CMOS Microwave Power Amplifier Design for Chireix Configurations

Laurens Bogaert, Joris Lambrecht

Supervisors: Prof. dr. ir. Dries Vande Ginste, Prof. dr. ir. Johan Bauwelinck
Counsellors: Ramses Pierco, Jochen Verbrugghe, Dr. ir. Guy Torfs

Master's dissertation submitted in order to obtain the academic degree of
Master of Science in Electrical Engineering

UNIVERSITEIT
GENT

CMOS Microwave Power Amplifier Design for Chireix Configurations

Laurens Bogaert, Joris Lambrecht

Supervisors: Prof. dr. ir. Dries Vande Ginste, Prof. dr. ir. Johan Bauwelinck
Counsellors: Ramses Pierco, Jochen Verbrugghe, Dr. ir. Guy Torfs

Master's dissertation submitted in order to obtain the academic degree of
Master of Science in Electrical Engineering

UNIVERSITEIT
GENT

Preface

Starting without experience in PA design and Cadence, this master thesis certainly has been a challenging and very interesting experience. Therefore, we would initially like to thank our supervisors Prof. dr. ir. Johan Bauwelinck and Prof. dr. ir. Dries Vande Ginste for giving us this opportunity. Special thanks go to our counsellors ir. Ramses Pierco, dr. ir. Jochen Verbrugghe and dr. ir. Guy Torfs for sharing their expertise on the subject with us, for guiding us through the design, and for proofreading part of the final report.

Additionally, we would like to thank ir. Jan Gillis for his help regarding some ICT problems, and ir. Gertjan Coudyzer and ir. Arno Vyncke for the helpful advice.

Next, special thanks are to be given to our parents and family, for giving us the opportunity to start and complete these studies and master dissertation.

We would also like to thank our fellow thesis students for being helpful and for creating a cheerful atmosphere in the thesis room: Hannes Ramon, Zeger Van de Vannet, Stef Vandermeeren, Bob Mertens, Thomas Deckmyn, Matthias Dewilde, Jelle Bailleul and Johannes Van Wonterghem.

Laurens Bogaert & Joris Lambrecht, May 2015

Admission to Loan

CMOS Microwave Power Amplifier Design for Chireix Configurations

by

Laurens BOGAERT & Joris LAMBRECHT

Master's Dissertation submitted to obtain the academic degree of
Master of Science in Electrical Engineering

Academic 2014–2015

Supervisors: Prof. dr. ir. Dries VANDE GINSTE, Prof. dr. ir. Johan BAUWELINCK
Counsellors: Ramses PIERCO, Jochen VERBRUGGHE, Dr. ir. Guy TORFS

Faculty of Engineering and Architecture
Ghent University

Departement of Information Technology
Chairman: Prof. Dr. Ir. Daniël DE ZUTTER

Summary

In this master dissertation, focus went to the design of a 15 GHz outphasing power amplifier with drivers in a 45 nm CMOS-GPDK. This PA is intended as a block in a 15 GHz outphasing transmitter. To obtain a higher efficiency at power backoff, single- and multilevel-LINC with load modulation is used initially. The desired output power is at least 20 dBm, at a peak drain efficiency $\eta_{d,peak}$ of at least 50 %. Two 0-dBm phase-modulated inputs are given to realize this outphasing.

Keywords

Outphasing, Load modulation, 45 nm CMOS, On-chip inductors, Power combination

CMOS Microwave Power Amplifier Design for Chireix Configurations

Laurens Bogaert, Joris Lambrecht

Supervisor(s): Prof. dr. ir. Dries Vande Ginste, Prof. dr. ir. Johan Bauwelinck, ir. Ramses Pierco, dr. ir. Jochen Verbrugghe, dr. ir. Guy Torfs

Abstract— **In this master dissertation, focus went to the design of a 15 GHz outphasing power amplifier with drivers in a 45 nm CMOS-GPDK. This PA is intended as a block in a 15 GHz outphasing transmitter. To obtain a higher efficiency at power backoff, single- and multilevel-LINC with load modulation is used initially. The desired output power is at least 20 dBm, at a peak drain efficiency $\eta_{d,peak}$ of at least 50 %. Two 0-dBm phase-modulated inputs are given to realize this outphasing.**

Keywords— **Chireix outphasing, Load modulation, 45 nm CMOS, On-chip inductors, Power combination**

I. INTRODUCTION

THE primary issue in power amplifier design is the trade-off between high efficiency and high linearity. High efficiency reduces the power-to-heat conversion and cooling requirements; for mobile users, it improves battery lifetime and size. High linearity enables more complex modulation schemes with amplitude modulation. The classical PA efficiency decreases quickly as the desired output power decreases ("output power backoff"). Depending on the probability density of the relative output amplitude, the resulting total average efficiency can be very low; especially with signals with high PAR. The concept of outphasing is to split a baseband signal into two constant-amplitude signals with a common and differential phase modulation. Both signals are amplified separately and recombined at the output, where the amplitude modulation is reconstructed. Amplitude linearity is no longer needed in the PA's, which enables the use of more efficient non-linear or switching PA's. The goal of this master dissertation was to obtain a peak output power at least 20 dBm at a total efficiency of at least 50 %. A differential class-$EF_{2,odd}$ PA output stage, as proposed in [2], was designed with 45 nm 1.1V NMOS-transistors. First, the concepts of outphasing, load modulation and ML-LINC are briefly explained. Afterwards, the design and results are presented.

II. OUTPHASING

The baseband signal $s(t) = I(t) + jQ(t) = A(t)e^{j\theta(t)}$ is split into two signals $s_1(t)$ and $s_2(t)$ with constant amplitude A_{max} (cfr. figure 1):

$$s_1(t) = A_{max}e^{j\theta(t)}e^{j\phi(t)} \qquad (1)$$

$$s_2(t) = A_{max}e^{j\theta(t)}e^{-j\phi(t)} \qquad (2)$$

$$\phi(t) = arccos(\frac{A(t)}{A_{max}}) \qquad (3)$$

At the output, the amplitude modulation is reconstructed with a gain factor of 2 (cfr. figure 1 (adapted from [3])):

$$s_{out}(t) = s_1(t) + s_2(t) = 2cos(\phi(t))A_{max}e^{j\theta(t)} \qquad (4)$$
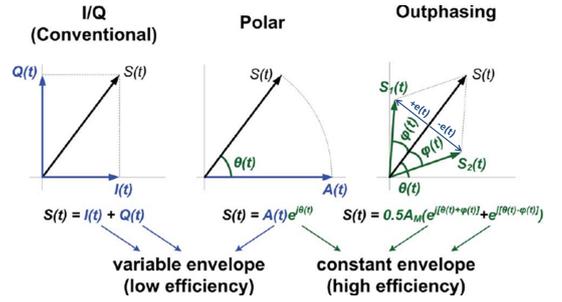
Fig. 1. Outphasing concept.

In ML-LINC (Multilevel-LINC), outphasing is combined with power supply control: below a certain P_{out} on, the supply is reduced and ϕ is reset to zero; resulting in a higher efficiency, with or without load modulation. AMO (Asymmetric Multi-Level Outphasing) ([4]) goes one step further by combining ML-LINC with asymmetric outphasing vectors, to guarantee a minimal ϕ_{avg} for all P_{out}. To combine the amplified signals $s_{1,out}(t)$ and $s_{2,out}(t)$, an isolating combiner (e.g. Wilkinson combiner) or non-isolating combiner (e.g. in [1]) can be used. The non-isolating combiner results in load modulation: the effective load seen by each PA depends on the other PA, and thus on the outphasing angle ϕ. As ϕ is increased, the real part of parallel equivalent of the load is increased, and the efficiency is improved with respect to the case without load modulation. A parallel reactance arises, which can be resonated away at $\phi = \phi_{comp}$ to improve the efficiency, as explained in [1] and [5] (p. 304).

III. POWER AMPLIFIER

A differential class-$EF_{2,odd}$ PA output stage, as proposed in [2] and demonstrated in figure 2 (from [2]), driven by an invertor stage followed common source-driver stage with LC-load was designed with 45 nm 1.1V NMOS-transistors. The use of a differential load is enabled through a balun. The RF-chokes are omitted: the inductance of the primaries of the balun suffices, because the design started from a class-E design with a finite DC-feed inductance. This same inductance is reused again to resonate the parasitic $C_{dd,tot,eff}$ away at the second harmonic, transforming the class-EF_{odd}-PA into an class-$EF_{2,odd}$-PA. As a first driver stage, an input balun followed by an invertor with a large DC-feedback resistor is used to provide non-linear gain

and a square-wave-like drive signal for the LC-drivers. The CS-LC-drivers use an extra inductance as a DC-feed and to resonate the large $C_{gg,eff}$ of the output stage transistors away at the fundamental frequency. An extra LC-tank was added to provide a large impedance at the third harmonic as well. Combined with the third-harmonic current generated by e.g. the square-wave-like input signal at the LC-drivers, this gives a sufficiently strong third harmonic in the input signal of the output stage. It was found that the amplitude of the third harmonic increases as the outphasing angle increases, resulting in a slightly larger efficiency because of the steeper edges in the drive signal. Only 45 nm - 1.1V NMOS-transistors have been used. Each invertor-NMOS has $W_{tot} = 40\mu m$, the invertor-PMOS has a total width of $80\mu m$. The LC-driver-NMOS has $W_{tot} = 60\mu m$, each output stage NMOS has a total width of $W_{tot} = 600\mu m$. This significant size difference is possible thanks to the inductive load, which resonates with the large $C_{gg,tot,eff}$.
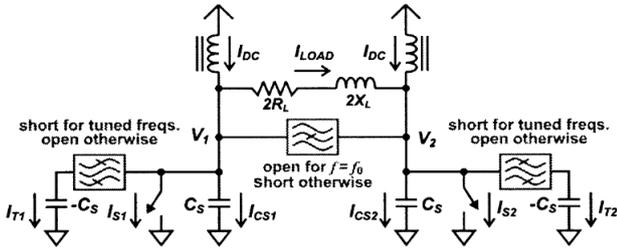
Fig. 2. Concept of a class-$EF_{even,odd}$-PA.
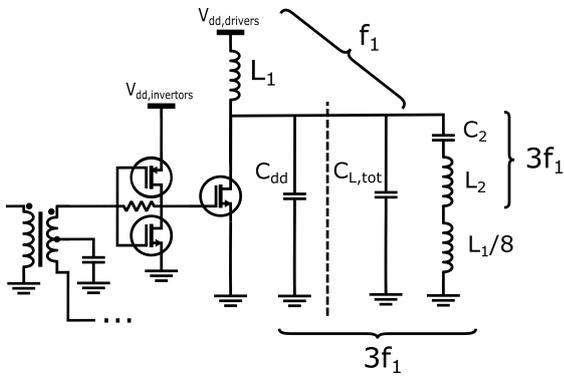
Fig. 3. Initial concept of the load network for the LC-driver. Although not exact, it is a good enough starting point: the admittance can now be calculated and set to zero at the fundamental and third harmonic. $C_{L,tot}$ represents the total effective input capacitance of the output stage transistors, some time-dependent contribution due to Miller effect is expected here as well.

The $V_{gd,breakdown}$-voltage is not specified in the 45 nm-GPDK. $V_{gd,breakdown} \approx 2V_{dd,nom}$, $V_{ds,breakdown} \approx (2...3)V_{dd,nom}$ is assumed in [6] and [7], and therefore also in this thesis.

Differential combining with a floating load is the most efficient because no combiner structure is involved. Differential combining through a balun to enable a single-ended output load is not efficient due to the rather high losses in the balun: even if the efficiency is e.g. 80 %, the total efficiency is reduced by this same factor. Therefore, the CM-combiner proposed in [1] is used. Off-chip 40Ω - $\frac{\lambda}{4}$-transmission lines are assumed

and an EM-model is simulated on a Rogers RO4003C-substrate ($\epsilon_{r,design} = 3.55$, $tan(\delta) = 0.0027$ at 10 GHz). On-chip $\frac{\lambda}{4}$-transmission lines are too long at 15 GHz, and very lossy, due to skin effect and a large $tan(\delta) = 0.1$ at 15 GHz. With ideal drivers and $V_{out,driver}$ in $[0.4V, 1.2V]$, each PA separately delivers 18.21 dBm into a 25 Ω load, at 60.71 % efficiency, with $V_{dg} <\approx V_{dg,breakdown} = 2.2V$. In the 1-level LINC setup with these ideal drivers, the PA supply had to be reduced to prevent V_{dg} from exceeding $V_{dg,breakdown}$ at higher outphasing angles. The specifications are almost met: $P_{out,max} = 19.99$ dBm at a peak efficiency of 54.74%, while $V_{gd,breakdown} < 2.2V$ for all outphasing angles, without compensation reactances jX_{comp}. The compensation inductances were not added because L_{comp} is outside of the realisable range, and the introduction of C_{comp}, L_{comp} lowers the peak efficiency below 50 %. The same PA was also simulated in a 4-level LINC setup, resulting in figure 4.
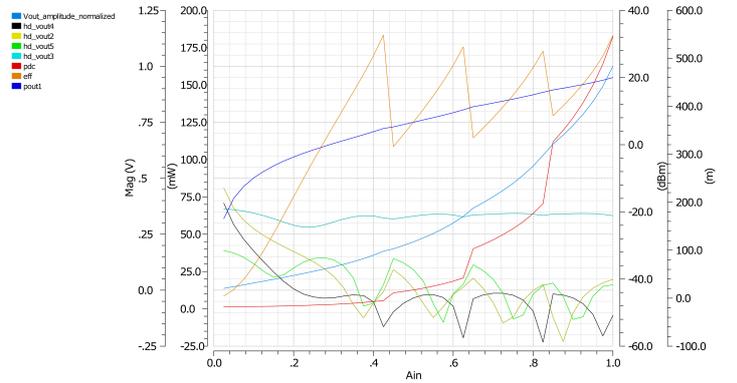
Fig. 4. PA performance with an off-chip CM-combiner and ideal drivers, in a 4-level LINC setup. All performance measures at power back-off (efficiency, harmonic distortion ...) were improved by increasing the number of levels from 1 to 4.

With the addition of the real drivers, a peak output power of 21.15 dBm is reached at an overall efficiency of 49.55 %. The invertors consume only 9.5 mW, the LC-drivers consume 22.64 mW. However, the specifications could not be not met without exceeding $V_{gd,breakdown}$ at a high P_{out} smaller than $P_{out,max}$. The effect of the load modulation is felt even into the LC-driver stage, resulting in asymmetric inputs at the output stage and a higher V_{gd}-peak. At this point, load modulation was abandoned in favour of the certainty of a fixed output load at each PA. The efficiency at output power back-off is improved by using 4-level-LINC system. An off-chip Wilkinson combiner, consisting of the same 40Ω - $\frac{\lambda}{4}$-transmission lines as in the CM-combiner and an extra 32 Ω-resistor, is used to present each PA with the same load as in the case with load modulation, while keeping the single-ended output load at 50 Ω. The results are given in figure 5. Because the transmission lines are identical, the maximal output power is also 21.15 dBm at an overall efficiency of 49.55 %; but the supply does not have to be reduced to prevent problems at output power back-off.

The efficiency of the PA output stages only is comparable as long as the outphasing angle is not too large. With load modulation, the efficiency drops less steeply with increasing outphasing angle because the PC-power decreases gradually since the real part of the load is increasing. The real drivers deliver a larger drive signal to the output stage than the assumed ideal drivers but
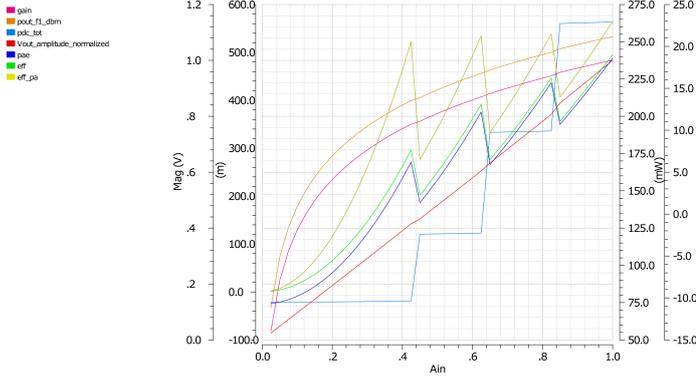
Fig. 5. Performance of the PA with real drivers, in a 4-level LINC system, without load modulation, and an ideal common-mode power combiner.
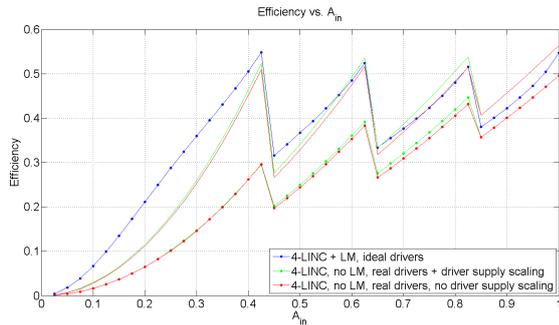
Fig. 6. Efficiency comparison of the 4-LINC outphasing PA with ideal drivers and load modulation, and the 4-LINC outphasing PA with real drivers and without load modulation (Wilkinson combiner). The efficiency of the output stage is also plotted separately in full lines.
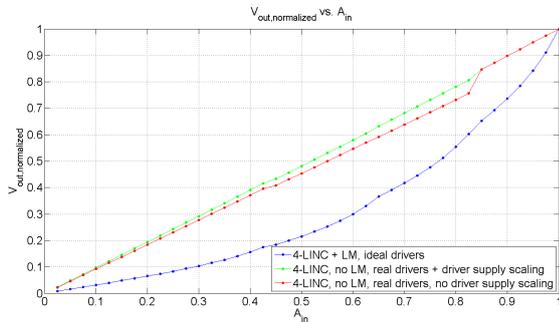
Fig. 7. Output amplitude comparison. As expected, the PA without load modulation is much more linear because the voltage division factor resulting for the unknown and non-linear PA output impedance and the variable load is not present. The only cause of amplitude distortion is a phase shift difference introduced by the PA's.

the rise and fall time are increased. Still, the efficiency peaks of the output stages are comparable. With the drivers included, the overall efficiency drops below the output stage efficiency, and this difference increases as the outphasing angle increases and $V_{dd,PA}$ is scaled down. Even though e.g. the $C_{dd,tot}$ is voltage-dependent, the output stage efficiency does not decrease with a decreasing supply voltage at first. An attempt was made to scale down the supply of the drivers, but this degrades the output stage efficiency and thus the overall efficiency, and the output linearity as well. The design without load modulation offers more power and a more linear output. This difference is quite significant,

considering that the output stages are almost identical.

IV. DESIGN OF THE PASSIVE COMPONENTS

During this work, gpdk45 was used for the initial design. However due to the requirement of a highly efficient amplifier and the fact that gpdk45 has a heavily doped substrate it was opted to change the process parameters, to resemble those of an RF CMOS substrate. Hence the design was performed with a ρ_{sub} of 10 Ω-cm instead of the extracted value of 0.01 Ω-cm for the gpdk45 process [7].

The final design consists of two inductors with completely different inductance values. Hence they were implemented via different techniques. Firstly, a 50 pH inductor is needed and since this value is small it is implemented by using a shorted stub with a length of 123.45 μm where the transmission line is designed as a coplanar waveguide (Fig. 8). To end up with a highly efficient amplifier, high Q-factors and a sufficiently high SRF are needed for the inductors. In this case, a Q-factor of 15.126 is achieved while the SRF is even higher than 300 GHz.

Fig. 8. Layout of a 50 pH shorted stub inductor

Secondly, 400 pH is needed and this value is too high to be implemented as a shorted stub inductor since this would result in a huge component. Hence the 400 pH inductor has been implemented as a two-turn spiral inductor (Fig. 9). Simulations provide 14.138 and 87 GHz as respective values for the Q-factor and the SRF of this component.
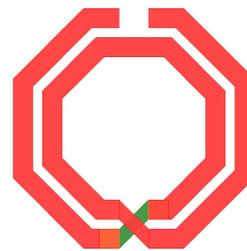
Fig. 9. Layout of a 400 pH inductor

V. DESIGN OF THE BALUNS

In the final configuration of the amplifier, two baluns are needed. Firstly, a power splitter has to convert the unbalanced signal entering the system to a differential input for the driver. This is done by a 50 Ω 1:1 balun (Fig. 10). After matching the balun with parallel capacitors, a total efficiency of 75.3 % is obtained, which is to be expected due to the limited Q-factor of the inductors used in the transformer design [8].
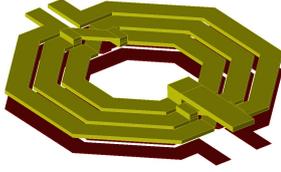
Fig. 10. Layout power splitter

During the course of this dissertation, different types of RF chokes were explored to find the most optimal one for the final EF_2 stage. Unfortunately, when the RF choke and output balun of the final stage are independently designed, only a suboptimal solution is obtained. During this dissertation this was solved by designing a balun and additional circuitry such that the equivalent system behaves the same at the operating frequency (Fig. 11) while the center tap is used instead of the RF choke to supply DC power to the final stage.
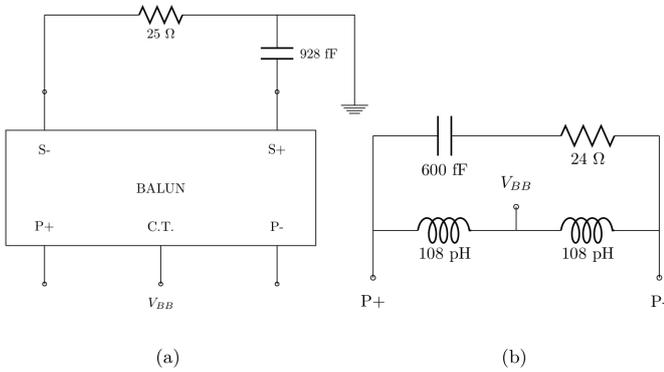
Fig. 11. Desired equivalence of the output balun circuitry

To be able to make the output balun very efficient (i.e. maximally obtainable efficiency equal to 79.3 % [8]) a combination of two techniques is used in this work. Firstly, one-turn primary and secondary windings are implemented since they can be made very wide resulting in the potential to obtain high Q-factors. Afterwards, the primary is shifted relative to the secondary to decrease the interwinding capacitances significantly (Fig. 12).
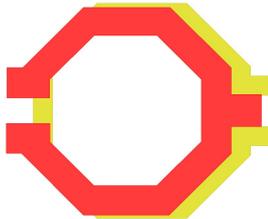
Fig. 12. Layout output balun of the final EF2 stage

VI. CONCLUSIONS AND FUTURE WORK

In this article, an outphasing PA with and without load modulation was presented and compared. When combined with ML-LINC, load modulation seems to lead to a sub-optimal design, even with ideal drivers: if no compensation reactances can be added, the supply has to be reduced to avoid V_{dg}-breakdown at moderate outphasing angles and the highest supply setting, reducing the peak output power and efficiency without a very substantial benefit in the efficiency at power back-off when compared to the ML-LINC system without load modulation. Giving up voltage margin in a rather low-voltage system, an output power reduction from 21.15 dBm to 19.99 dBm and a reduced linearity seems a rather large cost in exchange for a slower efficiency decrease and more efficiency at very low amplitudes; although this might become a significant advantage if the probability of these amplitudes is high enough. With load modulation, the DC power decreases gradually, and less DC-power is better for thermal reasons: the cooling requirements are reduced and the reliability and lifespan of the PA are most likely improved. The optimal choice will depend on the number of realisable levels and the probability density of the amplitude of the signals. In general, the ML-LINC system without load modulation seems the most attractive option, certainly when combined with AMO [4] and possibly unbalanced phase calibration [9]. As already mentioned earlier in this article, it is important to make passive components with high Q-factors when highly efficient RF circuits are needed. Hence by using a substrate with a higher resistivity ($\rho_{sub} > 10 \, \Omega$-cm) and therefore lower substrate losses, it will be possible to obtain higher efficiencies than the ones that are mentioned in this article.

REFERENCES

[1] Raab F.H., *Efficiency of Outphasing RF Power-Amplifier Systems*, IEEE Transactions on Communications, 33(10):1094–1099, 1985.
[2] Scott D. Kee; Ichiro Aoki; Ali Hajimiri; David B. Rutledge, *The Class-E/F Family of ZVS Switching Amplifiers*, IEEE Transactions on Microwave Theory and Techniques, 51(6):1677–1690, 2003.
[3] Dixian Zhao ; Kulkarni S. ; Reynaert P., *A 60-GHz Outphasing Transmitter in 40-nm CMOS*, IEEE Journal of Solid-State Circuits, 47(12):3172–3183, 2012.
[4] SungWon Chung; Godoy P.A.; Barton T.W. ; Huang E.W. ; Perreault D.J.; Dawson J.L., *Asymmetric Multilevel Outphasing Architecture for Multi-Standard Transmitters*, IEEE Radio Frequency Integrated Circuits Symposium, June 2009. RFIC 2009. p. 237-240
[5] Steve C. Cripps., *RF Power amplifiers for Wireless Communications.*
[6] Reynaert P., Steyaert M., *RF Power amplifiers for Mobile Communications.*
[7] Zisheng Li, *Analysis and Design of Highly Efficient Class-E Amplifiers for Indoor Ranging*, 2012-2013.
[8] Ichiro Aoki; Scott D. Kee; David B. Rutledge; Ali Hajimiri, *Distributed Active TransformerA New Power-Combining and Impedance-Transformation Technique*, IEEE Transactions on Microwave Theory and Techniques, 50(1):316–331, 2002.
[9] Joonhoi Hur ; Hyoungsoo Kim; Ockgoo Lee; Kwan-Woo Kim; Kyutae Lim; Bien F., *An Amplitude and Phase Mismatches Calibration Technique for the LINC-Transmitter With Unbalanced Phase Control*, IEEE Transactions on Vehicular Technology, 60(9):4184–4193, 10 Oct. 2011.

Contents

List of Figures

List of Tables

Acronyms

ADS Advanced Design System

ACPR Adjacent Channel Power Ratio

AMO Asymmetric Multilevel Outphasing

CF Crest Factor

CMRR Common Mode Rejection Ratio

DPD Digital Predistortion

EM Electromagnetic

EER Envelope Elimination and Reconstruction

EVM Error Vector Magnitude

ET Envelope Tracking

GMSK Gaussian Minimum Shift-Keying

GPDK Generic Process Design Kit

IC Integrated Circuits

LDO Low-Dropout Regulator

MOS Metal-Oxide-Semiconductor

ML-LINC Multilevel-LINC

MSE Mean Square Error

OFDM Orthogonal Frequency - Division Multiplexing

O-QPSK Offset-Quadrature-Phase-Shift-Keying

PA Power Amplifier

PAR Peak-to-Average-Ratio

PBO (Output) Power Backoff

PQF Peak Quality Factor

SRF Self Resonant Frequency

UMTS Universal Mobile Telecommunications System

W-CDMA Wideband Code Division Multiple Access

WiMax Worldwide Interoperability for Microwave Access

ZCS Zero-Current Switching

ZVS Zero-Voltage Switching

Part I

Introduction & Outphasing

Chapter 1

Introduction

The primary issue in power amplifier design is the trade-off between high efficiency and high linearity. High efficiency reduces the environmental and economical cost. Less power-to-heat conversion also reduces power supply and cooling requirements. This again reduces the power needed for cooling. For mobile users, it improves battery lifetime or battery size. High linearity allows the use of more complex modulation schemes with amplitude modulation. A trade-off exist because the classical PA efficiency decreases quickly as the desired output amplitude decreases ((Output) Power Backoff (PBO)). Depending on the probability density of the relative output amplitude, the resulting total average efficiency, which is the overlap integral of the probability density and the PA efficiency over the relative output power (cfr. figure 1.2, [1], p. 25.), can be very low; especially with signals with high Peak-to-Average-Ratio (PAR) or high Crest Factor (CF). Because of this low efficiency, the power amplifier will consume a large part of the total DC power consumption of the transmitter (cfr. figure 1.1, [2]).
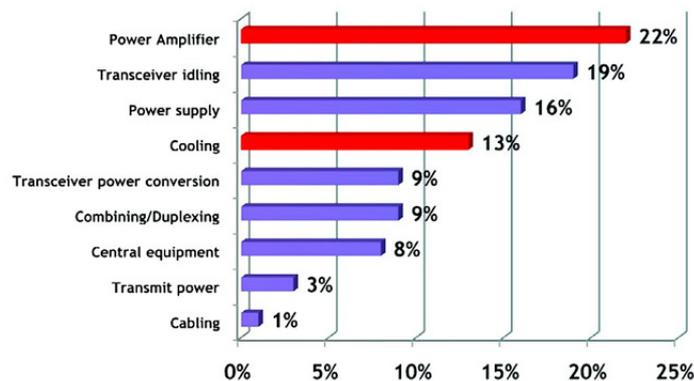
Figure 1.1: Basestation power consumption distribution.

Avoiding amplitude modulation would allow the power amplifier to constantly operate at maximum output power, where it is the most efficient. However, the constellation size, and thus the bit rate for a given bandwidth, then becomes limited by the phase resolution of the system, while the amplitude is not used at all. Furthermore, constant-amplitude symbols do not guarantee a constant-amplitude continuous time signal: the transmit pulse has to interpolate in between the symbols while staying on the unit circle. For example: 4-QAM filtered with a GMSK-pulse has a constant envelope, while 4-QAM with a square-root raised cosine pulse does not. In Offset-Quadrature-Phase-Shift-Keying (O-QPSK), the amplitude fluctuations are reduced by delaying the bit stream for e.g. the quadrature component by one bit period. This
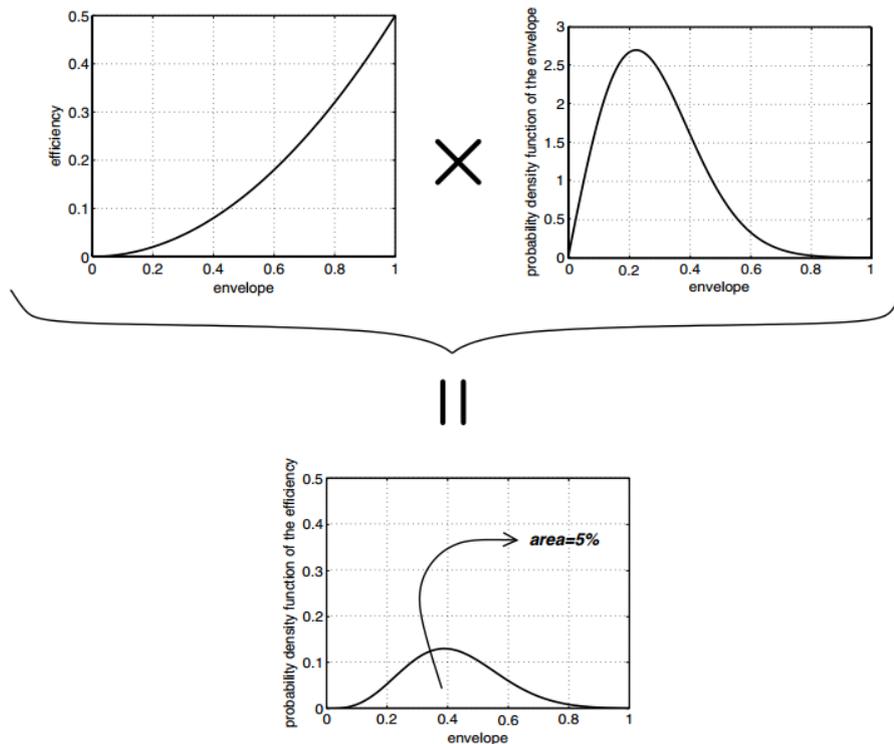
Figure 1.2: Typical overall PA efficiency example.

guarantees that only one bit changes per bit period. Because the symbols are Gray-mapped, all symbol transitions are forced to occur via an adjacent symbol. The amplitude fluctuations will be reduced even when filtered with a square-root raised cosine pulse, but some amplitude modulation still remains. With growing constellation sizes, amplitude modulation becomes inevitable. High PAR/CF-signals are being used today in W-CDMA (UMTS, 3G), CDMA2000 (3G) ... The use of OFDM, e.g. in WiMax (4G), LTE (4G) ..., offers a high spectral efficiency and a higher robustness against fading channels, but also implies a high PAR.

Several solutions to the fundamental trade-off exist:

- PAR/CF-reduction, Digital Predistortion (DPD):
  The communication standards could be adapted to produce signals with low PAR/CF. A low PAR implies that the PA can be driven into its saturation region, where it is most efficient, for a large fraction of the time. In and near this saturation region, the input-output relation is very non-linear, so the input of the PA has to be predistorted with the inverse non-linear function to obtain a linear output (DPD).

- Envelope Elimination and Reconstruction (EER) (Kahn technique), Envelope Tracking (ET), polar transmitter:
  These technologies are not identical, but the basic concept is similar: the amplitude modulation is separated from the baseband signal, resulting in two signal paths: a high-frequency phase-modulated carrier, and a much lower-frequency amplitude signal. This AM-signal is amplified by a "linear" low-frequency PA. The linearity-vs-efficiency trade-off is now pushed to lower frequencies, into the baseband instead of the passband, where a better overall solution should be found. The amplified AM-signal is used as the supply of a (switching) non-linear high-frequency PA which amplifies the phase-modulated carrier.

Essentially, the supply voltage of this PA is adjusted according to the desired instantaneous output amplitude. A disadvantage is that the amplitude and phase path will have to be matched very well to obtain the correct output; but path matching requirements are inherent to any architecture in which the signal is split. The distortion of the amplifier, e.g. the AM-PM and PM-PM function, are dependent on the supply, so the distortion-corrections have to be output amplitude-dependent as well. A very important issue is that the bandwidth of the amplitude is significantly larger than the bandwidth of the amplitude- and phase-modulated signal ([1], p. 159). The overall bandwidth could be restricted, e.g. max. 20-40 MHz ([3]), by the bandwidth of the amplitude modulator: the supply modulation is often done with a block similar to an Low-Dropout Regulator (LDO) ([1], p. 198).

- Doherty amplifier:
  This technique is primarily used in e.g. base stations and is based on the cooperation of a class AB and a class C PA, depending on the input level. At high input levels, both PA's, but mainly the peaking class C PA, contribute to a high output power. The class C PA is shut down gradually when the input level decreases. This is clearly an advantage: the class AB PA does not have to deliver the maximal total output power, so it can enter its saturation region well before the maximal total output power is reached, which means that the total efficiency at output power backoff can be increased. In the transition region from intermediate to high input levels, the class AB PA-gain drops because it is heavily saturated, but the class C PA-gain increases. To make this transition region linear, a non-isolating power combiner is used; resulting in "load pulling" or "load modulation", where the load seen by one PA is modified by the other PA. This load modulation principle results in higher efficiency and linearity ([4], p. 290).

- Outphasing:
  The outphasing power amplifier is also based on the separation of AM and PM and optionally also uses load modulation. In the basic outphasing PA, the choice seems straightforward since power combination through load modulation is the most efficient option (cfr. next paragraph).

- Asymmetric Multilevel Outphasing (AMO):
  AMO is an advanced outphasing architecture ([5]) which combines normal outphasing with power supply control and asymmetric outphasing vectors (different outphasing angles) to improve efficiency. AMO is further discussed in section 2.3.

The concept of outphasing is to split a signal which contains both amplitude and phase modulation into two constant-amplitude signals with a common and differential phase modulation. These constant-amplitude signals are amplified separately and recombined at the output; where the amplitude modulation is reconstructed. The input signal at the PA's does not contain amplitude modulation, so the PA's can be non-linear or switching PA's, where there is no direct or a very non-linear relationship between the input and output amplitude. The PA's could work at maximum output power and thus maximum efficiency, however this does not guarantee a high overall efficiency and it strongly depends on the type of combiner.

If an isolating combiner is used, the load seen by each PA is by definition constant and independent of the instantaneous output amplitude. This implies that both PA's deliver a constant amount of power and all the power that is not desired at the output will be dissipated in the combiner (e.g. in the isolation resistor if a Wilkinson combiner is used). This is clearly not efficient at all. Three solutions exist: inject the unnecessary output power back into the supply (outphasing energy recovering amplifier (OPERA), as described in e.g. [6]); use a non-isolating combiner, which is equivalent to load modulation (cfr. section 2.2); or avoid large outphasing angles in general, such as in the Multilevel-LINC (ML-LINC) systems (cfr. section 2.3) and AMO. In this thesis, we will focus on normal outphasing with load modulation first, and add

ML-LINC later. AMO could be part of future work.

This thesis focusses on the design of an outphasing power amplifier in the "generic" 45 nm Generic Process Design Kit (GPDK) provided by Cadence. The specifications are:

- Carrier frequency: 15 GHz

- 256-QAM with a minimal Error Vector Magnitude (EVM)

- At least 20 dBm output power

- At least 50 % peak overall efficiency

- The data rate was not strictly specified. The maximal data rate will be determined by the total bandwidth, which will most probably be restricted due to the use of harmonic resonators in the driver stages and PA output stage. For the theoretical simulations, we chose a symbol rate of 100 MBaud or 800 Mbit/s (cfr. section 2.4). In most transmitter systems; and certainly in this case, with a 15 GHz carrier; the relative bandwidth is small. Because the quality factors of the resonators are limited, the PA characteristics should not differ too much for frequencies that are relatively close to the carrier frequency. If necessary, equalisation could be used to improve the effective bandwidth and datarate.

- Two 0 dBm constant-amplitude, phase-modulated inputs are given.

- Temperature range: 0-80 °C

The work in this thesis is divided into multiple parts:

- Theoretical simulations of the outphasing system and load modulation

- Setup of a testbench in Cadence to determine estimates of EVM, Adjacent Channel Power Ratio (ACPR) ...

- Technology characterisation (passive and active)

- Choice of a PA class and system topology

- PA design: active and passive components

- Driver design

This work was assigned to two students: Laurens focussed on the characterisation of the substrate and design of passive components and combiners; Joris did the theoretical simulations, the characterisation of the active components and the PA and driver design. The same work distribution was used for writing this thesis.

Chapter 2

Outphasing

2.1 Outphasing concept
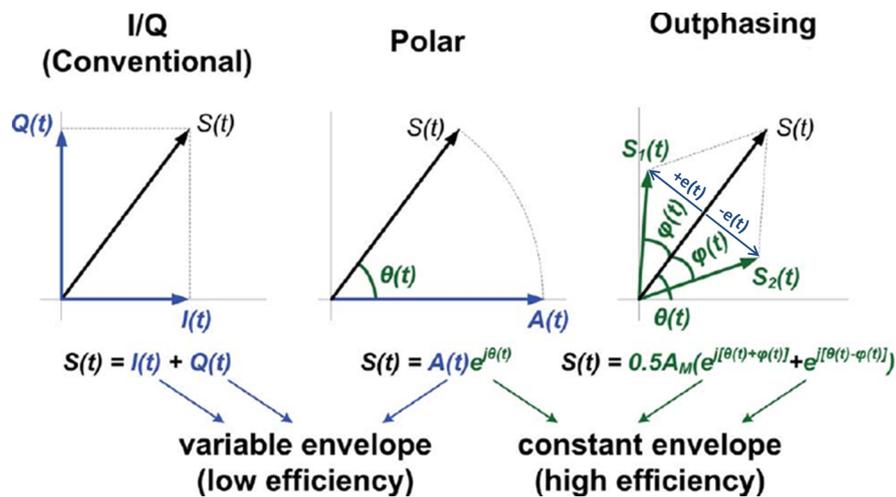
Figure 2.1: Outphasing concept.

The baseband signal $s(t) = I(t) + jQ(t) = A(t)e^{j\theta(t)}$ is split into two signals $s_1(t)$ and $s_2(t)$ with constant amplitude A_{max} (cfr. figure 2.1, adapted from [7]):

$$s_1(t) = A_{max}e^{j\theta(t)}e^{j\phi(t)} \tag{2.1}$$

$$s_2(t) = A_{max}e^{j\theta(t)}e^{-j\phi(t)} \tag{2.2}$$

At the output, $s_1(t)$ and $s_2(t)$ are added, and we obtain:

$$s_{out}(t) = A_{max}e^{j\theta(t)}[e^{j\phi(t)} + e^{-j\phi(t)}] = 2cos(\phi(t))A_{max}e^{j\theta(t)} \tag{2.3}$$

To reconstruct the amplitude modulation with a gain factor of 2, $\phi(t)$ has to be chosen to make sure that $cos(\phi(t)) \cdot A_{max} = A(t)$, so:

$$\phi(t) = arccos(\frac{A(t)}{A_{max}}) \tag{2.4}$$

In figure 2.1, we can see that $s_1(t)$ and $s_2(t)$ could also be written as the sum of $s(t)$ with an orthogonal "error" vector $e(t)$. $e(t)$ has an amplitude equal to $\sqrt{A_{max}^2 - A(t)^2}$ to ensure that $|s_1(t)| = |s_2(t)| = A_{max}$:

$$s_1(t) = s(t) + e(t) \tag{2.5}$$

$$s_2(t) = s(t) - e(t) \tag{2.6}$$

$$e(t) = A_{max}\sqrt{1 - \frac{A(t)^2}{A_{max}^2}}e^{j\theta(t)}e^{j\frac{\pi}{2}} \tag{2.7}$$

By setting (2.1) equal to (2.5) and solving for the real and imaginary parts, we obtain that both representations are equivalent if (2.4) is satisfied. The representation of (2.5) will be convenient for the explanation of scatterplots, spectra etc.; because it clearly separates the linear from the non-linear effects. In the representation of (2.1), the non-linearity of the arccos()-function is "hidden" in the phase factor $e^{j\phi(t)} = e^{j \cdot arccos(\frac{A(t)}{A_{max}})}$.

The decomposition of $s(t)$ into $s_1(t)$ and $s_2(t)$ is done by a signal component separator (SCS) block. The realisation of this block is outside of the scope of this thesis, a behavioral model will be used in the simulations. The implementations are often digital ([8], [9]), but analog ([10]) and mixed-signal SCS-designs also exist, which simultaneously offer high accuracy, very high speed and high power efficiency (e.g. 12-bit, 3.4 GS/s in [11]). Predistortion, equalisation and calibration can be integrated with the SCS.

2.2 Load modulation, reactance compensation

2.2.1 Differential vs. common mode combining

In the introduction, it was concluded that load modulation seems the most efficient way to combine the constant-amplitude signals $s_1(t)$ and $s_2(t)$. The goal of load modulation is to make the load seen by each PA dependent on the desired output amplitude. This is achieved by making the effective load, seen by one PA, dependent on the action of the other PA, so load modulation is only possible with a non-isolating power combiner.

In figure 2.2 (adapted from [12]), a common-mode combiner is used: the PA's are assumed to be (close to) ideal voltage generators, and their very low output impedance is inverted by the $\frac{\lambda}{4}$-transmission lines; transforming the PA's into current generators. The output currents are summed at the output. The common components in the currents are added and pushed into the load. The differential components do not flow through the load. The definitions of section 2.1 are directly compatible with common-mode combination. In our case, this type of combiner is not suitable on chip, since it is based on $\frac{\lambda}{4}$-transmission lines. This inherently leads to a more narrowband combiner. With the given substrate parameters, $\frac{\lambda}{4}$-TL's are not favourable on-chip and will be avoided when possible: at 15 GHz, they are still quite long ($\epsilon_r = 3.5 \rightarrow \frac{\lambda}{4}$=2.67 mm) and they suffer from high losses due to both the substrate ($tan(\delta) = 0.1$ at 15 GHz) and conductor losses (copper conductivity of only $2 \cdot 10^7 S/m$ combined with skin effect).

In figure 2.3, a differential load is used, the output power is proportional to $V_1(t) - V_2(t)$. To obtain a single-ended output, a transformer can be used, which should be quite broadband. We

can continue to use the same definitions as for common-mode combination by using $-s_2$ instead of s_2 at the input of the second PA.

The phase factor $e^{j\theta(t)}$ is common to s_1 and s_2, so we can temporarily set $\theta(t)$ to zero to look at the load modulation. If s_1 and $-s_2$ are perfectly in phase (the intended output amplitude is zero, $|e(t)| = A_{max}$); there is no voltage over the differential load so no current flows through it and both power amplifiers "see" an "open". If s_1 and $-s_2$ are out of phase ($|e(t)| = 0$) the maximal differential voltage is present over the load and maximal current flows. Both PAs see a load equal to $\frac{Z_{diff.}}{2}$. The phasors s_1 and s_2 add in phase and the maximal output amplitude is reached.
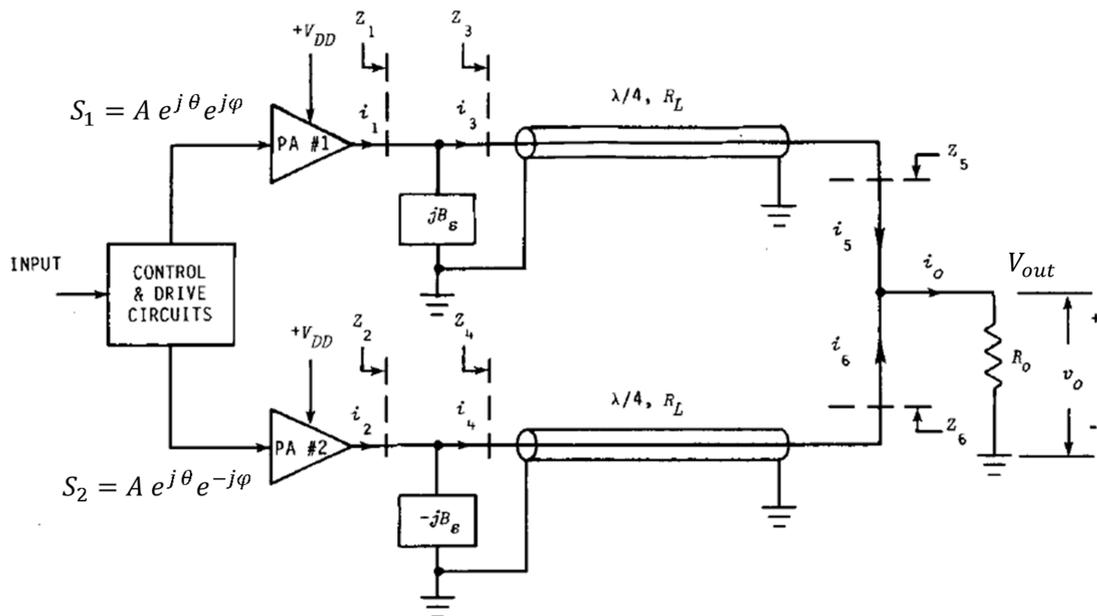
Figure 2.2: Outphasing (common mode) with a transmission line combiner.
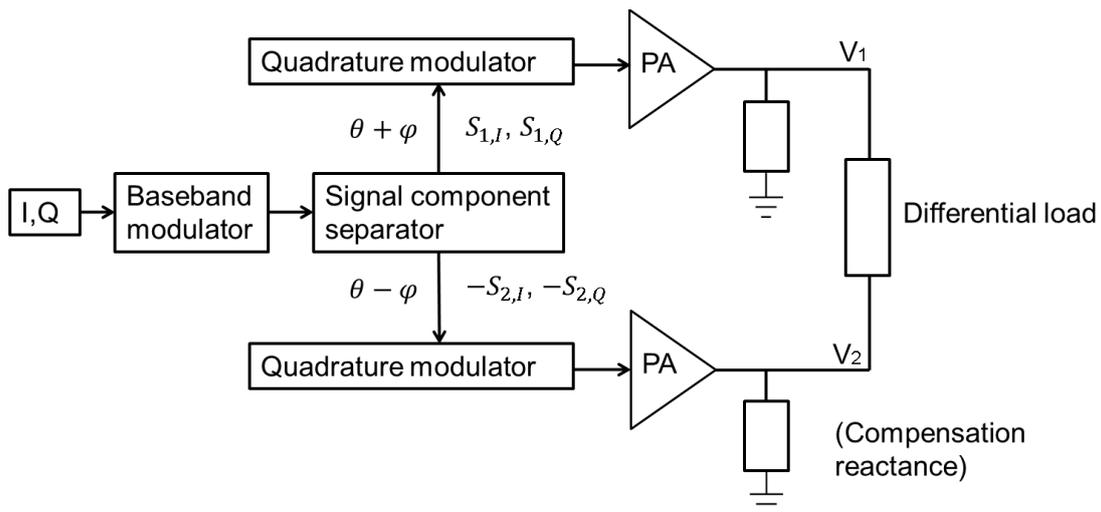
Figure 2.3: Outphasing (differential) block schematic.

