LUND UNIVERSITY

Frequency Generation and Baseband Filters for mm-Wave 5G and 6G Transceivers

Gannedahl, Rikard

2024

Frequency Generation and Baseband Filters for mm-Wave 5G and 6G Transceivers

Rikard Gannedahl

LUND
UNIVERSITY

Department of Electrical and Information Technology

Lund, May 2024

Department of Electrical and Information Technology

Lund University

ISBN 978-91-8104-074-6 (printed)

ISBN 978-91-8104-075-3 (digital)

ISSN 1654-790X, No.174

Printed in Sweden by *Tryckeriet i E-huset*, Lund University

Lund 2024

Abstract

In order to keep up with the constant demand for higher data rates, the fifth generation of mobile communication (5G) introduced the use of mm-wave carrier frequencies from 24 to 71 GHz. Plenty of new frequency spectrum then became available, which has allowed for channel bandwidths of several hundreds of MHz. 6G is projected to continue this trend with even higher carrier frequencies and wider bandwidths, reaching carrier frequencies above 100 GHz and bandwidths in the multi-GHz range. However, this creates new challenges for circuit designers, as the performance of radio circuits typically degrades with increasing frequency. Two critical components in radio transceivers whose performance are highly affected by the increase in carrier frequency and bandwidth are frequency generators and baseband filters. The former generates a local oscillator (LO) signal required for frequency translating the data signal to/from the baseband from/to the carrier frequency, and the latter is used to separate the desired signal from undesired interference and noise in a receiver and to prevent leakage of undesired spectrum content to nearby channels in a transmitter. The design of the frequency generation is also made much more complicated due to mm-wave 5G and 6G communication relying on beamforming, in which the signals from multiple antennas are combined to overcome the high path loss at these frequencies. The LO signal must then be distributed to multiple frequency converters while retaining a constant relative phase shift. If the beamforming is done using so-called LO beamforming, this relative phase shift must also be tunable in a very accurate manner.

In this thesis, five research papers are included; two concern mm-wave frequency generation, two concern multi-GHz integrated baseband filters, and one is about system-level simulations of beamforming receivers. The thesis is divided into two parts, with the first part providing an introduction and context to the conducted research, while the second part consists of the included papers.

Paper I presents a frequency generation circuit for 28-GHz LO beamforming sliding-IF transceivers. An external 7-GHz signal is first phase shifted by an injection-locked oscillator and then fed to an injection-locked frequency tripler. A harmonic mixer is used as a phase detector to measure the applied phase shift, enabling automatic phase tuning. The phase detector can also be used to automatically tune the oscillators to obtain injection lock. Additionally, a sliding-IF receiver is implemented to properly test the frequency generator.

Paper II presents a modular system-level testbench for sub-THz 6G beamforming receivers, implemented in MATLAB/Simulink. The testbench models

the analog circuit blocks with high fidelity, and can thus be used to investigate impacts of circuit non-idealities on the performance of the whole system. The effects of beam squint, ADC resolution, phase noise, baseband filter type, and interfering beams are simulated for a 32-element linear array.

Paper III describes a 28-GHz differential-to-quadrature injection-locked frequency tripler, intended for direct-conversion transceivers. A dual-injection scheme is used to maximize the harmonic rejection, and a mixed-signal feedback system minimizes the quadrature error by automatically tuning the frequency tripler so that its free-running oscillation frequency coincides with the third harmonic of the input signal.

In **Paper IV**, two differential multi-GHz 5^{th}-order integrated baseband filters are presented, one active Gm-C filter and one passive LC filter, intended for 6G applications, fabricated in a 22-nm FD-SOI CMOS process. The active filter is based on Nauta's transconductor and utilizes back-gate biasing to achieve state-of-the-art performance. The passive filter uses overlapping inductors, resulting in large mutual inductance and reduced footprint. Owing to this technique, the chip area of the passive filter is similar to that of most active filters, while providing clear benefits in power consumption and dynamic range.

In **Paper V**, the passive filter in Paper IV is further investigated and improved. A lumped model of the overlapping inductors is derived and used to develop a capacitive cancellation method to reduce stopband peaking, which is verified using EM simulations. To further improve the stopband performance, the overlapping inductors are redesigned in an 8-shape to reduce coupling between different inductors in the filter. The new filter reduces the stopband peaking by almost 30 dB compared to the original filter in Paper IV.

Populärvetenskaplig sammanfattning

För många är det säkert bekant att FM-radio kan sändas på frekvenser som 105,6 MHz, medan om pratar om Bluetooth och Wi-Fi nämns ibland 2.4-GHz- och 5-GHz bandet. Vad detta refererar till är den så kallade bärvågsfrekvensen. Kring denna bärvågsfrekvens upptar signalen man vill ta emot eller sända en viss bandbredd. Ju större bandbredd, desto högre datahastigheter kan man uppnå. FM-radio använder en bandbredd på 0,2 MHz, så radiostationen ovan kommer ta upp ett frekvensspann mellan 105,5 och 105,7 MH, medan nästa station kan sända på 105,7–105,9 MHz, och så vidare. Bluetooth och Wi-Fi har högre bandbredder, från enstaka MHz i Bluetooth till mer än 100 MHz i de senaste versionerna av Wi-Fi. Precis samma princip med bärvågsfrekvens och bandbredd används av våra telefoner när vi surfar, ringer eller SMS:ar.

I dag använder vi våra telefoner till alltmer dataintensiva uppgifter, såsom att streama film och musik och ringa videosamal, något som kräver allt större bandbredd. Problemet är att signalerna från två telefoner inte får ha överlappande frekvenser, vilket gör att de användbara frekvenserna håller på att ta slut. Därför tog man ett drastiskt beslut vid lanseringen av 5G: man öppnade upp för att använda så kallade millimetervågsfrekvenser, vilket är frekvenser mellan 24 och 300 GHz, cirka 10 till 100 gånger högre frekvenser än vad man tidigare använt i mobilkommunikation. Här finns enorma mängder med oanvända frekvenser, vilket löser bandbreddproblematiken. I 5G har man nöjt sig med bärvågsfrekvenser under 100 GHz, men i 6G förväntas frekvenser mellan 100 och 300 GHz också användas. Det finns dock en anledning till att man tidigare undvikit dessa frekvenser - förlusterna när signalen färdas från sändare till mottagare ökar drastiskt. För ett givet avstånd blir den mottagna signalstyrkan bara en hundradel så stor om man använder en tio gånger högre frekvens! Men det finns en lösning. Genom att använda flera antenner samtidigt som fångar upp mer av den sända signalen kan man öka den mottagna signalstyrkan. På samma sätt kan man även sända ut en starkare signal genom att använda flera antenner i sändaren. Kruxet är att när man använder flera antenner kommer signalen, likt ljuset från en ficklampa, skickas ut i en stråle och det är bara om mottagaren faktiskt befinner sig i den här strålen som signalen blir starkare. Som tur är kan strålens riktning styras genom att applicera en så kallad fasvridning vid varje antenn.

Bakom antennerna sitter radiochip som ansvarar för att förstärka, fasvrida och processa signalen, samt filtrera bort oönskade signaler till/från andra användare. I avhandlingen presenteras flera nya lösningar för dessa chip som är lämpliga för 5G-

och 6G-tillämpningar. Fokus har legat på tre olika områden: systemanalys, frekvensgenerering och kanalfiltrering.

Vad gäller systemanalys har vi utvecklat ett program för datorsimuleringar som kan användas i ett tidigt stadium i konstruktionsarbetet för att analysera hur olika kretselement påverkar prestandan hos ett stort antennsystem. Det unika med programmet är hur realistiskt kretselementen beter sig samt användarvänligheten. Tanken är att ingenjörer från olika discipliner ska kunna använda programmet för att analysera sina system.

Frekvensgenerering är en del av chipet som ansvarar för att generera den bärvågsfrekvens man vill ta emot eller sända på. Frekvensgeneratorn i en telefon där man vill använda millimetervågs-5G kan till exempel behöva generera 28 GHz. Vi har utvecklat två kretsar för detta där robusthet har varit huvudfokus. I radiokommunikation krävs ofta att chipet beter sig på ett väldigt precist sätt, men när man tillverkar chip kommer det alltid finnas slumpmässiga variationer. Därför måste man ofta utföra kalibreringar, något som tar tid och ökar kostnaderna. Våra chip kan i stället kalibrera sig själva. Den ena frekvensgeneratorn kan dessutom användas till att skapa den fasvridning som behövs för att styra signalstrålen.

Kanalfilter behövs i både sändare och mottagare och dess uppgift är att filtrera bort frekvenser som inte tillhör den egna signalen, som brus och andra signaler. När 6G lanseras förväntas bandbredden att öka till flera GHz och för den bandbredden finns det väldigt få exempel på kanalfilter i den vetenskapliga litteraturen. Vi har därför undersökt lämpligheten för två typer av filter till dessa tillämpningar; aktiva och passiva. Aktiva filter förbrukar energi, men kan göras väldigt små, medan passiva inte förbrukar någon effekt, men tar upp större chipyta. Genom att använda en innovativ teknik lyckades vi dock konstruera ett passivt filter som nästan är lika litet som jämförbara aktiva filter, vilket gör det till ett väldigt attraktivt val för 6G-tillämpningar.

Preface

This thesis summarizes the work I have done as a doctoral student at the *Integrated Electronics Systems* division at the Department of Electrical and Information Technology, Lund University, for the last couple of years. The primary focus of the work has been to design circuits for mm-wave 5G and 6G transceivers in CMOS technology and has been funded by the Excellence Center at Linköping – Lund in Information Technology (ELLIIT).

The thesis is split into two parts: Part I introduces the topics of the research papers and provides an overview of the current state of the research field, and Part II consists of the included research papers, as stated below.

List of Papers and Author's contributions

Paper I
R. Gannedahl and H. Sjöland, "An LO phase shifter with frequency tripling and phase detection in 28 nm FD-SOI CMOS for mm-wave 5G transceivers," in *Springer Analog Integrated Circuits and Signal Processing*, vol. 114, pp. 1-11, Jan 2023.
Author's contribution: I designed the circuits, performed the simulations, designed the layout, and wrote the manuscript, under supervision of the second author.

Paper II
R. Gannedahl, J. B. Asli, H. Sjöland and A. Alvandpour, "A Modular System-Level Testbench for 6G Beamforming Applications with Near Circuit-Level Fidelity," in *Proc. 2023 21st IEEE Interregional NEWCAS Conference (NEWCAS)*, Edinburgh, United Kingdom, 2023, pp. 1-5.
Author's contribution: I developed the analog part and most of the Matlab code for the testbench. I performed most simulations, co-wrote the manuscript with the second author, under supervision of the third and fourth authors, and I presented the work at the conference.

Paper III

R. Gannedahl and H. Sjöland, "A mm-Wave Differential-to-Quadrature Frequency Tripler with Automatic Locking and Quadrature Correction," in *Proc. 2023 IEEE Nordic Circuits and Systems Conference (NorCAS)*, Aalborg, Denmark, 2023, pp. 1-6.

Author's contribution: I designed the circuits, performed the simulations, designed the layout, and wrote the manuscript, under supervision of the second author. I also presented the work at the conference.

Paper IV

R. Gannedahl and H. Sjöland, "Active and Passive Integrated Filters for Multi-GHz 6G Baseband Applications," in *Proc. 2023 IEEE Asia-Pacific Microwave Conference (APMC)*, Taipei, Taiwan, 2023, pp. 524-526.

Author's contribution: I performed all the simulations, designed the layout, did all the measurements, and wrote the manuscript, under supervision of the second author. I also presented the work at the conference.

Paper V

R. Gannedahl and H. Sjöland, "Capacitive Cancellation in Compact Integrated Multi-GHz Differential Passive Baseband Filters." Accepted to *2024 22nd IEEE Interregional NEWCAS Conference (NEWCAS)*, Sherbrooke, Canada.

Author's contribution: I performed all the simulations, developed the lumped model and capacitive cancellation scheme, wrote the manuscript, under supervision of the second author. I will present the work at the conference.

Acknowledgments

As my PhD journey approaches its conclusion, there are many people I wish to express my gratitude to. Without their support, this thesis would never have come to fruition.

First and foremost, I would like to express my deepest gratitude to Henrik Sjöland, my main supervisor. Thank you for all your support, guidance and feedback during these years, going back all the way to my Master's thesis. You have always steered me in the right direction and your profound knowledge in all matters related to radio design has been truly inspiring. Thank you also to my co-supervisor, Pietro Andreani, for giving me new perspectives on circuit design and for our excellent cooperation in the analog IC course. Also, your impromptu mini-lectures in my office on obscure feedback details are always a delight.

In addition to my supervisors, I owe my gratitude to many researchers, both current and former, in the Integrated Electronic Systems group. Special thanks to my office mate, Iman Ghotbi, for his constant support and encouragement, and to Baktash Behmanesh, who always seems to have an answer to all my practical problems. I am also thankful to Masoud Nouripayam, who always manages to brighten my mood, and Hamid Karrari, who have made the countless hours we have spent together in Blåtunga less tedious. My appreciation also extends to Markus, Joachim, Ilayda, Arturo, Siyu, Sidra, Liang, Therese, Jonas, MJ, Lucas, Christian, Imad, Peng, Mojtaba, Dumitra, Lina, Sijia, and everyone else in the group who I have had the pleasure of working alongside with.

Outside of IES, there are many people who have significantly improved my day-to-day experience at the department, both professionally and socially. I would like to express my gratitude to Sirvan Abdollah Poor and Andreas Johansson for all their assistance and patience with me in the lab, and Erik Jonsson and Stefan Molund for helping me solve any computer-related issues. A huge thank you also to all the members of the second-table-on-the-right lunch crew for always bringing a smile to my face and taking my mind off my research. In particular, I want to thank Haorui, Hassan, Fatemeh, and Zahra.

Additionally, I want to thank Javad Bagheri Asli and Atila Alvandpour at Linköping University. It has been a great experience collaborating with you in our joint ELLIIT project, and hopefully we can produce some more research output in the near future.

In addition to my own research, I have done quite a lot of teaching over the years. While it has sometimes been very stressful and time-consuming, it has also been

very rewarding. So, thank you to all the students I have had the pleasure of teaching and supervising over the years. Seeing so many of you pursuing careers in circuit design, whether in academia or in industry, brings me great joy and pride.

Throughout these years, I have received endless support from my wonderful family. So, thank you to my mom Eva-Lena, my siblings Johan and Karin, their partners Josefin and Thomas, and my mother-in-law Anne. Also, I would not have been the person I am today without two of the most influential people in my life, my dad Per and father-in-law Peter, who were both taken from us at a much too young age.

Lastly, thank you Alexandra. I am forever grateful for all your love, support, encouragement, and patience throughout these years. This journey would never have been possible without you.

Rikard Gannedahl
Tågarp, April 2024

List of Acronyms and Abbreviations

AA	Anti-aliasing
ABF	Analog beamforming
ADC	Analog-to-digital converter
AM	Amplitude modulation
BB	Baseband
BER	Bit-error rate
DBF	Digital beamforming
DCO	Digitally controlled oscillator
DSP	Digital signal processor
EM	Electro-magnetic
EVM	Error vector magnitude
FDD	Frequency division duplex
FD-SOI	Fully-depleted silicon-on-insulator
FoM	Figure-of-merit
FR1	Frequency range 1
FR2	Frequency range 2
HBF	Hybrid beamforming
IB	In-band
IF	Intermediate frequency
IIP3	Input-referred third-order intercept point
ILFM	Injection-locked frequency multiplier
ILFT	Injection-locked frequency tripler
ILO	Injection-locked oscillator

LHP	Left-half-plane
LO	Local oscillator
LoS	Line-of-sight
LR	Locking range
NQS	Non-quasi static
OOB	Out-of-band
Op-amp	Operational amplifier
OSR	Oversampling ratio
PD	Phase detector
PL	Path loss
PLL	Phase-locked loop
PM	Phase modulation
PPF	Polyphase filter
PVT	Procees, voltage, and temperature
QILFT	Quadrature injection-locked frequency tripler
QO	Quadrature oscillator
RF	Radio frequency
RHP	Right-half-plane
rms	Root mean square
RX	Receiver
SFDR	Spurious-free dynamic range
SNR	Signal-to-noise ratio
TR	Tuning range
TTD	True time-delay
TX	Transmitter
VCO	Voltage-controlled oscillator

Table of Contents

Part I

Introduction

Chapter 1

Introduction and motivation

With 5G, we have seen the introduction of so-called mm-wave carrier frequencies in cellular communication. The term mm-wave refers to frequencies with wavelengths between 1 and 10 mm, which corresponds to a frequency range of 30-300 GHz[1]. While mm-waves have been used in radio circuits in the past, mainly for radar applications and point-to-point links, this is the first time it has been used for cellular communication. First out in mm-wave 5G was the so-called 28-GHz band (24.25-29.50 GHz), which has since been followed by several more bands, with the highest carrier frequency reaching 71 GHz [1]. With the introduction of 6G in a few years, the upper range of mm-wave frequencies (100-300 GHz), also referred to as sub-THz frequencies, is also expected to be used [2] [3] [4].

The reason for this increase in carrier frequency is quite simple – available bandwidth. Our constant need for higher cellular data rates has forced telecommunication companies to squeeze out every last drop of bandwidth from the regular sub-7 GHz frequencies, but it is still not enough for high-density urban areas [5]. There is simply not enough spectrum at these frequencies to satisfy the massive data rate and capacity demands. At mm-wave frequencies, on the other hand, large stretches of unoccupied spectrum are available, enabling very wide channel bandwidths, while simultaneously serving more users. For instance, the 28-GHz band allows for bandwidths of up to 400 MHz, while in a recent release, 3GPP Rel. 17 [1], a frequency band with channel bandwidths of up to 2 GHz was introduced, an unprecedented value in cellular communication. For sub-THz frequencies in 6G, channel bandwidths of 10 GHz or higher are expected [2] [4].

Given the availability of these massive channel bandwidths, why have we not utilized mm-waves before? It is because using these frequencies comes with several problems. The most apparent is the increased path loss, *PL*, which can be described by Friis' transmission formula [6]:

[1] For convenience, frequencies in the range 24-30 GHz are typically also considered mm-waves.

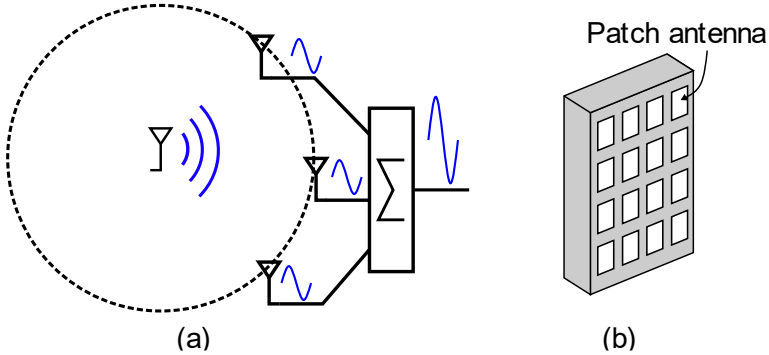


Figure 1: (a) Receiving antennas in a sphere around the transmitting antenna. (b) Planar antenna array.

$$PL = \left(\frac{4\pi D}{\lambda}\right)^2 = \left(\frac{4\pi D f}{c}\right)^2, \tag{1.1}$$

where λ and f are the wavelength and frequency of the carrier, respectively, D is the distance between transmitter (TX) and receiver (RX), and c is the speed of light. The path loss is the loss in signal power between the transmitter antenna and receiver antenna in free space and, as seen in Eq. (1.1), it increases quadratically with the carrier frequency. The reason for the increased path loss is reduced size of the receiver antenna aperture, i.e. the antenna area decreases with increasing frequency, thus capturing less of the radiated power. But if we use multiple receiver antennas, that independently receive the signal, and then combine their signals constructively, we can effectively increase the aperture size by a factor equal to the number of antennas. So, if we double the carrier frequency and at the same time increase from using one to four receive antennas, the aperture, and therefore the received power, will remain the same (for the same transmitted power density). However, the signals will only combine perfectly constructively if the antennas are located on a sphere around the transmitting antenna, since then they will all be at the same distance from the transmitter, causing the signals to reach each antenna at the same time, see Fig. 1a. This is of course impossible to implement practically when both the transmitter and receiver can move around. Instead, we place the antennas on a plane, in an arrangement referred to as an antenna array, see Fig. 1b. How can we then make all the signals from the antennas align so we can combine them constructively, when the propagation time from transmitter to receiver will be different for each antenna? We solve this by adding a variable time-delay, or more commonly a variable phase-shift, in the receiver path of each antenna, which allows us to align the signals from all antennas, no matter what direction the transmission is coming from. Antenna arrays can also be used in the transmitter, where the same technique of adding time-delays or phase-shifts to each antenna are applied, causing the signals to align in certain directions. The transmitted signal will then form a beam pattern, with the

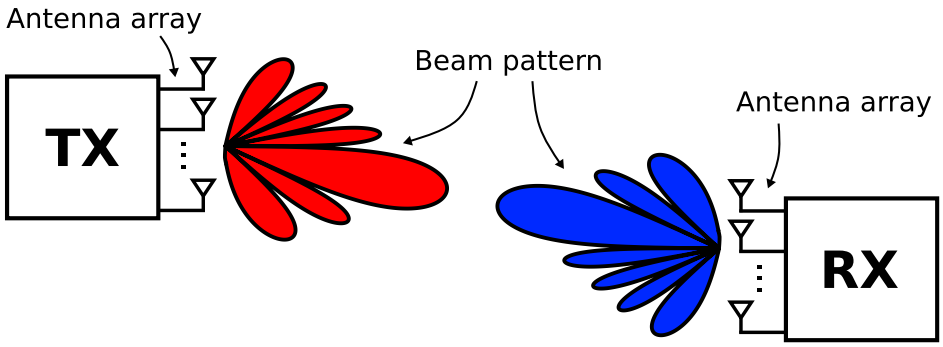


Figure 2: Beamforming in transmitter and receiver.

strongest beam pointing in a desired direction, typically towards the intended receiver, see Fig. 2; hence this is referred to as beamforming. The term beamforming is also used for the receiver, since it receives signals with higher gain in certain directions, forming receive beams, also shown in Fig. 2.

Traditionally, there has been three ways of separating users communicating with the same base station: in time, in frequency, and by using orthogonal codes [7]. However, the use of beams enables the separation of users also in space, since as long as the beams do not overlap, two or more users can communicate with the base station simultaneously using the same frequencies and codes. This is referred to as spatial multiplexing and greatly increases the capacity of the cell. This is one of the main reasons why beamforming is also used in 5G for sub-7 GHz frequencies.

The second issue with operating at mm-wave frequencies is that the performance of the radio circuits degrades. The high frequency means that parasitic capacitance must be kept to a minimum, resulting in transistors with minimum, or close to minimum, lengths. As any analog circuit designer can tell you, using short transistors results in poor g_m/I_D ratios, low output impedance, and high sensitivity to process variations, all resulting in worse performance than if long transistors are used. The parasitic capacitances making up a large portion of the total capacitance also makes implementing tunable circuits such as wideband local oscillators (LO) difficult, which is made worse by the poor quality factor of tuning capacitances at these frequencies [8] [9]. Furthermore, mm-wave frequencies are on the same order of magnitude as the transition frequency f_T of CMOS transistors, which typically worsens the performance of many radio frequency (RF) blocks. When it comes to receivers, operating closer to f_T limits the noise figure that can be reached, degrading receiver sensitivity [9] [10]. For transmitters, using thin oxide transistors with short channel length to provide sufficient gain at the high carrier frequencies, unfortunately limits the supply voltage that can be used and then also the output power that can be obtained [11]. Power amplifiers also suffer from significantly worse efficiency at these frequencies [11].

In addition to the difficultness of implementing wideband LOs mentioned above, there are several issues with the frequency generation related to its use in beamforming arrays. If there are multiple frequency up- or downconverters in the form of mixers in the beamforming array, which is typically the case, the LO signal must be distributed to each mixer. This will result in significant power losses, as routing losses are much higher at mm-wave compared to regular cellular frequencies [9], which must be compensated for by power-hungry buffers. The LO distribution must also not alter the relative phase between the LO signals to each frequency converters, as that would distort the beamforming. If *LO beamforming* is used (more on this in Chapter 2), this relative phase shift should be tunable in a known and very precise manner, requiring mm-wave phase shifters in the LO path. All of this should be achieved while keeping the phase noise level and power consumption low and the frequency tuning range wide, which is only possible with some innovative circuit designs. For instance, in Chapter 2 and 3, we will see how *injection-locking* can be used to both shift the phase and multiply the LO frequency. By using frequency multipliers, the LO frequency synthesizer can generate a lower frequency signal, which can be distributed with lower power. This can then be multiplied to the final LO frequency very close to the frequency converter, which not only reduces the routing losses, but also improves the phase noise.

It is not only the high carrier frequencies that are troublesome. The very large channel bandwidths can also be problematic, especially if multi-GHz bandwidths are used in future 6G implementations. In particular, the performance of the mixed-signal circuitry does not scale well for sampling frequencies in the GHz range and will thus likely become one of the most power-hungry parts in the signal chain. This puts stringent requirements on the analog baseband, to relax the requirements of the mixed-signal part as much as possible. For instance, as we will see in Chapter 4, by increasing the sharpness of the baseband filter, the sampling rate of the analog-to-digital converter (ADC) following it can be reduced, resulting in large power savings. However, implementing a high-order multi-GHz baseband filter with high dynamic range and low power consumption is not an easy task, as the performance of active filters deteriorates with increasing frequency. On the other hand, integrated passive filters get smaller, and the quality factor of integrated inductors increases with increasing cut-off frequency, making them a viable option for multi-GHz filters.

The aim of this thesis has been to develop circuits in CMOS for mm-wave beamforming 5G and 6G applications, focusing on frequency generation, system analysis, and baseband filtering. It is organized as follows:

Chapter 2 gives an overview of beamforming mm-wave transceivers, covering general theory and important aspects to consider when designing transceivers for large antenna arrays.

Chapter 3 discusses different ways to implement mm-wave frequency generation circuitry, explaining their strengths and weaknesses.

Chapter 4 discusses integrated active and passive filters for multi-GHz transceiver basebands, which are suitable for future 6G applications.

Chapter 5 summarizes and concludes the findings of the thesis and discusses potential future work to be done.

Paper I presents a phase shifter and frequency tripler for the 28-GHz band. A full sliding-IF receiver is also implemented to test the frequency generation scheme.

Paper II presents a modular testbench in MATLAB for simulating and analyzing beamforming receivers for sub-THz applications.

Paper III presents a differential-to-quadrature injection-locking frequency tripler for the 28-GHz band. Using feedback, the tripler automatically finds lock and corrects quadrature errors.

Paper IV presents active and passive multi-GHz baseband filters with state-of-the-art performance.

Paper V presents a capacitive cancellation technique to improve the stopband performance of the compact passive filter presented in Paper IV.

Chapter 2

Beamforming mm-Wave transceivers

The use of beamforming, both in the transmitter and receiver, greatly improves the achievable wireless communication distance. It also enables a fourth dimension of multiplexing – space (the three previous being time, frequency, and coding). However, the use of multiple antennas greatly increases the complexity of the system. This chapter covers the basics of beamforming, different transceiver architectures that can be used, circuit implementations, and non-idealities in the circuits and their impact on the system performance.

Beamforming theory

Imagine a linear array of *N* antennas with an antenna pitch *d*, see Fig. 3a. If we have a signal coming in from angle θ relative to the normal plane of the antennas, then the signal will first reach antenna 1 (assuming that θ is positive), then after a time-delay Δt it will reach antenna 2, and then after an additional time-delay Δt antenna 3, and so on. When the signal reaches the final antenna, it is delayed by $(N-1)\Delta t$ relative to the first antenna. Through basic trigonometry, and assuming that the signal is propagating with the speed of light *c*, it can be shown that:

$$\Delta t = \frac{d \cdot \sin\theta}{c} \tag{2.1}$$

Now, imagine that we can add an arbitrary time-delay after each antenna. If we then add a time-delay of $(N-1)\Delta t$ after the first antenna, a time-delay of $(N-2)\Delta t$ after the second antenna, and so on, the outputs after the delay cells will be perfectly synchronized. If we then combine all the delayed signals, they will add constructively, so the received total signal will have an amplitude *N* times higher than if only a single antenna was used, see Fig. 3b. This means that signal power increases by a factor N^2 (assuming an active combiner that combines the signals in voltage or current domain). Meanwhile, the noise from each antenna path will be

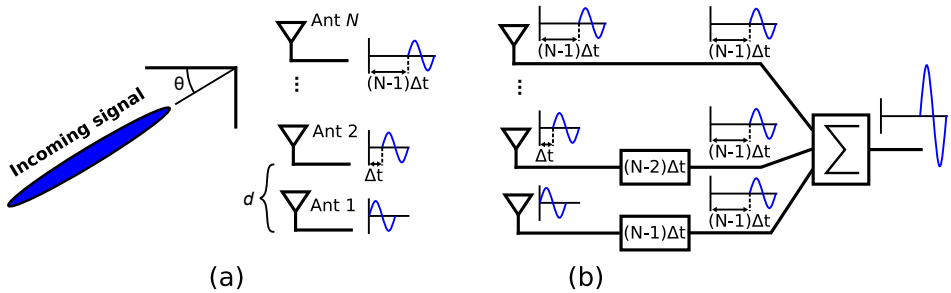


Figure 3: (a) Varying time-delays to the antennas. (b) Use of variable time-delays to make the signals add combine constructively.

uncorrelated, which means that the noise power increases by a factor *N* after the combination. So the total signal-to-noise ratio (SNR) increases by a factor *N*, which is the array gain.

We can use the exact same approach in transmitters to increase the transmitted power. If we apply the same time-delay as above in a transmitter, the signal from each antenna will add constructively in a beam along the angle θ (hence the term beamforming), increasing the transmitted power by N^2 in that direction, compared to transmitting the signal from a single antenna. It should be observed, however, that by using *N* antennas in the transmitter the total transmitted power has been increased *N* times, so that the power increase in the beam direction of N^2 can be seen as a factor *N* due to the increased total transmitted power multiplied by another factor *N* due to the array gain.

So, how do we implement the time-delay block required for beamforming? Usually, the answer is that we do not. While there are some examples in the literature [12] [13], the implementation of so-called true time-delay (TTD) cells that are tunable and accurate for mm-wave frequencies is a very difficult task. Instead, the more common approach is to use phase shifters to approximate the time-delay over the signal bandwidth. For a sine wave of frequency f_0, the phase shift $\Delta\phi$ in radians that correspond to a time-delay Δt is equal to:

$$\Delta\phi = 2\pi f_0 \Delta t, \tag{2.2}$$

which combined with Eq. (2.1) results in:

$$\Delta\phi = \frac{2\pi f_0 \cdot d \cdot \sin\theta}{c} \tag{2.3}$$

In fact, for a sine wave signal, this phase shift perfectly emulates a time-delay. However, for a modulated signal, the beamforming will only be perfect for the center frequency of the signal, while the beam will be distorted for the channel edges. This is known as *beam squint* and will be further discussed later in this chapter.

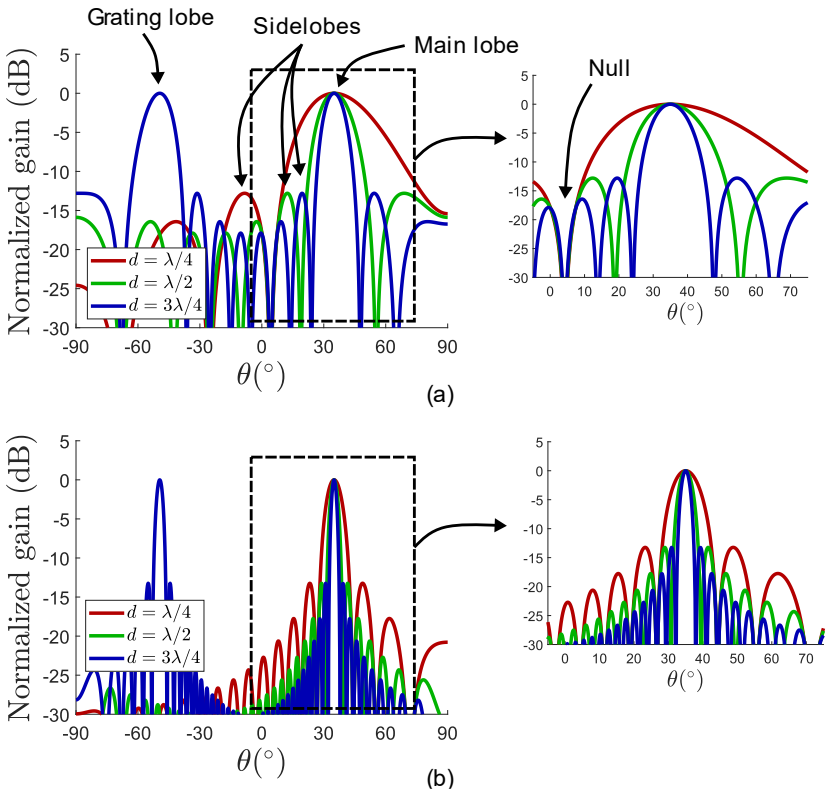


Figure 4: Normalized array factor for (a) 8 antennas and (b) 32 antennas, with the antenna pitch as a parameter.

So far, we have only considered the effect of beamforming in a particular desired direction. But what happens in the other directions? In general, a signal with frequency f arriving from angle θ with respect to the antenna array will be received with a complex gain r_n at antenna n:

$$r_n = a_n \exp\left(j(n-1)\frac{2\pi f \cdot d \cdot \sin\theta}{c}\right), \tag{2.4}$$

where a_n is the complex weight of antenna n, used to control the beam direction. The complex weight contains the phase shift as described, and it may also have different magnitudes to control the beam shape, also known as tapering. The total gain of the array after phase-shifting and signal combination becomes:

$$S(\theta) = \sum_{n=1}^{N} |a_n| \exp\left(j(n-1)\left[\frac{2\pi f \cdot d \cdot \sin\theta}{c} - \Delta\phi\right]\right) \tag{2.5}$$

By sweeping θ, we can now generate a plot of the so-called *array factor*, i.e. how much the antenna array amplifies or attenuates signals coming from different directions. This is done in Fig. 4 with the antenna pitch d as a parameter for two arrays with 8 and 32 elements, respectively. The desired beam direction is 35° and the magnitude $|a_n|$ is set to be the same (equal to unity) for all antennas. Several observations can be made from these plots. Firstly, the antenna pitch should be set to $\lambda/2$ to only have one main lobe. The undesired lobes formed when the pitch exceeds $\lambda/2$ is usually referred to as grating lobes. Secondly, even when using $d = \lambda/2$, lobes will be formed in other directions than the desired. These are referred to as sidelobes. It can be shown that without tapering the magnitude of the sidelobes will follow a sinc function and the first sidelobe will be 13.7 dB below the main lobe, independent of the number of antennas in the array [14]. Lastly, the more antennas there are, the narrower the beam will be.

That the first sidelobe is limited to 13.7 dB below the main lobe can in some scenarios be problematic, both in receivers and in transmitters. The sidelobe level can be lowered by using *tapering* in the array, i.e., the amplitude is lowered for the outer antenna elements according to a windowing function, at the cost of widening the main lobe. Fig. 5 plots the array gain for the same 8-element array as above, with and without a Dolph-Chebyshev window [15].

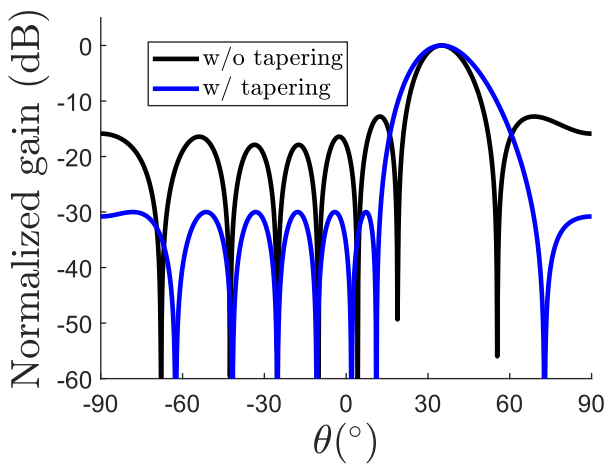


Figure 5: Tapering to reduce sidelobe levels.

In a cellular system, there will typically be multiple users that will cause interference to each other. When we generate the beam pattern, we want to place the main lobe towards the desired user, while placing nulls in the directions of other users that can be disturbed or cause interference, assuming that we can estimate where these other users are. This can be achieved by applying a method called zero-forcing [16], in which the complex weights **W** are given by pseudo-inverting the propagation channel:

$$\mathbf{W} = [\mathbf{w_1}, \dots, \mathbf{w_K}] = \mathbf{H}(\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}, \tag{2.6}$$

where $(\cdot)^{\mathrm{T}}$ is the matrix transpose and $\mathbf{H}$ is an $N \times K$ matrix, N being the number of receive antennas and K the number of incoming signals, both desired and interfering, and is given by:

$$\mathbf{H} = [\mathbf{h_1}, \dots, \mathbf{h_K}], \tag{2.7}$$

where $\mathbf{h_k}$ describes the propagation from signal k to the receive antenna. For a line-of-sight (LoS) scenario, which we are considering here, this will simply be:

$$\mathbf{h_k} = A_k\left[1 \; e^{j\pi \sin(\theta_k)} \; \dots \; e^{j\pi(N-1)\sin(\theta_k)}\right]^T, \tag{2.8}$$

where A_k is the magnitude and θ_k the direction of signal k, respectively. By applying weights $\mathbf{w_k}$ to the beamformer, the main lobe will be placed in direction θ_k, while nulls will be placed in the direction of all other signals present. Fig. 6 plots the array factor when zero-forcing is applied to an 8-element array, and a desired signal is coming from $\theta_1 = 35°$ and two interferers from $\theta_2 = -20°$ and $\theta_3 = 50°$, respectively.

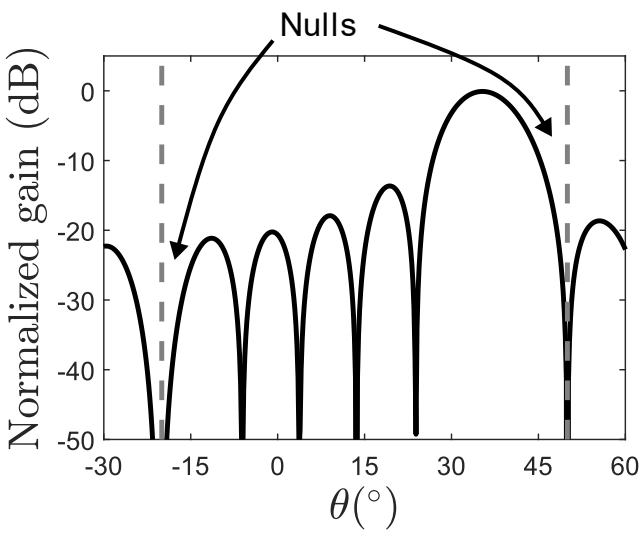


Figure 6: Null placement using zero-forcing.

Beamforming architectures

There are mainly three ways of implementing beamforming: fully in the digital domain, referred to as digital beamforming (DBF), fully in the analog domain, referred to as analog beamforming (ABF), and a mix of the two, referred to as hybrid

beamforming (HBF) [17]. There are advantages and disadvantages with every approach, and deciding which one will be the most efficient for a particular system is a very complex task, and highly dependent on many variables. This section will give an overview of these architectures. For convenience, it will focus on receivers, but most of the concepts covered are also applicable to transmitters.

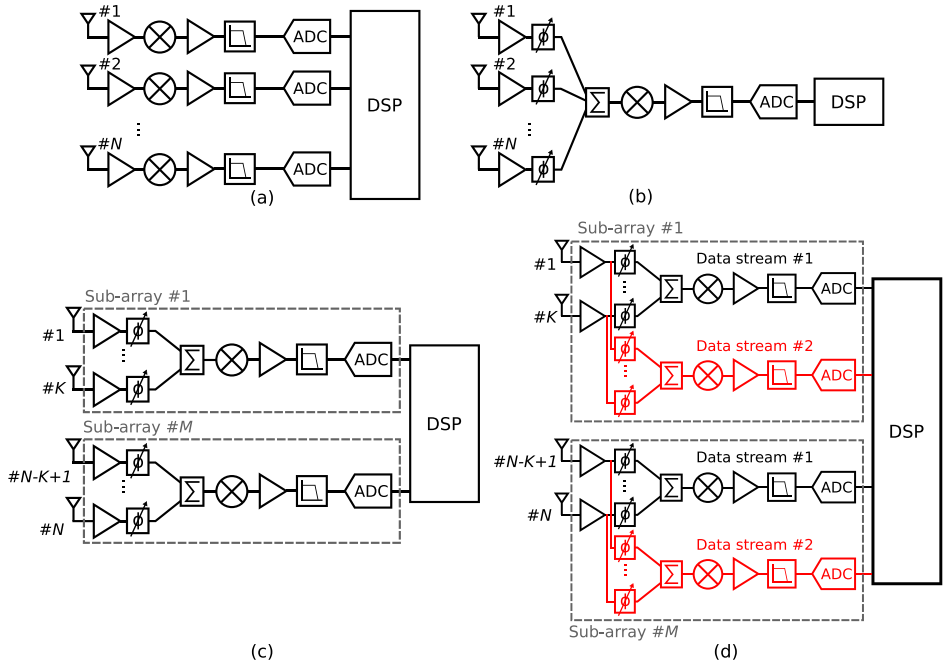


Figure 7: (a) Digital beamforming. (b) Analog beamforming. (c) Hybrid beamforming. (d) Hybrid beamforming with multiple data streams. Quadrature signals omitted for improved readability.

Digital beamforming

In digital beamforming, every antenna in the array is connected to a separate full receiver signal chain including ADC, see Fig. 7a. The application of beamforming weights and signal combination are performed entirely in the digital domain. There are several benefits of this approach. Firstly, as many data streams as there are antennas can be simultaneously processed in this way, resulting in the highest possible spectral efficiency. Effectively, this means that several beams can be used at the same time, receiving different signals from different directions. The shapes of these beams are also independent of each other and digitally programmable. Secondly, the beamforming accuracy is only limited by the digital baseband resolution. The beamforming can also be made wideband in the digital domain, accurately approximating also longer true time delays and thereby mitigating beam

squint. However, that the beamforming is done entirely in the digital domain greatly increases the power consumption of the digital signal processor (DSP) [18]. Also, the increased number of ADCs contribute to higher power consumption, combined with the fact that since the beam shapes are not formed until after being converted to the digital domain, every component in the analog receiver chain, including the ADCs, must be linear enough to handle in-band signals from all directions simultaneously. Lastly, another advantage of DBF is that it simplifies channel estimation since the entire channel is captured by the ADCs, while an analog or hybrid approach must scan the channel with the beam(s) [19].

Analog beamforming

In analog beamforming, the beamforming is performed entirely in the analog domain, i.e. all the phase shifters (or time-delay blocks) and signal combiners are implemented with analog blocks before the ADC, meaning that only one set of ADCs is required, no matter the number of antennas, see Fig. 7b. At a first glance, it may appear that this should lead to massive power savings for the mixed-signal circuitry compared to DBF. However, it is not quite that simple. The ADCs in ABF require a higher resolution than in DBF, since the SNR is higher, increasing the power consumption per ADC [20]. To what extent the ADC resolution can be reduced in DBF will be determined by the level of interference in the system, as no spatial filtering occurs prior to the ADC in DBF, thus requiring a higher ADC dynamic range than if only SNR was considered.

Depending on where in the RF chain the phase shift is applied, analog beamforming is divided into three sub-categories: *RF beamforming, intermediate frequency (IF)/baseband (BB) beamforming,* and *LO beamforming* [21]. These will be discussed in more detail later. Since the signals are combined in the analog domain, any interferers will, to some extent, be suppressed in comparison to the desired signal, easing the linearity requirements of any circuit blocks following the signal combination. The drawbacks are that only one data stream can be received at once and that the phase shifters will impact the signal quality in different ways, whether it be signal attenuation, noise, phase mismatch, nonlinearity, frequency response, etc.

Hybrid beamforming

Hybrid beamforming relies, as the name implies, both on analog and digital beamforming and is shown in Fig. 7c. The antenna array is divided into M sub-arrays with K inputs each, in which analog phase-shifting and signal combination is performed and the signal is then converted to the digital domain using ADCs. The outputs of each sub-array are subsequently combined using DBF.

The main advantage of HBF is that the signals are spatially filtered before the data converters, while still enabling multiple data streams to be processed. However, it should be noted that in most implementations of hybrid beamforming, each sub-array can only point its main lobe in one direction at the time, so not all antennas can always be utilized for all signals. For instance, if the array is split into 12 sub-arrays and we want to simultaneously receive signals from two users separated in direction by more than the beamwidth of the ABF sub-arrays, with approximately equal signal strength, six of the sub-arrays can be dedicated to one of the signals, and six to the other signal, reducing the array gain by 3 dB when compared to if all antennas are used for the signal. There are some published exceptions to this, where the hardware in each sub-array is multiplied so that multiple beams can be formed and therefore multiple data streams can be processed simultaneously, see Fig. 7d. In [22] and [23] dual data streams are supported, while in [24] as many as 16 data streams per sub-array are supported.

Also, HBF is less susceptible to beam squint than ABF in very large arrays, assuming the ABF is implemented with phase shifters and not TTD. The larger the array is, the longer the time delay should be for a given angle and, as will be explained later, beam squint gets worse the longer the time delay that is approximated with a phase shift. However, the digital beamforming part in HBF can approximate the time delay very well, so the beam squint will only affect the analog beamforming part of each sub-array. So, a linear 64-element HBF receiver with 4 sub-arrays will essentially experience the same beam squint as a 16-element ABF receiver, which is significantly less than a 64-element ABF receiver.

Architecture versus carrier frequency

The most important factor when it comes to selecting the beamforming architecture is the carrier frequency. At sub-7 GHz, the so-called 5G frequency range 1 (FR1) frequencies, digital beamforming is the *de facto* way of implementing beamforming. The rich channels between users and base station, i.e. there exists multiple paths of propagation, makes the spectral efficiency of DBF unparalleled. At the same time, the relatively low channel bandwidths make the power consumption of data converters and DSP manageable. On the other edge of the spectrum, we have the (anticipated) use of sub-THz carrier frequencies for 6G. Here, only LoS communication is expected, and the number of simultaneous users will be much lower than the number of antennas, meaning that the spectral efficiency will be similar for DBF and HBF. The large number of antennas required and the wide bandwidth anticipated would also make the power consumption of sub-THz 6G DBF equipment extremely high. To the author's knowledge, all published work on sub-THz antenna arrays in the literature relies on either analog or hybrid beamforming.

