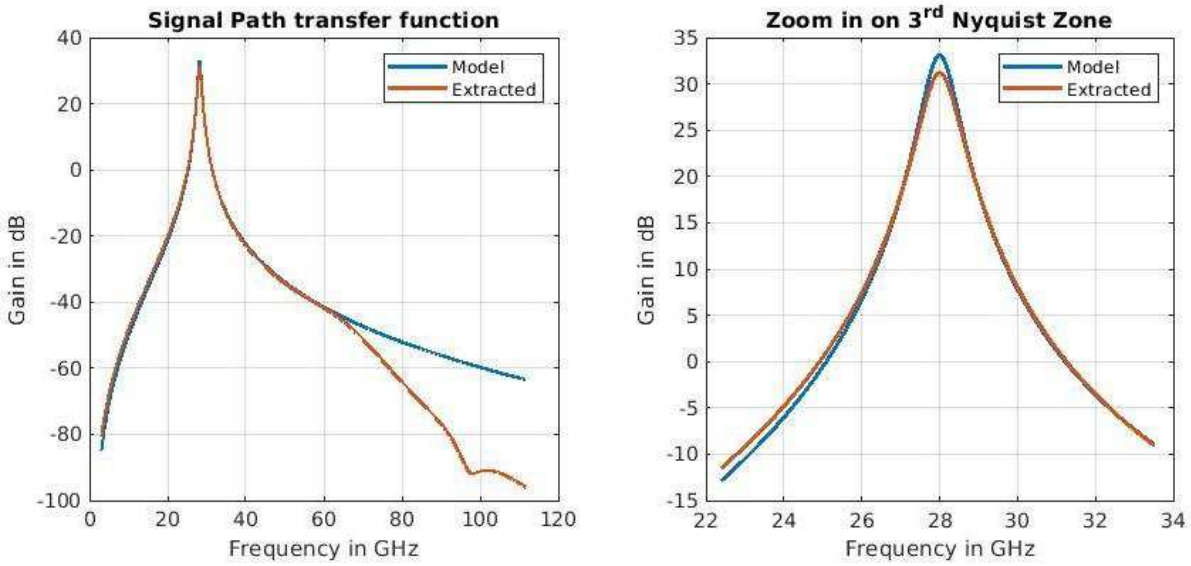

Figure 5-41: Signal Path transfer function from GSG input to comparator input

2dB are missing on the total gain. This loss is partially due to a lower-than-expected quality factors on the second resonator of about 19.7 instead of the 30 desired. This represents a reduction of the unit coefficient gain obtain from (5.17) of 3.65dB. This means, the product of the feed forward coefficient is effectively 1.7dB higher than expected. This will be accounted for when running the final optimization.

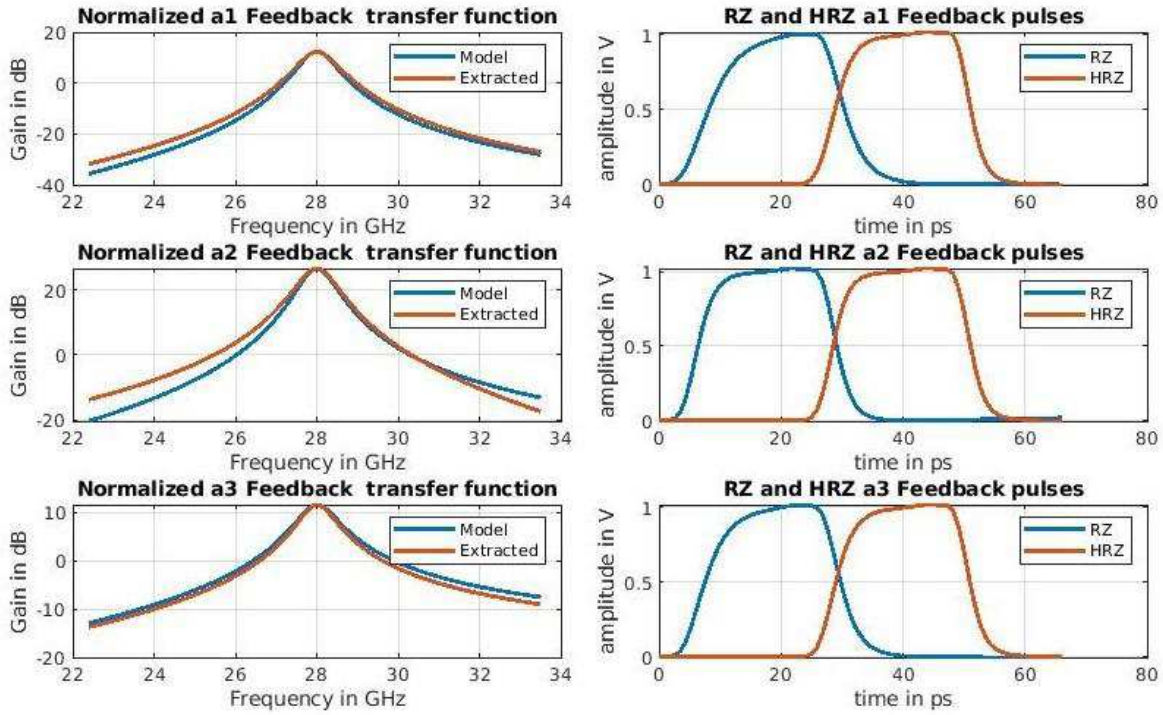

Figure 5-42: Left graphs, from top to bottom: First, second and third feedback transfer functions. Right graphs, from top to bottom: First, second and third feedback pulses for the RZ and HRZ DACs.

To account for as much implementation impairments as possible during the final optimization, the feedback transfer functions and the DAC pulse shapes from each feedback were extracted. They are depicted in Figure 5-42. The feedback transfer functions have a shape very close to the model, but the pulse shapes are quite different from an ideal RZ pulse. These pulses were extracted with some delay to properly capture the whole shape of the rising edge. Because their shapes are sufficiently different from the ideal case it is probably important to try to account for this difference.