For outphasing angles different from $\phi = 0$, the load seen by each PA is not real but complex. A reactance results from setting $V_{out,1} = Ae^{j\phi(t)}$ and $V_{out,2} = -Ae^{-j\phi(t)}$. The PA's deliver large output signals, but we still assume that they can be described as voltage generators with a finite output impedance. Output impedance is a small-signal concept, and in reality, the effective "output impedance" is time-variant and non-linear (so it should be measured at each harmonic of the fundamental frequency). Mismatch between the PA's and some asymmetry in the combiner will occur, so the effective output impedances (seen at the load) are not necessarily equal either. Still, this approximation is often considered reasonable and sometimes the output impedance is even omitted completely; because the PA is assumed to be heavily saturated and operating with "rail-to-rail" voltage swing. ([4], p.306). The linearity of the output impedances will also have an influence on the output signal itself: in the common-mode combiner, a non-linear series output impedance is transformed into a non-linear parallel impedance by the $\frac{\lambda}{4}$-transmission lines, and will form a current divider with the common-mode load. The common-mode output current will become a non-linear function of the input current. In the differential combiner, a non-linear voltage divider is formed directly by putting the load in series with the PA. In both cases, the output voltage will become a non-linear function of the input voltages, and distortion is introduced.

Some theoretical simulations and calculations were done to demonstrate the principle of outphasing and load modulation. In figure 2.4, differential combining with a transformer is assumed. The common-mode combining is demonstrated in section 2.2.2.
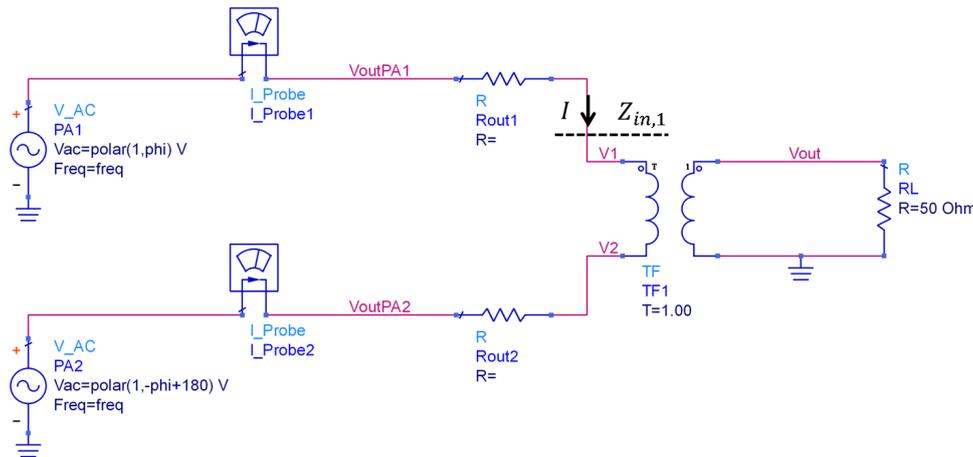
Figure 2.4: Idealized load modulation with a transformer.

In the schematic of figure 2.4, we assume that $V_{out,1} = e^{j\phi(t)}$, $V_{out,2} = -e^{-j\phi(t)}$, $Z_{in,1} = \frac{V_1}{I}$.

$$I = \frac{V_{out,1} - V_{out,2}}{Z_L + Z_{out,1} + Z_{out,2}} \tag{2.8}$$

$$Z_{in,1} = \frac{\frac{Z_L + Z_{out,2}}{Z_L + Z_{out,1} + Z_{out,2}}(V_{out,1} - V_{out,2}) + V_{out,2}}{I} = Z_L + Z_{out,2} + \frac{V_{out,2}}{V_{out,1} - V_{out,2}}(Z_L + Z_{out,1} + Z_{out,2}) \tag{2.9}$$

$$\frac{V_{out,2}}{V_{out,1} - V_{out,2}} = \frac{-e^{-j\phi(t)}}{e^{j\phi(t)} + e^{-j\phi(t)}} = \frac{-e^{-j\phi(t)}}{2cos(\phi(t))} = -\frac{1}{2}(1 - jtan(\phi(t))) \tag{2.10}$$

If we now assume that the output impedances are both equal to Z_{out}:

$$Z_{in,1} = Z_L + Z_{out} - \frac{1}{2}(1 - j \cdot tan(\phi))(2Z_{out} + Z_L) = \frac{Z_L}{2}[1 + j \cdot tan(\phi)(1 + 2\frac{Z_{out}}{Z_L})] \quad (2.11)$$

$V_{out,1}$ and $V_{out,2}$ are opposite and complex conjugate and the PA current flows in the opposite direction, so the input impedance seen by the second PA is complex conjugate:

$$Z_{in,2} = \frac{Z_L}{2}[1 - j \cdot tan(\phi)(1 + 2\frac{Z_{out}}{Z_L})] \quad (2.12)$$

If the output amplitude is decreased by increasing the outphasing angle, the input impedances become increasingly reactive (inductive at terminal 1, capacitive at terminal 2). With $A = (1 + 2\frac{Z_{out}}{Z_L})$, the admittance seen by the first PA is:

$$Y_{in,1} = \frac{1}{\frac{Z_L}{2}[1 + jA \cdot tan(\phi)]} = \frac{1 - jA \cdot tan(\phi)}{\frac{Z_L}{2}(1 + A^2 \tan(\phi)^2)} \quad (2.13)$$

$$\angle Y_{in,1} = -\angle Z_{in,1} = -atan(A \cdot tan(\phi)) \quad (2.14)$$

In the case of ideal voltage generators ($Z_{out} = 0$ and $A = 1$, with $\phi \in [0, \frac{\pi}{2}]$ and Z_{out}, Z_L real):

$$\angle Y_{in,1} = \phi(t) \quad (2.15)$$

To be able to provide a theoretical example, we can assume e.g. an ideal class B-PA. P_{out,f_1} is the output power at the fundamental frequency delivered by PA_1 to the load, I_1 is the fundamental output current. If we assume that the output voltage amplitude at the fundamental frequency f_1 does not change, and is still equal to V_{dc} in the ideal case, then the output power is proportional to the real part of the admittance of the load:

$$P_{out,f_1} = \frac{1}{2}\Re(V_1 I_1^*) = \frac{1}{2}\Re(Y_{in,1}^*)V_{dc}^2 \quad (2.16)$$

For the class B-PA, the DC current is related to the fundamental output current (and thus to the load admittance) by:

$$I_{dc} = \frac{2}{\pi}|I_1| \quad (2.17)$$

$$P_{dc} = V_{dc}\frac{2}{\pi}|I_1| = \frac{2}{\pi}V_{dc}^2|Y_{in,1}| \quad (2.18)$$

The drain efficiency becomes:

$$\eta_d = \frac{P_{out,f_1}}{P_{dc}} = \frac{\pi}{4}\frac{\Re(Y_{in,1})}{|Y_{in,1}|} = \eta_{d,ideal,classB} \cdot cos(\angle Y_{in,1})) \quad (2.19)$$

Because $\phi(t) = arccos(\frac{A(t)}{A_{max}})$, the PBO-drain efficiency becomes linear in A(t) (as demonstrated in figure 2.6) if $Z_{out} = 0$, $A = 1$:

$$\eta_d = \frac{\pi}{4} \frac{A(t)}{A_{max}} \tag{2.20}$$

From equation 2.13, a parallel equivalent for the input impedance can be found:

$$Y_{in,1} = \frac{1}{R_{parallel,in,1}} + \frac{1}{jX_{parallel,in,1}} \tag{2.21}$$

$$R_{parallel,in,1} = \frac{Z_L}{2}(1 + A^2 \tan(\phi)^2) \tag{2.22}$$

$$X_{parallel,in,1} = \frac{\frac{Z_L}{2}(1 + A^2 \tan(\phi)^2)}{A tan(\phi)} \tag{2.23}$$

$$R_{parallel,in,1} = R_{parallel,in,2}, X_{parallel,in,2} = -X_{parallel,in,1} \tag{2.24}$$

If $Z_{out} = 0$ and $A = 1$ and $Z_L = R_L$, we obtain:

$$R_{parallel,in,1} = \frac{R_L}{2cos(\phi)^2} \tag{2.25}$$

$$X_{parallel,in,1} = \frac{R_L}{sin(2\phi)} \tag{2.26}$$

The imaginary part of the input admittance is given by (with $Z_{out} = R_{out}$, $Z_L = R_L$):

$$\Im(Y_{in,1}) = \frac{-A tan(\phi)}{\frac{R_L}{2}(1 + A^2 \tan(\phi)^2)} = \frac{-A}{R_L} \frac{sin(2\phi)}{cos(\phi)^2 + A^2 sin(\phi)^2} \tag{2.27}$$

We can maximize the drain efficiency $\eta_d = \eta_{d,ideal,classB} \cdot cos(\angle Y_{in,1}))$ of the PA at a given outphasing angle ϕ_c by resonating out the parallel reactance. The load modulation is still present in parallel resistive load, so this compensation will not interfere with reconstruction of the output amplitude. $\Im(Y_{in,1})$ is negative (inductive) for every ϕ, so a parallel compensation capacitor is added. $\Im(Y_{in,2})$ is positive (capacitive), so a parallel inductor is necessary. In the case of zero output impedance, $A = 1$, these compensation reactances can be calculated directly from the input admittance:

$$C_c = \frac{-\Im(Y_{in,1(\phi_c)})}{\omega} = \frac{sin(2\phi)}{\omega R_L} \tag{2.28}$$

$$L_c = \frac{1}{\omega \Im(Y_{in,2(\phi_c)})} = \frac{R_L}{\omega sin(2\phi)} \tag{2.29}$$

If $A = 1$, the $\Im(Y_{in,1})$ is symmetrical in ϕ around $\frac{\pi}{4}$, so $\Im(Y_{in,1})$ will also become zero at $\frac{\pi}{2} - \phi_c$. At these outphasing angles, two efficiency maxima occur. In between, the efficiency drops. Increasing ϕ_c brings the efficiency peaks closer together. Given the probability density of the output signal, expressed as a function of ϕ; we could obtain the optimal ϕ_c to maximize the overall efficiency by maximizing the overlap integral as a function of ϕ_c.
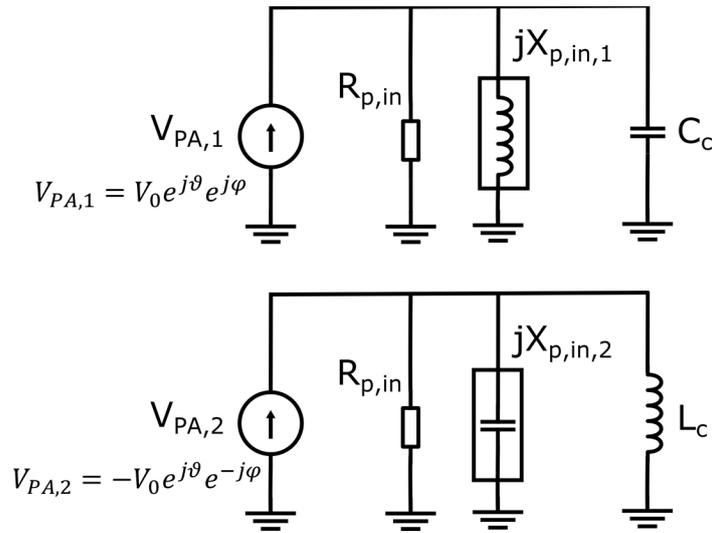
Figure 2.5: The effect of load modulation, seen by each PA, with the compensation reactances.

At 15 GHz, with $R_L = 50\Omega$, we obtain:

ϕ (°)	C_c (fF)	L_c (nH)
15	106.1033	1.0610
25	162.5597	0.6925
30	183.7763	0.6126
35	199.4090	0.5646
40	208.9827	0.5387
45	212.2066	0.5305

Table 2.1: C_c and L_c at various outphasing angles ϕ_c.

In this case, C_c can be made on chip for all ϕ_c, L_c is too large at very low ϕ_c; but almost realisable in the relevant ϕ_c-range (e.g. $\phi_c = 25°$). In general, an impedance transformation (e.g. with a transformer or an LC-matching network) could be done to bring both C_c and L_c into an acceptable range by scaling $R_{L,effective}$. Figure 2.7 illustrates the obtained efficiency when $Z_{out} = 0$, figure 2.8 is a similar characteristic obtained from ([4], p.308).
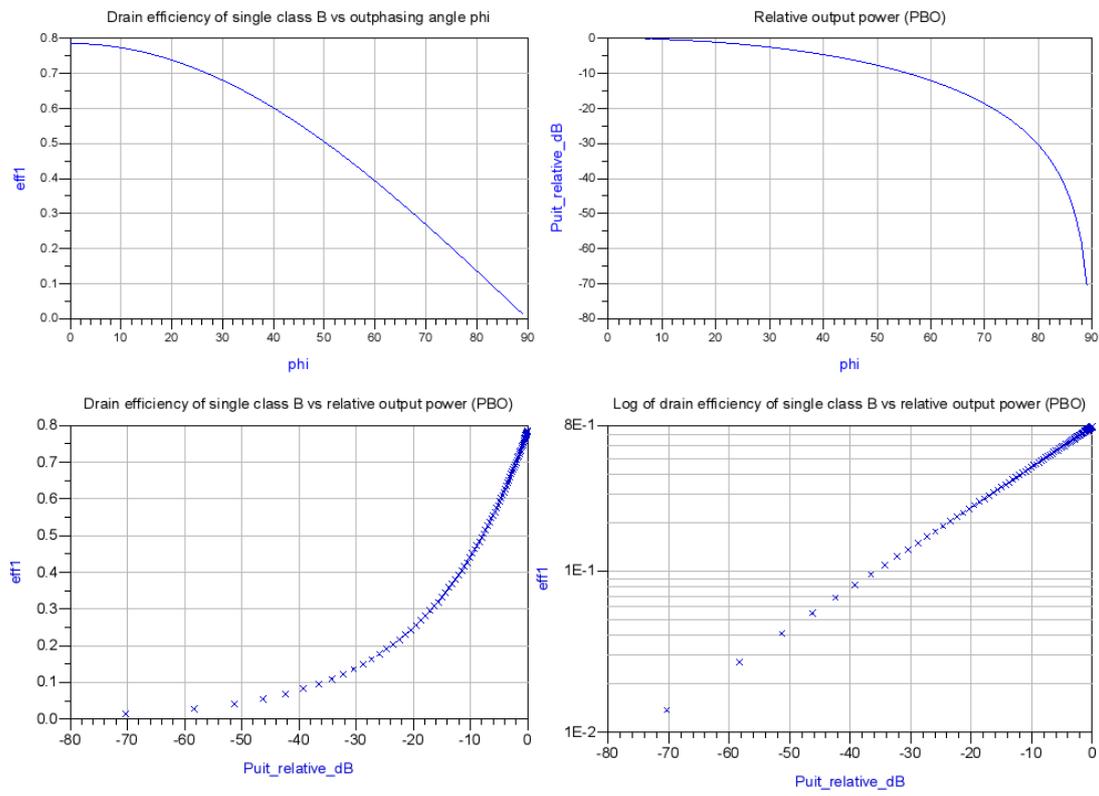
Figure 2.6: Simulated output power and efficiency in case of idealized class B PA's (ideal voltage generators), without compensation reactances.
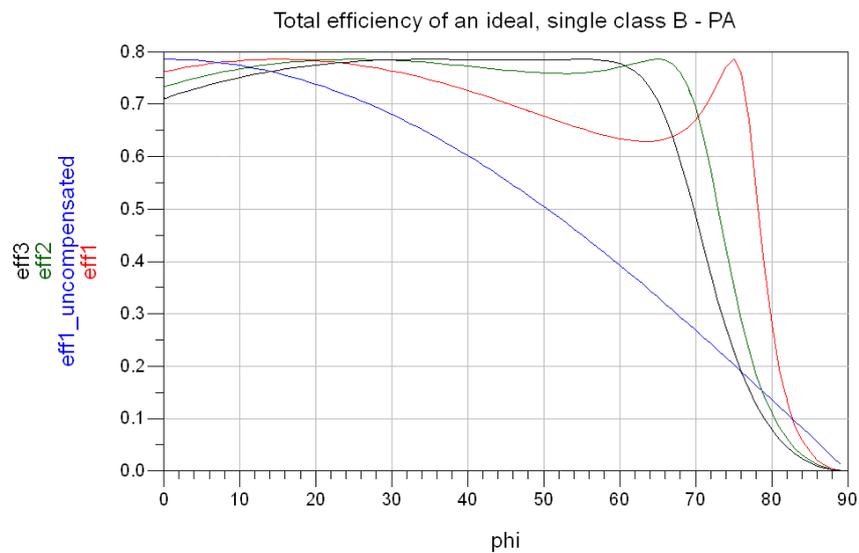
Figure 2.7: Efficiency comparison ($Z_0 = 0\Omega$) before and after compensation at $\phi_c = 15°$ (eff1), $\phi_c = 25°$ (eff2) and $\phi_c = 35°$ (eff3). To compare with figure 2.8, a small outphasing angle ϕ corresponds with a high output power and a less negative power backoff.
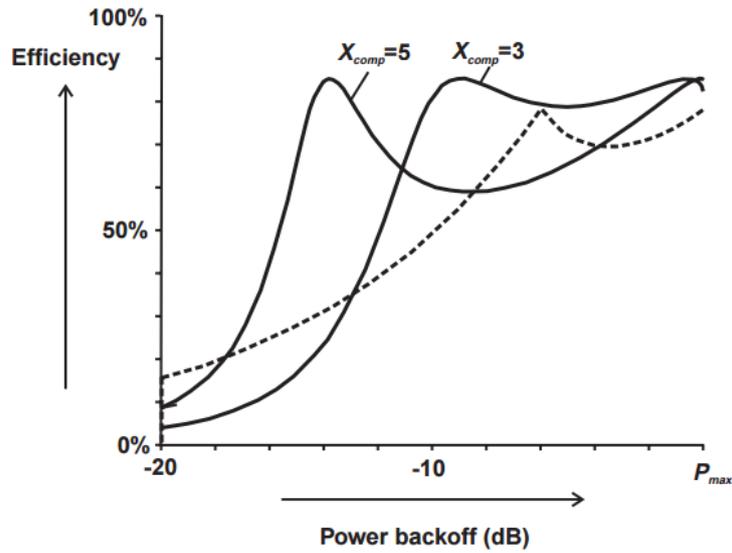
Figure 2.8: Typical PBO-characteristic in compensated outphasing PA with $R_{out} = 0$ Ohm.
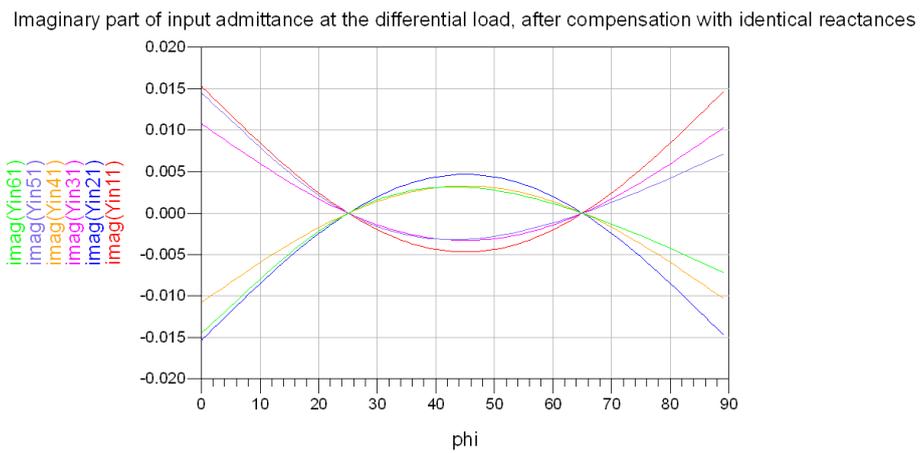
Figure 2.9: $\Im(Y_{in,1})$, $\Im(Y_{in,2})$ at the load for $Z_0 = 0\Omega$ ($Y_{in,11}$,$Y_{in,21}$), $Z_0 = 10\Omega$ ($Y_{in,31}$,$Y_{in,41}$) and $Z_0 = 25\Omega$ ($Y_{in,51}$,$Y_{in,61}$) when compensated at $\phi_c = 25°$.
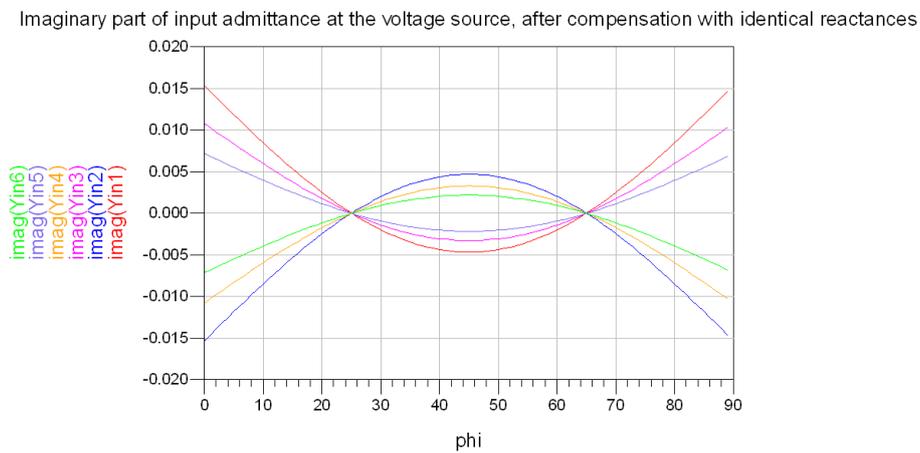
Figure 2.10: $\Im(Y_{in,1})$, $\Im(Y_{in,2})$ at the sources, for $Z_0 = 0\Omega$ ($Y_{in,11}$,$Y_{in,21}$), $Z_0 = 10\Omega$ ($Y_{in,31}$,$Y_{in,41}$) and $Z_0 = 25\Omega$ ($Y_{in,51}$,$Y_{in,61}$) when compensated at $\phi_c = 25°$.
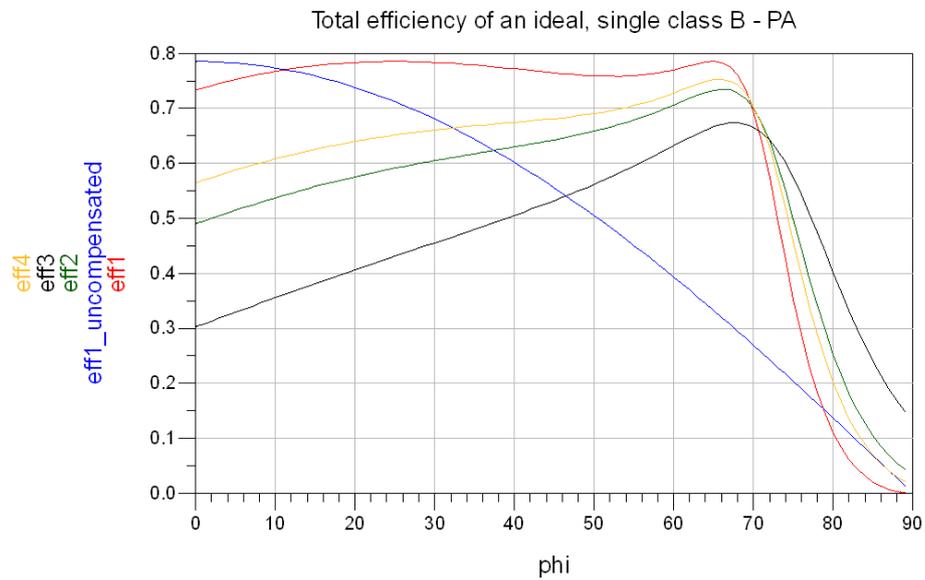
Figure 2.11: Efficiency comparison before and after compensation at $\phi_c = 25°$ for $R_L = 50\Omega$ and $Z_0 = 0\Omega$ (eff1), $Z_0 = 10\Omega$ (eff2) and $Z_0 = 25\Omega$ (eff3); and for $R_L = 200\Omega$ and $Z_0 = 25\Omega$ (eff4).
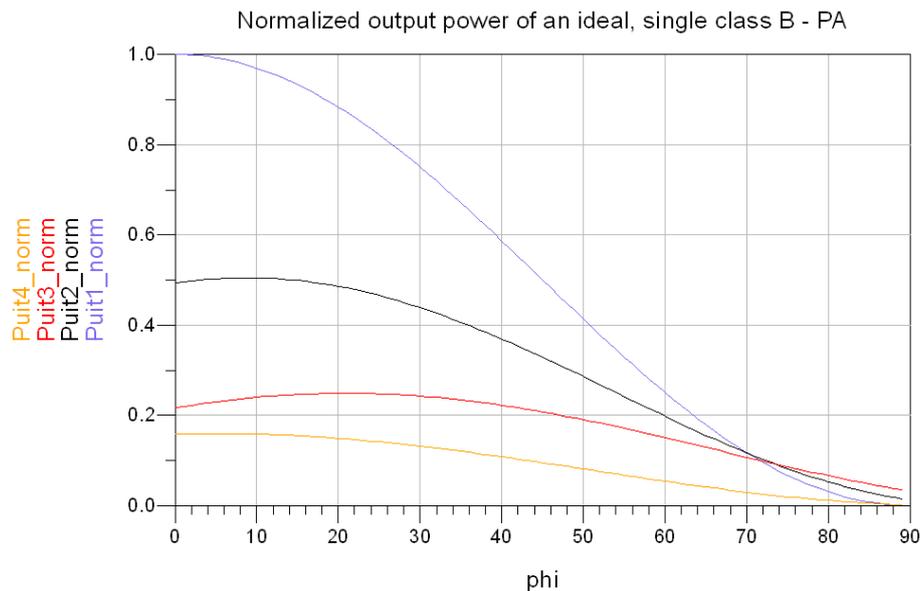
Figure 2.12: Output power when $Z_0 = 0\Omega$ ($P_{out,norm,1}$), $Z_0 = 10\Omega$ ($P_{out,norm,2}$) and $Z_0 = 25\Omega$ ($P_{out,norm,3}$), compensated at $\phi_c = 25°$; and for $R_L = 200\Omega$ and $Z_0 = 25\Omega$ ($P_{out,norm,4}$). The output power is not always strictly inversely proportional to ϕ any more, which will introduce distortion.

When $A = 1$, the PA's are ideal voltage sources, so the compensation reactances are independent: the input admittance at one PA is not influenced by the compensation reactance at the other PA. In general, this is no longer the case if the output impedance is non-zero. We add the output impedances (to ground) and the two unknown compensation impedances $Z_{c,1}$ and $Z_{c,2}$ into the schematic. By forcing the imaginary part of the input admittance to be zero at both PA-outputs simultaneously at $\phi = \phi_c$, we obtain two equations for $Z_{c,1}$ and $Z_{c,2}$.

If we assume the output impedances of both PA's to be equal, the circuit remains symmetrical and it is driven by (opposite) complex conjugate sources, so the complex conjugate symmetry remains as well. This eliminates one unknown: $Z_{c,2} = Z_{c,1}^*$. The goal is to present a real total load to the ideal voltage source "inside" the PA (as in the case when $Z_0 = 0$); but if we assume that the output impedance is real, this is equivalent to a real load at each PA output (so after the output impedance). Intuitively, we expect that the compensation reactances will not change in this special (but acceptable) case. This is confirmed by the equations and simulations (cfr. figure 2.9).
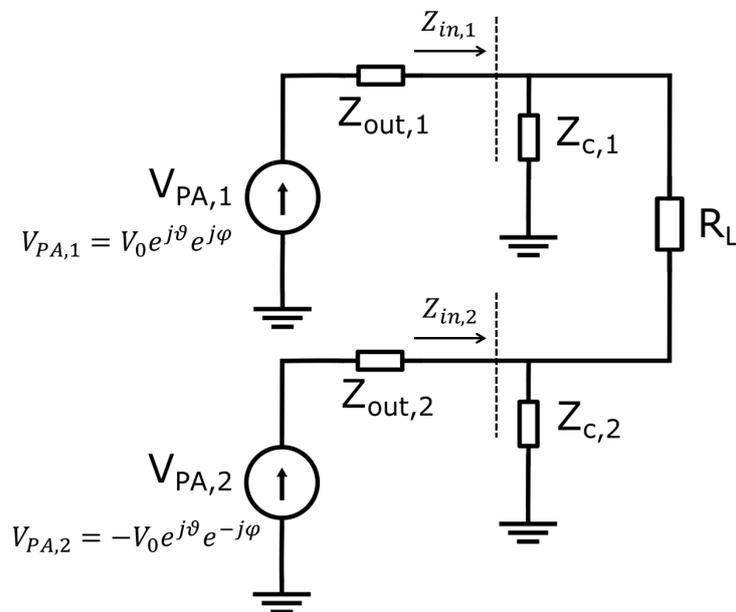
Figure 2.13: Modified schematic with output and compensation impedances.

When setting $Z_{c,2} = Z_{c,1}^*$ and assuming that $Z_{out} = R_0$, we obtain that $\Im(Y_{in,1})$ depends on R_0 but C_c does not:

$$\Im(Y_{in,1}) = \frac{\frac{R_L}{\omega C_c} - \frac{sin(2\phi)}{\omega^2 C_c^2}}{R_0^2 R_L + \frac{R_L + 2R_0}{\omega^2 C_c^2}} \rightarrow C_c = \frac{sin(2\phi)}{\omega R_L} \tag{2.30}$$

This implies that R_0 does not have to be known to provide C_c and L_c, so we know the optimal compensation reactance as long as the output impedance is real and identical for both PA's. This is demonstrated in figures 2.9 and 2.11. Increasing the output impedance will degrade the efficiency and output power: e.g. at $\phi = \phi_c$, the input admittance is real and a voltage divider is is formed between R_0 and $R_{parallel,in,1}$. Because $R_{parallel,in,1} = \frac{R_L}{2}(1 + A^2 tan(\phi)^2)$ is low at low ϕ, this effect dominates the efficiency increase that was gained by making $\Im(Y_{in}) = 0$. The efficiency becomes asymmetric in ϕ: at $\phi = 90° - \phi_c$, $\Im(Y_{in})$ is also zero but $R_{parallel,in,1}$ is much higher, so the efficiency stays high. An impedance transformation (e.g. with an 2:1-transformer)

can be used to increase $\frac{R_{L,eff}}{R_{L,eff}+R_{out}}$ and improve the efficiency ("eff4" in figure 2.11), especially at low ϕ, but because we assume ideal voltage generators, the output power increase inversely proportional with this impedance scaling ("$Pout4_{norm}$" in figure 2.12). C_c, L_c have to be scaled as well, resulting in unrealisable values in this case ($C_c \propto \frac{1}{R_{L,eff}}$, $L_c \propto R_{L,eff}$).

If the output impedance is non-linear, it could be specified at each harmonic, but it will in general not be possible to make the admittance real at each harmonic with only one discrete component C_c or L_c. In the theoretical model here, the voltage waveform is constant since it is fixed by an ideal voltage source, so if this voltage waveform contains harmonics and the load seen at these harmonics is not real, the efficiency drops, so more DC power will be consumed. The power in the harmonics is not useful output power, so the overall efficiency will degrade. To provide the correct compensation at a fixed ϕ_c but at different harmonics, a more complex network will be necessary.

If the output impedance at a given frequency is complex, $Z_0 = R_0 + jX_0$, the ideal load is not purely resistive because it should compensate for the output reactance, so $Z_{L,opt} = (R_{L,opt} - R_0) - jX_0$ to obtain a purely resistive load $R_{L,opt}$ at the ideal voltage source. Assuming that the output impedances are known but still identical for both PA's, $Z_{c,2} = Z_{c,1}^*$ should still hold and the same equations could be solved for the real and imaginary part of $Z_{c,1}$.

The effect of the compensation reactances on the efficiency and output power clearly depends on the value of the output impedances; which illustrates that their final effect cannot be well predicted and has to be determined by simulation.

2.2.2 Common mode combining

For completeness, we consider the admittances seen by the PA's in the common mode combiner, in the case of ideal voltage sources ([12]), with the same definitions for the PA inputs $s_1(t)$ and $s_2(t)$. We consider only the top path, since the circuit is symmetrical up to a complex conjugation. The ideal voltage generators with zero (or low) output impedance are transformed into (almost) ideal current generators with a very high output impedance, so the output currents are forced to flow through the load:

$$i_5 = I_m e^{j\phi} \tag{2.31}$$

$$i_6 = I_m e^{-j\phi} \tag{2.32}$$

$$V_{out} = 2R_L I_m cos(\phi) \tag{2.33}$$

$$Z_5 = \frac{V_{out}}{i_5} = 2R_L cos(\phi)e^{-j\phi} \tag{2.34}$$

The quarter-wavelength transmission lines with characteristic impedance Z_0 are assumed to be ideal:

$$Z_{in,1,CM} = \frac{Z_0^2}{Z_5} = \frac{Z_0^2}{2R_L cos(\phi)}e^{+j\phi} \tag{2.35}$$

$$Y_{in,1,CM} = \frac{2R_L}{Z_0^2}[cos(\phi)^2 - j\frac{sin(2\phi)}{2}] \tag{2.36}$$

The real part of the input admittance is modulated by ϕ and decreases to zero when ϕ goes to $\frac{\pi}{2}$, so the load modulation is present in $\Re(Y_{in,1,CM})$. Therefore, the negative imaginary part of the input admittance can be compensated without distorting the output, by a adding a parallel capacitor:

$$C_{c,CM} = \frac{sin(2\phi)R_L}{\omega Z_0^2} \tag{2.37}$$

Because of the complex conjugate symmetry, we also obtain:

$$L_{c,CM} = \frac{Z_0^2}{\omega sin(2\phi)R_L} \tag{2.38}$$

Comparing with the case of the differential combiner when $A = 1$ ($Z_{out} = 0$) gives:

$$Y_{in,1,DM} = \frac{cos(\phi)^2}{\frac{R_L}{2}} - j\frac{sin(2\phi)}{R_L} \tag{2.39}$$

$$Y_{in,2,DM} = Y_{in,1,DM}^* \tag{2.40}$$

$$C_{c,DM} = \frac{sin(2\phi)}{\omega R_L} \tag{2.41}$$

$$L_{c,DM} = \frac{R_L}{\omega sin(2\phi)} \tag{2.42}$$

We conclude that the outphasing systems appear very similar from the viewpoint of the PA's, the CM-compensation reactances can be obtained directly from the DM-compensation reactances by including the transformation $R_{L,eff} = \frac{Z_0^2}{R_L}$ of the $\frac{\lambda}{4}$-lines. The common-mode combiner introduces an additional degree of freedom: the characteristic impedance Z_0, which allows for a load transformation and the corresponding scaling of the compensation reactances, to bring these in a more realisable range, if necessary (e.g. increase C_c and decrease L_c at a fixed ϕ_c slightly by making $Z_0 < R_L$).

In the special and very convenient case of $R_L = 50\Omega = Z_0$, the admittances (and C_c, L_c) seen by the PA's are identical. This is confirmed in simulation: the exact same efficiency and output power plots are reproduced, with the same C_c, L_c, as demonstrated in figure 2.14, which shows the imaginary part of the input admittances after load compensation in the CM-combiner, when $R_L = 50\Omega = Z_0$ and $Z_{out} = 0$, for the cases described in table 2.1. All curves of the DM-combiner are reproduced in this case, with the exact same C_c and L_c. One important difference remains: when the output impedances are resistive but not zero, the efficiency of the differential combination is increased by increasing the (effective) load w.r.t. R_{out} to reduce the voltage division. For the common mode combiner, the load at the single-ended output has to decrease w.r.t. R_{out}, due to the $R_{L,eff} = \frac{Z_0^2}{R_L}$ - action of the $\frac{\lambda}{4}$-lines.
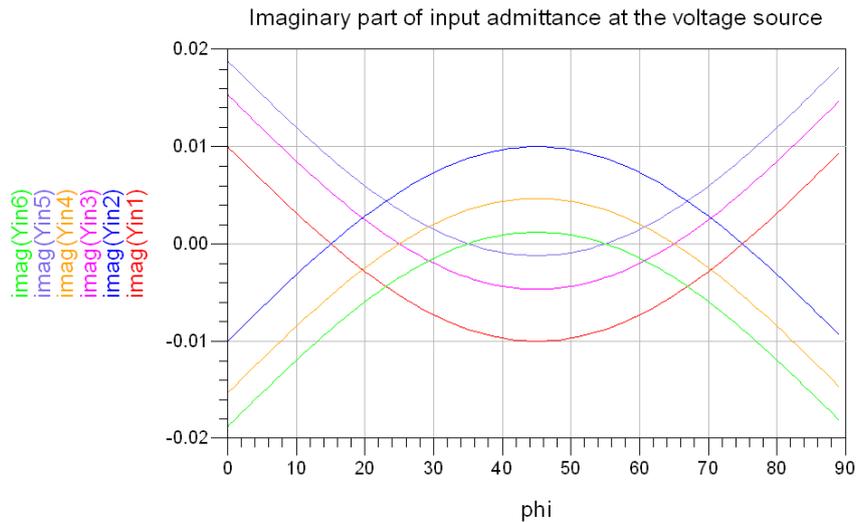
Figure 2.14: Imaginary part of the input admittance in the CM-combiner when $Z_{out} = 0$, for compensation at $\phi_c = 15°$ ($Y_{in,1}$, $Y_{in,2}$), $\phi_c = 25°$ ($Y_{in,3}$, $Y_{in,4}$) and $\phi_c = 35°$ ($Y_{in,5}$, $Y_{in,6}$) with C_c and L_c from table 2.1.

2.2.3 Effects of reactance compensation, power and efficiency tradeoffs

Because of the PA output impedance, adding compensation reactances will also deform the $P_{out}(\phi)$-characteristic, as demonstrated in figure 2.12. This deformation is deterministic, so if the $P_{out}(\phi)$-characteristic is known, it can be compensated for by the component separator (predistortion) by mapping the ideal ϕ on to the real ϕ which realizes the desired P_{out}. However, it is possible that, as a consequence of the compensation reactances, P_{out} is no longer strictly increasing with decreasing ϕ. This region will introduce even more distortion and degrade the EVM. To avoid this, this region could to be excluded completely and $\frac{A(t)}{A_{max}}$ could to be mapped onto a smaller range in ϕ. This requires more phase accuracy in the SCS: the effective slope of $V_{out}(\phi)$ has increased so the same small phase error will give a larger error in V_{out}. We could also clip the outphasing angle to a minimal value (the value, >0, where $P_{out}(\phi)$ is maximal) and still use the entire phase range. This will slightly improve the average output power. The EVM at high power (but $< P_{out,max}$) will degrade, but the overall PBO-efficiency will improve significantly because of the compensation. The potential efficiency gain increases if the original PAR/CRF of the input signals is (very) high and the original PBO-efficiency (very) low. The optimal tradeoff will strongly depend on the probability density and original PAR.

A similar reasoning can be applied to increase the average output power, even when P_{out} is strictly increasing with decreasing ϕ or when no compensation reactances are added. Because the probability of high output amplitudes is quite low, P_{out} can be increased to and clipped at $P_{out,max}$ from a certain power on, with a small EVM penalty. Again, this will improve the efficiency (certainly if no compensation reactances are used) because the PA's are more efficient at $P_{out,max}$, as described in [7] and demonstrated in figure 2.15 (from [7]).
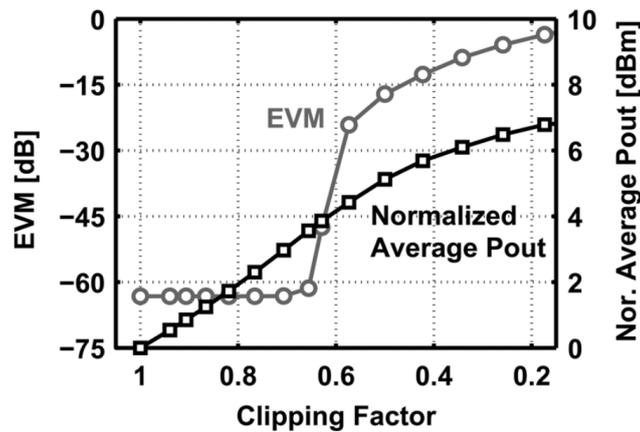
Figure 2.15: Example of average output power vs. EVM as a function of clipping angle.

2.3 Improved outphasing architecture: AMO

The basic outphasing architecture shows potential for a large efficiency increase, but it is not yet optimal. More advanced architectures have been developed. The AMO architecture proposed in [5] is based on an isolating Wilkinson combiner. Consequently, it does not benefit from load modulation; but it can still deliver a very high efficiency, possibly higher than the standard outphasing transmitter with load modulation.

When compensation reactances are omitted, the efficiency of the standard outphasing transmitter is low at low output power because low output power corresponds with high outphasing angles and strongly reactive loads seen by the PA's, which reduces the efficiency of each individual PA (cfr. section 2.2). When compensation reactances are added, the efficiency becomes dependent on the chosen compensation angle ϕ_c, but it still has minima at low or intermediate output power.

The AMO architecture contains two major modifications:

- Multilevel (ML-LINC):
  Large outphasing angles are avoided by making the PA supply dependent on the output amplitude, but only a discrete set of supply voltages is allowed. This supply voltage fixes the maximal PA output power, and intermediate output powers are obtained by increasing the outphasing angle; which will decrease the efficiency; but only until the next "level" is reached, at which point the supply is reduced, the outphasing angle is reset to zero and the efficiency is increased again. The efficiency vs. PBO plot (cfr. figure 2.17) now looks like a sawtooth: one ripple for each level, with a slope similar to the slope of the LINC-system at low outphasing angle, because we can expect that the plot will be as if multiple single-level plots, shifted to lower output powers, have been superimposed.
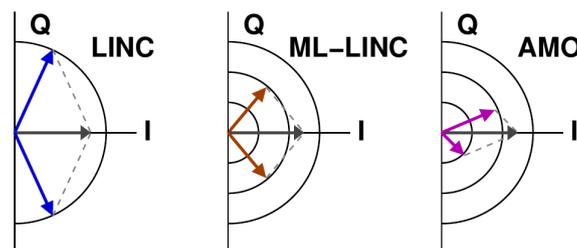
Figure 2.16: Extensions to the LINC-concept.
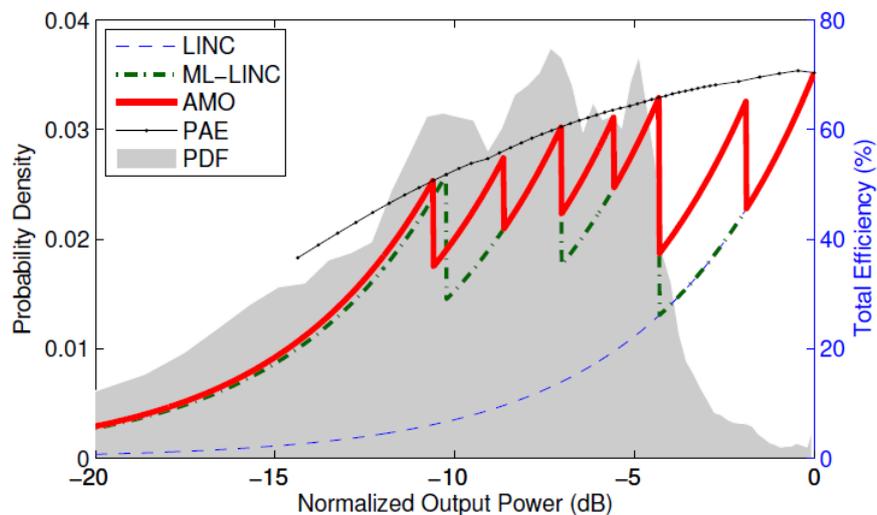
Figure 2.17: Efficiency comparison of (4-level) LINC-architectures, for WLAN 802.11g-signals.

- Asymmetric outphasing angles
  The ML-LINC-architecture already goes a long way (cfr. figure 2.17), but still does not guarantee optimal efficiency since the "average" outphasing angle is not guaranteed to be minimal at each output amplitude. This effect becomes important when the output level is slightly too high to be met with a lower supply setting: then a higher supply setting combined with a relatively large outphasing angle is necessary. Figure 2.17 (from [5]) confirms that it is at these points where the AMO-system obtains a higher efficiency than the ML-system. Here, both the AMO and ML-LINC architecture have four levels, but because of the outphasing angle asymmetry, the AMO transmitter has twice as much "effective" levels.
  To allow asymmetric outphasing angles while still achieving the correct phasor sum, one phasor has to be longer than the other: the supply setting has to be different for each PA (figure 2.16, from ([5])). In [5], a Wilkinson combiner is used, so the efficiency is maximized by choosing "adjacent" supply settings because this reduces the magnitude of the "loss" current through the isolation resistor.