The obvious question is then, at what frequency do we switch from digital beamforming to analog/hybrid beamforming? This is a highly debated topic in the

literature, and a complete discussion on this is far beyond the scope of this thesis. While there are many system analyses advocating for DBF in the mm-wave 5G, or frequency range 2 (FR2), frequencies [18] [25], arguing that the power consumption will be on similar levels as an HBF architecture while offering much higher spectral efficiency, actual circuit implementations of full scale DBF are limited in the literature, with only a few exceptions [26] [27] [28].

Phase shifters

As was explained in the previous section, phase shifters are required in ABF and HBF. This section covers phaser shifter imperfections, the different analog beamforming architectures and their effect on the phase shifter, and some examples of phase shifter implementations.

Phase resolution and phase error

The phase shifter will typically be controlled by a digital signal, meaning that the phase will be shifted in discrete steps. This step size, i.e. the phase resolution, should be as small as possible to generate the phase values given by Eq. (2.3) as closely as possible. When discussing phase shifters in the literature, the phase resolution is usually given as either the average phase step size or as n bits, meaning an average step size of $360°/2^n$. The deviation from this average step size, the phase error, is also important, and is usually given as a root mean square (rms) value for all phase settings.

Interestingly, for steering the main lobe accurately, neither the phase resolution nor the phase error is very critical. By applying a non-uniform phase shift over the array, the effective spatial resolution can be significantly improved. In [29] for instance, they show that even with a phase resolution of 22.5°, the spatial resolution can be less than 1° for an 8x8 array. For the phase error, assuming that it is uncorrelated and randomly distributed across the phase settings, it will average out if enough antennas are used, barely affecting the direction of the main lobe [30]. It should be noted though, that both limited phase resolution and phase errors will reduce the gain of the main lobe, albeit only slightly. Both non-idealities will, however, have a large impact when either applying sidelobe suppression [29] or when trying to form nulls in particular directions. Since the latter relies on the antenna signals perfectly cancelling each other, any deviation from the ideal phase setting will cause imperfect cancellation and part of the interfering signal in the intended null direction will then leak through.

Amplitude control and amplitude/phase orthogonality

When implementing a phased array where we only are concerned about the gain of the main lobe, all antennas should have equal magnitude of their weights. However, when implementing either sidelobe suppression through tapering or when implementing nulls, each antenna path will require a separate amplitude control. As with all real-world circuit implementations, this amplitude control will never be perfect, exhibiting both limited resolution and random errors, which will cause non-ideal array factors. Where in the signal chain this variable gain is implemented will depend on which ABF architecture is chosen, since it must be done prior to the signal combination. An important aspect of amplitude control is that it should be orthogonal to the phase setting, i.e., a change in the phase should not affect the amplitude and vice versa. Otherwise, unless very extensive calibrations have been performed, it will make it much harder to sustain sidelobe suppression and nulls while sweeping the beam angle [29].

Analog beamforming architectures

As mentioned in the section on analog beamforming, there are three main ways to introduce the required phase shift: RF beamforming, IF/BB beamforming, and LO beamforming.

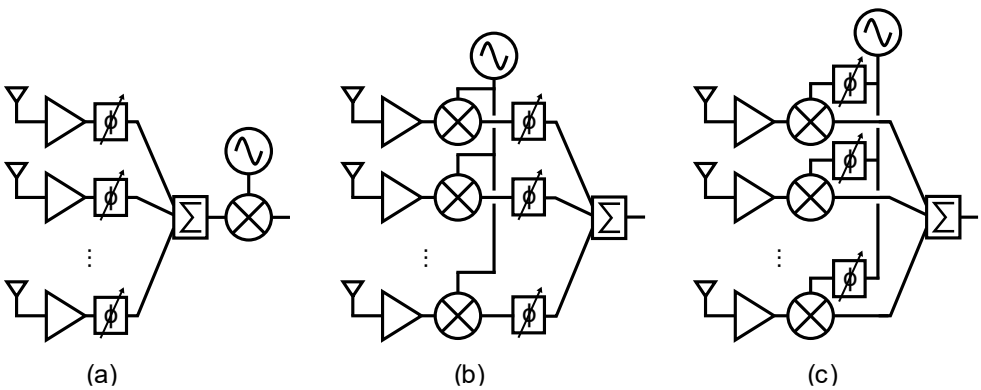


Figure 8: (a) RF beamforming. (b) IF/BB beamforming. (c) LO beamforming. Quadrature signals omitted for improved readability.

RF beamforming

In RF beamforming, the phase-shifting and signal combination are performed before the down-conversion mixer, see Fig. 8a. The major benefit of this approach is that only one set of mm-wave mixers is required. This greatly simplifies the LO distribution in the array since the mm-wave LO only has to be fed to one set of mixers. Additionally, interferers can be suppressed early on in the RF chain, reducing the linearity requirements of subsequent blocks.

The main disadvantage is that the phase shift is applied directly to the high frequency wideband signal. This can be done with either passive structures [29] [31], which typically attenuate the signal significantly requiring more gain stages in the signal path, or active components, which raise the noise and power consumption and cause non-linear distortion. The latter is typically done with *vector modulators* [32] [33] [34], which require quadrature phases of the signal, often generated by a passive quadrature hybrid. Due to the difficulty of implementing the phase shifters, the phase resolution is often limited in RF phase shifting. Other concerns are non-flat frequency response for amplitude and phase over the wide signal bandwidth, and the required chip area. To perform the signal combination, a Wilkinson combiner [35] is typically used, which thanks to its isolation resistors provides well matched input impedances also when the signals are misaligned in amplitude and phase.

IF/BB beamforming

In IF/BB phase shifting, the phase-shifting and signal combination are performed after the frequency down-conversion, requiring one down-conversion path per antenna, see Fig. 8b. The main benefit of this approach is the lower losses and higher accuracy when implementing the phase shifters and combiners at these frequencies, which is typically done with active components. The lower frequencies also means that the phase shifters will, assuming an inductorless design, occupy a very small area. An added benefit is that, at least for BB phase shifting, quadrature versions of the signals are available after the down-conversion, making it very easy to implement vector modulators [24] [36] [37] [38]. Just like DBF working on baseband signals, with analog IF/BB beamforming it is feasible to form multiple beams by using more than one set of active phase shifters per antenna branch. The lower effect of parasitics at baseband or IF frequencies, compared to RF, can allow multiple phase shifters to be used with high performance [24].

In addition to the large number of required mixers, IF/BB phase shifting, similar to RF phase shifting, relies on phase shifters in the signal path, meaning that amplitude errors in the phase shifters and non-flat frequency responses will directly affect the signal. Also, the phase shifters will add noise and distortion, reducing the dynamic range of the system.

LO beamforming

In LO phase shifting, it is not the signal itself that is phase shifted, but rather the LO signal of the mixer, see Fig. 8c. This phase shift is then passed to the signal when frequency converted by the mixer, since the mixer output signal phase is linearly dependent on the LO phase. This means that the signal combination is done after the mixer at either IF or BB frequencies, resulting in the same number of mixers as in IF/BB phase shifting. A benefit of this approach is that the conversion gain of the mixer is only weakly related to the amplitude of the LO signal, assuming that the LO amplitude is large enough. The decoupling of the phase-shifting from the signal

path also makes it easier to achieve high resolution phase shifters, since the phase shifters will not impact the noise and linearity. Furthermore, the LO signal is a single tone, so the phase shifter does not need to have a wideband flat frequency response, and it does not need to have a high linearity as there is no amplitude modulation of the signal. This significantly simplifies the design, enabling a more compact and low power phase shifter.

LO phase shifting is used in **Paper I** and **Paper II** and will therefore be the focus of the next section, which details implementations of LO phase shifters.

LO phase shifter implementations

A simple way of implementing an LO phase shifter is to use a ring oscillator and multiplexers. The inputs of each multiplexer are then connected to all output nodes of the ring oscillator, while the multiplexer outputs are each connected to a mixer LO input. The multiplexer can then be used to select which ring oscillator signals to use in the mixer, and a desired phase shift can be achieved. This was done in the first fully integrated silicon-based beamforming transceiver [21], where a ring oscillator built from eight differential amplifiers, yielding in total 16 different phases, was used. However, the phase resolution is limited by the number of stages in the oscillator, and ring oscillators typically have a worse phase noise performance than differential cross-coupled LC oscillators.

LO beamforming can, just like RF beamforming, also be done with passive components [39] and active vector modulators [40]. Another way of applying the phase shift, unique to LO phase shifting, is to do it directly in the phase-locked loop (PLL), by injecting an extra DC current into the charge pump [41] [42]. This makes the LO distribution very easy, since only a low-frequency reference signal needs to be distributed. It is also a very robust solution, since the phase shift will be proportional to the injected DC current, which in turn can be accurately controlled. However, it does mean that one PLL is required per antenna element, which will occupy a large area. Also, having multiple PLLs close to each other can cause pulling effects (more on this in Chapter 2).

One of the most common LO phase shift approaches is to use injection-locking [43] [44] [45] [46], which is when a signal is injected into an oscillator, forcing it to oscillate at the injected signal frequency f_{inj} instead of its own free-running oscillation frequency f_0. We refer to this as an injection-locked oscillator (ILO). Injection-locking will occur if the difference between f_{inj} and f_0 is smaller than the oscillator locking-range f_L [47]:

$$f_L = \frac{f_0}{2Q_{tank}} \frac{I_{inj}}{I_{osc}} \frac{1}{\sqrt{1 - \left(\frac{I_{inj}}{I_{osc}}\right)^2}}, \tag{2.9}$$

where Q_{tank} is the quality factor of the oscillator resonance tank, I_{inj} is the magnitude of the injected current, and I_{osc} is the magnitude of the oscillation current. Injection-locking will be discussed further in Chapter 3, but the most important property about injection-locking when it comes to implementing a phase shifter, is that the output signal of the oscillator will be phase shifted relative to the injected signal, given by [45]:

$$\Delta\phi = \sin^{-1}\left(\frac{f_0 - f_{inj}}{f_L}\right) \tag{2.10}$$

So, by tuning the free-running oscillation frequency, typically by using a varactor and/or switched capacitor cells, a desired phase shift can be achieved. However, the arcsin function is limited to an interval of $\pm 90°$. The arcsin function is also highly nonlinear close to the edges, making a wide range accurate phase shifting difficult to achieve. A common way to get around this problem is to have the ILO operating at lower frequency than the final LO frequency and feeding its output to a frequency multiplier, which multiplies the frequency by a factor M, see Fig. 9. A frequency multiplier can also be implemented using injection-locking [43] [46], referred to then as an injection-locked frequency multiplier (ILFM). The frequency multiplier does not only multiply the frequency of the input signal, but also the phase, as shown in Fig. 9. The term ϵ indicates any additional phase added by the multiplier itself. So, if for instance an injection-locked frequency tripler (ILFT) is used, the ILO only has to provide a phase shift of $\pm 60°$ to cover the whole 360° range, thus avoiding the most nonlinear regions of the arcsin function. Frequency multiplication is also beneficial from an LO distribution point-of-view, since it means that an LO signal with a lower frequency than the final LO frequency can be distributed, significantly reducing the losses. Frequency multiplication typically also results in better phase noise performance, than if the signal was directly generated at the final frequency, as will be discussed in Chapter 3.

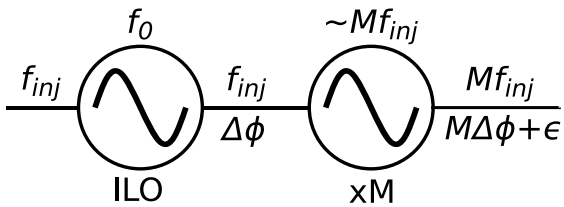


Figure 9: The use of frequency multipliers to extend the achievable phase tuning range.

One issue with using an ILO as phase shifter is its sensitivity to process, voltage, and temperature (PVT) variations, since all of these effects will change the resonance frequency of the tank, and thus the free-running frequency f_0, causing phase errors and potentially loss of injection-lock. This is especially problematic in conjunction with a frequency multiplier, given that the phase error will be multiplied by the multiplication factor. For this reason, some form of automatic calibration is required.

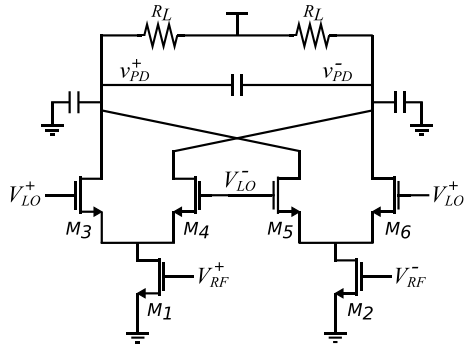


Figure 10: Schematic of an active mixer used as phase detector.

To perform an automatic calibration, the phase difference between each LO output must somehow be measured, which requires some form of phase detector (PD). A suitable implementation of a PD for mm-wave frequencies is an active mixer, see Fig. 10. If we multiply two signals of the same frequency, trigonometry tells us that the output becomes:

$$v_{PD} = A_1\cos(\omega t) \cdot A_2\cos(\omega t + \phi) = \frac{A_1 A_2}{2}(\cos\phi + \cos(2\omega t + \phi)) \tag{2.11}$$

By low-pass filtering the output, we get a DC output that is proportional to $\cos\phi$. In [43], they use this principle to automatically tune ILOs to achieve the desired phase shift. In total, they implement four LO phase shifters, LO_1-LO_4, with ILOs and ILFTs as in Fig. 9, and two PDs. Through multiplexers, each LO phase shifter can be connected to each PD input, see Fig. 11. They then apply a successive-approximation algorithm, which makes $(\phi_a - \phi_b) = (\phi_c - \phi_d)$ by tuning the ILOs so that $V_o = 0$ in Fig. 11. If $\phi_b = \phi_c$, i.e. V_b and V_c are connected to the same LO output, then $\phi_b = (\phi_a - \phi_d)/2$. So, if LO_1 and LO_2 are 90° apart, they can automatically tune LO_3 so that it is phase shifted 45° relative to LO_1, and then tune LO_4 so that it is phase shifted 22.5° relative to LO_1. However, while they do achieve very low phase error (0.93° rms), they only achieve a resolution of 22.5°. Additionally, it is not clear what happens if lock is lost during the automatic tuning, nor how the frequency tripler is tuned.

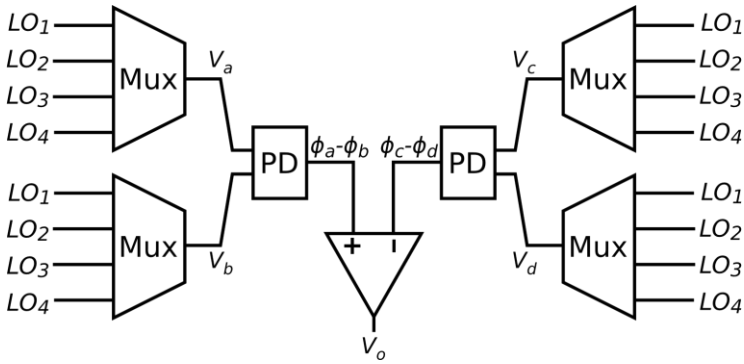


Figure 11: Automatic phase calibration scheme used in [43].

In **Paper I**, we use a similar approach, where an LO phase shifter is implemented using an ILO and an ILFT. But instead of comparing the phase between multiple ILFT outputs, the phase of the frequency tripler output, V_{FT}, is measured relative to the phase of the initial injected signal, V_{inj}, see Fig. 12a. This is possible thanks to harmonic mixing in the active mixer. V_{inj} is applied to the commutating pairs M_3-M_6 in Fig. 10, while the tripler output, V_{FT}, is applied to M_1 and M_2. If the signal to M_3-M_6 is large enough, the current steering will be that of a square wave, with significant third harmonic content. This third harmonic will mix with V_{FT}, generating a DC output proportional to $\cos\phi$. However, since we do not know the amplitude in Eq. (2.11), we cannot tell the absolute phase. To solve this, and removing the uncertainty of the sign of the phase shift due to $\cos\phi = \cos(-\phi)$, we can use a quadrature version of V_{inj} and feed it to a separate PD, see Fig. 12b. This yields:

$$v_{PD,Q} = A_1\cos(\omega t)\cdot A_2\sin(\omega t + \phi) = \frac{A_1A_2}{2}(\sin\phi + \sin(2\omega t + \phi)) \tag{2.12}$$

We can now calculate the phase shift as:

$$f(x) = \begin{cases} \tan^{-1}\dfrac{v_{PD,I}}{v_{PD,Q}}, & v_{PD,I} \geq 0 \\ \tan^{-1}\dfrac{v_{PD,I}}{v_{PD.Q}} + 180^{\circ}, & v_{PD,I} < 0 \end{cases} \tag{2.13}$$

The two filtered PD outputs can be digitized by an ADC and be used to calculate the obtained phase shift and tune the free-running frequency of both the ILO and the frequency multiplier to achieve the desired phase shift. Chapter 3 will explain how the phase detector also can be used to automatically find injection-lock.

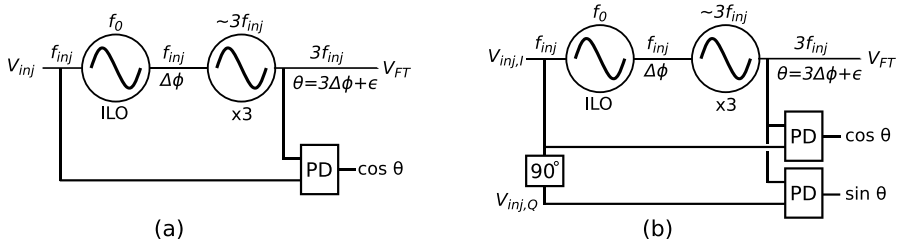


Figure 12: (a) Measuring the applied phase shift from the ILO and frequency multiplier. (b) Quadrature phase detection.

While the above technique is designed for a sliding-IF transceiver, it can also work for a direct-conversion transceiver. The only difference is that we then require a quadrature version of V_{FT} instead of V_{inj}, which can then be fed to two separate PDs. How quadrature signals can be generated at mm-wave frequencies will be discussed in Chapter 3.

Additional antenna array system considerations

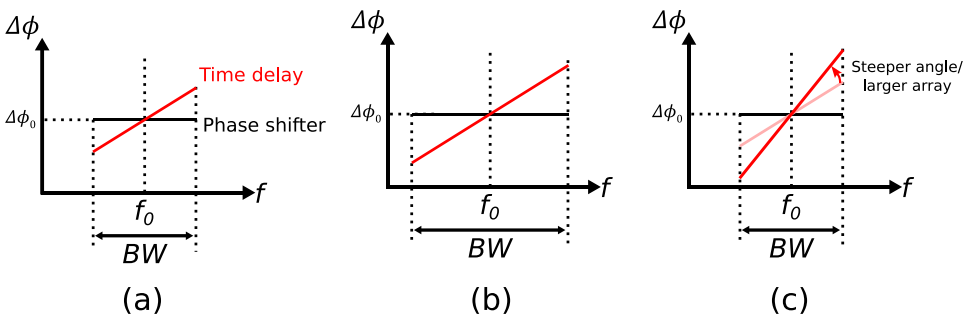


Figure 13: (a) Difference between a phase shifter and a TTD. (b) Impact of increasing the fractional bandwidth. (c) Impact of increasing the angle of arrival and/or array size.

Beam squint

The approximation of time-delays using phase shifters will cause the array factor to be distorted for wideband signals, as has been briefly mentioned earlier. This is because the applied phase shift from Eq. (2.5) is constant for all frequencies, while a time-delay is equivalent to a phase shift that varies linearly with the frequency, see Fig. 13a. The larger the fractional bandwidth, the worse the phase-shift approximation will be, as seen in Fig. 13b. The approximation will also be worse with steeper incident angles and larger arrays, as the time delays between the antenna signals will then increase. This corresponds to a steeper phase slope in Fig. 13c. Using Eq. (2.5), we can plot the array factor for $f = f_0$, $f = 1.1f_0$, and $f =$

$0.9f_0$, i.e. the center frequency and the two edge frequencies for signal with 20% fractional bandwidth, for two different arrays, one with 8 elements and one with 32 elements, and two different incident angles, 10° and 40°, see Fig. 14. It can be observed that the main lobes of the edge frequencies point in wrong directions in all four scenarios, but it is worse for large arrays and steep angles, as was predicted. This phenomenon is known as beam squint.

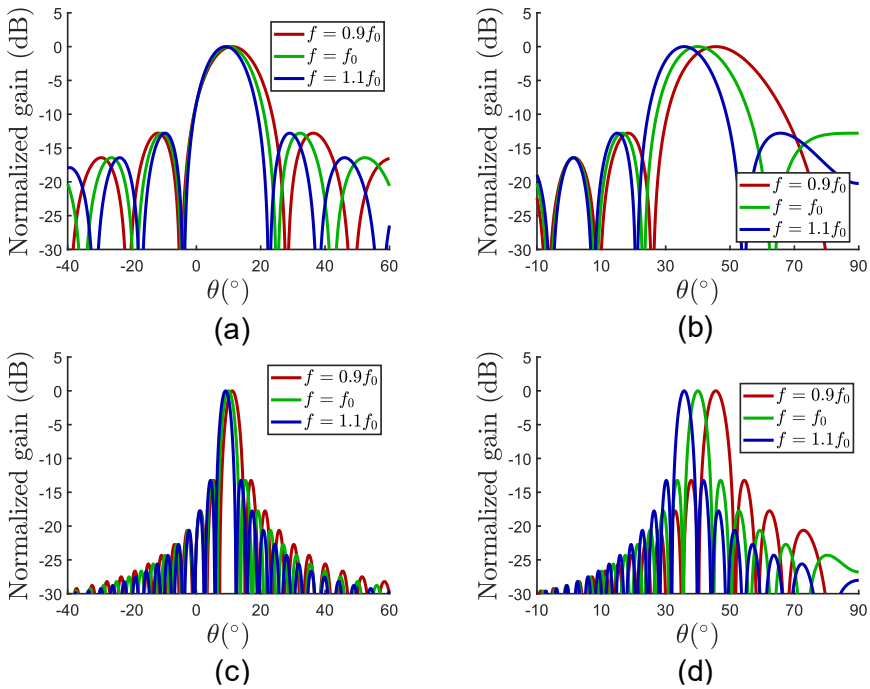


Figure 14: Effect of beam squint on different configurations. (a) 8 antennas, $\theta_0 = 10^\circ$. (b) 8 antennas, $\theta_0 = 40^\circ$. (c) 32 antennas, $\theta_0 = 10^\circ$. (d) 32 antennas, $\theta_0 = 40^\circ$.

Another way to visualize the effect of beam squint is to plot the normalized array factor in the desired direction versus the relative frequency, as has been done in Fig. 15 for the four scenarios above. As can be seen, the gain is not flat across the bandwidth, especially for the large array with large incident angle, degrading the SNR. While the gain flatness can be corrected for in the digital domain, this will not improve the SNR, only reduce the distortion.

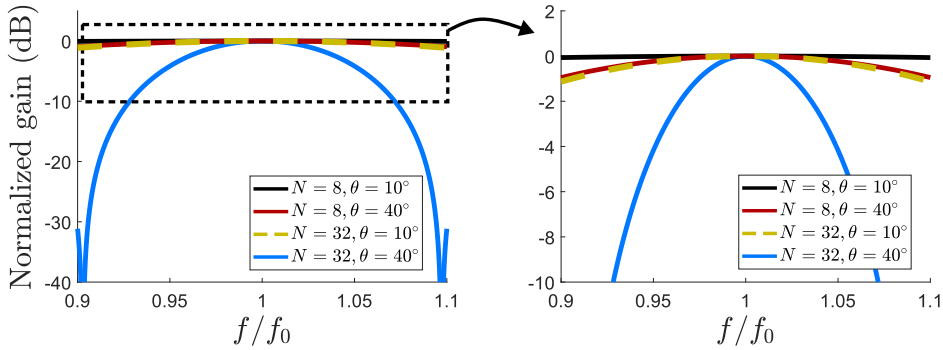


Figure 15: Array gain versus frequency.

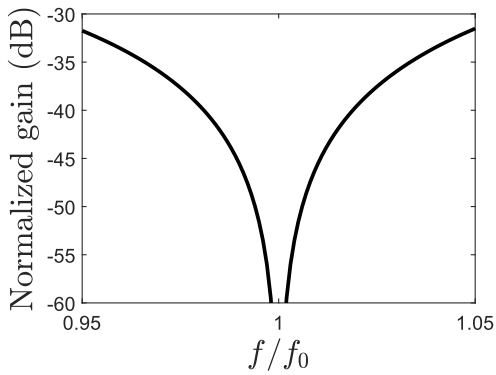


Figure 16: Array gain versus frequency for $\theta = -20°$.

Beam squint will also cause issues with the null-forming. This is illustrated in Fig. 16, where for an 8-element array, the array factor in the direction $\theta_{null} = -20°$ is plotted versus relative frequency when a null is placed at θ_{null}. Clearly, it is only the center of the frequency band that gets perfectly nulled, meaning that if an interferer is wideband, a significant part of its spectrum will not get well cancelled.

The fractional bandwidths in 5G are still relatively low, with the recent n263 band having the highest yet of 3.5% [1], meaning that beam squint has not been a significant issue. However, this number will increase with the introduction of sub-THz carrier frequencies in 6G. With the carrier frequencies and channel bandwidths currently being discussed, fractional bandwidths on the order of 10% are expected. At the same time, the higher carrier frequencies require larger antenna arrays to overcome the increased path loss, further worsening the beam squint problem.

The easiest way to combat the beam squint is to use hybrid beamforming with a relatively low number of antennas per sub-array. This ensures that the effect of beam squint is limited for each sub-array, while the appropriate time-delays can be implemented in the digital domain when combining the outputs from all sub-arrays.

Chip size versus antenna pitch

We know from earlier that the antenna pitch should be $\lambda/2$ to avoid gating lobes. For a 28-GHz signal, this corresponds to about 5 mm. If we instead move to 140 GHz, as will likely be a carrier frequency in 6G, this pitch is only 1 mm. This raises an interesting issue. If the dimensions per antenna element of the beamforming IC is larger than the antenna pitch in either direction, the length of the connection between the IC and each antenna cannot be uniform, resulting in uneven phase and amplitude shifts. Alternatively, the antenna pitch can be increased, with undesired grating lobes as a result. Using dummy antennas to increase the effective area of the antenna array also comes at the cost of grating lobes [48]. The main takeaway from this is that the area of a sub-THz transceiver should be made as small as possible, favoring direct-conversion architectures. Especially when transceivers for both horizontal and vertical polarization, i.e. two full transceivers, should fit under each antenna [29], the chip area limitation gets severe.

Phase noise

Phase noise is an interesting topic for beamformers, because, in addition to the usual effects it has in a transceiver (signal distortion due to symbol rotation, reciprocal mixing of interference, and loss of subcarrier orthogonality in an OFDM system), phase noise will affect the array factor [49]. If the LO signal is applied before the signal combination, then the phase noise will cause a different phase shift to each down-converting chain, causing non-ideal combination and cancellation. Similar to the phase shifter error, this will have much more impact on the nulls than the beams. It should be noted that this is not only problem in hybrid and analog beamforming but will have the same effect in digital beamforming. The only beamforming scheme that is insensitive to phase noise is RF beamforming, since the beamforming then is done in the RF domain prior to the up-/down-conversion.

However, if the phase error due to phase noise is the same for all down-converters, i.e. the phase noise is completely correlated, the relative phase between each path will be constant and the array factor will be unaffected [49] [50]. On the other hand, if the phase noise is completely correlated, no averaging of the phase noise will occur, demanding a high-performance LO to not distort the signal due to the phase noise effects not related to the array factor.

To achieve perfectly correlated phase noise, a single oscillator would have to be used for all mixers. This is not power efficient for a large array, since distributing a mm-wave signal over such a long distance would cause large losses. Instead, a low-frequency reference signal is typically distributed to multiple PLLs, which each then drive a smaller number of mixers in a sub-array. The phase noise in these sub-arrays will then be correlated, except for any buffers that will add uncorrelated noise. Using ILOs and ILFMs will also cause the phase noise to be correlated.

Fig. 17 shows the simulated SNR of a 48-element receiver, divided into 12 sub-arrays, versus the integrated phase noise jitter. The incoming signal has a carrier frequency of 100 GHz and a bandwidth of 10 GHz. Four scenarios are simulated: uncorrelated phase noise without interference, uncorrelated phase noise with interference, correlated phase noise within the sub-array without interference, and correlated phase noise within the sub-array with interference. The interference is 20 dB stronger than the desired signal and a null is placed in its direction. As seen in the plot, the best performance is achieved for uncorrelated phase noise without interference for all simulated phase noise levels. However, the correlated case performs better when the interference is present when the phase jitter exceeds 100 fs, due to the distortion of the array factor.

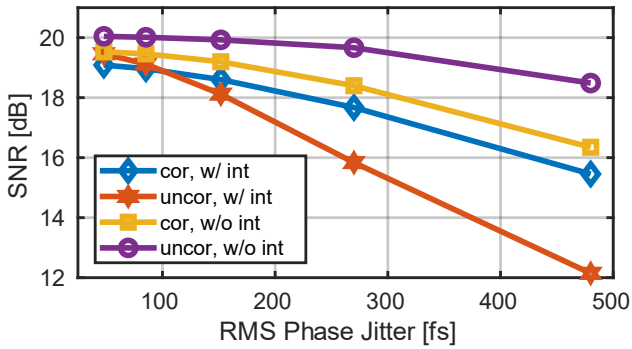


Figure 17: SNR versus phase noise jitter in a 48-element array.

Modular antenna array receiver testbench

As should be clear by now, the design of a beamforming antenna array transceiver is a highly complex task, with every architectural choice impacting the requirements of the different circuit blocks. Exactly what the requirements of these circuit blocks are and what kind of system-level impact circuit non-idealities will have is difficult to calculate, requiring system-level simulations where these non-idealities are properly modelled. While plenty of system-level simulations and analyses exist in the literature, they typically use an abstracted view of the analog circuitry. This raises three problems. Firstly, there is a risk that the model does not catch the impact of all important non-idealities, resulting in too optimistic results. Secondly, from a circuit designer's perspective, the abstracted view of the circuitry provides little insight in how to design the transceiver. Thirdly, the use of abstract models makes the system models less user-friendly, requiring vast prior knowledge to know how to adjust them to fit a specific use case. For these reasons, we developed a testbench in MATLAB and Simulink intended for large beamforming array receivers, which is presented in **Paper II**. The analog hardware is modelled in the *RF Blockset* [51]

environment in Simulink, which provides both a simulation engine for RF simulations and premade models of common RF blocks, such as amplifiers, mixers, filters, and phase shifters. For each block several parameters can be defined, such as noise figure and linearity, and in the case of a local oscillator, a phase noise profile. This enables system simulations with high accuracy, while Simulink's graphical user interface makes it user-friendly.

The testbench is based on a MATLAB wrapper that generates the input signal, containing both the desired signal and any interfering signals, and sets up all the circuit block parameters. The input signal is then passed to multiple receiver tiles, where each tile is a full ABF receiver implemented in Simulink using *RF Blockset*. The outputs from all the tiles are then combined in MATLAB, forming a large HBF receiver. This makes the testbench modular, as multiple Simulink files of different receiver architectures can be created and then chosen by simply changing a variable in the MATLAB wrapper. The number of tiles is also controlled by a variable in the wrapper. Each tile is simulated on a separate processor core, greatly reducing the simulation time.

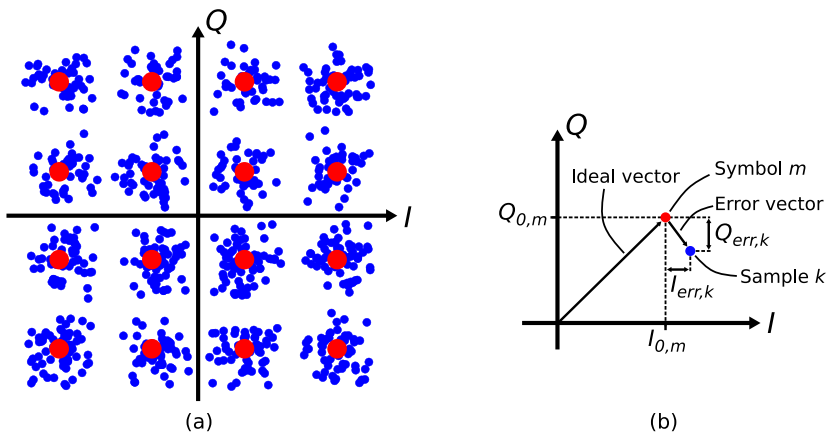


Figure 18: (a) Constellation diagram for 16-QAM, with ideal samples (red dots) and non-ideal samples (blue dots). (b) Definition of EVM.

To evaluate the performance of the receiver in the testbench, the *constellation diagram*, showing where in the IQ plane the sampled baseband symbols end up, is analyzed. Fig. 18a shows an example of the ideal constellation diagram for a 16-QAM modulation, along with the constellation diagram from a non-ideal receiver. The closer the samples are to the ideal points, the better the performance, and the lower the probability for demodulation errors. To quantify the closeness to the ideal points we use the error vector magnitude (EVM). It is defined as the rms length of the error vector, i.e. the distance between the sampled point and the ideal point in the constellation diagram, see Fig. 18b, normalized to the rms distance from the origin to the ideal points, and is typically given in percentage or decibels. For an *M*-QAM modulation, this becomes [52]:

$$EVM = \frac{\sqrt{\frac{1}{K}\sum_{k=0}^{K-1}\left(I_{err,k}^2 + Q_{err,k}^2\right)}}{\sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(I_{0,m}^2 + Q_{0,m}^2\right)}}, \tag{2.14}$$

where K is the number of samples, $I_{err,k}$ and $Q_{err,k}$ are the in-phase and quadrature error vector components of sample k, respectively, and $I_{0,m}$ and $Q_{0,m}$ are the ideal in-phase and quadrature vector components for the ideal symbol m, see Fig. 18b. The benefit of using EVM is that it captures the effect of all non-idealities, whether it be noise, nonlinearity, quadrature errors (more on this in Chapter 4), beam squint, phase noise, etc. Using some statistical analysis, EVM can be used to predict the bit-error rate (BER) of the system, i.e. how likely it is for the receiver to wrongly interpret a transmitted '0' as a '1', or vice versa. Table 1 presents the required EVM for a given BER for different modulation schemes [52]. Note that these requirements are for the total signal chain, i.e. the combined EVM from both the transmitter and receiver must be considered.

Table 1: Required EVM in dB for a given modulation and BER.

BER	4-QAM	16-QAM	64-QAM
10^{-3}	-9.8	-16.5	-22.5
10^{-4}	-11.4	-18.2	-24.3
10^{-5}	-12.6	-19.5	-26.6

Chapter 3

mm-Wave Frequency Generation

To perform the frequency translation that the vast majority of all RF transceivers rely on, an accurate and stable local oscillator signal is required. For a modern, sub-7 GHz, single-antenna, direct-conversion transceiver, the most common approach is to have a low-frequency reference clock, generated by an off-chip crystal oscillator, which is multiplied to the desired carrier frequency through a PLL. To obtain sufficient frequency resolution, multiplication by non-integer numbers can be performed using a so-called fractional-N PLL.

In a mm-wave, multi-antenna transceiver, there are significantly more aspects to consider; we must consider long interconnects for the clock distribution, correlated and uncorrelated phase noise, disturbances between different paths, and phase errors and drift between paths. Additionally, as will be shown in this chapter, the phase noise performance of mm-wave oscillators is typically worse than of their sub-7 GHz counterparts. This chapter covers different implementations for mm-wave frequency generation, discussing benefits and drawbacks of each approach.

Voltage-controlled oscillators

Most of this chapter will, in one way or another, be related to the most ubiquitous CMOS oscillator type for high-frequency applications, the cross-coupled LC oscillator, see Fig. 19a. This section will cover the basics of this circuit and bring up some issues and design considerations when operating at mm-wave frequencies.

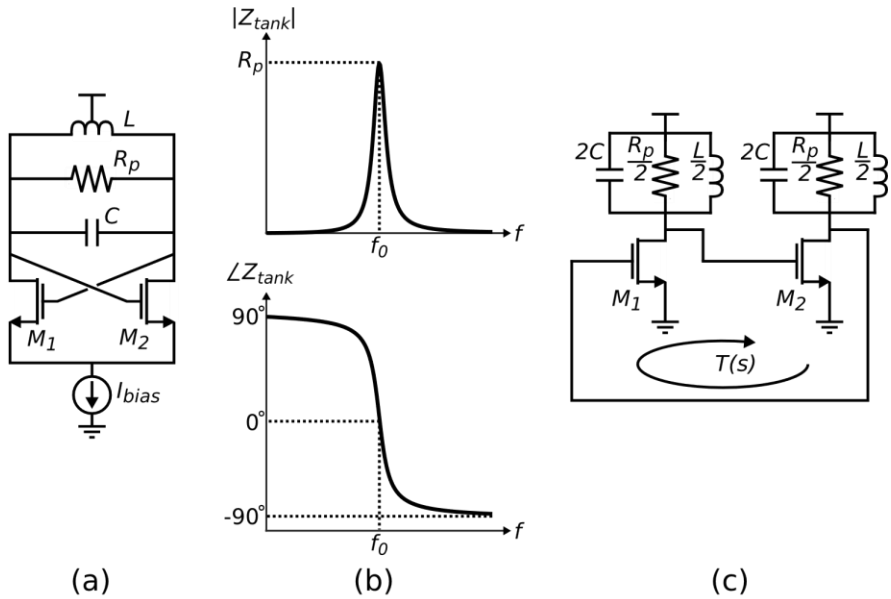


Figure 19: (a) The cross-coupled LC oscillator. (b) Impedance of the resonance tank. (c) Feedback view of the oscillator.

Basic theory

To understand the cross-coupled LC oscillator, we must first discuss the resonance tank, which, as seen in Fig. 19a, comprises an inductor L, a capacitor C, and a resistor R_P. R_P is not an explicit resistor but is used to model the losses of the tank due to the inductor and the capacitor not being an ideal inductance and capacitance, respectively. The tank impedance Z_{tank} will be given by:

$$Z_{tank}(s) = R_p||sL||\frac{1}{sC} \tag{3.1}$$

The magnitude and phase of Z_{tank} is plotted in Fig. 19b. At the so-called resonance frequency:

$$f_0 = \frac{\omega_0}{2\pi} = \frac{1}{2\pi\sqrt{LC}}, \tag{3.2}$$

the reactance of the inductor and of the capacitor perfectly cancel each other, resulting in $Z_{tank}(j2\pi f_0) = R_p$, which can clearly be seen in Fig. 19b. We can calculate R_P as:

$$R_p = Q_{tank}\omega L, \tag{3.3}$$

where Q_{tank} is the quality factor of the resonance tank and is given by:

$$Q_{tank} = \left(\frac{1}{Q_C} + \frac{1}{Q_L}\right)^{-1}, \tag{3.4}$$

where in turn Q_C and Q_L are the quality factors of the capacitive and inductive parts of the resonance tank, respectively. If the losses in these parts are modelled with a series resistance R_S, Q_C and Q_L can be calculated as:

$$Q_C = \frac{1}{R_S \omega C} \tag{3.5}$$

$$Q_L = \frac{\omega L}{R_S} \tag{3.6}$$

For oscillators in the single-digit GHz range, we typically have $Q_C \gg Q_L$, resulting in $Q_{tank} \approx Q_L$. However, as will be shown later, this is not always the case for mm-wave oscillators, especially in wideband operation.

The LC oscillator can be seen as feedback system, which is made clear by redrawing it as shown in Fig. 19c. Barkhausen's stability criterion tells us that for oscillation to occur, the loop gain $T(s)$ should have a phase shift of 360° and a magnitude equal to unity [53]. It can be seen in Fig. 19c that each stage will contribute $-g_m Z_{tank}/2$ to the loop gain, where g_m is the transconductance of M_1 and M_2, which should provide a phase shift of 180° for a total phase shift of 360°. This means that Z_{tank} must be real, which we have seen occurs at $\omega = \omega_0$. For the loop gain magnitude, we have:

$$|T(j\omega_0)| = \left(\frac{g_m R_p}{2}\right)^2 = 1 \tag{3.7}$$

This results in $g_m = 2/R_P$ to sustain oscillation. Usually, g_m is made significantly larger than $2/R_P$ to guarantee oscillation start-up across process and temperature variations.

It should be noted that for this analysis, we have assumed linear, small-signal operation. However, oscillators typically have large voltage swings and will thus behave in a non-linear way, with a transconductance that varies over the oscillation period. This is especially important for the analysis of phase noise. The small-signal analysis used so far predicts an infinitely growing amplitude if $g_m > 2/R_P$, which is of course unrealistic. In a real oscillator, the amplitude will grow until the effective transconductance $G_{\mathrm{m,eff}}$ is compressed due to non-linearities, reaching a point where $G_{\mathrm{m,eff}} R_p/2 = 1$, stabilizing the output amplitude. The final oscillation voltage will be given by:

$$V_{osc} = I_{osc} Z_{tank}, \tag{3.8}$$

where I_{osc} is the large-signal oscillation current. For a well-designed oscillator, the cross-coupled transistors should completely steer the bias current from one side to the other, resulting in I_{osc} being close to a square wave alternating between I_{bias} and 0 at each side of the resonance tank. The fundamental tone part of this square wave will have an amplitude of $(2/\pi)I_{bias}$ which, assuming the tank filters out all higher harmonics, yields a differential amplitude of [54]:

$$V_{osc} = \frac{2}{\pi} I_{bias} R_p \tag{3.9}$$

Phase noise

An oscillator should ideally generate a tone with perfectly stable amplitude and frequency, but due to noise in its components the oscillator frequency will vary over time. The output of an oscillator can be written as:

$$V(t) = A\sin(\omega_0 t + \phi(t)), \tag{3.10}$$

where A is the output amplitude, ω_0 is the noiseless oscillation angular frequency, t is the time, and $\phi(t)$ the so-called phase noise. The oscillator also has amplitude noise, so that also A fluctuates with time, but this is typically neglected due to the limiting properties of mixers making them less sensitive to amplitude noise. The instantaneous angular frequency of the oscillator will be given by:

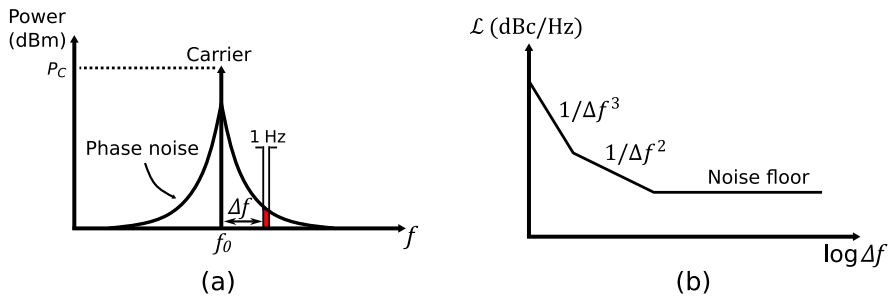


Figure 20: (a) Frequency spectrum of a noisy oscillator. (b) The different phase noise regions.

$$\omega_{inst} = \omega_0 + \frac{\partial \phi}{\partial t}, \tag{3.11}$$

So, the derivative of the phase noise will cause the oscillation frequency to deviate from its desired value, spreading out the spectral density of the signal into a skirt-shaped frequency profile around ω_0, see Fig. 20a. To quantify the noise, it is typically given as the noise power in 1 Hz bandwidth at a given offset Δf from f_0, also shown in Fig. 20a, normalized to the carrier power P_C. It is then typically

converted to decibels, resulting in the unit dBc/Hz, where dBc is decibels below the carrier.

Using the linear feedback model described earlier, Leeson derived the phase noise due to the tank losses [55]. However, he was unable to derive the noise injected by the active devices, only accounting for them with empirical fitting parameters. This resulted in the now very famous equation:

$$\mathcal{L}(\Delta f) = 10\log\left[\frac{kTFR_p}{A^2}\left(1+\left(\frac{f_0}{2Q_{tank}\Delta f}\right)^2\right)\left(1+\frac{f_c}{\Delta f}\right)\right], \qquad (3.12)$$

where k is Boltzmann's constant, T is the temperature in Kelvin, A is the oscillation amplitude, and F and f_c are fitting parameters called the oscillator's noise figure and flicker noise corner frequency, respectively. This equation predicts three distinct regions, plotted in Fig. 20b: one with a slope of $1/\Delta f^3$, called the flicker noise region, one with a slope of $1/\Delta f^2$, called the thermal noise region, and one region that is flat, referred to as the phase noise floor. For traditional oscillator design, the two former regions are the most important ones. However, for very wideband systems, as will likely be used in part of 6G, the third region can be of great importance, as the total noise stemming from this region will reach similar levels as the total noise from the two other regions.

Since Leeson's analysis, there have been a plethora of analyses that predict the phase noise in an oscillator without relying on the empiric fitting parameters, with [56] [57] [58] [59] being some of the most impactful examples. They are all based on treating the oscillator as a time-varying circuit, while Leeson relied on a time-invariant circuit. Still, the dependence on A and Q_{tank} in Leeson's equation remain true. So Q_{tank} and the amplitude should be maximized to achieve good phase noise performance.

Frequency tuning

For any practical application, the oscillation frequency must, to some degree, be tunable with a control signal. It is very common to use a voltage as the control signal, then the oscillator is called a voltage-controlled oscillator (VCO). To control the oscillation frequency, we need to adjust either the tank capacitance or tank inductance, or a combination of both. While there exist examples of inductance tuning, the former is by far the most popular approach. Capacitive tuning can broadly be divided into two categories: discrete tuning, in which capacitances are connected to the resonator through switches, allowing them to be turned on or off, and continuous tuning, in which the capacitance can be continuously tuned with an analog voltage using so called varactors, whose implementations will later be shown.

Wideband oscillators should preferably not be exclusively tuned by continuous tuning, since this would require a very large varactor. This is undesirable for several

reasons. Firstly, a large varactor results in a high control voltage-to-frequency gain, K_{VCO}. This means that noise and interference on the control voltage will strongly modulate the oscillation frequency, resulting in increased phase noise and spurs. Furthermore, the high K_{VCO} causes significant amplitude modulation (AM)-to-phase modulation (PM) within the oscillator, which causes amplitude noise from the current source to be converted into phase noise, thus increasing the total phase noise [60]. Lastly, varactors have a relatively poor C_{max}-to-C_{min} ratio [61], resulting in limited tuning range, as will be discussed below. Instead, a combination of discrete and continuous tuning is usually employed, where the discrete tuning is responsible for the coarse tuning, while the continuous part takes care of the fine frequency tuning. The continuous tuning should be large enough to cover (with some margin) the smallest frequency step of the discrete tuning, in order to avoid frequency gaps. There are also examples of oscillators that only use discrete tuning, so called digitally controlled oscillators (DCO).