These extracted pulses and transfer functions will be used as is by the system level model for the final optimization. The delays on the pulses will be subtracted from their respective ELD, compared to what was obtained previously, giving the initial ELD values for optimization. Also, the feedback transfer functions include already the feedback capacitance, hence the coefficients resulting from this final optimization will tell how it must be scaled relatively to its current value.

The results of this optimization are given in Figure 5-43. The STF is slightly different, which was to be expected since the feed forward and feedback transfer functions are not a perfect match. It remains very close to the original STF and is very satisfactory with a gain variation in the band of $\pm 0.2dB$. As expected, the anti-aliasing filtering is better for frequencies above 60GHz. The NTF is slightly shallower because of the second resonator reduced quality factor, but the impact is very minimal and unnoticeable on the SNR. Finally, the robustness to ELD variation of the original modulator is preserved.

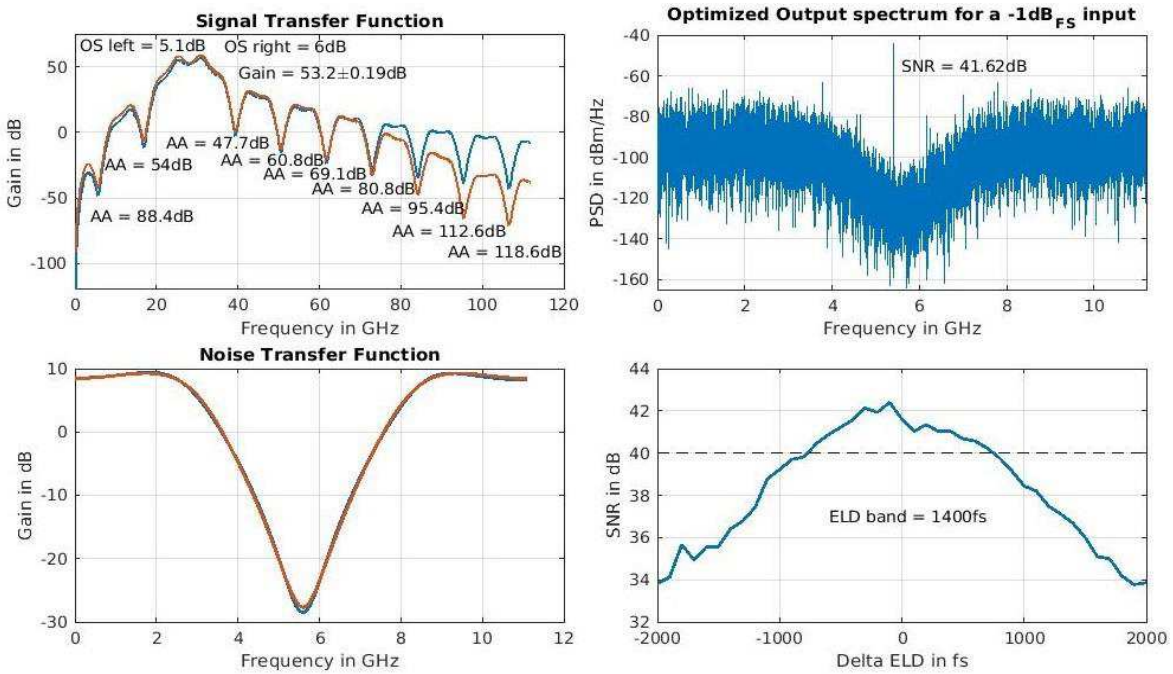

Figure 5-43: Performance summary of the SDM final optimization

One interesting outcome is on the different loop ELDs. From ELD_1 to ELD_3 respectively, they are simulated to be $2.47 \times T_S$, $2.37 \times T_S$ and $1.31 \times T_S$. As expected, ELD_3 is significantly reduced due to the delay of the real feedback pulse compared to the ideal RZ one. But this reduction is significantly less for ELD_1 and ELD_2. This gives some additional room for the implementation of the two most outer feedback paths.

Overall, the optimized modulator is very close to the target one and the explanation for the remaining differences is clear. Despite the feedback pulses being very different from an ideal RZ pulse, this does not prevent from matching the desired STF and NTF. Whatever might be the difference, it is compensated by adjusting the coefficients and individual ELDs during optimization.

5.2.3 Transient simulations

The optimization process just described allows to account for many implementation impairments in a very systematic way. Unfortunately, at the time of the chip tape out, time was missing to develop further this procedure and optimization was done by hand, leading to a largely sub-optimal result. The best simulation output spectrum is plotted in Figure 5-44. It results from a transient simulation without thermal noise, the receiver's thermal noise being specified separately from the quantization noise.

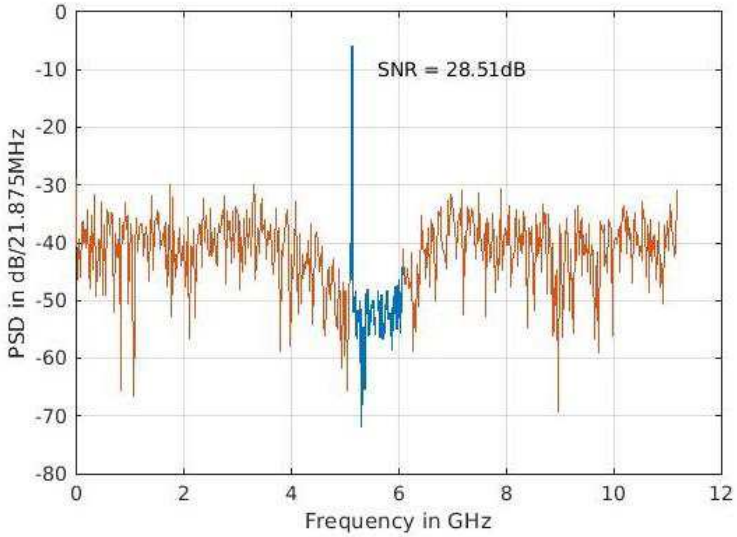

Figure 5-44 : Modulator output spectrum from transient post layout simulation

The performances are obviously not as good as expected. One could argue that it is barely better than a simple 1.5bit quantizer and an OSR of 11.2 without noise shaping. Such a configuration could pretend to have an SNR of about 20dB. Nonetheless, this shows that the approach, while being very challenging, is feasible.