Intuitively, a higher overall efficiency and more "sawteeth" are expected with an increasing number of levels. The ML-system is somewhat equivalent with a "discretized" polar PA/EER-PA combined with outphasing in between the levels. In the limit of a continuous number of levels, we would obtain the EER-system, but with the assumption that the amplitude control is obtained with 100 % efficiency, which is impossible in reality. The AMO-architecture was patented quite recently (2011, cfr. [13] and [14]) and is further developed by Eta Devices ([15], [3]).

2.4 Outphasing simulations in Matlab

Matlab simulations were done to determine the effect of gain and delay or phase mismatch, and to have a reference to compare with the results from the PA testbench in Cadence.

Gain differences can result from transistor asymmetry, power supply asymmetry (asymmetric power distribution) or asymmetry in the power combiner or other passive load components,

resulting in different terminations for the harmonic currents, causing asymmetric waveforms and a different voltage magnitude at the desired fundamental frequency. If the PA output is dependent on the input amplitude (which might be a very weak dependence for e.g. a switching amplifier and small amplitude deviations), asymmetry in the preceding blocks, e.g. the SCS, will result in different PA outputs. Even if the input amplitudes are identical, but not perfectly constant, the outputs might differ due the a difference in AM-AM characteristic between both PA's.

In this section, delay and phase differences are lumped together, because it will become clear that in this case (with $T << T_c$), the most dominant effect of a delay error τ is a phase error $\Delta\phi = -2\pi\frac{\tau}{T_c}$ introduced by the baseband (de)modulation. True delay differences might exist due to asymmetric layout or asymmetry in the power combiner. Even if this is not the case, a phase difference might still arise if the PM-PM characteristics of the PA's are not identical. Even though the origin of the effects might be different, the overall effect is very similar. Because it is not possible to estimate in advance how large the contribution of each cause might be, the total phase matching is expressed as a total delay matching instead, e.g. $\tau < \frac{T_c}{n}$. These requirements are (almost) equivalent to a total phase matching requirement $\Delta\phi < 2\frac{\pi}{n}$.

The carrier frequency is 15 GHz, the constellation is 256-QAM. A symbol period of 10 ns and a square root raised cosine pulse with rolloff factor 0.3 were chosen. This implies total RF-bandwidth of 130 MHz, a symbol rate of 100 Mbaud and a bit rate of 800 Mbit/s. The RF-bandwidth is comparable to the 100 MHz maximal bandwidth in LTE Advanced and the 160 MHz for IEEE 802.11 ac ([3]). For the PA-design, it was not strictly specified, but a choice has to be made for the simulations. The symbol rate is important because it determines the bandwidth and thus the EVM if the bandwidth is limited along the transmitter chain, but it does not influence the effects of gain and delay mismatch, which is the primary concern in this section.

Because of the sampling theorem for bandpass signals $(-f_L + (n-1)f_s \leq f_L,\ -f_H + nf_s \geq f_H)$, f_s can be quite low (n integer):

$$\frac{2f_H}{n} \leq f_s \leq \frac{2f_L}{n-1} \tag{2.43}$$

At the same time, f_s needs to be larger than the RF-bandwidth. To easily obtain the symbols by downsampling (without having to resample first), f_s has to be an integer multiple (k) of the symbol rate $\frac{1}{T}$. With $f_L = 14.935$ GHz and $f_H = 15.065$ GHz, searching for the lowest integer k for which both demands are satisfied gives $k = 8$ and $n = 38$. Sampling at $f_s = \frac{8}{T} = 800$ MHz makes the simulation quite fast, even for a large number of symbols (e.g. 10^6). If T is varied, the minimal k is first obtained by a script. $e(t)$ is a non-linear function of $s(t)$, so bandwidth expansion will occur, but we do not know in advance what the new bandwidth will become. With $f_s = 800$ MHz, the maximal RF-bandwidth that can be allowed while still satisfying equation 2.43 $((-f_L + (n-1)f_s \leq f_L,\ -f_H + nf_s \geq f_H,\ n = 38,\ f_L = f_c - \frac{BW}{2},\ f_H = f_c + \frac{BW}{2})$ is only 400 MHz (200 MHz single-sided bandwidth). This is a bandwidth expansion with more than a factor 3, but it is still not very large (e.g. when comparing with figure 2.18, which was made with a higher f_s). It was therefore checked and confirmed that the results in all cases (the ideal situation and with gain or delay error) did not improve by increasing the sample frequency (e.g. $k = 9, 11, 14, 16, 17, 18, 19, 21, 22..$ are allowed).

The impulse response duration has to be limited to L_f symbol periods. This is equivalent to multiplying the infinite impulse response with a rectangular pulse, or in frequency domain, taking the convolution of the ideal Fourier transform with a sinc; so sidelobes will occur. A realistic transmitter can suppress some of these with an output filter, to improve the ACPR

(Adjacent Channel Power Ratio). The effect in time domain is the introduction of ISI. The goal is to determine the effect of system imperfections, not of the finite impulse response duration, so L_f can be chosen high to minimize the linear ISI: $L_f = 40$. In the absence of gain and delay errors, this gives an EVM of -74 dB.

By using $[s_1(t), s_2(t)]$ for common-mode combiners and $[s_1(t), -s_2(t)]$ for differential combiners, the results become independent of the type of combination. Secondly, the presence of load modulation does not influence the ideal output, which in both cases equal to $s_1(t) + s_2(t)$. In an isolating combiner, the power of $e(t)$ is lost in the fixed load seen by each PA; in the non-isolating combiner, ideally, no power should be dissipated due to $e(t)$. Gain and delay errors are also included in $s_1(t) + s_2(t)$, so their effect on the output should be the same, regardless of load modulation. Consequently, with these definitions, the results of the theoretical simulations of this section should be independent of the realisation.

2.4.1 Ideal situation

The constant-amplitude signals $s_1(t) = s(t) + e(t)$ and $s_2(t) = s(t) - e(t)$ contain $e(t)$. $e(t)$ is derived from $s(t)$ in a very non-linear way, so it will have a wider power spectrum; because the non-linearity introduces frequency components which were not present in the input spectrum (cfr. figure 2.18).
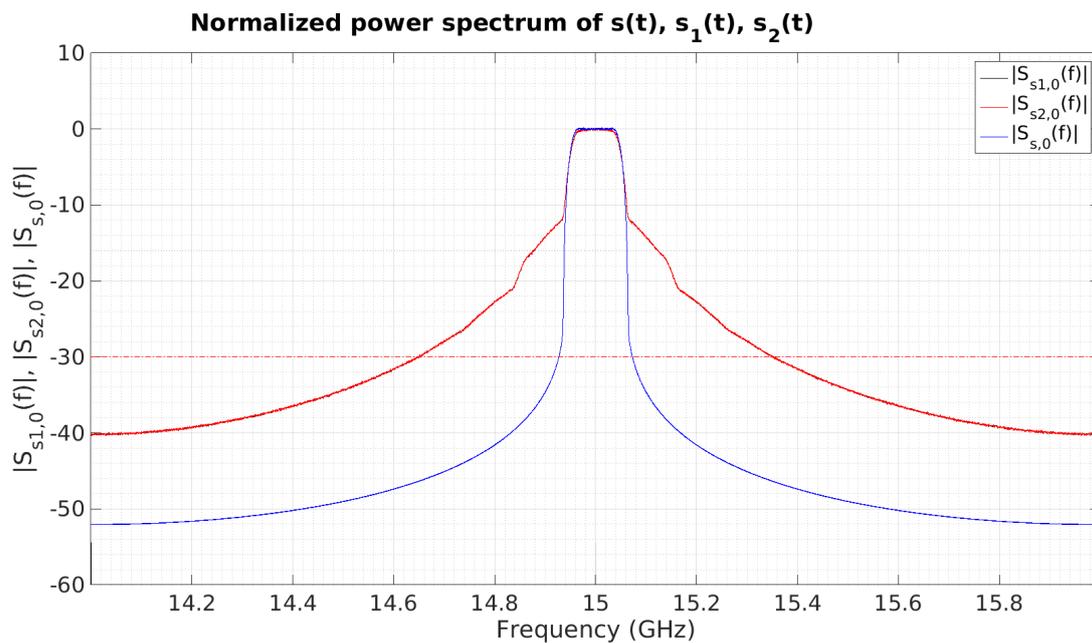
Figure 2.18: Normalized power spectrum (10^6 symbols).
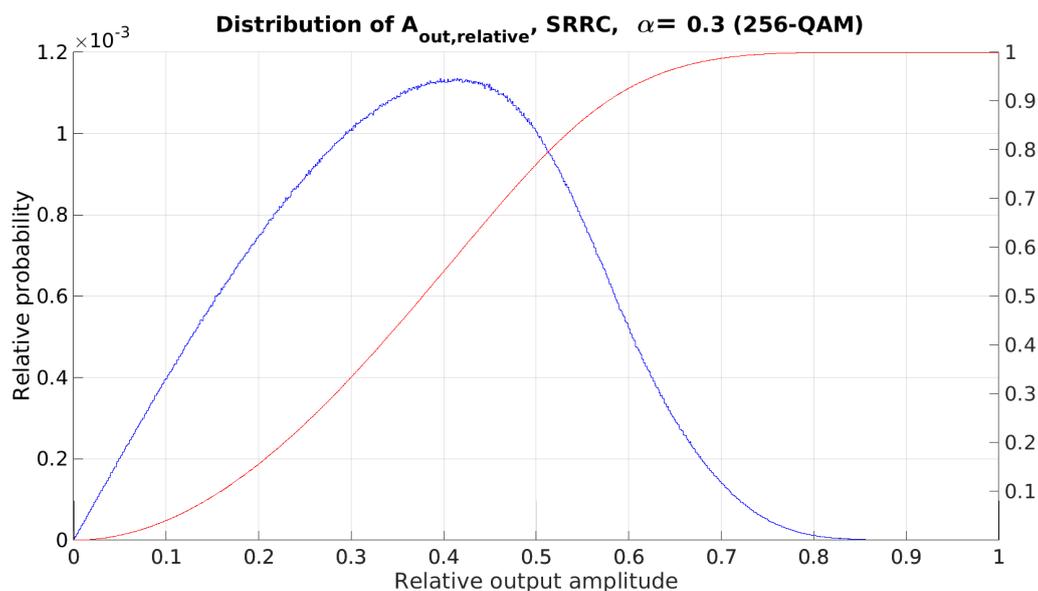
Figure 2.19: Simulated pdf and cpdf of the relative output power (linear). A large PAR of 7.8385 dB was found, due to the constellation type and the SRRC-filtering.
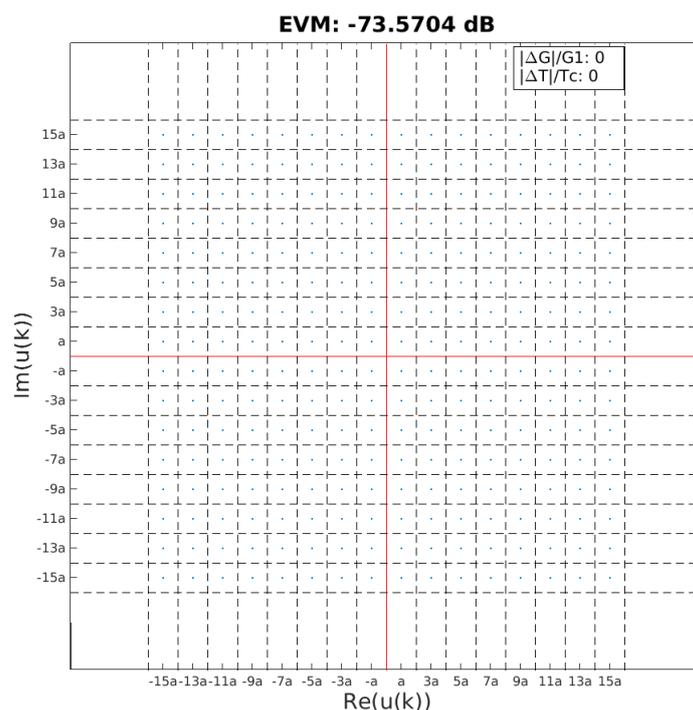
Figure 2.20: Simulated scatterplot and EVM of the ideal transmitter.

In a real receiver, the carrier phase and amplitude have to be estimated to perform coherent detection. It is therefore assumed that the receiver can correct the received scatterplot by rescaling and rotating, based on the location of the constellation edges. The EVM should be calculated after this auto-correction. Even then, constellation point-dependent errors can be observed. If one type of imperfection in the transmitter dominates the other, a predictable pattern appears. It is assumed that the receiver is not able to correct these errors because this would involve "pattern detection and correction" (detecting the center of each individual "point cloud" and performing a symbol-dependent correction), which is more complex. If the receiver

were able to do this, the EVM would only be limited by the total ISI (the size of the "point cloud").

2.4.2 Phase/delay difference

$$s_{1,BP}(t) = \Re[(s(t) + e(t))e^{j2\pi f_c t}] \tag{2.44}$$

$$s_{2,BP}(t) = \Re[(s(t) - e(t))e^{j2\pi f_c t}] \tag{2.45}$$

At this point, we assume that the gain of the PA's is identical, so it can be neglected (for now). If path 2 is delayed with respect to path 1, we obtain:

$$r(t) = s_{1,BP}(t) + s_{2,BP}(t - \tau) = \Re[(s(t) + s(t - \tau)e^{-j2\pi f_c \tau} + e(t) - e(t - \tau)e^{-j2\pi f_c \tau})e^{j2\pi f_c t}] \tag{2.46}$$

We assume a perfect bandpass demodulator, which gives the envelope v(t):

$$v(t) = s(t) + s(t - \tau)e^{-j2\pi f_c \tau} + e(t) - e(t - \tau)e^{-j2\pi f_c \tau} \tag{2.47}$$

The delay difference has to be small in order to produce a reasonable output. In reality, a symmetrical structure and layout should guarantee delay matching of the order of T_c. This is equivalent with guaranteeing that the effective path length difference will not exceed the shortest λ. The data rate is quite low with respect to the carrier frequency, resulting in a small relative bandwidth of $\frac{15.065 GHz}{14.935 GHz} = 1.0087$. This implies that λ_{15GHz} is a good measure for the path length matching requirements. At 15 GHz and with $\epsilon_r = 3.5$, $\lambda = 10.69$ mm. Matching up to (a certain fraction of) λ is certainly possible. Therefore, we will at first not consider delays larger than T_c.

The symbol period T (10 ns) is 150 times larger than T_c, so we can make the approximation that delays smaller than T_c will not affect the baseband signal: $s(t) \approx s(t-\tau)$. This also implies that $e(t) \approx e(t-\tau)$. In the simulation, we do take this delay into account (cfr. e.g. figure 2.24) by shifting the time vector of the transmit pulse for that path, so that the sampling rate does not have to change. No symbol shift is necessary because the delays are small with respect to T_c, so they are certainly smaller than $T_s = \frac{T}{8}$. However, with this approximation, the previous result can be simplified:

$$1 + e^{-j2\pi f_c \tau} = e^{-j\omega_c \frac{\tau}{2}}(e^{j\omega_c \frac{\tau}{2}} + e^{-j\omega_c \frac{\tau}{2}}) = 2cos(\omega_c \frac{\tau}{2})e^{-j\omega_c \frac{\tau}{2}} \tag{2.48}$$

$$1 - e^{-j2\pi \omega_c \tau} = 2jsin(\omega_c \frac{\tau}{2})e^{-j\omega_c \frac{\tau}{2}} \tag{2.49}$$

The same result will be obtained if a "realistic" demodulator is used (as in the simulations), $s_{BP}(t-\tau) = I(t-\tau)cos(2\pi f_c(t-\tau)) - Q(t-\tau)sin(2\pi f_c(t-\tau))$, so the bandpass demodulation can be done by:

$$I_{rec}(t) = 2(s_{BP}(t) \cdot cos(2\pi f_c t)) * h_{rec}(t) \tag{2.50}$$

$$Q_{rec}(t) = -2(s_{BP}(t) \cdot sin(2\pi f_c t)) * h_{rec}(t) \tag{2.51}$$

In this situation, $h_{rec}(t)$ simultaneously acts as a lowpass filter (to average out higher-frequency terms) and the baseband demodulator. We can decompose this filter into an ideal lowpass filter $h_{rec,LP}(t)$ and $h_{rec,BB}(t)$.

For $I_{rec}(t)$:

$$2[I(t-\tau)cos(2\pi f_c(t-\tau)) - Q(t-\tau)sin(2\pi f_c(t-\tau))] \cdot cos(2\pi f_c t))$$
$$= I(t-\tau)[cos(2\pi f_c \tau) + cos(2\pi f_c(2t-\tau)] - Q(t)[-sin(2\pi f_c \tau) + sin(2\pi f_c(2t-\tau)]$$

$h_{rec,LP}(t)$ filters out the double frequency components and we obtain: $I_{rec}(t) = I(t-\tau)cos(2\pi f_c \tau) + Q(t-\tau)sin(2\pi f_c \tau)$.

For $Q_{rec}(t)$:

$$-2[I(t-\tau)cos(2\pi f_c(t-\tau)) - Q(t-\tau)sin(2\pi f_c(t-\tau))] \cdot sin(2\pi f_c t))$$
$$= -I(t-\tau)[sin(2\pi f_c \tau) + sin(2\pi f_c(2t-\tau)] + Q(t-\tau)[cos(2\pi f_c \tau) - cos(2\pi f_c(2t-\tau)]$$

After filtering with $h_{rec,LP}(t)$, we obtain: $Q_{rec}(t) = Q(t-\tau)cos(2\pi f_c \tau) - I(t-\tau)sin(2\pi f_c \tau)$.

Comparing I_{rec}, Q_{rec} with $I(t)$, $Q(t)$:

$$I_{rec}(t) + jQ_{rec}(t) = e^{-j2\pi f_c \tau}[I(t-\tau) + jQ(t-\tau)] \approx e^{-j2\pi f_c \tau}[I(t) + jQ(t)] \tag{2.52}$$

We substitute the results into the received envelope:

$$v(t) \approx s(t)[1 + e^{-j2\pi f_c \tau}] + e(t)[1 - e^{-j2\pi f_c \tau}] = s(t)[2cos(\omega_c \frac{\tau}{2})e^{-j\omega_c \frac{\tau}{2}}] + e(t)[2j sin(\omega_c \frac{\tau}{2})e^{-j\omega_c \frac{\tau}{2}}] \tag{2.53}$$

The definition of $e(t)$ was:

$$e(t) = j\sqrt{\frac{|s(t)|^2_{max}}{|s(t)|^2} - 1} \cdot s(t) \tag{2.54}$$

The envelope can now be expressed as a function of $s(t)$ only:

$$v(t) \approx 2s(t)e^{-j\omega_c \frac{\tau}{2}} \left(cos(\omega_c \frac{\tau}{2}) - sin(\omega_c \frac{\tau}{2})\sqrt{\frac{|s(t)|^2_{max}}{|s(t)|^2} - 1} \right) \tag{2.55}$$

We now apply the receive filter and sample the output:

$$u(t) = [s(t) + s(t-\tau)e^{-j2\pi f_c \tau}] * h_{rec}(t) + [e(t) - e(t-\tau)e^{-j2\pi f_c \tau}] * h_{rec}(t), u(k) = u(kT) \tag{2.56}$$

$$u(t) \approx (s(t)[1 + e^{-j2\pi f_c \tau}]) * h_{rec}(t) + (e(t)[1 - e^{-j2\pi f_c \tau}]) * h_{rec}(t) \tag{2.57}$$

The delay has an effect through the baseband-demodulation:

- The term $s(t - \tau) * h_{rec}(t)$ will introduce linear ISI, because $[h_{tr}(t) * h_{rec}(t)](t = kT) = \delta(k - L_f)$ holds only when h_{tr} is not delayed. The receiver cannot distinguish the individual terms in $u(t)$, so the problem cannot be solved by sampling at $t = L_f T + \tau$ instead of $t = L_f T$.

- Because $e(t)$ is a non-linear function of $s(t)$ and $h_{tr}(t)$, analytically calculating $e(t) * h_{rec}(t)$ becomes a problem. However, we do know that $s(t)$, and thus $e(t)$, depends on "all" "random" previous symbols (a limited number of previous symbols in reality since L_f is limited), which will give an undesired "random" term: $e(t) * h_{rec}(t)$ and $e(t - \tau) * h_{rec}(t)$ introduce non-linear ISI.

For the effects of the bandpass demodulation, we temporarily neglect the linear ISI and consider the approximations of $u(t)$ and $v(t)$:

- Symbol rotation: is multiplied with $e^{-j\omega_c \frac{\tau}{2}}$, so the received symbols will be rotated over $-\omega_c \frac{\tau}{2}$.

- A term proportional to e(t) appears: the "error" signal "leaks through" because it is not perfectly cancelled at the output, which will broaden the power spectrum and increase the ACPR (Adjacent Channel Power Ratio, the ratio of leakage power into adjacent bands with respect to the main channel power) (cfr. figure 2.25).

- The amplitude of $u(t)$, and thus the amplitude of the received symbols, becomes dependent on the delay or phase difference. $e(t)$ is orthogonal (counter-clockwise) to $s(t)$, so $j \cdot e(t)$ is opposite to $s(t)$; and because $|e(t)|$ is larger when $|s(t)|$ is smaller ($|s(t)|^2 + |e(t)|^2 = A_{max}^2 = |s_1(t)|^2 = |s_2(t)|^2$), $|s(t)|$ will decrease more when $|s(t)|$ is small. Any deviation from the ideal situation causes $e(t)$ to appear in the output, so in general, because $|e(t)|$ is larger when $|s(t)|$ is smaller, all effects (scaling, rotation, and also ISI) will be larger for the smaller constellation symbols.

To explain the scatterplots more clearly, we estimate u(k) by neglecting the ISI temporarily, replacing $s(t)$ by $a(k)$ and $e(t)$ by $e(k)$:

$$u_{estimate}(k) \approx 2a(k)e^{-j\omega_c \frac{\tau}{2}} \left(cos(\omega_c \frac{\tau}{2}) - sin(\omega_c \frac{\tau}{2}) \sqrt{\frac{1}{|a(k)|^2} - 1} \right) \qquad (2.58)$$

$$\alpha_2 = \left(cos(\omega_c \frac{\tau}{2}) - sin(\omega_c \frac{\tau}{2}) \sqrt{\frac{1}{|a(k)|^2} - 1} \right) \qquad (2.59)$$

When $|a(k)|$ is small, $|u(k)|$ will be reduced more.

When path 1 is delayed, the sign of $e(t) - e(t - \tau)$ is reversed, so we would obtain:

$$u_{estimate}(k) \approx 2a(k)e^{-j\omega_c \frac{\tau}{2}} \left(cos(\omega_c \frac{\tau}{2}) + sin(\omega_c \frac{\tau}{2}) \sqrt{\frac{1}{|a(k)|^2} - 1} \right) \qquad (2.60)$$

$$\alpha_1 = \left(cos(\omega_c \frac{\tau}{2}) + sin(\omega_c \frac{\tau}{2}) \sqrt{\frac{1}{|a(k)|^2} - 1} \right) \qquad (2.61)$$

When $|a(k)|$ is small, $|u(k)|$ will be increased more.

Based on these results, we expect that the average received power will also decrease (with respect to the ideal situation) when path 2 is delayed and increase when path 1 is delayed. An approximation of this power change can be obtained from the approximation of $u(k)$:

$$E[|\alpha_2|^2] = cos(\omega_c \frac{\tau}{2})^2 + E\left(\left(\frac{1}{|a(k)|^2} - 1\right)\right) \cdot sin(\omega_c \frac{\tau}{2})^2 - 2sin(\omega_c \frac{\tau}{2})cos(\omega_c \frac{\tau}{2}) \cdot E\left(\sqrt{\left(\frac{1}{|a(k)|^2} - 1\right)}\right)$$
$$(2.62)$$

$$E[|\alpha_1|^2] = cos(\omega_c \frac{\tau}{2})^2 + E\left(\left(\frac{1}{|a(k)|^2} - 1\right)\right) \cdot sin(\omega_c \frac{\tau}{2})^2 + 2sin(\omega_c \frac{\tau}{2})cos(\omega_c \frac{\tau}{2}) \cdot E\left(\sqrt{\left(\frac{1}{|a(k)|^2}\right) - 1}\right)$$
$$(2.63)$$

We now calculate $E\left(\left(\frac{1}{|a(k)|^2} - 1\right)\right)$ and $E\left(\sqrt{\left(\frac{1}{|a(k)|^2} - 1\right)}\right)$ numerically and obtain the approximate average power change. Tables 2.2 and 2.3 compare this prediction with the obtained power from the simulations. The relative power changes are not identical, but the trend is the same. The relative power change can also be predicted visually, as demonstrated in figure 2.21.

Delay	$\frac{P_{rec}}{P_{rec}(\tau=0)}$	$E[|\alpha_2|^2]$
0	1	1
$\frac{T_c}{50}$	0.7381	0.7864
$\frac{T_c}{30}$	0.6592	0.6768
$\frac{T_c}{20}$	0.5059	0.5781

Table 2.2: Comparison of the theoretical and measured relative average power change, when path 2 is delayed.

Delay	$\frac{P_{rec}}{P_{rec}(\tau=0)}$	$E[|\alpha_1|^2]$
0	1	1
$\frac{T_c}{50}$	1.2575	1.2696
$\frac{T_c}{30}$	1.44	1.4784
$\frac{T_c}{20}$	1.682	1.7696

Table 2.3: Comparison of the theoretical and measured relative average power change, when path 1 is delayed.

If the transmitter would be able to estimate the delay (τ_{est}), the rotation introduced by $e^{-j2\pi f_c \tau}$ in the bandpass demodulation could be anticipated by "multiplying" the constant-envelope baseband signal of that path with $e^{+j2\pi f_c \tau_{est}}$. Because the modulation is a linear function of the symbols ($s(t) = \sum_k a(k)h_{tr}(t-kT)$), this multiplication can even be done on the symbols for that path, which makes the implementation very simple. $e(t)$ is defined as being orthogonal to $s(t)$, so all phase modulation or phase offset of $s(t)$ appears in $e(t)$ as well. We then obtain:

$$v(t) = s(t) + s(t-\tau)e^{j2\pi f_c \tau_{est}}e^{-j2\pi f_c \tau} + e(t) - e(t-\tau)e^{j2\pi f_c \tau_{est}}e^{-j2\pi f_c \tau} \qquad (2.64)$$

In the ideal case, $\tau_{est} = \tau$:

$$v(t) = s(t) + s(t - \tau) + e(t) - e(t - \tau) \tag{2.65}$$

Because the symbol rate is much lower than the carrier frequency and we assumed that $\tau < T_c$, we have that: $s(t) \approx s(t - \tau)$ and $e(t) \approx e(t - \tau)$, so a near-perfect output is expected. This is confirmed in figure 2.24. Some linear ISI remains from the demodulation of $s(t) + s(t - \tau)$, and some non-linear ISI is caused by the demodulation of $e(t) - e(t - \tau)$, but the improvement in EVM is very large, as expected, since τ is of the order of $\frac{1}{f_c}$ but the variation in time of s(t) and e(t) referenced with respect to T instead of $\frac{1}{f_c}$.
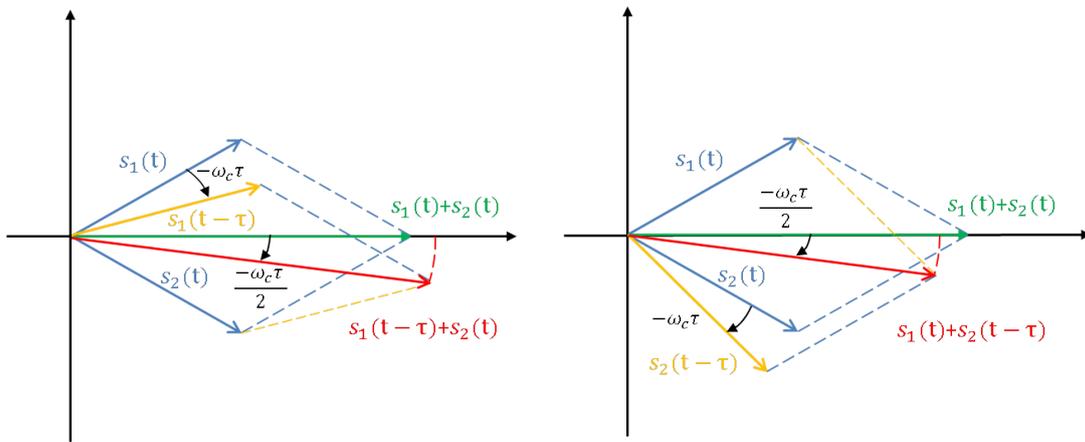
Figure 2.21: Illustration of the effect of a delay difference on the phasor sum.
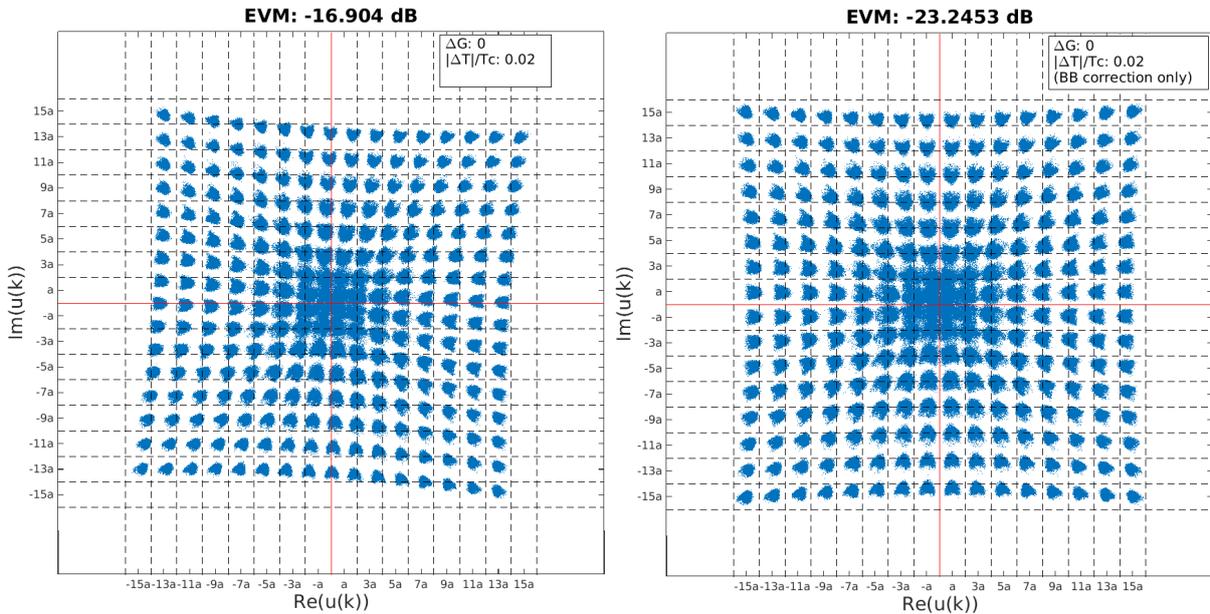
Figure 2.22: Simulated scatterplot, without (left) and with (right) receiver corrections, path 2 delayed by $\frac{T_c}{50} = 1.333ps$.
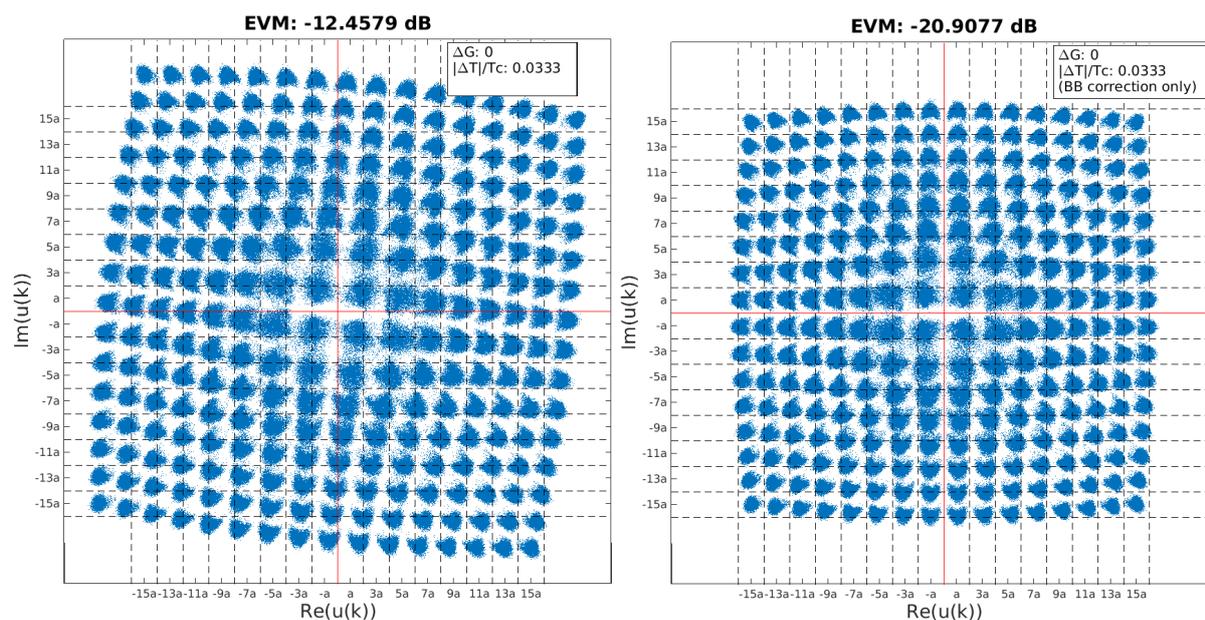
Figure 2.23: Simulated scatterplot, without (left) and with (right) receiver corrections, path 1 delayed by $\frac{T_c}{30} = 2.222ps$.
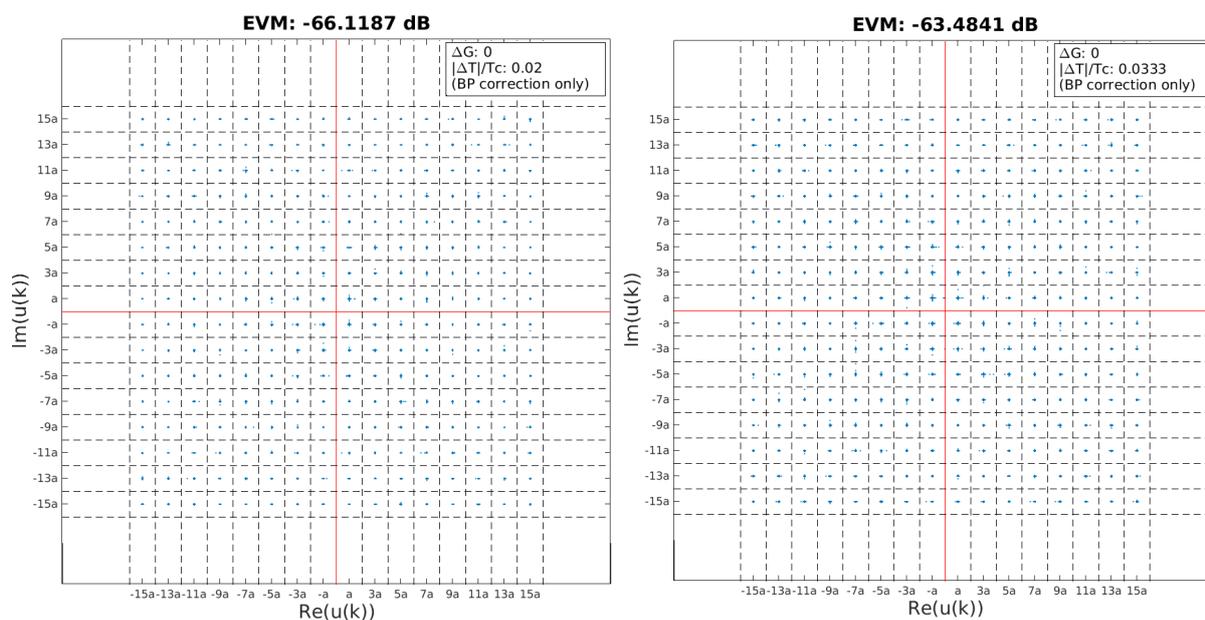
Figure 2.24: Simulated scatterplots with bandpass (phase) correction. In the left figure, path 2 is delayed by $\frac{T_c}{50} = 1.333ps$. In the right figure, path 1 is delayed by $\frac{T_c}{30} = 2.222ps$. Some linear and non-linear ISI remains because $s(t)$ and $s(t-\tau)$ are not exactly equal, and $e(t)$ and $e(t-\tau)$ do not cancel perfectly at the output.
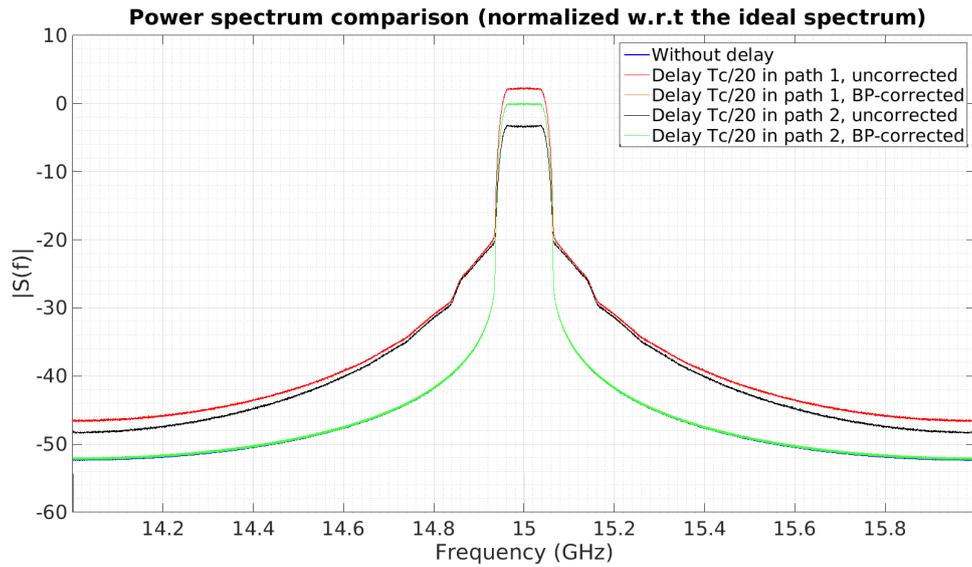
Figure 2.25: Simulated power spectrum in the case of a delay difference. When the bandpass correction (the rotation) is applied, the spectrum almost coincides with the ideal power spectrum.

2.4.3 Gain difference

To estimate the effect of a gain difference, we temporarily set the ideal gain to 1 and introduce an extra term α ($0 < \alpha < 1$) in the path with the highest gain, e.g. path 1:

$$v(t) = (1 + \alpha)(s(t) + e(t)) + (s(t) - e(t)) = (2 + \alpha)s(t) + \alpha e(t) \qquad (2.66)$$

Again, e(t) "leaks through", resulting in bandwidth expansion and ACPR-degradation. The useful signal power increases as α increases, but the baseband demodulation of e(t) will again result in non-linear ISI and a higher EVM. The baseband demodulation of s(t) is "perfect", "without" linear ISI. We temporarily neglect the ISI to explain the scatterplots: an orthogonal "error vector" e(t) is added to s(t), so if path 1 has the high gain, the symbols will be rotated counter-clockwise. Because $|e(t)|$ increases as $|s(t)|$ decreases, this effect will be more apparent for the inner symbols. If path 2 has the highest gain, the conclusion still holds, but all deviations are in the opposite direction. Figures 2.26 and 2.27 demonstrate this effect, but it is also clear that the ISI increases very rapidly with increasing α.
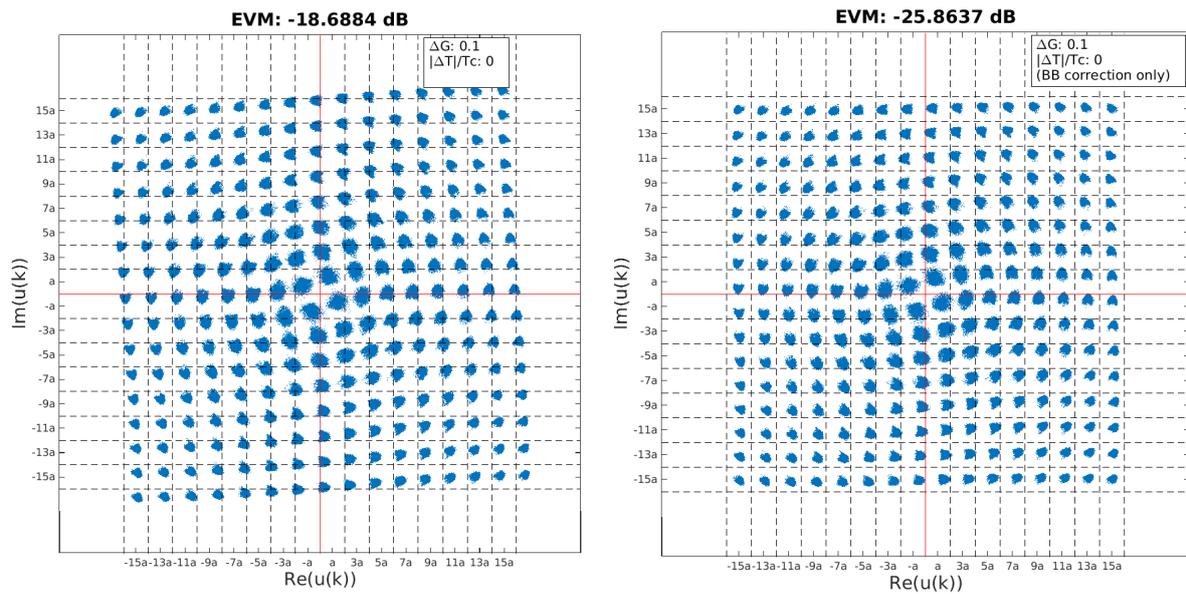
Figure 2.26: Simulated scatterplot with $G_1 = 1.1$ and $G_2 = 1$, without (left) and with (right) receiver corrections.
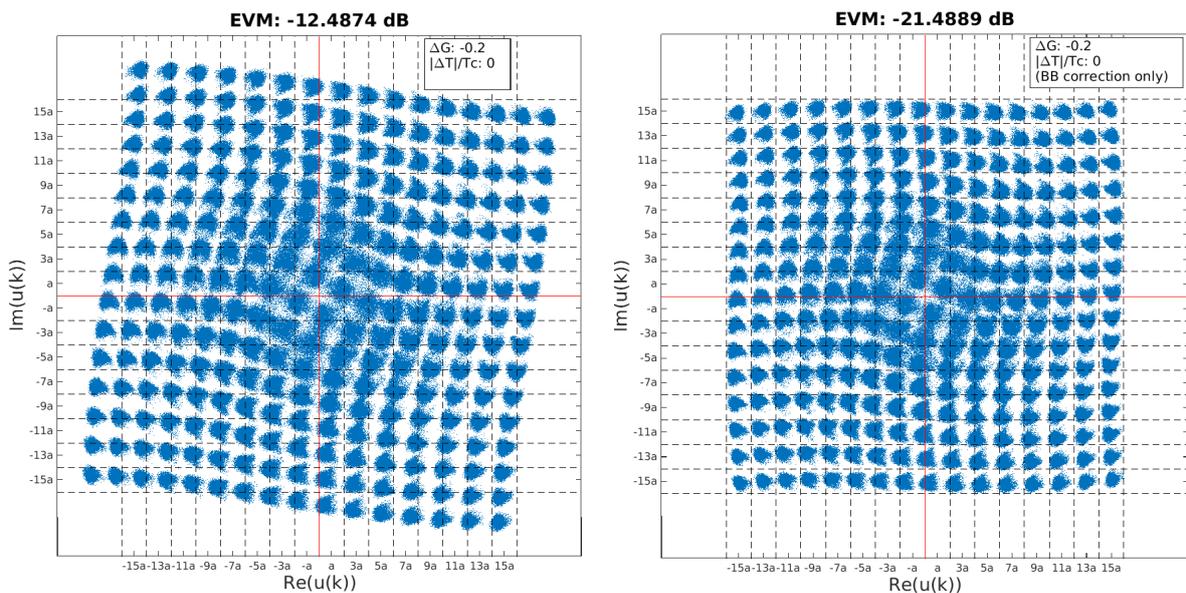
Figure 2.27: Simulated scatterplot with $G_1 = 1$ and $G_2 = 1.2$, without (left) and with (right) receiver corrections.
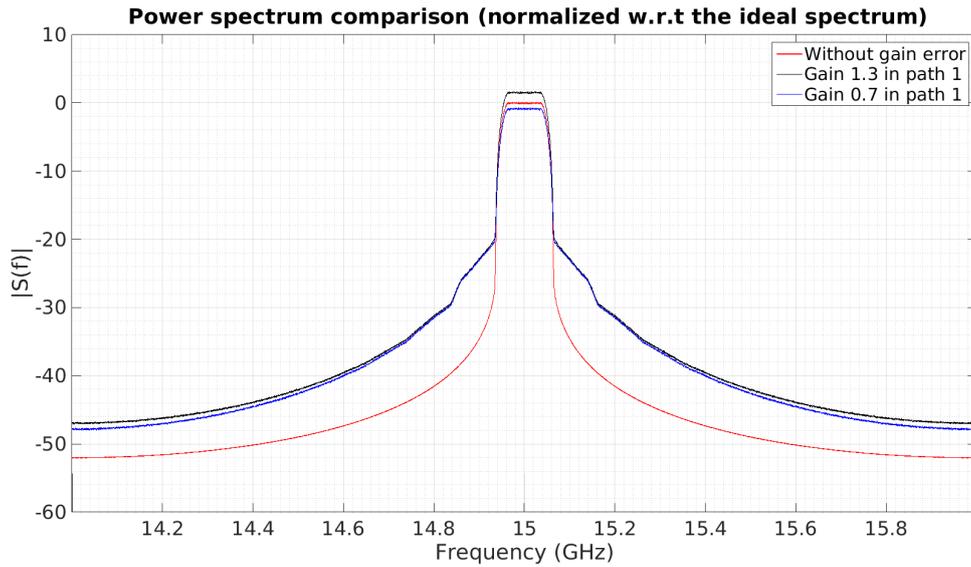
Figure 2.28: Simulated power spectrum in the case of a gain difference.