Designing a mm-wave oscillator with a wide tuning range is troublesome for two main reasons. Firstly, the high frequency requires use of small inductors, on the order of 25-250 pH, to resonate with the parasitic capacitances, coming from the inductors themselves, the cross-coupled transistors, the interconnect, tuning circuitry, and output buffers. Adding more tuning circuitry for wider bandwidth results in more parasitic capacitance, and the inductors may then get so small that parasitic inductances in the interconnects can be comparable in size, reducing the quality factor and creating additional resonances affecting the circuit behavior. Additionally, for very low inductances, below about 100 pH at 30 GHz, $Q_L$rapidly deteriorates [62].

The second reason is related to the quality factor of the tuning circuitry. As was explained earlier, for oscillators operating at GHz frequencies, we typically assume $Q_{tank} = Q_L$. If this was the case for mm-wave oscillators, we would expect the Q_{tank} to improve with frequency since Q_L increases[2]. However, from Eq. (3.5), we observe that Q_C decreases with increasing frequency for a given R_S, eventually severely limiting Q_{tank}. While fixed capacitors can achieve a very low series resistance, this is not the case for tunable capacitors, whether they are discretely or continuously tuned. Assuming the tunable capacitance C_{tune} has a quality factor Q_{tune} and the fixed capacitance C has an infinite quality factor, Q_{tank} will be given by [61]:

$$Q_{tank} = \left(\frac{1}{Q_L} + \frac{C_{tune}}{C_{tune} + C}\frac{1}{Q_{tune}}\right)^{-1} \tag{3.13}$$

[2] The increase will not be linear with frequency as predicted by Eq. (3.6), due to the skin-effect [92] causing the equivalent series resistance to increase with frequency. Still, we do expect Q_L to increase with frequency.

That is, the larger the portion of the total capacitances that is tunable, the larger the impact on Q_{tank}, and the wider the frequency tuning range we need, the larger the portion of the tunable capacitance must be. Q_{tank} is then reduced and so is the output amplitude, and in the worst case the oscillation even stops. If possible, more bias current can then be used to restore the oscillation amplitude, at the cost of increased power consumption, but if the cross-coupled transistors must be enlarged, there will be more parasitic capacitance reducing the tuning range. Even if the amplitude can be restored, it should be noted that the reduced tank quality factor will result in increased phase noise.

It should also be noted that the assumption that the quality factor of the fixed capacitance is very high is not necessarily true at mm-wave frequencies, since a significant portion of it will not come from intentional capacitors, but rather parasitic capacitances, which can have a substantial series resistance. For instance, the gate resistance will be in series with any parasitic capacitance stemming from transistor gates connected to the resonator tank. Given that minimum length are typically used at these frequencies to maximize the speed of the transistors, this resistance can be significant.

Discrete tuning

The easiest way to implement a discrete tuning cell is to simply connect a capacitor C_{SW} in series with a transistor, which is controlled by a digital signal V_C, either acting as a closed or open switch, see Fig. 21a. When the transistor is on, i.e. the switch is closed, the channel will have a non-zero resistance R_{SW}, see Fig. 21b, given by [63]:

$$R_{SW} = \left(k \frac{W_{SW}}{L_{SW}} V_{ov}\right)^{-1} \tag{3.14}$$

where k is a process-dependent constant, W_{SW} and L_{SW} the width and length of the transistor, respectively, and $V_{ov} = V_{GS} - V_t$ is the overdrive voltage, where V_{GS} is the gate-source voltage and V_t is the threshold voltage. This results in a finite quality factor:

$$Q_{SW} = \frac{1}{\omega C_{SW} R_{SW}} \tag{3.15}$$

From Eq. (3.14), we observe that by using minimum length and increasing the width of the transistor, we can reduce R_{SW}, thereby improving the quality factor. We should also use a transistor with as low threshold voltage as possible to maximize the overdrive. Why not then simply make the transistor so wide that Q_{SW} becomes so high that it is negligible in comparison to the tank quality factor? The problem is

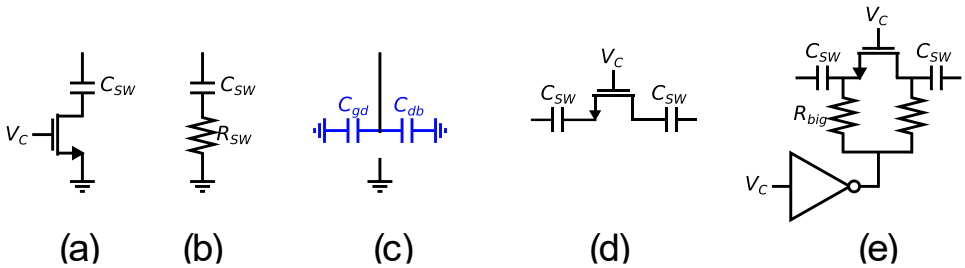


Figure 21: (a) Switched capacitor cell. Equivalent circuit when V_C is pulled (b) high, and (c) low. (d) Differential switched capacitor cell. (e) Differential cell with common-mode control.

parasitic capacitances in the off-state. When the transistor is off, the capacitance will approximately be given by the gate-to-drain capacitance C_{gd} and drain-to-bulk capacitance C_{db} of the transistor in parallel, see Fig. 21c, i.e. $C_{off} \approx C_{gd} + C_{db}$. By doubling the transistor width, C_{gd} and C_{db} also double, reducing the C_{on}/C_{off} ratio. This means that the achievable tuning range is lowered, since the ratio of the highest to lowest oscillation frequency will be:

$$\frac{f_{high}}{f_{low}} = \frac{2\pi}{2\pi}\sqrt{\frac{L(C_1 + C_{on})}{L(C_1 + C_{off})}} = \sqrt{\frac{C_1 + C_{on}}{C_1 + C_{off}}}, \tag{3.16}$$

where f_{high} and f_{low} are highest and lowest free-running oscillation frequencies, respectively, and C_1 the total static capacitance, excluding C_{off}. Clearly, this ratio increases with a higher C_{on}/C_{off} ratio. As an example, if we have $C_{off} = 0.5C_1$ and $C_{on}/C_{off} = 2$, the ratio becomes ~1.15. But if we instead have $C_{on}/C_{off} = 4$ while keeping the C_{off}-to-C_1 ratio constant, it increases to ~1.41.

By making the switched capacitor cell differential, see Fig. 21d, we can double the Q_{SW} for the same C_{on}/C_{off} ratio [63]. This is because the capacitors share the same switch, effectively halving its resistance. An issue with the switched capacitor cell is that the node between capacitor and transistor is floating. When the cell is on, we want the DC voltage to be 0 V to maximize the overdrive, thereby minimizing R_{SW}. However, when the switch is off, the DC voltage should not be 0 V. This is because most of the oscillation signal will be present in this node, even when the switch is off. The voltage will therefore drop below 0 V for half the oscillation cycle, which, if the amplitude is large enough, can forward bias substrate diodes, significantly degrading the Q [54]. This can be solved by connecting an inverted version of the control signal to these nodes through large resistors, see Fig. 21e [63], which will force the DC voltage to be 0 V when the switch is on and VDD when it is off. By making the resistors large, they will have a negligible impact on the Q.

Continuous tuning

For the continuous tuning, we can use the nonlinearity of the gate capacitance of a transistor. By connecting the body, drain and source terminals of a transistor together and applying a DC voltage V_{ctrl} to this node, while keeping the DC voltage of the gate fixed (typically at VDD when it is connected to the resonator tank of an NMOS cross-coupled oscillator), see Fig. 22a, we can vary the capacitance. Fig. 22b shows a typical varactor capacitance, C_{var}, versus V_{ctrl} behaviour for an accumulation-mode transistor, which is preferable over the regular inversion-mode transistor for these applications [64]. As seen in the figure, the whole tuning range may not be achievable due to the voltage limitations of the technology, contributing to the problem mentioned earlier with poor C_{max}-to-C_{min} ratio. Additionally, if the transistors are made short to maximize the quality factor, a significant part of the total gate capacitance will be due to parasitic overlap capacitance, which is not tunable. Thus, this will further limit the achievable C_{max}-to-C_{min} ratio [64].

Given that the gate of the varactor is connected to the oscillator tank, which has a voltage V_{osc} that varies significantly over one oscillation period, the effective varactor capacitance, $C_{v,eff}$, seen by the tank will not quite follow the plot shown in Fig. 22b. As the gate voltage varies, so will the instantaneous capacitance, and since the C-V characteristic is non-linear, $C_{v,eff}$ will depend on the oscillation amplitude. It is tempting to assume that $C_{v,eff}$ will simply be the average of the instantaneous capacitance over one oscillation period, but this is not the case [65]. Instead, $C_{v,eff}$ will be given by [65]:

$$C_{v,eff} = C_{v,0} - \frac{1}{2} C_{v,2}, \tag{3.17}$$

where $C_{v,0}$ is the time-average capacitance and $C_{v,2}$ the second-order Fourier coefficient of the function $C_{var}(V_{var}(t))$, where $V_{var}(t) = V_{osc}(t) - V_{ctrl}$, i.e. the time-varying voltage across the varactor. This will smooth out the C-V characteristic, with a higher amplitude resulting in a more linear tuning curve [65] [66], as shown in Fig. 22c. While this more linear tuning simplifies the PLL design, it even further reduces the already poor C_{max}-to-C_{min} ratio, at least in a modern, low-voltage CMOS process. $C_{v,eff}$'s dependence on the oscillation amplitude is also the reason why varactors cause AM-to-PM conversion, since amplitude variations will alter $C_{v,eff}$, which in turn shifts the oscillation frequency [65]. As was noted earlier, this turns amplitude noise from the current source into phase noise [60]. Interestingly, at the V_{ctrl} where $C_{v,eff} = C_{var}$, the effective capacitance is independent of the amplitude, and thus the noise will be minimized [65].

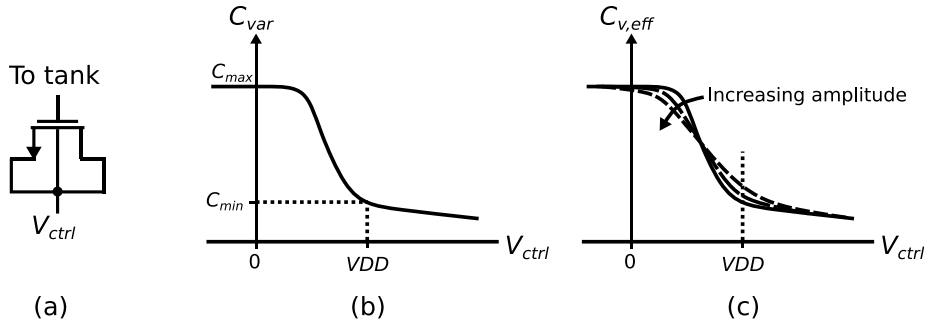


Figure 22: (a) Varactor schematic. (b) Capacitance versus control voltage in a varactor. (c) Effective capacitance for different amplitudes.

Quadrature frequency generation

Modern cellular communication relies on quadrature modulation which requires quadrature LO signals, i.e. two identical signals, but with a 90° phase shift between them. This enables separation of negative and positive frequency content at the baseband. However, if the LO signals are not perfectly separated by 90° and/or have non-equal amplitudes, the separation will not be perfect, causing the negative and positive frequency baseband signals to interfere with each other. This in turn results in signal distortion. How well the transceiver can separate the positive and negative baseband frequencies is indicated by the *image-rejection ratio* (IRR) [67]:

$$IRR = \frac{1 + 2(1+\epsilon)\cos\alpha + (1+\epsilon)^2}{1 - 2(1+\epsilon)\cos\alpha + (1+\epsilon)^2}, \tag{3.18}$$

where ϵ is the difference between the I and Q amplitudes, normalized to the average amplitude, and α is the quadrature phase error. Fig. 23 shows a contour plot of the resulting IRR for varying phase and amplitude errors. The required IRR value will depend on the modulation order, with for instance 64-QAM requiring a higher IRR than 16-QAM.

The choice of transceiver architecture will determine at what frequency the quadrature signals must be generated: for a direct-conversion architecture, this will be at the carrier frequency, while in a heterodyne architecture, the quadrature signals are only required at the IF frequency. This is very important for mm-wave systems, since implementing accurate quadrature signals at these frequencies, as will be shown shortly, is far from trivial.

For generating the quadrature signals, there are three common ways: using a polyphase filter, using a quadrature oscillator, and generating a signal at twice the desired frequency and then using a frequency divide-by-2 circuit to generate quadrature signals at the desired frequency [68] [67] [69]. For mm-waves, a fourth

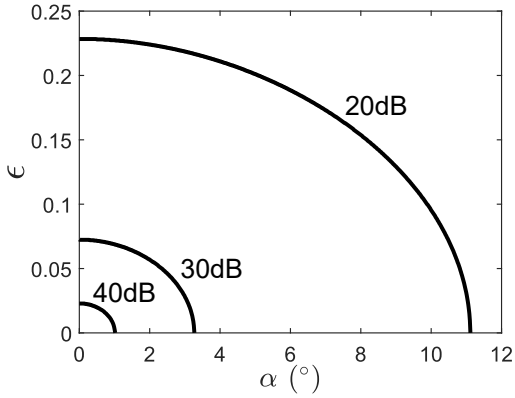


Figure 23: Tolerable amplitude and phase errors to achieve 20, 30, and 40 dB of IRR.

option also exists, quadrature hybrids [70] [71]. Each one of them has its advantages and disadvantages.

Divide-by-2 circuits to generate quadrature signals are based on the principle that it is not only the frequency that is divided by 2, but also the phase. A differential signal has two phases, 0° and 180°, so when that signal is divided by two, the phases will instead be 0° and 90°, i.e. the signals will be in quadrature. Divide-by-2 circuits are very common at sub-7 GHz frequencies, because (a) implementing an oscillator at twice these frequencies can be done with very limited penalties, and (b) the dividers are also relatively easy to implement at these frequencies, as they can be implemented with digital circuit blocks such as flip-flops, yielding very high quadrature accuracy and low power consumption [54]. However, these points do not hold true for mm-wave frequencies. A VCO operating at, for instance, 56 GHz to generate 28-GHz signals will have worse performance than a 28-GHz VCO, especially if wideband operation is targeted. The divide-by-2 circuit will also not be possible to implement with conventional digital blocks, making extracting accurate quadrature signals more difficult. For these reasons, the divide-by-2 approach has seen little use in the literature for mm-wave frequencies. One exception is [72], where they do manage to generate quadrature 28-GHz signals using this technique with relatively low power consumption (4.7 mW for the divider). However, the quadrature error is 8.5° at the output, an unacceptable number.

Polyphase filters (PPF) generate quadrature signals by passing the signal through an *RC* network, see Fig. 24a. The quadrature accuracy of a PPF is frequency dependent, with perfect quadrature occurring only at $f_1 = 1/(2\pi RC)$, assuming perfect matching between the resistors and between the capacitors. To obtain high quadrature accuracy over a wider bandwidth, multiple PPF stages can be cascaded. Fig. 24b shows the difference in IRR response for a single-stage and 3-stage PPF. The usable PPF bandwidth can, for a given acceptable IRR, be optimized by having a different RC constant for each stage [73]. An example of this is also shown in Fig. 24b, where a wider PPF bandwidth is achieved if an IRR > 40 dB is targeted. The

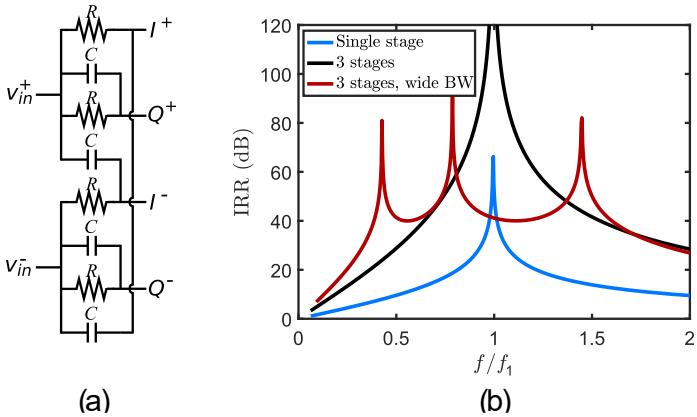


Figure 24: (a) Single-stage PPF. (b) IRR for different PPF configurations.

issue with cascading multiple stages is that each stage contributes losses [68], especially at mm-waves, which must be compensated for by power-hungry buffers. Additionally, the PPF will at mm-wave frequencies be sensitive to process variations, due to small physical size of components and relatively large parasitics [67] [69], typically requiring some form of tuning, as for instance is done in [74].

Quadrature hybrids are less sensitive to process variations and typically have less insertion loss than PPF at mm-wave frequencies. Quadrature hybrids are based on transmission lines[3] and can be implemented in multiple ways, such as branch-line [70] [75], coupled-line [71], and Lange couplers [76]. Each implementation has their advantages and disadvantages, but in common for all of them is a relatively large area consumption. This is especially true at the lower range of the mm-wave spectrum, as the dimensions will be proportional to the wavelength. Additionally, they have a limited bandwidth if a high IRR is targeted.

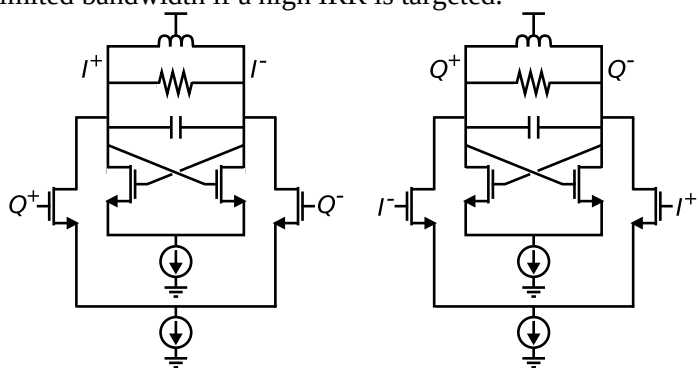


Figure 25: The basic quadrature oscillator.

[3] The transmission lines can be replaced by lumped-element counterparts [75]. This is typically done at sub-10-GHz frequencies, as implementing transmission lines would be too area consuming.

Lastly, quadrature oscillators are built from two separate oscillators that are locked to each other, forcing quadrature oscillation, see Fig. 25. The main issue is that it does not operate at the frequency of the impedance peak of the resonator, yielding worse phase noise performance [54] [77]. The coupling itself also limits the performance in several ways. The injection transistors' finite output resistance can degrade the tank *Q*, while also capacitively loading the other tank, limiting the achievable tuning range. Quadrature oscillators will be discussed further later in this chapter.

While the direct-conversion transceiver is the most popular sub-7 GHz architecture nowadays, the difficult implementation of mm-wave quadrature frequency generation, among other reasons, has made heterodyne architectures the more popular choice for mm-wave implementations [78]. Still, there are many benefits to using a direct-conversion architecture, such as smaller area, fewer components, no concern of image frequencies, and wider bandwidth [69] [74].

Injection-locking and frequency multiplication

As was noted in Chapter 2, injection-locking is when an oscillator is forced to oscillate at the frequency of an injected signal, referred to as an injection-locked oscillator (ILO), and it can be used to implement both phase shifters and frequency multipliers. Additionally, it can be used in frequency dividers and it is the mechanism behind quadrature oscillators. In this section, a more thorough description of the process will be given.

Consider the simplified LC oscillator shown in Fig. 26a. Barkhausen's stability criterion tells us that the phase shift around the loop must be zero. This is provided by the combination of the negative unity amplifier and transistor, implying that no additional phase shift can be added by the resonance tank, i.e., the oscillator will oscillate at the resonance frequency of the tank, which is then the free-running frequency of the oscillator ω_0. Let us now inject a current through transistor M_2 with magnitude $|I_{inj}|$ and frequency ω_{inj}, see Fig. 26b. To begin with, we assume $\omega_{inj} = \omega_0$. This means that the resonance tank adds no phase shift, and thus the injected current must be in phase with I_{osc}, shown in the phasor plot in Fig. 26b. This also means that a larger current will pass through R_P, increasing V_{osc}. If we then increase ω_{inj} slightly, so that the oscillator remains locked to the injected frequency, the oscillator is no longer oscillating at its peak resonance frequency. As seen in Fig. 19b, the resonance tank will then add a non-zero phase shift of ϕ_{tank}. To preserve a phase shift of 180° between V_{osc} and I_{osc}, I_{osc} and I_{inj} can no longer be in phase, but must have a phase difference θ so that the tank current I_T is phase shifted by $\phi = -\phi_{tank}$ relative to I_{osc}, see Fig. 26c. An interesting observation

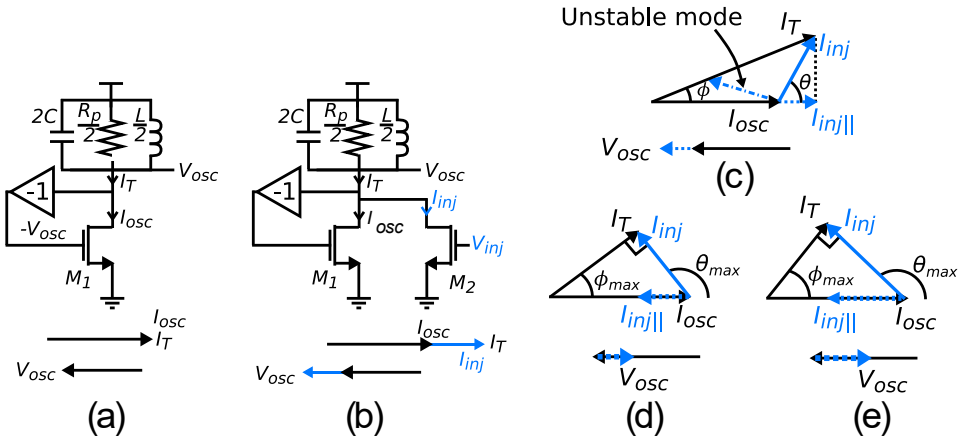


Figure 26: Injection-locking. (a) Simplified oscillator without injection. (b) Injection with $\omega_{inj} = \omega_0$. (c) Phasor diagram when $\omega_{inj} \neq \omega_0$. (d) Phasor diagram at the edge of lock. (e) The effect of stronger injection.

from this figure is that there are two solutions for θ for which $\phi = -\phi_{tank}$, also shown in Fig. 26c. However, it can be shown that it is only the solution with the smaller absolute angle of θ that is stable [79]. Also note that V_{osc} must remain in anti-phase with I_{osc}, since $I_{osc} = -G_{m1}V_{osc}$. This means that to understand what happens to V_{osc}, we must consider the part of I_{inj} that is parallel with I_{osc}, referred to as $I_{inj||}$, following the notations in [79]. In the scenario shown in Fig. 26c, $I_{inj||}$ is in phase with I_{osc}, resulting in a larger V_{osc} than in the free-running case, albeit smaller than if $\omega_{inj} = \omega_0$.

The higher we make ω_{inj}, the larger the phase shift in the resonance tank will be and the larger θ must be to counteract this. As the magnitude of I_{inj} is fixed, its phasor will move in a circle. This means that the maximum achievable angle of $\phi = \phi_{max}$ is when I_T and I_{inj} are orthogonal, see Fig. 26d. Using trigonometry, we can calculate:

$$\tan \phi_{max} = \frac{I_{inj}}{I_T} \tag{3.19}$$

$$\cos \theta_{max} = -\frac{I_{inj}}{I_{osc}} \tag{3.20}$$

If the difference $\Delta\omega = \omega_{inj} - \omega_0$ is so large that $|\phi_{tank}| > |\phi_{max}|$, the injected current can no longer compensate the added phase shift and lock is lost. However, the oscillator will not fully return to ω_0 when this happens due to an effect called pulling, which will be discussed later. It can easily be seen that a larger I_{inj} results in larger achievable ϕ_{max}, see Fig. 26e, which in turn yields a larger locking range

(LR), i.e. a larger maximum $\Delta\omega = \Delta\omega_{max} = \omega_L$ which still results in injection lock. To calculate ω_L, we need to consider how ϕ_{tank} behaves in the vicinity of ω_0. It can be shown that [79]:

$$\tan\phi_{tank} = Q_{tank}\left(\frac{\omega_0}{\omega_{inj}} - \frac{\omega_{inj}}{\omega_0}\right) \tag{3.21}$$

If we assume $|\Delta\omega| \ll \omega_0$, this can be approximated as:

$$\tan\phi_{tank} \approx -2Q_{tank}\left(\frac{\Delta\omega}{\omega_0}\right) \tag{3.22}$$

At the edge of lock, we have $\phi_{tank} = -\phi_{max}$, so by combining Eq. (3.19) and Eq. (3.22), and replacing I_T with $\sqrt{I_{osc}^2 - I_{inj}^2}$, we eventually get [47]:

$$\omega_L = \frac{\omega_0}{2Q_{tank}}\frac{I_{inj}}{I_{osc}}\frac{1}{\sqrt{1 - \frac{I_{inj}^2}{I_{osc}^2}}} \tag{3.23}$$

ω_L will be symmetric about ω_0, so the total LR will be $\omega_0 \pm \omega_L$.[4]

From Fig. 26d and Fig. 26e, we can also observe that $I_{inj||}$ in these scenarios will be in anti-phase with I_{osc}, implying a V_{osc} smaller than in the free-running case.

Phase shifting

As was noted in Chapter 2, injection-locking can be used as a phase shifter, as is done in **Paper I.** The phase difference θ between the injected signal and output signal becomes, assuming $I_{osc} \gg I_{inj}$ [47]:

$$\sin\theta \approx \frac{2Q_{tank}}{\omega_0}\frac{I_{inj}}{I_{osc}}\Delta\omega \approx \frac{\Delta\omega}{\omega_L} \tag{3.24}$$

This predicts an achievable phase shift of up to $\pm 90^\circ$. However, if the assumption $I_{osc} \gg I_{inj}$ does not hold, then a slightly larger phase shift is achievable [47]. This can also be seen in Fig. 26d, where θ must be greater than 90° to achieve ϕ_{max}.

One issue with using an ILO for phase shifting is that the output amplitude will vary with the phase shift setting, as was shown earlier. This breaks the desired orthogonality between phase and amplitude settings in a beamformer. To solve this issue, we implemented a peak detector [80] in **Paper I**, which measures the

[4] The locking range will in general actually not be symmetric about ω_0. ϕ_{max}, on the other hand, will always be symmetric about 0°, so this error comes from the approximation done in Eq. (3.22). For a more thorough discussion on this, the reader is referred to [79].

amplitude of ILO and compares it to a reference value. If the amplitude is lower than the reference value, the bias current is increased, which we know from Eq. (3.14) increases the amplitude. If the amplitude is higher than the reference value, the bias current is instead lowered. The injection bias current is also increased/decreased with the same relative amount as the oscillator bias current, to minimize the impact on the applied phase shift.

LO pulling

What happens if we gradually increase $\Delta\omega$ until it is larger than ω_L? As noted earlier, the oscillator will not fully return to ω_0, but will still be influenced by the injected signal. This is known as *LO pulling* and is often a highly undesirable effect. For instance, in a frequency division duplex (FDD) system, the transmitter and receiver operate on two slightly different frequencies, which necessitates two oscillators operating at different frequencies. These oscillators can pull on each other, causing distortion in both the transmitted and received signal. However, in **Paper III**, we take advantage of this effect to detect if the oscillator is locked, as will be explained later in this chapter.

LO pulling can be divided into two sub-categories: *quasi-lock* and *fast beat* [47]. Before explaining these effects, we define a frequency ω_B:

$$\omega_B = \sqrt{\Delta\omega^2 - \omega_L^2} \tag{3.25}$$

Quasi-lock occurs when the injected signal is very close to the LR. Over a period of $2\pi/\omega_B$, the oscillator will lock to injected signal for part of the period, but then slip to the frequency $\omega_{\mathrm{inj}} + \omega_{\mathrm{L}}$, see Fig. 27a. Analyzing the frequency spectra reveals that tones of signal will be situated at $\omega_{\mathrm{inj}} + n\omega_{\mathrm{B}}$ for $n = 0,1,2$ ..., with ω_{inj} being the strongest tone, also shown in Fig. 27a.

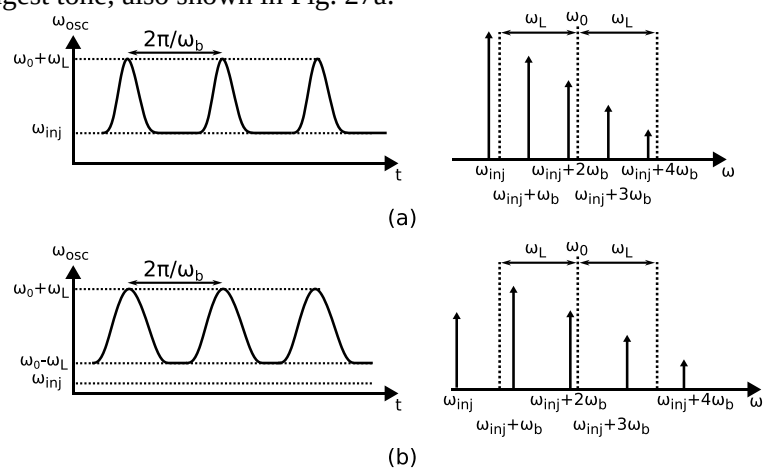


Figure 27: (a) Transient and frequency spectrum during quasi-lock. (b) Transient and frequency spectrum during fast beat.

Fast beat occurs when $\Delta\omega$ is significantly larger than ω_L. The instantaneous frequency will have similar behavior as in the quasi-lock case, but it will never reach ω_{inj}. It will instead be limited to $\omega_0 - \omega_{\mathrm{L}}$. Once again, the tones are given by $\omega_{\mathrm{inj}} + n\omega_{\mathrm{B}}$ for $n = 0,1,2$ ..., but this time $\omega_{\mathrm{inj}} + \omega_{\mathrm{B}}$ will be the strongest tone, see Fig. 27b.

Frequency multiplication

Injection-locking is readily used for frequency multiplication, i.e. the oscillator locks onto the frequency $M \cdot \omega_{inj}$, where M is the multiplication factor. If the signal is injected differentially (Fig. 26b), M should be an odd number, whereas if it is done in a common-mode node, M should be an even number.

The principle behind the frequency multiplication is that the non-linearity of the injection transistors will generate harmonics of the injected signal. If the resonance frequency of the tank is close enough to the frequency of the desired harmonic, the oscillator will lock to this harmonic. This means that the injection transistor should be as non-linear as possible to achieve a reasonable LR. Hence, a short transistor biased in class-C or class-D is usually preferable. Still, even if these techniques are used, a frequency multiplier will always have a worse LR than a regular ILO for a given bias current. This makes them susceptible to PVT variations, as it is easier for such variations to alter the free-running frequency to such a degree that lock is lost.

Frequency multipliers has become a very popular choice for implementing the frequency generation in large, mm-wave beamforming transceivers [29] [30] [81] [82] [83], mainly for three reasons. Firstly, it simplifies the clock distribution, since the multipliers can be placed very close to their respective mixer, while only a ω_0/M signal needs to be distributed. Secondly, it typically improves the phase noise performance in mm-wave applications, as will be explained in the next subsection. Lastly, in the case of direct-conversion transceivers, it reduces the pulling from the power amplifier on the LO [78].

In both **Paper I** and **Paper III**, injection-locked frequency triplers (ILFT) are used. As mentioned above, frequency multipliers are sensitive to PVT variations, unless the injection circuitry consumes an excessive amount of power. To combat this, the implementations in both papers rely on automatic tuning of the tripler. The implementation in **Paper III** will be discussed later. In **Paper I**, there is also an ILO that should be tuned independently of the ILFT. In Chapter 2, it was explained how two PDs are used to measure the phase shift applied by the ILO, but interestingly, we can also use the PDs to automatically find lock of both the ILO and ILFT. It is done as follows: Assume that we inject a signal V_{inj} to which neither the ILO nor the ILFT are initially locked, and thus they are oscillating at their respective free-running frequencies, f_{ILO} and f_{ILFT} (ignoring any pulling effects), see Fig. 28a. Since $3f_{inj} \neq f_{ILFT}$, no DC voltage will be generated by either of the PDs. If we then start tuning f_{ILO}, the ILO will eventually lock, see Fig. 28b. This causes the

ILO to output a significant voltage at f_{inj}. This signal will be so strong that part of its third harmonic will leak through the ILFT to the PD and mix with f_{inj}, causing a non-zero DC voltage. This will cause $v_{PD,I} - v_{PD,Q}$ to diverge from 0 V, and when the difference is greater than some certain threshold, we can tell that the ILO is locked. Next, f_{ILFT} is tuned, see Fig. 28c. The absolute value of the PD outputs, i.e. $|v_{PD}| = \sqrt{v_{PD,I}^2 + v_{PD,Q}^2}$, will be proportional to the amplitude of the signal at $3f_{inj}$ in the ILFT. The maximum amplitude will happen when $f_{ILFT} = 3f_{inj}$. Thus, we simply tune f_{ILFT} until $|v_{PD}|$ reaches its maximum value, at which point the ILFT is tuned to its ideal frequency.

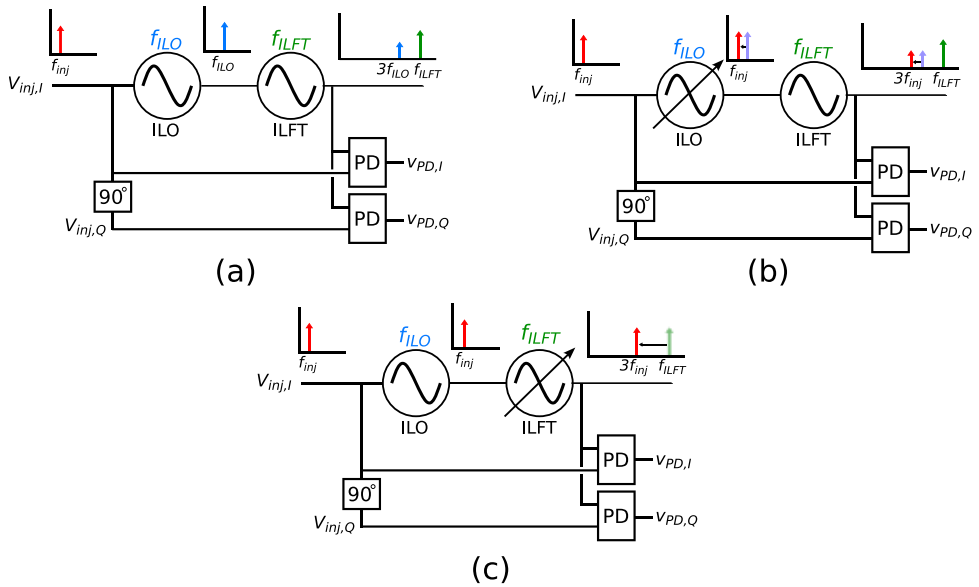


Figure 28: Automatic tuning for lock. (a) Initial state. (b) Tuning of the ILO. (c) Tuning of the ILFT.

Phase noise in injection-locking

An interesting property of injection-locking from a frequency generation point-of-view is that phase noise will also be locked to the input signal, i.e. the phase noise profile of the injection-locked oscillator will follow that of the injected signal, as long as the phase noise is within the locking bandwidth [47]. This is also true if the signal is multiplied, although the phase noise will then increase by a factor of the multiplication factor M squared. That is, a frequency tripler will increase the phase noise level of the injected signal by a factor of $3^2 = 9$, or 9.5 dB. From Leeson's equation (Eq. (3.12)), we can observe that the same phase noise increase is expected if we triple ω_0 of an oscillator, assuming that all other parameters remain constant. However, as has been explained earlier, Q_{tank} is typically significantly lower at

mm-wave frequencies due to the frequency tuning circuitry. So, we can expect a lower frequency oscillator followed by a mm-wave frequency multiplier to have better phase noise performance than an oscillator directly generating the mm-wave frequencies.

One drawback of using frequency multipliers is that the phase noise floor will also be multiplied by M^2. The phase noise floor has previously been of little concern to in-channel reciprocal mixing and symbol jitter since when integrating the total phase noise up to offsets corresponding to half the channel bandwidth, its contribution has been negligible. However, as was noted earlier, this will no longer be the case in sub-THz 6G, with channel bandwidths in the multi-GHz range. In fact, the phase noise floor can then even be the dominant source of clock jitter, especially if frequency multipliers are used [84] [85].

Another thing to consider with injection-locking is that the phase noise will be correlated with the phase noise of the injected signal. As discussed in Chapter 2, this is beneficial from the perspective of array factor and ability to generate deep nulls, but a drawback if instead the symbol distortion due to clock jitter is the main concern.

Series injection

So far, the injected signal has been injected in *parallel* with oscillator transconductor. It is also possible to inject a signal in *series* with the transconductor, see Fig. 29a. While not as common as its parallel counterpart, series injection has been used for frequency multipliers [69] [86] [87] and dividers [88], and quadrature oscillators [69] [86] [89] [90].

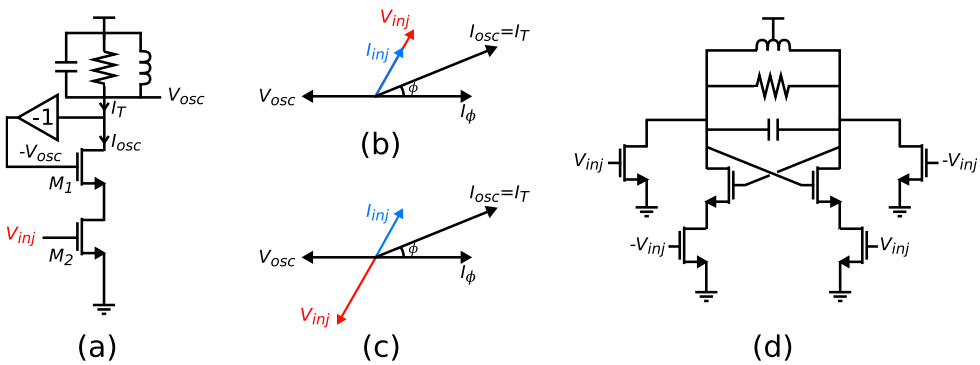


Figure 29: (a) Series injection. (b) Phasor diagram for $n = 1,5,9$ ... (c) Phasor diagram for $n = 3,7,11$ ... (d) Dual-injection.

The major benefit of series injection is that it separates the injection transistors from the resonance tank, removing the tank Q degradation and parasitic capacitances associated with these transistors. However, series injection causes

source degeneration of the cross-coupled transistors, lowering the effective transconductance [91]. In [91], the authors analyze the mechanism of series injection when the injection transistor is biased in class-D. They show that the tank current I_T, similar to parallel injection, can be split into two currents; one in anti-phase with the oscillation voltage, which they refer to as I_ϕ, and one due to the injected signal, called I_{inj}, see Fig. 29b. Their analysis shows that I_{inj} will alternate between being in phase and in anti-phase with the injected voltage, V_{inj}, depending on which harmonic is considered. Harmonics $n = 1,5,9$ ... will be in phase (Fig. 29b), while harmonics $n = 3,7,11$ ... will be in anti-phase[5] (Fig. 29c). This is in stark contrast with parallel injection, where I_{inj} will always be in phase with V_{inj} for all harmonics. Based on this analysis, they proposed a frequency tripler with dual injection, using both series and parallel injection, see Fig. 29d. The parallel injection transistors and the series injection transistors are driven in anti-phase. This results in the 3rd harmonics of the two injection methods being in-phase, while the fundamental tones are in anti-phase, which improves the locking range and harmonic rejection. A similar technique is also used in [69] and in **Paper III**, although the series injection transistors are driven in class-AB, resulting in less third harmonic content.

Quadrature signals

Quadrature oscillators

Quadrature oscillators (QO) are fundamentally injection-locked systems, since two oscillators are locked to each other. To understand why they will oscillate in quadrate, the typical QO (Fig. 25) is redrawn in a simplified way, see Fig. 30a, where $T(s)$ is the free-running loop gain of each oscillator. From this, it can be shown that [61]:

$$(X^2 + Y^2)(T(s) - 1) = 0 \tag{3.26}$$

As will be shown shortly, $T(s) - 1 \neq 0$ at the frequency of oscillation, implying:

$$X^2 + Y^2 = 0 \Longrightarrow X = \pm jY \tag{3.27}$$

That is, the voltages must be in quadrature. If we draw the phasors, something interesting can be observed. The signals are in quadrature, $\theta = 90^\circ$, and thus $\phi \neq 0^\circ$, see Fig. 30b, which means that the oscillator must oscillate at frequency $\omega_{QO} \neq$

[5] Since they are analyzing a differential circuit, only odd harmonics are considered.

ω_0 to generate a non-zero phase shift. For this reason, $T(j\omega_{QO})$ must be complex and thus cannot be equal to 1.

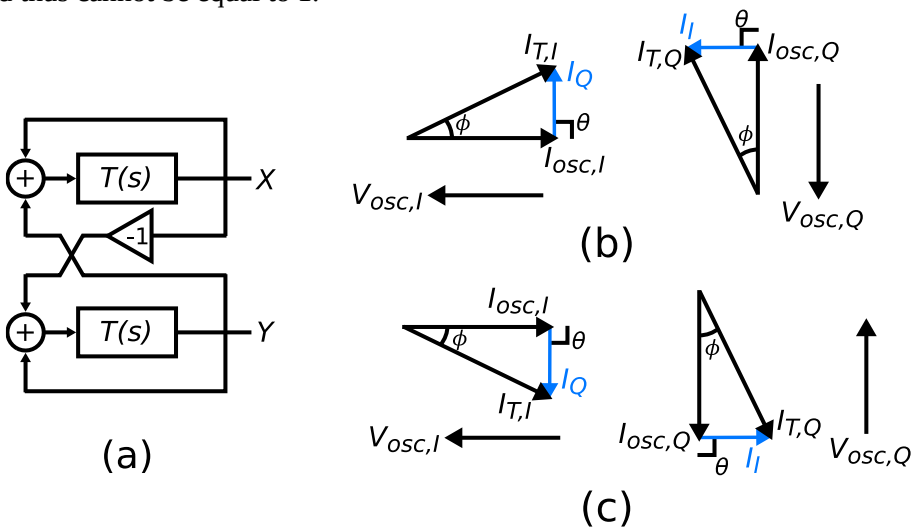


Figure 30: (a) Simplified view of a QO. (b) Phasor diagram of a QO. (c) Alternative oscillation mode.

Interestingly, there is another solution that also leads to quadrature signals, as shown in Fig. 30c. That is, the QO has two modes of oscillation, one above ω_0 (Fig. 30b) and one below ω_0 (Fig. 30c), which has also been observed in measurements [77]. However, the mode above ω_0 tends to prevail due to it having a slightly higher amplitude than the other mode [92].

Quadrature frequency multipliers

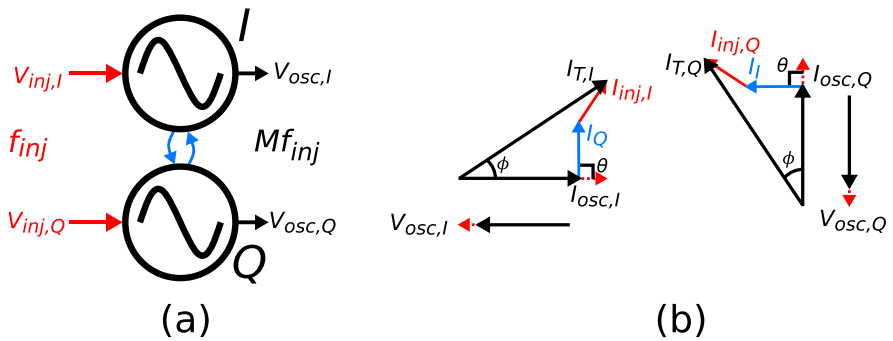


Figure 31: Quadrature frequency multiplier. (a) Overview. (b) Phasor diagram.

As was explained earlier in this chapter, it is difficult to implement quadrature frequency generation at mm-wave frequencies with acceptable performance. But we have also seen that injection-locked frequency multiplication can yield good phase

noise performance. This technique can also be applied to quadrature oscillators, thus improving on one of their major drawbacks. The most straightforward way of doing this is to generate quadrature signals at a lower frequency f_{inj} and use both the I and Q signals to lock the frequency multiplier [86] [93] [94] [95], see Fig. 31. This approach has three fundamental problems: Firstly, any quadrature phase error at f_{inj} will be multiplied by M, forcing very high quadrature accuracy at the lower frequency to achieve acceptable performance. Secondly, the quadrature signal at f_{inj} will in general have worse phase noise performance compared to a differential oscillator at the same frequency, limiting the final achievable phase noise. Lastly, distributing a quadrature signal severely complicates the layout compared to a differential signal. In [95], they solve the two latter problems by distributing a differential signal and generating the quadrature injection signal with a PPF right next to the quadrature injection-locked frequency tripler (QILFT). However, the first problem remains. In [94], they use a similar approach, but add a quadrature coupler at the output to improve the quadrature accuracy, although its effect is not quantified. Additionally, the circuit is very large, occupying a chip area of 0.84 mm^2 (including pads).

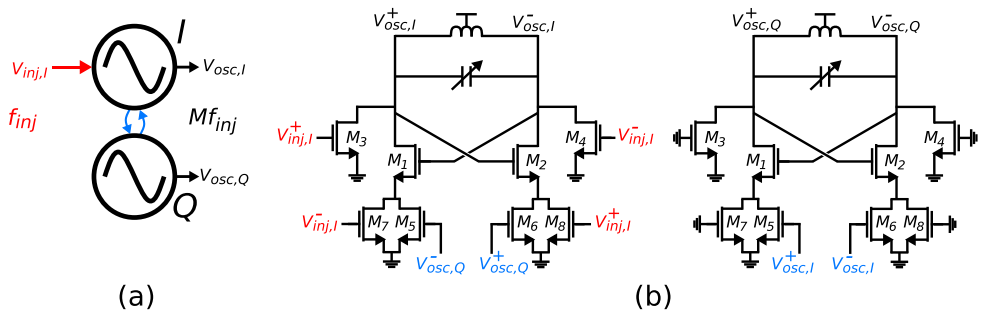


Figure 32: Differential-to-quadrature frequency multiplier. (a) Overview. (b) Schematic (biasing circuitry omitted).