The input DR was scaled such that the quantization noise is 12dB below the thermal noise. Since this is roughly the degradation showed by the electrical simulation; it means the quantization noise is about equal to the thermal noise, and will contribute to a 3dB NF degradation, going from the targeted 10dB to 13dB. Because the targeted quantization noise contribution was almost zero, the large drop in performances of the ADC only brings an acceptable system performance degradation that keeps it functional. The maximum output SNR is in the 25dB range for a single receiver, and a simple 8 element beamformer, bringing 9dB of SNR improvement, would allow the use of a 64-QAM modulation with an effective data rate of 5Gbps over the 1GHz bandwidth of the channel. While this is below expectation it is more than enough for a proof of concept.

The sub-optimal system optimization is only a portion of the performance loss explanation. More generally there are three main sources of degradation. The first one has to do with the very nature of the quantizer. The second one is sub-optimal setting in simulation. The last one is sub-optimal optimization and implementation.

5.2.3.1 Quantizer limitations

The quantizer comparator architecture was chosen for its ability to perform a fast comparison. This was mandatory in order to meet the challenging feedback loop timing. It was chosen to exploit its intrinsic sampling ability. This choice was motivated by getting rid of a sample and hold circuit and all the circuitry required to control it. Some optimization was made such as setting the source follower output common mode high for larger current in the comparator input pair. This current is then integrated on

the parasitic gate capacitance of the reset NMOS transistors. These transistors then inject this integrated image of the input signal into the regenerative latch until it diverges, which marks the sampling instant.

This mechanism introduces two effects. First, the input signal undergoes an integration that is not taken into account. In itself this is not a blocking point, it needs only to be accounted for in the modulators model used for optimization. This integration is done at best on a window as wide as the rising edge (around 7ps). This would have a low pass characteristic but with a very high cut off frequency. This would most likely have a minor effect, but the analysis needs to be done to confirm this intuitive answer. This is left to future work.

The second effect is the signal dependency of the sampling instant. This has multiple effects like introducing some non-linearity or signal dependent aperture jitter. While this is undesirable, it probably has a limited impact since happening at the quantizer level. Hence it will be attenuated by the NTF. The most critical effect is that the sampling instant is not properly define. In the present case, where the feedback timing is critical, having a signal dependent variability in the sampling instant makes it difficult to ensure this timing accuracy. There is probably an average sampling instant within the sampling clock cycle but, even in simulation, it is very difficult to evaluate.

This aspect was clearly overlooked when choosing the comparator architecture. It is left to future work to investigate for architectures with a fast enough comparison time and a properly defined sampling instant.

5.2.3.2 Sub-optimal simulation settings

In order to compensate for process variations, and also to have some room for experimental investigations, a large amount of configurability was introduced in the design. The center frequency and quality factors of each resonator are controllable. The feedforward coefficient and the feedback delays are adjustable and so on. The end result is that the modulator has a total of 144 control bits corresponding to the tuning of 17 design parameters. Obviously, that is way too many possible configurations to test all of them individually.

Thanks to the acquired understanding of the system, very efficient methods can be elaborated to find a near-optimal, if not optimal, configuration by testing only a limited amount of the possible configurations. In simulation though, this is still a challenge since a single post layout simulation of about 50ns, to perform a meager 1024 points DFT, takes about two days. Hence, only a very limited amount of these simulations could be run, most likely leading to only a sub-optimal configuration.

In practice, the only parameters that were adjusted are the feedback delays. As previously explained, this was necessary because the quantizer sampling instant is not properly defined. No time was left to investigate the potential impact of the feedforward coefficient of the resonator's quality factors for example. It is very likely that an optimal configuration that will lead to better performances exists, but it is simply too time consuming to find it in simulation. On the contrary, this is perfectly feasible in the lab while measuring the chip and will be done once it is available for testing. Hopefully, this will be eased thanks to the added testing features and the proposed calibration procedure.

5.2.3.3 Sub-optimal optimization and implementation

The chip was sent for fabrication on a Multi-Project Wafer run (MPW), hence the tape out date was fix and the design time limited. Other than the sub-optimal system optimization mentioned above, the design of the digital portion of the loop was also rushed, and too few design optimization iterations on post layout simulation were run. This results in frequent corruption of the feedback pulses caused mainly by misalignment between the clock and data provided to the DACs. This happens mostly when the quantizer regeneration time is a little too long. The lack of optimization is especially costly because running digital gates at 22.4GHz is really at the limit of the CMOS 28nm FDSOI process used for this implementation.

The most outer feedback loop imperfection will have the strongest impact on the SQNR, since the in-band noise it introduces will see no shaping at all. This could have been mitigated with a conjunction of a better system and implementation optimization. In the implemented design, the ELD of this most outer loop is only about $2.25 \times T_S$, while the ELD on the most inner loop is $1.39 \times T_S$. With a difference of less than one clock cycle between the two, it was chosen to implement the delay simply with inverters.

The improved ELD optimization methodology provides a value of $2.47 \times T_S$ for ELD_1 and $1.31 \times T_S$ for ELD_3. Now, with a difference of $1.16 \times T_S$, this delay could have been implemented using a latch. This would have allowed to deliver a data independent of the comparator regeneration time to the most outer feedback DACs and to significantly reduce the performance loss.