2.4.4 Combination or gain and delay (phase) error

In reality, gain and phase errors will exist simultaneously. Table 2.4 summarizes the results, and makes clear that the matching between both paths indeed has to be very tight. From the central part of the table, it is also clear that the effect of the delay is more dominant than the effect of the gain. Matching up to a phase(delay) difference of e.g. $\tau = \frac{T_c}{100}$ or $\Delta\phi = 3.6°$ introduces about 16 dB more EVM than matching up to $+/- 1$ % gain difference. When $\tau = \frac{T_c}{100}$, the EVM increase due to $+/- 1$ % gain difference is less than 1 dB.

$[\tau(\Delta\phi),\text{ path}]$	$G_1 = 0.9$	$G_1 = 0.95$	$G_1 = 0.99$	$G_1 = 1$	$G_1 = 1.01$	$G_1 = 1.05$	$G_1 = 1.1$
$\frac{T_c}{50}(7.2°),1$	-22.5449	-23.9363	-24.8644	-24.8442	-24.902	-24.1324	-22.838
$\frac{T_c}{100}(3.6°),1$	-24.9813	-28.799	-30.3659	-30.315	-30.4755	-28.7664	-25.1659
$\frac{T_c}{180}(2°),1$	-25.2333	-30.686	-34.9439	-35.478	-35.2624	-30.0182	-26.1934
$\frac{T_c}{360}(1°),1$	-25.8829	-31.1964	-40.1625	-41.4199	-40.0887	-31.8893	-26.5818
No delay	-25.7762	-31.8435	-46.5018	**-73.5609**	-45.4232	-31.7344	-26.372
$\frac{T_c}{360}(1°),2$	-25.2936	-31.813	-39.9421	-40.8623	-39.5277	-32.5986	-26.1669
$\frac{T_c}{180}(2°),2$	-25.4906	-30.451	-34.5851	-35.0854	-34.7532	-30.6226	-25.402
$\frac{T_c}{100}(3.6°),2$	-24.2123	-27.4919	-29.7536	-30.1003	-29.3627	-27.9322	-24.047
$\frac{T_c}{50}(7.2°),2$	-21.2467	-22.8554	-23.6507	-23.5805	-23.3696	-22.8362	-22.1057

Table 2.4: EVM comparison for various combinations of ΔG and τ (or $\Delta\phi$).

2.4.5 Bandwidth limitations, bandwidth difference

As demonstrated in figure 2.18, the constant envelope signals $s_1(t)$ and $s_2(t)$ have a much wider bandwidth than $s(t)$. This bandwidth needs to be provided in the entire path from baseband to upconversion and in the bandpass section. The transfer function of this signal chain could be summarized in one bandpass filter at 15 GHz, or more conveniently, a lowpass filter in the baseband path, $H_{LP,tot}(f)$. Additional phase shifts (e.g. a difference in phase

shift introduced by the PA, due to mismatch in the AM-PM or PM-PM characteristic of the PA's), delays, attenuations, gain mismatch... can all be included in these transfer functions $H_{LP,tot,1}(f)$ and $H_{LP,tot,2}(f)$. In the previous sections, the frequency dependence was neglected, reducing $H_{LP,tot,1}(f)$ and $H_{LP,tot,2}(f)$ to $H_{LP,tot,1}$ and $H_{LP,tot,2}$, and we focussed on the effect of a difference in $|H_{LP,tot,i}|$ and $\angle H_{LP,tot,i}$. In this section, $H_{LP,tot,i}(f)$ is assumed to be a first order lowpass filter and mismatch in $H_{LP,tot,i}(f)$ is considered. In reality, $H_{LP,tot,i}(f)$ is most likely a higher-order filter, but the results here are just for illustration.

$$s(t) = s_1(t) * h_{LP,tot,1}(t) + s_2(t) * h_{LP,tot,2}(t) \tag{2.67}$$

If $h_{LP,tot,1}(t) = h_{LP,tot,2}(t)$, $e(t)$ still cancels completely:

$$s(t) = (s(t) + e(t)) * h_{LP,tot,1}(t) + (s(t) - e(t)) * h_{LP,tot,2}(t) = 2s(t) * h_{LP,tot}(t) \tag{2.68}$$

This formula is not exact: intuitively, we can see that limiting the bandwidth will make the amplitude of $s_1(t)$ and $s_2(t)$ variable, because the high frequency components of $e(t)$, necessary to compensate for the envelope variation of $s(t)$, are attenuated. This time-variant envelope variation will lead to a time-variant error term $\Delta\phi$ in the phase due to the AM-PM characteristic of the PA's; but this effect is not included in equation 2.68. If the bandwidth limitation and AM-PM characteristic are identical in both paths, the phasors $s_1(t)$ and $s_2(t)$ are rotated over the same angle $\Delta\phi$, so $\Delta\phi$ directly contributes to the phase error in the phasor sum $s(t)$.

Due to the AM-AM characteristic of the PA, undesired AM is expected on the PA outputs by limiting the bandwidth. If both paths suffer from the same bandwidth limitation and AM-AM-function, this undesired AM will directly appear in the output. If the non-idealities in the paths are different, asymmetric amplitude variations will appear on the output; so both amplitude and phase errors are introduced in the phasor sum $s(t)$.

For switching amplifiers, we expect that the output amplitude is only slightly dependent on the input amplitude, as long as the deviations on the input amplitude are small: the amplifier will still switch, if the input amplitude does not decrease too severely. In other words: the AM-AM characteristic should be quite flat around the ideal input operating point. With this assumption, the constant-envelope character of $s_1(t)$ and $s_2(t)$, which was lost in the path to the PA, will be "restored" by the PA. This conclusion will only hold if the bandwidth limitation is not too severe and not to asymmetric, and if the AM-AM curve is sufficiently flat. Finally, we remark that a flat AM-AM-curve might be convenient in this case, but it might be less convenient when we try to compensate for a gain difference (cfr. section 2.7).

Table 2.5 contains the results for a symmetrical (baseband) bandwidth limitation. We can try to model the effect of the "restoration" of the constant amplitude of $s_1(t)$ and $s_2(t)$ at the PA output, the effect is consistent but only very minimal.

Condition	EVM
$BW = 10$ GHz	-52.8748
$BW = 10$ GHz, with restoration	-53.271
$BW = 1$ GHz	-32.8951
$BW = 1$ GHz, with restoration	-33.1996
$BW = 0.8$ GHz	-30.9701
$BW = 0.8$ GHz, with restoration	-31.2372
$BW = 0.5$ GHz	-26.8811
$BW = 0.5$ GHz, with restoration	-27.0811

Table 2.5: EVM comparison, for a symmetrical bandwidth limitation.

In the scatterplots, we observe linear, circular ISI. This shape of the ISI is independent from the symbol location because $e(t)$ is eliminated completely, as expected from equation 2.68.

Table 2.6 contains the results for an asymmetrical (baseband) bandwidth limitation. Due to "leak through" from $e(t)$ to the output, non-linear ISI is added to the linear ISI from the bandwidth limitation. This also introduces a rotation and scaling of the symbols, as described in the previous sections. Similar scatterplots, but with more ISI, are reproduced. Therefore, the EVM is slightly worse than in the case of the EVM of the smallest symmetrical bandwidth from table 2.5. For the simulations, the bandwidth asymmetry was chosen very large, but the effects stay small.

Condition	EVM
$BW_1 = 10$ GHz, $BW_2 = 1$ GHz	-31.223
$BW_1 = 10$ GHz, $BW_2 = 1$ GHz, with restoration	-32.3358
$BW_1 = 1$ GHz, $BW_2 = 0.5$ GHz	-27.3969
$BW_1 = 1$ GHz, $BW_2 = 0.5$ GHz, with restoration	-27.6841

Table 2.6: EVM comparison, for an asymmetrical bandwidth limitation.

Finally, we consider cases with all three non-idealities. From table 2.4, we know that $\tau < \frac{T_c}{100}$ is necessary to obtain a useful output. If we add e.g. $G1 = 1.05$ and $BW = 2$ GHz, we obtain figure 2.29.

When $G1 = 1.01$, $\frac{T_c}{100}$, $BW_1 = 1$ GHz and $BW_2 = 2$ GHz, we obtain figure 2.30. By adding the bandpass (phase) correction, almost only linear ISI is observed.
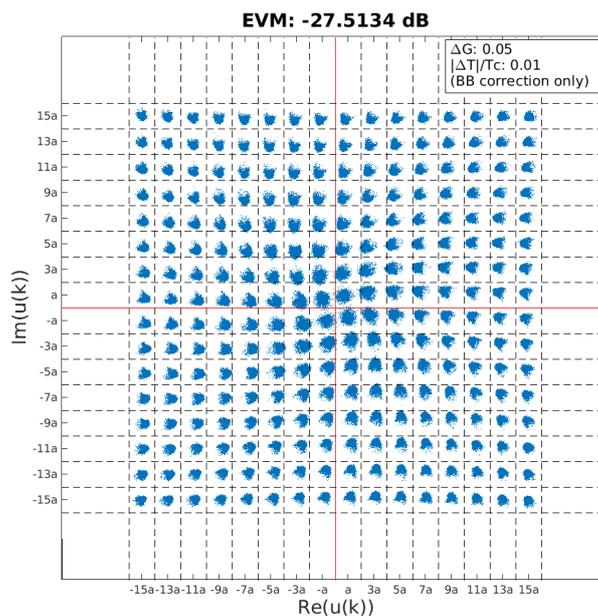
Figure 2.29: Simulated scatterplot, path 2 delayed by $\frac{T_c}{100}$, $G1 = 1.05$ and $BW = 2$ GHz.
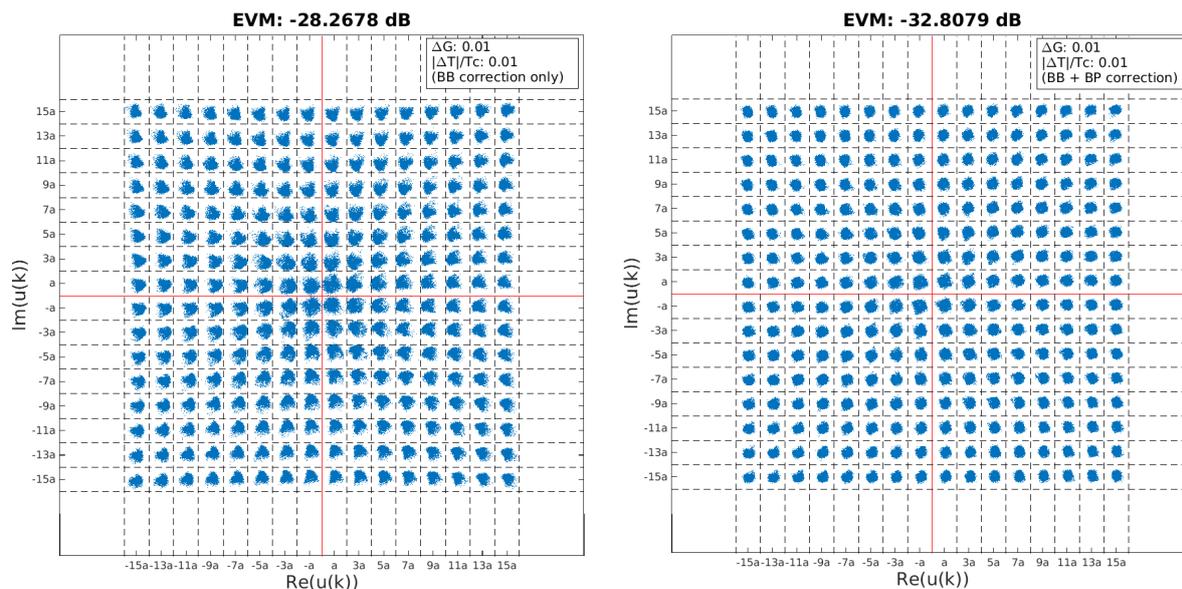
Figure 2.30: Simulated scatterplot, path 2 delayed by $\frac{T_c}{100}$, $G1 = 1.01$, $BW_1 = 1$ GHz and $BW_2 = 2$ GHz; and without (left) and with (right) bandpass/phase correction.

2.5 Outphasing testbench in Cadence

To determine the EVM accurately, a large number of symbols have to be transmitted. Due to the high carrier frequency, a direct simulation in time domain will take a lot of time for a given symbol interval. Fortunately, spectreRF supports an "envlp" or envelope simulation, in which the simulator tries to skip carrier cycles while still accurately simulating the envelope, so that no information is lost. Two options are available: shooting, which is a time-domain based approach suited for switching systems; and a harmonic-balance based simulation. The envlp-simulation does not only provide the transmitted envelope: the power spectrum, main

channel power, ACPR, ... can be obtained as well.

Initially, the baseband sources (rfVsource) which are already present in the rfLib, and which are configured with .bin-files, were used to test the concept and obtain an output EVM and ACPR, power spectrum... Some of the 802.11 standards are implemented in spectreRF: the user has to select the modulation source, the standard and some parameters; and calling the EVM-function after the envlp simulation gives the received constellation diagram and an EVM estimate. Because the simulator knows the standard, it knows the transmit pulse, so it knows how to demodulate the envelope and sample it afterwards. Unfortunately, none of the already implemented standards support 256-QAM constellations yet. The .bin-files cannot be read, so the information on how a standard is specified to the simulator is not available to the user.

At this point, we had to define our own "standard" by choosing the symbol rate and transmit/receive pulse. There are no specifications on coding because we are only interested in symbols, not in the BER. The data files for the baseband sources were created in Matlab. These files are used by the behavioral model of the component separator (written in Verilog A). The SCS-output is then upconverted to f_c, and the envlp simulation proceeds correctly. The shape of the output envelope is correct, but because the simulator does not know the used transmit pulse; it cannot demodulate it the envelope. The constellation plot which appears after calling the evmQAM()-function appeared to be the sampled envelope instead of the samples of the baseband demodulated envelope.

Attempts were made to do the baseband demodulation in an Ocean-script, because it would be convenient to do the entire EVM simulation and processing by executing a single script. First, a time domain convolution of the envelope with the receiver filter was attempted, but this proved very slow and the output was not as expected. Second, the calculation of the convolution was attempted in the frequency domain as the inverse Fourier transform of the product of the Fourier transforms of the envelope and the filter, but this did not work either.

The help of Cadence support was requested, and simultaneously, the simulated envelope was imported in Matlab with the help of the JVSpectre class, provided by dr.ir. Jochen Verbrugghe. This allowed the baseband demodulation and symbol correction (scaling, rotation) to be done in Matlab.

2.6 Comparison: testbench vs. simulation

The next step is to compare the output of the testbench with the output from the Matlab simulations (10^6 symbols) in the case of an ideal PA, as demonstrated in table 2.7. The harmonic balance - envlp simulation was used, with oversample factor 2 and 5 harmonics. This is a rather low number of harmonics, but it does not trigger a warning, the output is correct, and this number can be increased when real, non-linear PA's are used instead of perfect amplifiers. Data files containing 10^5 symbols, filtered with the same transmit pulse (SRRC, $L_f = 40$, $T = 10$ ns, $\alpha = 0.3$) are generated in Matlab and saved as .pwl-files. The number of symbols is quite large but not very large, to keep the simulation time acceptable.

Both simulations are not exactly identical due to multiple factors. Very small rounding errors occur when exporting the data to the .pwl-files, so the precision in Matlab is higher. The EVM is an average, so the results might differ slightly because the used data in the testbench is independent from the data of the Matlab simulation and a smaller number of symbols is used in the testbench. The data is also sampled at a different sampling rate: f_s is at least $\frac{8}{T}$ in Matlab because bandpass signals are represented. In the envlp simulation, the specified data is baseband data, so we need to satisfy the sampling theorem in for baseband signals instead

of bandpass signals. f_s was limited to $\frac{4}{T}$ (which is two times faster than strictly necessary) to speed up the simulation by allowing a larger (fixed) output time step, equal to $\frac{1}{f_s}$. These effects are combined but the relative error is very small, except for the ideal simulation, when the EVM is already very small. Overall, the EVM, scatterplots and power spectra agree well.

Condition	Simulated EVM	EVM from testbench
Ideal	-74.0196	-66.9322
$G_1 = 0.99$	-46.9557	-45.1866
$G_1 = 1.01$	-45.7595	-45.0655
$G_1 = 0.95$	-32.1999	-32.1606
$G_1 = 1.05$	-32.2725	-32.3370
$\frac{T_c}{100}$, path 2	-29.9193	-30.2281
$\frac{T_c}{50}$, path 2	-23.5385	-23.9027
$\frac{T_c}{30}$, path 2	-18.1532	-18.3265
$\frac{T_c}{20}$, path 2	-13.0024	-13.2977
$\frac{T_c}{100}$, path 1	-30.6144	-30.8213
$\frac{T_c}{50}$, path 1	-24.7511	-25.0830
$\frac{T_c}{30}$, path 1	-20.9809	-21.0490
$\frac{T_c}{20}$, path 1	-18.5581	-18.066

Table 2.7: EVM comparison. The corresponding power spectra are plotted below.
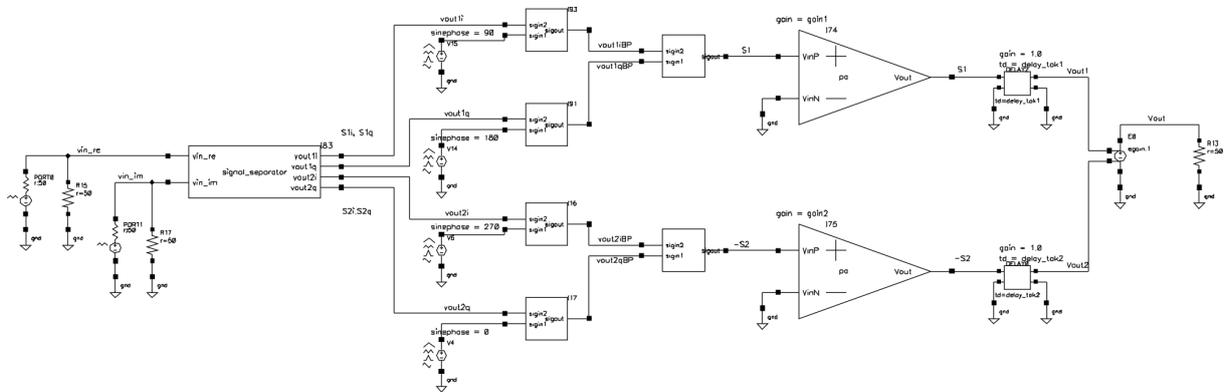
Figure 2.31: Schematic of the outphasing testbench in Cadence, with ideal amplifiers.
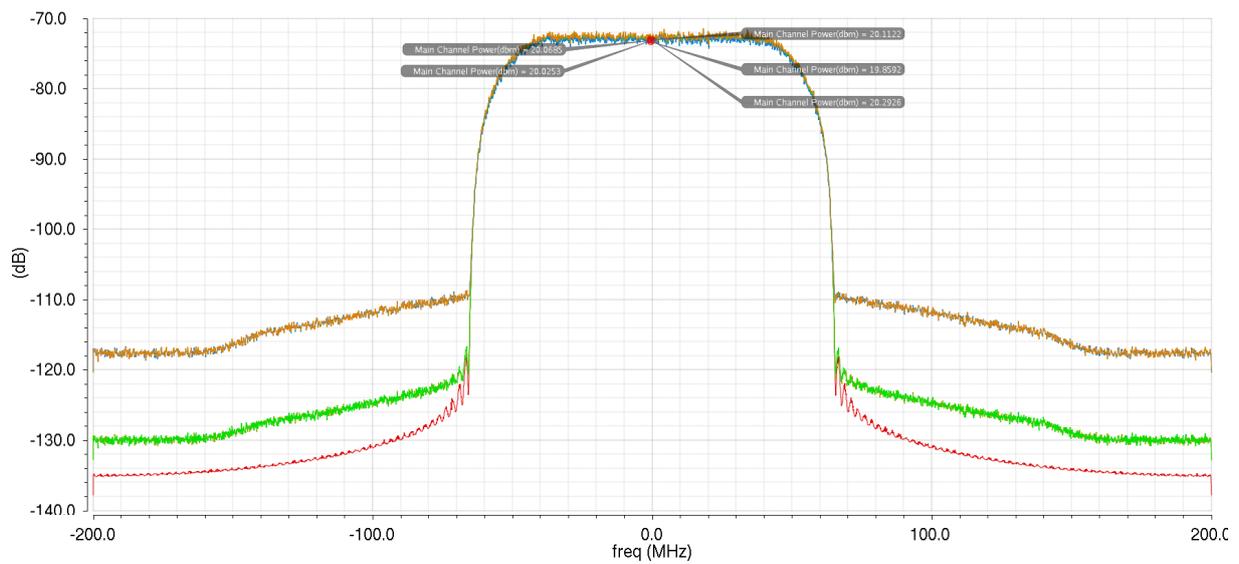
Figure 2.32: Simulated power spectrum, for $G_1 = 1$ (red) $G_1 = 0.99$, $G_1 = 1.01$, $G_1 = 0.95$ and $G_1 = 1.05$.
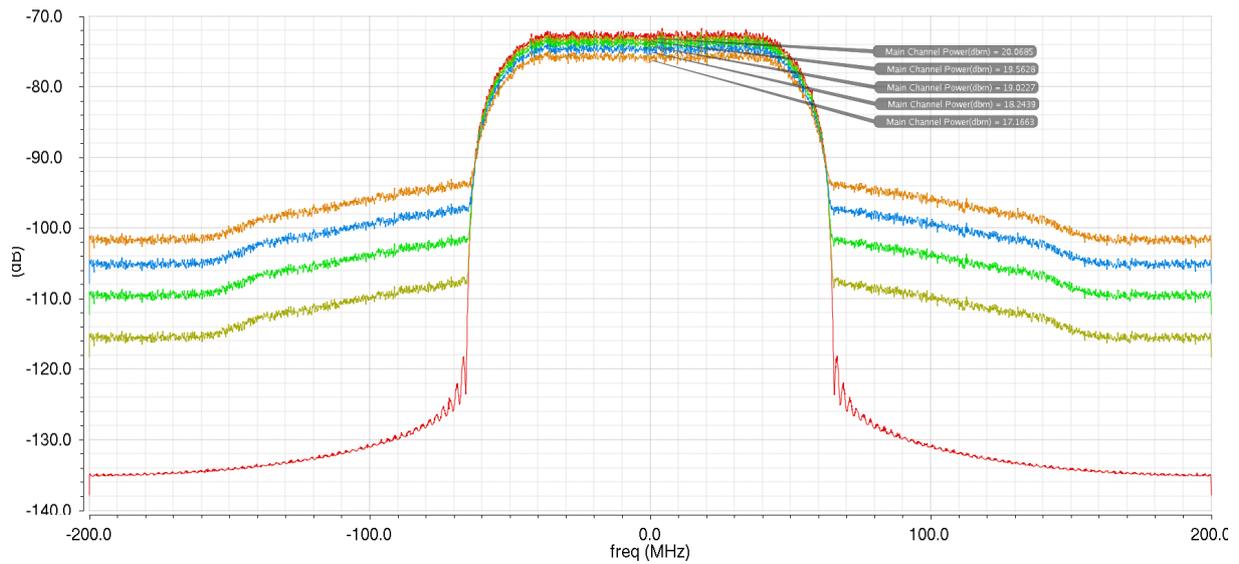
Figure 2.33: Simulated power spectrum, for the ideal case without delay (red) for a delay of $\frac{T_c}{100}$ (gold), $\frac{T_c}{50}$ (green), $\frac{T_c}{30}$ (blue) and $\frac{T_c}{20}$ (orange) in path 2.
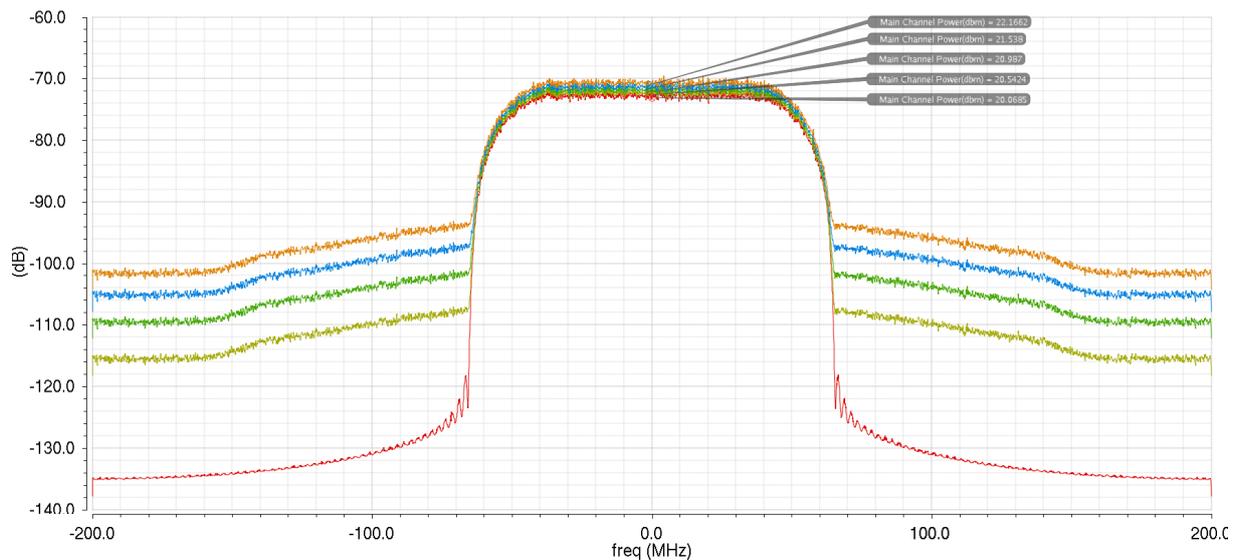
Figure 2.34: Simulated power spectrum, for the ideal case without delay (red) for a delay of $\frac{T_c}{100}$ (gold), $\frac{T_c}{50}$ (green), $\frac{T_c}{30}$ (blue) and $\frac{T_c}{20}$ (orange) in path 1.

2.7 Estimation and correction of gain and phase errors

A major disadvantage of the outphasing architecture is the need to closely match two high-frequency signal paths in order to obtain good output quality. This necessity was confirmed by the theoretical simulations in the previous sections. In reality, the physical layout should be as symmetrical as possible in order to avoid that mismatch is created in the first place; but even then; gain and delay calibration will be necessary. A mismatch calibration block and algorithm are therefore essential parts of the outphasing amplifier. The implementation of these calibration and matching blocks is outside of the scope of this thesis, but in the following section, some options, found in literature, will be discussed.

2.7.1 Error estimation

Accurately estimating the mismatch is a crucial first step in the calibration. Based on the results of the previous sections, some conclusions can be made. Afterwards, we will compare with realistic calibration systems.

Both gain and delay error contribute to the amplitude and phase error at the output. Detecting the exact gain and phase error from a constellation diagram involves "pattern recognition" and requires demodulation, which increases the complexity. They can however be decoupled by adding an amplitude detector at the output of each PA to detect the gain difference. Asymmetry in the combiner might cause an asymmetric phasor sum even with perfect input phasors, but these effects should already be captured in the reference values from a start-up procedure which can use test symbols. Therefore, we can now assume that the gain difference is corrected perfectly. Still, some symbol rotation and scaling will remain due to delay difference; the rotation angle is related to the magnitude of the delay but it does not indicate which branch is delayed: the rotation is clockwise in both cases. From the amplitude detection of the separate PA outputs (they provide a constant output amplitude in the ideal case), we know what the ideal average output power should be. We could now add a power detector at the output, and the sign of

the relative power change ΔP_{rec} of P_{out} w.r.t $P_{out,expected}$ could be used to indicate which path should be corrected.

In real calibration systems, feedback from the PA output back to the calibration block and signal component separator is added. The estimation can be based on the actual transmitted signal itself, or on the transmission of test symbols. Fast periodic recalibration based on test symbols could be unfavourable because these test sequences interfere with the transmitter operation.

Another major difference between calibration systems is the hardware in the feedback path: the most obvious solution is to add an entire downconversion feedback path (as in [16], [17]) to obtain the transmitted symbols, but this introduces extra complexity. This is is not always necessary: solutions based on power detection at the PA output only exist as well, but these might involve test symbols ([18]).

In [19], the demodulated output is used to build a linearized model of the entire system from baseband to output, including the non-linear amplifiers, based on Mean Square Error (MSE)-minimisation. When this model is known, linear equalizers are determined with the same MMSE-method. This algorithm is used continuously, making the calibration adaptive.

In [16], observation vectors are defined by grouping N samples of s_1, s_2 and the demodulated output. The correction is done by multiplying (the samples of) path 2 with γ, which simultaneously corrects for gain and phase imbalance. The estimated optimal value of γ is determined by solving equations which minimize the MSE based on the observations. Because of the AM-AM and AM-PM of the amplifier, the phase and gain of the PA in path 2, which now has $\gamma s_2(t)$ as an input, will not be exactly equal to the $\phi + \Delta\phi$ and $G + \Delta G$ from the case when the input was γ, so the algorithm has to be iterative. Increasing the number of samples per symbol period increases the convergence speed.

Figure 2.35 demonstrates the PA-characteristics that were assumed in [17] (p. 387).
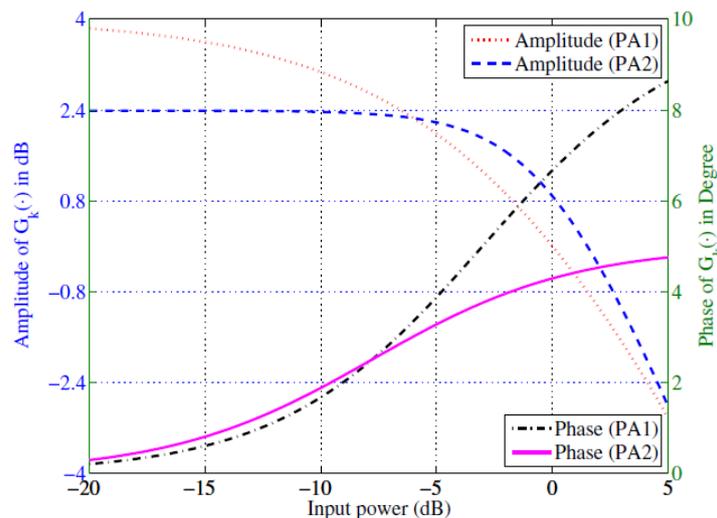
Figure 2.35: Example of assumed PA-characteristics.

2.7.2 Error correction

Once the estimates of ΔG and $\Delta\phi$ (and τ) are known, corrections can be made. After the simulation of the effect of a delay error, we already proposed to anticipate the rotation due to

$e^{-j2\pi f_c \tau}$ in one of the paths by rotating the symbols for that path in the opposite direction. As demonstrated in the previous section, this simple correction is already very effective because the carrier frequency is much higher than the symbol period, so the effect of the delay in the baseband is very limited if the rotation is corrected. If we wish to eliminate the linear ISI as well, we can only delay the non-delayed path and multiply the associated symbols with $e^{-j2\pi f_c \tau}$, to cancel $e(t)$ at the output. The receiver will not suffer from this correction because the channel also introduces delay, and each receiver has to estimate the carrier phase for coherent detection. However, this might not be realisable because the delays are very small.

Correcting a gain difference (e.g. due to transistor or biasing mismatch) is much more problematic: we intend to use high-efficiency amplifiers, which are either saturated or switched to reduce the voltage-current overlap over the active element. In this case, the output amplitude is determined by the supply voltage instead of the input amplitude (e.g. for a class-E PA), as long as the input amplitude is high enough to satisfy the switching requirement. This implies that performing the gain correction in the baseband path has (almost) no effect at all (in this case), and it seems at first that we have no option than to correct in the RF path.

Three solutions exist:

- The PA output amplitude could be controlled by making supply voltage of each PA adjustable over a given range. Two major disadvantages arise: good cancellation of $e(t)$ will could require quite fine control. Secondly, the distortion characteristics (PM-PM, AM-PM...) of the PA's are dependent on the supply voltage. This implies that correcting the gain difference by making the supply of one PA different from the other, will introduce a differential phase shift and will interfere with the correction of the phase/delay. This makes the convergence of the calibration algorithm more difficult and slower.

- For certain switching PA's, the output amplitude can be adjusted per path by controlling the duty cycle of the switching input waveform for each path. For example, NXP's outphasing systems use patented duty cycle control technology. ([20], p.12). In this case, the duty cycle control even enables the same PA to be used in different frequency bands (patent: US 8174322 B2).

- The most promising and generally applicable option seems to be avoiding amplitude adjustments entirely and controlling the output amplitude by introducing phase imbalance (cfr. figure 2.36, from [18]): either $s_1(t)$ or $s_2(t)$ is intentionally rotated over an angle $\phi_{calibration}$. This changes the amplitude and phase of $s_1(t) + s_2(t)$; but the phase change can be compensated for directly by rotating both $s_1(t)$ and $s_2(t)$ over the opposite angle. Because the amplitudes cannot be changed, distortion will occur if the intended output amplitude is smaller than ΔGA_{max}: we can then only rotate the outphasing vectors to be opposite, but still a vector with length ΔGA_{max} leaks through. This is compensated in advance: once ΔG is known, the amplitude of the baseband signals is not allowed to decrease below ΔGA_{max} ("vector hole punching").

In [18], an estimation of $\Delta\phi$ and ΔG is done based on a test sequence of only 5 cases and output power detection, no downconversion is used. The sign of the amplitude mismatch is determined by temporarily shutting down one PA at a time and comparing the total output power. Next, the output power is calculated for three other cases: in-phase inputs (maximal output power), out-of-phase inputs (minimal output power) and in-quadrature-inputs. The three measured output powers depend on ΔG and the sine or cosine of $\Delta\phi$, so these equations can be solved to obtain ΔG and $\Delta\phi$. $\Delta\phi$ is compensated first, and then the new outphasing angles are calculated which correct for the gain imbalance as well (cosine law). Because both the estimation and the correction do not involve amplitude control, the iterative steps and convergence problems due to AM-AM and AM-PM are avoided, and the entire correction can be done in the baseband

section. Still, if e.g. the PM-PM characteristics of the PA's are not equal up to a constant $\Delta\phi$, the phase difference depends on the input phase but the correction that is used, does not; so some mismatch remains, unless the phase of the test vectors is swept over the entire range and the $\Delta\phi$ is known for every input phase.

Finally, we remark that this approach is very similar to the AMO-system, in which phase asymmetry was introduced even in the ideal case, to improve the efficiency. Combining AMO with this type of calibration seems a perfect match, but in that case, the calibration will have to be done for each supply setting (for each level) to take the supply-dependence of the PM-PM functions into account.
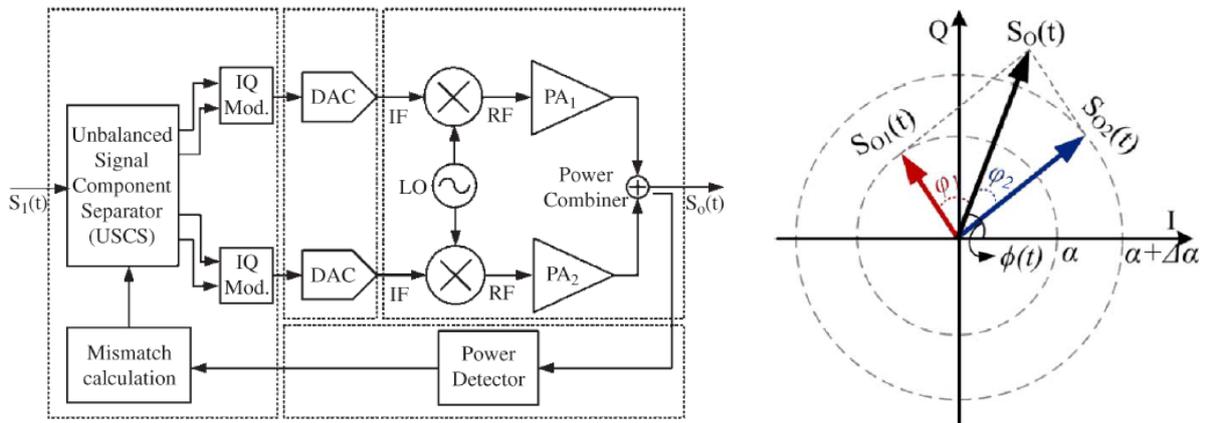
Figure 2.36: System block diagram for unbalanced phase calibration.

Part II

Active components

Chapter 3

Active components

3.1   EKV-(B) model

The EKV-B model, as used in [21], provides a continuous expression for the drain current in the saturation region (pentode region) over all inversion regions by introducing an inversion coefficient IC ([22]):

$$I_{DS} = I_0 \frac{W}{L} IC \tag{3.1}$$

The technology current I_0 is given by:

$$I_0 = 2n_0 \mu_0 C'_{ox} U_T^2 \tag{3.2}$$

μ_0 is the carrier mobility, n_0 is the substrate factor, C'_{ox} is the gate-oxide capacitance per unit area and U_T is the thermal voltage:

$$U_T = \frac{kT}{q} \tag{3.3}$$

The substrate factor n models a loss of coupling efficiency between the gate and the channel. In weak inversion, a capacitive voltage divider from the gate to the channel is formed because of the gate-oxide, depletion, and interface state capacitances.

$$n, WI \approx 1 + \frac{C'_{dep}}{C'_{ox}} \tag{3.4}$$

The inversion coefficient is determined by this coupling factor n, the threshold voltage V_{th} and V_{GS}:

$$IC = \left(ln \left(1 + e^{\frac{V_{GS} - V_{th}}{2nU_T}} \right) \right)^2 \tag{3.5}$$

The regions of inversion with the corresponding IC are given in figure 3.1 ([22], p.54).
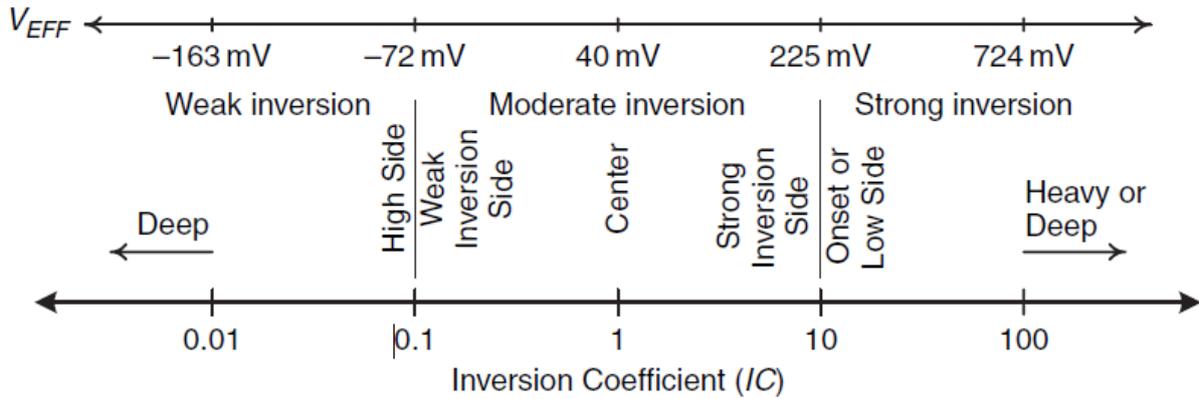
Figure 3.1: Regions of inversion, with the corresponding IC ($V_{EFF} = V_{GS} - V_{th}$).

The saturation voltage $V_{DS,sat}$ is given by:

$$V_{DS,sat} = 2U_T\sqrt{IC + 0.25} + 3U_T \tag{3.6}$$

In weak inversion, $V_{DS,sat} \approx 4U_T$; in strong inversion, $V_{DS,sat} \approx 2U_T\sqrt{IC + 0.25} \approx \frac{V_{GS} - V_T}{n}$.

With this definition of IC, the transconductance efficiency $\frac{g_m}{I_{DS}}$ is given by:

$$\frac{g_m}{I_{DS}} = \frac{1}{nU_T(\sqrt{IC + 0.25} + 0.5)} = \frac{2}{n(V_{DS,sat} - 2U_T)} \tag{3.7}$$

In weak inversion, $\frac{g_m}{I_{DS}}$ is independent of IC, and consequently, of I_{DS} as well:

$$\frac{g_m}{I_{DS}}, WI \approx \frac{1}{nU_T} \tag{3.8}$$

Well into strong inversion, but before velocity saturation occurs, another asymptote exists:

$$\frac{g_m}{I_{DS}}, SI \approx \frac{1}{nU_T\sqrt{IC}} \tag{3.9}$$

In the velocity saturation region, $\frac{g_m}{I_{DS}}$ becomes inversely proportional to I_{DS} and IC:

$$\frac{g_m}{I_{DS}}, SI, v_{sat} \propto \frac{1}{IC} \tag{3.10}$$

When $\log_{10}\left(\frac{g_m}{I_{DS}}\right)$ is plotted as a function of $\log_{10}\left(I_{DS}\right)$, these approximations result in three asymptotes, with slope zero, slope $\frac{-1}{2}$ and slope -1 respectively. The intersection of the SI-asymptote and the SI-with-velocity-saturation asymptote occurs at a critical inversion coefficient IC_{CRIT}. To model I_{DS} more accurately near and in the velocity saturation region, IC is replaced by $IC\left(1 + \frac{IC}{IC_{CRIT}}\right)$ ([22]):

$$I_{DS} = I_0\frac{W}{L}IC\left(1 + \frac{IC}{IC_{CRIT}}\right) \tag{3.11}$$

Therefore, the real transconductance efficiency at IC_{CRIT} is only a fraction $\frac{1}{\sqrt{2}}$ (or 3 dB less) of the $\frac{g_m}{I_{DS}}$ that would be obtained if there was no velocity saturation and VFMR.

Beyond IC_{CRIT}, velocity saturation effects and VFMR (vertical field mobility reduction) effects occur, which can reduce the slope of $\log_{10}\left(\frac{g_m}{I_{DS}}\right)$ [$\log_{10} I_{DS}$] below -1, especially in short-channel transistors. As V_{GS} increases, V_{DS} must be increased as well to keep the transistor in the pentode region. The resulting horizontal electric field reaches a critical field strength E_{CRIT}, and the drift velocity of the charge carriers saturates. This velocity saturation implies a saturation of the transconductance g_m, so $\frac{g_m}{I_{DS}} \propto \frac{1}{I_{DS}}$. The strong electric field orthogonal to the channel, caused by the high V_G (and with $V_{SB} = 0$, a high V_{GB}), attracts the charge carriers towards the insulator-inversion layer interface, where the mobility reduces due to interactions with interface states. Due to the high saturation voltage and poor $\frac{g_m}{I_{DS}}$, this region of inversion is not favourable. Choosing a very low level of inversion, deep into weak inversion, gives a high g_m for a given current but also implies a very wide transistor (high $\frac{W}{L}$), resulting in large parasitic capacitances. Achieving high bandwidth therefore implies a trade-off between g_m and $C_{parasitic}$. In general, a good trade-off between a good $\frac{g_m}{I_{DS}}$, a low current I_{DS} and reasonable bandwidth is found in or near the "moderate inversion" region, around $IC \approx 1$ and with an intermediate $\frac{g_m}{I_{DS}}$; but the optimal choice of IC depends on the situation. A point in the MI-region could be used as a starting point for the optimisation.

3.2 Model parameter extraction

To extract n, I_0 and V_{th}, we inject current into the drain of the test-transistor (cfr. figure 3.2) and we keep the transistor in the pentode region, where the model is applicable, by forcing V_{GS} and V_{DS} to be equal by an ideal voltage controlled voltage source. We save all DC model parameters, currents and voltages. These are imported into Matlab with the aid of the provided JVSpectre class.

The choice of the dimensions of the test-transistor is a trade-off: we need to be able to distinguish all inversion regions clearly, so the transistor has to enter the velocity saturation region before a breakdown-voltage is reached (cfr. section 3.6). This implies a short channel length and a relatively small width for a given current sweep. On the other hand, the channel length cannot be too short or the transistor will enter velocity saturation too soon; which will not allow us to separate asymptote with slope $\frac{-1}{2}\log_{10}\left(\frac{g_m}{I_{DS}}\right)$-plot from the velocity saturation asymptote with slope -1. Therefore, the channel length should be long enough. In this case (cfr. figure 3.3), L is not very large: e.g. $L >\approx 2L_{min}$.

From the asymptotes in the $\log_{10}\left(\frac{g_m}{I_{DS}}\right)$ [$\log_{10} I_{DS}$]-plot, the technology current I_0, coupling factor n and IC_{CRIT} can be obtained:

- In deep weak inversion, we obtain n for a given temperature by: $n = \frac{1}{(\frac{g_m}{I_{DS}})U_T}$.

- At the intersection of the weak-inversion and strong-inversion asymptote, we know from equations 3.8 and 3.9 that $IC = 1$, so: $I_0 = \frac{I_{DS}}{\frac{W}{L}}$.

- To calculate IC, we need V_{th}, which can be derived from extrapolation of the strong inversion current. The traditional formula for the strong-inversion current in the pentode region is: $I_{DS} = \frac{\mu C'_{OX}}{2}\frac{W}{L}(V_{GS} - V_{th})^2$. In strong inversion, I_{DS} must be proportional to $(V_{GS} - V_{th})^2$, regardless of the model; so by plotting $\sqrt{I_{DS}}$ as a function of V_{GS}, drawing

a linear tangent line and extrapolating this line to the intersection with the V_{GS}-axis, we obtain an approximation of V_{th}.