These issues would all be resolved if we instead could distribute a differential signal and inject that into a frequency multiplier which outputs quadrature signals, see Fig. 32a. This is the concept of [69], which is also used in [81] and **Paper III**. A differential signal f_{inj} is injected into the I core of a QO operating at f_{Qosc}, which is in the vicinity of $3f_{inj}$, locking the QO to $3f_{inj}$. For the injection, the dual injection scheme described earlier is used and the quadrature coupling is realized through series injection, see Fig. 32b. The Q core, not subjected to the injection signal, has dummy injection transistors to minimize the mismatch between the cores. While this eliminates all the previously described issues, it adds two new ones: phase and amplitude accuracy. To understand why, we once again turn to the phasor description. Without the injected signal, we have the scenario shown in Fig. 33a, where $\phi_I = \phi_Q = \phi_1$. Next, the injection signal is applied to the I core. In

general, I_{inj} will not be in phase with $I_{T,I}$, which changes ϕ_I to ϕ_2, see Fig. 33b. Since the Q core is locked to the I core, they have the same oscillation frequency, which makes $\phi_Q = \phi_I = \phi_2$. This can only be achieved if the I and Q cores are not in quadrature, see Fig. 33b[6]. The only way to avoid this quadrature error is if $I_{inj,I}$ is perfectly in phase with $I_{T,I}$, see Fig. 33c, which happens if $3f_{inj} = f_{QO}$, since ϕ_I will then remain equal to ϕ_1. The QO must therefore always be tuned to exactly, or at least very close to, the third harmonic of the injected frequency for it to be usable.

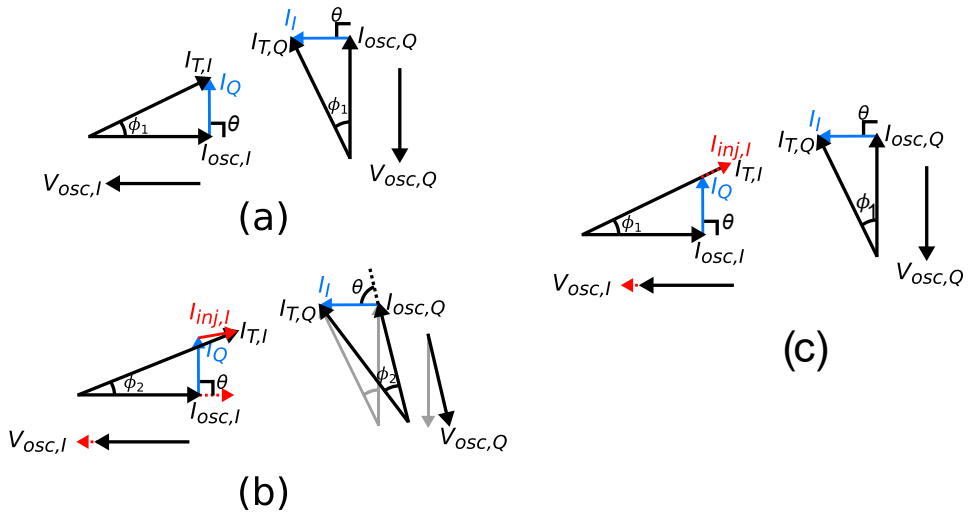


Figure 33: (a) Phasors without injection. (b) Phasors when $3f_{inj} \neq f_{QO}$. (c) Phasors when $3f_{inj} = f_{QO}$.

If we assume that $3f_{inj} = f_{QO}$, then it can be observed from Fig. 33c that $V_{osc,I}$ will be larger than $V_{osc,Q}$. Thus, there will be an amplitude mismatch between I and Q, even when the QO is perfectly tuned. This amplitude mismatch will cause a small quadrature phase error, since I_I should be slightly larger than I_Q. However, it should be clear that there exists an f_{QO} very close to $3f_{inj}$ for which the phase error due to amplitude mismatch cancels the phase error due to $f_{QO} \neq 3f_{inj}$ so that the signals are in perfect quadrature.

In [69], they argue that the use of dual injection can, if the injection transistors are sized correctly, remove this amplitude error. However, they only consider the fundamental tone at f_{inj} to be the source of error and provide no simulations or measurements to strengthen their case, while the above analysis clearly shows amplitude errors due to $3f_{inj}$. Still, the cancellation of the f_{inj} does remove some amplitude and phase error. One way to reduce the amplitude error is to weaken the

[6] The final ϕ_I and ϕ_Q will not be exactly as shown in Fig. 33b, since ϕ_I is drawn assuming $\theta_I = 90°$, which we have shown is not the case. The general principle is still true though.

injection current, at the cost of a smaller LR, while the quadrature coupling is made relatively strong.

To solve the issue with quadrature phase error, we implemented a feedback system in **Paper III**, which is shown in Fig. 34. A phase detector, implemented as an active mixer, measures the quadrature phase error α between the I and Q outputs. The PD outputs a DC voltage proportional to $\sin\alpha$, which is digitized by an ADC and processed by a simple DSP. The DSP will then turn on or off switched capacitor cells in QO until $\sin\alpha$ is close to 0, which is the equivalent of the oscillator being tuned so that $3f_{inj} \approx f_{QO}$. This has assumed that the QO was locked to the injected signal. What happens then if the injected signal is outside the LR? Initially, one might think that there should be no signal detected by the phase detector, since the tones present would be $3f_{inj}$ and f_{QO}, and that $f_{QO,I}$ and $f_{QO,Q}$ are in quadrature. This would result in $V_{PD} = 0$V and the feedback system would interpretate that as the QO being perfectly tuned to $3f_{inj}$. However, this is not the case. As was explained earlier, even when the injected signal is outside the LR, it will still pull the oscillator, causing the oscillator to shift its oscillation frequency. The pulling will act on the I oscillator, generating what is essentially a mismatch between the I and Q oscillators. This results in a quadrature phase error, which can be detected by the PD and used to tune the oscillator until lock is found. An added benefit of the automatic tuning scheme is that the LR, and therefore I_{inj}, can be kept relatively small. As explained above, this results in a low quadrature amplitude error.

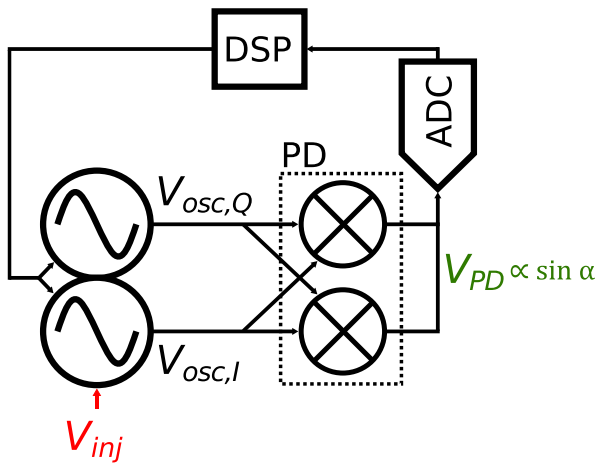


Figure 34: The feedback system used in **Paper III**.

Chapter 4

Integrated Wideband Baseband Filters

The Nyquist-Shannon sampling theorem is known to every electrical engineer and states that to be able to perfectly reconstruct a signal with its highest frequency content at f_H, the sampling rate f_s must be:

$$f_s > 2f_H \tag{4.1}$$

Any spectrum content above $f_s/2$ will be folded after the sampling, in a process referred to as aliasing, and distort the wanted signal, as seen in Fig. 35. This is where the baseband filter comes in. Its task is to filter spectrum content above $f_s/2$, also known as the Nyquist rate f_N, to avoid aliasing when sampling, which is why it is called an anti-aliasing (AA) filter. Since this is applied at the baseband, it is also often referred to as a baseband filter.

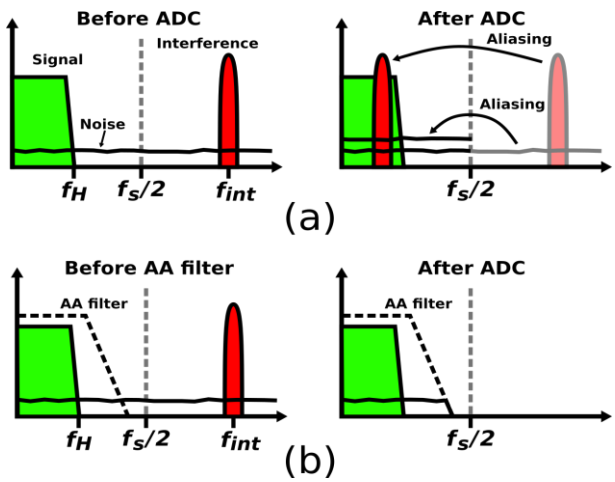


Figure 35: The effect of aliasing (a) without and (b) with an AA filter.

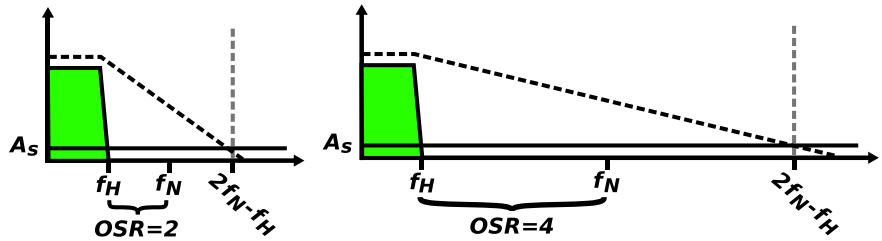


Figure 36: The effect of OSR on required sharpness of the filter.

The sharpness of the filter will be determined by its order and type, and the requirements can be determined by the data converter specifications and how much distortion the system can tolerate. More specifically, the most important parameter will be the converter's oversampling ratio (*OSR*), i.e., the ratio between the Nyquist rate $f_N = f_s/2$ and f_H. Fig. 36 illustrates two scenarios, one where $OSR = 2$ and one where $OSR = 4$. The filter must reach an attenuation of $|A_s|$ at $2f_N - f_H$ to not add excessive noise in the frequency channel of the desired signal. Clearly, the second scenario can get away with a less sharp filter, and hence a lower filter order can be used. The noise and interference between f_H and f_N can subsequently be filtered in the digital domain. However, if the AA filter can also suppress some of the interference at these frequencies, it will reduce the ADC dynamic range requirements.

In most wireless communication standards, the channel bandwidths are limited to the MHz range. Given the improvements in data converters with semiconductor process scaling, and that their power consumption scales linearly with sampling rate at these frequencies [96], this means that modern transceivers can operate with a significant *OSR* with negligible increase to the total power budget. This does not only relax the AA filter requirements, but also improves the converter dynamic range [97]. However, this will not be case for multi-GHz bandwidth millimeter-wave 6G receivers. There, the ADCs are expected to be among the most power-hungry parts of the receivers, and while the Walden FoM [98] predicts a linear relation between power and sampling rate, this is only true up to the 100-MHz range [99], after which power increases closer to quadratically with sampling rate. This means that high-order AA filters will be required in sub-THz 6G. At the same time, the filters should not limit the receiver dynamic range, all while being as compact as possible. The latter is of extra importance in sub-THz antenna arrays if the baseband circuitry is placed in the chip(s) adjacent to the antenna array, since, as noted in Chapter 2, for these chips the dimensions per channel must be smaller than the antenna pitch.

This chapter covers the design of both active and passive integrated baseband filters for multi-GHz bandwidth applications. But first we start with some general filter theory.

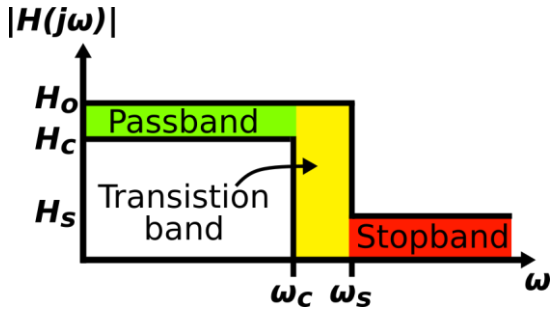


Figure 37: General specifications for a lowpass filter.

General filter theory

An ideal lowpass filter has a transfer function given by:

$$|H(j\omega)| = \begin{cases} H_0 & \text{for } |\omega| \leq \omega_c \\ 0 & \text{for } |\omega| > \omega_c \end{cases} \tag{4.2}$$

where ω_c is the cut-off frequency. This is known as a brick-wall filter, which of course is impossible to implement in a real circuit. A realizable filter will have a transfer function [100]:

$$|H(j\omega)| = \frac{A}{\sqrt{1 + k^2 F(\omega)^2}}, \tag{4.3}$$

where $F(\omega)$ is a polynomial or rational function of order N, and A and k are constants. The higher the order N is, the faster the roll-off can be, which asymptotically can reach a value of up to $20N$ dB/decade. For any lowpass filter, we define a *passband* below ω_c, a *stopband* above ω_s, and a *transition band* between ω_c and ω_s, according to Fig. 37. In the passband, the gain is allowed to vary between H_o and H_c, while in the stopband, the gain must be below H_s. The filter designer must find a transfer function H that fulfils these specifications. The most common approach taken is to pick from a wide range of already established filter functions, each of which has some distinct properties and will always yield a realizable filter. Examples of these are Butterworth, Chebyshev, inverse Chebyshev, Bessel, Pascal, Cauer, and elliptic filters. The most commonly used for multi-GHz lowpass filters in the literature are Butterworth and Chebyshev, which will therefore be the focus of this section. Both Butterworth and Chebyshev are *all-pole* filters, since they do not have any zeros in their transfer functions.

The general transfer function of a Butterworth filter can be written as [100]:

$$|H(j\omega)| = \frac{H_o}{\sqrt{1 + \left(\frac{\omega}{\omega_c}\right)^{2N}}}, \tag{4.4}$$

where ω_c in this case is the 3-dB angular frequency. The Butterworth filter transfer function magnitude is monotonically decaying with frequency and all derivatives of $|H(j\omega)|$ are equal to zero at $\omega = 0$, a property referred to as *maximally flat magnitude*. The poles of the Butterworth filter are given by:

$$s_n = \omega_c e^{j\left(\frac{2n+N-1}{2N}\pi\right)}, \qquad \text{for } n = 1, 2, \ldots, N \tag{4.5}$$

This corresponds to poles placed at equidistant angles around a half-circle in the left half-plane, see Fig. 38a.

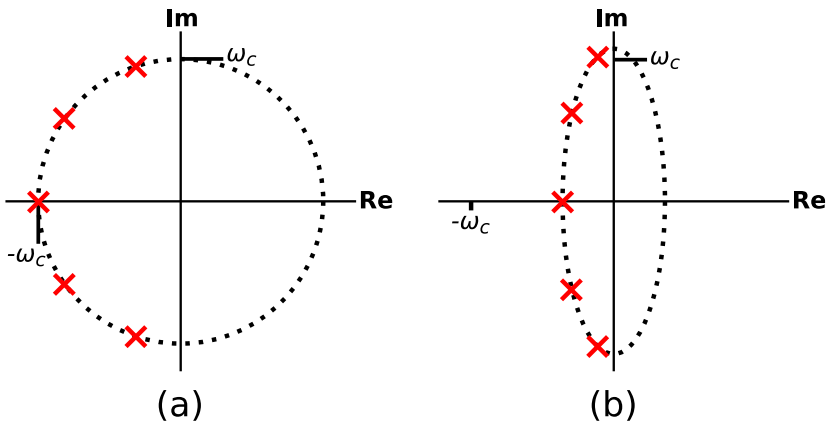


Figure 38: Poles in a 5th-order (a) Butterworth filter, (b) Chebyshev filter.

An n^{th}-order Chebyshev filter has a general transfer function given by [100]:

$$|H(j\omega)| = \frac{H_o}{\sqrt{1 + \varepsilon^2 C_N^2(\omega/\omega_c)}}, \tag{4.6}$$

where ε is known as the ripple factor and $C_N(\omega)$ is the Chebyshev polynomial, given by:

$$C_N(\Omega) = \cos(N \cos^{-1}(\Omega)) \tag{4.7}$$

While this may not appear to be a polynomial, it can be shown that Eq. (4.7) can be rewritten in a recursive manner [100]:

$$C_N(\Omega) = 2\Omega C_{N-1}(\Omega) - C_{N-2}(\Omega) \tag{4.8}$$

Setting $N = 0$ and $N = 1$ in Eq. (4.7) yield $C_0(\Omega) = 1$ and $C_1(\Omega) = \Omega$, respectively, from which Eq. (4.8) can be used to find the polynomials of higher order N. For instance, for $N = 2 - 4$, we get:

$$C_2(\Omega) = 2\Omega^2 - 1$$
$$C_3(\Omega) = 4\Omega^3 - 3\Omega$$
$$C_4(\Omega) = 8\Omega^4 - 8\Omega^2 + 1$$

A Chebyshev filter has a sharper roll-off near ω_c than a Butterworth filter, at the cost of allowing in-band ripple, which will cause the in-band magnitude to vary between H_o and $H_o/\sqrt{1+\varepsilon^2}$. It should be noted that ω_c here is not equal to the 3-dB bandwidth, but it is rather the frequency where $|H(j\omega_c)| = H_o/\sqrt{1+\varepsilon^2}$, which will be below the 3-dB bandwidth when the ripple is less than 3 dB. However, for simplicity and easier comparison, ω_c will be assumed to be equal to $\omega_{-3\mathrm{dB}}$ for the rest of this chapter. The Chebyshev filter also has worse group delay characteristics than a Butterworth filter, i.e. the phase differs more from an ideal linear phase-versus-frequency behavior. The poles of the Chebyshev filter are given by:

$$s_n = \omega_c \sin\frac{(2N+2n-1)\pi}{2N}\sinh\left(\frac{1}{N}\sinh^{-1}\frac{1}{\varepsilon}\right) + j\omega_c \cos\frac{(2N+2n-1)\pi}{2N}\cosh\left(\frac{1}{N}\sinh^{-1}\frac{1}{\varepsilon}\right) \tag{4.9}$$

This results in poles placed on an ellipse, see Fig. 38b.

As we have seen, both Butterworth and Chebyshev filters use complex poles. In fact, all common filter functions have complex poles. This causes peaking in the internal nodes of the filters. The higher the quality factor *Q* of the pole, given by the ratio of the imaginary to the real part, the larger the peaking. A Chebyshev filter will have poles with higher *Q*, and therefore more peaking, than a Butterworth filter of the same order, as can be seen when comparing Fig. 38a and Fig. 38b. This will have an adverse effect on the linearity of an active filter. Additionally, the higher the order of the filter, the higher the *Q* the worst poles will have. Fig. 39 shows an example of the peaking in the various nodes in a 5th-order Chebyshev filter. If gyrators are used to implement the filter, which will be discussed later, each inductor will add an additional node where peaking also can be observed.

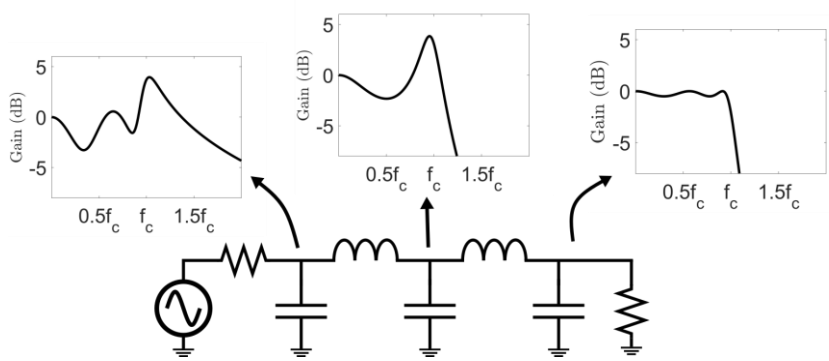


Figure 39: Peaking in the nodes of a 5th-order Chebyshev filter.

The reader might wonder now, why the need for complex poles? By simply cascading *N* single-pole *RC* filters, we could create a filter with only real poles and an asymptotic attenuation slope of $N \cdot 20$ dB/decade, the same as a filter with *N* complex poles. The problem is what happens near the cut-off frequency, which the following example illustrates: Imagine that we want a 5th-order filter with a cut-off frequency f_c. Where should we then place the real poles? If we place them at $-f_c$, then each pole will cause the transfer function to drop by 3 dB at f_c, resulting in $|A(f_c)| = -15$ dB, a much too low value. It turns out that we should place them at $-2.60f_c$ to achieve $|A(f_c)| = -3$ dB. This will instead cause a very flat filter response around f_c, which is also highly undesirable. A filter with complex poles such as a Butterworth, on the other hand, can have a relatively flat passband, while still achieving very sharp attenuation right after the cut-off frequency. Fig. 40 shows the transfer function of a filter with five real poles at $-f_c$, five real poles at $-2.60f_c$, and a 5th-order Butterworth filter with a cut-off frequency f_c. While the three filters asymptotically have the same roll-off (100 dB/decade), the Butterworth filter clearly has significantly sharper roll-off close to the cut-off frequency, while simultaneously having a flatter passband.

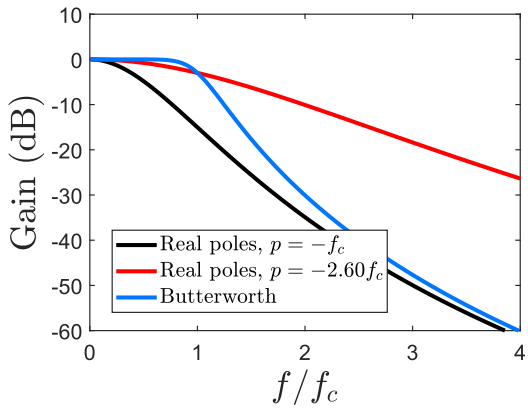


Figure 40: Transfer function of filters with real poles versus a Butterworth filter.

To implement the filters on chip we have two main options; either an active filter, where we use active devices together with capacitors and, if necessary, resistors to place poles in the correct position, or passive filters, in which resistors, capacitors and inductors are used to generate the complex poles.

Active integrated filters

There are multiple active filter topologies, but for multi-GHz filters, mainly three types have been used in the literature: active-RC filters, active inductor filters, and Gm-C filters. These names refer to the main building blocks of each filter.

The active-RC filters use operational amplifiers (op-amps) with RC feedback to generate the desired transfer function, often in the form of a biquadratic equation, more on this later. While the feedback improves the linearity of the filter, it also means that an op-amp with extremely large gain-bandwidth product would be required for a multi-GHz filter, while of course still being stable. This is very difficult to implement, especially in CMOS technology. For this reason, the active-RC low-pass filter with the highest measured cut-off frequency in CMOS, to the author's knowledge, is limited to 1.76 GHz [101]. By using SiGe BiCMOS technologies, where higher g_m/I_D and transition frequencies f_T are achievable, active-RC filters with significantly higher cut-off frequencies have been demonstrated [102] [103] [104].

In active inductor filters, the building blocks incorporate an active structure whose impedance increases linearly with frequency, mimicking an inductor[7]. This inductance can then be used to implement a biquadratic equation. Example of techniques to implement the active inductors in GHz-filtering applications are feedforward [105], feedback [106], and cross-coupling [107] [108].

Gm-C filters, contrary to active-RC filters, are operated in open-loop with transconductors charging and discharging capacitors, forming integrators. The open-loop nature of Gm-C filters makes them very suitable for high frequency applications. As will be shown in the next subsections, the Gm-C blocks can be utilized to generate biquadratic transfer functions, but also to emulate inductors using gyrators, making a direct synthesis from a passive filter possible.

When going from a filter specification to an actual implementation, there are different synthesis methods. In the next two subsections, we will cover two methods known as cascaded biquad synthesis and gyrator synthesis.

Cascaded biquad synthesis

Biquads are a group of circuit blocks that exhibit biquadratic transfer functions given by [53]:

$$T(s) = \frac{a_2 s^2 + a_1 s + a_0}{s^2 + s\left(\frac{\omega_0}{Q}\right) + \omega_0^2} \tag{4.10}$$

Since we are focusing on all-pole lowpass filters, we will assume $a_1 = a_2 = 0$. This will be a two-pole system with natural angular frequency ω_0 and pole quality factor Q, and low-frequency gain a_0/ω_0^2. These parameters can be set by changing the values of the components in the biquad. To generate a filter of order N, one can then

[7] As will be seen shortly, Gm-C filters can also utilize active inductors, in the form of gyrators. However, we make a distinction here between the active inductors, which are formed within the fundamental building block, and a gyrator, which is built from multiple building blocks.

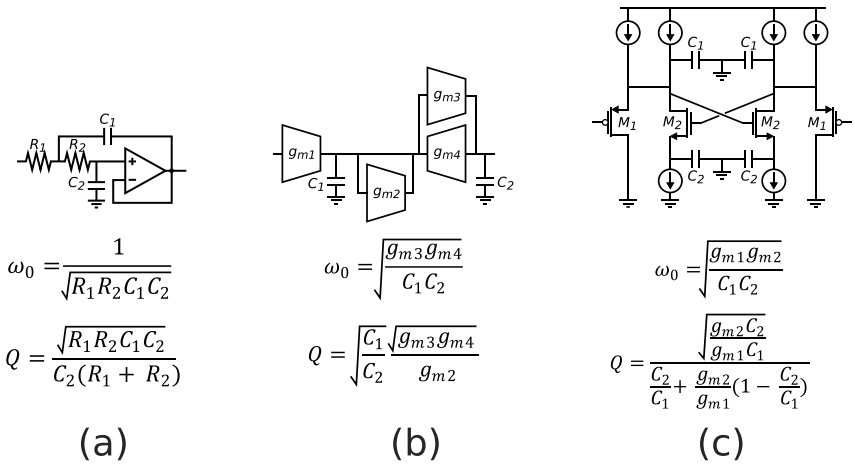


Figure 41: Examples of (a) Sallen-Key biquad, (b) Gm-C biquad, and (c) active inductor biquad.

simply cascade *N*/2 biquads, where each biquad is designed to generate one complex pole-pair, so that the total transfer function gets the desired characteristic.

Given the rather general equation of a biquad, it should not come as a surprise that multiple circuit configurations exist for its implementation, but not all are suitable for very high frequencies. For GHz low-pass filters, the Sallen-Key biquad [101] [102] [103], Multiple-Feedback biquad [104], both examples of active-RC filters, Gm-C biquads [109] [110] [111], and active inductor biquads [105] [106] [107] [108] have all been used. Fig. 41 shows some examples of these, including equations for their ω_0 and Q.

The main issue with biquad filters is their high sensitivity to process variations and mismatch, especially if high-*Q* poles are required [112] [107] [101].

LC ladder synthesis

LC ladder synthesis is based on emulating a passive LC ladder filter with active circuitry. A major benefit of this is that LC ladder filters have a low sensitivity to process variations, a characteristic that also carries over to the active implementation [113]. It starts with an ideal passive filter with resistors, inductors, and capacitors in a ladder configuration that generates the desired filter function. This can be generated using tables [100] or, more commonly nowadays, computer software, such as [114]. Next, each inductor is replaced with a *gyrator*, see Fig. 42a. By applying a test source to the input, the input impedance of this circuit can easily be shown to be:

$$Z_{in} = \frac{v_{in}}{i_{in}} = s\frac{C_{gyr}}{G_m^2} \tag{4.11}$$

Thus, we have emulated an inductor with inductance $L_{gyr} = C_{gyr}/G_m^2$. However, this only emulates a grounded inductor, but a floating inductor is often required. To emulate such an inductor, we extend the gyrator as shown in Fig. 42b. The inductance is still given by $L_{gyr} = C_{gyr}/G_m^2$. Furthermore, most high-speed integrated filters use a differential design, which we can take advantage of in generating the negative transconductance. Fig. 42c. shows how the circuit can be adjusted to emulate a differential floating inductor. For matching purposes, it is also beneficial to replace the resistors with unity-feedback transconductors with transconductance $G_m = 1/R$ [113].

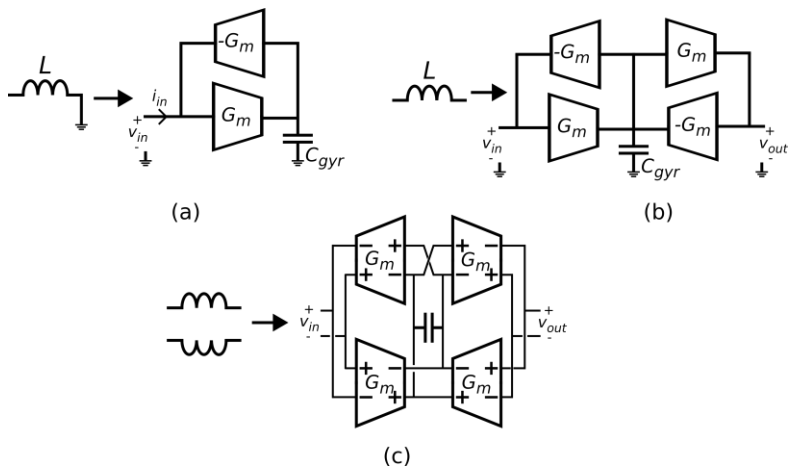


Figure 42: Gyrators to emulate (a) inductor to ground, (b) floating inductor, and (c) differential, floating inductor.

Examples of multi-GHz filters that are designed using LC ladder synthesis can be found in [115] [116] [117], and in **Paper IV**.

Impedance scaling

If all the transconductors in a Gm-C filter with their output connected to a certain node and all capacitors in that node are scaled with a factor β, see Fig. 43a, the noise originating from this node will be scaled by a factor $1/\beta$ [118]. If instead all the transconductors with their *inputs* connected to a certain node and all capacitors in that node are scaled with a factor α, see Fig. 43b, then the voltage-level in that node will be scaled by a factor $1/\alpha$, reducing the distortion originating from that node [118]. This is known as impedance scaling since the impedance level in each node is altered. It is important to note that this does not change the overall transfer function of the filter [100]. Naturally, increasing the transconductance by a factor α or β will increase the power consumption of that transconductor by the same factor.

Since each node will have a unique transfer function to the output, and thus contribute different amounts of noise and distortion, it is reasonable to assume that for a given power budget, there exists an optimum way to scale each node. This is exactly what is addressed in [118], where an algorithm for optimizing the dynamic range of Gm-C filters is derived. In summary, they show that first optimizing the filter for noise performance and then for linearity results in a filter with optimal dynamic range.

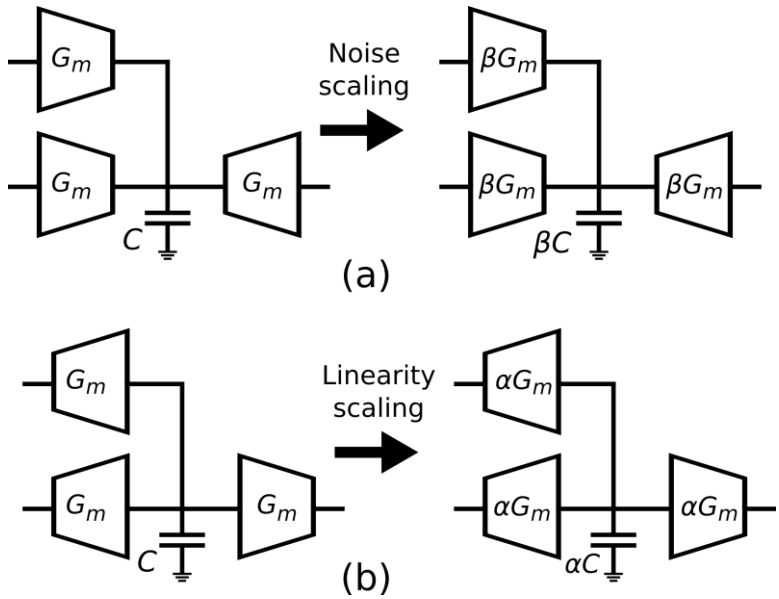


Figure 43: Impedance scaling for (a) improved noise performance, and (b) for improved linearity performance.

While the authors note that this procedure has been derived for Gm-C filters, they claim that it should also be applicable to active-RC filters.

Transconductor design

An ideal transconductor should have infinite DC voltage gain and a transconductance equal to G_m across all frequencies. This yields a perfect integrator in combination with a capacitor, see Fig. 44 (black curve). This is of course not achievable in a real implementation. The DC voltage gain will always be limited by the finite output resistance R_0, and the transconductance will eventually differ from its low-frequency value due to zeros and/or poles, also shown in Fig. 44 (red curve). Still, we should strive to design a transconductor that performs as close to the ideal case as possible.

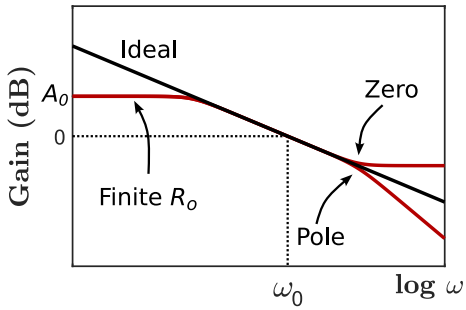


Figure 44: Voltage gain of ideal (black curve) and non-ideal (red curve) Gm-C integrator.

How will these non-idealities then affect the filter transfer function and how well must the real transconductor approximate an ideal transconductor in order for these effects to be negligible? This will depend on the exact filter transfer function, but in general, the higher the Q of the filter poles, the better the transconductor must be [112]. As an example of the transconductor requirements, we simulate a 4.9-GHz, 0.8-dB ripple 5th-order Chebyshev filter, implemented using LC ladder synthesis (Fig. 45), with various imperfections using *Verilog-A* and observe the impact on the filter transfer function. First, the DC voltage gain $A_0 = G_m R_0$ was swept, see Fig. 46a. It can be seen in the figure that we at least require $A_0 > 40$ dB to not cause significant distortion of the transfer function. Achieving this in a modern CMOS technology is very difficult, especially at these high frequencies where short transistors must be used to limit the parasitic capacitances. One way to boost the output impedance is to use cascodes. However, this results in internal nodes, i.e. nodes that are not either the input, output, or supply nodes, which in turn will introduce poles. This brings us to the next simulation, where a left-half-plane (LHP) real-valued pole was introduced and whose absolute value was swept from 50 GHz to 350 GHz, see Fig. 46b. For $|f_p| = 50$ GHz, the transfer function is heavily distorted with excessive peaking, but even for $|f_p| = 350$ GHz, a frequency more than 50 times higher than the cut-off frequency, some peaking can still be observed. This rules out the use of designs with internal nodes. Lastly, the pole was removed and a right-half-plane (RHP) zero was introduced and swept from 50 GHz to 350 GHz, see Fig. 46c. A RHP zero is more likely to show up in the transconductance than a LHP zero, as will be explained later. As can be seen in the figures, an LHP pole and an RHP zero with same magnitude result in an almost identical transfer function. Thus, we can conclude that it is the phase shift due to the zero/pole that is causing the problems, not its effect on the magnitude.

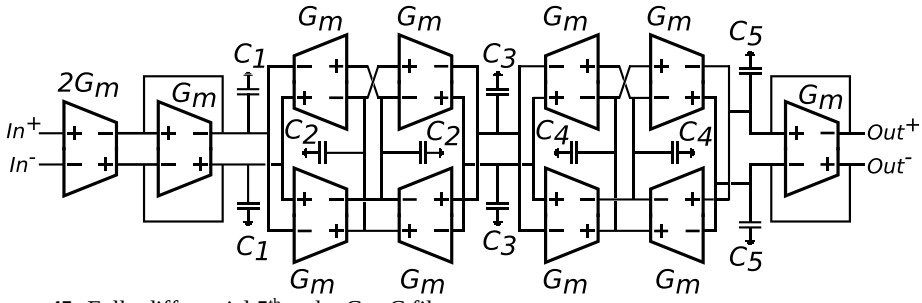


Figure 45: Fully differential 5th-order Gm-C filter.

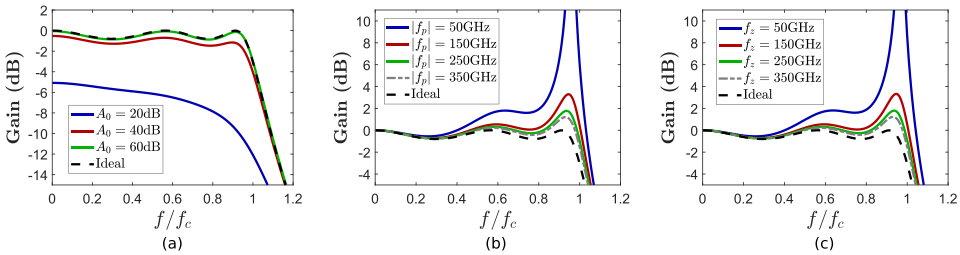


Figure 46: Impact of non-ideal transconductor. (a) Finite R_o. (b) LHP pole. (c) RHP zero.

Differential amplifier

The most straight-forward implementation of a transconductor is to use a regular differential amplifier with active loads, see Fig. 47a, which is done in [109] [111] [117]. While the simplicity of the circuit makes it easy to implement, it comes with some drawbacks. Firstly, the triple-stacked transistors limit the voltage headroom and therefore the linearity. Secondly, in a fully differential application, it requires a separate common-mode feedback. Lastly, for high-frequency applications, the transistor lengths must be kept short to limit the parasitic capacitances. This will cause a relatively low output resistance of the transconductor, which will impact the overall transfer function, as discussed earlier. In [117], they counteract this by adding a negative resistance circuit, implemented with cross-coupled transistors, in parallel with the load, see Fig. 47b, greatly boosting the output resistance.

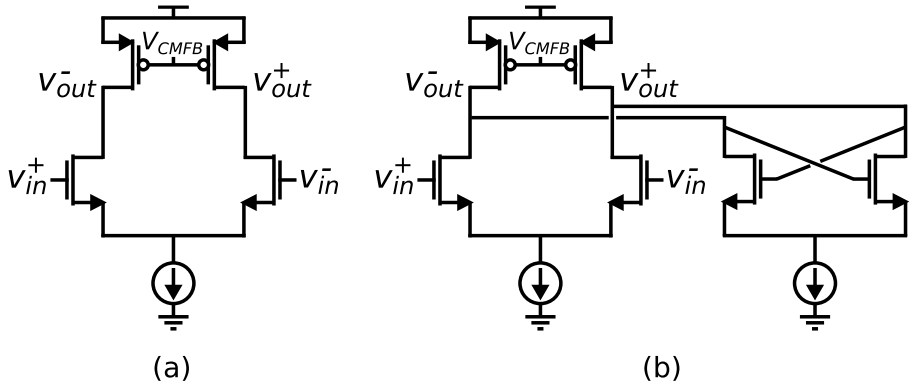


Figure 47: Differential transconductor with active load. (a) Regular implementation. (b) With a negative resistance added.

Nauta transconductor

The most common transconductor implementation for wideband CMOS filters is likely the Nauta transconductor [112], shown in Fig. 48, which is also used in **Paper IV**. The main advantage of this design is that it only contains two signal nodes, the input and output, resulting in that all parasitic capacitances from signal nodes to signal ground can be absorbed in the filter capacitors. So as long as the filter capacitors are large enough, the transconductor will act as a proper integrator to very high frequencies. We still have overlap capacitances between the input and output, but the effect of these will be cancelled in a differential floating gyrator [112]. The other major benefit is that the transconductor can be tuned to achieve a very high differential output resistance. To understand this we analyze the circuit in Fig. 48. For a differential signal, the output conductance will be [112]:

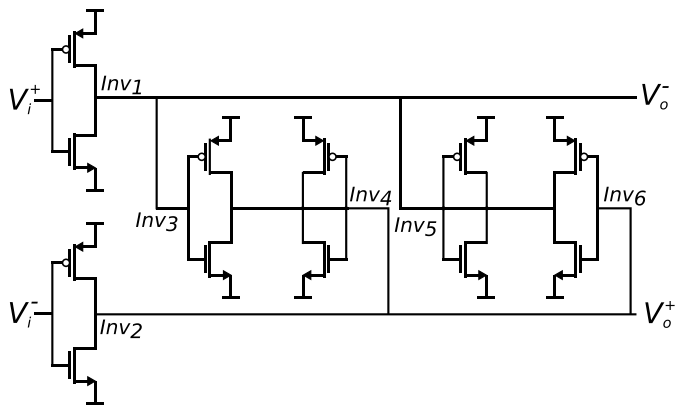


Figure 48: Schematic of the Nauta transconductor.

$$g_{o-} = g_{o1} + g_{o5} + g_{o6} + g_{m5} - g_{m6} \quad (4.12)$$

$$g_{o+} = g_{o2} + g_{o3} + g_{o4} + g_{m4} - g_{m3} \quad (4.13)$$

where g_{mx} and g_{ox} are the transconductance and output conductance, respectively, of inverter Inv_x. Thus, if we tune the circuit so that $g_{o1} + g_{o5} + g_{o6} + g_{m5} = g_{m6}$ and $g_{o2} + g_{o3} + g_{o4} + g_{m4} = g_{m3}$, the differential output conductance becomes very small, yielding an almost perfect transconductor. This tuning can be realized by feeding a separate supply voltage to Inv_4 and Inv_5, which is made slightly lower than the nominal supply voltage, lowering g_{m4} and g_{m5}. This tuning, referred to as *Q* tuning, can be performed automatically by using a master VCO in a feedback loop that controls the supply voltage [113]. Inv_4 and Inv_5 can also be designed with smaller transistor widths, compared to Inv_3 and Inv_6, but this makes the filter susceptible to process variations. In [119], they use this approach and claim that the process technology is so robust that no tuning is necessary, although they do not provide any data to show this. Also, this filter has a cut-off frequency of 450 MHz, which means that larger transistors can be used than in a multi-GHz filter. So even if the claim is true for that particular filter, it is not necessarily true for multi-GHz filters.

In Nauta's initial paper [112], inverters Inv_1, Inv_2, Inv_3, and Inv_6 all had the same dimensions, while Inv_4 and Inv_5 were made slightly narrower. However, Inv_3- Inv_6 can actually be made significantly smaller than Inv_1 and Inv_2, which reduces power consumption of the transconductor and the capacitance in the output nodes. What sets the lower limit for the sizes of Inv_3- Inv_6 is the common-mode stability. It can be shown that to ensure common-mode stability, we must have [115]:

$$g_{m3} + g_{m4} \geq 0.66 g_{m1} \quad (4.14)$$

As previously mentioned, the transconductor will have a very large bandwidth, since it has no internal nodes. What will ultimately limit the bandwidth is non-quasi static (NQS) behavior of the transistors , which will cause a delay between the input voltage and the output current [120]. This delay can be modelled with a RHP zero in the transconductance [113]. This zero is typically situated well above the cut-off frequency of the filter, but as was shown earlier, even a zero that is orders of magnitude higher than the cut-off frequency will still distort the transfer function. The Chebyshev filter in Fig. 45 was implemented with Nauta transconductors and simulated, see Fig. 49a (black curve). Because of this NQS zero, peaking of almost 3 dB occurs near the cut-off frequency, an unacceptable amount. Fortunately, if a resistor is placed in series with the filter capacitors (Fig. 49b), this will cause an LHP zero [113]. If the resistor is sized so that the two zeros have the same frequency magnitude, the phase shift will be completely cancelled, resulting in the desired transfer function, see Fig. 49b (red curve), a technique applied in **Paper IV.**

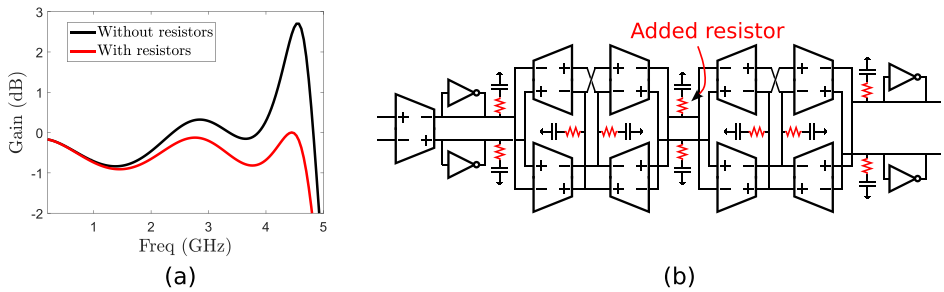


Figure 49: (a) Filter transfer with and without added resistors. (b) Resistors added in series with filter capacitors.

The linearity performance of the transconductor when used in a filter is mainly determined by the linearity of its *V-I* conversion, and the linearity of its output conductance [113]. The former improves with higher overdrive voltage, while the latter worsens with higher overdrive voltage, since that causes the transistors to operate closer to the triode region. This means that there is a threshold voltage that yields optimum linearity, something that can be used to the fully in a fully-depleted silicon-on-insulator (FD-SOI) technology, where the threshold voltage has a stronger dependence on the back-gate voltage than in a regular bulk CMOS process [121]. In addition to this, the back-gate has a larger voltage span in an FD-SOI process than in a bulk process. In [119], the back-gate voltage is tuned to optimize the filter linearity across a large supply voltage range. This technique is also used in **Paper IV**, where we also use the back-gate voltage for both frequency tuning and *Q* tuning. The latter is done by connecting the back-gates of the Inv_4 and Inv_5 transistors to separate back-gate voltages V'_{BBN} and V'_{BBP}, see Fig. 50, instead of the nominal V_{BBN} and V_{BBP}. By making $V'_{BBN} < V_{BBN}$ and $|V'_{BBP}| < |V_{BBP}|$, the same effect as lowering the supply voltage to Inv_4 and Inv_5 is achieved without the need for a separate voltage regulator.

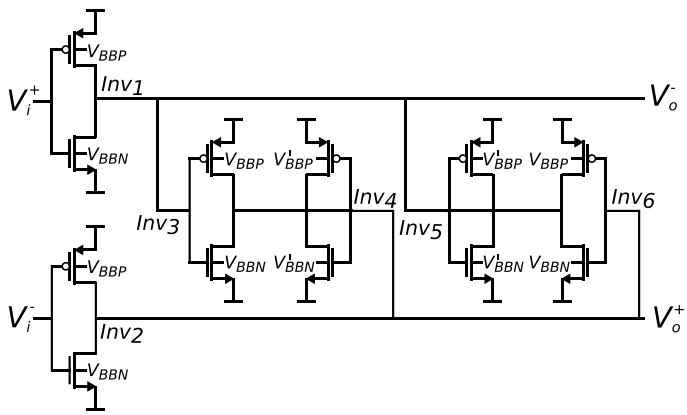


Figure 50: Nauta transconductor with back-gate tuning.