A second mistake was done while trying to achieve the targeted power efficiency. An initial target of 15mW was split in two halves, 7.5mW for the feedforward analog path and 7.5mA for the feedback digital path. The design was started by the analog part. To fit into this power budget, it required to constrain the amount of trans-conductance on the gm-cells, therefore limiting the maximum achievable values for c_1 and c_2 feedforward coefficients. The direct consequence is a reduced amplitude at the quantizer input which leads to more frequent long regeneration times and corrupted feedback pulses. It turned out to be impossible to keep the power budget of the digital part within 7.5mW. High speed logic running at 22.5GHz rapidly draws a significant amount of power and the final power consumption ended up around 40mW, for a total of about 50mW. This is still a very attractive number for a full 1GHz bandwidth receiver at 28GHz center frequency, including the analog to digital conversion. But the power split becomes very uneven. It would have probably been interesting to implement larger c_1 and c_2 coefficients, even at the cost of higher power, to obtain larger amplitudes at the quantizer input. Simply doubling the power consumption on the analog could double c_1 and c_2, quadrupling the quantizer input amplitude, while increasing the overall power consumption by only 15%. This estimation is far too quick but gives an idea of the tradeoff that was poorly made in the early design stage.

This also questions the choice of having a purely passive first resonator. This choice was mostly done to have only two active stages for better power efficiency. The downside was to have many constraints imposed on that first resonator, input matching, noise performances, single ended to differential conversion and adequate value for capacitive feedback. Unable to properly satisfy all of them, the noise performance ended up sacrificed.

This also have the downside of sending back some of the out of band shaped noise towards the antenna. Even though it is at a very low level and out of band, if all the antennas of the array start radiating some out of band power, it can exceed the specified limit. All signals should be uncorrelated and should not form any beam, but it would be hard to ensure it at all times.

The architectural choices were made with the available information at the time and with sound arguments. Using the acquired experience since then, it appears it would be sensible to leave the input matching network out of the loop and have an active first resonator. That would allow for a better quality factor of this resonator while de-correlating some of the design challenges. All of that would come at a limited cost on the total power consumption, while benefiting from a significant performance improvement.

5.2.4 Conclusion

The presented optimization methodology allows for a fairly systematic way to take into account and compensate for multiple implementation impairments such as transfer function deviation from ideal model, feedback pulse shape or variation of the feedforward coefficients. This compensation is achieved by adjusting the feedback coefficients and timing using the tools developed for the optimization of the

ideal model. Because of time limitation, this methodology was not developed at the time of the tape out. As a result, the design sent for fabrication is somewhat sub-optimal. Nonetheless, the hope for this design is to be good enough for a proof of concept.

The systematic optimization process that was developed afterward was used to evaluate the quality of the actual design and initiate new thoughts on future improvements. While overall the architecture seems good, it could benefit from some tweaks around the first resonator. The implementation would benefit from more optimization, in particular on the digital feedback loop that was unfortunately developed under a stringent time constraint. Finally, some investigations on a different quantizer architecture could bring significant improvements.

5.3 Test Chip Top Level

The receiver was embedded on a test chip. The testing purpose is two folds, testing the receiver by itself and test it as the element of a digital beamformer. In particular, it is important to confirm that this architecture would be good enough to exploit the targeted efficient beamforming using a combination of discrete time delays and digital phase shifts. Eight of the receivers presented above were integrated on a single chip to form an eight-channel digital beamformer. First, the block diagram of the top level of test chip will be detailed, followed by the top view of the layout.

5.3.1 Top Level Block Diagram

The test chip top level block diagram is given in Figure 5-45. A left to right description will be made first, following the signal path, and then, from right to left, following the control path. The eight analog inputs arrive on the left, each of them feeding a Single Receiver (SRx). Thanks to time the interleaved nature of the quantizer, the SRx outputs are already de-multiplexed by a factor of two. This signals then go through two layers of de-multiplexing by eight. This results in a total de-multiplexing factor of 128 and a 256-bit wide data bus, clocked at 175MHz for each SRx. This is an adequate rate to interface with a standard digital interface. This digital interface is the portion on the right, circled by the red dotted line in Figure 5-45.

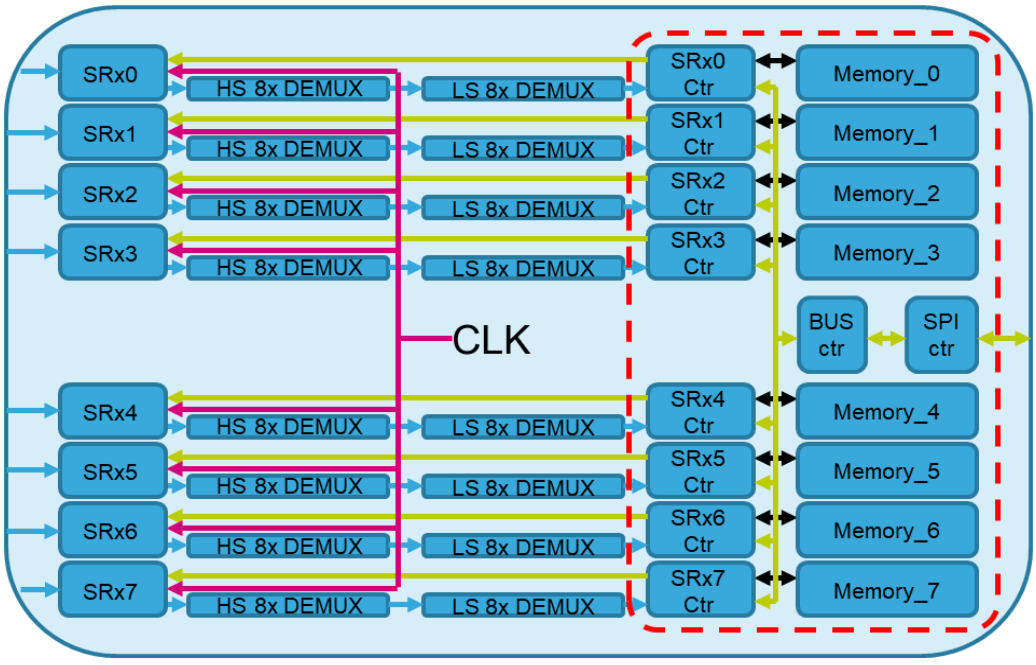

Figure 5-45: Test chip top level block diagram

The purpose of this digital interface is to store the data from an acquisition that will fill an internal memory. Once this is done, this internal memory can be read at a slower rate from the outside. The data, when entering this digital interface, are first received by an individual SRx controller that will store them in a dedicated 128kB memory. The SRx controller can then send the stored data to the outside world using the common bus and the Serial Parallel Interface (SPI).