- Now that we have obtained all necessary parameters, we can calculate I_{DS} based on equation 3.1 and compare with the measurement. This provides verification, e.g. for V_{th}.
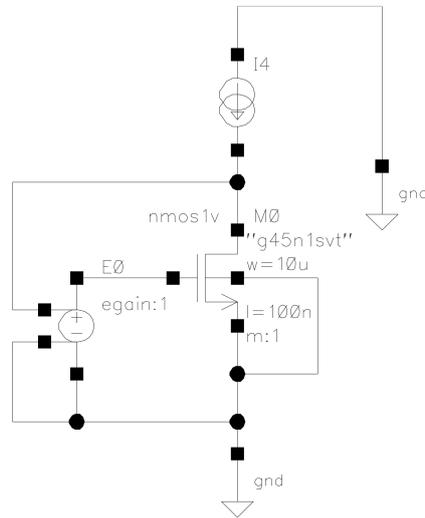
Figure 3.2: Test schematic of a 1.1V standard-V_{th} NMOS.
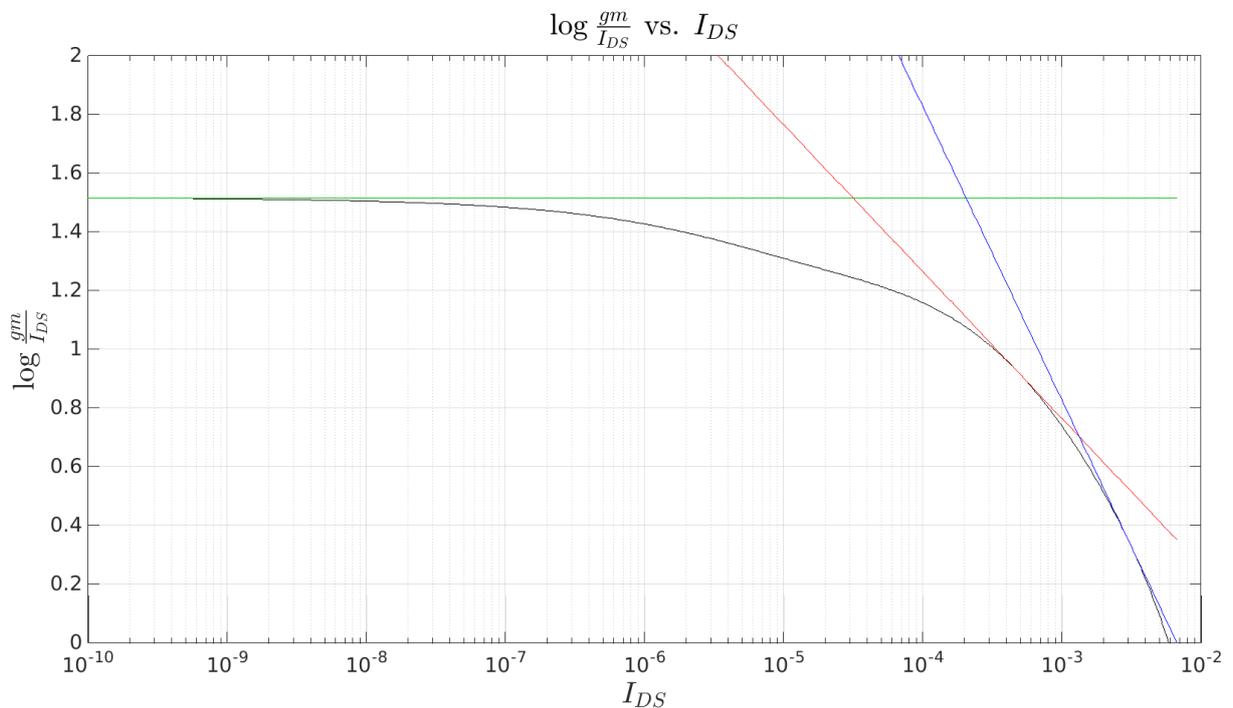
Figure 3.3: $\log_{10}\left(\frac{g_m}{I_{DS}}\right)$ as a function of I_{DS} (logarithmically), with the three asymptotes, for a 1.1 V standard-V_{th} NMOS with $\frac{W}{L} = \frac{10\mu m}{0.1\mu m} = 100$.
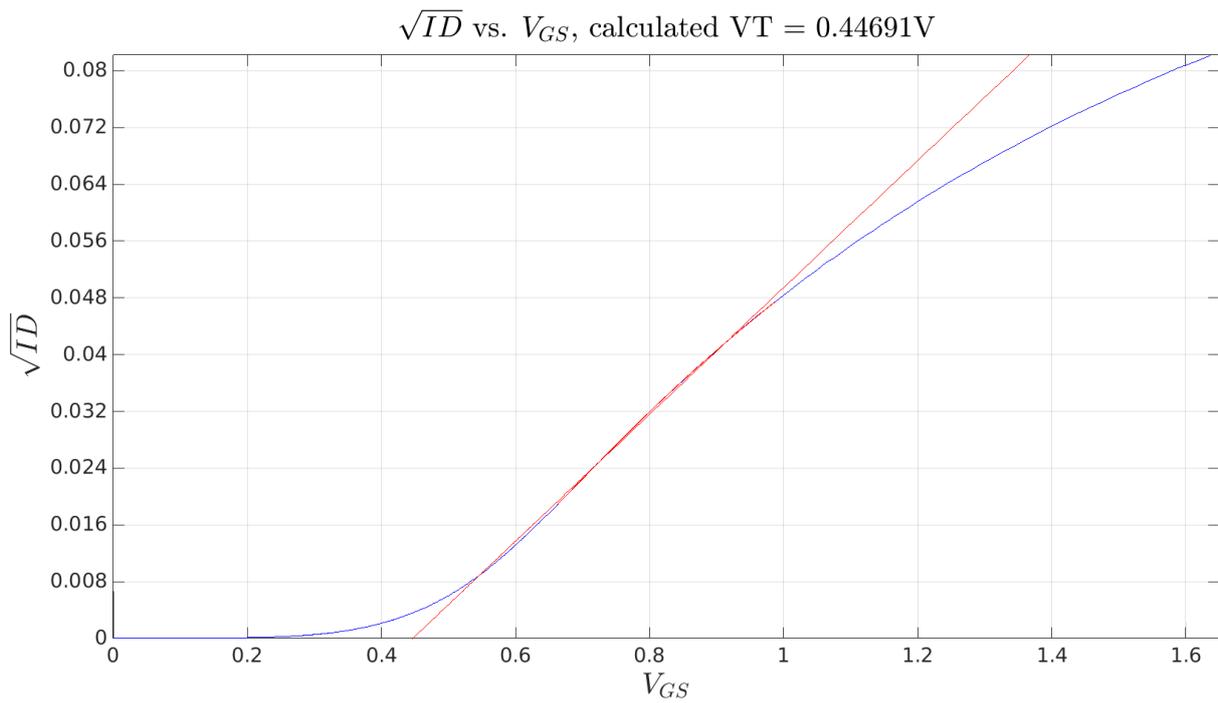
Figure 3.4: $\sqrt{I_{DS}}(V_{GS})$, for a 1.1 V standard-V_{th} NMOS with $\frac{W}{L} = \frac{10\mu m}{0.1\mu m} = 100$.
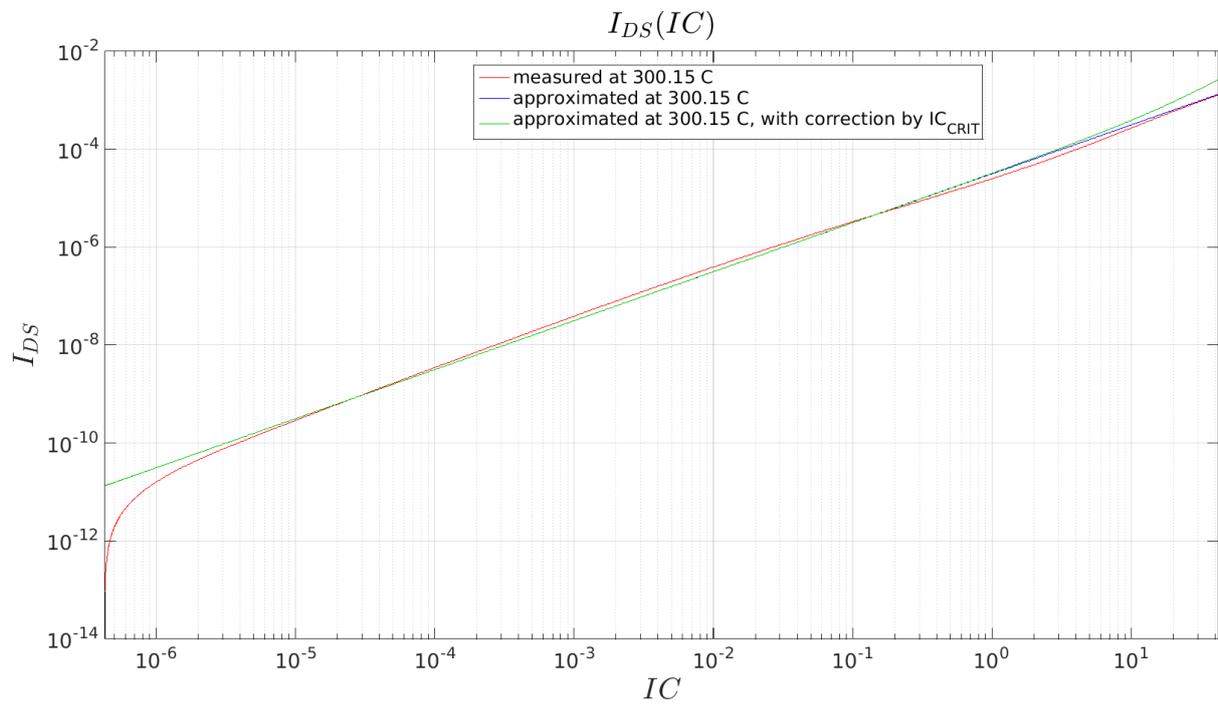
Figure 3.5: Comparison of the measured I_{DS} and the I_{DS} calculated based on the estimated I_0, n and V_{th}; for a 1.1 V standard-V_{th} NMOS with $\frac{W}{L} = \frac{10\mu m}{0.1\mu m} = 100$.

3.3 Parameter summary

Type	W (μm)	L (μm)	$I_0(nA)$	n	$V_T(V)$	IC_{CRIT}	$I_{CRIT}(mA)$
1.1V, svt	1	0.1	339.605	1.178	0.4353	41.594	0.14
1.1V, svt	10	0.1	313.912	1.178	0.4469	42.188	1.3243
1.1V, svt	100	0.1	400.848	1.178	0.4219	40.274	16.144
1.8V, svt	1	0.3	413.585	1.249	0.45	92.802	0.12794
1.8V, svt	10	0.3	406.853	1.248	0.46	94.121	1.2764
1.8V, svt	100	0.3	458.683	1.249	0.46	90.284	13.804
1.1V, lvt	1	0.1	350.511	1.182	0.3458	48.562	0.17022
1.1V, lvt	10	0.1	325.297	1.182	0.3569	48.721	1.5849
1.1V, lvt	100	0.1	417.331	1.184	0.3336	46.4	19.364
1.1V, hvt	1	0.1	319.362	1.228	0.5169	36.874	0.11776
1.1V, hvt	10	0.1	294.536	1.228	0.5271	37.572	1.1066
1.1V, hvt	100	0.1	377.741	1.228	0.5014	35.794	13.521
1.1V, nat	3	0.3	566.197	1.269	0.101	93.332	0.52845
1.1V, nat	30	0.3	641.813	1.257	0.0984	90.697	5.821
1.1V, nat	300	0.3	757.609	1.279	0.0881	72.369	54.828
1.8V, nat	10	1	1974.147	2.359	-0.0194	(111.076)	2.1928
1.8V, nat	100	1	2110.319	2.388	-0.0234	(108.306)	22.856
1.8V, nat	1000	1	2146.571	2.394	0.00260	(41.424)	88.92

Table 3.1: Parameter summary for the NMOS-transistors (svt = standard-V_T, lvt = low-V_T, hvt = high-V_T, nat = native).

The IC_{CRIT} of the 1.8V native NMOS-transistors is not accurate: within the given voltage range, no velocity-saturation region was found in the $\frac{g_m}{I_{DS}}$-plot. IC_{CRIT} was then automatically calculated by "fitting" the "slope -1"-asymptote to the lowest $\frac{g_m}{I_{DS}}$-points, but this is not accurate as the intersection point with the SI-asymptote now becomes dependent on the maximal current or IC which is considered; and "IC_{CRIT}" then becomes dependent on $\frac{W}{L}$ as well; which should not be the case. The IC_{CRIT} of the 1.1 V standard-V_{th} NMOS-transistors is also high, but in this case, a "slope -1"-asymptote was found. The IC_{CRIT} is in this case is indeed approximately independent of $\frac{W}{L}$.

Type	W (μm)	L (μm)	$I_0(nA)$	n	$V_T(V)$	IC_{CRIT}	$I_{CRIT}(mA)$
1.1V, svt	1	0.1	262.1974	1.1389	0.4166	37.7895	0.09908
1.1V, svt	10	0.1	276.2039	1.1556	0.4197	34.8153	0.9616
1.1V, svt	100	0.1	241.1187	1.1556	0.4272	39.7896	9.594
1.8V, svt	1	0.3	290.0229	1.132	0.455	35.6196	0.034435
1.8V, svt	10	0.3	287.3204	1.1319	0.455	33.8653	0.03243
1.8V, svt	100	0.3	223.0272	1.1319	0.455	37.7366	2.8054
1.1V, lvt	1	0.1	269.1636	1.1687	0.335	34.2755	0.092257
1.1V, lvt	10	0.1	285.4579	1.1884	0.335	31.5833	0.90157
1.1V, lvt	100	0.1	257.3668	1.1884	0.335	35.8466	9.2257
1.1V, hvt	1	0.1	205.407	1.1389	0.492	30.9303	0.063533
1.1V, hvt	10	0.1	215.9129	1.1557	0.492	29.0216	0.62661
1.1V, hvt	100	0.1	192.4287	1.1556	0.492	34.3345	6.6069

Table 3.2: Parameter summary for the PMOS-transistors (svt = standard-V_T, lvt = low-V_T, hvt = high-V_T). All voltages and current are noted as positive voltages and currents.

The individual parameters do not always agree very well with the parameters provided by the model: e.g. for the 1.1V standard V_{th}-NMOS, the model gives $V_{th} \approx 0.515V$ instead of the derived $V_{th} \approx 0.4219V$. However, the used BSIM4-model is a much more complicated model than the formulas given in section 3.1, and it is physics-based, rather than based on fitting formulas on simulated graphs. The V_{th} given by the model is dependent on L (e.g. $V_{th} \approx 0.534V$ with $L = 0.1\mu m$ for $\frac{W}{L} = 10, 100, 1000$ and $V_{th} \approx 0.39V$ for $L = 1\mu m$). The set of parameters given in tables 3.3 and 3.4 does give a good approximation of the current in the relevant regions (e.g. not for very small currents), which was the only intention of the curve fitting.

At first sight, it seems strange that the ratio of the technology currents of e.g. the 1.1V standard V_{th}-NMOS and PMOS is only about max. $\frac{400nA}{241nA} \approx 1.66$, while typically $\frac{\mu_n}{\mu_p} \approx 3$ (e.g. 3.1 in normal Si at 300 K with $\mu_n = 1400\frac{cm^2}{V \cdot s}$ and $\mu_p = 450\frac{cm^2}{V \cdot s}$, and $\frac{\frac{K_P}{n},NMOS}{\frac{K_P}{n},PMOS} \approx 3.4$ in the 0.35 μm-technology in [23]). Because of equation 3.2 and because the derived n_{NMOS} is only slightly larger than n_{PMOS}, the ratio of the technology currents is expected to be about equal to this ratio of mobilities. However, the mobilities are strongly dependent on doping and $\frac{\mu_n}{\mu_p}$ decreases as the doping concentration increases.

This ratio of $I_{0,NMOS}$ and $I_{0,PMOS}$ was checked in simulation: a 1.1V standard V_{th}-NMOS with $\frac{W}{L} = \frac{100\mu m}{0.1\mu m}$ and a 1.1V standard V_{th}-NMOS with $\frac{W}{L} = \frac{180\mu m}{0.1\mu m}$ were simulated in the identical situation as figure 3.2; with $V_{GS,NMOS} = V_{in} + \Delta V_{th}$ and $V_{GS,PMOS} = V_{DD} - V_{in}$; and V_{in} was swept. $\Delta V_{th} = V_{th,NMOS} - |V_{th,PMOS}|$ was based on the threshold voltages provided by the model. For $L = 0.1\mu m$, we obtained $\Delta V_{th} \approx 0.515 - 0.489 = 26mV$. In this case, with $\frac{W}{L}, PMOS = 1.8 \cdot \frac{W}{L}, NMOS$, overlapping currents (with the opposite sign) were found. With $L = 1\mu m$, $\Delta V_{th} \approx 0.39 - 0.33 = 60mV$, it was found that $\frac{W}{L}, PMOS = 2 \cdot \frac{W}{L}, NMOS$ is necessary to obtain the same currents. This approximately confirms the obtained ratio of the I_0's. The same was done for 1.8 V standard-V_{th}-transistors.

3.4   High-frequency current modelling

While design a class B-PA, it was found that, from a certain frequency on, the predicted current for large $\frac{W}{L}$ becomes incorrect, as shown in figure 3.6. In this test circuit, a standard - V_{th} 1.8 V NMOS with $W = 200\mu m$ and $L = 150nm$ is used. At 1 GHz, the drain current appears to be correct: the gate is biased at V_{th} and a sine is applied, so I_{DS} is approximately a positive half-sine. When the frequency in this same test circuit is increased to 15 GHz, non-physical current waveforms appear, making the correct PA operation impossible, the voltages are almost constant. This problem was solved by decreasing $\frac{W}{L}$ by a factor 10 and placing 10 identical transistors in parallel. Afterwards, $\frac{W}{L}$ was decreased even further, giving better results. Although the resulting circuit is not exactly the same, the discrepancy in the behaviour should not be this large. The same problem was also confirmed in the first class-E designs that were made. This problem was reported to Cadence support, where it was confirmed that this is most likely due to an issue in the model at high frequencies.

We decided to continue with a large number of parallel transistors with a lower $\frac{W}{L}$-ratio. At first sight, this creates a problem with the parasitic capacitances: by sufficiently increasing the number of fingers for a fixed and high $\frac{W}{L}$-ratio, the effective drain and source area, and the parasitic capacitances associated with it, can be roughly halved. This is now not possible: we have to keep the $\frac{W}{L}$ and the finger size constant, so increasing the multiplier also increases the parasitic capacitances with the same factor. This is however not too unrealistic: at high frequencies, it is often observed that the gate poly resistance becomes too large to enable a large number of small fingers because of the resulting long gate length, which makes $V_{GS,effective}$ decrease along the length of the gate. A wider transistor is then also realised by putting multiple transistors in parallel. The parasitics then become quite large, as demonstrated in figures 3.8 and 3.9.
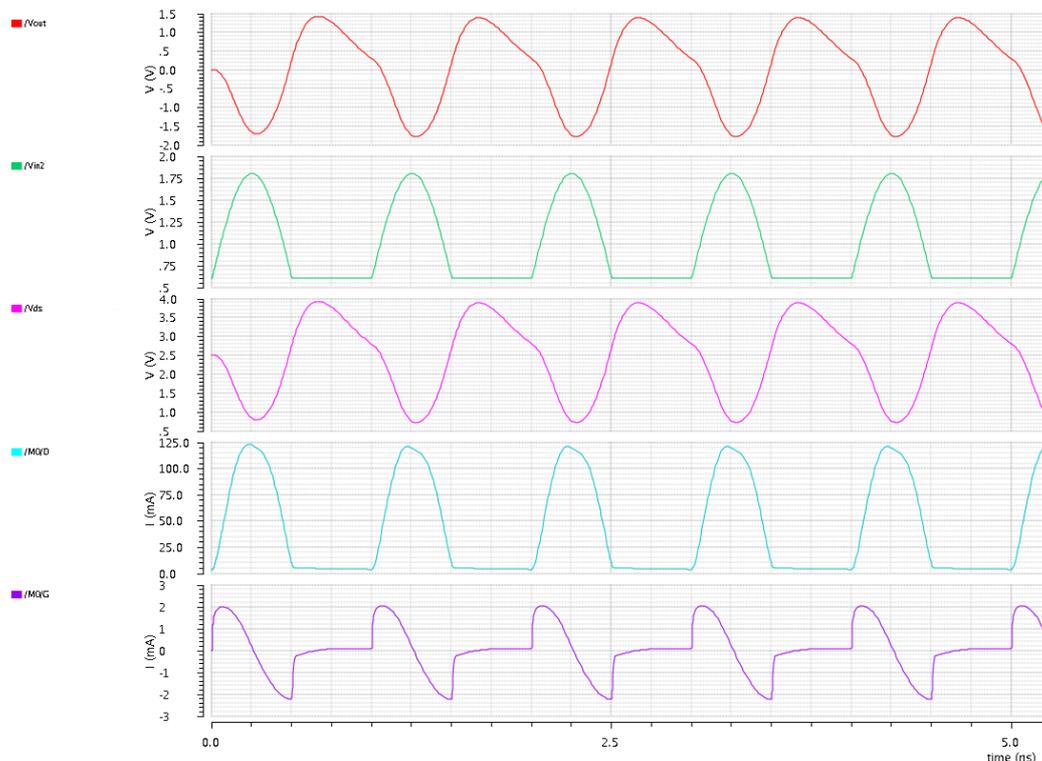
Figure 3.6: Waveforms in an early 1 GHz class B - test case.
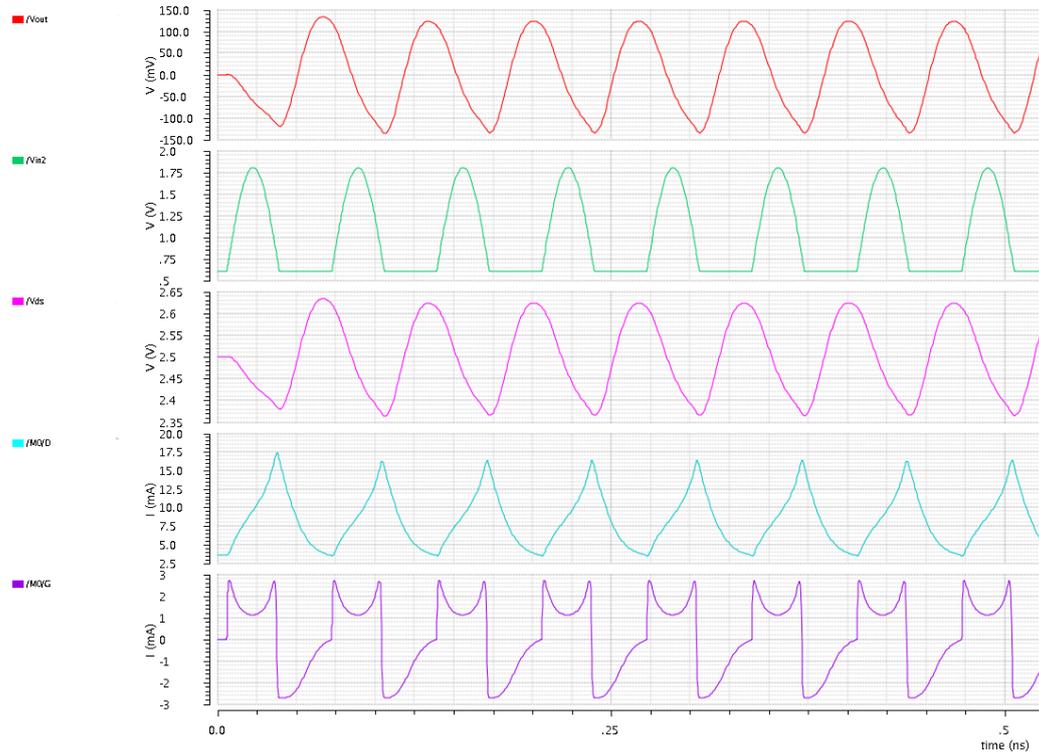
Figure 3.7: Waveforms in the same class B - test case, at 15 GHz.

3.5 Parasitic capacitances

The parasitic capacitances were determined from the same simulation, and checked separately.

In the triode region, the parasitic capacitances can be approximated by:

$$C_{gs} = C_{gso} \cdot W + \frac{1}{2} C_{ox} W L \tag{3.12}$$

$$C_{gd} = C_{gdo} \cdot W + \frac{1}{2} C_{ox} W L \tag{3.13}$$

In the pentode region:

$$C_{gs} = C_{gso} \cdot W + \frac{2}{3} C_{ox} W L \tag{3.14}$$

$$C_{gd} = C_{gdo} \cdot W \tag{3.15}$$

The source and drain capacitance are composed of a constant term and a voltage-dependent term:

$$C_{db,tot} \approx C_{db,o} W + C_{jd} W \approx C_{db,tot,avg} \cdot W \tag{3.16}$$

$$C_{sb,tot} \approx C_{sb,o}W + C_{js}W \approx C_{sb,tot,avg} \cdot W \tag{3.17}$$
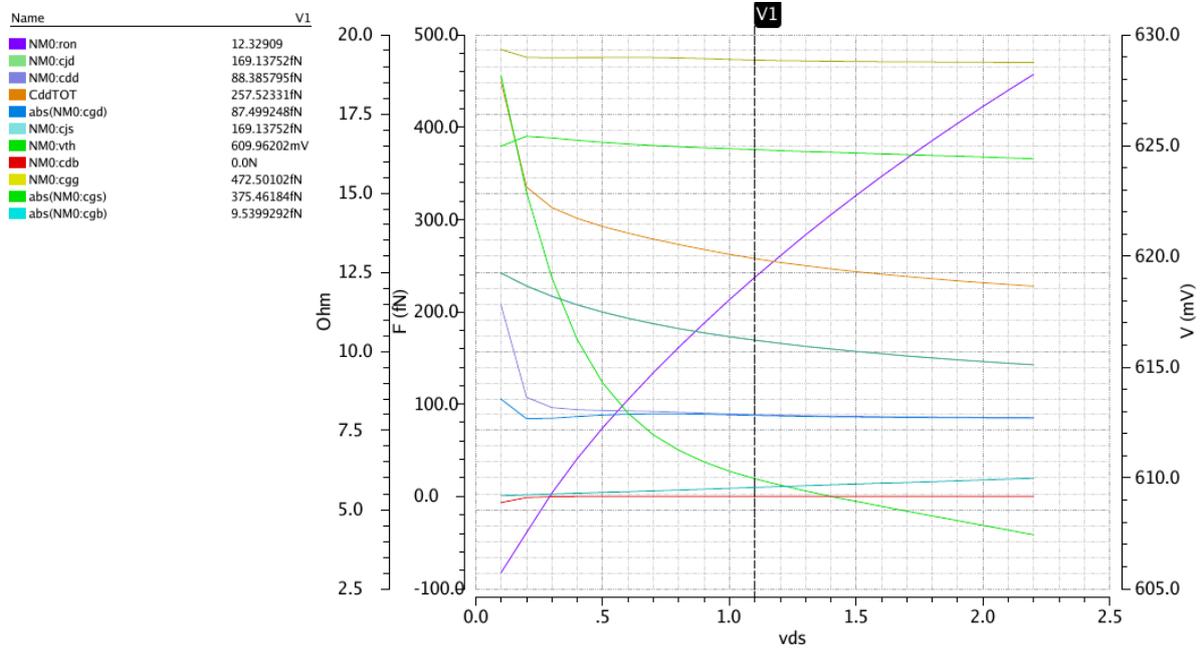
Figure 3.8: Capacitances provided by the model, for $V_{GS} = 0.8V$, as a function of V_{DS}, for a large 1.1V standard-V_{th} NMOS, with $\frac{W}{L} = \frac{2\mu m}{45nm}$ (1 finger) and multiplier 300.
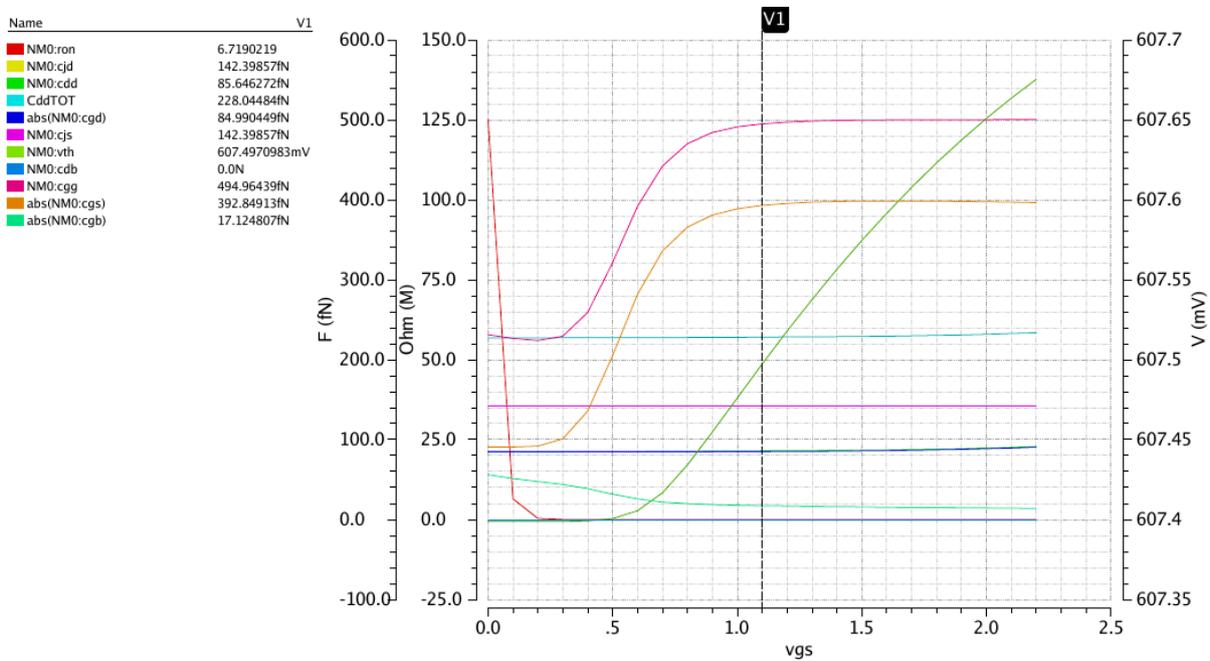
Figure 3.9: Capacitances provided by the model, for $V_{DS} = 2.2V$, as a function of V_{GS}, for a large 1.1V standard-V_{th} NMOS, with $\frac{W}{L} = \frac{2\mu m}{45nm}$ (1 finger) and multiplier 300.

Modelling the entire voltage dependence is not necessary in the first design steps, therefore, some simplifications were made:

- C_{gd} is approximately independent of IC, so the average is calculated.

- The intrinsic source-bulk and drain-bulk capacitances "cdb" and "csb" are very small (cfr. figures 3.8 and 3.9), they are neglected. The junction capacitances "cjs" and "cjd" (they are equal in figure 3.8) are voltage-dependent: e.g. the inverse biased drain-bulk diode behaves as a varicap. They are large and therefore not neglected, but the voltage dependence is averaged out.

- As demonstrated in figure 3.9, where the transistor is continuously in pentode because V_{DS} is high, C_{gs} increases strongly with increasing V_{GS}, especially in the region around V_{th}; because of the build-up of the inversion charge-channel. Because $C_{gs} \approx C_{gso} \cdot W + \frac{2}{3} C_{ox} W L$ for a transistor in pentode and in inversion, the "effective" $C_{ox,eff}$ is calculated by $C_{ox} = \frac{C_{gs,highIC} - C_{gs,lowIC}}{\frac{2}{3} \frac{W}{L}}$.

Type	W	L	$C_{gso}(F/m)$	$C_{ox}(F/m^2)$	$C_{gdo}(F/m)$	$C_{js,avg}(F/m)$	$C_{db,tot,avg}(F/m)$
1.1V, svt	1	0.1	$1.498 \cdot 10^{-10}$	0.0178	$1.509 \cdot 10^{-10}$	$4.398 \cdot 10^{-10}$	$3.754 \cdot 10^{-10}$
1.1V, svt	10	0.1	$1.500 \cdot 10^{-10}$	0.0178	$1.513 \cdot 10^{-10}$	$4.275 \cdot 10^{-10}$	$3.675 \cdot 10^{-10}$
1.1V, svt	100	0.1	$1.497 \cdot 10^{-10}$	0.0177	$1.513 \cdot 10^{-10}$	$3.467 \cdot 10^{-10}$	$2.797 \cdot 10^{-10}$
1.8V, svt	1	0.3	$1.686 \cdot 10^{-10}$	0.0119	$1.165 \cdot 10^{-10}$	$3.794 \cdot 10^{-10}$	$3.352 \cdot 10^{-10}$
1.8V, svt	10	0.3	$1.686 \cdot 10^{-10}$	0.0120	$1.165 \cdot 10^{-10}$	$3.519 \cdot 10^{-10}$	$3.114 \cdot 10^{-10}$
1.8V, svt	100	0.3	$1.682 \cdot 10^{-10}$	0.0119	$1.165 \cdot 10^{-10}$	$2.356 \cdot 10^{-10}$	$1.788 \cdot 10^{-10}$
1.1V, lvt	1	0.1	$1.553 \cdot 10^{-10}$	0.0169	$1.587 \cdot 10^{-10}$	$3.705 \cdot 10^{-10}$	$3.299 \cdot 10^{-10}$
1.1V, lvt	10	0.1	$1.553 \cdot 10^{-10}$	0.0169	$1.589 \cdot 10^{-10}$	$3.561 \cdot 10^{-10}$	$3.183 \cdot 10^{-10}$
1.1V, lvt	100	0.1	$1.551 \cdot 10^{-10}$	0.0168	$1.593 \cdot 10^{-10}$	$2.768 \cdot 10^{-10}$	$2.272 \cdot 10^{-10}$
1.1V, hvt	1	0.1	$1.406 \cdot 10^{-10}$	0.0169	$1.417 \cdot 10^{-10}$	$5.056 \cdot 10^{-10}$	$4.309 \cdot 10^{-10}$
1.1V, hvt	10	0.1	$1.411 \cdot 10^{-10}$	0.0169	$1.418 \cdot 10^{-10}$	$4.950 \cdot 10^{-10}$	$4.256 \cdot 10^{-10}$
1.1V, hvt	100	0.1	$1.409 \cdot 10^{-10}$	0.0168	$1.416 \cdot 10^{-10}$	$4.410 \cdot 10^{-10}$	$3.398 \cdot 10^{-10}$
1.1V, nat	3	0.3	$3.717 \cdot 10^{-10}$	0.0166	$3.822 \cdot 10^{-10}$	$3.238 \cdot 10^{-10}$	$3.082 \cdot 10^{-10}$
1.1V, nat	30	0.3	$3.778 \cdot 10^{-10}$	0.0165	$3.879 \cdot 10^{-10}$	$2.503 \cdot 10^{-10}$	$2.367 \cdot 10^{-10}$
1.1V, nat	300	0.3	$4.092 \cdot 10^{-10}$	0.0159	$4.045 \cdot 10^{-10}$	$2.233 \cdot 10^{-10}$	$2.074 \cdot 10^{-10}$
1.8V, nat	10	1	$1.587 \cdot 10^{-10}$	0.0088	$7.856 \cdot 10^{-10}$	$3.499 \cdot 10^{-10}$	$3.377 \cdot 10^{-10}$
1.8V, nat	100	1	$1.643 \cdot 10^{-10}$	0.0087	$8.030 \cdot 10^{-10}$	$1.745 \cdot 10^{-10}$	$1.222 \cdot 10^{-10}$
1.8V, nat	1000	1	$1.619 \cdot 10^{-10}$	0.0079	$8.376 \cdot 10^{-10}$	$1.350 \cdot 10^{-10}$	$1.265 \cdot 10^{-10}$

Table 3.3: Capacitance parameter summary for the NMOS-transistors (svt = standard-V_T, lvt = low-V_T, hvt = high-V_T, width and length in μm).

Type	W	L	$C_{gso}(F/m)$	$C_{ox}(F/m^2)$	$C_{gdo}(F/m)$	$C_{js,avg}(F/m)$	$C_{db,tot,avg}(F/m)$
1.1V, svt	1	0.1	$1.963 \cdot 10^{-10}$	0.0161	$1.951 \cdot 10^{-10}$	$4.749 \cdot 10^{-10}$	$4.011 \cdot 10^{-10}$
1.1V, svt	10	0.1	$1.961 \cdot 10^{-10}$	0.0161	$1.951 \cdot 10^{-10}$	$4.591 \cdot 10^{-10}$	$3.920 \cdot 10^{-10}$
1.1V, svt	100	0.1	$1.961 \cdot 10^{-10}$	0.0161	$1.951 \cdot 10^{-10}$	$3.609 \cdot 10^{-10}$	$2.791 \cdot 10^{-10}$
1.8V, svt	1	0.3	$2.168 \cdot 10^{-10}$	0.0148	$2.146 \cdot 10^{-10}$	$5.209 \cdot 10^{-10}$	$4.304 \cdot 10^{-10}$
1.8V, svt	10	0.3	$2.169 \cdot 10^{-10}$	0.0148	$2.145 \cdot 10^{-10}$	$4.913 \cdot 10^{-10}$	$4.094 \cdot 10^{-10}$
1.8V, svt	100	0.3	$2.169 \cdot 10^{-10}$	0.0148	$2.144 \cdot 10^{-10}$	$3.635 \cdot 10^{-10}$	$2.665 \cdot 10^{-10}$
1.1V, lvt	1	0.1	$1.884 \cdot 10^{-10}$	0.0162	$1.896 \cdot 10^{-10}$	$4.119 \cdot 10^{-10}$	$3.581 \cdot 10^{-10}$
1.1V, lvt	10	0.1	$1.883 \cdot 10^{-10}$	0.0162	$1.896 \cdot 10^{-10}$	$3.941 \cdot 10^{-10}$	$3.443 \cdot 10^{-10}$
1.1V, lvt	100	0.1	$1.883 \cdot 10^{-10}$	0.0162	$1.896 \cdot 10^{-10}$	$2.967 \cdot 10^{-10}$	$2.301 \cdot 10^{-10}$
1.1V, hvt	1	0.1	$1.682 \cdot 10^{-10}$	0.0149	$1.668 \cdot 10^{-10}$	$3.981 \cdot 10^{-10}$	$3.145 \cdot 10^{-10}$
1.1V, hvt	10	0.1	$1.680 \cdot 10^{-10}$	0.0149	$1.668 \cdot 10^{-10}$	$3.853 \cdot 10^{-10}$	$3.087 \cdot 10^{-10}$
1.1V, hvt	100	0.1	$1.679 \cdot 10^{-10}$	0.0149	$1.668 \cdot 10^{-10}$	$3.382 \cdot 10^{-10}$	$2.559 \cdot 10^{-10}$

Table 3.4: Capacitance parameter summary for the PMOS-transistors (svt = standard-V_T, lvt = low-V_T, hvt = high-V_T, width and length in μm).

In tables 3.3 and 3.4, we consistently see that the $C_{sb,tot,avg}$ and $C_{db,tot,avg}$ are lower for the highest $\frac{W}{L}$-transistors, because these have 10 fingers, while the others only have one. For the 1.8 native NMOS, the number of fingers was chosen equal to the $\frac{W}{L}$-ratio, to test the dependence of the capacitance on the number of fingers. When the number of fingers is large enough, the area of source and drain is approximately halved, so we expect roughly half of the $C_{db,tot,avg}$ that was found for only one finger. The other parameters are consistent and independent of $\frac{W}{L}$ (when rounded). Finally, we check for the case of the 1.1 V standard-V_{th} NMOS of figures 3.8 and 3.9:

$$C_{gs} \approx C_{gso} \cdot 600\mu m + \frac{2}{3} C_{ox} \cdot 600\mu m \cdot 45nm \approx 408fF \tag{3.18}$$

This approximately agrees with the measured $375fF$ at 1.1 V. Because $C_{gg} = C_{gs} + C_{gd} + C_{gb}$, we expect $C_{gg} \approx 408fF + C_{gdo}W \approx 499fF$, which is quite close to the measured 472 fF. The transistor has only one finger, so we expect $C_{jd} \approx C_{db,tot,avg}W \approx 225fF$, which is in between the measured C_{jd} and $C_{dd,TOT}$. The obtained $C_{db,tot,avg}$ is larger than expected because the capacitances are extracted from the same current simulation as the other parameters. The current was swept from a very low value on, where $V_{DS} = V_{GS}$ is very low, and C_{jd} decreases with increasing drain voltage. It was decided not to set up a separate simulation: C_{jd} is not estimated correctly, but we are interested in the total parasitic capacitance at the drain, $C_{dd,TOT}$. It was checked that the values predicted by $C_{db,tot,avg}W$ are close enough (e.g. $12,5\%$ error in this case, at 1.1 V) to the measured $C_{dd,TOT} = C_{dd} + C_{jd}$ for a first design step, and $C_{dd,TOT}$ is also voltage-dependent. We also calculate a corrected $C_{db,tot,avg} = \frac{C_{dd,TOT} - C_{gd}}{W} = \frac{257fF - 87fF}{W} = 2.833 \cdot 10^{-10}F/m$. When the exact value of C_{dd} was needed, it was simulated first. The total transistor output impedance is very important for the termination of the harmonics, and was checked with an S-parameter simulation at a certain bias setting first (cfr. section 6.6.).

We conclude that for the PA design, we will not use the native transistors because they have almost zero threshold voltage, which is not convenient for a limited-conduction angle PA: the gate would have to be driven very low to turn the transistor off. Their minimal size is larger, so they will have a higher C_{gs}. PMOS-transistors are not used either: they have more than twice as much parasitics for the same R_{on} or I_{DS} when compared to an NMOS (cfr. the ratio of the I_0's).

3.6   Breakdown-voltages

The 45-nm GPDK is a fictional CMOS process, provided by Cadence for demonstration and research purposes. The breakdown voltages specified in the model documents are not realistic, e.g. several parameters are left at the default values of the BSIM4-model. For example: the drain diode breakdown voltage, represented by the "bvd"-parameter, is left at 10 V for all transistor types, which is not realistic for either the 1.1V or 1.8V transistors. The "wBvds"-parameter specified in the BSIM4-model in ADS represents the drain-source breakdown voltage warning threshold, but is not specified in the GPDK045-models. This question was asked to Cadence support, and it was confirmed that these parameters are not set to a more realistic value because the process is fictional. The nominal supply voltage that is specified for a transistor type is a typical voltage for a transistor of similar geometry (e.g. in other processes), but these might deviate in reality since e.g. the implants or the gate-oxide thickness might be adjusted to make the transistors more robust.

In a power amplifier, the output stage transistors are most often the largest transistors, making them the most expensive in terms of chip area and input power that has to be spent to drive them. To maximize the output power, efficiency, gain, PAE... a relatively large supply voltage and maximal voltage swing over the output transistors is desired. The transistors should be driven to the maximal possible voltage stress while still guaranteeing some long-term reliability.

The typical values used as breakdown-voltages are $V_{gd,max} < 2V_{DD,nominal}$ and $V_{ds,max} < (2...3)V_{DD,nominal}$. These values were also found in [24] and seem to be commonly used values. Because of the lack of correct breakdown-specifications in the GPDK-documentation and because no other documentation with specific values was found, we chose to keep the V_{gd} and V_{ds} to these limits as well. Power dissipation limitations are also not specified, and were not used: to obtain an efficient PA, we have to minimize the I_{DS} - V_{DS}-overlap in time, so $P_{dissipated}$ has to be small. Furthermore, the $V_{gd,max} < 2V_{DD,nominal}$-requirement is already quite restrictive: the maximal drain voltage peak occurs when the transistor is switched off and V_{GS} is sufficiently below V_{th} (e.g. $\approx < 0.4V$). In this worst-case situation, the $V_{ds,max} < (2...3)V_{DD,nominal}$ and $V_{gd,max} < 2V_{DD,nominal}$-requirements apply simultaneously, so $V_{gd,max} < 2V_{DD,nominal}$ is the most restrictive.

Part III

Passive components

Chapter 4

Passive Components

4.1 Introduction

To make an amplifier, both active and passive components are needed. In this chapter, we will elaborate on the passive components which are available in CMOS design. For the different types of components there will be several options, each one with their own advantages and disadvantages [25][26].

The fact that the components are used at high frequencies will impose new design challenges but it will also provide new possibilities (e.g. the ability to use on-chip components). Due to the high frequencies, even small parasitics might no longer be negligible. Hence for high frequency design, the minimization of parasitics becomes one of the most dominant design criteria. This will eliminate the possibility to use off-chip components as they would introduce excessive parasitics which would alter the behaviour of the circuit.

However, this is not the only advantage of on-chip components. When on-chip components are used it will be easier to make a good differential circuit, although matched off-chip components are also commercially available. While designing matched values will be easier, on-chip components suffer from high tolerances and thus the effective component value might differ significantly from the desired value. Another disadvantage of using on-chip components is the fact that the Q-factor of the on-chip inductors is rather limited.

4.2 Substrate

To design discrete components, one can use the spice models from the gpdk library. However this gpdk library is incomplete (e.g. it lacks models for transformers) and thus a first step in the design of the passive components consisted of determining the layer stack of the given technology. Most of the parameters of the stack were already documented but the substrate parameters and loss tangent were not available in the gpdk documentation and thus it was necessary to extract them by fitting the spice model to Advanced Design System (ADS) simulations. This fitting was done by using an inductor and resulted in a loss tangent of 0.1 for the different dielectric layers and a substrate of 100 μm thickness with a resistivity of 0.01 Ω-cm.

When looking at the extracted substrate parameters, it is clear that the substrate resistivity is rather low. A substrate with a low ρ_{sub} has plenty of advantages but is to be avoided when one wants to build an efficient RF power amplifier. Hence we have opted to alter the substrate

resistivity in our design to 10 Ω-cm [24], which is a value that is often documented in literature [27] concerning sub-micron RF devices (while the original, heavily doped substrate is used more often in case of digital circuitry). The design process for the passive components will be approximately the same when a lower value for the substrate resistivity is chosen but will lead to inferior results.

To demonstrate why we have chosen a lightly doped substrate (i.e. high ρ_{sub}) over a heavily doped substrate (i.e. low ρ_{sub}) the advantages and disadvantages of both will be given. Heavily doped substrate is mostly used when digital devices need to be designed. This is due to the fact that this type of substrate protects the circuit from latch-up and hot electrons. Another important reason why heavily doped substrates are used is because this kind of substrate behaves as a single-point ground reference which leads to a reduction in common impedance coupling and ground bounce. This is essential because digital devices generate plenty of switching noise which might couple to the (sensitive) analog devices when no single-point ground reference is present. On the other hand, when RF analog circuits (GHz range and up) need to be designed, a higher ρ_{sub} is to be preferred. This is due to the fact that a heavily doped substrate will strongly deteriorate the RF performance and electrical isolation of the analog components.

Since the goal of this thesis is to design an efficient RF amplifier, we come to the conclusion that using a high ρ_{sub} will lead to a better device.

4.3 Resistors

Because resistors introduce losses (which decreases the efficiency and may also lead to heat evacuation problems), an IC designer should try to minimize the number of resistors. Nevertheless, it is still interesting to look at how they can be designed, as resistors aren't always avoidable. One way to get a resistor is by using very thin metal films.
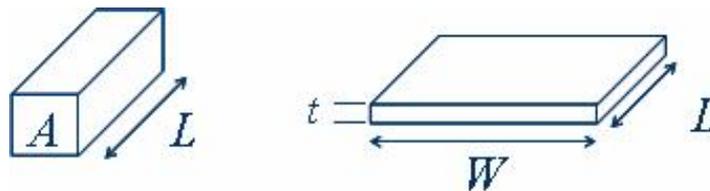
Figure 4.1: Metal film

The resistance obtained by using metal films can be calculated by using Pouillet's law (equation (4.1), where ρ and σ respectively denote the resistivity and conductivity of the metal film).

$$R = \rho \times \frac{L}{A} = \frac{1}{\sigma} \times \frac{L}{A} \tag{4.1}$$

Unfortunately, ultra thin metal layers are not available in gpdk45. Thus this type of resistor can't be considered in the design of our Power Amplifier (PA).