Figure-of-merit

As with most circuit blocks, figure-of-merits (FoMs) are used to compare the performance of different baseband filters. Throughout the years, several FoMs have been proposed. The most basic one only considers the power consumption P_{DC}, number of poles N, and cut-off frequency f_c [115] [107] [108]:

$$FoM_1 = \frac{P_{DC}}{N \cdot f_c} \tag{4.15}$$

This FoM does not consider the dynamic range of the filter, which of course is a very critical property. Eq. (4.15) can be altered as follows to account for this [122] [105]:

$$FoM_2 = \frac{P_{DC}}{N \cdot f_c \cdot SFDR}, \tag{4.16}$$

where $SFDR$ is the spurious-free dynamic range, given by:

$$SFDR = \left(\frac{IIP3}{N_0}\right)^{2/3} \tag{4.17}$$

where N_0 is the integrated in-band noise and $IIP3$ the input-referred third-order intercept point.

For filters that focus on cut-off frequency tuning, the tuning range TR, defined as the ratio between the highest and lowest cut-off frequency, can also be taken into account [108] [123]:

$$FoM_3 = \frac{P_{DC}}{N \cdot f_c \cdot SFDR \cdot TR} \tag{4.18}$$

FoM_2 and FoM_3 only consider the linearity at a single frequency point, and where this point is chosen is rather arbitrary. As explained earlier, the complex poles will cause internal peaking at different frequencies, resulting in vastly different IIP3 depending on at which frequency it is measured, with the worst performance usually being close to the cut-off frequency. To remedy this, in [105], an FoM which takes the frequency where the IIP3 is measured, f_{IIP3}, into consideration is proposed:

$$FoM_4 = \frac{P_{DC}}{N \cdot f_c \cdot SFDR \cdot \frac{f_{IIP3}}{f_c}} = \frac{P_{DC}}{N \cdot SFDR \cdot f_{IIP3}} \tag{4.19}$$

However, this is only appropriate if the IIP3 gets approximately linearly worse with frequency, which is true in [105], but is in general not true. Typically, IIP3 will have a relatively flat characteristic at lower frequencies, and it is not until it gets close to the cut-off frequency that it rapidly deteriorates.

In addition to this, all the presented FoMs scale linearly with filter order. This makes sense from a pure power consumption perspective, but an increase in order

will also reduce the dynamic range, since each stage contributes noise and non-linearity, and causes more peaking due to nodes with higher Q at higher filter order. So, if two filters have similar FoM at the same IIP3 frequency, the filter with the highest order should be considered better.

Another note is that none of proposed FoMs consider out-of-band (OOB) IIP3, only in-band (IB) IIP3, despite that OOB-IIP3 is usually more relevant to filter design for receivers, since the interferers will typically be outside the channel bandwidth. In fact, most published works on multi-GHz filters in the literature do not even present the OOB-IIP3.

Passive integrated filters

Until this point, we have discussed active filters and their shortcomings. Passive filters, on the other hand, consume no power, are noiseless (assuming lossless inductors and capacitors), and are perfectly linear (except if there are nonlinear capacitances or inductances). So why do we not always use passive filters?

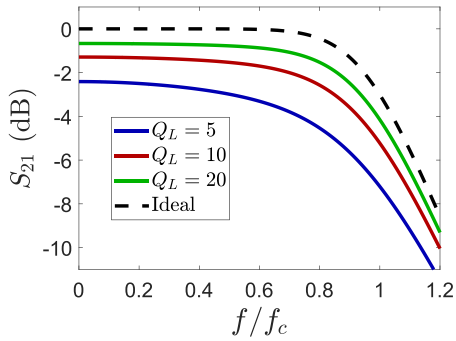


Figure 51: Impact of finite Q_L for a 5th-order 200-MHz Butterworth filter.

Assume that we want to design a 5th-order Butterworth baseband filter for a 5G FR2 transceiver, with a cut-off frequency of 200 MHz and 50-Ω input and output impedances. The inductors would then be 64.4 nH. Not only would implementing such an inductor on-chip occupy a very large chip area, but the quality factor of an integrated inductor, Q_L, will be very limited at such low frequencies. Similar to how the finite output resistance of a transconductor distorts the transfer function, so will a low Q_L, in particular for filters with high-Q poles. Fig. 51 shows a sweep of Q_L for a 200-MHz, 5th-order Butterworth filter. As seen in the figure, a limited Q_L causes both losses at low frequency and a droop in the transfer function. Even for $Q_L = 10$, a completely unrealistic value to achieve at these frequencies, these effects can still be observed to some degree. A passive filter would therefore have to be implemented off-chip, greatly adding to the cost. For this reason, baseband filters have almost exclusively been implemented as active integrated filters in mobile

communication transceivers. It is not until the latest releases of 5G, with channel bandwidths up to 2 GHz, and WiGig/IEEE 802.11ad, with a bandwidth of 2.16 GHz, that examples of integrated passive lowpass filters have been published [124] [125] [126].

If we then consider a multi-GHz baseband filter for sub-THz 6G communication, the situation changes drastically. If we design the Butterworth filter as above, but target a cut-off frequency of 4.7 GHz, each inductor will instead have a value of 2.74 nH, a much more reasonable value to integrate on-chip. The achievable Q_L is also significantly higher at these frequencies. It is therefore reasonable to assume that we will see an increase in popularity of passive baseband filters in the years to come.

Compact differential passive filters

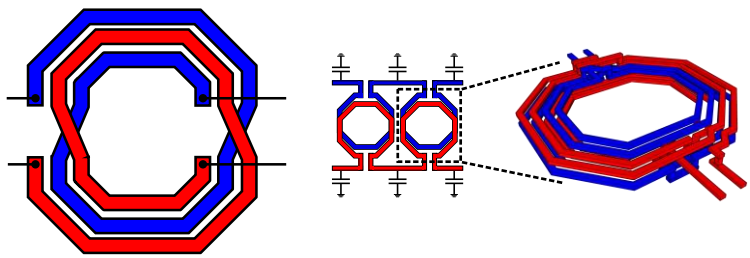


Figure 52: Overlapping inductor used in (a) [124] and (b) **Paper IV.**

Considering the 4.7-GHz Butterworth filter mentioned above, using separate inductors would require a chip area of approximately 0.16 mm^2 for a differential filter. This is still quite a bit larger than the typical area of active filters with similar cut-off frequencies, which ranges approximately from 0.01 mm^2 to 0.10 mm^2. As was pointed out in Chapter 2, the area of the chip is critical in sub-THz antenna arrays, so it should be minimized. To make the footprint smaller, we can, in the case of a differential filter, utilize mutual inductance. This is done in [124], where two 1.5-turn inductors are intertwined, but not overlapping, see Fig. 52a. To make the inductive coupling even stronger, the inductors can be placed so that they completely overlap, see Fig. 52b, which is what we use in **Paper IV**. Furthermore, by doing this, it is easier to use more turns, making the design even more compact. By using this technique, the final footprint of the 4.7-GHz Butterworth filter is reduced to just 0.07 mm^2. The measured transfer function of this filter is shown in Fig. 53a. The filter follows the ideal transfer function well with two exceptions: a droop in the passband and a peaking behavior in the stopband. The former is due to a combination of limited Q_L and so-called skin effect [92], which causes the resistive losses to increase with increasing frequency, and the latter is due to parasitic capacitances in the inductors, as will be discussed more in detail later.

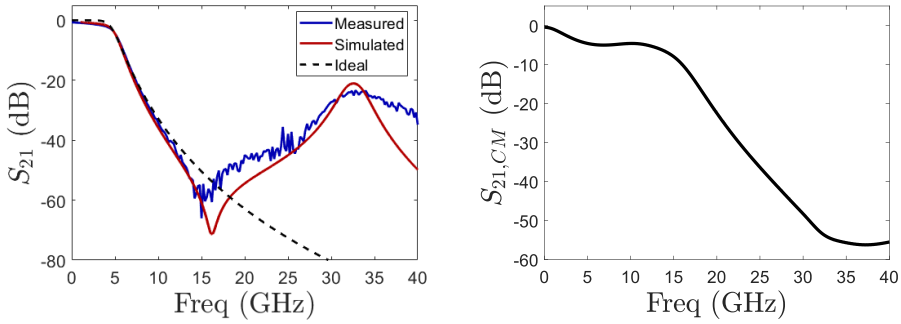


Figure 53: Measured and simulated transfer function of the compact filter in **Paper IV**. (a) Differential mode. (b) Common mode (simulated only).

One drawback of this compact filter is that it relies on differential signals to achieve the correct inductance. For a common-mode signal, the currents will flow in opposite directions in the overlapping metal traces, cancelling most of the magnetic fields and thereby lowering the inductance significantly. Fig. 53b shows the simulated common-mode transfer function, and as seen, no significant filtering occurs below 15 GHz.

Input impedance and input buffer

The filters described so far have been designed for a 50-Ω input and output impedance, which is the typical value if the filter is supposed to interface with off-chip components. However, since this filter is supposed to be on-chip, and interface with other on-chip components, the designer is free to choose impedance. In a current-mode receiver or transmitter, the relatively low input impedance assumed so far is typically not an issue since the stage before the filter can be designed for this. An example of this is seen in [125], where a 3rd-order elliptic filter is placed after a current-mode passive mixer. The equivalent output resistance of the mixer, which also becomes the source resistance for the filter, is only 87 Ω. In a voltage-mode transmitter or receiver, on the other hand, a high input impedance is preferable, since a low input impedance would be very hard to drive for the previous stage. But increasing the input and output impedance also increases the values of the inductances by the same ratio, which in turn means a larger area. Given that the main drawback of passive filters is the area, this is not an attractive solution. The other option is to use an input buffer with high input impedance, as for instance done in [124]. This buffer will of course consume power and limit the dynamic range of the filter. The question is then, can a multi-GHz passive filter with an input buffer still perform better than an active filter with similar cut-off frequency?

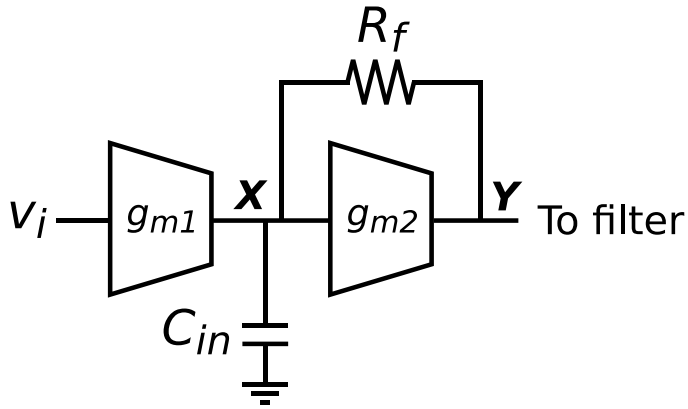


Figure 54: Input buffer for a passive filter.

One way of implementing a high-input, low-output impedance buffer is to use an open-loop transconductance amplifier followed by a resistive shunt feedback stage, see Fig. 54. It is easy to show that the output impedance, neglecting the output resistance of the amplifiers, will be:

$$Z_o = \frac{1 + sC_{in}R_f}{G_{m2} + sC_{in}} \tag{4.20}$$

For low frequencies, this can be approximated as $Z_o \approx 1/G_{m2}$. Thus, we simply scale the second amplifier so that $G_{m2} = 1/50$ S to get the correct 50Ω impedance. By shorting the output, the low frequency transconductance can be calculated as:

$$G_{m,tot} = \frac{i_o}{v_i} = G_{m1}(1 - R_f G_{m2}) \tag{4.21}$$

The unloaded voltage gain is then given by:

$$A_V = G_{m,tot} Z_o \approx \frac{G_{m1}}{G_{m2}}(1 - R_f G_{m2}) \tag{4.22}$$

If we then make $R_f G_{m2} \gg 1$, this simplifies to:

$$A_V \approx -R_f G_{m1} \tag{4.23}$$

Assuming that the source and load resistances are the same in the filter, we get a 0 dB passband magnitude by setting $A_V = -2$.

This buffer was implemented, with transconductors replaced with inverters, and simulated with the compact filter described above in **Paper IV**. The average in-band noise was 2.8 nV/$\sqrt{\text{Hz}}$, worst-case in-band IIP3 3.2 dBV, and the OOB IIP3 varied between -1.9 and 2.6 dBV, while consuming 8.5 mW from a 0.8 V supply. This gives an $FoM_2 = 2.45$ aJ. To the author's knowledge, this outperforms all published multi-GHz active filters by at least an order of magnitude. This is an expected result, since the number of active devices contributing to noise and nonlinearity are much

fewer than for a fully active filter. Additionally, the buffer only has to deal with peaking in nodes *X* and *Y* in Fig. 54, which have less peaking than the nodes inside the filter. In fact, the peaking for the first node in a Butterworth filter does not occur until after the cut-off frequency. This is why the buffered filter has better in-band IIP3 than OOB IIP3.

Since the buffer only has to deal with the peaking in the first filter node (and the peaking it is causing in node *X* due to the feedback through R_f), it is only weakly dependent on the order of the filter. That is, we could add another inductor stage, creating a 7th order filter, and the performance of the buffer would barely change. If we do the same in an active filter, the power consumption, noise, and nonlinearity would all increase, meaning that the case for passive filters in multi-GHz applications is even stronger for high-order filters.

An input buffer can also be beneficial if the filter should be tunable. If we want to tune f_c by a factor x while keeping the source and load resistances constant and retaining the filter characteristic, we must scale all inductances and capacitances by a factor $1/x$. However, as noted in Chapter 3, while tuning capacitances is commonly done, tuning inductances can be very difficult, especially if the overlapping inductor described in the previous section is used. But if we instead scale the source and load resistances by a factor x and the capacitances by $1/x^2$, while keeping the inductance constant, we achieve the same effect. From Eq. (4.20), we observe that the buffer source resistance can be scaled by changing G_{m2}. This can for instance be done by switching in and out unit cell transconductances.

Capacitive cancellation

The behavior of on-chip inductors differs, unfortunately, significantly from an ideal inductance due to parasitics. A common model of an on-chip inductor is shown in Fig. 55a [92]. One of the most important parasitics for filter design is C_p, which models inter-turn capacitance. This capacitance will cause zeros in the stopband, limiting the achievable attenuation.

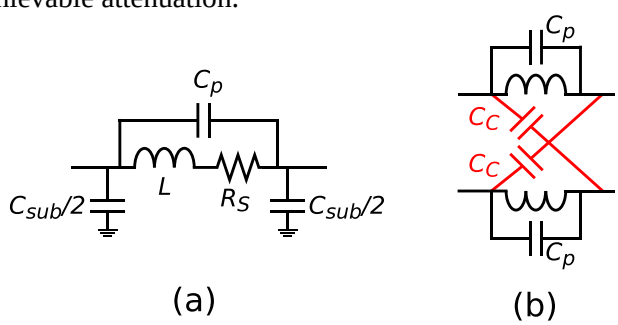


Figure 55: (a) Typical lumped model of an inductor. (b) Capacitive cancellation of C_p.

Fortunately, these capacitances can easily be dealt with in a differential filter [127]. By cross-coupling capacitors C_C as shown in Fig. 55b, we get a differential voltage gain as follows:

$$A(s) = \frac{1 + s^2 L(C_p - C_C)}{1 + s\frac{L}{Z} + s^2 L(C_p + C_C)} \tag{4.24}$$

where Z is the load impedance seen by the inductor. If we set $C_C = C_p$, Eq. (4.24) simplifies to:

$$A(s)|_{C_C=C_p} = \frac{1}{1 + s\frac{L}{Z} + s^2 2LC_p} = \frac{Z}{Z + sL + s^2 2LC_p Z} \tag{4.25}$$

The zero is now gone, and the voltage gain is very similar to the ideal transfer function, only differing by a term $s^2 2LC_p Z$ in the denominator.

The ultra-compact filter presented earlier also suffers from limited stopband attenuation, as seen in Fig. 53a. However, the approach described above does not work here. This is because the dominant parasitic is not the inter-turn capacitance in each inductor, but rather the capacitances between the overlapping inductors. In **Paper V**, we developed a lumped model of the overlapping inductors, which, using nodal analysis, revealed that two zeros and two high-Q poles were responsible for the stop-band peaking, see Fig. 56a. Note that C_p here models the overlapping parasitic capacitance between the sub-inductors, not the inter-turn capacitance. Furthermore, the analysis showed that by placing a capacitor $C_C = 4C_p$ between approximately the midpoint and output of each sub-inductor, the two poles and high-Q poles could be neutralized, also shown in Fig. 56a. Fig. 56b shows the voltage gain of the lumped model and electro-magnetically (EM) simulated overlapping inductor, with and without C_C. Clearly, C_C greatly improves the stopband behavior.

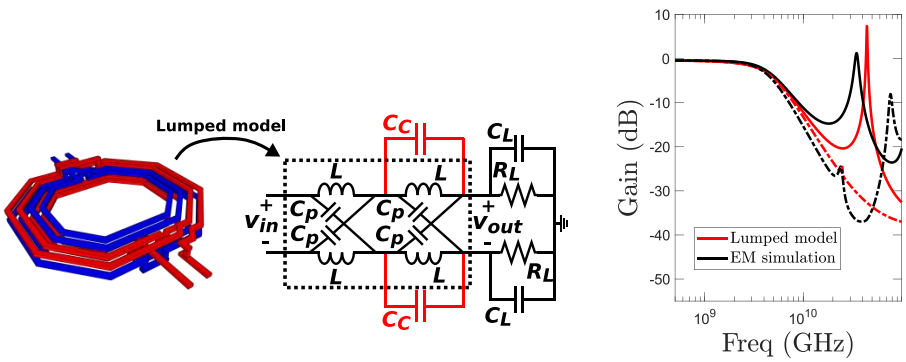


Figure 56: (a) Lumped model with capacitive cancellation. (b) Voltage gain without (solid) and with (dashed) C_C.

However, when the technique was applied to the full filters, the results were not as good as expected. This is due to magnetic coupling between the two inductor pairs. To reduce this, the inductors were redesigned using an 8-shape [128] [129], see Fig. 57a. This greatly improved the transfer function, as seen in Fig. 57b, with an increased attenuation of at least 20 dB between 22 and 36 GHz.

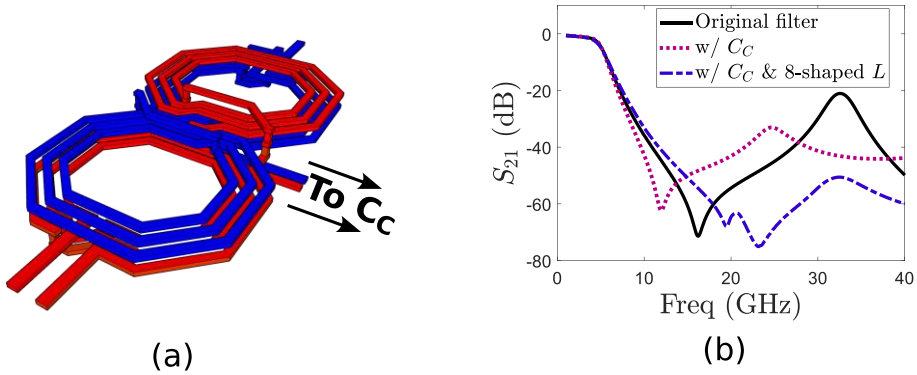


Figure 57: (a) 8-shaped overlapping inductor. (b) EM simulated filter.

Chapter 5

Conclusions and Future Work

The use of mm-wave carrier frequencies, beamforming, and large signal bandwidths in 5G and 6G have created both new challenges and opportunities for circuit designers. This thesis has presented both system-level and circuit-level considerations and solutions for these applications.

The large number of antenna elements used in the beamforming transceiver arrays makes for a very complex system design, with multiple architectural choices to be made. Given the complexity, it can be hard to accurately predict the impact of various circuit non-idealities and how it correlates with the architecture. Therefore, a simulation testbench for large beamforming arrays has been developed in Simulink and MATLAB. The novelty of the testbench is its ability to simulate relatively large systems, while also modelling the non-idealities of individual circuit blocks. It can therefore be used for comparing different system architectures and choices, and also for extracting circuit level specifications on each circuit block for a specific system architecture. Such specifications can for instance be LO phase noise levels, noise figure and linearity of any amplifier and mixer, phase shifter resolution and accuracy, and sharpness and cut-off frequency of baseband filters.

The design of mm-wave frequency generation entails multiple problems, such as tuning circuitry with low quality factor, high sensitivity to PVT variations, and poor quadrature accuracy. For antenna arrays, one also must consider how to distribute the LO signal, since routing mm-wave signals over long distances will cause significant power losses. With these problems in mind, two frequency generation schemes for the 28-GHz band have been designed, both based on injection-locked frequency triplers and with a focus on automatic calibration. One is intended for sliding-IF transceivers, while the other is for direct-conversion. Both utilize phase detectors for their automatic calibration, where in the former, the phase detector measures the applied phase shift, while in the latter, it is the quadrature phase error that is measured. In both cases, the phase detectors can also be used to detect if the circuitry has obtained injection-lock. The automatic calibration makes both circuits very robust with regards to PVT variations, which otherwise is a major problem when using injection-locking, while the use of frequency triplers greatly simplifies the clock distribution in large antenna arrays.

With the anticipated use of multi-GHz signal bandwidths for some 6G implementations, baseband filter design becomes a key challenge. Two integrated baseband filters, one active Gm-C filter and one passive filter, have therefore been designed for such bandwidths. While the active filter achieves state-of-the-art performance among comparable filters, it is greatly outperformed by the passive filter. By utilizing mutual inductance between the differential filter halves, the main drawback of using a passive filter, the large chip area, is minimized. However, capacitive coupling between the overlapping inductors caused stopband peaking in the filter response. To mitigate this, a capacitive cancellation technique was developed, and the inductors were redesigned using an 8-shape to reduce magnetic coupling between them. Given the excellent performance, it is likely that we will see a significant increase in the popularity of integrated passive filters for future multi-GHz bandwidth wireless applications.

Building upon the research presented in this thesis, many topics for future research can be identified:

- The testbench presented in **Paper II** can be used to further investigate the impact of analog non-idealities on large beamforming receivers. For instance, the optimum ADC resolution is a highly debated topic in the literature, which hopefully is a topic the testbench can help shed some light on. Additionally, the testbench can be expanded to also cover transmitters.

- Further investigations of the quadrature frequency tripler in **Paper III**, as exactly what happens when the *I* oscillator is not locked is not fully understood yet. Interestingly, the phase detector outputs a negative voltage both if $3f_{inj} < f_0 - f_L$ and if $3f_{inj} > f_0 + f_L$. This means that the quadrature phase error has the same sign in both cases, which is also confirmed by simulations. Intuitively, one would expect the sign to be different for the two cases, since the injected signal is pulling the *I* oscillator in different directions.

- By combining the techniques in **Paper I** and **Paper III**, a quadrature LO phase shifter suitable for direct-conversion receivers could be developed.

- The proposed capacitive cancellation technique to improve the stopband performance of passive filters in **Paper V** should be tested and verified on silicon. It could also be of interest to place multiple filters in close proximity to each other, to investigate potential effects of magnetic coupling between them when used in a phased array system.

- The input buffer for the passive filter in **Paper IV** should also be tested on silicon to properly compare its performance with active filters. Additionally, if some peaking in the buffer transfer function can be introduced, it could potentially counteract the droop of the passive filter.

References

[1] 3GPP, “3GPP TR 21.917 V17.0.1,” 16 January 2023. [Online]. Available: https://www.etsi.org/deliver/etsi_tr/121900_121999/121917/17.00.01_60/.

[2] M. Latva-aho and K. Leppänen, (Eds.), “Key drivers and reserach challenges for 6G ubiquitous wireless intelligence,” 6G Research Visions, No. 1. University of Oulu, 2019. [Online]. Available: http://urn.fi/urn:isbn:9789526223544.

[3] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjöland and F. Tufvesson, “Towards Power Efficient 6G Sub-THz Transmission,” *6G Wireless Systems: Vision, Requirements, Challenges, Insights, and Opportunities*, vol. 109, no. 7, pp. 1166-1199, Proceedings of the IEEE.

[4] H. Halbauer and T. Wild, “Towards Power Efficient 6G Sub-THz Transmission,” in *2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, Porto, Portugal, 2021.

[5] Ericsson AB, “Leveraging the potential of 5G millimeter wave,” 2021. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/further-insights/leveraging-the-potential-of-5g-millimeter-wave.

[6] H. T. Friis, “A Note on a Simple Transmission Formula,” *Proceedings of the IRE,* vol. 34, no. 5, pp. 254-256, 1946.

[7] A. F. Molisch, Wireless Communications, 2nd ed., John Wiley & Sons Ltd., 2011.

[8] V. Issakov, “The State of the Art in CMOS VCOs: Mm-Wave VCOs in Advanced CMOS Technology Nodes,” *IEEE Microwave Magazine,* vol. 20, no. 12, pp. 59-71, 2019.

[9] S. Andersson, L. Sundström and S. Mattisson, “Design considerations for 5G mm-wave receivers,” in *2017 Fifth International Workshop on Cloud*

Technologies and Energy Efficiency in Mobile Communication Networks (CLEEN), Turin, Italy, 2017.

[10] A. M. Niknejad, S. Thyagarajan, E. Alon, Y. Wang and C. Hull, "A circuit designer's guide to 5G mm-wave," in *2015 IEEE Custom Integrated Circuits Conference (CICC)*, San Jose, USA, 2015.

[11] A. M. Niknejad, D. Chowdhury and J. Chen, "Design of CMOS Power Amplifiers," *IEEE Transactions on Microwave Theory and Techniques,* vol. 60, no. 6, pp. 1784-1796, 2012.

[12] A. Karakuzulu, W. A. Ahmad, D. Kissinger and A. Malignaggi, "A Four-Channel Bidirectional D-Band Phased-Array Transceiver for 200 Gb/s 6G Wireless Communications in a 130-nm BiCMOS Technology," *IEEE Journal of Solid-State Circuits,* vol. 58, no. 5, pp. 1310-1322, 2023.

[13] R. Rotman, M. Tur and L. Yaron, "True Time Delay in Phased Arrays," *Proceedings of the IEEE,* vol. 104, no. 3, pp. 504-518, 2016.

[14] B. Sadhu and L. Rexberg, "Phase Arrays for 5G Millimeter-Wave Communications," in *Millimeter-Wave Circuits for 5G and Radar,* Cambridge, UK, Cambridge University Press, 2019.

[15] R. Hansen, "Proceedings of the IEEE," *Array pattern control and synthesis,* vol. 80, no. 1, pp. 141-151, 1992.

[16] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications,* vol. 24, no. 3, pp. 528-541, 2006.

[17] I. Ahmed, H. Khammari, A. Shahid, A. Musa, K. S. Kim, E. D. Poorter and I. Moerman, "A Survey on Hybrid Beamforming Techniques in 5G: Architecture and System Model Perspectives," *IEEE Communications Surveys & Tutorials,* vol. 20, no. 4, pp. 3060-3097, 2018.

[18] P. Skrimponis, S. Dutta, M. Mezzavilla, S. Rangan, S. H. Mirfarshbafan, C. Studer, J. Buckwalter and M. Rodwell, "Power Consumption Analysis for Mobile MmWave and Sub-THz Receivers," in *2020 2nd 6G Wireless Summit (6G SUMMIT),* Levi, Finland, 2020.

[19] J. Mirzaei, S. ShahbazPanahi, F. Sohrabi and R. Adve, "Hybrid Analog and Digital Beamforming Design for Channel Estimation in Correlated Massive MIMO Systems," *IEEE Transactions on Signal Processing,* vol. 69, pp. 5784-5800, 2021.

[20] C. T. Rodenbeck, M. Martinez, J. B. Beun, J. Silva-Martinez, A. İ. Karşilayan and R. Liechty, “When Less Is More … Few Bit ADCs in RF Systems,” *IEEE Access,* vol. 7, pp. 12035-12046, 2019.

[21] H. Hashemi, X. Guan, A. Komijani and A. Hajimiri, “A 24-GHz SiGe phased-array receiver-LO phase-shifting approach,” *IEEE Transactions on Microwave Theory and Techniques,* vol. 53, no. 2, pp. 614-626, 2005.

[22] S. Mondal and J. Paramesh, “A Reconfigurable 28-/37-GHz MMSE-Adaptive Hybrid-Beamforming Receiver for Carrier Aggregation and Multi-Standard MIMO Communication,” *IEEE Journal of Solid-State Circuits,* vol. 54, no. 5, pp. 1391-1406, 2019.

[23] R. A. Shaheen, R. Akbar, A. Sethi, J. P. Aikio, T. Rahkonen and A. Pärssinen, “A 45nm CMOS SOI, four element phased array receiver supporting two MIMO channels for 5G,” in *2017 IEEE Nordic Circuits and Systems Conference (NORCAS): NORCHIP and International Symposium of System-on-Chip (SoC),* Linköping, Sweden, 2017.

[24] E. Naviasky, L. Iotti, G. LaCaille, B. Nikolić, E. Alon and A. M. Niknejad, “A 71-to-86-GHz 16-Element by 16-Beam Multi-User Beamforming Integrated Receiver Sub-Array for Massive MIMO,” *IEEE Journal of Solid-State Circuits,* vol. 56, no. 12, pp. 3811-3826, 2021.

[25] S. Dutta, C. N. Barati, D. Ramirez, A. Dhananjay, J. F. Buckwalter and S. Rangan, “A Case for Digital Beamforming at mmWave,” *IEEE Transactions on Wireless Communications,* vol. 19, no. 2, pp. 756-770, 2020.

[26] B. Yang, Z. Yu, J. Lan, R. Zhang, J. Zhou and W. Hong, “Digital Beamforming-Based Massive MIMO Transceiver for 5G Millimeter-Wave Communications,” *IEEE Transactions on Microwave Theory and Techniques,* vol. 66, no. 7, pp. 3403-3418, 2018.

[27] R. Lu, C. Weston, D. Weyer, F. Buhler, D. Lambalot and M. P. Flynn, “A 16-Element Fully Integrated 28-GHz Digital RX Beamforming Receiver,” *IEEE Journal of Solid-State Circuits,* vol. 56, no. 5, pp. 1374-1386, 2021.

[28] N. Tawa, T. Kuwabara, Y. Maruta, M. Tanio and T. Kaneko, “28 GHz Downlink Multi-User MIMO Experimental Verification Using 360 Element Digital AAS for 5G Massive MIMO,” in *2018 48th European Microwave Conference (EuMC),* Madrid, Spain, 2018.

[29] 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," *IEEE Journal of Solid-State Circuits,* vol. 52, no. 12, pp. 3373-3391, 2017.

[30] Y. Wang *et al.*, "A 39-GHz 64-Element Phased-Array Transceiver With Built-In Phase and Amplitude Calibrations for Large-Array 5G NR in 65-nm CMOS," *IEEE Journal of Solid-State Circuits,* vol. 55, no. 5, pp. 1249-1269, 2020.

[31] R. Garg and A. S. Natarajan, "A 28-GHz Low-Power Phased-Array Receiver Front-End With 360° RTPS Phase Shift Range," *IEEE Transactions on Microwave Theory and Techniques,* vol. 65, no. 11, pp. 4703-4714, 2017.

[32] K. Khalaf *et al.*, "A 60-GHz 8-Way Phased-Array Front-End With T/R Switching and Calibration-Free Beamsteering in 28-nm CMOS," *IEEE Journal of Solid-State Circuits,* vol. 53, no. 7, pp. 2001-2011, 2018.

[33] Y. Yu, P. G. M. Baltus, A. d. Graauw, E. v. d. Heijden, C. S. Vaucher and A. H. M. v. Roermund, "A 60 GHz Phase Shifter Integrated With LNA and PA in 65 nm CMOS for Phased Array Systems," *IEEE Journal of Solid-State Circuits,* vol. 45, no. 9, pp. 1697-1709, 2010.

[34] K. Kibaroglu, M. Sayginer and G. M. Rebeiz, "A Low-Cost Scalable 32-Element 28-GHz Phased Array Transceiver for 5G Communication Links Based on a 2×2 Beamformer Flip-Chip Unit Cell," *IEEE Journal of Solid-State Circuits,* vol. 53, no. 5, pp. 1260-1274, 2020.

[35] K.-J. Koh and G. M. Rebeiz, "A Q -Band Four-Element Phased-Array Front-End Receiver With Integrated Wilkinson Power Combiners in 0.18- μm SiGe BiCMOS Technology," *IEEE Transactions on Microwave Theory and Techniques,* vol. 56, no. 9, pp. 2046-2053, 2008.

[36] M. Tabesh, J. Chen, C. Marcu, L. Kong, S. Kang, A. M. Niknejad and E. Alon, "A 65 nm CMOS 4-Element Sub-34 mW/Element 60 GHz Phased-Array Transceiver," *IEEE Journal of Solid-State Circuits,* vol. 46, no. 12, pp. 3018-3032, 2011.

[37] G. Mangraviti *et al.*, "13.5 A 4-antenna-path beamforming transceiver for 60GHz multi-Gb/s communication in 28nm CMOS," in *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, San Francisco, USA, 2016.

[38] S. Kundu and J. Paramesh, "A Compact, Supply-Voltage Scalable 45–66 GHz Baseband-Combining CMOS Phased-Array Receiver," *IEEE Journal of Solid-State Circuits,* vol. 50, no. 2, pp. 527-542, 2015.

[39] C. Lu, M. K. Matters-Kammerer, R. Mahmoudi, P. G. Baltus, E. Habekotté, K. v. Hartingsveldt and F. v. d. Wilt, "A 48 GHz 6-bit LO-path phase shifter in 40-nm CMOS for 60 GHz applications," in *2013 Proceedings of the ESSCIRC (ESSCIRC)*, Bucharest, Romania, 2013.

[40] Y. Soliman and R. Mason, "Application of Subharmonic Injection Locking of LC Oscillators to LO-Based Phase-Shifting Phased-Array Architectures," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, no. 12, pp. 3475-3484, 2010.

[41] A. Axholt and H. Sjöland, "A 60 GHz receiver front-end with PLL based phase controlled LO generation for phased-arrays," *Analog Integr Circ Sig Process*, vol. 80, pp. 23-32, 2014.

[42] S. Ek *et al.*, "A 28-nm FD-SOI 115-fs Jitter PLL-Based LO System for 24–30-GHz Sliding-IF 5G Transceivers," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 7, pp. 1988-2000, 2018.

[43] L. Wu, A. Li and H. C. Luong, "A 4-Path 42.8-to-49.5 GHz LO Generation With Automatic Phase Tuning for 60 GHz Phased-Array Receivers," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 10, pp. 2309-2322, 2013.

[44] L. Wu, H. F. Leung, A. Li and H. C. Luong, "A 4-Element 60-GHz CMOS Phased-Array Receiver With Beamforming Calibration," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 64, no. 3, pp. 642-652, 2017.

[45] S. Patnaik and R. Harjani, "A 24-GHz phased-array receiver in 0.13-µm CMOS using an 8-GHz LO," in *2010 IEEE Radio Frequency Integrated Circuits Symposium*, Anaheim, CA, USA, 2010.

[46] N. Ebrahimi, P.-Y. Wu, M. Bagheri and J. F. Buckwalter, "A 71–86-GHz Phased Array Transceiver Using Wideband Injection-Locked Oscillator Phase Shifters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 2, pp. 346-361, 2017.

[47] B. Razavi, "A study of injection locking and pulling in oscillators," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 9, pp. 1415-1424, 2004.

[48] B. Sadhu, X. Gu and A. Valdes-Garcia, "The More (Antennas), the Merrier: A Survey of Silicon-Based mm-Wave Phased Arrays Using Multi-IC Scaling," *IEEE Microwave Magazine*, vol. 20, no. 12, pp. 32-50, 2019.

[49] G. LaCaille, A. Puglielli, E. Alon, B. Nikolic and A. Niknejad, “Optimizing the LO Distribution Architecture of mm-Wave Massive MIMO Receivers,” 2019. [Online]. Available: https://arxiv.org/abs/1911.01339.

[50] M. E. Rasekh, M. Abdelghany, U. Madhowz and M. Rodwell, “Phase noise analysis for mmwave massive MIMO: a design framework for scaling via tiled architectures,” in *2019 53rd Annual Conference on Information Sciences and Systems (CISS)*, Baltimore, USA, 2019.

[51] “RF Blockset - MATLAB,” MathWorks, [Online]. Available: https://www.mathworks.com/products/rf-blockset.html. [Accessed 21 April 2024].

[52] R. A. Shafik, M. S. Rahman and A. R. Islam, “On the Extended Relationships Among EVM, BER and SNR as Performance Metrics,” in *2006 International Conference on Electrical and Computer Engineering*, Dhaka, Bangladesh, 2006.

[53] A. Sedra, K. Smith, T. C. Carusone and V. Gaudet, Microelectronic Circuits, 8th ed., New York City: Oxford University Press, 2021.

[54] B. Razavi, Design of CMOS Phase-Locked Loops, Cambridge: Cambridge University Press, 2020.

[55] D. Leeson, “A simple model of feedback oscillator noise spectrum,” *Proceedings of the IEEE,* vol. 54, no. 2, pp. 329-330, 1966.

[56] A. Hajimiri and T. Lee, “A general theory of phase noise in electrical oscillators,” *IEEE Journal of Solid-State Circuits,* vol. 33, no. 2, pp. 179-194, 1998.

[57] D. Murphy, J. J. Rael and A. A. Abidi, “Phase Noise in LC Oscillators: A Phasor-Based Analysis of a General Result and of Loaded Q,” *IEEE Transactions on Circuits and Systems I: Regular Papers,* vol. 57, no. 6, pp. 1187-1203, 2010.

[58] P. Andreani, X. Wang, L. Vandi and A. Fard, “A study of phase noise in colpitts and LC-tank CMOS oscillators,” *IEEE Journal of Solid-State Circuits,* vol. 40, no. 5, pp. 1107-1118, 2005.

[59] F. Pepe and P. Andreani, “A General Theory of Phase Noise in Transconductor-Based Harmonic Oscillators,” *IEEE Transactions on Circuits and Systems I: Regular Papers,* vol. 64, no. 2, pp. 432-445, 2017.

[60] E. Hegazi, H. Sjöland and A. Abidi, “A filtering technique to lower LC oscillator phase noise,” *IEEE Journal of Solid-State Circuits,* vol. 36, no. 12, pp. 1921-1930, 2001.

[61] B. Razavi, RF Microelectronics, 2nd ed., Upper Saddle River, NJ: Pearson, 2012.

[62] S. Elabd, S. Balasubramanian, Q. Wu, T. Quach, A. Mattamana and W. Khalil, “Analytical and Experimental Study of Wide Tuning Range mm-Wave CMOS LC-VCOs,” *IEEE Transactions on Circuits and Systems I: Regular Papers,* vol. 61, no. 5, pp. 1343-1354, 2014.

[63] H. Sjöland, “Improved switched tuning of differential CMOS VCOs,” *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing,* vol. 49, no. 5, pp. 352-355, 2002.

[64] P. Andreani and S. Mattisson, “On the use of MOS varactors in RF VCOs,” *IEEE Journal of Solid-State Circuits,* vol. 35, no. 6, pp. 905-910, 2000.

[65] E. Hegazi and A. Abidi, “Varactor characteristics, oscillator tuning curves, and AM-FM conversion,” *IEEE Journal of Solid-State Circuits,* vol. 38, no. 6, pp. 1033-1039, 2003.

[66] R. Bunch and S. Raman, “Large-signal analysis of MOS varactors in CMOS -G/sub m/ LC VCOs,” *IEEE Journal of Solid-State Circuits,* vol. 38, no. 8, pp. 1325-1332, 2003.

[67] S. Kulkarni, D. Zhao and P. Reynaert, “Design of an Optimal Layout Polyphase Filter for Millimeter-Wave Quadrature LO Generation,” *IEEE Transactions on Circuits and Systems II: Express Briefs,* vol. 60, no. 4, pp. 202-206, 2013.

[68] T. Zhang, M. Taghivand and J. C. Rudell, “A 55–70GHz two-stage tunable polyphase filter with feedback control for quadrature generation with <2° and <0.32dB phase/amplitude imbalance in 28nm CMOS process,” in *ESSCIRC Conference 2015 - 41st European Solid-State Circuits Conference (ESSCIRC),* Graz, Austria, 2015.

[69] A. Musa, R. Murakami, T. Sato, W. Chaivipas, K. Okada and A. Matsuzawa, “A Low Phase Noise Quadrature Injection Locked Frequency Synthesizer for MM-Wave Applications,” *IEEE Journal of Solid-State Circuits,* vol. 46, no. 11, pp. 2635-2649, 2011.

[70] C. Marcu *et al.*, "A 90 nm CMOS Low-Power 60 GHz Transceiver With Integrated Baseband Circuitry," *IEEE Journal of Solid-State Circuits,* vol. 44, no. 12, pp. 3434-3447, 2009.

[71] M. Frounchi, A. Alizadeh, C. T. Coen and J. D. Cressler, "A Low-Loss Broadband Quadrature Signal Generation Network for High Image Rejection at Millimeter-Wave Frequencies," *IEEE Transactions on Microwave Theory and Techniques,* vol. 66, no. 12, pp. 5336-5346, 2018.

[72] P. K. Khanna, Y. Zhao, M. Forghani and B. Razavi, "A Low-Power 28-GHz Beamforming Receiver with On-Chip LO Synthesis," in *ESSCIRC 2023-IEEE 49th European Solid State Circuits Conference (ESSCIRC)*, Lisbon, Portugal, 2023.

[73] J. Kaukovuori, K. Stadius, J. Ryynanen and K. A. I. Halonen, "Analysis and Design of Passive Polyphase Filters," *IEEE Transactions on Circuits and Systems I: Regular Papers,* vol. 55, no. 10, pp. 3023-3037, 2008.

[74] F. Piri, M. Bassi, N. R. Lacaita, A. Mazzanti and F. Svelto, "A PVT-Tolerant >40-dB IRR, 44% Fractional-Bandwidth Ultra-Wideband mm-Wave Quadrature LO Generator for 5G Networks in 55-nm CMOS," *IEEE Journal of Solid-State Circuits,* vol. 53, no. 12, pp. 3576-3586, 2018.

[75] D. Ozis, J. Paramesh and D. J. Allstot, "Integrated Quadrature Couplers and Their Application in Image-Reject Receivers," *IEEE Journal of Solid-State Circuits,* vol. 44, no. 5, pp. 1464-1476, 2009.

[76] N. Hosseinzadeh and J. F. Buckwalter, "A compact, 37% fractional bandwidth millimeter-wave phase shifter using a wideband lange coupler for 60-GHz and E-band systems," in *2017 IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS),* Miami, USA, 2017.

[77] A. Mirzaei, M. E. Heidari, R. Bagheri, S. Chehrazi and A. A. Abidi, "The Quadrature LC Oscillator: A Complete Portrait Based on Injection Locking," *IEEE Journal of Solid-State Circuits,* vol. 42, no. 9, pp. 1916-1932, 2007.

[78] P. Sagazio, S. Callender, W. Shin, O. Orhan, S. Pellerano and C. Hull, "Architecture and Circuit Choices for 5G Millimeter-Wave Beamforming Transceivers," *IEEE Communications Magazine,* vol. 56, no. 12, pp. 186-192, 2018.

[79] B. Hong and A. Hajimiri, "A Phasor-Based Analysis of Sinusoidal Injection Locking in LC and Ring Oscillators," *IEEE Transactions on Circuits and Systems I: Regular Papers,* vol. 66, no. 1, pp. 355-368, 2019.

[80] S. Sanielevici, K. Cioffi, B. Ahrari, P. Stephenson, D. Skoglund and M. Zargari, “A 900-MHz transceiver chipset for two-way paging applications,” *IEEE Journal of Solid-State Circuits,* vol. 33, no. 12, pp. 2160-2168, 1998.

[81] Hong-Teuk Kim, *et al.* , “A 28-GHz CMOS Direct Conversion Transceiver With Packaged 2×4 Antenna Array for 5G Cellular System,” *IEEE Journal of Solid-State Circuits,* vol. 53, no. 5, pp. 1245-1259, 2018.

[82] J. Pang *et al.*, “A 28-GHz CMOS Phased-Array Transceiver Based on LO Phase-Shifting Architecture With Gain Invariant Phase Tuning for 5G New Radio,” *IEEE Journal of Solid-State Circuits,* vol. 54, no. 5, pp. 1228-1242, 2019.

[83] J. D. Dunworth *et al.*, “A 28GHz Bulk-CMOS dual-polarization phased-array transceiver with 24 channels for 5G user and basestation equipment,” in *2018 IEEE International Solid-State Circuits Conference - (ISSCC)*, San Francisco, USA, 2018.

[84] J. Antes and I. Kallfass, “Performance Estimation for Broadband Multi-Gigabit Millimeter- and Sub-Millimeter-Wave Wireless Communication Links,” *IEEE Transactions on Microwave Theory and Techniques,* vol. 63, no. 10, pp. 3288-3299, 2015.

[85] J. Chen, D. Kuylenstierna, S. E. Gunnarsson, Z. S. He, T. Eriksson, T. Swahn and H. Zirath, “Influence of White LO Noise on Wideband Communication,” *IEEE Transactions on Microwave Theory and Techniques,* vol. 66, no. 7, pp. 3349-3359, 2018.

[86] G. Mangraviti *et al.*, “A 52–66GHz subharmonically injection-locked quadrature oscillator with 10GHz locking range in 40nm LP CMOS,” in *2012 IEEE Radio Frequency Integrated Circuits Symposium*, Montreal, Canada, 2012.

[87] M.-C. Chen and C.-Y. Wu, “Design and Analysis of CMOS Subharmonic Injection-Locked Frequency Triplers,” *IEEE Transactions on Microwave Theory and Techniques,* vol. 56, no. 8, pp. 1869-1878, 2008.

[88] A. Mazzanti, P. Uggetti and F. Svelto, “Analysis and design of injection-locked LC dividers for quadrature generation,” *IEEE Journal of Solid-State Circuits,* vol. 39, no. 9, pp. 1425-1433, 2004.

[89] J.-H. Chang and C.-K. Kim, “A symmetrical 6-GHz fully integrated cascode coupling CMOS LC quadrature VCO,” *IEEE Microwave and Wireless Components Letters,* vol. 15, no. 10, pp. 670-672, 2005.

[90] S.-L. Jang, T.-S. Lee, C.-W. Hsue and C.-W. Chang, “A Low Voltage and Low Power Bottom-Series Coupled Quadrature VCO,” *IEEE Microwave and Wireless Components Letters*, vol. 19, no. 11, pp. 722-724, 2009.

[91] L. Iotti, G. LaCaille and A. M. Niknejad, “A Dual-Injection Technique for mm-Wave Injection-Locked Frequency Multipliers,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 12, pp. 5417-5428, 2021.

[92] H. Darabi, Radio Frequency Integrated Circuits and Systems, Cambridge: Cambridge University Press, 2015.