Now following the control path, it goes as follow. All controllers in the test chip are a chain of slaves. The SPI controller is slave to the external master controller. The BUS controller is slave to the internal SPI controller and the SRx controllers are slave to the BUS controller. The SPI controller receives a 16-bit address and a 16-bit payload. The address first 4 bits will tell the BUS controller which SRx to communicate with. The last 8 bits of the address and the payload represent the data to be sent to that SRx. This can be done in unicast mode, addressing only one SRx controller, or in broadcast mode, addressing all SRx controllers at the same time. The 8 address bits tells which SRx register must be written, and the 16-bit payload is what should be written in that register. There are a total of 144 register bits that correspond to the SRx control bits and 4 bits that correspond to a command register. Based on that command register value, the SRx controller can perform different tasks such as launching an acquisition, reading or writing the memory or sending data to the BUS, the SPI controllers and finally the outside world. Thanks to this approach each channel was made fully independent to ease debugging.

The digital interface will not be described further, but this shows that even the simple task of storing an acquisition in a memory and reading it later at a slow rate require some level of complexity for this digital interface.

Overall, the test chip embeds a total of 1MB of memory which allows to store a sequence of $23.4\mu s$ of the full 8 channels beamformer output for a total of 512ksample per channel. These data can then be treated externally to perform the digital processing portion of the beamforming. For a 100MHz channel, this corresponds to receiving about 2000 symbols. This is enough to perform some meaningful statistical measurements such as Error Vector Modulation (EVM) on 16-QAM modulation and maybe even 64-QAM ones.

The last thing is the clock distribution. It is part of the challenge of beamforming systems using large antenna array. Here, this difficulty was not covered because of the time constraint. The chip is fed by a differential 22.4GHz clock. This clock is then passively slip in 8 to feed the individual channels. This approach is made possible by deporting the challenge on the external clock driver that must provide a clock with a power of about 10dBm. A fully integrated approach would require adding a PLL and an active distribution network to properly feed the receivers. The added power consumption would then be split between each channel to have a more realistic evaluation of the power consumption per channel.

5.3.2 Top level layout

The top layout of a single receiver is given in Figure 5-46. On top of the analog feedforward layout that was provided in Figure 5-39, Figure 5-46 displays the layout of the digital part of the loop as well that the bumps that will allow the connection with a flip-chip package. The analog and digital part of the loop have separated single power supply bumps. The three left bumps are the Ground Signal Ground (GSG) input and all the remaining bumps are for ground connection. The overall strategy is to have as many ground bumps as possible to provide a reference with the lowest possible impedance, and to decouple everything versus this reference.

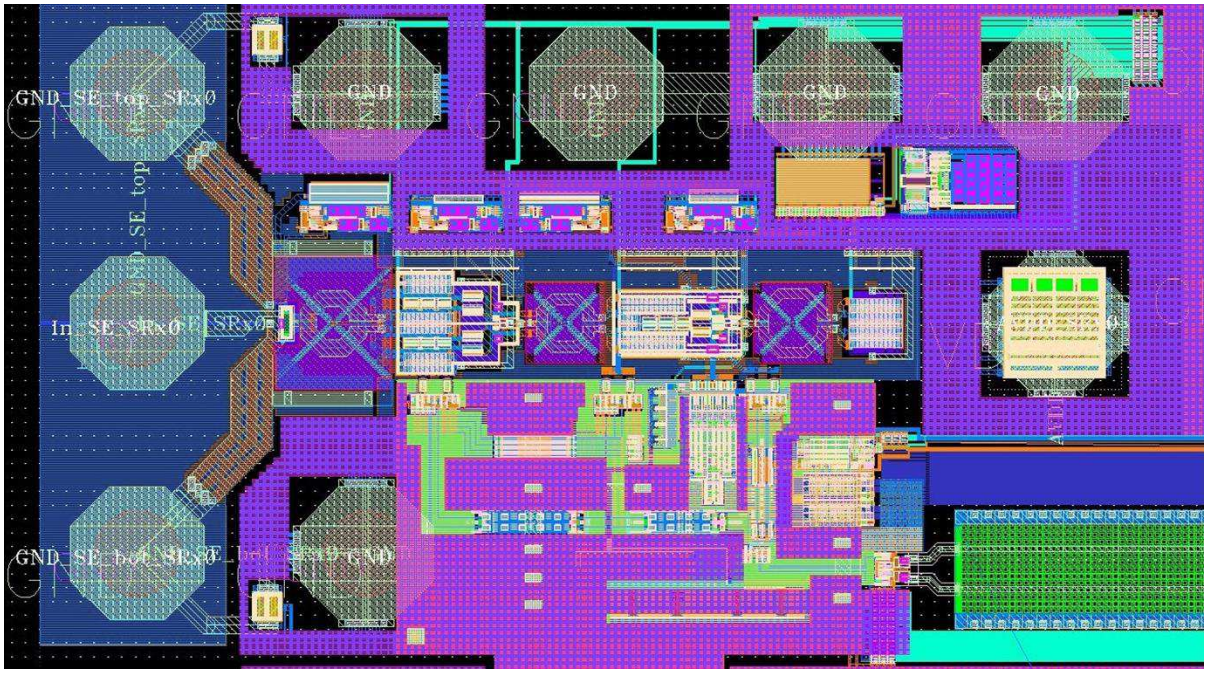

Figure 5-46: Single Receiver's layout

Figure 5-47 displays the top layout of the whole test chip. For better display, it has been rotated clockwise by 90° compared to its block diagram counterpart from Figure 5-45. The RF inputs are now on the bottom side and the digital interface on the top, and the clock distribution sits in the middle. The die size is $2.5 \times 3.8 = 9.5mm^2$ with a split of about half of it for the receivers and the clock distribution and the other half for the digital interface.