Before looking at the options available in the given technology, it is important to first define the concept of sheet resistance. The sheet resistance R_{sh} of a certain layer, expressed in ohms per square, is the resistance of a square structure implemented in that layer. By rewriting Pouillet's

law (4.1), it becomes obvious why this definition is useful.

$$
\begin{aligned}
R &= \rho \times \frac{L}{A} \\
&= \rho \times \frac{L}{Wt} \\
&= \frac{\rho}{t} \times \frac{L}{W} \\
&= R_{sh} \times \frac{L}{W}
\end{aligned}
\tag{4.2}
$$

Essentially this formula consists of two terms. The first term of this equation is technology dependent and can not be changed by the Integrated Circuits (IC) designer (only by the process engineer). Hence, this ratio is a technology dependent parameter which can be given for every layer of interest. It can be seen that this first term is the sheet resistance as exactly this fraction is obtained when one computes the resistance of a square resistor. For the different types of resistors which will be given in the course of this section, the sheet resistance will be mentioned as it is an important parameter. Small sheet resistances will lead to big resistors, especially when high resistor values are needed. This is to be avoided and thus one has to take into account the R_{sh} when choosing how to implement a given resistor. A last remark concerning this concept is about how those unit squares are to be chosen. It is advantageous to choose the unit squares as small as possible to minimize the surface area of the resistor. But other phenomena (e.g. electromigration, ...) might also play a role in the dimensioning and might require a larger unit square.
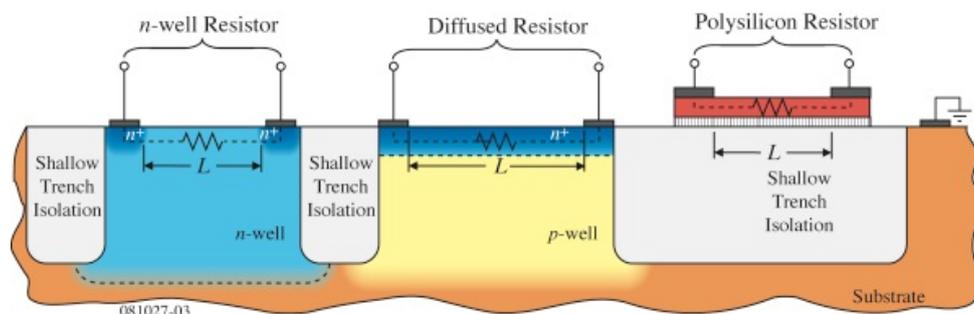
Figure 4.2: Resistor types available in GPDK45

As mentioned before, thin film metals are not available in the gpdk45 substrate stack. But there are still plenty of other possibilities (figure 4.2 [28]). First of all, a resistor can be realised by using non-salicided polysilicon. This layer has a moderate sheet resistance of approximately 500 Ω per square (while salicided polysilicon is low-ohmic and behaves more like a metal). A non-salicided diffused resistor is a variant with a slightly smaller R_{sh} of approximately 200 Ω per square. The word diffused means that the resistor is made in the same way as drain and source islands are made in a Metal-Oxide-Semiconductor (MOS) transistor. There are 2 reasons why this variant is less preferable, namely the fact that the poly resistor will be smaller (when sufficiently high resistance values are required) and that the diffused resistor will be slightly less linear.

To make high resistance values, a p-well resistor can be considered since the p-well has the highest sheet resistance for the gpdk45 process (i.e. approximately 1 kΩ per square). But there are some good reasons to avoid using a p-well despite the fact that resistors can be made small

by using this layer. One of the main problems from which they suffer is that the shielding to
the substrate and to neighbouring components is worse than for the alternatives. This means
that p-well resistors will be quite noisy since all disturbances and noise from the substrate can
be coupled directly onto the resistors. A second very important reason to avoid this kind of
resistors is that the temperature coefficient and the voltage modulation coefficient (i.e. the
dependence of the resistor value on the voltage applied over the terminals of the resistor) are
rather high as compared to the alternatives. As we work at high frequencies, there will also be
a third major problem when using this type of resistors. Due to higher parasitic capacitances,
this variant might be less suitable at high frequencies. Because of all these disadvantages, it is
reasonable to assume that for small up to moderate resistor values, unsalicided poly will be the
best option.

4.4   Capacitors

4.4.1   Applications

Capacitors are essential building blocks in IC design. They can be used to isolate stages for
biasing purposes and as decoupling capacitors. These types of capacitors are typically rather
large, but small capacitances are also needed. Those small-valued capacitors can for example
be used as part of resonating LC tanks, LC baluns and matching circuits

It is important to mention that for high frequencies (e.g. 15 GHz) the typically needed ca-
pacitance values will be in the femtofarad range. This indicates why off-chip components are
inadequate.

4.4.2   Types of capacitors

Essentially, there are 3 categories of capacitors. First of all, one can make a (parallel plate)
capacitor by taking 2 conductive layers and putting a dielectric in between them. The next
category consists of capacitors based on MOS transistors and the last kind uses the depletion
layer of a pn junction to realise a capacitor.

MIM/MOM capacitors

The easiest way to make a capacitor is by putting a dielectric in between 2 conductive layers
(i.e. the electrodes). This is called a Metal-Insulator-Metal (MIM) capacitor.

In this paragraph, two versions will be compared. First of all the dedicated MIM capacitor
consisting of a metal layer, silicon dioxide and an inter-metal conductor. The reason why a
dedicated inter-metal layer is used, is because the capacitance density is in a first approximation
inversely proportional to the distance between the 2 electrodes. Secondly, a variant consisting
of a stack of several metal layers (i.e. a MOM capacitor) will be considered (figure 4.3).

Figure 4.3: MOM capacitor

When comparing both variants, one can see that the dedicated capacitor is indeed a lot better than the stacked variant. Simulation shows that in both cases the Self Resonant Frequency (SRF) will be very large ($SRF_{MIM} = 195$ GHz and $SRF_{MOM} = 224$ GHz), so both types are usable at a high frequency. But when comparing the Q-factor, simulation points out that the dedicated variant resembles much more an ideal capacitor (e.g. when designing a $100\,\mu\text{m}^2$ capacitor, the Q-factor of the dedicated capacitor will be approximately 400 while only 8.25 is achieved in the case of the metal stack). This can be caused by the fact that the lower metal layers of the MOM capacitor are too close to the substrate, resulting in a high portion of leakage at the operating frequency.

In case of a capacitor, the Q-factor is defined as the impedance of the capacitor at the given frequency divided by the equivalent series resistance (formula 4.3).

$$Q = \frac{\frac{1}{\omega C}}{R} = \frac{1}{\omega C R} \tag{4.3}$$

A last point of difference between the variants is the capacitance density. When looking at a $100\,\mu\text{m}^2$ capacitor, simulation shows another advantage of the dedicated capacitor ($122\,\text{fF}$ as compared to $62.5\,\text{fF}$ in case of the stacked version).

As a consequence, the MIM version provided by GPDK will be our first choice to realise a capacitor. Earlier in this section, it was mentioned that this will lead to high Q-factors for the capacitors. Hence, when designing a certain part of the amplifier (e.g. an LC tank), capacitors can initially be approximated as being ideal.

The only problem concerning MIM capacitors is the limited capacitance density. This can be solved by making a different type of capacitor. But while these other capacitors might have a larger capacitance density, they will be modulated by the voltage applied over its terminals. An alternative to make a higher capacitance without suffering from this voltage dependence, is to make comb capacitors. This would however make the component value much more sensitive to process variations.

A comb capacitor was designed on the uppermost metal layer, to give an indication on the capacitance density of such a capacitor. This resulted in a capacitance density of 24.2 fF per $100\ \mu m^2$ and a Q-factor of 8.43. Consequently, this capacitance density is still lower than the value for the MOM and MIM capacitors but can be increased by implementing this comb capacitor on multiple metal layers.

To find out the capacitance between the terminals of the dedicated MIM capacitor, one can use a simple formula (4.4). This equation consists of 2 terms, first the capacitance due to the field lines in an ideal parallel plate capacitor. Secondly a contribution can be seen concerning fringe field. While the first term is proportional to the surface area (denoted by S) of the electrodes, the second term will be proportional to their perimeter (denoted by P).

$$C[fF] = S[\mu m^2] \times 1.025 \ \frac{fF}{\mu m^2} + P[\mu m] \times 0.2425 \ \frac{fF}{\mu m} \qquad (4.4)$$

MOS capacitors

Since the capacitance density of MIM capacitors is rather low, those capacitors will be big when used to design a large capacitance value. This can be a problem when trying to design decoupling capacitors. MOS capacitors can be used to make these high capacitance values without sacrificing too much space. These capacitors are essentially MOS transistors where the drain and source are shorted and form the connection to the channel (lower electrode) while the gate acts as the upper electrode. The fact that this offers the possibility to get much higher capacitance values when using the same surface area comes at a cost. The capacitance value is fairly voltage dependent (as can be seen in figure 4.4 for a MOS transistor with a surface area of $100 \, \mu m^2$), particularly when the MOS transistor is in weak inversion or depletion. Hence, this type of capacitors is only useful when designing a large valued capacitance of which the exact value isn't important but for which we know that the applied voltage over the capacitor is large enough and almost constant (e.g. to design a decoupling capacitor). Apart from the dependence on the biasing voltage, there will also be a significant influence of the temperature. When it is assumed that the MOS capacitor is forward biased by e.g. 1 Volt a relative variation of 1.13% of the capacitance value is obtained when changing the environment temperature from 0 to 80 °C.
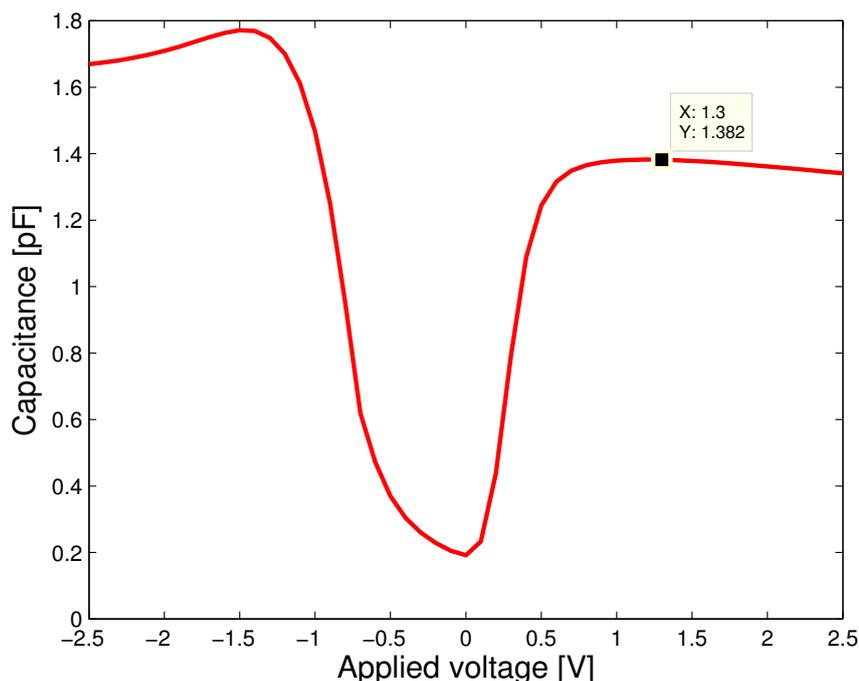
Figure 4.4: Capacitance in function of the voltage over the MOS capacitor

When comparing the capacitance achieved with this MOS capacitor to the value obtained by a MIM capacitor of the same size, it is clear that the MOS capacitor realises densities which are approximately 10 times larger. This is due to the fact that the gate oxide is a very thin dielectric.

Junction capacitors

A last method to design a capacitor is by using the depletion layer of a pn junction. Although the depletion layer behaves as a capacitor (separation of charges), it is not a good way to design a given capacitance value. This can be seen by the fact that the width of the depletion layer will be modulated by the voltage over the pn junction. Hence, the capacitance value realised by the junction will vary significantly when changing the voltage applied over the junction (as can be seen in figure 4.5 for a pn junction with a surface area of $100\,\mu m^2$). This property will limit the use of pn junctions when designing capacitors.
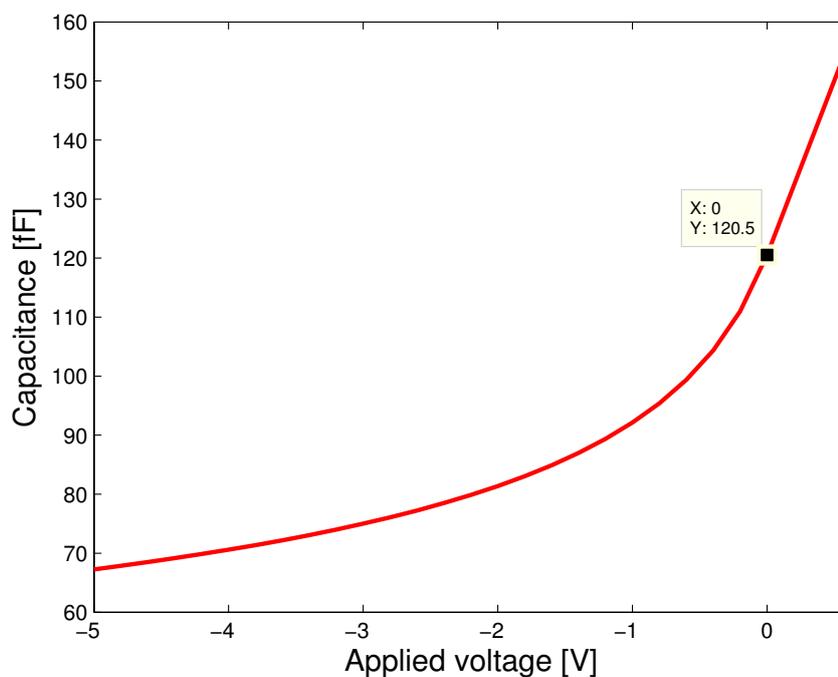
Figure 4.5: Junction capacitance in function of the voltage over the junction

An application where such an effect is desired, is in the case of a varactor. Such a varactor can for example be used to change the oscillation frequency of an LC tank by changing the voltage applied over the tank.

4.5 Inductors

4.5.1 Applications

Previously in this chapter, a short overview of two different passive components (i.e. resistors and capacitors) was given. Another building block that is essential in RF design are inductors.

For the given technology the inductance values that are realisable will be in the picohenry range. Inductors can be used for diverse applications: A first example is an RF choke (for which a large inductance value is needed). This is a component which ideally behaves as a perfect short at DC and as a perfect open for all other frequencies. An inductor can also be used as part of an LC tank, LC balun, etc.

4.5.2 Figures of merit

During the design of an inductor, the design engineer has to be able to compare several versions of an inductor. To find out what version is the best, different figures of merit are computed: the surface area, DC losses, Q-factor, SRF, the dependence of the inductance on the frequency, etc. The 3 last figures of merit will be discussed in the following paragraphs. Earlier in this thesis, a comparison was given between the case of a heavily doped substrate on the one hand and a lightly doped substrate on the other and the difference between the two will also be shown for the different figures of merit.

Q-factor

The principle of the Q-factor will be explained in this section and for this purpose figure 4.6 was added.
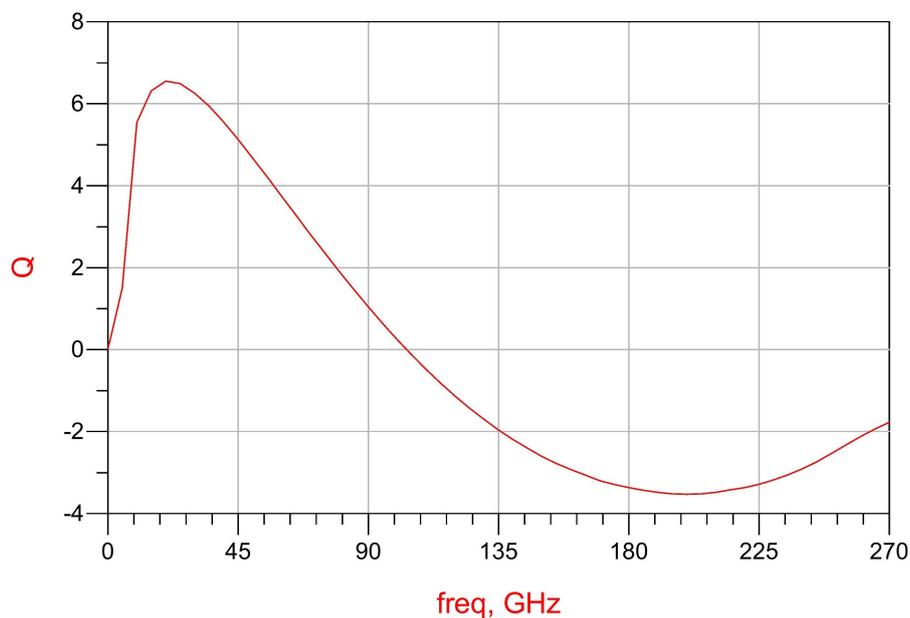
Figure 4.6: Q-factor in the case of a heavily doped substrate

In this graph formula 4.5 was used to compute the quality factor for a given inductor.

$$Q = \frac{\omega L}{R} = \frac{stored\ energy}{dissipated\ energy} \tag{4.5}$$

It is important to mention that this Q-factor is frequency dependent and will be positive for frequencies lower than the SRF (the SRF will be explained in the next section). For a frequency

which is a lot lower than the SRF the maximum Q-factor will be achieved. This maximum Q-factor is called the Peak Quality Factor (PQF) and it is desirable to design the inductor to have its PQF at a frequency which lies close to the operating frequency of the system.

In case of figure 4.6 the PQF is equal to 6.548 while the SRF (i.e. frequency where the Q-factor crosses zero) is approximately 100 GHz. When simulating the same layout for a lightly doped substrate, the PQF approximately doubles (from 6.548 to 15.787). The fact that the Q-factor approximately doubles when increasing the resistivity of the substrate is the reason why a lightly doped substrate was preferred.

Since the goal of this project is to make an efficient amplifier, it is essential to maximize the Q-factor under certain constraints. Hence it is important to know how the Q-factor can be influenced and what mechanisms are the cause of the losses (as losses will lower the Q factor). These loss mechanisms can be subdivided into two groups, caused by resistive effects and capacitive effects.

The first resistive effect that introduces losses (and therefore reduces the Q-factor) are metal losses. This results from the fact that metals have a finite conductivity and thereby dissipate some energy. To minimize these losses it is important to use the metal with the lowest sheet resistance. In the gpdk stack this will correspond to the uppermost metal layer. The DC resistance of a metal track will be given by pouillet's law (equation 4.1) but when the frequency increases, one has to take into account that the effective cross-section of the track will decrease due to skin and proximity effect (i.e. current crowding) and thus that the AC resistance will be somewhat higher.

Skin effect is the tendency for AC currents to flow at the boundaries of the metal track resulting in a decrease of the effective cross section of the conductor. This skin effect can be taken into account (equation 4.6) by using an extra parameter, namely the skin depth δ.

$$R_{AC} = \frac{l}{w\delta\sigma(1 - exp(\frac{-t}{\delta}))} \tag{4.6a}$$

$$\delta = \sqrt{\frac{2}{\omega\mu_0\sigma}} \tag{4.6b}$$

In formula 4.6 w,t and l respectively denote the width, thickness and length of the metal track while the conductivity is denoted by σ. Additional variables used in the formula for the skin depth are ω (radial frequency) and μ_0 (permeability of vacuum).
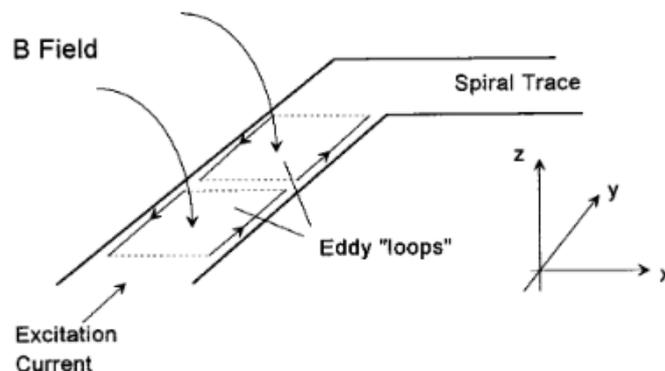
Figure 4.7: Proximity effect: current crowding

While skin effect is also present in the case of a single line, current crowding due to proximity effect is only present when two or more current carrying conductors lie close to each other (figure 4.7 [29]). To understand current crowding, one can look at how a magnetic field is formed by a current carrying inductor. When observing the magnetic field formed by the inductor it becomes obvious that the strongest magnetic fields will be induced in the center of the spiral. Since this magnetic field will induce eddy currents on the metal tracks the original current profile will be altered by the presence of the magnetic field. If special attention is given to the current profile within a single track of the spiral, it can be seen that maximal current densities are to be found at the inner side of the track (i.e. the side closest to the center of the spiral). This effect will be larger for the inner turns than for the outer turns and that is why it might be advantageous to use a tapered layout (which will be discussed later on in this chapter).
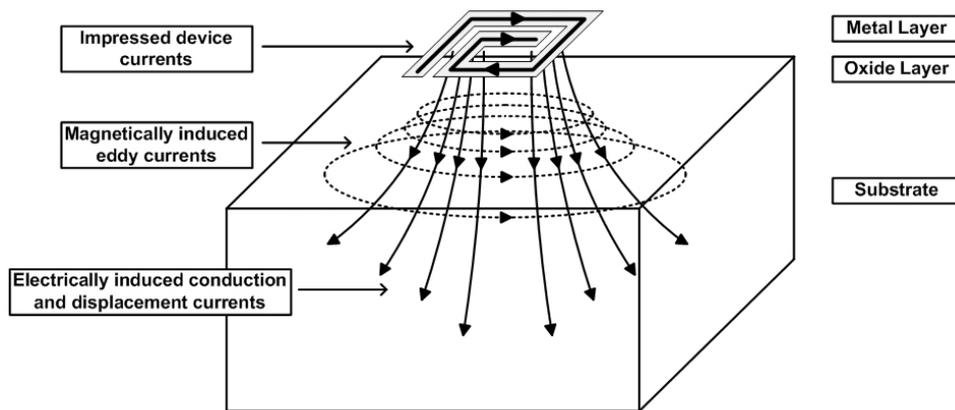
Figure 4.8: Loss mechanisms (electric and magnetic losses)

The second group of resistive losses is due to the presence of the substrate. These losses can be further subdivided into losses originating from the electric coupling and losses originating from the magnetic coupling (figure 4.8 [30]) to the substrate. The latter is due to the fact that varying magnetic fields will introduce eddy currents in neighbouring conductors (e.g. the substrate), thereby countering the desired effect of the inductor by introducing a magnetic field in the opposite direction (as formulated in Lenz law). Since those losses will be highest when the substrate acts like a conductor it becomes clear why the PQF was a lot higher for a lightly doped substrate than for a heavily doped substrate. Apart from the decrease of the Q-factor, an opposing magnetic field will also give rise to a reduction in the apparent inductance value.

Secondly, there will be a contribution from the electric field. The electric fields caused by the inductor give rise to substrate currents and since the substrate has a finite resistivity, ohmic losses will appear due to these substrate currents.

Finally there are some capacitive effects that will also degrade the performance of the spiral inductor. As mentioned in the previous paragraph, the substrate will be conductive and thus this substrate will act as the second plate of a parallel plate capacitor (where the spiral inductor will be the top plate). The problem caused by this parasitic capacitor is that it might resonate with the inductor itself, setting an upper limit on the frequency for which the inductor can be used. A second capacitive effect that decreases the performance of the coil is the inter-winding capacitance between neighboring lines (but this part is often negligible) as well as the capacitance due to the underpass or overpass part(s) of the inductor.

To conclude this section a final remark concerning figure 4.6 has to be made. For frequencies under the PQF frequency, the metal losses will dominate while substrate losses will dominate for

the frequencies above this PQF frequency. Since it is good practice to design an inductor which has its PQF around the operating frequency one can alter the width of the track to change the metal losses and thus shift the frequency at which the PQF is obtained.

Self Resonant Frequency

When designing an inductor, it is important that the device still works at the desired frequency. Hence, one should make sure that the inductor still behaves like an inductor at the operating frequency. To find out for what frequency range the device can be used, the SRF is defined as the frequency at which the device starts behaving capacitively. In contrast to the large variation of the Q-factor between heavily doped and lightly doped substrate, the SRF only shifted slightly when altering the ρ_{sub} in our simulations.

The parasitic effects leading to resonance with the inductor will be the capacitance originating from the parallel plate capacitor between coil and substrate C_{ox} on the one hand and the inter-winding and overlap capacitance C_{inter} on the other (4.7).

$$SRF = \frac{1}{2\pi\sqrt{L_s C_{tot}}} \tag{4.7a}$$

$$C_{tot} = C_{inter} + C_{ox} \tag{4.7b}$$

Inductance value

It is not only important to have an operating frequency which is lower than the SRF to make sure that the device still acts as an inductor, but it is equally important that the Q factor is high enough to make efficient devices. Hence we need an SRF which is sufficiently higher than the operating frequency. But this is not the only reason for taking a sufficiently high SRF. When taking a look at figure 4.9, it is clear that the inductance value only remains constant for frequencies which are sufficiently low since the inductance changes a lot around the SRF.
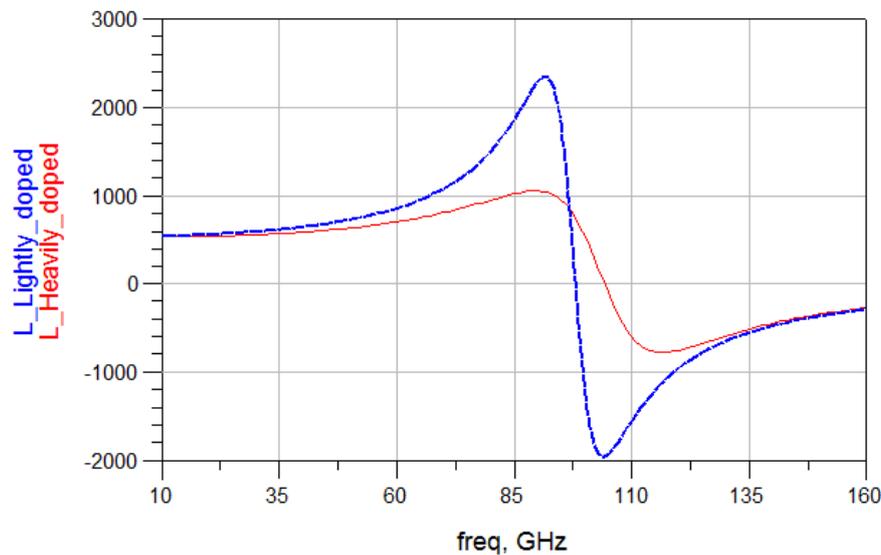
Figure 4.9: Inductance value [pH] in the case of a heavily and a lightly doped substrate

When comparing the simulation results for a heavily doped substrate with the results for a lightly doped substrate, it can be noticed that the inductance is slightly higher in case of a lightly doped (i.e. more resistive) substrate (560.6 pH as compared to 546.67 pH for a heavily doped substrate). The reason for this is that the eddy currents are more suppressed in case of a lightly doped variant, resulting in less influence from the magnetic field induced by the eddy currents.

4.5.3 Topology

Standard design criteria

When trying to design an optimal inductor, one has to take into account the consequences of the design choices on the figures of merit presented in the previous section. It was stated that the metal losses could be minimized by using the metal layer with the lowest sheet resistance. In the given technology this corresponds to the uppermost metal layer which gives rise to a second advantage. When the distance between the spiral and the substrate is maximized, the C_{ox} will be minimized since the capacitance is approximately inversely proportional to this distance. Hence the Q-factor as well as the SRF can be optimized by using the uppermost metal layer.

Another important aspect to take into account when designing an inductor is that the space between the inductor and the substrate should be left blank. Inserting components in between can deteriorate the performance of the system (due to interference and eddy currents introduced by the magnetic field of the coil). An interesting exception to this rule are ground shields (which will be discussed later on in this chapter) since they try to exclude the substrate losses from the system by introducing a more ideal ground termination for the electric fields of the spiral inductor.

Finally, some remarks will be made on the different dimension parameters. An important parameter will be the spacing between adjacent lines. This value should be minimized to maximize

the magnetic coupling between neighboring lines unless the component needs a very high SRF. While increasing the spacing might raise the SRF, it is important not to exaggerate while doing this. The reason for this is that the effect on the SRF starts to become negligible while the Q-factor keeps on decreasing when increasing the spacing. Another parameter that is important for the design of an inductor is the track width. When tracks are too narrow the metal losses will be quite high and when the tracks are too wide the SRF will decrease significantly and current crowding effects will become more important. Hence the upper limit of the width will be controlled by skin effect and the SRF while the lower limit depends on the maximal current density (based on electromigration).

The dependence of the upper limit on the skin effect can be explained as follows. Normally, an increase in the track width will result in a larger effective cross-section which will give rise to a lower series resistance. On the other hand, the losses to the substrate will increase when making the tracks wider. Since the inductors need to work at a high frequency (i.e. 15 GHz) skin effect will be important and an increase in the thickness of the width might not result in a reasonable increase in the effective cross-section. When this phenomenon is combined with the increase in substrate losses, a decrease of the Q-factor might appear when making the tracks wider. Whether the Q-factor will decrease or increase when widening the tracks will thus depend on the starting track width.

Shape of the spiral

First of all, it is important to mention why spirals should be used instead of regular tracks. The main reason for this choice is that spiral inductors not only get their inductance value from self-inductance of the tracks but also from mutual coupling between the different parts of the spiral resulting in a higher inductance density. Since gpdk45 only allows angles of 45 and 90 degrees, the ways to build this spiral are rather limited. Subsequently, the only options to consider for this technology are an octagon and a square. While the square is easier to design, the octagon gives rise to better performances. Firstly, the inductance density of an octagonal inductor will be higher than for the square variant resulting in a smaller chip area. Secondly, since the angles are more rounded in case of an octagon the corners will have less influence (with regard to current crowding and reflection) and will suffer less from electromigration [31] (leading to an increase in lifetime). As a consequence of the reduction in current crowding effects, the Q-factor will be higher leading to a more efficient amplifier. Because of these reasons we have opted to use only octagonal inductors in the design of the amplifier.
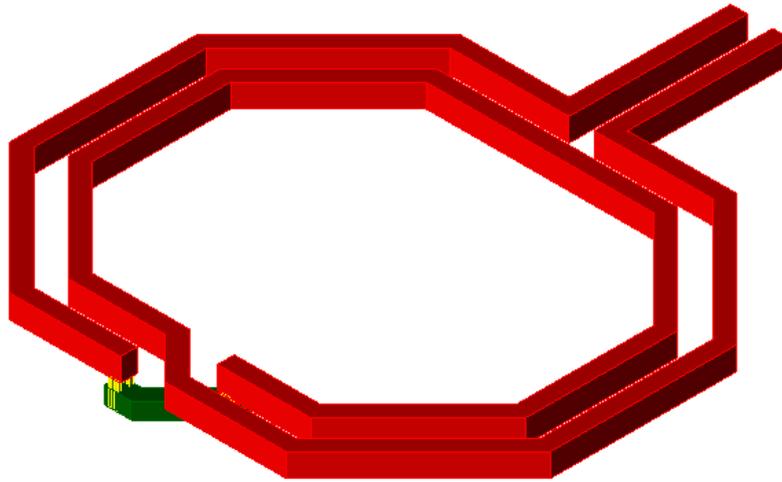
Figure 4.10: Octagonal spiral inductor

Apart from the previous considerations, one also has to choose between asymmetrical and symmetrical spirals. In case of a differential circuit, the latter will of course be optimal with regard to performances as well as the surface area required for the design. The reason why symmetrical inductors will give rise to better performances is because the geometrical, electric and magnetic center of the inductor will coincide, resulting in a higher mutual inductance and hence a higher total inductance. When designing the output and input balun of the outphasing structure it is even essential that symmetrical inductors are used since both branches need to undergo the same load. The biggest disadvantage of a symmetrical circuit is the reduction in SRF caused by an increase in ac potential difference between neighbouring turns of the symmetrical spiral inductor as compared to the asymmetrical variant [32].

Tapering

During the explanation of the proximity effect, it was stated that the strongest magnetic field can be found in the center of the spiral. This has as a consequence that the proximity effect is strongest in the inner turns of the coil. Hence it is advantageous to make these inner turns (slightly) smaller than the outer turns. Due to current crowding, the highest current densities [33] can be found at the outer parts of the tracks and this will be especially the case for the inner tracks (due to the fact that eddy currents will be largest in these turns). Hence the inner tracks can be narrowed to decrease the substrate coupling (which will also result in a higher SRF) without significantly increasing the metal losses of the tracks since approximately no current flows through the center of these inner tracks.
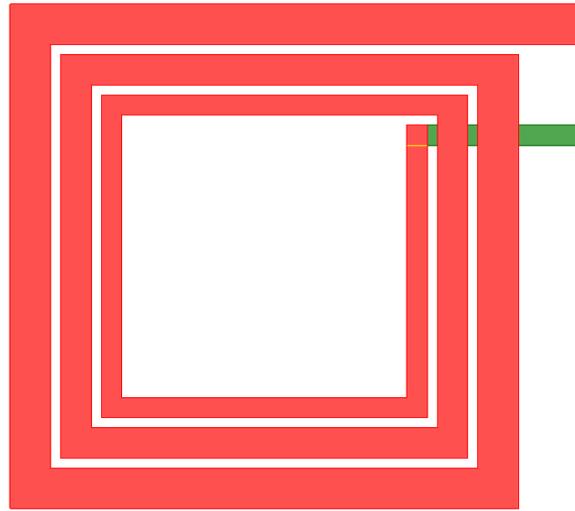
Figure 4.11: Tapered inductor

Series connected stack inductor

When combining two (or more) spiral inductors in series (figure 4.12) some interesting properties can be obtained. The inductance value will approximately scale with $(number\ of\ layers)^2$ while the series resistance only scales linearly with the amount of layers leading to an increase of the Q-factor when adding more spirals in series. This is caused by mutual coupling between the series stacked spirals. Another reason to use series connected spirals is because of the fact that this is an effective solution to make the inductor more compact. On the other hand there will be a major increase in parasitic capacitance (particularly in interwinding capacitance), resulting in a rather strong reduction of SRF. This will oppose the increase of the Q-factor and might even result in a lower Q-factor than the one that is obtained without a series connection. Hence this topology should only be used at sufficiently low frequencies.
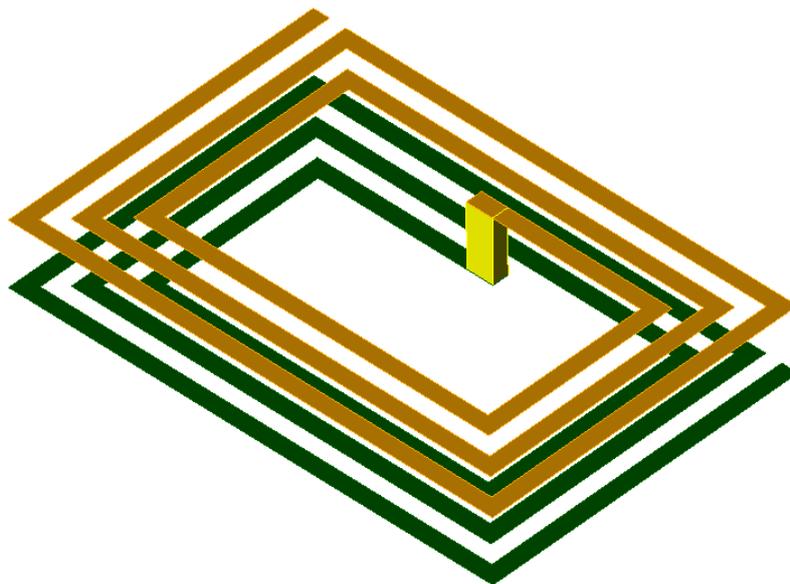
Figure 4.12: Series connected stack inductor

Shunt connected stacked inductor

Another variant on the standard spiral is obtained when one combines two (or more) spirals in shunt (figure 4.13). The advantage of such a construction is the reduction in metal losses, possibly leading to an increase of the Q-factor. The biggest disadvantage of this variant is again the reduction in SRF, although less severe than for series stacked inductors.
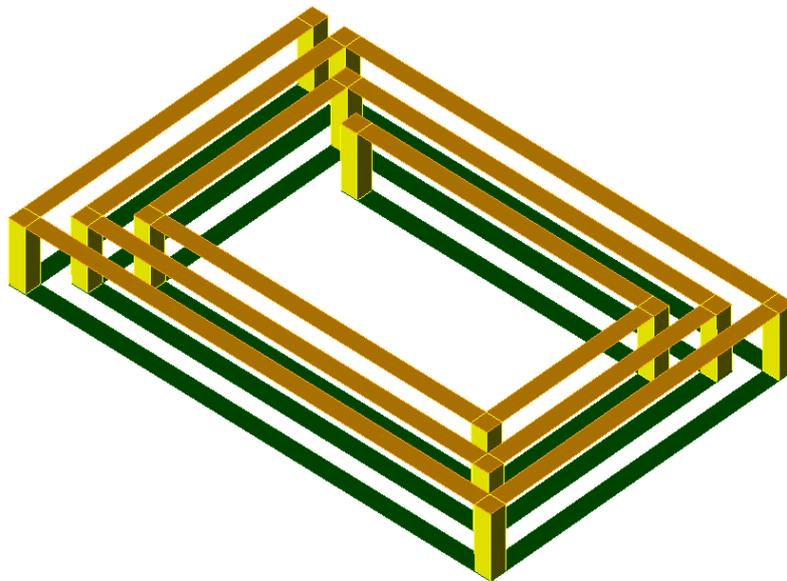
Figure 4.13: Shunt connected stack inductor

4.5.4 Equivalent model

The modelling of inductors will be rather complex as opposed to for example resistors and parallel plate capacitors. In literature one can find plenty of equivalent circuits that try to model these inductors but are only valid for certain topologies. Some of them are empirically determined while others originate from mathematical models. Additionally, there is a difference in the frequency range over which the models can be used (narrowband and broadband models are available).

In the case of gpdk45, the equivalent model given in figure 4.14 is used. This model consists of several components given in the form of empirically determined formulas and is only valid for frequencies well below the SRF.
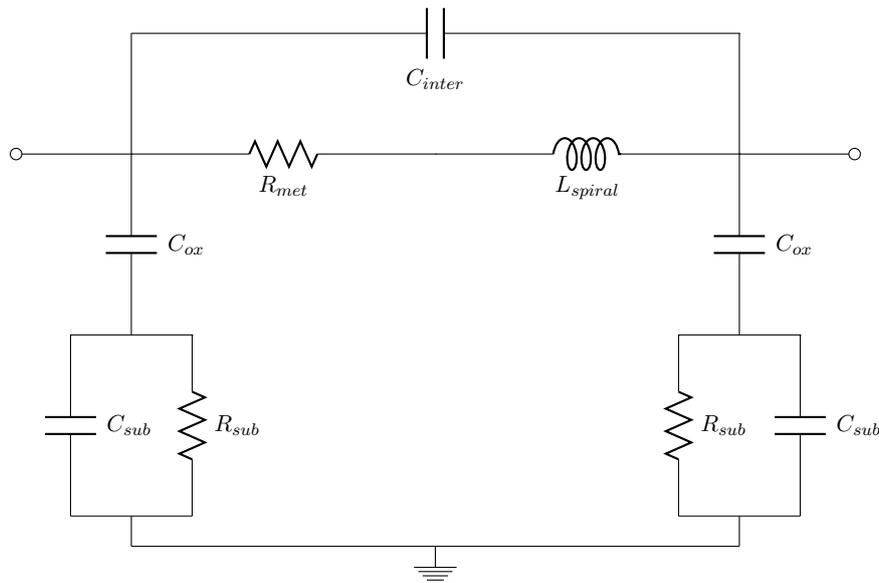
Figure 4.14: Equivalent model of a spiral inductor

Such a model might facilitate the design process because initial guesses for the dimensions can be derived from the model but for a final design, full wave simulations are still obliged. Another reason why this is a useful tool is because one can easily derive the influence of increasing or decreasing a certain parameter in the layout. The effect of the different parameters on the Q-factor, the SRF and the inductance value are respectively given in figures 4.15, 4.16 and 4.17 (for which r, nr, w and s respectively denote the inner radius, number of turns, track width and spacing between adjacent turns). For these curves, default values of w = 4 μm, s = 1.5 μm, r = 15 μm and nr = 2 are used for the parameters that are kept constant.

A quick remark has to be made concerning the fact that these results where derived from the model with a heavily doped substrate. When one would make the same graphs for a lightly doped substrate, one would approximately find the same results with the exception of the fact that the Q-factor and the SRF will be higher. The plots for the Q-factor and inductance value are given at 15 GHz, which is the operating frequency. When those values are negative, the SRF will be smaller than 15 GHz, rendering the inductor useless.

Some interesting remarks can be made by observing the different plots. First of all, it becomes clear that increasing the inner radius or the number of turns to increase L (while keeping the SRF sufficiently above 15 GHz) will give rise to a reduction of the Q-factor. Hence it is important to choose the number of turns and the inner radius as small as possible for a given inductance value. Unfortunately, these two requirements are contradictory and thus it will be important not to exaggerate with the number of turns nor the inner radius. Secondly, the SRF will decrease when the track width increases (due to higher C_{ox}) but the Q-factor will initially rise when increasing the track width. This is because metal losses are dominant for narrow tracks while substrate losses only become dominant when wide tracks are chosen (and for those wide tracks it can be seen that the Q-factor will decrease with increasing track width). When looking at the effect of the track width on the inductance value, one can notice that the influence of the track width is limited. The track width is thus essentially determined by looking at the trade-off between metal losses and substrate losses on the one hand and the SRF on the other. Another parameter that only has minor influence on the inductance value is the spacing. Since the spacing also has little impact on the SRF, the choice of the optimal spacing will mostly depend on the Q-factor. When observing the plot that depicts the relation between the spacing and Q-factor, it become

obvious that a small spacing will be a good choice for an initial design. This originates from the fact that the inductance is a bit higher when the spacing is decreased (due to an increase of the magnetic coupling between adjacent tracks).