[93] R. Dumont, M. D. Matos, A. Cathelin and Y. Deval, “A 5G 65-nm PD-SOI CMOS 23.2-to-28.8 GHz Low-Jitter Quadrature-Coupled Injection-Locked Digitally-Controlled Oscillator,” in *2022 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, Denver. USA, 2022.

[94] J. Zhang, H. Liu, C. Zhao and K. Kang, “A 22.8-to-43.2GHz tuning-less injection-locked frequency tripler using injection-current boosting with 76.4% locking range for multiband 5G applications,” in *2018 IEEE International Solid-State Circuits Conference - (ISSCC)*, San Francisco, USA, 2018.

[95] W. L. Chan and J. R. Long, “A 56–65 GHz Injection-Locked Frequency Tripler With Quadrature Outputs in 90-nm CMOS,” *IEEE Journal of Solid-State Circuits*, vol. 43, no. 12, pp. 2739-2746, 2008.

[96] T. Sundstrom, B. Murmann and C. Svensson, “Power Dissipation Bounds for High-Speed Nyquist Analog-to-Digital Converters,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 56, no. 3, pp. 509-518, 2009.

[97] F. Maloberti, Data Converters, Dordrecht, The Netherlands: Springer, 2010.

[98] R. Walden, “Analog-to-digital converter survey and analysis,” *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 4, pp. 539-550, 1999.

[99] B. Murmann, “ADC performance survey 1997-2023,” 2023. [Online]. Available: https://github.com/bmurmann/ADC-survey.

[100] H. G. Dimopoulos, Analog Electronic Filters - Theory, Design and Synthesis, Dordrecht: Springer, 2011.

[101] P. Wambacq, V. Giannini, K. Scheir, W. V. Thillo and Y. Rolain, “A fifth-order 880MHz/1.76GHz active lowpass filter for 60GHz communications in

40nm digital CMOS," in *2010 Proceedings of ESSCIRC*, Seville, Spain, 2010.

[102] F. Centurelli, P. Monsurrò, G. Scotti, P. Tommasino and Trifiletti, "10-GHz Fully Differential Sallen–Key Lowpass Biquad Filters in 55nm SiGe BiCMOS Technology," *Electronics*, vol. 9, no. 4, p. 563, 2020.

[103] C. Bocciarelli, F. Centurelli, P. Monsurrò, V. Spinogatti and A. Trifiletti, "A 17 GHz inductorless low-pass filter based on a quasi-Sallen–Key approach," *Int J Circ Theor Appl*, vol. 51, no. 11, pp. 5066-5084, 2023.

[104] B. Jalali, M. Moretto, A. Singh, S. Shahramian and Y. Baeyens, "An Analog Wide-Bandwidth Baseband Chain for 12Gbps 256QAM Direct-Conversion Receiver," in *2018 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, Philadelphia, USA, 2018.

[105] K. Sohal, D. Manstretta and R. Castello, "A 2nd Order Current-Mode Filter with 14dB Variable Gain and 650MHz to 1GHz Tuning-Range in 28nm CMOS," in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, Daegu, 2021.

[106] H. Xiao and R. Schaumann, "Very-high-frequency lowpass filter based on a CMOS active inductor," in *2002 IEEE International Symposium on Circuits and Systems (ISCAS)*, Phoenix-Scottsdale, AZ, USA, 2002.

[107] F. Centurelli, P. Monsurrò, G. Scotti, P. Tommasino and A. Trifiletti, "A SiGe HBT 6th-Order 10 GHz Inductor-Less Anti-Aliasing Low-Pass Filter for High-Speed ATI Digitizers," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 69, no. 1, pp. 100-113, 2022.

[108] Y. Chen, P.-I. Mak, S. D'Amico, L. Zhang, H. Qian and Y. Wang, "A Single-Branch Third-Order Pole–Zero Low-Pass Filter With 0.014-mm2 Die Size and 0.8-kHz (1.25-nW) to 0.94-GHz (3.99-mW) Bandwidth–Power Scalability," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 60, no. 11, pp. 761-765, 2013.

[109] D. Baranauskas, D. Zelenin, M. Bussmann, S. Elahmadi, J. K. Edwards and C. A. Gill, "A 1.6–3.2-GHz Sixth-Order +13.1-dBm OIP3 Linear Phase gm-C Filter for Fiber-Optic EDC Receivers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, no. 5, pp. 1314-1322, 2010.

[110] T.-Y. Lo and C.-C. Hung, "A 1 GHz OTA-based low-pass filter with a high-speed automatic tuning scheme," in *2007 IEEE Asian Solid-State Circuits Conference*, Jeju, 2007.

[111] V. Szortyka, K. Raczkowski, M. Kuijk and P. Wambacq, “A Wideband Beamforming Lowpass Filter for 60 GHz Phased-Array Receivers,” *IEEE Transactions on Circuits and Systems I: Regular Papers,* vol. 62, no. 9, pp. 2324-2333, 2015.

[112] B. Nauta, “A CMOS transconductance-C filter technique for very high frequencies,” *IEEE Journal of Solid-State Circuits,* vol. 27, no. 2, pp. 142-153, 1992.

[113] B. Nauta, Analog CMOS Filters for Very High Frequencies, Norwell, Massaschuesetts: Kluwer Academic Publishers, 1993.

[114] “LC Filter Design Tool,” Marki Microwave, [Online]. Available: https://markimicrowave.com/technical-resources/tools/lc-filter-design-tool/.

[115] F. Houfaf, M. Egot, A. Kaiser, A. Cathelin and B. Nauta, “A 65nm CMOS 1-to-10GHz tunable continuous-time low-pass filter for high-data-rate communications,” in *2012 IEEE International Solid-State Circuits Conference*, 2012.

[116] Y. Lu, R. Krithivasan, W.-m. L. Kuo, X. Li, J. D. Cressler, H. Gustat and B. Heinemann, “A 70 MHz - 4.1 GHz 5th-Order Elliptic gm-C Low-Pass Filter in Complementary SiGe Technology,” in *2006 Bipolar/BiCMOS Circuits and Technology Meeting*, Maastricht, Netherlands, 2006.

[117] N. Sabatino, G. Minoia, M. Roche, D. Baldi, E. Temporiti and A. Mazzanti, “A 5th order gm-C low-pass filter with ±3% cut-off frequency accuracy and 220MHz to 3.3GHz tuning-range in 28nm LP CMOS,” in *ESSCIRC 2014 - 40th European Solid State Circuits Conference (ESSCIRC)*, Venice Lido, Italy, 2014.

[118] Y. Palaskas and Y. Tsividis, “Dynamic Range Optimization of Weakly Nonlinear, Fully Balanced, Gm-C Filters With Power Dissipation Constraints,” *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing,* vol. 50, no. 10, pp. 714-727, 2003.

[119] J. Lechevallier, R. Struiksma, H. Sherry, A. Cathelin, E. Klumperink and B. Nauta, “5.5 A forward-body-bias tuned 450MHz Gm-C 3rd-order low-pass filter in 28nm UTBB FD-SOI with >1dBVp IIP3 over a 0.7-to-1V supply,” in *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers,* 2015.

[120] Y. Tsividis and C. McAndrew, The MOS Transistor, 3rd ed., New York: Oxford University Press, 2012.

[121] R. Carter *et al.*, “22nm FDSOI technology for emerging mobile, Internet-of-Things, and RF applications,” in *2016 IEEE International Electron Devices Meeting (IEDM)*, San Francisco, 2016.

[122] A. Yoshizawa and Y. P. Tsividis, “Anti-Blocker Design Techniques for MOSFET-C Filters for Direct Conversion Receivers,” *IEEE Journal of Solid State Circuits*, vol. 37, no. 3, pp. 357-364, 2002.

[123] D. Chamla, A. Kaiser, A. Cathelin and D. Belot, “A Gm-C low-pass filter for zero-IF mobile applications with a very wide tuning range,” *IEEE Journal of Solid-State Circuits*, vol. 40, no. 7, pp. 1443-1450, 2005.

[124] H.-C. Park *et al.* , “4.1 A 39GHz-Band CMOS 16-Channel Phased-Array Transceiver IC with a Companion Dual-Stream IF Transceiver IC for 5G NR Base-Station Applications,” in *2020 IEEE International Solid-State Circuits Conference - (ISSCC)*, San Francisco, 2020.

[125] P. Song and H. Hashemi, “mm-Wave Mixer-First Receiver With Selective Passive Wideband Low-Pass Filtering,” *IEEE Journal of Solid-State Circuits*, vol. 56, no. 5, pp. 1454-1463, 2021.

[126] N. Saito *et al.*, “A Fully Integrated 60-GHz CMOS Transceiver Chipset Based on WiGig/IEEE 802.11ad With Built-In Self Calibration for Mobile Usage,” *IEEE Journal of Solid-State Circuits*, vol. 48, no. 12, pp. 3146-3159, 2013.

[127] S. Wang, F. C. Lee and J. D. v. Wyk, “Design of Inductor Winding Capacitance Cancellation for EMI Suppression,” *IEEE Transactions on Power Electronics*, vol. 21, no. 6, pp. 1825-1832, 2006.

[128] T. Mattsson, “Method of and inductor layout for reduced VCO coupling”. United States Patent US7151430B2, 19 December 2006.

[129] P. Andreani, K. Kozmin, P. Sandrup, M. Nilsson and T. Mattsson, “A TX VCO for WCDMA/EDGE in 90 nm RF CMOS,” *IEEE Journal of Solid-State Circuits*, vol. 46, no. 7, pp. 1618-1626, 2011.

Part II

Included papers

Paper I:

An LO phase shifter with frequency tripling and phase detection in 28 nm FD-SOI CMOS for mm-wave 5G transceivers

"An LO phase shifter with frequency tripling and phase detection in 28 nm FD-SOI CMOS for mm-wave 5G transceivers," in *Springer Analog Integrated Circuits and Signal Processing*, vol. 114, pp. 1-11, Jan 2023.

Analog Integrated Circuits and Signal Processing (2023) 114:1–11
https://doi.org/10.1007/s10470-022-02126-8

An LO phase shifter with frequency tripling and phase detection in 28 nm FD-SOI CMOS for mm-wave 5G transceivers

Rikard Gannedahl[1] · **Henrik Sjöland**[1]

Received: 5 June 2022 / Revised: 10 October 2022 / Accepted: 26 December 2022 / Published online: 10 January 2023

Abstract
This paper presents an LO phase shifter with frequency tripling for 28-GHz 5G transceivers. The phase shifting and frequency tripling are achieved using an injection-locked oscillator and injection-locked frequency tripler, respectively. A phase detector based on third harmonic mixing is also implemented and is used to detect the applied phase shift, supporting automatic calibration of the phase shifter. Additionally, an algorithm to automatically tune the oscillators to their respective locking frequency is presented. To test the phase shifter, a 24–30-GHz sliding-IF receiver is implemented. Simulations show that a $> 360°$ tuning range over the full 24–30 GHz span is achieved, with a gain variation of 0.11 dB or less, and that the phase detector has an rms phase error of $< 2.5°$. The circuit is implemented in a 28nm FD-SOI CMOS process and the entire chip measures 1080 µm × 1080 µm, including pads, and consumes 27–29 mW from a 1 V supply.

Keywords Beamforming · mmWave · Phase shifter · Frequency multiplier · Injection locking

1 Introduction

To account for the ever-increasing needs for mobile data, the fifth generation of cellular network technology (5G) has enabled the use of mm-wave frequencies for mobile communication, i.e., frequencies between 30 and 300 GHz. Since the mm-wave spectrum is mostly unoccupied, very large bandwidths can be allocated to each user [1]. For instance, the first commercial bands operating at 24–30 GHz and 37–43.5 GHz offer up to 400 MHz of bandwidth per user, enabling unprecedented data rates in mobile communication [2]. However, the mm-wave communication suffers from significantly higher path loss than its sub-6GHz counterpart [3].

This necessitates the use of antenna arrays, see Fig. 1, in which tens to hundreds of antenna elements are used to focus the transmitted or received power in a certain direction, in a process called beamforming, thus greatly improving the achievable communication distance [4]. The direction of the main beam, or lobe, can be controlled by applying an appropriate phase shift to each antenna element signal. In addition to the main lobe, there will be nulls, where the signal is completely canceled, and sidelobes, see Fig. 1.

The phase shift can either be implemented in the digital domain, referred to as digital beamforming, in the analog domain, referred to as analog beamforming, or in both domains, referred to as hybrid beamforming [5]. Digital beamforming results in the highest system capacity, since all degrees of freedom can be utilized in the channel, but it also requires a full RF chain and a data converter for each antenna element, causing high power consumption [1]. On top of this, it requires extremely fast digital signal processing due to the huge amount of data generated, further increasing the power consumption.

Analog beamforming, on the other hand, only requires a single data converter, significantly reducing the power consumption. However, this means that only a single beam can be created at the time, resulting in a poor utilization of the frequency spectrum resources [5].

Hybrid beamforming has proven to be an efficient middle ground for mm-wave communication, almost reaching the system capacity of digital beamforming, while consuming less power [5, 6]. Thus, analog phase shifters are typically required for mm-wave communication. For simplicity of discussion, the rest of this paper will focus on phase shifting

✉ Rikard Gannedahl
rikard.gannedahl@eit.lth.se

Henrik Sjöland
henrik.sjoland@eit.lth.se

[1] Department of Electrical and Information Technology, Lund University, John Erikssons väg 4, 223 63 Lund, Sweden

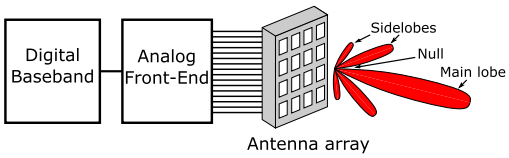


Fig. 1 A beamforming antenna array

in receivers. However, in general, the same concepts apply equally well to transmitters.

The analog phase shift can be applied to the received signal at baseband (BB), at mm-wave frequencies, or to the local oscillator (LO) signal of the mixers [7]. The major benefit of mm-wave phase shifting is that the signals are combined before they are down-converted. This means that only a single mixer is required, making the LO signal routing very simple. However, the phase shifter and combiner must operate at mm-wave frequencies, resulting in high power consumption and/or losses, and any amplitude-variation with phase setting will directly affect the signal. IF/BB phase shifting also suffers from this direct relation between phase-shifter gain and signal amplitude, but is typically easier to implement due to the lower operating frequency. LO phase shifting has, due to the weak relation between mixer gain and LO amplitude, a much lower sensitivity to amplitude-variations of the phase shifter. Both IF/BB and LO phase shifting require one mixer per antenna element, complicating the LO distribution, especially for very large arrays. However, much research has recently focused on combining a large number of smaller integrated circuits (ICs), each with a limited number of antenna elements, in a so-called tiled approach, thereby improving yield and modularity while reducing cost [4]. This means that the penalty of using LO or IF/BB phase shifting in terms of LO distribution becomes less significant compared to mm-wave phase shifting and combining.

An important aspect of an analog phase shifter is phase-amplitude control orthogonality, that is, it should be possible to control the amplitude and phase of each antenna element independently [8]. If the amplitude of each antenna element can be controlled prior to combining the signals, tapering can be used to reduce the amplitude of the sidelobes, at the cost of widening the main beam [9]. However, the sidelobe suppression will be limited by the phase resolution of each phase shifter. In [8], it is shown that for an 8x8 antenna array, the sidelobe suppression will degrade by about 5 dB if the phase shifters have a resolution of 22.5°, compared to using phase shifters with infinite resolution. Phase shifters with 5° resolution, on the other hand, only degrade the sidelobe suppression by about 1 dB. Interestingly, in both cases, the beam direction resolution is less than 1° when non-uniform phase settings are used.

Phase-amplitude control orthogonality is also important for the achievable peak-to-null ratio. If the amplitude varies with varying phase setting, or vice versa, the signals will not perfectly cancel in the null direction. The same is true if the actual phase shift deviates from the desired phase setting. In [10], it is shown that to achieve a 30 dB peak-to-null ratio in a four-element array, the rms phase error and amplitude variation must be less than 2° and ±1.5 dB, respectively.

Due to process, voltage, and temperature variations, this kind of performance is typically only achieved with a time-consuming and costly manual calibration. To speed up that process or even completely circumvent it, several designs with with either built-in self-tests (BIST), automatic calibration schemes or very robust design, requiring little to no calibration, have been proposed [10–15]. Wu et al. [10, 11] achieve excellent phase accuracy and amplitude stability with automatic calibration, but the phase resolution is limited to 22.5°. On the other hand, Inac et al. [12] uses a BIST and achieves a phase resolution of 11.25°, but the rms phase error is about 4°. Yin et al. [13] implements phase shifters with a resolution of 5.6° that only requires calibration at one frequency to cover all of its intended frequencies (23.5–29.5 GHz), but the amplitude variation and phase error is still 1.1 dB and 4.8°, respectively. [14] claims a design robust enough to not require any calibration. However, while the design does achieve an uncalibrated phase resolution and amplitude variation of less than 6.1° and ± 0.8 dB, respectively, the phase error is significant. While not explicitly stated, based on the presented plots it appears to be several degrees. Lastly, [15] achieves an extraordinary rms phase error of 0.08° and rms amplitude error of 0.01 dB after automatic calibration, with a phase resolution of 0.05°. However, their calibration is based on connecting each transmitter output to each receiver input through switches, degrading noise performance and potentially causing cross-talk. For some frequencies, their noise figure (NF) is as high as 11 dB with a gain of − 3 dB, meaning that any circuitry added at the baseband will severely degrade the NF. Additionally, it severely complicates the layout.

In this work, a 24–30 GHz LO phase shifter intended for a hybrid beamforming array is presented, see Fig. 2. The phase shift is accomplished using an injection-locked oscillator (ILO) followed by an injection-locked frequency tripler (ILFT), similar to the work in [10, 11]. A phase detector (PD) is added for built-in measurements of the phase shift and to automatically find the frequency control settings to lock the ILO and ILFT. Additionally, a 28-GHz receiver is implemented to verify the performance of the phase shifter. This paper is an extended version of the work presented in [16].

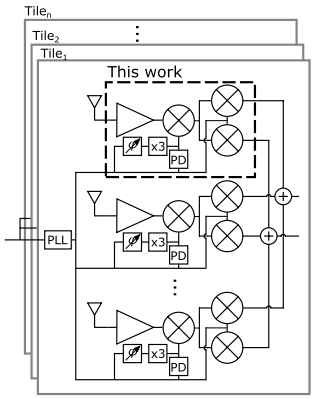


Fig. 2 LO beamforming receiver architecture

2 Injection-locked phase shifter and phase detector

If an oscillator with a free-running oscillation frequency f_0 is injected with a signal at frequency f_{inj}, the oscillator can be forced to oscillate at frequency f_{inj} [17]. The oscillator is then said to be injection-locked. Injection-locking will occur if the difference between f_0 and f_{inj} is smaller than the one-sided locking range f_L, given by [18]:

$$f_L = \frac{f_0}{2Q} \cdot \frac{I_{inj}}{I_{osc}} \cdot \frac{1}{\sqrt{1 - \frac{I_{inj}^2}{I_{osc}^2}}}, \tag{1}$$

where Q is the quality factor of the resonator in the oscillator, and I_{inj} and I_{osc} are the magnitudes of the injected and free-running oscillation currents, respectively.

Since the injection-locked oscillator (ILO) will not oscillate at the resonance frequency of its tank, the ILO output must be phase-shifted relative to the injected signal in order to sustain a 360° phase shift when going through the oscillator loop [18]. This phase shift can be approximated by [18]:

$$\Delta\phi \approx \arcsin\left(\frac{f_0 - f_{inj}}{f_L}\right) \tag{2}$$

Thus, by changing the free-running frequency of the ILO, for instance by using a varactor, this phase shift can electronically be controlled. An issue with this approach is that Eq. 2 is limited to phase shifts of up to ± 90°, while for a phased array, phase shifts of up to ± 180° are required. This can be solved by using a frequency tripler [10, 19], which also triples the phase shift, extending the achievable phase shift to ± 270°. The frequency tripler can also be implemented as an injection-locked oscillator, but with a resonance frequency three times that of the phase-shifting ILO, so that it locks to the third harmonic of the injected signal.

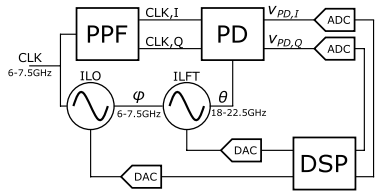


Fig. 3 Architecture of the proposed LO phase shifter

In addition to improving the phase shifting range, using a tripler has two added benefits, both related to that the central PLL seen in Fig. 2 only needs to operate at 1/3 of the final frequency. Firstly, it makes the frequency distribution more power efficient, since lower frequency means less losses [20], thus requiring less buffering. The buffers themselves will also be more power-efficient at lower frequencies. Secondly, the phase noise will improve. This is because the phase noise of ILO will ideally follow the phase noise of the PLL, while the frequency tripler will follow the phase noise of the ILO, multiplied by a factor 3^2, corresponding to an addition of 9.5 dB [10]. Since VCOs, due to poor varactor quality factor, typically have worse figure-of-merit at higher frequencies, a PLL operating at the final frequency would most likely have worse phase noise performance than a PLL operating at one-third of the frequency followed by a frequency tripler.

If a single PLL is used for multiple ILOs and ILFTs, as in Fig 2, the phase noise in each antenna element path will be correlated inside the injection-locked bandwidth of the oscillators. While correlated noise is typically something to be avoided in circuit design, it may actually be an advantage in multi-user beamforming applications [21]. This is because correlated phase noise will affect the phase of each antenna element signal the same, causing the relative phase difference between antenna elements to be unaffected, thus not impacting the shape of the beams and nulls. On the other hand, uncorrelated phase noise will affect the phase of each antenna element signal differently, thus distorting the shape of the beams and nulls.

Figure 3 shows our proposed architecture for the LO phase shifter with frequency tripling and phase detection. It comprises an ILO, an ILFT, a polyphase filter (PPF), a phase detector, two ADCs, two DACs, and a DSP. Note that the converters and DSP are not implemented in this work. A 6–7.5-GHz external clock is injected into the ILO, which, assuming that injection-locking occurs (more on that later), adds a phase shift ϕ. This signal is then injected into the ILFT, which outputs an 18–22.5-GHz signal with a phase

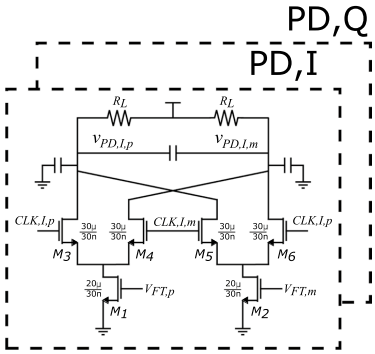


Fig. 4 Schematic of the *I*-part of the phase detector

shift of $\theta = 3\phi + \epsilon$, where ϵ is the phase shift added by the ILFT, relative to the initial clock. The external clock is also injected into a PPF, generating quadrature signals, *CLK*, *I* and *CLK*, *Q*. These quadrature clock signals and the ILFT output are then fed to the phase detector.

The phase detector is implemented as two Gilbert mixers, one for the *I*-signal and one for the *Q*-signal, see Fig. 4. The output from the ILFT is fed to the common source transistors M_1 and M_2, while the output from the PPF is fed to the commutating pairs, M_3–M_6. The commutating pairs are sized so that the current is completely steered from one branch to the other, creating a large third harmonic. This third harmonic mixes with ω_{ILFT} and, assuming that the current in each commutating pair is a perfect square wave and only accounting for the third harmonic, results in an output:

$$\begin{aligned} v_{PD,I} &= g_{m1} R_L V_{FT} \cos(3\omega_{clk} t + \theta) \frac{4}{3\pi} \cos(3\omega_{clk}) \\ &= \frac{2}{3\pi} g_{m1} R_L V_{FT} (\cos(\theta) + \cos(6\omega_{clk} + \theta)), \end{aligned} \tag{3}$$

$$\begin{aligned} v_{PD,Q} &= g_{m1} R_L V_{FT} \cos(3\omega_{clk} t + \theta) \frac{4}{3\pi} \sin(3\omega_{clk}) \\ &= \frac{2}{3\pi} g_{m1} R_L V_{FT} (\sin(\theta) - \sin(6\omega_{clk} + \theta)), \end{aligned} \tag{4}$$

where g_{m1} is the transconductance of M_1 and M_2, R_L is the load resistor, and V_{FT} is the output amplitude of the frequency tripler.

After low-pass filtering the two outputs and converting them to digital signals using the ADCs, the phase can, using simple digital processing, be calculated as:

$$\theta = \begin{cases} \arctan\left(\frac{v_{PD,Q}}{v_{PD,I}}\right), & \text{if } v_{PD,I} \geq 0 \\ \arctan\left(\frac{v_{PD,Q}}{v_{PD,I}}\right) + 180°, & \text{otherwise} \end{cases} \tag{5}$$

Based on this phase measurement, the ILO can then be tuned with a DAC to achieve the desired phase shift without any further calibration, see Fig. 3.

However, this only works if the ILO and ILFT are injection-locked, but the phase detector can also be used to automatically tune the oscillators to achieve lock. The following example, in which the oscillators are tuned to lock to an injected signal with frequency ω_{clk}, illustrates this.

1. Initially, the free-running frequencies of the ILO (ω_{ILO}) and the ILFT (ω_{ILFT}) are tuned to their lowest settings, i.e. the varactor voltages $V_{tune,ILO}$ and $V_{tune,ILFT}$ are set to 0 V.[1] This means that $\omega_{clk} > \omega_{ILO}$ and $3\omega_{clk} > \omega_{ILFT}$, see the top part of Fig. 5(a). This means that in the phase detector, $\omega_{clk,I}$ and $\omega_{clk,Q}$ are mixed with ω_{ILFT}. Since ω_{ILFT} is *not* a harmonic of ω_{clk}, no DC output will be generated in the PD and thus $v_{PD,I} - v_{PD,Q} \approx 0$V, see Fig. 5(a).
2. Next, $V_{tune,ILO}$ is increased, shifting ω_{ILO} up in frequency, see Fig. 5(a). With $V_{tune,ILO} = 0.2$V, locking of ILO has yet to occur, and the PD still outputs close to 0 V.[2] But when $V_{tune,ILO}$ reaches about 0.3V, the ILO locks, see Fig. 5(b). This drastically increases the amplitude of the signal at ω_{clk} that is injected into ILFT, which in turn means that a significant portion of this signal leaks through the ILFT to the PD, creating a DC output when mixed with $\omega_{clk,I}$ and $\omega_{clk,Q}$, see the bottom part of Fig. 5(b). Thus, when $v_{PD,I} - v_{PD,Q}$ starts to diverge significantly from 0V, the ILO is locked.
3. Lastly, the ILFT needs to be tuned. For optimum performance, $3\omega_{clk}$ should be exactly equal to ω_{ILFT}, since this gives the highest amplitude. Combining the low-pass filtered results from (3) and (4) with the Pythagorean identity gives:

$$\sqrt{v_{PD,I}^2 + v_{PD,Q}^2} = \frac{2}{3\pi} g_{m1} R_D V_{FT} \tag{6}$$

That is, $\sqrt{v_{PD,I}^2 + v_{PD,Q}^2}$ will be proportional to V_{FT}, which, as noted above, reaches its peak value when $3\omega_{clk} = \omega_{ILFT}$. Thus, $V_{tune,ILFT}$ is swept until $\sqrt{v_{PD,I}^2 + v_{PD,Q}^2}$ is maximized, see Fig. 5(c).

[1] The final design also incorporates switched capacitors, but for simplicity, only varactors are included in this example.

[2] $v_{PD,I}$–$v_{PD,Q}$ will not be exactly 0 V due to a small portion of the injected signal's fundamental tone leaking through the ILO and the ILFT and mixing with the PPF output.

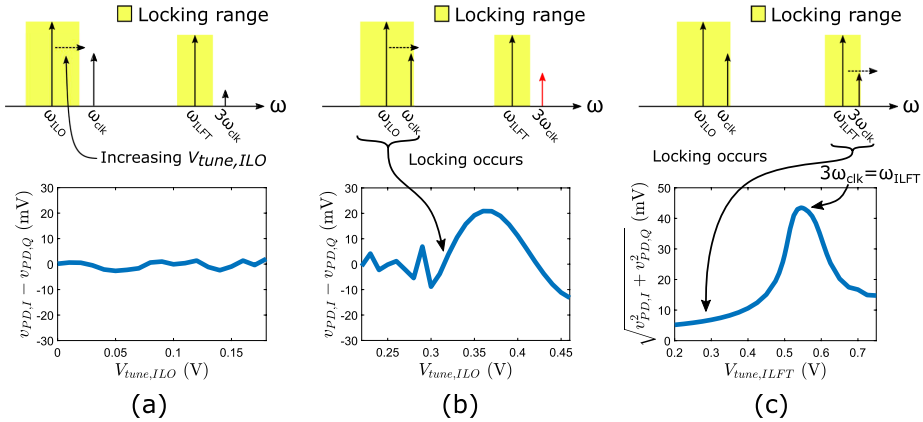


Fig. 5 Scheme for using the PD to detect injection-lock for both the ILO and the ILFT. **a** $V_{tune,ILO}$ starts at 0V and is continuously increased, which increases ω_{ILO}. **b** $V_{tune,ILO}$ reaches a high enough value to achieve lock, causing $v_{PD,I}$ and $v_{PD,Q}$ to diverge. **c** $V_{tune,ILFT}$ is then swept until $\sqrt{v_{PD,I}^2 + v_{PD,Q}^2}$ reaches its peak value

Since the phase can be tuned continuously with a varactor, the phase resolution will only be limited by the resolution of the ILO DAC. The phase detector ADCs will similarly cause a phase error due to the quantization noise. This leads to the question, what are the ADC and DAC resolutions required to make the quantization effects negligible, and what power consumption can be expected for the converters? We start by analyzing the DAC controlling the ILO. Simulations show that the steepest slope for phase vs $V_{tune,ILO}$ is about 1.5°/mV. Since the phase gets tripled in the ILFT, the resulting slope will be 4.5°/mV. Assuming a full-range voltage of 1V, this corresponds to a minimum DAC resolution of 12 bits to achieve about 1° resolution. Since only a DC voltage is required to control the phase, the sampling rate of the DAC only has to be high enough for the beam to track a moving target, which requires phase update intervals on the order of milliseconds [22]. In [23], a 12-bit DAC with a 112 kS/s sample rate is presented. The DAC consumes 50.8 µW, while occupying an area of only 270 µm^2. Thus, the ILO DAC can be implemented with negligible impact on the total power consumption and area of the LO generation circuit, while achieving a worst-case phase resolution of about 1° and more than sufficient phase update rate.

The ILFT DAC needs a resolution of about 1 mV, or 10 bits, to find the optimal tuning voltage of the tripler. The speed of the ILFT DAC is more relaxed than for the ILO DAC, since this voltage will not change with phase setting, and thus only has to be fast enough to counteract any drift in free-running oscillation frequency. Thus, the power and area consumption will be even smaller than for the ILO DAC.

The ADCs will cause phase errors due to quantization noise, which can easily be investigated using MATLAB or similar software, by simply quantizing the ideal $V_{PD,I}$ and $V_{PD,Q}$ signals for phases varying from 0° to 360°. Then, by comparing $\arctan(V_{PD,Q,quant}/V_{PD,Q,quant})$ to the initial phase, the phase error for various ADC resolutions can be found. Figure 6 shows the rms phase error versus number of bits of the ADC. As seen, for resolutions above 9 bits, the rms error will be less than 0.1°, assuming full-swing input to the ADC. The sampling rate should be on the same order as the DAC. As an example of an ADC that can be used, in [24], a 10 MS/s calibration-free ADC with 11 effective number of bits (ENOB) is presented, consuming 0.41 mW and occupying 0.04mm^2. Reducing the rate to kS/s should provide significant power reductions, making the impact of the ADCs on the total power consumption negligible.

Another source of phase error is mismatch in the PPF and between the *I*- and *Q*-part of the phase detector. To investigate this, a 500-samples Monte Carlo simulation was

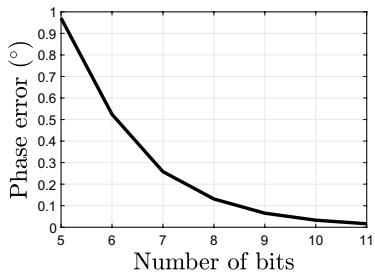


Fig. 6 Rms phase error due to ADC quantization versus number of ADC bits

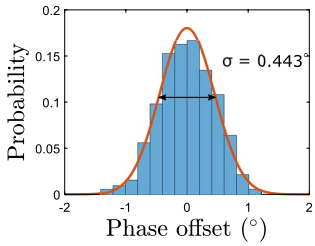


Fig. 7 Phase detector error due to mismatches

performed. Figure 7 shows the resulting histogram of the phase offset. As seen, the rms phase error is about 0.45°. This error is uncorrelated with the phase error due to the ADCs, and will dominate the total random phase error, which will be about 0.5°. There is, however, also a deterministic phase error due to unwanted mixing that will dominate the total phase error of the circuit, as will be seen in Sect. 4.

3 Circuit implementation

The LO phase shifter is implemented in STMicroelectronics' 28 nm FD-SOI CMOS process and uses a 1 V supply. To properly load the phase shifter and test its capabilities, a 24–30 GHz sliding-IF receiver is also implemented. Figure 8 shows the block schematic of the full circuit.

3.1 LO phase shifter

The implemented LO phase shifter consists of an ILO, a peak detector, an ILFT, a polyphase filter followed by digital logic to generate 25% duty cycle pulses, a PD, and a buffer, as seen in Fig. 8. The ILO is implemented as a regular differential cross-coupled LC oscillator, but with two additional

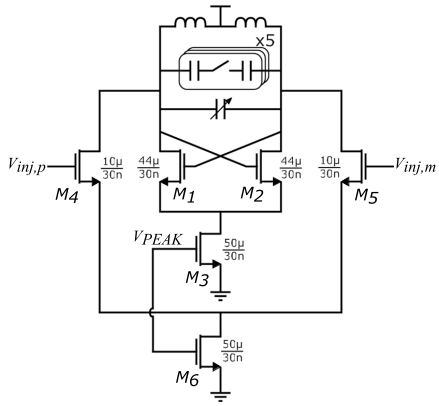


Fig. 9 Schematic of the injection-locked oscillator

injection transistors, see Fig. 9. Using five unary-weighted switched capacitor cells and a varactor, a tuning range between 5.6 and 8.4 GHz is achieved. The varactor is sized so that its tuning capacitance is considerably larger than the capacitance step of one switched capacitor cell. This is done so that a $\pm 60°$ phase shift can be achieved using only the varactor, no matter the frequency. If the gates of the current sources M_3 and M_6 would have been connected to fixed DC voltages, the ILO output amplitude would vary significantly with phase setting, since a large phase shift would correspond to the oscillator operating far from its resonance frequency. This would in turn cause the ILFT amplitude and therefore also the RX gain and noise figure to vary with phase setting. To counteract this, M_3 and M_6 are instead connected to the output of the peak detector, V_{PEAK}, which regulates the ILO amplitude to be almost constant. While it would be enough to only connect M_3 to the peak detector to obtain constant amplitude, the I_{inj}/I_{osc} ratio, and thereby also

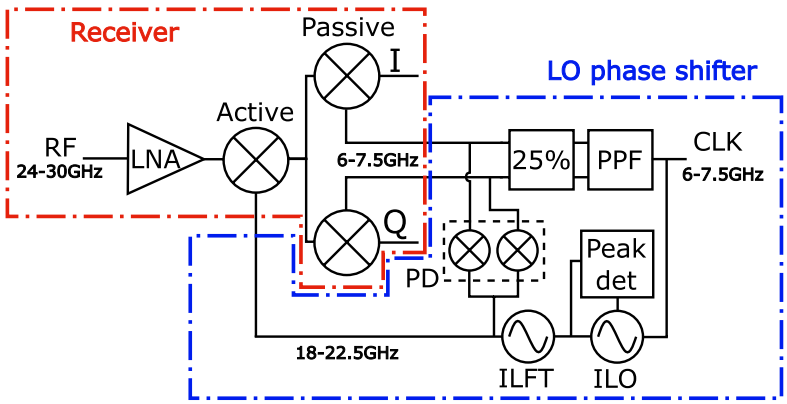


Fig. 8 Block diagram of the full circuit. Differential signals drawn as single lines for improved readability

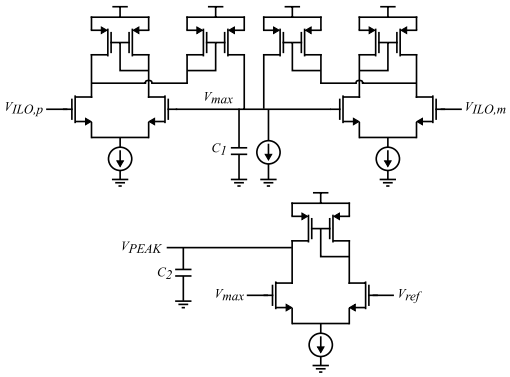


Fig. 10 Schematic of the peak detector

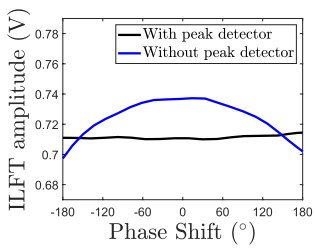


Fig. 11 Impact of peak detector on ILFT amplitude

f_L, is kept approximately constant by also connecting M_6 to V_{PEAK}. Otherwise, there would be a risk of losing injection-lock for certain phase settings. The peak detector is based on the design presented in [25] and is shown in Fig. 10. Figure 11 shows the output amplitude of the ILFT versus phase shift at 21 GHz, with and without the peak detector feedback. The variation in amplitude is reduced from 5.7% when not using the peak detector to 0.7% when using the peak detector. The ILO and peak detector consume between 1.2 and 1.7 mW combined, depending on frequency and phase setting.

The ILFT uses the same architecture as the ILO. However, to maximize the third harmonic current, the injection transistors are biased in weak inversion. Additionally, since the ILFT free-running oscillation frequency is not changed once lock is achieved, there is no need for peak detection feedback, so the current-source transistors are biased with fixed DC gate voltages. A tuning range between 17.4 and 23 GHz is achieved with three unary-weighted switched capacitor cells and a varactor. An inverter-based buffer is also implemented between the ILFT and the receiver. This is done to reduce the capacitive load of the ILFT, and to

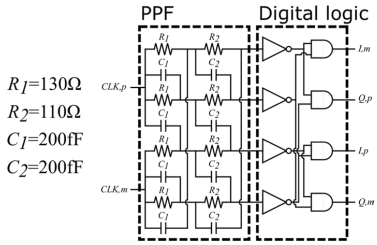


Fig. 12 Schematic of the PPF and digital logic

reduce frequency pulling of the ILFT when a large interferer is present in the RX. The ILFT and buffer combined consume 6–8 mW, depending on frequency.

The phase detector schematic was shown in Fig. 4. The transistors should be as large as possible to minimize the effect of mismatch and flicker noise, but the size must be limited not to load the ILFT too much and to prevent roll-of before 22.5 GHz, the highest output frequency of the ILFT. The total power consumption of both mixers is 1.2 mW.

Lastly, to generate the quadrature signals, a two-stage PPF is used, see Fig. 12. The resistors used are polysilicon resistors, while the capacitors are metal-oxide-metal capacitors. The PPF is followed by a simple digital circuit to generate pulse waves with 25% duty cycle, which consumes 7 mW.

3.2 Receiver

The full schematic of the receiver is shown in Fig. 13. It comprises an on-chip balun, an LNA, an active mixer and quadrature passive mixers. As can be seen, the LNA is a typical inductively source-degenerated cascode LNA, with the load tuned to 28 GHz. The active mixer is double-balanced, implemented as a Gilbert cell, and is driven by the ILFT, with an *LC* load tuned to 7 GHz. The image frequency is situated around 14 GHz, far from the desired signal and will be heavily filtered by the antenna, input matching, and LNA output. Thus, no explicit image filter is required. The LNA and Gilbert mixer combined consumes 10 mW.

The two quadrature mixers are implemented with a double-balanced architecture and are driven by the 25% duty-cycle quadrature clock, providing isolation between the the mixer *I* and *Q* outputs, providing quadrature baseband signals.

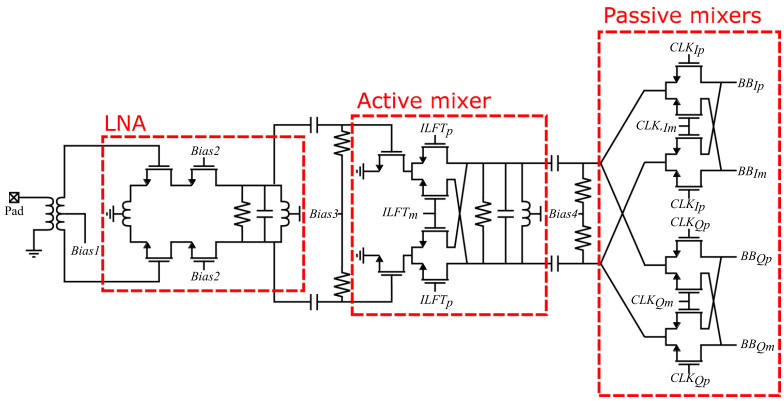


Fig. 13 Schematic of the sliding-IF receiver

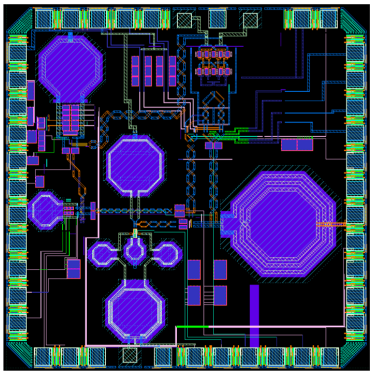


Fig. 14 Layout of the full circuit

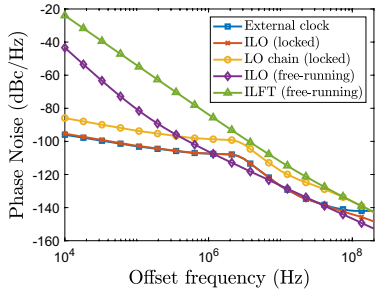


Fig. 15 Phase noise in the LO generation

4 Post-layout simulation results

The following simulations were performed after post-layout parasitic extractions using Cadence QRC. Inductors, transformers and longer interconnects were modeled using Keysight Momentum. The full layout is shown in Fig. 14, and measures 1080 µm × 1080 µm, including pads. The entire IC comsumes 27–29 mW, depending on phase setting and frequency.

To simulate the phase noise of the LO chain, a phase noise profile based on the 7GHz integer-N PLL in [26] was added to the external clock signal. Figure 15 plots the phase noise of the external clock, and the outputs of the ILO and ILFT. Also plotted are the phase noise profiles for ILO and ILFT when free-running. As can be seen, the ILO perfectly follows the external clock for low offset frequencies, and the phase noise of the ILFT is about 9.5 dB above the clock phase noise, matching the theory. However, for large offset frequencies, the ILO and ILFT phase noise profiles start to deviate from the external clock phase noise and instead start to follow the free-running phase noise. This occurs when the frequency offset exceeds the one-sided locking-range. The phase noise at these frequency offsets will then be uncorrelated. However, this should not cause any significant beam distortion due to the low phase noise levels at these offsets. In [18], it is noted that the phase noise performance of an injection-locked oscillator will be worse when operating far from the free-running oscillation frequency due to the change in tank impedance, which is the case when applying a phase shift of ±60°. However, the simulated phase noise varies by less than 1 dB across all phase settings.

Figure 16(a)–(c) show the phase detector measured phase versus the actual BB output phase, at RF input frequencies 24 GHz, 26.8 GHz and 30 GHz, respectively. Also shown is the phase error, i.e. the difference between the detected and the actual phase. For all three frequencies, the phase shifter achieves more than 360° of phase shift range. The rms phase

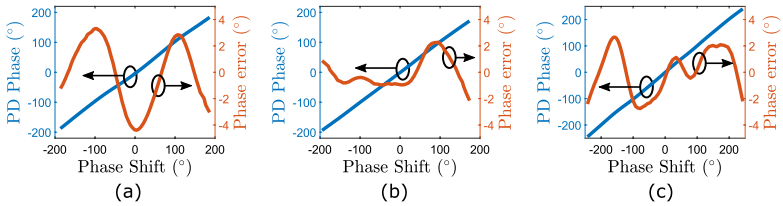


Fig. 16 Detected phase shift and phase error versus actual phase shift at **a** 24 GHz, **b** 26.8 GHz and **c** 30 GHz

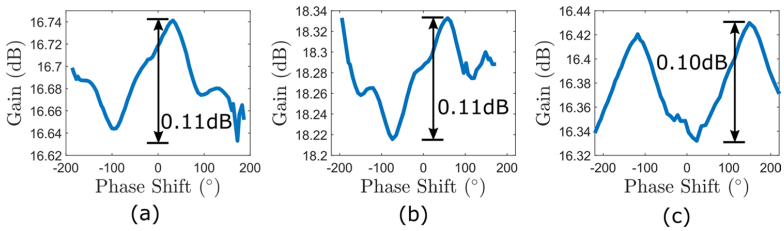


Fig. 17 Receiver gain versus phase shift at **a** 24 GHz, **b** 26.8 GHz and **c** 30 GHz

Table 1 Performance comparison of phase-shifting receivers

	This work[a]	[11]	[12]	[15]	[13][b]
Freq (GHz)	24–30	57–66	8.5–11.5	35–42	23.5–29.5
Phase resolution (°)	Cont.[c]	22.5	11.25	0.05	5.6
Rms phase error (°)	2.4	± 0.6[d]	4	0.08	3.6
Gain variation (dB)	0.11	2.2	2	0.08	0.6
Max gain (dB)	16.3–18.3	16.5–21	9–17	− 10 to 2	19
NF (dB)	4.8–6.1	5.8–8.4	3.5	7–11[e]	5.5[e]
Pwr consump./channel (mW)	27–29	60	36	125	156
Technology	28 nm FD-SOI	65 nm CMOS	130 nm CMOS	65 nm CMOS	180 nm BiCMOS

[a] Simulated data only

[b] Only RF part, no down-conversion

[c] Phase tuning is done with varactor and is therefore limited by the voltage control resolution

[d] Rms value not stated, only maximum variation

[e] Including T/R switch

errors are 2.4°, 1.1°, and 1.7°, respectively. This is not random phase error, but rather deterministic due to unwanted mixing in the phase detector. Part of the ILO signal injected into the ILFT will leak through to the PD and mix with the PPF output. Since these signals are at the same frequency, this will generate a DC component, distorting the desired DC component. Figure 17(a)–(c) show the receiver gain versus phase shift for the same simulations. As can be seen in the plots, the receiver gain varies between 16.3 and 18.3 dB versus input frequency, but the maximum gain variation versus phase shift for a given frequency is only 0.11 dB, proving the usefulness of the peak detector feedback. The noise figure at these frequencies are 5.6 dB, 4.8 dB, and 6.1 dB, respectively, with negligible variation with phase setting.