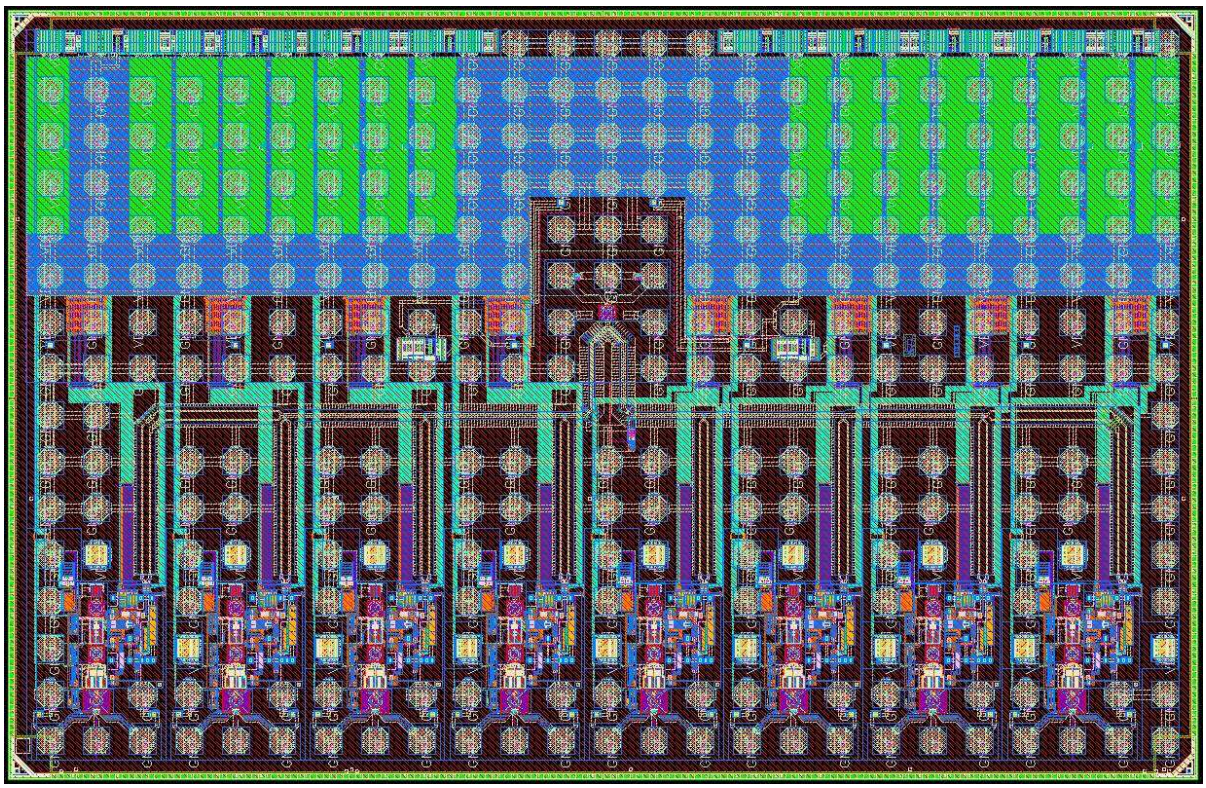

Figure 5-47: Test chip layout

The die was fabricated using the 10-metal layer FDSOI 28nm CMOS process from STMicroelectronics. It will be assembled with a custom flip-chip package currently under design. The package chip will then be measured using a custom test board and test software. An initial design of the test board was done as well as a draft version of the software. As soon as the package will be available, everything should be ready for testing.

5.3.3 State of the Art comparison

Comparison with the state of the art is always a difficult task since no contribution target the exact same specifications and applications. Many Figures of Merits have been used to help with the matter. Nonetheless, it is difficult in the current case since there are very little contributions of full receivers including the ADC at 28GHz.

On one hand, contributions propose either the RF front end or the ADC, and only rarely both. There are also full TRX systems, but they are not monolithically integrated, and their contributions rely more on the challenge of assembling a functional system, than on the performances of the building blocks. For that reason, they will not be considered here.

On the other hand, since the proposed architecture merges the RF front end with the ADC, it is impossible to separate the performances of these two parts for a fair comparison. For these reasons the following analysis is to be taken with a grain of salt. Its purpose is to highlight the potential of the proposed architecture more than to quantify precisely the potential performance improvements.

Table 5-5: State of the art comparison

	[5-10]	[5-11]	[5-12]	[5-13]	This Work
Architecture	HBF sub-array	ABF	HBF Fully connected	DBF	DBF
NB of elements	32	8	8	16	8
NB of beams	2	1	2	4	N/A
True time delay	no	no	no	yes	yes
Input	RF	RF	RF	IF	RF
Output	Analog IF @3GHz	Analog Base Band	Analog low-IF @50MHz	Digital Base Band Beamformed	Digital Base Band raw
Monolithic	yes	yes	yes	yes	yes
Process	130-nm SiGe BiCMOS	28nm CMOS	65nm CMOS	40nm CMOS	28nm FDSOI CMOS
Fc	28GHz	25.8GHz-28GHz configurable	25GHz-30GHz configurable	1GHz	28GHz
BW	1.5GHz	500MHz	100MHz	100MHz	1GHz
NF (dB)	6.6	6.7	7.3dB	N/A	13dB
IIP3 (dBm)	-11.9*	-59.9	-19.4*	N/A	-25dBm**
Power per element per beam (mW)	51.55	50	21.25	7.1	50

* Estimated from 1dB compression point plus 9.6dB

** Target value, not evaluated in simulation

The design proposed in [5-10] is an RF analog beamformer with two 16 elements sub-arrays used to produce two beams using a dually polarized antenna array. In the same article they also demonstrate a module with 4 chips and 64 in package antennas providing the possibility to have from 8 separate beams of 16 elements each to 2 separate beams of 64 elements each, providing interesting flexibility despite their HBF sub-array architecture. Overall, they demonstrate very good performances in all areas. Their power consumption is in the range of what was estimated in Chapter 3. Their output is analog at a 3GHz IF, meaning that some base band processing and conversion to digital is missing for a complete picture. The chip is implemented in a 130nm SiGe BiCMOS process which is pretty good for RF but les adequate for ADC and almost incompatible with digital processing. This means that this approach cannot propose a fully integrated solution. Also, their sub-array approach does not allow to scale sufficiently the total throughput when increasing the number of antennas and the number of beams.