Lastly, one would expect the inductance to rise when the inner radius r is increased. However by observing figure 4.17, one can see that this is no longer valid when the radius is increased above a certain value. This is caused by the fact that for these values of r the application frequency is too close to the SRF (figure 4.16).
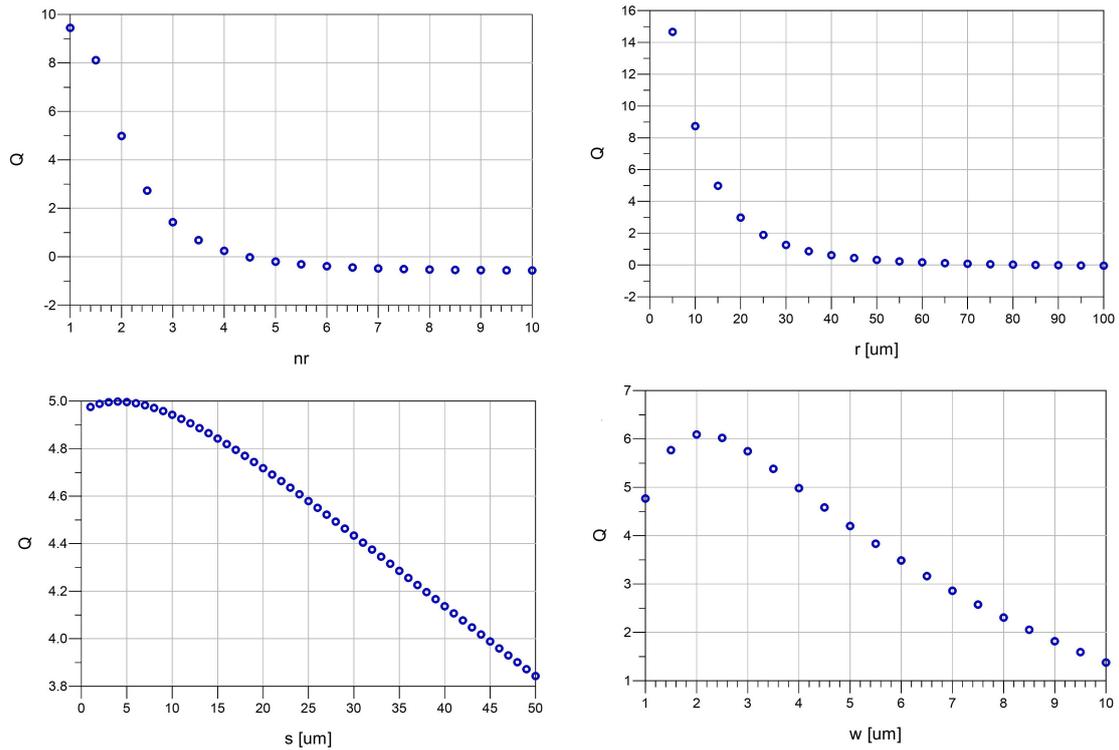
Figure 4.15: Influence on the Q-factor (at 15 GHz)
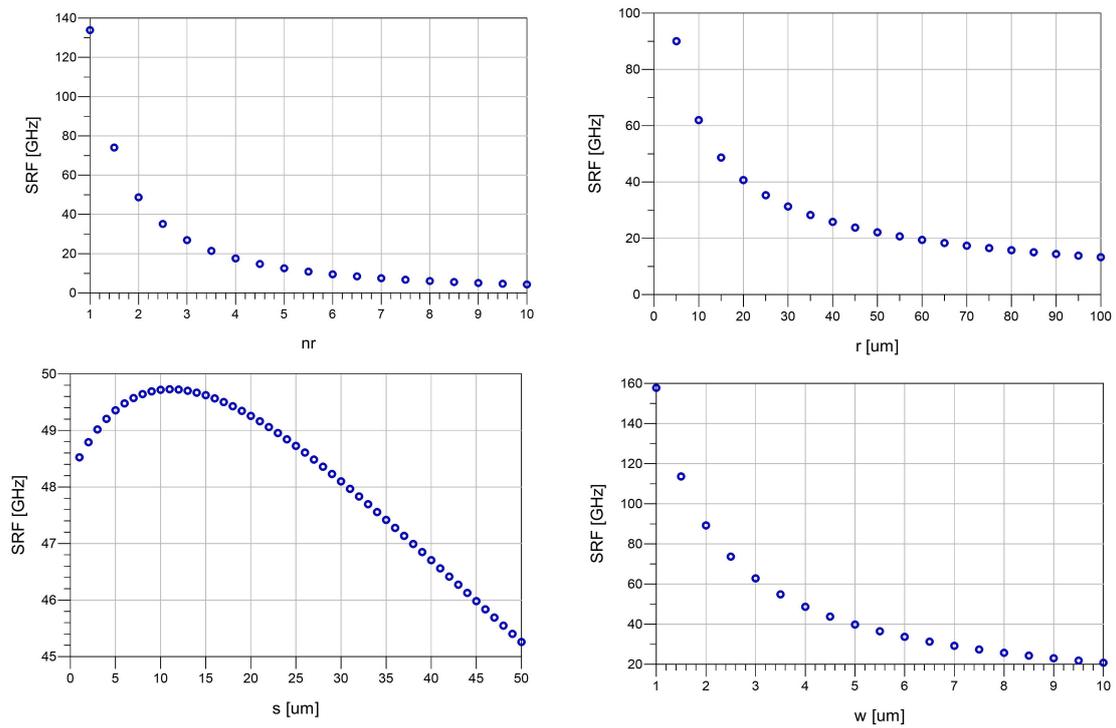
Figure 4.16: Influence on the Self Resonant Frequency
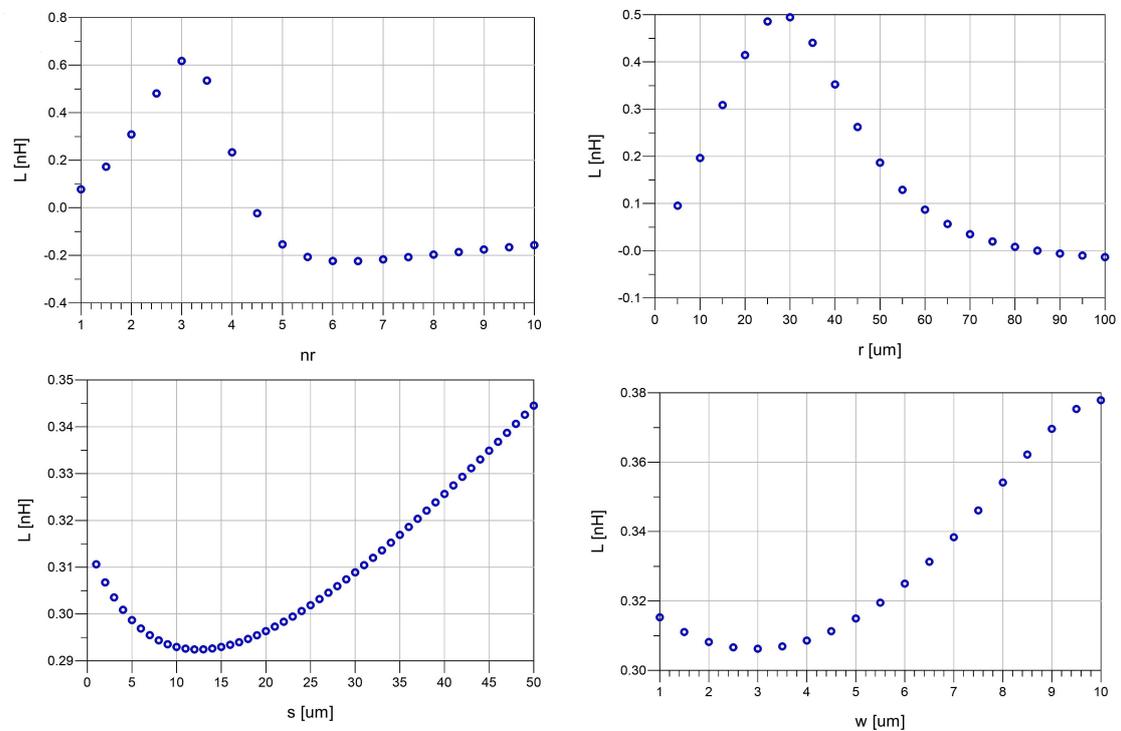
Figure 4.17: Influence on the inductance value (at 15 GHz)

When no such model as in figure 4.14 is available and a quick guess is needed for the inductance value, one can make use of one of the many formulas in literature (e.g. formula 4.8 [34], where n, d_{out}, d_{in} and μ_0 respectively denote the number of turns, the outer and inner diameter and

the permeability of vacuum) :

$$L \approx \frac{1.07\mu_0 n^2 d_{avg}}{2} \left[ln(2.29/\rho) + 0.19\rho \right] \tag{4.8a}$$

$$d_{avg} = \frac{d_{in} + d_{out}}{2} \tag{4.8b}$$

$$\rho = \frac{d_{out} - d_{in}}{d_{in} + d_{out}} \ (fill \ factor) \tag{4.8c}$$

When this formula is used for the example inductor (figure 4.9) a value of 522.16 pH is found, which is a good approximation of the value that results from simulations (i.e. only an error of approximately 5%).

4.5.5   Patterned ground shield

The different loss mechanisms were discussed earlier in this chapter, including the substrate losses. Since those losses result in a large degradation of the quality factor (especially when a heavily doped substrate is used) it is interesting to look at a possible solution to increase the quality factor without having to change the process parameters.
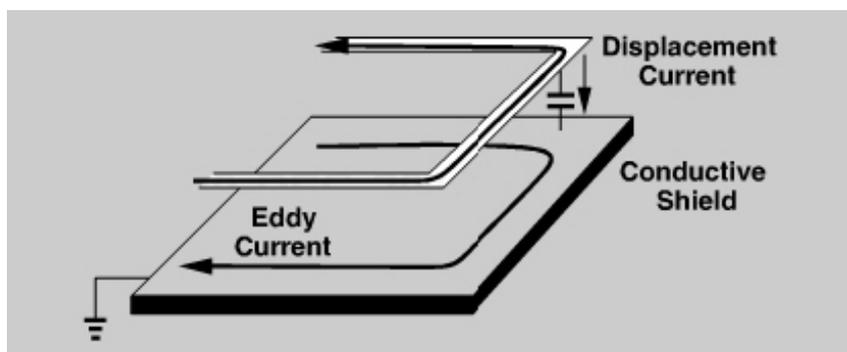
Figure 4.18: Solid ground shield

One of the solutions for this problem is a ground shield (figure 4.18 [25]) designed on a lower level metal layer. Without such a shield, there will be significant energy losses due to the penetration of the electric field into the lossy silicon substrate. Hence, the goal of the shield is to terminate the electric fields caused by the inductor before they reach the substrate.

However, new problems will arise when using a solid ground shield. The fact that a highly conductive layer is placed very close to the coil would mean that strong eddy currents are induced in the shield. Thus when a solid ground shield is placed underneath the coil, the substrate losses are cancelled out (or replaced by smaller shield losses) but the substantial eddy currents underneath the coil will give rise to an opposing magnetic field resulting in a reduction of the apparent inductance value. Consequently the Q-factor will decrease despite the fact that the total amount of losses is reduced.
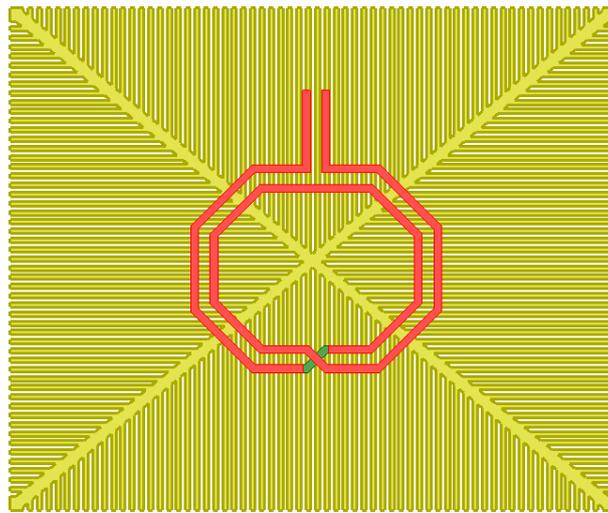
Figure 4.19: Patterned ground shield

An alternative that takes this problem into account is a patterned ground shield (e.g. figure 4.19). This ground shield has the advantage that it provides a low-impedant return path for the electric field while keeping the eddy currents to a minimum. Since gaps are introduced in the shield, some electric field lines might couple to the substrate, resulting in an increase of the substrate losses as compared to the case where a solid shield. Nevertheless, the Q-factor will increase when using a patterned shield instead of a solid shield.

Since the ground shield will act as a parallel plate capacitor with the coil itself, this solution is not suited for very high frequencies as it effectively lowers the SRF. This problem can possibly be solved by letting the shield float, but the most optimal shield still depends on the inductor layout and the operating frequency.

4.6 Transmission lines

4.6.1 Introduction

Apart from lumped components and transistors, an amplifier might also make use of transmission lines. This type of component will for instance be used in a class F power amplifier. Transmission lines help to convert impedances in a frequency dependent manner but are also useful to transport signals over a long distance.

(a) Coplanar waveguide

(b) Microstrip
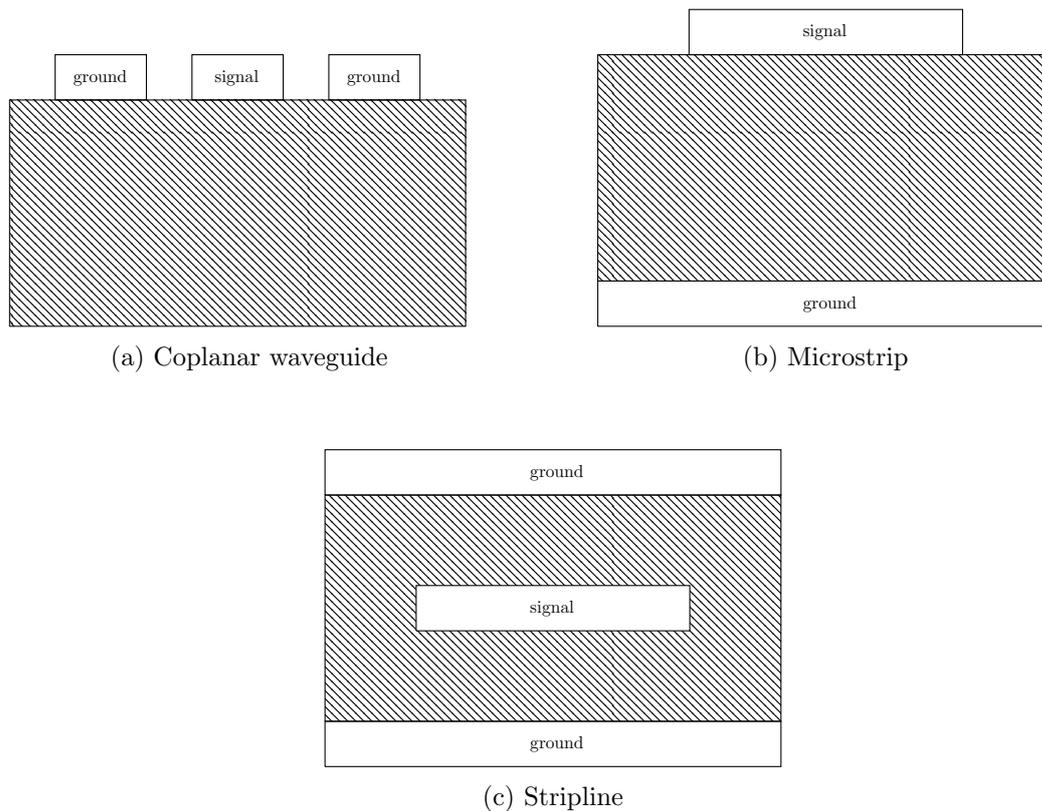
(c) Stripline

Figure 4.20: Types of transmission lines

The different types of transmission lines that are realisable on chip, are drawn in figure 4.20. Striplines are inconvenient to use on-chip for two reasons. Firstly, the signal line will preferably be positioned on one of the uppermost layers since the those layers are least resistive and can handle high current densities. Another reason why striplines are to be avoided is because they result in a high capacitance between ground and signal path, providing only low characteristic impedances. An advantage of the stripline is that it will result in superior isolation when compared to microstrip transmission lines. On the other hand, if the coplanar waveguide is well designed, this type of transmission line will also provide a good isolation to the environment.

There still remains the choice between the other two variants, namely the coplanar waveguide and the microstrip transmission line. While the latter results in a lower coupling to the substrate (at least when a large enough ground plane is used), the coplanar waveguide will be less susceptible to the folding of the waveguide. Since the uppermost metal layers have the lowest R_{sh} for the given technology, an additional advantage is obtained for coplanar waveguides, namely the fact that the ground plane losses will be lower. The optimal type of transmission line will thus depend on the design criteria set for the transmission line. Full wave simulations will be needed to make a well informed choice between the two resulting designs.

For the design of the waveguides, it is important to have highly conductive ground and signal lines to approximate the ideal behaviour of a transmission line. In case of a microstrip or a stripline, this will result in the use of the lowest metal layer instead of the polysilicon layer to implement the (bottom) ground plane. This is cause by the fact that poly has a sheet resistance which is orders of magnitude larger than for the lowest metal layer. Some of the other metal layers will have an even lower sheet resistance but as they lie closer to the other conductor(s), choosing one of them as the (bottom) ground layer will deteriorate the behaviour of the transmission line even

further. The reason why the distance between the different conductors should be maximized is that leakage between the conductors (originating from the losses in the dielectrics) will alter the behaviour of the transmission line.

4.6.2 Quarter wavelength transmission lines

Transmission lines will convert impedances in a frequency dependent manner. This conversion will depend on the length of the transmission line relative to the wavelength. An interesting conversion between impedances takes places when designing the transmission line to have a length of a quarter wavelength at the operating frequency (figure 4.21).
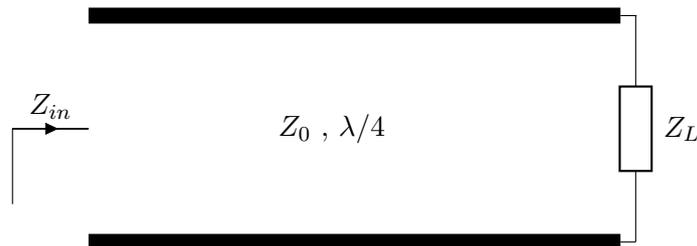
Figure 4.21: $\lambda/4$ transmission line

When this condition is fulfilled, impedances can be transformed by applying equation 4.9 (Z_0 denotes the characteristic impedance of the transmission line). This component can for instance be used to make sure that a short acts as an open at the operating frequency (and every odd multiple of the operating frequency). An application of this type of conversion is the substitution of an RF choke by a $\lambda/4$ transmission line since it will convert an ac short to an open at 15 GHz. As this might be an interesting component, we will elaborate on this subject further on in this chapter.

$$Z_{IN} = \frac{Z_0^2}{Z_L} \tag{4.9}$$

Equivalent to the other components mentioned in this chapter, ideal transmission lines are impossible to obtain and thus the transformation mentioned in formula 4.9 will only be approximately valid. For example, when a quarter wavelength transmission line is needed as a substitute for an RF choke (i.e. the ac termination of the transmission line is a short), the ideal case would result in a periodic circular movement on the unit circle of the Smith chart. However, simulations provides a rather different result (figure 4.22). Starting at the left part of the spiral (i.e. at DC), one can notice that the ideal behaviour of zero ohm DC resistance is not obtained (for this example 4.26 Ω is found to be the DC resistance). When increasing the frequency a spiral can be found which converges to the center by periodically approximating an open and a short for the different harmonics. However, the approximation of opens and shorts deteriorates when increasing the order of harmonic. When designing a quarter wavelength for a class E/F amplifier, it is essential that the behaviour of the transmission line is approximately ideal for the first couple of harmonics but at higher frequencies the quality of the waveguide is allowed to be inferior.
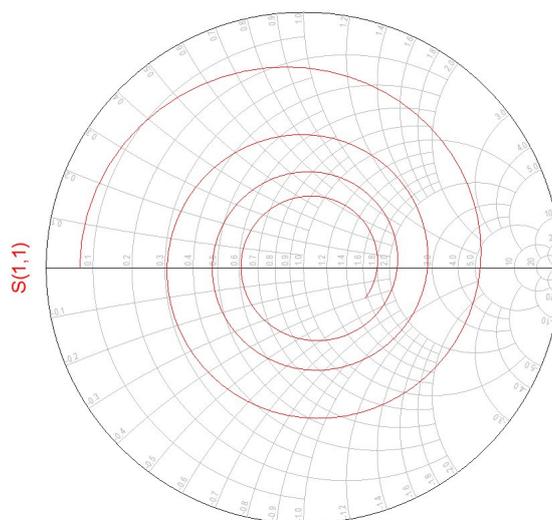
Figure 4.22: Behaviour of a $\lambda/4$ transmission line shorted at the output ($Z_0 = 60.86\ \Omega$)

4.7 Design of Components

4.7.1 Inductors

High inductance values

When a circuit has to be designed, it is important to know which values can be achieved for the discrete components used in the circuit. After taking a look at figures 4.15, 4.16, 4.17 it becomes clear that a high inductance value will be hard to achieve. This results from the fact that a large number of turns will be needed to make a high inductance value, resulting in a low Q-factor (i.e. very lossy components) and a low SRF (i.e. these components are not usable at high enough frequencies).

Since a 600 pH inductor was needed at a certain moment in the design process of a class EF_2 amplifier, it was important to find out the quality factor and SRF that is achievable for this high inductance value. Luckily, this was an inductor that doesn't carry a large AC current and ideally even no DC current. Hence it was possible to choose the track width quite small (which results in lower capacitive losses, effectively increasing the SRF).

	inner radius	track width	spacing	turns	surface area
Single layer	18.14 μm	5 μm	2.5 μm	3	5776 μm^2
Series stacked (2 layers)	11.42 μm	5.6 μm	2.2 μm	2	2485 μm^2

Table 4.1: Dimensions of the 600 pH octagonal inductors

Earlier in this chapter, the concept of a series connected inductor was briefly explained. The advantages of this topology are that it results in lower metal losses and a more compact solution (due to a large mutual inductance between the 2 spirals). The big drawback of a series connected multilayer inductor is the fact that the SRF decreases. After simulation of both variants (table 4.1) the results (table 4.2) indicate that the single layer inductor has more appealing properties for our application. This is due to the fact that the stacked series inductor suffers significantly from interwinding capacitance.

	L	Q-factor	SRF	Max. current	R_{DC}
Single layer	585 pH	11.67	80 GHz	AC: 80 mA / DC: 40 mA	2.722 Ω
Series stacked	596 pH	7.338	43 GHz	AC: 89.6 mA / DC: 44.8 mA	1.971 Ω

Table 4.2: Figures of merit of the 600 pH octagonal inductors

The Q-factor and SRF of the series stacked inductor can be slightly improved by implementing the bottom spiral on a lower metal layer in the substrate stack. However, due to numerous problems (e.g. higher sheet resistance, lower maximal current density, closer to the substrate, ...) the improvements made by doing this shift won't be as great as it would seem at first glance.

When one spiral of the series stacked inductor is simulated as a stand-alone inductor, it results in an inductance of 153 pH. This proves that the inductance value of the stack approximately scales with the square of the number of series stacked spirals. Although it must be mentioned that this law will result in an overestimation of the effective total inductance of the stack.

Ground shield

Previously, it was mentioned that using a shield to decrease the substrate losses might improve or deteriorate the high frequency behaviour, depending on the inductor. This will be shown by comparing the high frequency behaviour with and without shield for two different inductors.

First of all, a 200 pH inductor was needed in the design of a class E and class F amplifier and thus the different implementations of a shield are compared for this inductor. The inductor was implemented by using a one-and-a-half turn octagonal spiral inductor with the following dimensions: the inner radius, track width and spacing were respectively 19.7, 14.38 and 2 μm. The same inductance value was also implemented with 2 turns but this resulted in a lower Q-factor (9.819 as compared to 13.739 in case of one-and-a-half turns) and thus it can be concluded that it is optimal to minimize the number of turns given a certain inductance value. The reason why the Q-factor is higher for the one-and-a-half turn variant is because a smaller amount of turns will allow wider tracks as such effectively lowering the metal losses of the inductor.

Different types of shields (floating as well as grounded variants) were tested for this inductor. The first type is a solid ground shield while the other two are patterned ground shields as in figure 4.19 (first implemented in the bottom metal layer and next in the poly layer). The results from the respective simulations are shown in table 4.3.

	L	Q-factor	SRF
no shield	206.7 pH	13.739	121 GHz
solid shield (bottom metal layer), floating	121.5 pH	7.63	162 GHz
solid shield (bottom metal layer), grounded	109.8 pH	6.742	161 GHz
patterned ground shield (bottom metal layer), floating	206.5 pH	14.323	118 GHz
patterned ground shield (bottom metal layer), grounded	200.2 pH	12.330	118 GHz
patterned ground shield (poly layer), floating	206.6 pH	14.557	119 GHz
patterned ground shield (poly layer), grounded	206.6 pH	14.450	119 GHz

Table 4.3: Shielding of a spiral inductor

From these results, some interesting conclusion can be drawn. When comparing the inductance

value for the different types of shield it becomes obvious that the effective inductance will decrease a lot when eddy currents might appear on the shield, since these currents will give rise to an opposing magnetic field. This is especially the case when a solid shield is used but might also be noticeable when a patterned ground shield, consisting of metal, is used. Because the apparent inductance decreases for these types of shields, the Q-factor will be lowered as compared to the case where no shield is present. The quality of a patterned ground shield will depend on the grid size and when the ground shield has been well designed eddy currents will be minimized while decreasing the substrate losses as much as possible.

For this specific inductor used at 15 GHz, the patterned ground shields implemented in the poly layer seem to be the best choices since the SRF and inductance aren't significantly reduced while the Q-factor improves when adding the shield. Since the accurate simulation of a shield takes a long time to complete it is best to only take this possibility into account when making a final design of the components. It is however important to remark that using a shield will not always be better (e.g. during the design of one of the inductors, namely a 120 pH RF choke, simulations provided a Q-factor of 14.377 for the stand alone variant while the most optimal shield only resulted in a Q-factor of 10.015).

Another interesting phenomenon is that the floating shields seem to work better at this frequency than their respective grounded variant. This is probably due to the fact that the shield acts as a parallel plate capacitor with the coil itself. But since the floating shield will not be connected to a certain voltage, this parasitic effect will not be as visible as in the case of a grounded shield. Since poly is a high-ohmic material, the center of this shield will appear to be floating even when it is grounded resulting in the fact that the difference between a grounded poly shield and a floating variant is not as high as it is in the case of a metal patterned ground shield. Since the distance from poly to spiral is higher than from the bottom metal layer to the spiral the parasitic effects will be lower even further favoring the use of a poly shield.

RF Chokes in differential circuits

In the class EF_2 amplifier a 108 pH RF choke is needed. In this section, several topologies (figure 4.23) are compared to find out the most optimal way to make an RF choke of 108 pH in a differential circuit. Topologies shown in figures 4.23c and 4.23d are in fact single inductors of approximately 216 pH where the center tap is used to provide a power supply connection. Since the 2 branches are slightly asymmetric in both topologies (due to the undercross part) the center tap is also placed in a slightly asymmetric manner.

(a) separate

(b) separate with undercross

(c) combined

(d) 8-shaped
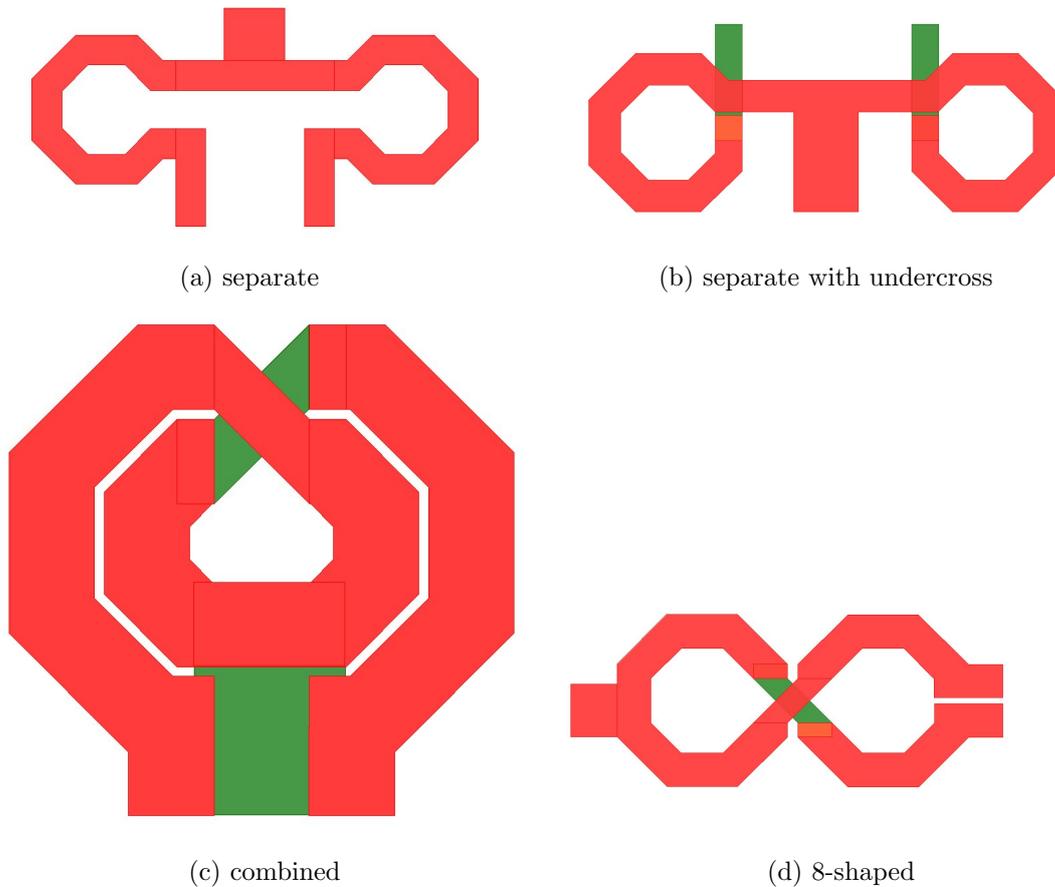
Figure 4.23: RF chokes for differential circuits

Important figures of merit for an RF choke are the DC resistance, Q-factor and SRF but there are also other criteria that might make one topology superior with regard to the other (for example the surface area which has a direct relation to the cost of the chip).

	L	Q-factor	R_{DC}	surface area
figure 4.23a	106.987/107.343 pH	13.675/13.861	0.450/0.448 Ω	24640 μm^2
figure 4.23b	112.839/108.508 pH	12.331 / 11.522	0.408/0.459 Ω	19724 μm^2
figure 4.23c	106.774/105.737 pH	8.521/8.343	0.374/0.280 Ω	13824 μm^2
figure 4.23d	112.307/106.394 pH	13.044/12.547	0.331/0.431 Ω	19946 μm^2

Table 4.4: Figures of merit for the differential RF choke topologies

When comparing the four different topologies (table 4.4, where the properties for both branches are given) it becomes clear that the combined inductor results in the worst Q-factor. The reason for this phenomenon is that the spiral has a wide metal track at its center and this is exactly the place where the magnetic field is the strongest. Due to this strong magnetic field large proximity effects will be induced in these inner tracks. A solution to this problem might be to use more narrow lines at the center of the coil but due to electromigration this is not a valid option. It is important to notice that the RF chokes were designed to be able to carry sufficient (DC) current.

The best topologies seem to be figure 4.23a and 4.23d. When comparing both, one would expect the first one to be the most optimal RF choke but since the DC resistance is smaller for the

8-shaped RF choke the latter results in a more efficient amplifier according to simulations via Cadence. Cadence also proved that it is best to avoid the combined RF choke for the given application, unless the RF choke needs to be as compact as possible.

Due to the overlap part for the crossing, the Q-factor is a bit smaller for the 8-shaped RF choke than for the topology where both windings are separated. Subsequently the high frequency behaviour of the 8-shaped RF choke will be inferior as compared to the case where the chokes for the differential branches are separated.

The biggest advantage of the 8-shaped topology concerns its interference with the neighbouring components. Since the current flows in an opposing way through the two circles, the magnetic fields and hence eddy currents induced by the 2 current carrying loops will approximately cancel each other out at a far enough distance from the choke. This is extra helpful in the given application, since an antenna is placed on-chip and thus it is desirable that the active circuit on itself doesn't radiate significantly.

While this section was about selecting the best RF choke, it can be shown that this will only result in a suboptimal solution. A better way to tackle this problem is to combine the output balun and the RF choke and use the center tap to provide connection to the power supply. Subsequently this will be an important topic in the next chapter.

Final result

In the final design only two unique inductors are used apart from the RF chokes. The desired inductance for the first component is 400 pH while the inductor should be able to withstand an AC current of at least 75 mA. The final structure realising this inductance value is shown in figure 4.24 (and the dimensions of this layout are mentioned in table 4.5).
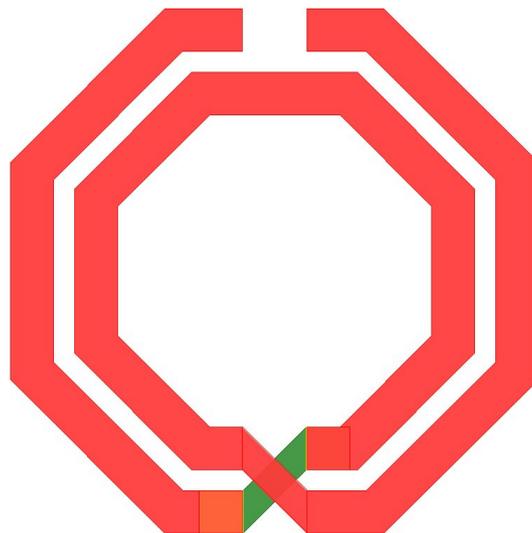
Figure 4.24: Layout of a 400 pH inductor

inner radius	28.41 μm
track width	7.8 μm
spacing	3.8 μm
turns	2
surface area	9120 μm^2

Table 4.5: Dimensions of a 400 pH octagonal inductor

No shielding is used since the use of a patterned ground shield gives rise to a decrease in quality for this specific inductor. At 15 GHz, the use of a shield maximally leads to a Q-factor of 13.769 while the inductor without shield gives rise to a Q-factor of 14.138 (table 4.6).

inductance (15 GHz)	398.6 pH
Q-factor (15 GHz)	14.138
SRF	87 GHz
max. current	124.8 (AC) /62.4 (DC) mA
R_{DC}	1.441 Ω

Table 4.6: Figures of merit of a 400 pH octagonal inductor

Secondly, an inductor of 50 pH is needed which should withstand 120 mA. Since the inductance value is small, a shorted stub can be used. The realised inductance will depend on the length of the stub (equation 4.10 [35]) relative to the wavelength.

$$
\begin{align}
L_{eq} &= Z_0 \frac{1}{f} \frac{l}{\lambda} \tag{4.10} \\
&= \frac{Z_0 l}{c/n} \tag{4.11} \\
&= \frac{Z_0 l \sqrt{\epsilon_r}}{c} \tag{4.12}
\end{align}
$$

By substituting the characteristic impedance and ϵ_r of a coplanar waveguide (which will be discussed in the next section) it can be found that a length of 139 μm is needed to realize a 50 pH inductor. Since the track width and spacing of the final layout (figure 4.25 and table 4.7) differ from the dimensions used for the coplanar waveguide in the next section, the permittivity and characteristic impedance will be slightly different. Hence the required length is not 139 μm but 123.45 μm. In the layout shown in figure 4.25, the middle track will be one terminal of the inductor while the outer tracks should be shorted and form the second terminal.

The fact that the permittivity depends on the spacing can be explained by looking at the electric field lines. When a different spacing between signal and ground lines is used, a different field distribution can be found. And thus a different effective substrate stack will be seen when the spacing is changed, resulting in an alternation of the permittivity of the coplanar waveguide.
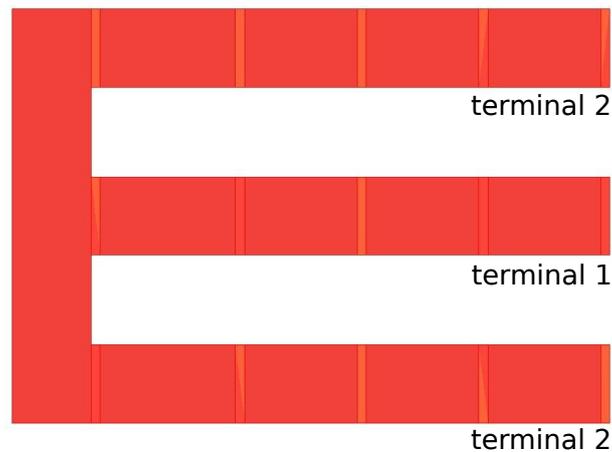
Figure 4.25: Layout of a 50 pH shorted stub inductor

stub length	123.45 μm
track width	17.5 μm
spacing	20 μm
stacked layers	3

Table 4.7: Dimensions of a 50 pH shorted stub inductor

This kind of inductor is usable at high frequencies (i.e. a high SRF is obtained) but will quickly become too large when an average to high inductance value is needed.

In an initial design the track width was chosen to be the minimal track width needed for the given current density since large widths will result in a large device. But this initial layout resulted in a rather low Q-factor (i.e. only 9.98). Consequently the track width and spacing was increased to obtain a higher Q-factor (table 4.8) at the cost of an increase in surface area. The increase in track width is not only advantageous for the Q-factor but also to decrease the DC resistance. The latter is rather essential since this 50 pH inductor should connect the circuitry to the power supply.

inductance (15 GHz)	53.76 pH
Q-factor (15 GHz)	15.126
SRF	>300 GHz
max. current	630 (AC) /315 (DC) mA
R_{DC}	0.076 Ω

Table 4.8: Figures of merit of a 50 pH shorted stub inductor

4.7.2 $\lambda/4$ transmission line

Unfolded transmission lines

Earlier in this chapter several transmission line topologies were discussed. Striplines are to be avoided, hence we will only compare microstrip transmission lines to coplanar waveguides. Initially the unfolded quarter wavelengths will be compared after which the influence of folding will be discussed for the different topologies.

To be able to make a quarter wavelength transmission line, it is important to find out the wavelength in vacuum and the relative permittivity of the waveguide. The first one is easy to compute (5 mm in the case of a 15 GHz fundamental frequency) while the latter has to be found empirically since a stack with different ϵ_r materials is used. As a matter of fact the relative permittivity (ϵ_r) will even depend on the type of transmission line. In case of the coplanar waveguide and microstrip used in the simulations, the respective ϵ_r values are found to be 3.12 and 2.689.

Since a transmission line rotates the impedance in a circular manner on the Smith chart when the characteristic impedance is chosen as reference impedance of the chart, it is important to find out the characteristic impedance of both transmission lines. The tested coplanar waveguide and microstrip respectively correspond with a Z_0 of 60.86 and 61.31 Ω. The exact value of the characteristic impedance is interesting when the behaviour of the transmission line needs to be described but when the substitution of an RF choke is the only goal of the design of the transmission line, the effective characteristic impedance will not have a large impact. In that case, the only specifications that need to be fulfilled are the fact that it should have a small DC resistance and that it should approximate an open at the fundamental frequency.

The microstrip waveguide mentioned in the previous paragraphs uses the bottom metal layer as its reference plane (i.e. the plane where the return currents flow). The alternative using the polysilicon layer as a return path was also tested. This alternative results in a DC resistance which is orders of magnitude larger due to the fact that poly is a bad conductor (even when it is salicided). As a consequence, this variant is to be avoided since it is desirable that the transmission line presents the load as good as possible at the input when looking at DC. It is important to mention that connecting the 2 sides of the reference plane to a more ideal ground will result in the fact that both variants (poly and bottom metal layer as reference plane) will behave approximately the same. This is caused by the fact that DC currents will essentially flow through this more ideal ground resulting in a DC behaviour which will mostly be influenced by the signal plane and a high frequency behaviour that is comparable for both reference planes since the distance of the reference plane to the signal plane will be approximately the same for both variants.

When the transmission line is terminated by a short the behaviour depicted in figure 4.26 is obtained. On this plot the microstrip and coplanar waveguide are respectively indicated by a dashed and a solid line. When taking a look at the Smith chart a large difference in low frequency behaviour can be noted. While the coplanar waveguide has a DC resistance of only 4.27 Ω, the microstrip transmission line has an inferior DC behaviour (namely a DC resistance of 8.784 Ω). This is a consequence of the fact that the return path for the microstrip flows along the rather resistive bottom metal layer while the return path for the coplanar waveguide traverses the low-ohmic upper metal layer. However the high frequency behaviour doesn't differ substantially between both topologies. Consequently the coplanar waveguide technique is to be preferred due to its DC behaviour (and the fact that it is more suited for folding).
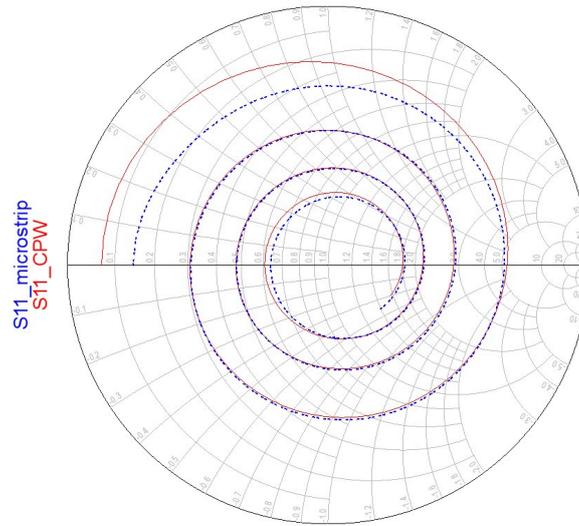
Figure 4.26: S_{11} for a shorted 15 GHz quarter wavelength coplanar waveguide and microstrip

In terms of surface area needed to implement the transmission lines, both variants perform bad (as can be seen in table 4.9) because of the fact that the wavelength is still rather large at 15 GHz. The coplanar waveguide needs to be placed between 2 ground lines and thus uses quite a lot of surface area. On the other hand we have the microstrip transmission line which will be slightly more compact but still rather large. This is caused by the fact that the ground plane needs to be wide enough to catch most of the electric field lines. Hence the required width of the ground plane will depend on the distance between the ground plane and the signal line.

	Surface area
Coplanar waveguide	104710 μm^2
Microstrip	96474 μm^2
8-shaped RF choke	19946 μm^2

Table 4.9: A comparison between the surface areas of unfolded transmission lines and an 8-shaped RF choke

Folded transmission lines

In the previous section it was proven that the DC behaviour of a coplanar waveguide is superior to the DC behaviour of a microstrip. This is however not the only reason why our final design of a transmission line will be a coplanar waveguide. Another important advantage of the architecture of a coplanar waveguide is the fact that the ground lines will shield the signal line from the environment making it more suitable to fold.

For simulation purposes, a coplanar waveguide and a microstrip were designed by using the same signal line section, rendering it possible to compare the effect of the reference plane for both topologies. When comparing the results of this simulation, it can be seen that the difference between the DC resistances of both topologies is less striking than it was in the case of unfolded transmission lines. The reason for this is the fact that the current return path at DC is a lot shorter since the return path directly runs from the output port to the input port. As a consequence, the DC path along the reference plane will be quite short and thus the DC behaviour will be largely dominated by how the signal line is constructed. Resulting from the fact that the return path has shortened, folding will give rise to a superior DC behaviour as

compared to the unfolded variant (i.e. 2.9 and 3.1 Ω respectively for the coplanar waveguide and microstrip). The coplanar waveguide will still be optimal with respect to the DC resistance, although less prominent than it is in the case of the unfolded transmission line.

Since the same structure is used in both topologies to implement the signal line, both waveguides works as a quarter wavelength transmission line at slightly different frequencies. The reason for this can be found in the previous section, namely the difference between the permittivities of the different topologies. Since the coplanar waveguide has a higher ϵ_r this transmission line will resonate at a lower frequency than the microstrip.

When comparing the high frequency behaviour of the two topologies this change in resonant frequency is not the only thing that meets the eye. The folded coplanar waveguide will behave approximately like the unfolded variant. This is due to the fact that the ground lines shield the signal line from the environment, which is in this case the remaining parts of the same signal line (e.g. proximity effects are suppressed). In the case of a microstrip, simulations indicate that the deterioration of the high frequency behaviour as compared to the unfolded variant will be much worse (table 4.10).

	unfolded	folded
Coplanar waveguide	336.75 Ω	331.78 Ω
Microstrip	318.153 Ω	299.63 Ω

Table 4.10: The effect of folding on R_{15GHz} for a shorted transmission line

Coplanar waveguide: parallel tracks

Since the DC resistance of a transmission line is extremely important for the quality of the waveguide it is advantageous to decrease this value. When placing resistors in parallel, the equivalent resistance of the parallel circuit will be smaller than the smallest resistor value. Equivalently, the DC resistance of the transmission line will decrease when connecting transmission lines (that are implemented on different metal layers) in parallel.

To prove this concept three variants of a coplanar waveguide were tested. The difference between the 3 variants is the amount of parallel tracks (implemented for the signal as well as ground lines). The DC resistance of the different transmission lines are mentioned in table 4.11. It is obvious that increasing the number of parallel tracks even further will lower the DC resistance for every track added in parallel to the design. Of course the amount of parallel layers is limited to the number of metal layers in the stack and the maximum allowable coupling to the substrate. When the variants are simulated with a short as termination one can observe that increasing the number of parallel layers will make sure that opens become harder to realize at every odd multiple of the fundamental frequency. Hence to be able to realize decent opens and shorts, the number of parallel tracks in our final design is chosen to be three.

amount of parallel tracks	R_{DC}	R_{15GHz}	R_{45GHz}
1	8.926 Ω	426.07 Ω	200.72 Ω
2	4.249 Ω	386.95 Ω	180.94 Ω
3	2.832 Ω	331.79 Ω	160.95 Ω

Table 4.11: Parallel tracks in a coplanar waveguide

Corners

One of the implicit specifications of a component is that, under normal circumstances, its lifetime surpasses the duration of use of the device in which it is used. Consequently octagonal corners are desirable to improve the lifetime of the device since those bends are less prone to failure than 90 degree corners [31]. This design trick was already mentioned in the part on inductors and can again be used when designing transmission lines. However, this is not the only reason why octagonal corners are to be preferred. First of all, the DC resistance can be proven to be slightly smaller in the case of the octagonal corners (in our case the difference in DC resistance is as large as 4%).

Another reason why octagonal corners should be used is because of the improvement in the high frequency behaviour. Since octagonal corners are less disruptive features in the transmission line, the reflections on these corners will be smaller, resulting in a high frequency behaviour which is much more alike to the behaviour of an unfolded transmission line.

Final result

When combining the elements mentioned in the previous sections, a relatively good transmission line can be made (figure 4.27). This layout uses octagonal bends and stacks three parallel layers to make sure that the final DC resistance is sufficiently small. The three uppermost metal layers are used since they can conduct the highest current densities and because the sheet resistances of the top metal layers are the lowest that can be found in the given stack.

484.2 µm

287.3 µm
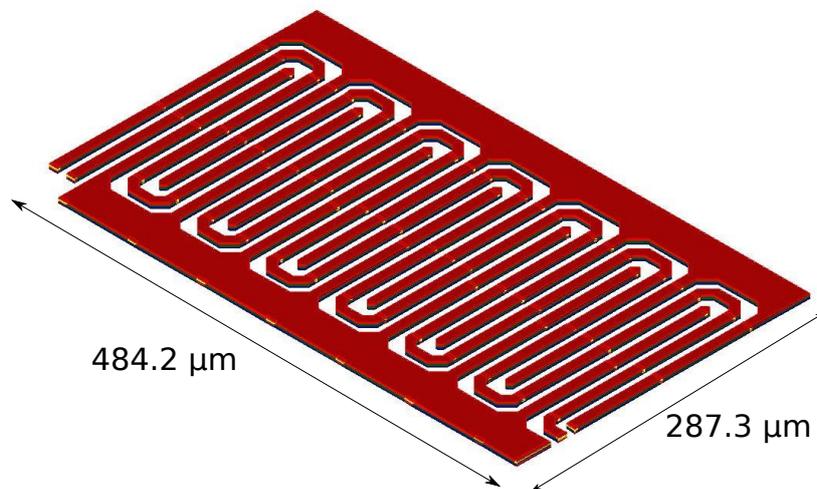
Figure 4.27: Layout coplanar waveguide

Full wave simulations indicate that the DC resistance achieved with this structure is 2.832 Ω while the characteristic impedance of the transmission line is equal to 62.366 Ω. Since the transmission line might be used as an alternative for an RF choke, it is important to look at the reflection coefficient of the structure when the output is terminated by a short. This reflection coefficient is depicted in figure 4.28, where the characteristic impedance of the transmission line is used as the center of the Smith chart. When this transmission line needs to take over the role of the RF choke it is important that a large current is able to flow through the metal. Due to the stacking of multiple layers and a relatively wide metal track (i.e. 8 μm) the transmission line can conduct up to 144 mA DC current (288 mA AC current) before the current density might become too high.

Additionally, the tolerance of the ground was tested by misaligning the ground planes relative to the signal line (i.e. the signal line was shifted by 0.5 μm in the direction of the longest dimension of the coplanar waveguide). Of course, no change in the DC resistance was observable but also the high frequency behaviour did not differ significantly (e.g. R_{15GHz} changed from 331.786 Ω to 331.754 Ω). Consequently it is clear that the ground plane has been well designed.
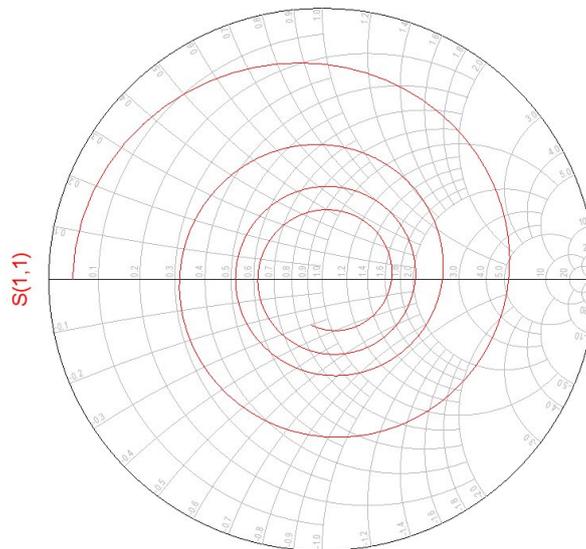
Figure 4.28: S_{11} of the final coplanar waveguide terminated with a short

Efficiency is not the only thing that is important when an antenna-on-chip is made. To gain enough profit, it is essential to take into account the surface area used by the amplifier. In the case of this coplanar waveguide, the occupied surface area will be 0.13937 mm^2. Luckily the ground track lying around the signal line shields the transmission line from the environment (but not from the substrate) making sure that neighbouring components can be placed relatively close to the transmission line without suffering too much from its proximity.