The performance of the circuit is summarized in Table 1, where it is also compared with other receivers with either automatic phase calibration, built-in self test or limited phase calibration. While this work should not be directly compared to the other works presented in Table 1, given that this work is only simulated, the table still gives an indication of the proposed LO phase shifter's ability to drive a mm-wave receiver with competitve performance.

5 Conclusion

An LO phase shifter for 28-GHz 5G transceivers is presented. It features a frequency tripler and a phase detector based on third harmonic mixing to support automatic phase tuning. An algorithm to automatically detect injection-locking using the output of the phase detector is also presented.

The performance of the phase shifter was tested using a 24–30 GHz sliding-IF receiver. The receiver achieves a gain of 16.3–18.3 dB and a noise figure between 4.8 and 6.1 dB, proving the driving capabilities of the LO circuit. Owing to a peak detector-based feedback in the phase shifter, the gain variation of the receiver is only 0.11 dB across all phase settings. The rms difference between the output phase of the receiver and the phase detector, i.e. the phase error, is 2.4°.

Acknowledgements The authors would like to thank ELLIIT for funding this project. The authors would also like to thank STMicroelectronics for providing access to their technology.

Funding Open access funding provided by Lund University. This work was funded by the Excellence Center at Linköping - Lund in Information Technology (ELLIIT).

Data availability statement The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declaration

Conflict of interest The authors declare that there are no conflict of interest regarding the publication of this paper.

References

1. Xiao, M., Mumtaz, S., Huang, Y., Dai, L., Li, Y., Matthaiou, M., Karagiannidis, G. K., Björnson, E., Yang, K., Chih-Lin, I., & Ghosh, A. (2017). Millimeter wave communications for future mobile networks. *IEEE Journal on Selected Areas in Communications, 35*(9), 1909–1935. https://doi.org/10.1109/JSAC.2017.2719924
2. Leveraging the potential of 5G millimeter wave. Technical report, Ericsson (2021). Retrieved May 28, 2022, from https://www.ericsson.com/490025/assets/local/reports-papers/further-insights/doc/leveraging-the-potential-of-5g-millimeter-wave.pdf.
3. Rappaport, T. S., Sun, S., Mayzus, R., Zhao, H., Azar, Y., Wang, K., Wong, G. N., Schulz, J. K., Samimi, M., & Gutierrez, F. (2013). Millimeter wave mobile communications for 5G cellular: It will work! *IEEE Access, 1*, 335–349. https://doi.org/10.1109/ACCESS.2013.2260813
4. Sadhu, B., Gu, X., & Valdes-Garcia, A. (2019). The more (antennas), the merrier: A survey of silicon-based mm-wave phased arrays using multi-IC scaling. *IEEE Microwave Magazine, 20*(12), 32–50. https://doi.org/10.1109/MMM.2019.2941632
5. Molisch, A. F., Ratnam, V. V., Han, S., Li, Z., Nguyen, S. L. H., Li, L., & Haneda, K. (2017). Hybrid beamforming for massive MIMO: A survey. *IEEE Communications Magazine, 55*(9), 134–141. https://doi.org/10.1109/MCOM.2017.1600400
6. Ahmed, I., Khammari, H., Shahid, A., Musa, A., Kim, K. S., De Poorter, E., & Moerman, I. (2018). A survey on hybrid beamforming techniques in 5G: Architecture and system model perspectives. *IEEE Communications Surveys & Tutorials, 20*(4), 3060–3097. https://doi.org/10.1109/COMST.2018.2843719
7. Hashemi, H., Guan, X., Komijani, A., & Hajimiri, A. (2005). A 24-GHz SiGe phased-array receiver-LO phase-shifting approach. *IEEE Transactions on Microwave Theory and Techniques, 53*(2), 614–626. https://doi.org/10.1109/TMTT.2004.841218
8. Sadhu, B., Tousi, Y., Hallin, J., Sahl, S., Reynolds, S. K., Renström, Ö., Sjögren, K., Haapalahti, O., Mazor, N., Bokinge, B., Weibull, G., Bengtsson, H., Carlinger, A., Westesson, E., Thillberg, J.-E., Rexberg, L., Yeck, M., Gu, X., Ferriss, M., … Valdes-Garcia, A. (2017). A 28-GHz 32-element TRX phased-array IC with concurrent dual-polarized operation and orthogonal phase and gain control for 5G communications. *IEEE Journal of Solid-State Circuits, 52*(12), 3373–3391. https://doi.org/10.1109/JSSC.2017.2766211
9. Haupt, R. (1985). Reducing grating lobes due to subarray amplitude tapering. *IEEE Transactions on Antennas and Propagation, 33*(8), 846–850. https://doi.org/10.1109/TAP.1985.1143682
10. Wu, L., Li, A., & Luong, H. C. (2013). A 4-path 42.8-to-49.5 GHz LO generation with automatic phase tuning for 60 GHz phased-array receivers. *IEEE Journal of Solid-State Circuits, 48*(10), 2309–2322. https://doi.org/10.1109/JSSC.2013.2269855
11. Wu, L., Leung, H. F., Li, A., & Luong, H. C. (2017). A 4-element 60-GHz CMOS phased-array receiver with beamforming calibration. *IEEE Transactions on Circuits and Systems I: Regular Papers, 64*(3), 642–652. https://doi.org/10.1109/TCSI.2016.2612232
12. Inac, O., Shin, D., & Rebeiz, G. M. (2012). A phased array RFIC with built-in self-test capabilities. *IEEE Transactions on Microwave Theory and Techniques, 60*(1), 139–148. https://doi.org/10.1109/TMTT.2011.2170704
13. Yin, Y., Ustundag, B., Kibaroglu, K., Sayginer, M., & Rebeiz, G. M. (2021). Wideband 23.5-29.5-GHz phased arrays for multi-standard 5G applications and carrier aggregation. *IEEE Transactions on Microwave Theory and Techniques, 69*(1), 235–247. https://doi.org/10.1109/TMTT.2020.3024217
14. Khalaf, K., Vaesen, K., Brebels, S., Mangraviti, G., Libois, M., Soens, C., Van Thillo, W., & Wambacq, P. (2018). A 60-GHz 8-way phased-array front-end with T/R switching and calibration-free beamsteering in 28-nm CMOS. *IEEE Journal of Solid-State Circuits, 53*(7), 2001–2011. https://doi.org/10.1109/JSSC.2018.2822676
15. Wang, Y., Wu, R., Pang, J., You, D., Fadila, A. A., Saengchan, R., Fu, X., Matsumoto, D., Nakamura, T., Kubozoe, R., Kawabuchi, M., Liu, B., Zhang, H., Qiu, J., Liu, H., Oshima, N., Motoi, K., Hori, S., Kunihiro, K., … Okada, K. (2020). A 39-GHz 64-element phased-array transceiver with built-in phase and amplitude calibrations for large-array 5G NR in 65-nm CMOS. *IEEE Journal of Solid-State Circuits, 55*(5), 1249–1269. https://doi.org/10.1109/JSSC.2020.2980509
16. Gannedahl, R., & Sjöland, H. (2021). An LO frequency tripler with phase shifter and detector in 28nm FD-SOI CMOS for 28-GHz transceivers. In *2021 IEEE nordic circuits and systems conference (NorCAS)* (pp. 1–7). https://doi.org/10.1109/NorCAS53631.2021.9599854.
17. Adler, R. (1946). A study of locking phenomena in oscillators. *Proceedings of the IRE, 34*(6), 351–357. https://doi.org/10.1109/JRPROC.1946.229930

18. Razavi, B. (2004). A study of injection locking and pulling in oscillators. *IEEE Journal of Solid-State Circuits, 39*(9), 1415–1424. https://doi.org/10.1109/JSSC.2004.831608
19. Axholt, A., & Sjöland, H. (2010). A 90 nm CMOS 14.5 GHz Injection Locked LO generator with digital phase control. In *2010 IEEE MTT-S international microwave symposium* (pp. 1004–1007). https://doi.org/10.1109/MWSYM.2010.5515588.
20. Vecchi, F., Repossi, M., Eyssa, W., Arcioni, P., & Svelto, F. (2009). Design of low-loss transmission lines in scaled CMOS by accurate electromagnetic simulations. *IEEE Journal of Solid-State Circuits, 44*(9), 2605–2615. https://doi.org/10.1109/JSSC.2009.2023277
21. LaCaille, G., Puglielli, A., Alon, E., Nikolic, B., & Niknejad, A. (2019). Optimizing the LO distribution architecture of mm-wave massive MIMO receivers. arXiv:1911.01339.
22. Cudak, M., Kovarik, T., Thomas, T. A., Ghosh, A., Kishiyama, Y., & Nakamura, T. (2014). Experimental mm wave 5G cellular system. In *2014 IEEE globecom workshops (GC Wkshps)* (pp. 377–381). https://doi.org/10.1109/GLOCOMW.2014.7063460.
23. Aiello, O., Crovetti, P., & Alioto, M. (2019). Standard cell-based ultra-compact DACs in 40-nm CMOS. *IEEE Access, 7*, 126479–126488. https://doi.org/10.1109/ACCESS.2019.2938737
24. Hsu, C.-W., Chang, S.-J., Huang, C.-P., Chang, L.-J., Shyu, Y.-T., Hou, C.-H., Tseng, H.-A., Kung, C.-Y., & Hu, H.-J. (2018). A 12-b 40-MS/s calibration-free SAR ADC. *IEEE Transactions on Circuits and Systems I: Regular Papers, 65*(3), 881–890. https://doi.org/10.1109/TCSI.2017.2771364
25. Sanielevici, S. A., Cioffi, K. R., Ahrari, B., Stephenson, P. S., Skoglund, D. L., & Zargari, M. (1998). A 900-MHz transceiver chipset for two-way paging applications. *IEEE Journal of Solid-State Circuits, 33*(12), 2160–2168. https://doi.org/10.1109/4.735700
26. Lee, C. -H., Kabalican, L., Ge, Y., Kwantono, H., Unruh, G., Chambers, M., & Fujimori, I. (2014). A 2.7GHz to 7GHz fractional-N LCPLL utilizing multimetal layer SoC technology in 28nm CMOS. In *2014 symposium on VLSI circuits digest of technical papers* (pp. 1–2). https://doi.org/10.1109/VLSIC.2014.6858390.

Rikard Gannedahl received the M.Sc. in electrical engineering from Lund University, Lund, Sweden, in 2018. He is currently pursuing the Ph.D. degree with the Analog RF Group, Department of Electrical and Information Technology, Lund University. His main research interest is mm-wave CMOS design for 5G and 6G applications, with a focus on receivers and frequency generation. His Master's thesis topic was mm-wave digitally controlled oscillators, and was done at the RF-ASIC Group, Ericsson AB, Lund, Sweden.

Henrik Sjöland received the M.Sc. degree in electrical engineering from Lund University, Sweden, in 1994, and the PhD degree from the same university in 1997. In 1999 he was a postdoc at UCLA on a Fulbright scholarship. He has been an associate professor at Lund University since year 2000, and a full professor since 2008. Since 2002 he is also part time employed at Ericsson Research, where he is currently a Senior Specialist. He has authored or co-authored about 200 international peer reviewed journal and conference papers and holds patents on about 50 different inventions. Henrik Sjöland is currently an associate editor of IEEE Transactions on Circuits and Systems – I, and he has previously been an associate editor of IEEE Transactions on Circuits and Systems – II. He is a Senior Member of IEEE. He has successfully been the main supervisor of 14 PhD students to receive their degrees. His research interests include design of radio frequency, microwave, and mm wave integrated circuits, primarily in CMOS technology.

Paper II:

A Modular System-Level Testbench for 6G Beamforming Applications with Near Circuit-Level Fidelity

Rikard Gannedahl, Javad Bagheri Asli, Henrik Sjöland, Atila Alvandpour,

”A Modular System-Level Testbench for 6G Beamforming Applications with Near Circuit-Level Fidelity,” *in Proc. 21st IEEE Interregional NEWCAS Conference (NEWCAS)*, Edinburgh, United Kingdom, 2023, pp. 1-5.

A Modular System-level Testbench for 6G Beamforming Applications with Near Circuit-Level Fidelity

Rikard Gannedahl[1], Javad Bagheri Asli[2], Henrik Sjöland[1], Atila Alvandpour[2]

[1]Department of Electrical and Information Technology, Lund University, Lund, Sweden
[2]Department of Electrical Engineering, Linköping University, Linköping, Sweden

rikard.gannedahl@eit.lth.se, javad.bagheri.asli@liu.se, henrik.sjoland@eit.lth.se, atila.alvandpour@liu.se

***Abstract*—Sub-THz frequencies are tomorrow's hot research area in mobile communication. However, in this range of frequencies the systems are complex, and it is hard to explore various system architectures and correlate the system-level solutions with circuit-level performances and requirements. This paper presents a scalable testbench in MATLAB/Simulink for sub-THz hybrid beamforming receivers. The testbench models analog and mixed signal blocks with high fidelity, enabling system level simulations with circuit-level imperfections. A receiver with multiple 4-element subarrays is simulated in the testbench, and the impact of phase noise, beam squint, phase shifter inaccuracies, ADC resolution, and more are investigated. Additionally, a Mueller-Müller symbol synchronizer is implemented to achieve symbol-rate sampling.**

Keywords—6G, Sub-THz, Beamforming, System-level testbench.

I. Introduction

The utilization of millimeter-wave frequencies was introduced in 5G, and 6G is expected to continue this trend by pushing the carrier frequency up to the sub-THz range (0.1-1 THz). The higher carrier frequency allows channel bandwidths of 10s of GHz to be used, enabling unprecedented data rates [1]. However, the antenna area is very small at these frequencies, so the received power will be very low. A well-adopted approach to compensate the low received power is using massive antenna arrays. This approach efficiently combines power in space by forming beams, improving link range, and enabling spatial filtering [2]. Fig. 1 illustrates a so-called hybrid beamforming implementation in which the beam is formed in both the analog and digital domains [3]. While the antenna arrays provide many advantages, the system complexity increases, and many essential aspects must be considered when operating at sub-THz frequencies.

In 6G, the fractional bandwidth, i.e., channel bandwidth divided by the carrier frequency, is expected to be significantly larger than in previous generations [1], at least for the lower sub-THz range. Applying frequency-independent phase shifts to the antenna element signals to steer the beam will then result in poor approximations of time delays at the frequency edges of the channel, which will cause uneven beamforming gain across the channel bandwidth, an effect known as beam squint [4]. The beam squint gets worse with increasing array size and beam angle. However, true time delays, free from beam squint effects, can be implemented by signal processing in the digital domain. Therefore, in the case of hybrid beamforming, each analog beamforming subarray should be small enough to cause negligible beam squint. Then

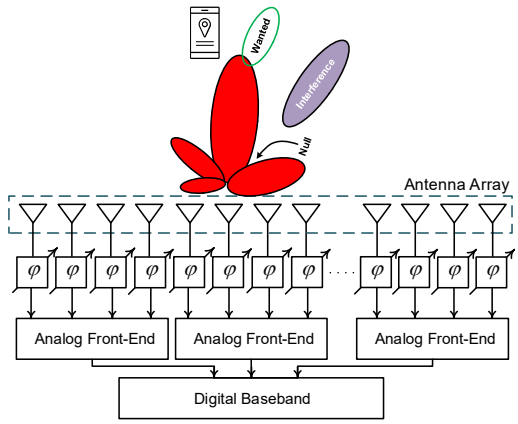


Fig. 1. Hybrid beamforming with small analog sub arrays.

the outputs of the subarrays can be combined with an appropriate time delay in the digital domain (Fig. 1).

Another important aspect for sub-THz phased arrays is the area of each sub-array integrated circuit (IC). To avoid grating lobes [2], the antenna pitch should be $\lambda/2$, where λ is the carrier frequency wavelength. This means that the area per antenna element in the array scales as $1/f^2$. On the other hand, it is difficult to scale down the IC area as much with frequency, which implies that at high frequencies the IC area risks being larger than the antenna array. This would mean that the antenna elements would have to be spaced by more than $\lambda/2$ [2].

When designing a sub-THz beamforming receiver, there are many choices to be made; total number of antennas, number of antennas per sub-array, phased array and receiver architectures, analog-to-digital converter (ADC) resolution and oversampling, etc. All these choices will have an impact on the requirements of each sub-block circuit in the receiver, and vice versa, i.e., limits and non-idealities of the sub-block circuits can greatly impact the system level performance. While in the literature, there are many system level analyses of beamforming receivers [1] [5], they typically consider the receiver as a black box with some non-idealities. In some cases, the fidelity is improved by considering the Figure-of-Merits of important sub-blocks in the receiver to analyze and optimize the power consumption [6] [7]. However, both approaches provide limited insight for a circuit designer on the required specifications of each sub-block, and does not consi-

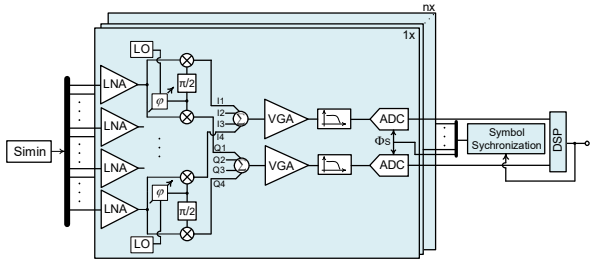


Fig. 2. A block schematic of 4-elemet IC tile.

der the interaction of multiple nonidealities, such as the effect of limited interferer cancellation on ADC resolution.

In this work, we have therefore developed a testbench with near circuit-level fidelity in MATLAB and Simulink for simulating beamforming sub-THz receivers. The testbench can be used by both circuit designers and communication engineers, requiring limited prior knowledge.

Using this testbench, we have simulated an analog sub-array composed of a 4-antenna element, LO phase-shifting, direct-conversion receiver, referred to as 4-element IC tile henceforth. Signals from multiple of this 4-element IC tile are then combined in the digital domain to form a massive hybrid beamforming array. The use of a direct-conversion architecture minimizes the area and as will be seen in Section III, using only four antennas per sub-array causes negligible beam squint.

II. MATLAB/SIMULINK IMPLEMENTATION

The testbench is implemented as follows. A wideband single-carrier signal with a specified power is generated using a random integer sequencer and passed through a raised-cosine transmit filter. Throughout this paper, the signal will be 16-QAM modulated with a 10 GHz bandwidth, and the carrier frequency is 100 GHz. The channel is assumed to be free space. Presently the testbench only considers a 2D environment, so the antenna array is linear. Depending on the angle of arrival (AoA), θ, each antenna element in the array will receive the signal with a different delay. To model this delay of the modulated carrier signal, an FFT is used on the baseband signal and a phase shift is applied according to:

$$\varphi_n[k] = (n-1)\pi \sin\theta \cdot \frac{f[k]-f_c}{f_c}, \qquad (1)$$

where n is the index of the antenna, k is the index of the FFT sub-carrier, f_c is the carrier frequency, and f is a linearly spaced vector stretching from $-f_s/2$ to $f_s/2$, where f_s is the sampling frequency of the FFT. The signal is then transformed back to the time-domain using an IFFT.

The signal is then passed to a Simulink environment, containing the receiver. The analog blocks of the receiver, i.e., all blocks preceding the ADC, are modelled using Simulink's *RF Blockset* library [8]. This library provides both a simulation engine intended for RF simulations, and models of common RF blocks and their non-idealities. When converting the Simulink signal to the *RF Blockset* environment, the baseband signal is up-converted to the carrier frequency, passed through the analog blocks, and then converted back to the Simulink baseband signal environment. Lastly, the signal is digitized by an ADC, which has been implemented in Simulink as a time-interleaved pipelined ADC, where each sub-ADC has a sampling rate of 1GS/s. It considers nonidealities such as multiplying DAC amplifier gain variation and slew rate limitations, comparator reference error, etc. To cover the wide signal bandwidth, 10 sub-ADCs are interleaved.

While in this paper we focus on the 4-element IC tile mentioned in Section I, it should be noted that the testbench is architecture-agnostic, that is, any architecture with an arbitrary number of antenna elements can easily be implemented and simulated using the Simulink GUI.

A. 4-element IC Tile

A block schematic of the 4-element IC tile is shown in Fig. 2. It uses an LO phase-shifting, direct-conversion architecture. The circuit blocks modelled are low-noise amplifiers (LNA), phase shifters, mixers, variable-gain baseband amplifiers (VGA) and filters, local oscillators, and ADCs. Also, a simple digital signal processing (DSP) block, in which the output from each tile is combined, is implemented. In *RF blockset*, the gain, noise figure, *IIP3*, and input and output impedances of each block are modelled. It also models the phase noise of the LO.

B. Symbol Synchronization

Given the extremely large bandwidth, oversampling will be very expensive in terms of power consumption of both the ADC and the DSP. Thus, there should ideally only be one complex sample per symbol, i.e., the sampling rate should be the same as the symbol rate. However, this means that the sampling must be synchronized with the symbols. The most popular symbol synchronization algorithm for single-sample-per-symbol is the Mueller-Müller (MM) algorithm [9]. However, the MM algorithm requires the *SNR* to be higher than a certain threshold to function, which, in general, the output from each tile will not fulfil. The *SNR* will not reach this required level until all signals have been combined in the digital domain. Since the symbol delay to each tile will be different, each ADC will require a different sampling clock phase. However, when steering the beam in a certain direction θ, the difference in sampling time between each tile can be calculated as:

$$\Delta\tau = \frac{d \cdot \sin\theta}{c} \cdot N\Big|_{d=\frac{\lambda}{2}} = \frac{\sin\theta}{2f_c} \cdot N, \qquad (2)$$

where N is the number of elements per tile, d is the antenna element spacing, and c is the speed of light. These delays can be implemented by phase shifting, assuming the clock is sinusoidal, or time delaying the ADC sampling clock signal.

Thus, the symbol time synchronization can be split into two parts; one relative timing between tiles that depends on AoA, and one common timing from MM synchronization block, which together estimate the correct sampling time for the different tiles. This is shown in Fig. 3(a), where the output of the MM block is fed to the first ADC sampling clock, and then a time delay $\Delta\tau$ is added to the clock signal between each ADC.

III. Simulation Results

A. Symbol Synchronization

To verify the operation of the symbol synchronizer, the error vector magnitude (*EVM*) was simulated versus the symbol timing, see Fig. 3(b). This shows the *EVM* of a received signal with and without the MM synchronization block, where 0 and 1 are perfect symbol delay synchronization, and 0.5 is sampling halfway between symbols.

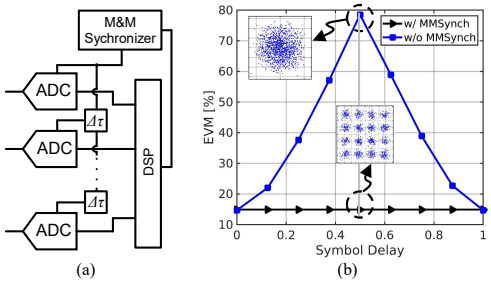


Fig. 3. The MM symbol sychronization algorithm, (a) block diagram, and (b) *EVM* of the received signal *vs*. symbol timming with/without symbol sychronization.

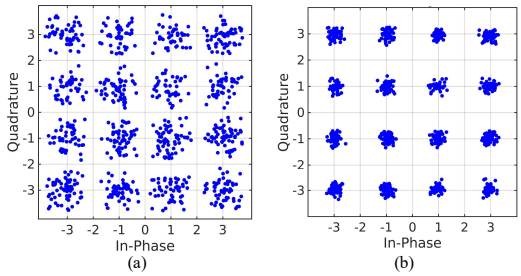


Fig. 4. The constellation diagram for (a) 32-elements tile and (b) eight 4-element tiles, for a signal with a 40° AoA.

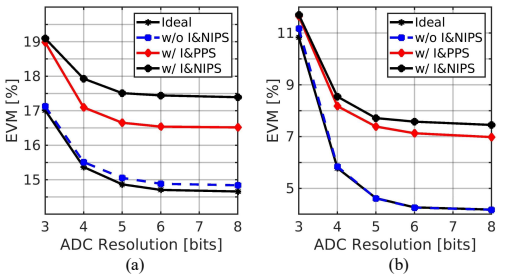


Fig. 5. *EVM vs*. the ADC resolution for input powers of (a) -55 dBm and (b) -40 dBm.

B. Beam Squint

To test the effect of beam squint, a single tile with 32 elements and two ADCs (one for the *I*-path and one for the *Q*-path) is compared to a setup with eight 4-element tiles. Fig. 4 shows the constellation diagram for (a) the 1×32 setup and (b) the 8×4 setup for a signal with a 40° AoA. Clearly, the beam squint causes very large distortion in the 1×32 setup, while it has a negligible effect on the 8×4 setup.

C. ADC Resolution

A heavily debated subject in the literature is the ADC resolution in beamforming receivers [7] [10] [11]. Since the output of multiple ADCs are combined, the requirements on each ADC are relaxed, compared with a single channel implementation [10]. However, the ADC still needs to have a large enough dynamic range to account for interferers, which can even be in-band, from other directions than the desired signal. If the direction of an interfering beam is known, a null can be placed in its direction by each tile in the phased array. However, due to beam squint and phase shifter error and limited resolution, the interferer can only be partially cancelled, and a larger dynamic range and thereby higher ADC resolution is expected to be required. To investigate this, the *EVM* of a receiver with six 4-element tiles versus ADC resolution is simulated for two levels of input power, -55 dBm and -40 dBm. For both input power levels, four scenarios were considered: ideal phase shifters without interferer ("Ideal"), non-ideal phase shifters without interferer ("w/o I&NIPS"), perfect phase shifters with interferer ("w/ I&PPS"), and non-ideal phase shifters with interferer ("w/ I&NIPS"). The results are shown in Fig. 5. The desired signal has an AoA of 15º, while the interferer has an AoA of -20º. The interferer is an in-band, 10-GHz, 16-QAM signal and is 20 dB stronger than the desired signal. Non-ideal phase shifters have a resolution of 5 bits and an rms error of 2°. The VGA in each scenario was set so that the peak signal level was about 4 dB below the ADC full-scale.

As shown in Fig. 5(a), for an input power of -55 dBm, the receiver is limited by noise to an *EVM* of about 15%. This is approximately the required *EVM* for 16-QAM demodulation [5], so this input power corresponds to the sensitivity of the receiver. In the ideal case, the optimum ADC resolution appears to be between 5 and 6 bits, beyond that there would be diminishing returns. Using non-ideal phase shifters without an interferer present gives a similar result. This is expected since non-ideal phase shifters will have a limited impact on the main lobe. On the other hand, when an interferer is present, the *EVM* deteriorates noticeably, especially when using a non-ideal phase shifter. However, the optimum ADC resolution still appears to be between 5 and 6 bits. For an input power of -40 dBm (Fig. 5(b)), the results are similar, except a lower *EVM* is reachable. However, to reach the floor of the *EVM*, a slightly higher ADC resolution, about 6 bits, is required compared to the lower input power case.

D. EVM vs. received power

To investigate the impact of analog non-idealities on the link budget, *EVM* is plotted versus received input power *PRX*, see Fig. 6, for the same receiver as in III-B. The ADC resolution is set to 8 bits. In Case I, only thermal noise in the receiver is considered, in which case it is expected that the *EVM* drops with increasing input power. However, the *EVM* reaches a floor of about 3.2%. This is because the baseband

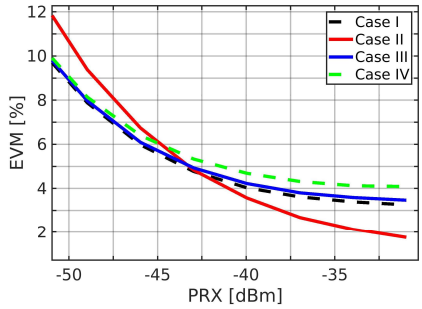


Fig. 6. *EVM vs.* received input power for various cases.

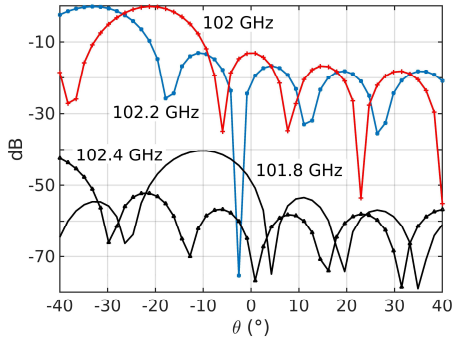


Fig. 7. Interferers and *IM3* components *vs.* scan angle.

filter, a 5th order Butterworth filter, does not have a uniform group-delay in the passband, causing intersymbol interference (ISI). This can either be resolved by using an equalizer in the digital domain, at the cost of higher power consumption in the DSP, or by using a Bessel filter. The Bessel filter has a flat in-band group-delay but also a wider transition band, causing less filtering of noise and interferers. This filter is used in Case II, where the cut-off frequency is also increased to ensure no ISI occurs. As seen in Fig. 6, at low input power, this receiver performs worse than Case I, as more noise is integrated, but at higher *SNR* the performance is better due to the lack of ISI.

In Case III and Case IV, the effect of phase noise is simulated. In both cases, the initial Butterworth filter is once again used. In Case III, each pair of *I/Q* mixers are fed with a separate LO, making the phase noise uncorrelated between each antenna path. In Case IV, one common LO drives all mixers in each sub-block, making the phase noise correlated. In both cases the same phase noise profile was used for all LOs. As discussed in [12], the phase noise floor is the most critical part for wideband signals. Here, a noise floor of –125 dBc/Hz was chosen, a rather high level. This noise floor extends to an offset of 5 GHz from the LO frequency. As expected, the uncorrelated phase noise causes less *EVM*. However, since the performance of each LO here is assumed to be the same, the power consumption of the LO generation will be approximately four times higher in Case III than in Case IV, but this is not considering the extra buffers required in Case IV. Also of note is that for *EVM* values of around 8% and above, the phase noise has a negligible effect in both cases, despite the high phase noise level chosen. Therefore, if only a 16-QAM modulation is targeted, the phase noise requirement can be quite relaxed. However, this may not be applicable if reciprocal mixing of interferers is considered, or if multi-carrier modulation such OFDM is used.

E. Non-linearity versus interferer angle

To ensure that the testbench generates valid results with respect to non-linearity, the setup from [13] was replicated. Two tones, one at 102 GHz and one at 102.2 GHz, were generated and applied to a receiver comprised of two 4-element tiles. The receiver had a total *IIP3* of -17 dBm. The 102-GHz and 102.2-GHz tones have AoA of -21° and -33°, respectively. Both tones have an input power of -37 dBm. Because of the non-linearity in the receiver, *IM3* components at 101.8 GHz and 102.4 GHz will be generated, whose magnitudes will depend on the scan angle θ, that is, the angle in which the receiver is "looking". The peak directions of the *IM3* components will in general not coincide with directions of the actual interferers. This was simulated by sweeping the scan angle of receiver with the two interferers present. Fig. 7 shows the magnitude of the two interferers and the *IM3* components, normalized to the peak level of the interferers, versus scan angle. This figure perfectly matches Fig. 9 in [13].

IV. Conclusion

In this work, a system-level testbench with near-circuit level fidelity has been developed in MATLAB and Simulink. It can be used to investigate the impact of non-ideal analog hardware on the system level performance or to find specifications of analog blocks based on system level requirements.

To demonstrate its capabilities, a 4-element IC tile was implemented and simulated in various configurations. It is shown that for high *SNR* applications, phase noise and filter group-delay can have a significant impact. The combination of non-ideal phase shifters and limited ADC resolution in presence of interferers is also simulated and shown to severely affect the performance. Additionally, to minimize ADC and DSP power consumption, a symbol-rate sampling is used with an MM symbol synchronizer, in combination with beam direction dependent time skew of the tile sampling clocks.

Acknowledgment

We would like to show our gratitude to the Excellence Center at Linköping-Lund in Information Technology (ELLIIT) for supporting this work.

References

[1] K. Rikkinen, *et al.*, "THz Radio Communication: Link Budget Analysis toward 6G," *IEEE Communications Magazine*, 2020.

[2] B. Sadhu, *et al.*, "The More (Antennas), the Merrier: A Survey of Silicon-Based mm-Wave Phased Arrays Using Multi-IC Scaling," *IEEE Microwave Magazine*, 2019.

[3] A. F. Molisch, *et al.*, "Hybrid Beamforming for Massive MIMO: A Survey," *IEEE Communications Magazine*, vol. 55, nr 9, pp. 134-141, 2017.

[4] R. Rotman, M. Tur and L. Yaron, "True Time Delay in Phased Arrays," Proceedings of the IEEE, vol. 104, nr 3, pp. 504-518, 2016.

[5] T. Tuovinen, N. Tervo and A. Pärssinen, "Analyzing 5G RF System Performance and Relation to Link Budget for Directive MIMO," *IEEE Transactions on Antennas and Propagation*, 2017.

[6] P. Skrimponis, et al., "Towards Energy Efficient Mobile Wireless Receivers Above 100 GHz," IEEE Access, vol. 9, pp. 20704-20716, 2021.

[7] E. Björnson, M. Matthaiou och M. Debbah, "Massive MIMO with Non-Ideal Arbitrary Arrays: Hardware Scaling Laws and Circuit-

Aware Design," IEEE Transactions on Wireless Communications, vol. 14, nr 8, pp. 4353-4368, 2015.

[8] "RF Blockset - MATLAB & Simulink," MathWorks, [Online]. Available: https://www.mathworks.com/products/rf-blockset.html.

[9] Q. Chaudhari, Wireless Communications from the Ground Up, Melbourne, Australia, 2018.

[10] C. T. Rodenbeck, *et al.*, "When Less Is More... Few Bit ADCs in RF Systems," *IEEE Access*, 2019.

[11] K. Roth, et al., "A Comparison of Hybrid Beamforming and Digital Beamforming With Low-Resolution ADCs for Multiple Users and Imperfect CSI," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, nr 3, pp. 484-498, 2018.

[12] J. Chen, *et al.*, "Influence of White LO Noise on Wideband Communication," *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, nr 7, 2018.

[13] B. Rupakula and G. M. Rebeiz, "Third-Order Intermodulation Effects and System Sensitivity Degradation in Receive-Mode 5G Phased Arrays in the Presence of Multiple Interferers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, nr 12, pp. 5780-5795, 2018.

Paper III:

A mm-Wave Differential-to-Quadrature Frequency Tripler with Automatic Locking and Quadrature Correction

Rikard Gannedahl, Henrik Sjöland

"A mm-Wave Differential-to-Quadrature Frequency Tripler with Automatic Locking and Quadrature Correction," in *Proc. 2023 IEEE Nordic Circuits and Systems Conference (NorCAS)*, Aalborg, Denmark, 2023, pp. 1-6

A mm-Wave Differential-to-Quadrature Frequency Tripler with Automatic Locking and Quadrature Correction

Rikard Gannedahl, Henrik Sjöland
Dept. of Electrical and Information Technology - Lund University, Sweden
{rikard.gannedahl, henrik.sjoland}@eit.lth.se

***Abstract*—This paper presents an injection-locked frequency tripler for 28-GHz direct-conversion transceivers, which takes a differential input and outputs a quadrature signal. A feedback system is added to ensure that lock occurs and to keep the quadrature error below 1° across operating frequencies and process corners. Simulations show that the tripler can cover the entire 24-30 GHz band across all corners, while consuming 32 mW from a 0.8 V supply. The circuit is designed and simulated in a 22nm FD-SOI CMOS process and occupies 0.2 mm^2 active chip area.**

***Index Terms*—mm-Wave, frequency tripler, quadrature, injection locking, FD-SOI**

I. Introduction

Since the early 2000's, the direct-conversion transceiver (DCT) has been the most popular architecture for mobile communications, thanks to its compact size due to lack of off-chip image and IF filters. However, with the introduction of mm-wave frequencies in 5G, the DCT has seen a decline in popularity due to its requirement of mm-wave quadrature local oscillator (LO) signals, which are difficult to generate with sufficient quadrature accuracy [1].

The three most common ways of generating quadrature signals are: a) using a polyphase filter (PPF), b) generating a differential signal at twice the desired frequency, followed by a divide-by-2 circuit, and c) using a quadrature voltage-controlled oscillator (QVCO). While a passive PPF consumes no power and adds negligible amount of phase noise, wideband designs are lossy and require power consuming amplifiers to obtain sufficient quadrature LO signal level. The divide-by-2 approach is a very common method for sub-6 GHz applications, since it generates very accurate quadrature signals and can be relatively wideband. However, it is less suitable for mm-wave applications, due to poor VCO and divider performance at these frequencies. Lastly, a QVCO provides accurate and wideband quadrature signals, but has worse phase noise performance than a regular VCO.

In [2], the authors proposed a novel quadrature injection-locked frequency tripler (QILFT), in which a differential signal at 1/3 of the desired output frequency is injected into only one halve of a quadrature oscillator, see Fig. 1, and locking to the third harmonic. Normally, this would cause a quadrature amplitude and phase error due to the two oscillator halves now being asymmetric. However, in [2], they use what they refer to as double-sided injection, in which the signal gets injected both in parallel through M_3 and M_4 and in series through M_7 and M_8, but with opposite phase, which reduces the quadrature error. The explanation given in [2] for this improved performance is that when the signal injected through the parallel path is at its maximum, which increases the oscillation amplitude instantaneously, the injected tail current is simultaneously at its minimum, which lowers the oscillation amplitude, thus minimizing the amplitude imbalance between the I and Q halves. A more rigorous explanation is given in [3]. There, the authors show that for parallel injection, the third harmonic of the injected current has the same phase as the fundamental. For series injection, on the other hand, the third harmonic is in anti-phase to the fundamental tone. This means that if the parallel injection and series injection are in anti-phase, the fundamental tones are partially or completely cancelled, which reduces the amplitude error and the modulation sidebands at $3f_{inj} \pm f_{inj}$, as well as at f_{inj}, while the third harmonics are added constructively, thus increasing the locking range.

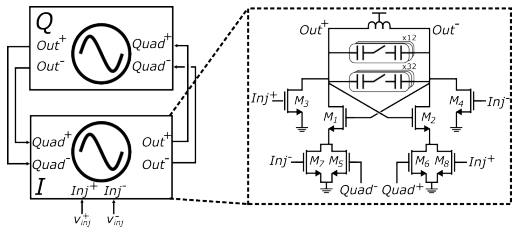


Fig. 1: The quadrature injection-locked frequency tripler proposed in [2].

However, there is a major drawback to this approach. Unless the injected signal is exactly one-third of the free-running oscillation frequency, f_{QILFT}, there will be a quadrature phase error. The greater the difference between $3f_{inj}$ and f_{QILFT}, the greater this error will be, requiring time-consuming and expensive calibration of the frequency tripler [2] [4]. In this work, we build upon the idea in [2] and add a feedback

979-8-3503-3757-0/23/$31.00 ©2023 IEEE

system, so that the quadrature error is minimized across the entire frequency band. The feedback system comprises a phase detector (PD), a 1.5-bit flash analog-to-digital converter (ADC), and a simple digital signal processing (DSP) block. The circuit is designed for the 5G FR2 28-GHz spectrum (24.25-29.50 GHz) and is implemented and simulated in a 22nm Fully-Depleted Silicon-on-Insulator (FD-SOI) CMOS technology.

II. System Overview

Fig. 2 shows a block overview of the system, which works as follows. A differential signal at frequency f_{inj} is injected in the I-part of the QILFT, and a reset pulse is sent to the DSP, setting the QILFT to its highest free-running frequency. The output of the QILFT is fed to a phase detector (PD), implemented as two Gilbert mixers. The PD measures the quadrature phase error between the two oscillator halves and outputs a differential DC voltage proportional to it. The output of the PD is fed to a differential 1.5-bit Flash ADC, where it is compared to a differential reference voltage. The ADC has three possible outputs; '11' if the PD output is above the positive reference voltage, '01' if it is between the reference voltages, and '00' if it is below the negative value of the reference voltage. The ADC output is sent to the digital signal processor, which acts as a counter. As long as the DSP receives a '00', it will keep counting up and turning on switched capacitor cells in the QILFT until $f_{QILFT} = 3f_{inj}$, at which point the ADC outputs '01'. Likewise, if the DSP receives '11', it will count down, turning off switched capacitor cells until it receives '01'.

At a first glance, it may appear that this feedback loop only works if the QILFT is injection-locked to the input signal, since it is only then the PD would produce a DC output carrying information about the frequency mis-tuning. However, this is not the case. Even if the QILFT does not get locked, pulling will still occur [5]. This will cause more phase shift in the I-oscillator, than in the Q-oscillator, resulting in a quadrature error which can be detected by the phase detector. While this will cause the PD to have a non-zero DC

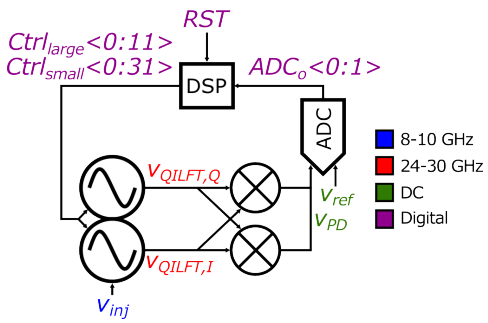


Fig. 2: Block diagram of system. Differential signals are drawn as single-ended for improved readability.

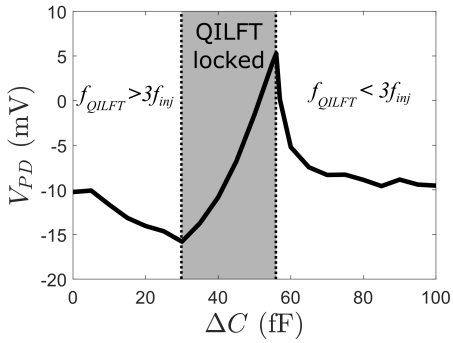


Fig. 3: Phase detector output versus extra tank capacitance.

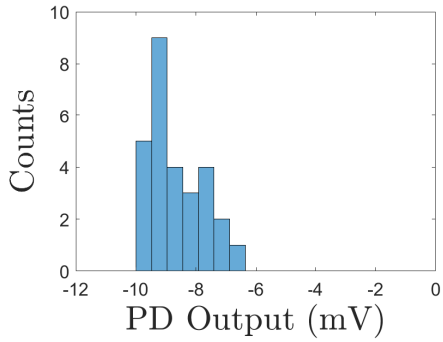


Fig. 4: Histogram of PD output with $f_{inj} = 8$ GHz when lock has not occurred.

output, it does not necessarily mean that the DC output will have the correct sign for the control loop. That is, the effect of LO pulling can actually cause the feedback to increase the gap between $3f_{inj}$ and f_{QILFT}, instead of decreasing it. Fig. 3 shows an example of the PD output versus tank capacitance. As seen in the figure, when f_{QILFT} is higher than $3f_{inj}$, the PD outputs a negative voltage, which would cause the feedback to increase the capacitance, which is the correct response. Interestingly, when f_{QILFT} is lower than $3f_{inj}$, the PD also outputs a negative voltage, which results in an incorrect response, i.e., the feedback would lower f_{QILFT} even further. This is why the locking sequence is initialized by a reset pulse, setting f_{QILFT} to its highest point, which is then sequentially lowered until ideal tuning is achieved.

In addition to the phase error due to LO pulling, there will also always be a quadrature error due to random mismatch between the I and Q oscillator. If this random error is of the same order as the deterministic error due to LO pulling, there is a risk that the feedback system does not detect that lock has not occurred, or even steers the oscillator in the

wrong direction. To ensure that the effect of LO pulling is significantly stronger than that of the random mismatch, a 28-samples Monte Carlo simulation was performed, where random mismatches were added to the two oscillators and the phase detector. The oscillators were set to their highest oscillation frequency and an 8-GHz, 600-$mV_{p,diff}$ signal was injected in the I oscillator. This corresponds to the worst case scenario, since this is the lowest injection frequency that will be used and the further away the injected signal is from the free-running oscillation frequency, the weaker the effect of LO pulling is. Fig. 4 shows a histogram of the differential output voltage of the phase detector. The mean value is -8.7 mV, with a standard deviation of 1.0 mV. This is plenty of margin to -1.8 mV, the level that is needed for the feedback network to detect that lock has not occurred, as will be discussed further in Section III-C.

III. Circuit Implementations

A. Quadrature Injection-Locked Frequency Tripler

The schematic of the QILFT was shown in Fig. 1. As seen in the figure, it comprises two LC oscillators, one I part and one Q part. The injected signal is applied to transistors M_3, M_4, M_7, and M_8 in the I part only. However, these transistors are still present in the Q part as well, but only with a DC voltage applied to them, acting as dummy transistors. Quadrature coupling is done through transistors M_5 and M_6.

The resonance tank in each oscillator comprises a center-tapped differential inductor and two capacitor banks; one for coarse tuning with 12 switched capacitor cells, and one for fine tuning with 32 switched capacitor cells. Fig. 5 shows the tuning range across process corners. As can be seen, the QILFT can cover the entire 24.25-29.50 GHz band.

B. Phase Detector

The phase detector is implemented as two Gilbert cells, with their outputs combined in the current domain, and is shown in Fig. 6. The input transistors, M_1-M_4, are biased in class C to maximize the gain for limited DC current. This makes the

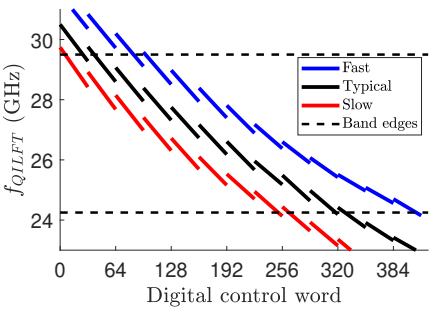


Fig. 5: Free-running tuning range of oscillator across process corners.

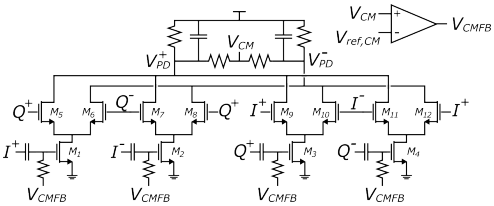


Fig. 6: Schematic of the phase detector.