The solution proposed in [5-11] is very similar but with a pure ABF architecture and a Homodyne approach with a single die and an antenna in package strategy. This is inadequate for base stations but could be of interest for UE. Their design seems to suffer in terms of linearity, but this can probably be fixed in the design. Their proposition also does not include the ADC, but it is implemented in a 28nm CMOS process which is also good for that purpose and could also include some digital processing, even though it might be limited to implement the full modem.

These two solutions are the only ones that provide only instantaneous bandwidth to be compatible with a base station. The solution proposed in [5-12] is a fully connected HBF with a 100MHz of bandwidth but its center frequency is tunable over a wide range. This is clearly the kind of solution that would fit a UE application. As for the two first designs the ADC is missing and the 65nm CMOS process used is probably good enough for the ADC but inadequate for digital processing.

Finally, [5-13] is at the other end of the spectrum, proposing only the ADC and the DBF processing. As already mentioned, the work presented in this manuscript is inspired in several ways by this contribution. Their design shines because both the ADC and the DBF processing are implemented. The BSP technic used demonstrate both low power and true time delay capability. Their biggest limitation is their bandwidth and would probably be a significant challenge to push further than they already did with their architecture. Their solution highlights two strong points of DBF, its accuracy and is potential to scale up. Just by implementing four beams, the power consumption per element per beam falls to 7mW. Of course, the RF front end is missing, but any additional consumption per element gets divided by four. Since this approach also does not need any phase shifting of weighting in the RF front end, it is very likely that it could be made very low power.

Their DBF processing consumes 49mW per beam. This number can be used to estimate DBF processing for the solution proposed in this manuscript. The bandwidth is 10 times larger, but it is made of only 8 elements, half of their solution, and the OSR is only 11.2 instead of 20. The processing for one beam can be estimated to be about 137mW per beam. Targeting 12 beams as per the system analysis would lead to 1.65W for the digital processing. Adding the 50mW*8 of the front end gives 2W of total power consumption, or 21mW per element per beam. Scaling up to 256 antennas this gives 65W. This is of the right order of magnitude compared to the power available on a streetlamp, even though it is probably still too high for a four-sector cell, also needing some additional power for the modems. Nonetheless, this clearly shows that the proposed approach is realistic and only needs some reasonable power efficiency improvement to achieve the targets of 5G.

5.4 CONCLUSION

In this chapter, starting from the architecture and implementation was described step by step. Along the way, it was ensured that the choices that were made were compatible with the architecture. This was necessary mostly to confirm the feasibility of two choices, the use of capacitive feedback DACs and

that of transformer-based resonators. This last choice proves to be very fruitful. Along the classical advantages of providing a simple way to bias circuits and to offer reduced magnetic coupling with its surrounding, a new usage was discovered, allowing for significant power saving. This, with several other design tricks, allowed to reach the desired performances for the building blocks.

After working at the building block level, all of them were assembled to form the modulator. From electrical simulations, its various transfer functions of interest and feedback pulse shapes were extracted. These extracted characteristics were injected in place of their ideal equivalent in the architectural model to run the final optimization. This step allows to adjust the feedback coefficients and ELDs to a near optimal solution that account for many implementation impairments.

The final test was to run a transient simulation of the modulator. While the output spectrum clearly shows noise shaping, the performances are relatively far from expectations. Several reasons were identified to explain this discrepancy. Overall, the implementation is lacking post layout optimization. This was in part due to the limited time available before the tape out date, but it is not the main reason. It was shown that the original power consumption target was too optimistic and led to maybe the wrong choice of using a fully passive first stage. Also, the chosen quantizer topology suffers from a sampling instant dependency to the input signal. Since the architecture requires an accurate feedback timing, it is very likely that this has a detrimental effect on the performances. The confirmation of these hypothesis is left for future work as well as finding solutions for them.

Despite the limited performance, the proposed implementation was deemed good enough for a proof of concept and sent for fabrication. The test chip includes eight receivers, embedded test memory and a digital interface. This forms an eight-channel digital beamformer that has been built in the FDSOI 28nm CMOS process from STMicroelectronics. A custom package to measure the prototype is currently in fabrication, hoping it will allow to gather more inputs for future evolutions of the design.

5.5 REFERENCES

[5-1] W. Zhuo et al., "A capacitor cross-coupled common-gate low-noise amplifier," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 52, no. 12, pp. 875-879, Dec. 2005

[5-2] Xiaoyong Li, S. Shekhar and D. J. Allstot, "G_m-boosted common-gate LNA and differential colpitts VCO/QVCO in 0.18-/spl mu/m CMOS," in IEEE Journal of Solid-State Circuits, vol. 40, no. 12, pp. 2609-2619, Dec. 2005

[5-3] M. Vigilante and P. Reynaert, "On the Design of Wideband Transformer-Based Fourth Order Matching Networks for E-Band Receivers in 28-nm CMOS," in IEEE Journal of Solid-State Circuits, vol. 52, no. 8, pp. 2071-2082, Aug. 2017

[5-4] J. Harrison, M. Nesselroth, R. Mamuad, A. Behzad, A. Adams and S. Avery, "An LC bandpass ΔΣ ADC with 70dB SNDR over 20MHz bandwidth using CMOS DACs," 2012 IEEE International Solid-State Circuits Conference, San Francisco, CA, 2012, pp. 146-148

[5-5] K. Kwok and H. C. Luong, "Ultra-low-Voltage high-performance CMOS VCOs using transformer feedback," in IEEE Journal of Solid-State Circuits, vol. 40, no. 3, pp. 652-660, March 2005.