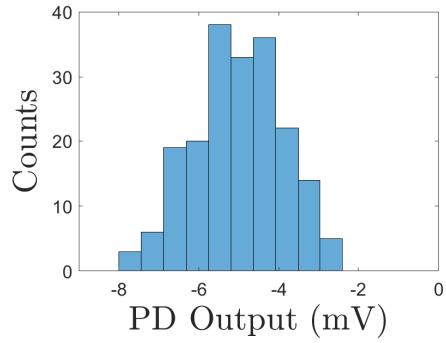


Fig. 7: Monte Carlo simulation of phase detector with an input signal with a 1° phase error.

output voltage, especially the common-mode level, sensitive to process variations and input amplitude. To combat this, common-mode feedback is used to control the bias voltage of M_1-M_4.

Since the quadrature accuracy is mainly limited by how accurately the phase detector can detect the error, the transistors in the phase detector should be made as large as possible to minimize the mismatch, but still small enough to not excessively load the frequency tripler. Transistors M_1-M_4 have dimensions 192 µm/20 nm, and M_5-M_{12} have dimensions 24 µm/20 nm. As the oscillation amplitude is significantly larger than what the PD requires, capacitive voltage division is used to reduce both the input amplitude and capacitive loading. In total, the PD loads each oscillator output node with a capacitance of 120 fF. For comparison, the total node capacitance varies between 560 fF and 880 fF, depending on frequency setting. Fig. 7 shows a 200-point Monte Carlo simulation of the phase detector with a 1° phase error input signal. The mean output value is 5.0 mV and the standard deviation is 1.1 mV, meaning that the phase detector can reliably detect a 1° phase error.

C. Flash ADC

The differential flash ADC is shown in Fig. 8. The PD outputs are fed to two fully differential comparators, along with the two reference voltages. Three different references voltages ($High$, Mid, Low) can be selected depending on

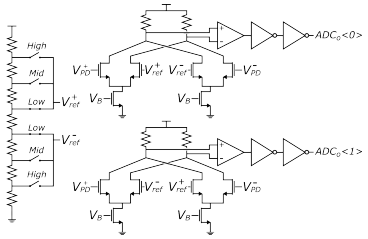


Fig. 8: Schematic of the Flash ADC and reference ladder.

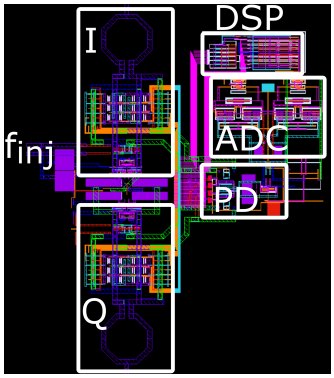


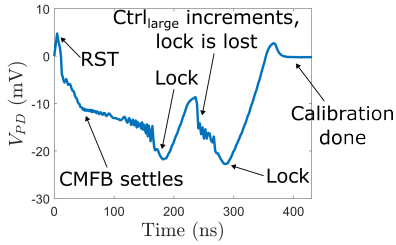
Fig. 9: Layout of whole system.

which output frequency is targeted in the oscillators. For lower frequencies, the phase will shift less per each small capacitor cell turned on, so the static phase error can be reduced by using a lower $V_{ref,diff}$ for these frequencies. The selectable levels are 1.8 mV, 2.3 mV, and 2.8 mV.

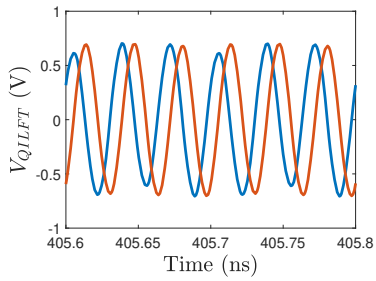
D. DSP

The DSP acts as a counter using flip-flops and low complexity logic, with an external clock controlling the refresh rate. If it receives '00' from the ADC, it will turn on the small tuning capacitances one-by-one until all 32 are on, at which point all are turned off and $Ctrl_{large}$ is incremented by one. If it instead receives '11', the operation is reversed. If '01' is received, the digital output is unchanged. The DSP also has a reset pin, in which a positive flank will cause all outputs to be forced to zero.

The clock period needs to be long enough that the phase detector and ADC have enough time to settle to their the final value after the DSP has changed the control setting. On the other hand, a too slow clock means that the system will take excessively long time to lock and settle. Simulations show robust operation for a clock frequency of 160 MHz, resulting in a worst-case scenario of 2.6 µs to achieve lock and calibration.



(a)



(b)

Fig. 10: Simulated transient for $f_{inj} = 10$GHz. (a) Output of PD. (b) Waveforms once feedback has settled.

IV. Full System Simulation Results

The following simulations were performed following a post-layout parasitic extraction using Calibre xACT. Inductors and interconnects were modelled using Keysight Momentum. The layout of the full system is shown in Fig. 9. The active area is 420 µm x 470 µm and the entire circuit consumes 32 mW from a 0.8 V supply, out of which the QILFT consumes 30 mW.

A. Automatic tuning

Fig. 10(a) shows the phase detector output for a transient simulation when a 600-$mV_{p,diff}$ 10-GHz signal is injected. As seen in the figure, the tuning is initiated with a reset signal, setting the free-running frequency to its highest point. At this point, lock has not occured, but as explained in Section II, the PD output still becomes negative due to LO pulling. This causes the feedback system to start turning on switched capacitor cells in the oscillators. At around 180 ns, lock is achieved, but there is still a large quadrature error. The feedback system then continues until the PD output is within the specified reference voltages, which happens at around 370 ns, at which point the quadrature error is minimized and the tuning is done. Fig. 10(b) shows the I and Q waveforms at this point. The output amplitude is 680 mV and the quadrature error is 0.4°. Some amplitude modulation can be observed in the I waveform, stemming from the injected 10-GHz signal.

TABLE I: Performance comparison of 28-GHz quadrature frequency multipliers.

	This work[a]	ISSCC'18 [6]	TMTT'19 [7]	TCAS-I'19 [8]
Freq [GHz]	24-30	22.8-43.2	20.4-36.6	22.6-24.8
Multiplication factor	3	3	1.5	3
Automatic tuning	Yes	Tuning-less	No	No
Output power variation [dB]	3.7	>30	6	-
Power consumption [mW]	32	19.2[b]	81.6[c]	10.4
Area [mm^2]	0.20	0.84	0.27	0.13
Technology	22nm FD-SOI	65nm CMOS	45nm CMOS	130nm CMOS

[a]Simulated data only.
[b]Including PPF buffers, but excluding output buffers.
[c]Including output buffers.

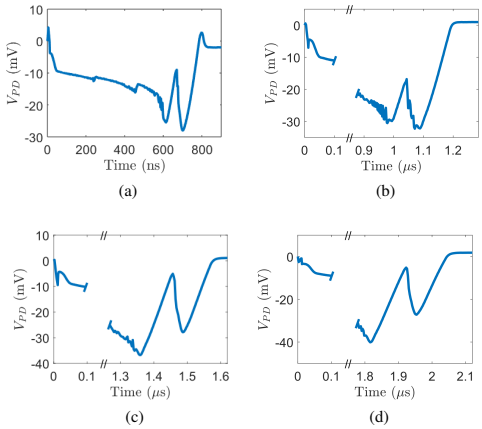


Fig. 11: Simulated transient for (a) $f_{inj} = 9.5$GHz, (b) $f_{inj} = 9$GHz, (c) $f_{inj} = 8.5$GHz, and (d) $f_{inj} = 8$GHz.

Fig. 11 shows the PD output for four additional injection frequencies; 9.5 GHz, 9 GHz, 8.5 GHz, and 8 GHz. As seen in figures, the PD output voltage follows a similar pattern as in Fig. 10(a) and settles at a final voltage within the reference voltages, close to 0 mV, for all four scenarios. As expected, the lower the injection frequency, the longer it takes for the feedback system to settle, since the QILFT always starts at its highest self-oscillating frequency when the tuning is initiated.

Fig. 12 shows the simulated quadrature error (without mismatch) and output amplitude versus injection frequency after the tuning has settled. The quadrature error is below 0.5° for all simulated frequencies, while the amplitude varies from around 450 mV to 680 mV. The worst-case harmonic rejection ratio is 19.4 dB.

B. Phase noise

Since the output phase noise of the QILFT will depend on the phase noise of the input signal, the phase noise profile from [6] was added to the input signal. Fig. 13 shows the simulated input phase noise, output phase noise, and free-running phase noise at 27 GHz ($f_{inj} = 9$GHz). As expected, the output phase

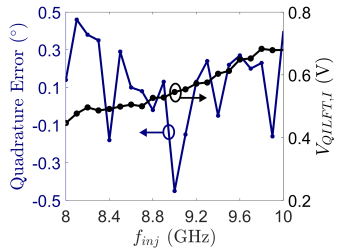


Fig. 12: Simulated quadrature error and output amplitude versus injection frequency.

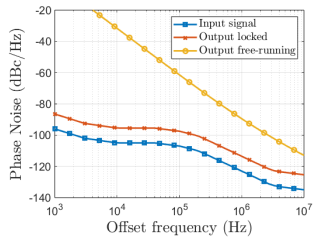


Fig. 13: Phase noise profile at $f_{QILFT} = 27$ GHz ($f_{inj} = 9$ GHz).

noise tracks the input phase noise with a 9.5 dB difference due to the frequency tripling.

C. Comparison

Table I summarizes the performance of the proposed quadrature frequency tripler and compares it to other state-of-the-art quadrature frequency multipliers. While [6] and [7] both achieve a greater tuning range than the proposed tripler, it is at the cost of either higher power consumption or larger area consumption. Additionally, the tripler in [6] suffers from significant output power variation.

V. Conclusion

A differential-to-quadrature injection-locking frequency tripler for mm-wave direct-conversion transceivers has been

presented. A novel feedback system is used to automatically find injection-lock and tune the tripler for optimal quadrature accuracy. Simulations show that it can provide wideband and accurate quadrature signals without the need for any calibration, while being compact and robust to process variations.

ACKNOWLEDGMENT

The authors would like to thank GlobalFoundries for PDK access and the Excellence Center at Linköping - Lund in Information Technology (ELLIIT) for funding this work.

REFERENCES

[1] L. Li *et al.*, "The path to 5G: mmWave aspects," *Journal of Communications and Information Networks*, vol. 1, no. 2, pp. 1–18, 2016.

[2] A. Musa *et al.*, "A Low Phase Noise Quadrature Injection Locked Frequency Synthesizer for MM-Wave Applications," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 11, pp. 2635–2649, 2011.

[3] L. Iotti, G. LaCaille, and A. M. Niknejad, "A Dual-Injection Technique for mm-Wave Injection-Locked Frequency Multipliers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 12, pp. 5417–5428, 2021.

[4] H.-T. Kim *et al.*, "A 28-GHz CMOS Direct Conversion Transceiver With Packaged 2×4 Antenna Array for 5G Cellular System," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 5, pp. 1245–1259, 2018.

[5] B. Razavi, "A study of injection locking and pulling in oscillators," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 9, pp. 1415–1424, 2004.

[6] J. Zhang *et al.*, "A 22.8-to-43.2GHz tuning-less injection-locked frequency tripler using injection-current boosting with 76.4for multiband 5G applications," in *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, 2018, pp. 370–372.

[7] S. Shin *et al.*, "Wide Locking-Range Frequency Multiplier by 1.5 Employing Quadrature Injection-Locked Frequency Tripler With Embedded Notch Filtering," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 12, pp. 4791–4802, 2019.

[8] D. Shin and K.-J. Koh, "24-GHz Injection-Locked Frequency Tripler With Third-Harmonic Quadrature Phase Generator," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 8, pp. 2898–2906, 2019.

Paper IV:

Active and Passive Integrated Filters for Multi-GHz 6G Baseband Applications

Rikard Gannedahl, Henrik Sjöland

"Active and Passive Integrated Filters for Multi-GHz 6G Baseband Applications," in *Proc. 2023 Asia-Pacific Microwave Conference (APMC)*, Taipei, Taiwan, 2023, pp. 524-526

Active and Passive Integrated Filters for Multi-GHz 6G Baseband Applications

Rikard Gannedahl, Henrik Sjöland
Dept. of Electrical and Information Technology - Lund University, Sweden
{rikard.gannedahl, henrik.sjoland}@eit.lth.se

***Abstract*—This paper presents the design and comparison of an active G_m-C and a passive baseband filter for 6G applications. Both presented circuits are 5th-order differential low-pass filters with 3-dB cut-off frequencies of 4.9 and 4.7 GHz, respectively, implemented in a 22nm FD-SOI CMOS technology. The active filter uses back-gate tuning to optimize the transfer function, and achieves a measured in-band IIP3 between -1.0 and -9.1 dBV_p, out-of-band IIP3 between 1.0 and 15 dBV_p, and input-referred noise of 6.55 $nV/\sqrt{Hz}$, while consuming 19.9 mW from a 0.8 V supply and occupying 0.05 mm^2 core chip area. By utilizing mutual inductance, the footprint of the passive filter is minimized, occupying only 0.07 mm^2.**

***Index Terms*—Analog filters, multi-GHz, 6G, FD-SOI**

I. Introduction

With data rates exceeding 100 Gbps per user being targeted in the sixth generation of mobile communication (6G), the channel bandwidths will have to be extended to the multi-GHz range [1]. This necessitates multi-GHz baseband filters, which should have a sharp filter characteristic, high dynamic range (DR), low power consumption, and small size.

Given passive filters' large area consumption, on-chip baseband filters have typically been implemented as active filters in previous mobile communication generations. But since the passive component values scale inversely with cut-off frequency (f_c), the area penalty is not as severe for multi-GHz filters. At same time, active filter performance decreases with increasing f_c. Thus, the question arises as to when on-chip passive filters become a better choice than active ones. To investigate this, we have designed, measured and compared two multi-GHz 5th-order filters, one active and one passive, in a 22nm Fully-Depleted Silicon-on-Insulator (FD-SOI) CMOS technology.

II. Active filter

The active filter is a G_m-C filter based on the well-known transconductor described in [2], shown in Fig. 1, which consists of six inverters, Inv_1-Inv_6, and has no internal nodes. By tuning the transconductance of Inv_4 and Inv_5, the differential output conductance can be minimized, resulting in very high unloaded differential voltage gain. This tuning is typically done by feeding a separate supply voltage to Inv_4 and Inv_5, which is slightly lower than the regular VDD, something that requires a separate low-dropout regulator. We propose instead to use the back-gate to alter the transconductance,

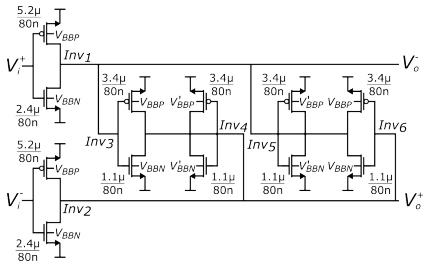


Fig. 1: Schematic of the unit transconductor.

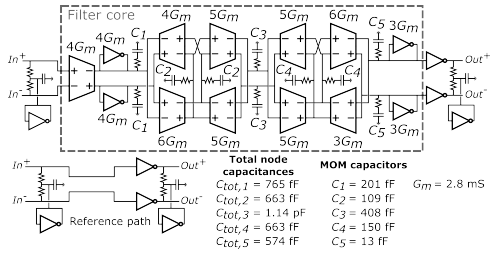


Fig. 2: Complete active filter schematic.

taking advantage of the stronger back-gate control in an FD-SOI process compared to a regular bulk process [3]. As seen in Fig. 1, the transistors in Inv_4 and Inv_5 are connected to separate back-gate voltages, V'_{BBP} and V'_{BBN}, whose absolute values are made slightly smaller than the regular ones, resulting in a higher threshold voltage and thereby lower g_m. Back-gate tuning is also used in [4], but it is only used for optimizing the linearity and tuning the filter cut-off frequency f_c. Process variations are assumed to be small enough that no output conductance tuning is required, an assumption not applicable to multi-GHz filters.

For the active filter, a 4.9-GHz 0.8-dB in-band ripple Chebyshev implementation was chosen. A Chebyshev filter does not only have a sharper roll-off than Butterworth and Bessel filters, it also has larger filter capacitances. This allows for larger transistors to be used, thereby giving the designer more flexibility. The filter schematic is shown in Fig. 2. The filter has been impedance scaled for optimum dynamic range, according

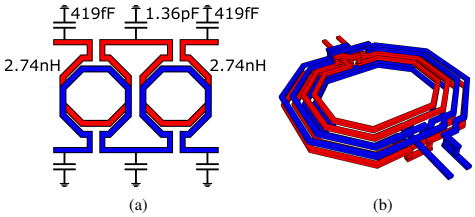


Fig. 3: (a) Simplified illustration of the passive filter. (b) 3D view of the inductors.

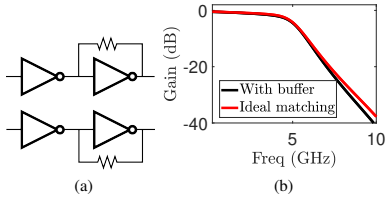


Fig. 4: Buffer implementation for passive filters. (a) Block diagram. (b) Transfer function with buffer.

to the method described in [5]. Resistors have been placed in series with the filter capacitors, which creates left-half-plane (LHP) zeros. This is done to cancel the phase shift from right-half-plane (RHP) zeros, stemming from non quasi-static effects [2]. Simulations show that without the resistors, in-band peaking of almost 3 dB occurs. In addition, as suggested in [2], a reference path is added to be able to de-embed the filter from the rest of the measurement setup. The filter core measures about 0.05 mm^2.

III. Passive Filters

A. Filter design

For the passive filter, a differential 5-pole 4.7-GHz Butterworth filters have been implemented. Fig. 3a shows a simplified illustration of the design. The two paths are mirrored vertically and placed so that the inductors overlap. For a differential signal, the currents will therefore flow in the same direction in the inductors, causing the magnetic fields to add constructively, effectively nearly doubling the inductance of each inductor. Due to this effect, the filter occupies 0.07 mm^2. A more detailed 3D rendition of the overlapping inductors is shown in Fig. 3b.

B. Input buffer

The passive filter is designed assuming single-ended input and output impedances of 50 Ω. Such low impedances may or may not be desirable, depending on the application. In a current-mode transmitter or receiver, the DAC and transimpedance amplifier, respectively, can be matched to this impedance. However, in a voltage-mode receiver or transmitter, a high filter input impedance is needed, requiring an input buffer. An example of such a buffer is shown in Fig. 4a. It comprises an inverter, followed by a resistive-feedback inverter, which is sized so that $R_o \approx 1/G_m = 50\Omega$. Given that this buffer is only necessary in some applications, we opted to only simulate its performance and not include it in the manufactured circuit, to focus on the core passive filter performance. Fig. 4b shows the simulated transfer function when using this buffer, compared with ideal matching. The simulated worst-case in-band IIP3 is 3.2 dBV$_p$, worst-case out-of-band IIP3 is -1.9 dBV$_p$ for an IM3 component located at 3 GHz, and average in-band input-referred noise (IRN) is 2.8 nV/$\sqrt{\text{Hz}}$, while consuming 8.5 mW from 0.8 V supply.

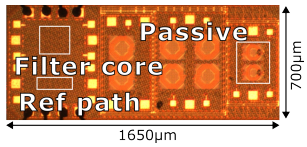


Fig. 5: The manufactured chip with active and passive filter.

IV. Measurement Results

The manufactured chip is shown in Fig. 5. Both filters were measured using SGS probes. The transfer functions were measured with a 4-port VNA, Rohde & Schwarz ZVA67, with true differential signal option. Two-tone signals were generated using two signal generators, Agilent E8257D, followed by a power combiner and a 180° splitter, and non-linearity was measured with a spectrum analyzer, Rohde & Schwarz FSW85. The same spectrum analyzer was used for the noise measurement, where three wideband amplifiers were cascaded ahead of the instrument to minimize its noise contribution.

A. Active filter

Fig. 6a shows the measured transfer function of the active filter, alongside the ideal transfer function. By using only the back-gate voltage, the cut-off frequency can be tuned from 2.5 GHz to 6.4 GHz, also shown in Fig. 6a. At the nominal 4.9 GHz, the entire chip consumes 21.6 mW from a 0.8 V supply. After subtracting the simulated power consumption of the reference path, output buffer and common-mode generators, the resulting power consumption of the filter core is 19.9 mW.

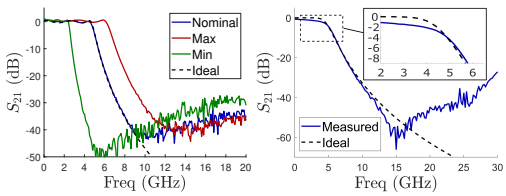
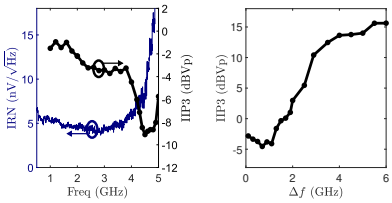


Fig. 6: Measured transfer functions of (a) active filter, and (b) passive filter.

TABLE I: Comparison of state-of-the-art multi-GHz low-pass filters.

	This work Active	**This work** Passive	**This work** Passive w/ buffer	ISSCC '12 [6]	ESSCIRC '14 [7]	TCAS-I '22 [8]
f_c | -3dB [GHz]	4.9	4.7	4.7	4.7	3.3	10.3
Order	5	5	5	3	5	6
Technology	22nm FDSOI	22nm FDSOI	22nm FDSOI	65nm CMOS	28nm CMOS	55nm BiCMOS
Type	Chebyshev	Butterworth	Butterworth	Chebyshev	Chebyshev	Custom
In-band gain [dB]	0.5	-0.7	-0.7	2.7	-1[c]	-0.2
Droop @ $f_c/2$ [dB]	0	0.6	0.6	0[c]	0[c]	0.7[c]
In-band IIP3 [dBV]	-1.5@$f_c/5$ -8.0@f_c	N/A	3.2[b]	-3@f_c	-8@f_c	6.7@$f_c/5$
Out-of-band IIP3 [dBV]	1.0-15	N/A	-1.9-2.6[b]	—	—	—
Inp. ref. noise [nV/$\sqrt{\text{Hz}}$]	6.55	N/A	2.8[b]	6.61	7.0	15.8
Min. attenuation $\{2f_c, 4f_c\}$ [dB]	32	31	37[b]	22[c]	—	38[c]
Pwr consump. [mW]	19.9	N/A	8.5[b]	19	30	43
Filter area [mm^2]	0.05	0.07	0.07	0.01	0.09	0.02
Pwr per pole per Hz [mW/GHz]	0.81	N/A	0.36[b]	1.35	1.81	0.70
FoM[a] [aJ]	36.2@$f_c/5$ 98.1@f_c	N/A	2.45[b]	74.3@f_c	184@f_c	48.1@$f_c/5$

[a] FoM = $P_{DC}/(\text{Order} \cdot f_c \cdot SFDR)$ [9]. [b] Simulated data. [c] Estimated from plots.



(a) (b)

Fig. 7: Measured (a) input-referred noise and in-band IIP3, and (b) out-of-band IIP3 of the active filter.

Fig. 7a shows the IRN and in-band IIP3 versus frequency, measured using two tones separated by 10 MHz. The average in-band IRN is 6.55 nV/$\sqrt{\text{Hz}}$, and the in-band IIP3 varies between -1.0 and -9.1 dBV$_\text{p}$. Out-of-band IIP3 was measured by fixing the lower IM3 component at 3 GHz and applying two tones at frequency offsets of Δf and $2\Delta f$, respectively, from this frequency, and then sweeping Δf, see Fig. 7b. For $\Delta f \geq 1.9$GHz, where both tones are outside the passband, the IIP3 increases from 1.0 to 15.6 dBV$_\text{p}$.

B. Passive filter

Fig. 6b shows the transfer functions of the passive filter, along with the ideal transfer function. The in-band S_{21} is slightly lower than 0 dB due to resistive losses, which increase at higher frequencies because of skin effect, causing some droop. The filter follows the transfer function of the ideal filter closely up to about 12 GHz, where stop-band zeros causes deviation, likely due to parasitic capacitances in the inductors.

C. Filter Comparison

Table I compares the presented active filter, passive filter, and the passive filter with a simulated voltage buffer, with state-of-the-art multi-GHz filters. The active filter achieves an excellent figure of merit (FoM) when compared with other multi-GHz filters. While the filter in [6] achieves a slightly better FoM, it has a lower filter order (3$^\text{rd}$), which results in less internal voltage peaking near f_c, which is not accounted for in the FoM. Additionally, each filter stage contributes noise and non-linearities, worsening the total DR. However, even when accounting for the voltage buffer, the passive filter outperforms all active filters with a simulated FoM of 2.45 aJ, while only occupying a slightly larger area.

V. Conclusion

The presented active G_m-C filter with back-gate tuning achieves state-of-the-art performance for multi-GHz applications. However, unless versatility and area consumption are of main concern, the described passive filter greatly outperforms the active one in terms of figure of merit, even when accounting for a separate input buffer.

Acknowledgment

The authors would like to thank GlobalFoundries for the chip fabrication and ELLIIT for funding this work.

References

[1] "Key drivers and research challenges for 6G ubiquitous wireless intelligence," White Paper, Oulu University, Sept 2019.

[2] B. Nauta, *Analog CMOS Filters for Very High Frequencies*, 1st ed., M. Ismail, Ed. New York, USA: Springer New York, 1993.

[3] R. Carter *et al.*, "22nm FDSOI technology for emerging mobile, Internet-of-Things, and RF applications," in *2016 IEEE International Electron Devices Meeting (IEDM)*, 2016, pp. 2.2.1–2.2.4.

[4] J. Lechevallier *et al.*, "A forward-body-bias tuned 450MHz Gm-C 3rd-order low-pass filter in 28nm UTBB FD-SOI with >1dBVp IIP3 over a 0.7-to-1V supply," in *IEEE ISSCC Dig. Tech. Papers*, 2015, pp. 1–3.

[5] Y. Palaskas and Y. Tsividis, "Dynamic range optimization of weakly non-linear, fully balanced, Gm-C filters with power dissipation constraints," *IEEE Trans. Circuits Syst. II: Analog and Digital Signal Processing*, vol. 50, no. 10, pp. 714–727, 2003.

[6] F. Houfaf *et al.*, "A 65nm CMOS 1-to-10GHz tunable continuous-time low-pass filter for high-data-rate communications," in *IEEE ISSCC Dig. Tech. Papers*, 2012, pp. 362–364.

[7] N. Sabatino *et al.*, "A 5th order gm-C low-pass filter with ±3% cut-off frequency accuracy and 220MHz to 3.3GHz tuning-range in 28nm LP CMOS," in *Proc. ESSCIRC*, 2014, pp. 351–354.

[8] F. Centurelli *et al.*, "A SiGe HBT 6th-Order 10 GHz Inductor-Less Anti-Aliasing Low-Pass Filter for High-Speed ATI Digitizers," *IEEE Trans. Circuits Syst. I: Regular Papers*, vol. 69, no. 1, pp. 100–113, 2022.

[9] A. Yoshizawa and Y. Tsividis, "Anti-blocker design techniques for MOSFET-C filters for direct conversion receivers," *IEEE J. Solid-State Circuits*, vol. 37, no. 3, pp. 357–364, 2002.

Paper V:

Capacitive Cancellation in Compact Integrated Multi-GHz Differential Passive Baseband Filters

Rikard Gannedahl, Henrik Sjöland

"Capacitive Cancellation in Compact Integrated Multi-GHz Differential Passive Baseband Filters." Accepted to *2024 22nd IEEE Interregional NEWCAS Conference (NEWCAS)*, Sherbrooke, Canada.

Capacitive Cancellation in Compact Integrated Multi-GHz Differential Passive Baseband Filters

Rikard Gannedahl, Henrik Sjöland
Dept. of Electrical and Information Technology - Lund University, Sweden
{rikard.gannedahl, henrik.sjoland}@eit.lth.se

***Abstract*—The anticipated introduction of multi-GHz channel bandwidths in 6G makes integrated passive baseband filters very attractive. This work presents techniques to obtain high stop-band attenuation in a compact 4.7-GHz 5th-order passive baseband filter. A lumped model of the filter inductors is derived, and a capacitive cancellation technique is proposed, along with techniques to reduce inductive coupling, which are verified using EM simulations. The proposed techniques improve the stop-band attenuation by up to 30 dB.**

***Index Terms*—Low-pass filters, 6G, integrated filters, capacitive cancellation**

I. Introduction

While the exact specifications for the sixth generation of mobile communication (6G) are yet to be finalized, the trend of ever-increasing data rates is clear and will push channel bandwidths into the multi-GHz range and carrier frequencies into the sub-THz range [1], [2]. The anti-aliasing baseband filter preceding the analog-to-digital converter (ADC) in receivers or following the digital-to-analog converter (DAC) in transmitters will thus also require a cut-off frequency in the multi-GHz range. For previous mobile communication generations, integrated baseband filters have almost exclusively been implemented as active ones due to the unreasonably large footprint of integrated inductors, with some notable exceptions in [3] and [4], where in both cases 800-MHz passive low-pass filter are implemented on-chip, at the cost of substantial area consumption.

However, the higher cut-off frequencies in 6G means smaller filter components, making integrated passive filters a viable choice. For instance, for a 50-Ω 4.7-GHz 5th-order Butterworth filter, the inductors should be 2.74 nH, as shown in Fig. 1a, which can easily be integrated on chip. Still, if this is implemented using regular separated inductors, a differential filter would occupy about 0.16 mm^2. This is considerably larger than similar active filters, where footprints of 0.01-0.05 mm^2 are typical [5]–[8]. In sub-THz antennna arrays, the size of the chip is of particular importance, due to the very small antenna pitch at these frequencies. The chip dimensions per antenna channel must be smaller than the antenna pitch to avoid grating lobes [9], and it is therefore critical that the filter is as small as possible. Fortunately, the footprint of the aforementioned 4.7-GHz filter can be substantially reduced by utilizing mutual inductance, which we demonstrated in [5]. In that work, the two filter paths were mirrored along the horizontal axis and placed so that inductors overlap, see

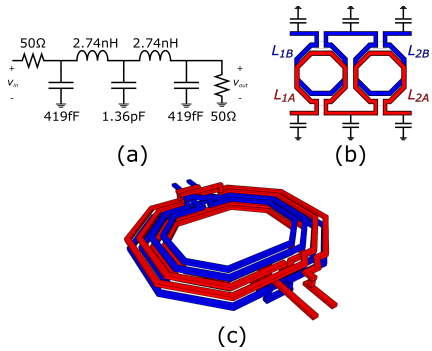


Fig. 1: The 4.7-GHz Butterworth filter in [5]. (a) Ideal schematic. (b) Conceptual filter implementation. (c) 3D rendition of overlapping inductors.

Fig. 1b for a conceptual drawing. Thus, for a differential signal, the currents will flow in the same direction, causing the magnetic fields to add constructively, effectively almost doubling the inductance. In addition to this, area is saved since the inductors are placed on top of each other. Owing to these two effects, this filter is only 0.07 mm^2, smaller than even a single-ended passive filter implementation. A more detailed 3D rendition of the overlapping inductor pair is shown in Fig. 1c. As seen in the figure, the inductors were implemented using four turns each.

This filter was shown to greatly outperform state-of-the-art active filters in terms of dynamic range and power consumption, even when considering an input buffer [5]. However, one issue with this filter can be seen in Fig. 2, which shows both the measured and the electro-magnetically (EM) simulated transfer function of the filter. Above 15 GHz there is a peaking behaviour in the stop-band due to parasitics, limiting the attenuation to only about 24 dB at 33 GHz. This causes issues in both receivers and transmitters; for the former, it can make the ADC susceptible to interferers and increases the noise aliasing, reducing the signal-to-noise ratio, while in the latter, it can cause undesired spectrum content from the DAC to be transmitted to other frequency channels. In this work, we therefore investigate the origins of this stop-band peaking and present methods to mitigate it.

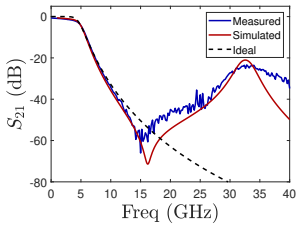


Fig. 2: Measured and simulated transfer function of the Butterworth filter in [5].

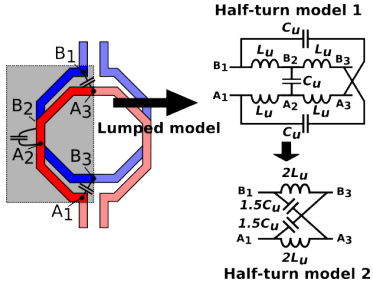


Fig. 3: Derivation of a lumped model of one half-turn.

II. Lumped Inductor Model

In order to analyze the parasitics of the inductor, a lumped model is created. Given that the two sub-inductors are placed directly on top of each other, the capacitance between the overlapping metal traces will be the dominant parasitic. Fig. 3 shows how an approximate lumped model can be created for a half-turn (HT) of the inductor pair if we only consider this parasitic capacitance. Each half-turn is modelled with four unit-inductors L_u and three unit-capacitors C_u, two of which are cross-coupled since the traces of each sub-inductor run in opposite directions. This model can be simplified to only contain two inductors and two cross-coupled inductors (Fig. 3, bottom right), with little loss in fidelity, as will be seen shortly. The full four-turn inductors can subsequently be built from eight cascaded half-turn models, see Fig. 4a. However, given the large number of components in the model, a nodal analysis would provide limited insight about the behaviour of this circuit, and thus further simplifications are necessary. This is done in Fig. 4b, where the full inductor is split into just two segments, each with a cross-coupled capacitor pair. Fig. 5a shows the voltage gain of the three models for $C_u = 1.85$fF and $L_u = 151$pH, and an EM simulation of the overlapping inductor pair, when they are all loaded with $R_L = 50\Omega$ and $C_L = 419$fF, i.e. the same as load as the second inductor in the filter sees. The inductor models based on half-turn model 1 and 2 show good agreement with the EM simulated model, with stop-band peaking at around 35 GHz, which is the cause of the limited stop-band attenuation seen in Fig. 2. Model 3 also has stop-band peaking, but at a slightly higher frequency at around 44 GHz. Still, it is reasonable to assume that the underlying mechanism for this peaking is the same in all three models and the real inductor pair. If we can then develop a method to reduce the peaking for model 3, it is likely also effective for the other cases. We thus proceed with an analysis of model 3.

Nodal analysis reveals that the differential-mode voltage gain of model 3 is given by Eq. (1), see the bottom of the next page. This system has four real zeros (two at 260 Grad/s and two at -260 Grad/s) and two complex-conjugate pole pairs, see Fig. 5b. One of the pole pairs has a very high Q value, and in conjunction with the zeros, that pole pair is causing the peaking.

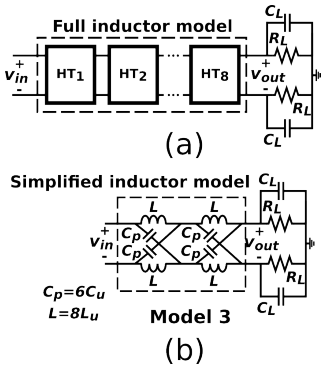


Fig. 4: (a) Lumped model built from half-turn models. (b) Simplified lumped model of the overlapping inductor.

III. Capacitance Cancellation

For filters with regular separated integrated inductors, non-ideal stop-band behaviour can typically be traced to inter-turn parasitic capacitances that can be modelled with a capacitor between the input and output in parallel to the inductor [10], causing transfer zeros. In a differential circuit, the effect of these capacitances can easily be neutralized by cross-coupling capacitors between the input of one inductor and the output of the other inductor, and vice versa [11]. This is, however, not applicable to the overlapping inductors, since the main parasitic capacitances are then already cross-coupled, as shown in the previous section. To cancel the effect of these capacitances, we should then instead add capacitors in parallel with the inductors.

If we add a capacitor C_C as shown in Fig. 6, the voltage gain becomes the expression in Eq. (2). It can clearly be seen that two of the zeros can be removed by setting $C_C = C_p$, but this does not suppress the effect of the problematic pole pair. If we instead set $C_C = 4C_p$, the two zeros cancel with two poles, as shown in Eq. (3). Now, the expression only comprises two real zeros and one (relatively) low-Q complex pole pair, and the stop-band peaking has been eliminated. Thus, this analysis shows that a capacitor can be added between the mid-

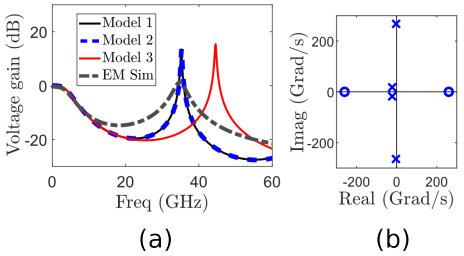


Fig. 5: (a) Voltage gain of the lumped models and EM simulated inductor. (b) Pole-zero plot of model 3.

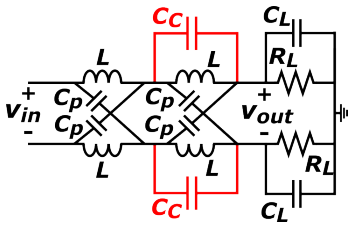


Fig. 6: Capacitive cancellation scheme.

point of each sub-inductor and its corresponding output node to eliminate the stop-band peaking.

Fig. 7a shows the voltage gain of model 3 and an EM simulated overlapping inductor, with and without the capacitive cancellation. A *fracpole* component [12] is added to the lumped model to emulate the skin effect. It can clearly be seen in the figure that the proposed capacitive cancellation technique shows great merit.

The simulation was repeated, but for a load that corresponds to the load seen by the first inductor in the Butterworth filter. While the analysis in this section was carried out for the second inductor in the filter, it can be seen in Fig. 7b that the proposed method is also effective for the first inductor.

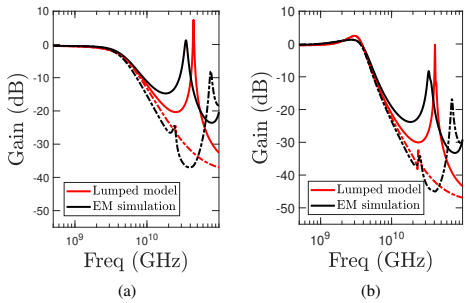


Fig. 7: Voltage gain of the lumped model and EM simulated inductor, without (solid lines) and with (dashed lines) C_C, with loads corresponding to the (a) second inductor, and (b) first inductor of the filter.

IV. Filter Simulations

Next, the capacitive cancellation technique was applied to the full Butterworth filter. Fig. 8 shows the simulated transfer function of the original filter (black curve), and with capacitive cancellation when the inductor pair in different filter sections is individually EM simulated (red dashed curve), i.e. the mutual inductances between L_1 and L_2 are not considered. Clearly, the added capacitors greatly improve the stop-band behaviour. However, when the full layout is EM simulated, the transfer function deteriorates, also plotted in Fig. 8 (purple dotted curve), due to mutual coupling between L_1 and L_2.

A proven way to reduce coupling between multiple inductors is to use 8-shaped inductors [13], [14], at the cost of a slightly larger area and lower quality factor. Therefore, a new overlapping inductor pair was designed using this technique, of which a conceptual drawing is shown in Fig. 9a, while a more detailed 3D rendition is shown in Fig. 9b. In the latter, the taps used to connect C_C are also shown. As seen in the figures, mutual inductance is still utilized to minimize the inductor footprint. At 4.7 GHz, the 8-shaped inductor has a differential Q value of 10.3, while the initial inductor has a differential Q value of 12.3, so slightly higher in-band losses are expected. The final layout using the 8-shaped inductors is shown in Fig. 10 and measures 0.07 mm^2, the same size as

$$A_V = \frac{(1-s^2LC_p)^2}{2(1+s^2LC_p)(1+s\frac{L}{R_L}+s^2L(C_p+C_L))-(1-s^2LC_p)^2} \tag{1}$$

$$A_{V,CC} = \frac{(1-s^2LC_p)(1+s^2L(C_C-C_p))}{2(1+s^2L(C_p+\frac{C_C}{2}))(1+s\frac{L}{R_L}+s^2L(C_p+C_L+C_C))-(1+s^2L(C_C-C_p))^2} \tag{2}$$

$$A_{V,CC}|_{C_C=4C_p} = \frac{(1-s^2LC_p)(1+s^23LC_p)}{2(1+s^23LC_p)(1+s\frac{L}{R_L}+s^2L(5C_p+C_L))-(1+s^23LC_p)^2} = \frac{(1-s^2LC_p)}{1+s\frac{2L}{R_L}+s^2L(7C_p+2C_L)} \tag{3}$$

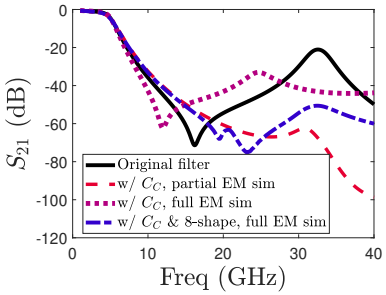
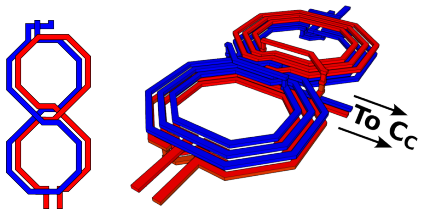


Fig. 8: EM simulated transfer functions of different filters.



(a) (b)

Fig. 9: 8-shaped inductor pair. (a) Conceptual figure. (b) 3D rendition.

the original filter, so the increase in inductor size is negligible. Ground rings were also added around the inductors to further reduce the coupling. The simulated transfer function for the filter using 8-shaped inductors and capacitive cancellation is plotted in Fig. 8 (blue dash-dot curve). Between 22 and 36 GHz, the attenuation is improved by at least 20 dB, with a peak improvement of almost 30 dB at 33 GHz, compared with the original filter. Due to the slightly lower Q, the attenuation is about 0.5 dB higher at 4.7 GHz in the new filter.

Lastly, the common-mode transfer function of both the original filter and the new filter with 8-shaped inductors was simulated, see Fig. 11. It can be seen in the figure that the new filter has better common-mode filtering than the original filter between 10 and 20 GHz, while it is worse above 20 GHz. For most scenarios, the transfer function of the new filter is preferable. Despite this improvement, the common-mode filtering is still considerably worse than the differential filtering, with no significant attenuation occurring below about 11 GHz. This is because the currents flow in opposite directions in the overlapping traces in the inductor pair for a common-mode signal, cancelling most of the magnetic field, hence reducing the inductance. This is the main disadvantage of using the compact, overlapping inductors, and needs to be considered in the system design.

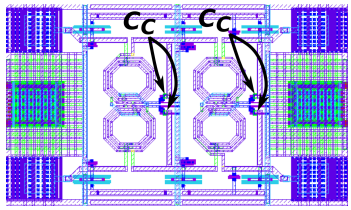


Fig. 10: Layout of the filter using 8-shaped inductors.

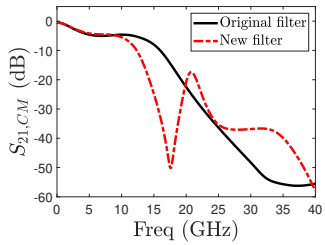


Fig. 11: EM simulated common-mode gain of the original filter and filter with 8-shaped inductors.

V. Conclusion

This work has presented capacitive cancellation and inductive coupling-reducing techniques to improve the stop-band attenuation in a very compact multi-GHz passive baseband filter, which shows excellent merit. While the results are only simulated, the close agreement between EM simulations, lumped models, and the measured original filter makes us confident in the effectiveness of the proposed approach.

Acknowledgment

The authors would like to thank GlobalFoundries for fabricating the initial chip and the Excellence Center at Linköping – Lund in Information Technology (ELLIIT) for funding this work.

References

[1] T. S. Rappaport *et al.*, "Wireless Communications and Applications Above 100 GHz: Opportunities and Challenges for 6G and Beyond," *IEEE Access*, vol. 7, pp. 78 729–78 757, 2019.

[2] N. Rajatheva *et al.*, "White paper on broadband connectivity in 6G," *arXiv preprint arXiv:2004.14247*, 2020.

[3] H.-C. Park *et al.*, "4.1 A 39GHz-Band CMOS 16-Channel Phased-Array Transceiver IC with a Companion Dual-Stream IF Transceiver IC for 5G NR Base-Station Applications," in *2020 IEEE International Solid-State Circuits Conference - (ISSCC)*, 2020, pp. 76–78.

[4] P. Song and H. Hashemi, "mm-wave mixer-first receiver with selective passive wideband low-pass filtering," *IEEE Journal of Solid-State Circuits*, vol. 56, no. 5, pp. 1454–1463, 2021.

[5] R. Gannedahl and H. Sjöland, "Active and Passive Integrated Filters for Multi-GHz 6G Baseband Applications," in *2023 IEEE Asia-Pacific Microwave Conference (APMC), in press*, 2023.

[6] F. Houfaf *et al.*, "A 65nm CMOS 1-to-10GHz tunable continuous-time low-pass filter for high-data-rate communications," in *IEEE ISSCC Dig. Tech. Papers*, 2012, pp. 362–364.

[7] P. Wambacq *et al.*, "A fifth-order 880MHz/1.76GHz active lowpass filter for 60GHz communications in 40nm digital CMOS," in *2010 Proceedings of ESSCIRC*, 2010, pp. 350–353.
[8] F. Centurelli *et al.*, "A SiGe HBT 6th-Order 10 GHz Inductor-Less Anti-Aliasing Low-Pass Filter for High-Speed ATI Digitizers," *IEEE Trans. Circuits Syst. I: Regular Papers*, vol. 69, no. 1, pp. 100–113, 2022.
[9] B. Sadhu, X. Gu, and A. Valdes-Garcia, "The More (Antennas), the Merrier: A Survey of Silicon-Based mm-Wave Phased Arrays Using Multi-IC Scaling," *IEEE Microwave Magazine*, vol. 20, no. 12, pp. 32–50, 2019.
[10] A. Massarini, M. Kazimierczuk, and G. Grandi, "Lumped parameter models for single- and multiple-layer inductors," in *PESC Record. 27th Annual IEEE Power Electronics Specialists Conference*, vol. 1, 1996, pp. 295–301 vol.1.
[11] S. Wang, F. C. Lee, and J. D. van Wyk, "Design of Inductor Winding Capacitance Cancellation for EMI Suppression," *IEEE Transactions on Power Electronics*, vol. 21, no. 6, pp. 1825–1832, 2006.
[12] K. Kundert, "Modeling Skin Effect in Inductors," Oct. 2001.
[13] T. Mattsson, "Method of and inductor layout for reduced VCO coupling," Dec. 19 2006, US Patent 7,151,430.
[14] P. Andreani *et al.*, "A TX VCO for WCDMA/EDGE in 90 nm RF CMOS," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 7, pp. 1618–1626, 2011.