[5-6] J. Yuan and C. Svensson, "High-speed CMOS circuit technique," in IEEE Journal of Solid-State Circuits, vol. 24, no. 1, pp. 62-70, Feb. 1989.

[5-7] D. Schinkel, E. Mensink, E. Klumperink, E. van Tuijl and B. Nauta, "A Double-Tail Latch-Type Voltage Sense Amplifier with 18ps Setup+Hold Time," 2007 IEEE International Solid-State Circuits Conference. Digest of Technical Papers, San Francisco, CA, 2007, pp. 314-605.

[5-8] M. van Elzakker, E. van Tuijl, P. Geraedts, D. Schinkel, E. Klumperink and B. Nauta, "A 1.9μW 4.4fJ/Conversion-step 10b 1MS/s Charge-Redistribution ADC," 2008 IEEE International Solid-State Circuits Conference - Digest of Technical Papers, San Francisco, CA, 2008, pp. 244-610.

[5-9] S. Le Tual, P. N. Singh, C. Curis and P. Dautriche, "22.3 A 20GHz-BW 6b 10GS/s 32mW time-interleaved SAR ADC with Master T&H in 28nm UTBB FDSOI technology," 2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), San Francisco, CA, 2014, pp. 382-383.

[5-10] B. Sadhu et al., "A 28-GHz 32-Element TRX Phased-Array IC With Concurrent Dual-Polarized Operation and Orthogonal Phase and Gain Control for 5G Communications," in IEEE Journal of Solid-State Circuits, vol. 52, no. 12, pp. 3373-3391, Dec. 2017.

[5-11] H. Kim et al., "A 28-GHz CMOS Direct Conversion Transceiver With Packaged 2×4 Antenna Array for 5G Cellular System," in IEEE Journal of Solid-State Circuits, vol. 53, no. 5, pp. 1245-1259, May 2018.

[5-12] S. Mondal, R. Singh, A. I. Hussein and J. Paramesh, "A 25–30 GHz Fully-Connected Hybrid Beamforming Receiver for MIMO Communication," in IEEE Journal of Solid-State Circuits, vol. 53, no. 5, pp. 1275-1287, May 2018.

[5-13] S. Jang, R. Lu, J. Jeong and M. P. Flynn, "A 1-GHz 16-Element Four-Beam True-Time-Delay Digital Beamformer," in IEEE Journal of Solid-State Circuits, vol. 54, no. 5, pp. 1304-1314, May 2019.

6 CONCLUSION

In this manuscript, a solution for millimeter wave 5G small base station receivers using digital beamforming with a large antenna array was proposed. This solution is based on a band pass sigma-delta modulator receiver performing RF sampling.

To propose a solution with adequate performances for the targeted application, in chapter 2, starting from 5G key performance indicators for enhanced Mobile Broad Band, an in-depth system analysis was performed. The result of this analysis allowed to state working hypotheses, mainly through establishing the link budget and evaluating the multiple operator interferences. Another important outcome is that the implementation challenge is not only on the receiver's analog part, but also on the large amount of digital processing to be performed under stringent timing constraints.

In chapter 3, it was investigated if the current state of the art was adequate to build a receiver compatible with the system performance required. This chapter focus was on the analog part of the receiver. First, the receiver as whole was specified in detail. Then, using a Near-ZIF architecture, the building blocks were specified. The following state of the art survey showed no blocking point for such a receiver.

The purpose of chapter 3 was to ensure the system analysis results were compatible with a physical implementation, but it did not address the digital processing challenge. This was investigated in chapter 4, through the proposition of a different receiver's architecture. It was shown that sigma-delta modulators had the potential to significantly reduce the digital processing complexity. It is possible thanks to two major properties, over-sampling, and low resolution quantizer. The over-sampling allows to implement free discrete time delays for wide band beamforming. The low resolution quantizer allows for cheap multiplexer-based multiplication, significantly lowering the digital processing complexity.

The proposed architecture pushes this idea to its limit by performing direct sampling of the RF signal. The receiver, for a single antenna element, is then reduced to a sigma-delta modulator. The proposed modulator is a band pass continuous time one. To keep the sampling frequency at an acceptable level, the receivers perform sub-sampling, digitizing the signal in its third Nyquist zone. One important result is the new excess loop delay compensation technique developed for sub-sampling modulators. Combined with an "$f_s/4$" architecture, this technique relaxes the timing constraints and is key for the proposed architecture.

Finally, Chapter 5 describes the proposed implementation of this architecture. Despite the architecture was optimized to ease implementation, it remained a difficult challenge. Two major implementation choices are at the heart of the proposed design. The first one, is the use of transformer-based resonators and the second one, is the use of a two-time interleaved quantizer. The transformers provide an additional degree of freedom through their coupling factor. It was used either to reduce power consumption, or to adjust the feedback elements size for a realistic implementation. The two-time interleaved quantizer is possible thanks to the "$f_s/4$" architecture and allow to significantly relax the required comparison time.

The proposed architecture was sent for fabrication, in ST 28nm FDSOI CMOS process, on a test chip made of eight parallel receivers. The samples have been manufactured but have not been tested yet because of the need for a custom package, currently under development. The top-level simulation of the modulator, using post layout extraction of the transistors and electromagnetic extraction of the passive devices, were performed. Despite lower performances than expected, these results demonstrate the feasibility of the proposed architecture. An initial analysis of the performance loss led to some interesting ideas for future implementations. While this work tried to be as extensive as possible, it is clearly incomplete and leaves the door open for many more improvements and investigations in all the covered areas.