When a decision is made based on the DC resistance, it is clear that the RF choke implemented via a spiral inductor will be better. However, due to the bad approximation of an open at 15 GHz, the possibility exists that a transmission line will result in a better RF choke. A technique which offers a decent DC resistance as well as a good approximation of an open at 15 GHz will be explored in the next chapter. This technique will be based on an LC lattice structure.

4.7.3 Resistors

In the square wave generator at the input of the chip, feedback resistors are desired. Since the resistance value is limited, non-salicided polysilicon is used instead of a p-well resistor. The resistance value needed for this device is approximately 3 kΩ. This results in a rectangular of for example 1 μm by 4.62 μm. In the final design, a resistor of 1 by 5 micron was used (3.25 kΩ) since the exact resistance value is not that important.

Chapter 5

Interstage connections & Power combiners

5.1   Introduction

To be able to achieve the specifications set for the design of the PA it might be needed to implement different stages. As the last stage will most likely determine the total efficiency of the PA, this one will need to be highly efficient. To drive this final stage, a certain input power will be needed which might not be available at the output of the signal generator. A solution to get enough power at the input of this final stage, is to add one or more driver stages between the signal generator and the final stage.

The interconnection of the different stages is not always trivial. Since the impedance levels can differ significantly for the input of one stage and the output of the previous stage, it might be necessary to combine them via a transformer. Removing the interstage transformer in such a system might introduce major reflections, rendering the amplifier inefficient. In this chapter, a comparison will be given between 2 techniques to do this interstage matching. Firstly, LC lattice transformers will be briefly discussed and afterwards a variant will be described which uses two coupled inductors (which is typically called monolithic transformers). The fact that different impedance levels can be matched, makes sure that stages can be optimized without paying attention to the circuitry to which they should be connected.

Transformers are not only useful to connect stages, they will also be used to split the power at the input and combine the power of the different branches at the output. The latter effectively increases the output power.

5.2   LC circuits

The first method of transforming impedance levels, does this transformation by using discrete components designed according to the rules mentioned in previous chapter. Consequently the design process of an LC circuit will by quite easy. Unfortunately, this technique also gives rise to several disadvantages when compared to the monolithic transformers. Especially the fact that the solution takes in a lot of space and that both branches of the balanced port will see a different circuit when this LC transformer is used as a balun, will make sure that LC matching should be avoided when possible.
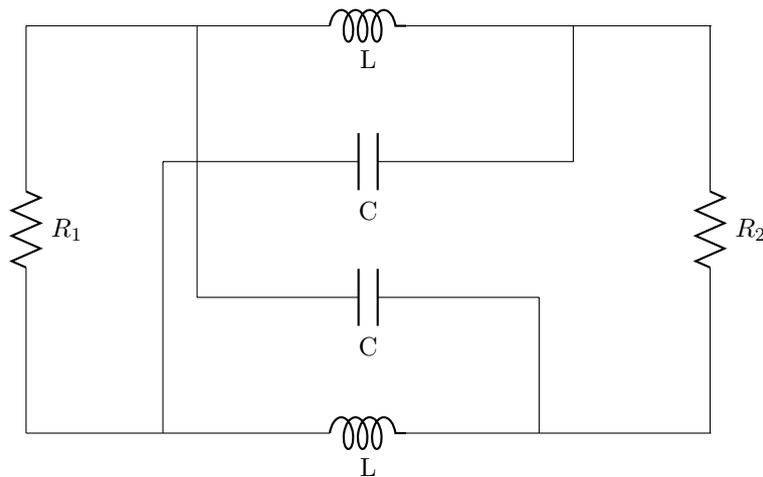
Figure 5.1: Lattice LC transformer

When it is desirable that a load impedance R_2 is seen as R_1 at the input port, the circuit described in figure 5.1 [36][37] can be used. The values for the inductors (L) and capacitors (C) can be easily computed when the resistor values R_1 and R_2 are known (equation 5.1, where the parameter f denotes the frequency for which the transformation circuit is designed, namely the operating frequency).

$$L = \frac{\sqrt{R_1 R_2}}{2\pi f} \tag{5.1a}$$

$$C = \frac{1}{2\pi f \times \sqrt{R_1 R_2}} \tag{5.1b}$$

When one of the terminals of the load is connected to ground (possibly via a decouple capacitor) figure 5.1 can be redrawn in such a way that two separate branches appear (figure 5.2). An interesting remark that can be made concerning this schematic is the fact that both terminals at the input see a different circuit when the frequency differs from the operating frequency. In the schematic drawn in figure 5.2 the upper branch connects to the load via a lowpass filter while the lower branch connects to it via a highpass filter. As already mentioned, this is one of the reasons why LC baluns are to be avoided when designing a differential circuit.
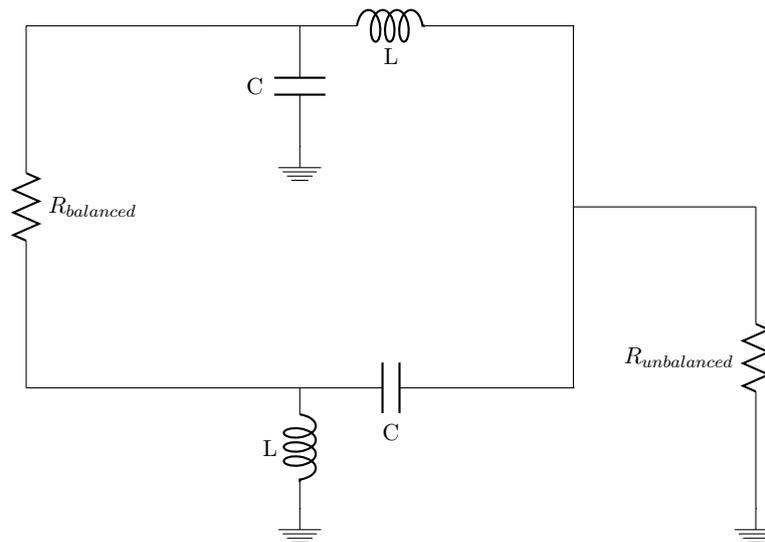
Figure 5.2: LC balun

It is important to mention that figure 5.1 is not the only valid option to transform impedances with an LC circuit. Another LC technique to achieve this goal starts from a matching circuit used for single ended impedance matching (e.g. the LC circuit depicted in figure 5.3 [38]). To transform a differential load of $2 \times R_2$ to $2 \times R_1$ one should first transform a single ended load R_2 to R_1. The differential LC circuit will then consist of 2 inductors with the same value as in the single ended matching circuit and with a capacitor realizing half the capacitance value as compared to the single ended circuit. Unfortunately, this type of circuit can't be used as a balun.

(a) Single ended

(b) Differential
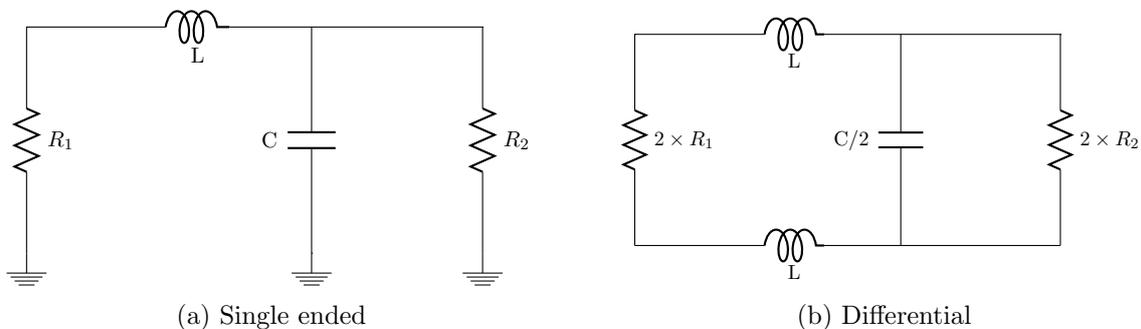
Figure 5.3: LC matching $(R_2 > R_1)$

The component values used in figure 5.3 are the ones that are expressed in equation 5.2 [38]. To make the matching circuit as efficient as possible it is important to include the parasitic capacitances of the inductors in the computation of the explicit capacitor. As a result, the optimal C value will differ from the one that is found when evaluating equation 5.2.

$$Q = \sqrt{\frac{R_2}{R_1} - 1} \tag{5.2a}$$

$$L[nH] = \frac{0.159 \times R_2}{f[GHz] \times Q} \tag{5.2b}$$

$$C[pF] = \frac{159 \times Q}{f[GHz] \times R_2} \tag{5.2c}$$

Although LC circuits are designed for a given frequency, a broadband solution is obtained when R_1 and R_2 do not differ immensely. This is caused by the fact that a small transformation ratio will result in a low Q-factor of the matching system resulting in a large bandwidth. Hence the bandwidth of an LC matching circuit might be larger than the bandwidth of a monolithic transformer due to the fact that monolithic transformers suffer from potentially large interwinding capacitances. Nevertheless, when a high transformation ratio is needed, monolithic transformers will provide a broader passband.

5.3 Monolithic transformers

5.3.1 Working principle

Earlier in this chapter, LC baluns/transformers were discussed. Another way to implement a transformer is by using two (AC) coupled inductors. Those monolithic transformers work due to the fact that both inductors can create a magnetic field as well as sense the magnetic field induced by the other coil. A schematic that indicates the basic components of such a transformer is shown in figure 5.4. While the turn ratio in this schematic is N_p:N_s, it is not needed to have N_p identical turns for the primary and N_s identical turns for the secondary as long as the inductance ratio is chosen correctly. When desired, a center tap can be added to the primary and/or secondary to add a connection to ground or to a power supply.
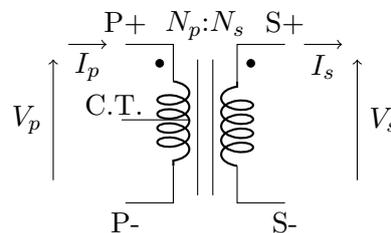
Figure 5.4: Basic transformer

To describe the relations of the currents, voltages and impedances, sign conventions have to be made. The direction of positive currents and voltages are indicated on figure 5.4. When those conventions are used, equation 5.3 is found to describe the transformer operation. It is important to note that these equations are only valid when assuming an ideal transformer with coupling factor 1.

$$\frac{V_p}{V_s} = \frac{N_p}{N_s} \tag{5.3a}$$

$$\frac{I_p}{I_s} = \frac{N_s}{N_p} \tag{5.3b}$$

Starting from these relations, it can be shown that the input impedance Z_{in} (seen at the primary winding) is a rescaled version of the load impedance Z_L that is connected between the terminals of the secondary winding (equation 5.4). Since inductance values scale quadratically with the number of turns, the ratio between the two inductance values should correspond with the desired impedance transformation ratio.

$$\frac{Z_{in}}{Z_L} = \frac{V_p}{I_p} \times \frac{I_s}{V_s} = \frac{N_p^2}{N_s^2} = \frac{L_p}{L_s} \tag{5.4}$$

In practice, no ideal coupling between the primary and secondary is obtained and the effective coupling depends on the class of transformers. The amount of signal coupled from one spiral of the transformer to the other can be characterized by the use of a coupling factor k (formula 5.5, where M denotes the mutual inductance between the two coils). This value will be larger than zero and smaller than one. Off-chip transformers can easily attain a coupling factor which approximates unity as opposed to on-chip transformers. The reason for this is that off-chip transformers make use of a high μ material to capture the magnetic field lines while the core of an on-chip transformer is for example SiO_2 which means that a relevant fraction of the field lines will couple to the environment. The coupling factor that can actually be obtained on-chip will depend on the geometry of the transformer which will be discussed later on in this chapter.

$$k = \frac{M}{\sqrt{L_1 L_2}} \tag{5.5}$$

In the formula for the coupling factor the variable M is used. This value is the mutual inductance between the two coils. By using the self inductances and the mutual inductance, formulas for the relations between voltages and currents can be given (formula 5.6). These equations prove that a time varying current in one inductor will induce an additional current in the other. Since the current has to vary in time to be coupled via the induced magnetic field, a monolithic transformer won't pass DC signals.

$$V_p = L_p \frac{\partial I_p}{\partial t} - M \frac{\partial I_s}{\partial t} \tag{5.6a}$$

$$V_s = L_s \frac{\partial I_s}{\partial t} - M \frac{\partial I_p}{\partial t} \tag{5.6b}$$

In figure 5.4 dots were used to provide info on the polarity of the transformer. Since the sign of the coupling term in equation 5.6 depends on the direction of the current, it is important to find out the polarity of a monolithic transformer (figure 5.5) or an unexpected 180 degree shift might occur.
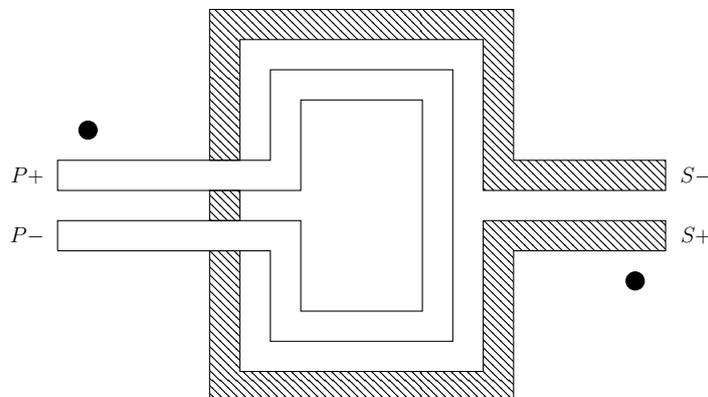
Figure 5.5: Polarity of a stacked monolithic transformer

5.3.2 Monolithic baluns

Baluns are a special type of transformers and aim to transform differential signals to unbalanced signals or vice versa. In this thesis, baluns were made by providing the right connections to a transformer with the correct turn ratio. In other words, when a 1:n balun is needed, a 1:n transformer is designed and the S+ terminal is shorted to ground. This is however a suboptimal solution since the design does not take the voltage profile into account. In literature (e.g [39]), transformer geometries are proposed that are designed to function as a balun. The cited example geometry has been implemented in this thesis for the required 50-to-50 Ω balun, but the final layout of this balun results in inferior behaviour as compared to the 50-to-50 Ω transformer where the S+ terminal is shorted to ground. The reason for this is that a 432 pH 1-turn inductor is needed for the implementation of this geometry which gives rise to 2 problems. First of all a significantly larger surface area is needed as compared to the case where a regular transformer is used as a balun. Secondly a large diameter is needed to obtain this inductance value with only 1 turn and this results in a decrease in the Q-factor (figure 4.15). Hence the profit in efficiency due to the implementation of this type of geometry is countered by the decrease in Q-factor due to an increase in inner diameter.

5.3.3 Transformer classes

Earlier in this chapter, monolithic transformers were defined as two coupled inductors. However it was not specified how these inductors are placed relative to each other such that the coupling would be realised. Roughly speaking, one can subdivide the monolithic transformers in three classes (figure 5.6 [40]) where each class has its own set of advantages and disadvantages (table 5.1). As long as one can live with the high interwinding capacitances (and hence the low SRF) the stacked transformer will be the most optimal type since it is a compact solution with a high coupling factor (resulting in a high efficiency when qualitative inductors are used). Another advantage of a stacked transformer is the fact that inherent autoshielding is present, meaning that the upper coil will be shielded from the substrate by the lower coil. This will result in a lower coupling to the substrate and consequently to a higher efficiency of the final transformer.

(a) Tapped   (b) Interleaved   (c) Stacked
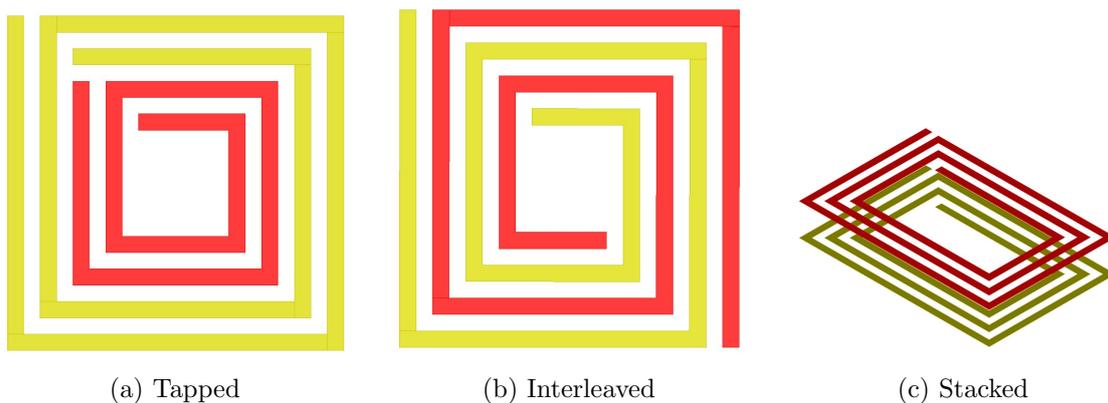
Figure 5.6: Types of transformers

	Inductance density	Coupling factor	SRF
Tapped	Mid	Low	High
Interleaved	Low	Mid	High
Stacked	High	High	Low

Table 5.1: Types of transformers: properties

When the high interwinding capacitances of a stacked transformer form an obstacle, the primary and secondary winding can be shifted relative to each other. The disadvantages of shifting the primary with regard to the secondary is that the coupling of the upper spiral to the substrate increases and that the coupling factor k decreases. These two trends can potentially lower the total efficiency. However, the efficiency might increase when a small shift is applied. This is caused by the increase in the Q-factor of the windings resulting from the decrease in interwinding capacitance.

In older technologies, stacked transformers might be hard to realise due to the limited number of metal layers. The stacked transformer requires at least two layers for the implementation of the transformer itself while additional layers are needed to implement the crossings. In the gpdk45 technology this will not cause further problems since 11 metal layers are present.

When a sufficiently high number of metal layers is available it is easy to make transformers with a turn ratio of 2:1, 3:1, ..., n:1. This is for example done by designing the secondary first and stacking n of those in series to create the primary. However this will most likely not result in a ratio which is exactly equal to n:1 due to imperfect coupling. Another problem arising from this technique is the fact that the transformation ratio is limited. There are two reasons for this, firstly the limited number of metal layers present in the stack and secondly the fact that lower metal layers are to be avoided for several reasons (high sheet resistance, low maximum current density and high capacitive coupling to the substrate).

5.3.4 Design process of a stacked transformer

The design of a transformer requires two essential steps. First the transformer itself is implemented after which matching is applied to the circuit to improve the efficiency of the power transfer. For the design of the transformer itself one needs to find the required L_p and L_s which will depend on the load value and the desired transformation ratio. Subsequently, the inductors are designed according to the design rules mentioned in the previous chapter.
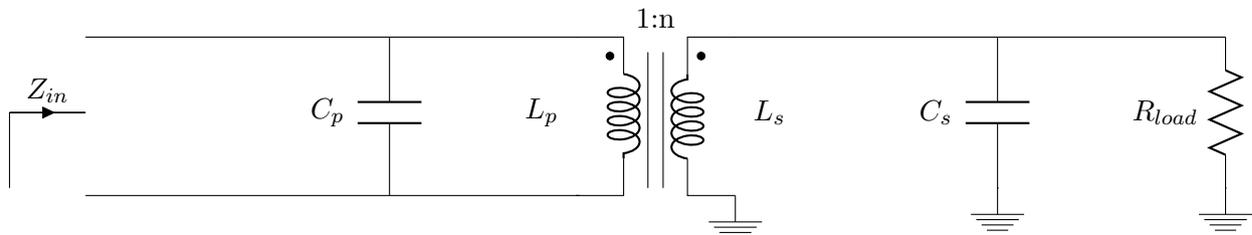
Figure 5.7: Basic design circuit of a transformer

In figure 5.7 the transformer itself is depicted together with the matching circuit. This schematic consists of a transformer with a 1:n turn ratio which will ideally result in a 1:n^2 impedance transformation (equation 5.7).

$$Z_{in} = \frac{R_{load}}{n^2} \tag{5.7}$$

To choose the value of L_p that is best suited for the desired impedance transformation, formula 5.8 can be used [41].

$$\omega_o L_p = \frac{1}{\sqrt{\frac{1}{Q_s^2} + \frac{Q_p}{Q_s}k^2}} \times \frac{R_{load}}{n^2} \tag{5.8}$$

Next, the inductance value of the secondary inductor is easily obtained when the desired transformation ratio is known (equation 5.9).

$$L_s = L_p \times n^2 \tag{5.9}$$

For the second part of the design, values for C_s and C_p are needed to resonate out (part of) the reactance. Starting values for these capacitors are computed taking simultaneous matching of a 2 port network into account [42, p. 572-573] where port 1 and 2 will respectively consist of the primary and secondary terminals. While no special attention was spent on baluns during the design process of the transformer itself, the balun operation is taken into account during the implementation of the matching network (i.e. S-parameters needed for simultaneous matching are derived with the S+ terminal shorted to ground). Of course there is a limit to the elements that can be used in the matching circuit (e.g. series capacitors are prohibited when a DC connection to the center tap is required). Final C_s and C_p values can be made different from the values derived by simultaneous matching when a net reactance is desired at the primary.

The efficiency of the final impedance conversion will thus depend on the transformer itself and on the matching circuit. For this efficiency an upper bound can be determined (equation 5.10 [41]) which will be a function of the Q-factor of the inductors as well as the coupling between them.

$$\eta_{max} = \frac{1}{1 + 2 \times \frac{1}{Q_1 Q_2 k^2} + 2 \times \sqrt{(1 + \frac{1}{Q_1 Q_2 k^2}) \times \frac{1}{Q_1 Q_2 k^2}}} \tag{5.10}$$

When this formula is computed for different values of k and Q (assuming both inductors to have the same Q-factor which is highly unlikely), it can be seen (figure 5.8) that the maximally achievable efficiency will increase when Q and/or k is increased. Hence a stacked transformer will most likely result in the most efficient transformation.
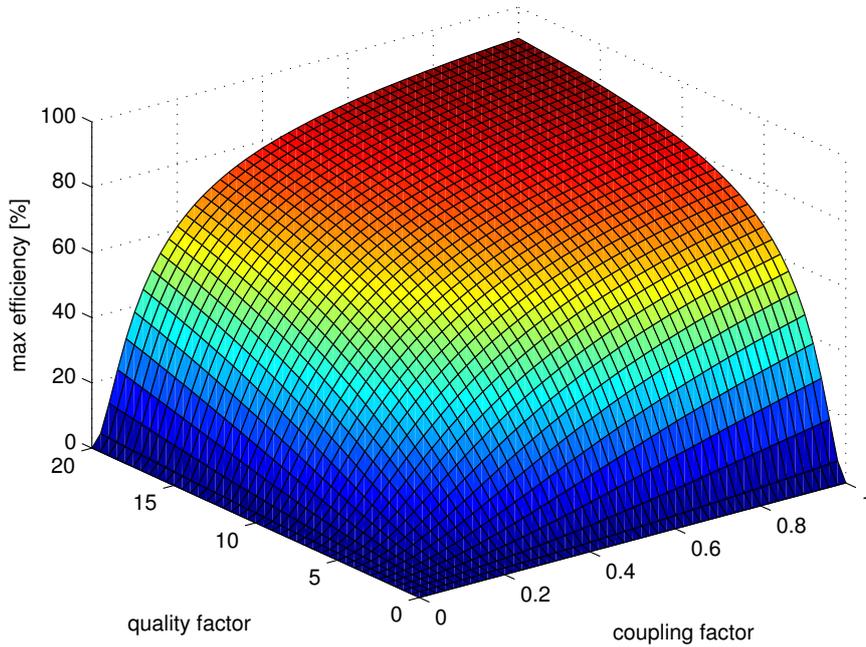
Figure 5.8: Upper bound on the efficiency of a transformer

Since stacking usually provides the maximum efficiency, stacked transformers will be our first choice. In a stacked transformer, an extra degree of freedom is present, namely the choice of which winding that should be put on top of the other. The most optimal way to use this degree of freedom is by putting the highest inductance on top of the lowest ("auto-shielding"). The reason for this is that the highest voltages will be induced in the inductor with the largest inductance value, resulting in a potentially higher coupling to the substrate when this auto-shielding technique is not applied.

5.3.5   Simulation technique to determine the coupling factor

Since the coupling factor k is an important parameter to describe a transformer, it is essential that this value can be easily deduced from simple simulations. For this purpose it is interesting to write down the relation between V_p and I_p in 2 distinct cases (equation 5.11), namely when the secondary is left open and when the secondary terminals are shorted. Those are also the relations that one would find when simulating inductors with a respective value of L_1 and $(L_1 - \frac{M^2}{L_2})$.

$$V_{open} = j\omega L_1 I_1 \tag{5.11a}$$

$$V_{short} = j\omega(L_1 - \frac{M^2}{L_2})I_1 \triangleq j\omega L_{short}I_1 \tag{5.11b}$$

Two quick simulations can thus provide the values for L_1 and L_{short}. Since k is a function of M, a next step to find the coupling factor will be to rewrite the formula for L_{short} (equation 5.12).

$$M^2 = L_2(L_1 - L_{short})$$ (5.12)

In a final step, this formula for the mutual inductance is substituted in the equation for the coupling factor (equation 5.5). This will result in a simple method to find k (equation 5.13) where L_{short} and L_1 are respectively the inductance measured for a shorted secondary coil and for an open at the secondary.

$$k = \sqrt{1 - \frac{L_{short}}{L_1}}$$ (5.13)

5.4   Power combiners and splitters

Transformers are not the only components that can be used to split and combine the power at respectively the input and output of the amplifier. However, most other options (e.g. a Wilkinson combiner, figure 5.9) make use of quarter wavelength transmission lines. Unfortunately those transmission lines are rather large at 15 GHz. As a consequence, no further attention will be paid to these types of splitters/combiners.
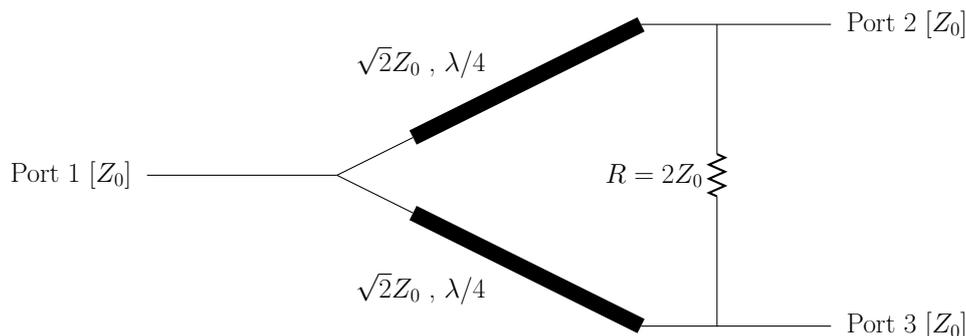
Figure 5.9: Wilkinson power combiner/splitter

5.5   Mixed-mode S parameters

The S parameters relevant for transformer design are not the regular single ended S-parameters. While transformers are 4 port circuits (5 port or 6 port when the primary and/or secondary winding has a center tap) it is more interesting to look at the equivalent network were both primary ports are combined and both secondary ports are combined to get a mixed mode network (figure 5.10). To convert the S-parameters of a 4 port network to the mixed mode S parameters of the equivalent mixed mode 2 port network, matrix multiplication can be used [43]. Those mixed mode S parameters will then give an indication on how differential and common mode steering at the input will be handled towards the output.

(a) Single ended S parameters

(b) Mixed mode S parameters
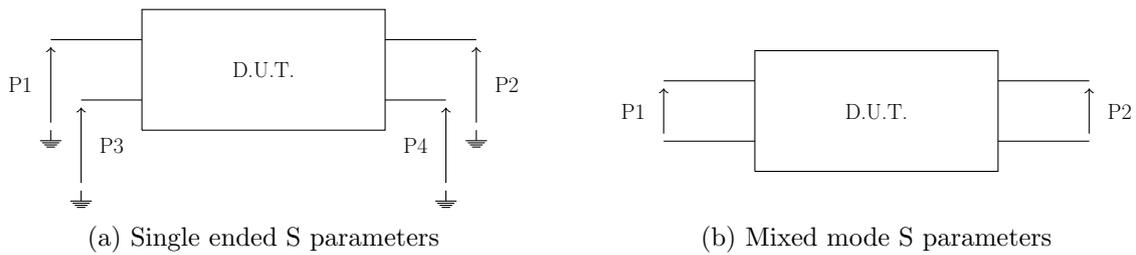
Figure 5.10: S parameters: single ended versus mixed mode

5.6 Design of Components

5.6.1 Introduction

In the remaining part of this chapter, several transformers and baluns will be discussed. Essentially, those transformers consist of inductors and capacitors. Since these capacitors result in an approximately ideal behaviour (with a high Q-factor), they will be replaced by ideal components to simplify the calculations.

Earlier in this chapter, two types of transformers were discussed, namely LC and monolithic transformers. A First step in the design of LC transformers is the design of the inductor. Subsequently the right capacitance value is chosen, which might differ from the ideal value (formula 5.1) when the parasitic capacitances from the inductors are taken into account. Special attention should be paid when comparing the surface areas in the next sections since the total area of an LC transformer is computed as the sum of the surface areas needed to implement the stand alone components. Hence the surface area mentioned for the LC transformers is a vast underestimation since a real design requires a considerable spacing between the components to be able to neglect the interference between them.

The design process for the second type of transformers (i.e. the monolithic transformers) can be found earlier in this chapter. However an important remark has to be made concerning the given layer stack. Since the two upper metal layers have the lowest sheet resistances and are furthest from the substrate, one would expect them to result in the most efficient transformers. Although the Q-factor is indeed higher for these windings, using the upper two metal layers will not always result in the most efficient amplifiers. The reason for this is that the η_{max} depends on the coupling factor as well as the Q-factor. By looking at the substrate stack, one can observe that the thickness of the dielectric between the two uppermost metal layers is rather large which gives rise to a lower coupling factor, as such resulting in a potential decrease of the maximum efficiency. Hence it might be advantageous to use the 2nd and 3rd metal layer (as seen from the top) instead of the two uppermost layers. Of course this trick is not available when the transformer needs to be able to withstand high currents. To prove the original statement, the output combiner at the end of this chapter was simulated in both cases and an efficiency gain of 1.4 % was obtained by using the 2nd and 3rd metal layer (as seen from the top).

5.6.2 LC balun as an RF choke

To implement an RF choke, one can use a spiral inductor. This will result in an excellent low frequency behaviour but since there is an upper limit to the inductance value that can be achieved, the RF choke won't present a good enough approximation of an open at the application

frequency. A solution to this problem was discussed in the previous chapter (i.e. using a quarter wavelength transmission line) but in that case a mediocre DC behaviour is obtained.

A lattice LC structure (figure 5.11) can be used to obtain decent DC behaviour and RF behaviour at the same time.
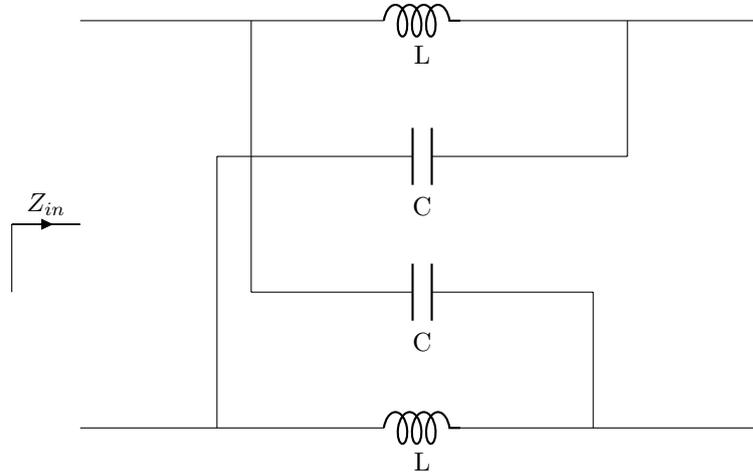
Figure 5.11: Lattice LC transformer as an RF choke replacement

To be able to compare this circuit with the other solutions in this thesis, simulations were run in case of a 113 pH inductor (where the inductance value was chosen arbitrarily). As the device needs to represent an open at the operating frequency, a 996.3 fF capacitor is needed.

The simulation results indicate that the DC resistance (0.8 Ω) is higher than for a straightforward implementation (0.381 Ω) but is lower than when a transmission line is used to implement the RF choke (2.832 Ω). Regarding the behaviour at 15 GHz, this device is superior to both variants. The biggest disadvantage of this implementation is however the fact that two inductors are needed which gives rise to a large total surface area. In the example configuration a total surface area of 22852 μm^2 is required, apart from the spacing that will be needed to make sure that the different components have a negligible influence on each others behaviour.

5.6.3 A comparison between LC and monolithic transformers/baluns

2:1 Transformer

In this section, 2 types of transformers will be compared. Each of them will approximately provide a transformation of a 50 Ω load to a 25 Ω input impedance. To make a good comparison between both, each will be designed to accommodate the same current as this will determine the width of the inductors as such influencing the Q-factor and consequently the efficiency of the transformer.

Monolithic 2:1 transformer: In a first step of the design, the required inductance values are computed via formulae 5.8 and 5.9. An unknown constant is present in equation 5.8 and thus an initial guess of this value is needed to find out the desired value for L_p. When assuming the Q- and coupling factor to be approximately 13 and 0.75, the desired values for L_p and L_s are found to be respectively 351.83 and 703.66 pH. The effective implementation results in inductance values of respectively 356.2 and 705 pH and consists of two turns for the primary winding and

three turns for the secondary where the secondary is split in two series connected spirals, namely one turn below the primary and two turns in a metal layer above the primary.

To find out the efficiency of this transformer, matching needs to be performed, but let's first take a look at the maximally achievable efficiency for the given monolithic transformer (equation 5.10). First the Q-factors of the different inductors are needed (namely $Q_p = 9.417$ and $Q_s = 9.155$). Secondly the coupling factor is computed via the method derived earlier in this chapter. The resulting coupling factor (i.e. 0.842) gives rise to a maximally obtainable efficiency of 77.48% for the given transformer.

The matching circuit itself consists of capacitors. To make sure that a DC connection can be made to the center tap (to connect with a ground or supply voltage), only parallel capacitors should be used. For the given transformer, the capacitors C_p and C_s will have respective values of 256 fF and 128 fF and are connected as in figure 5.7.

Unfortunately, the results indicate that the unknown constant was overestimated and thus an iterative process can be used to optimize the design when needed. Since this transformer is not used in the final design, no extra iterations were done. Nevertheless, the 4:1 transformer in one of the following sections will make use of the Q-factor and coupling factor of this device to make a better guess regarding the constant in formula 5.8.

LC lattice 2:1 transformer: To make a 2:1 transformer with an LC lattice (figure 5.1), two inductors of 375.13 pH and two capacitors of 300.11 fF are needed (equation 5.1). The inductor used in the simulations has an effective inductance value of 376.4 pH, a Q-factor of 13.852 and an SRF of 93 GHz. Due to the nonidealities a reactive part appears at the primary when the secondary is connected to a load of 50 Ω. To make the apparent load at the input purely real the capacitance values are increased to 345 fF. Consequently the apparent transformation seems to be only 22:50 and this can be brought back to a 24:50 Ω transformation by increasing the inductance (to 400 pH) and reducing the capacitance (to 330 fF).

Conclusion: The resulting figures of merit can be found in table 5.2. From extra simulations, other observations can be made that are not shown in this table. First of all, it is important that the impedances seen at the positive and negative primary terminal are approximately the same and rather frequency independent (for the frequency band of interest). Simulations prove that this is the case for both variants and that the LC transformation results in a slightly superior behaviour regarding this topic.

Secondly, the transformer should have a frequency flat differential to differential mixed mode S-parameter $S_{DD,21}$ to make sure that the signal is not heavily distorted at the secondary port. Additionally, the $S_{DD,21}$ curve needs to lie close to 0 dB to minimize the amount of losses in the transformer. The opposite holds for the harmonics. Since harmonics should be suppressed as much as possible, it is interesting to have a transformer that blocks most of the harmonic power presented at the primary or secondary. One of the most important exceptions to this requirement is when a square wave needs to pass the transformer without significant distortion. In that case, the transfer function of the transformer should have a flat passband which is as broad as possible. It seems that for the given transformers, the monolithic variant will be superior when it comes to the suppression of the harmonics (table 5.3). Hence the monolithic transformer will result in a more linear device.

	Monolithic	LC lattice
Efficiency (15 GHz)	76.6 %	92.9 %
Bandwidth ($|S_{11}| <$ -10 dB)	29 GHz	80.5 GHz
Surface area	5940 μm^2 (incl. matching)	$2 \times 6887 + 2 \times 270.71 \ \mu m^2$

Table 5.2: A comparison between an LC and a monolithic 2:1 transformer

$S_{DD,21}$	Monolithic	LC lattice
15 GHz	-1.18 dB	-0.32 dB
30 GHz	-1.989 dB	-0.389 dB
45 GHz	-4.751 dB	-0.482 dB

Table 5.3: Attenuation of a differential signal: A comparison between an LC and a monolithic 2:1 transformer

It is important to notice that the efficiency in table 5.2 was computed without looking at the output spectrum. Consequently this efficiency will be an upper bound on the effective efficiency since a certain fraction of the output power shall lie outside the useful frequency band. In case of the monolithic 2:1 transformer a final efficiency of 76.6 % is obtained. This value lies close to the upper limit of 77.48 % and thus it can be concluded that the matching circuit used for this device has been well designed. However the efficiency after matching is still vastly lower than for the LC transformer. The biggest issues with the latter is the fact that it is enormous and that the excessive bandwidth might result into problems (e.g. suppression of the harmonics is too low). Due to the absence of a center tap in case of an LC lattice, an additional RF choke will be needed to provide connection to a power supply or ground, deteriorating the total efficiency of the system. A final major issue in case of an LC transformer is the fact that no DC isolation is present between the circuit at the primary and the secondary and hence both circuits should be co-optimized regarding the biasing circuit, resulting in a sub-optimal solution.

1:4 Balun

In this section, 3 types of baluns will be discussed. Each of them will approximately provide a transformation of an unbalanced 100 Ω load to a 25 Ω input impedance and all of them will have the same current limit. The difference with the devices from the previous section is the fact that one of the secondary terminals will be connected to ground, resulting in a different voltage profile which shall deteriorate the behaviour of the transformer.

Monolithic 1:4 Balun: The design process for this balun is approximately the same as for the monolithic 1:2 transformer and thus it won't be discussed as extensively. In the final design, the inductance values of the primary and secondary are found to be respectively 321.1 pH and 1.246 nH, according to simulations. Since the secondary inductor is larger than in the case of the 1:2 transformer from the previous section, the interwinding capacitances will be higher resulting in a lower Q-factor. This is not only true for the secondary but also for the primary. The effective Q-factors of this device are $Q_p = 8.556$ and $Q_s = 7.867$ which is significantly lower than in the case of the stand alone inductors. The layout of this balun is approximately the same as for the 1:2 transformer. The primary will again be sandwiched between the two spirals of the secondary, hence the structure from figure 5.12 is obtained. In the ideal case, a series connection of two identical spirals result in an increase with a factor of 4 as compared to the

inductance of one spiral. Since the coupling between the two spirals of the secondary is not ideal, the two secondary spirals will be made slightly larger than the primary.

(a) Top view   (b) Tilted view
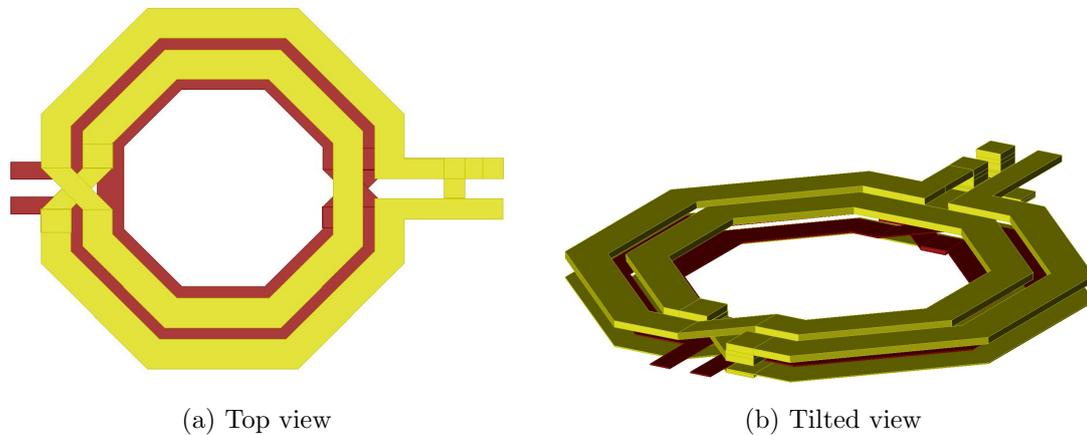
Figure 5.12: Monolithic 1:4 balun: layout

To find out the maximally achievable efficiency after matching, one should first try to determine the coupling factor. Since the coupling factor of this balun equals 0.862, the maximum efficiency will be 75.448 % which is slightly lower than for the 1:2 transformer and this is caused by the increase in interwinding capacitance. To apply power matching to this circuit, the configuration of figure 5.7 is used with C_p equal to 215 fF and C_s equal to 67.25 fF.

LC lattice 1:4 Balun: Since the design process of a 1:4 balun will exactly correspond with the design process of a 1:2 transformer, only the final values for the circuit depicted in figure 5.1 will be given in this paragraph. The inductors that are used have an inductance value of 530.5 pH, a quality factor of 13.326 and an SRF of 92 GHz. To make sure that no reactive part is seen at the input of the balun, the capacitors should be approximately equal to 216 fF.

An interesting remark concerning these LC baluns is the fact that the circuit only depends on the application frequency and the characteristic impedance of the balun. The characteristic impedance of the LC transformer/balun is defined as the geometric mean of the two port values that should be matched. For example, one can use the same circuit to match 25 Ω to 100 Ω or to match 10 Ω to 250 Ω. Of course there will be a difference in the resulting behaviour, since the latter transformation will be more narrowband than the first.

Guanella 1:4 transformer: Additionally, there are certain winding ratios that can be obtained by intelligently connecting 1:1 transformers, which is especially helpful off-chip or when large transformation ratios are desired. An example of a circuit that combines two 1:1 transformers to implement a 1:4 balun is the Guanella transformer, depicted in figure 5.13 [44].
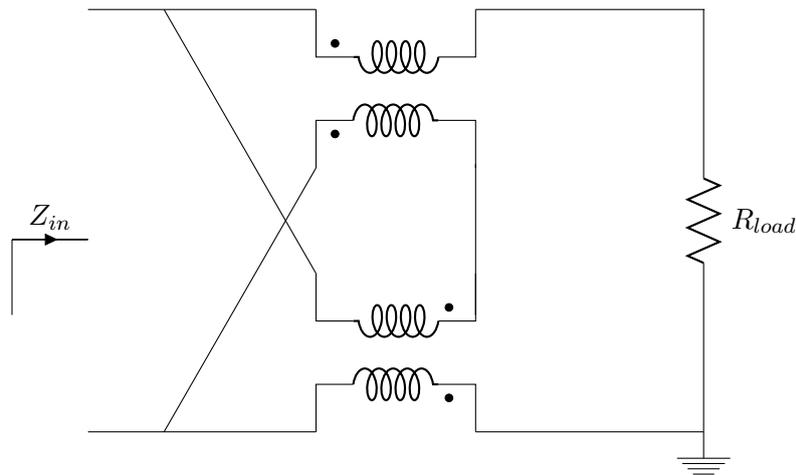
Figure 5.13: Guanella 1:4 balun

Conclusion: In this paragraph, only the monolithic transformer and LC lattice transformer will be compared (table 5.4). An extra figure of merit that is added in this table is the Common Mode Rejection Ratio (CMRR) which is a quality measure for the selectivity of the balun. The CMRR is defined as the differential mode gain relative to the common mode gain and should be as large as possible since a high differential mode gain is desirable while common mode steering should ideally have no impact on the output. The computation of the CMRR is trivial when mixed mode S parameters are derived for the balun and since the fourth port of the transformer is now shorted to ground, 3 port mixed mode S-parameters are needed [45]. It can be seen in table 5.4 that the CMRR is significantly higher in case of a monolithic transformer resulting in less distortion as compared to the ideal behaviour when a 15 GHz differential signal as well as a 15 GHz common mode signal is present at the balanced port of the balun.

When the mixed mode S-parameters are derived, it is important to take a look at the $S_{SD,21}$ curve. This curve indicates the behaviour of the circuit when a differential signal is applied at the balanced port. Since it is desirable to have as little distortion as possible, it is important to have a flat $S_{SD,21}$ curve around the operating frequency. For the two variants that are compared in table 5.4, this requirement approximately holds.

	Monolithic	LC lattice
Efficiency (15 GHz)	69.9 %	83.5 %
CMRR (15 GHz)	37.569 dB	29.587 dB
Bandwidth ($|S_{11}| <$ -10 dB)	34 GHz	15 GHz
Surface area	4924 μm^2 (incl. matching)	$2 \times 8731 + 2 \times 185.4 \ \mu m^2$

Table 5.4: A comparison between an LC and a monolithic 4:1 transformer

If the effectively obtained efficiency is compared to the upper limit which was computed earlier in this chapter (namely 75.448 %), it can be seen that the matching circuit does not suffice. There are two ways to solve this problem but each of these solutions will bring forth new problems. A first solution is to take the balun operation into account during the design process of the transformer. Special geometries can be used to make sure that the geometric, magnetic and electric center of the spirals will coincide rendering the impedance transformation more efficient. However this will come at the cost of an increase in circuit complexity, surface area, etc. and the design of those geometries is not always straightforward. A second method to improve the

power transfer between the primary and the secondary is to use a complete matching network instead of only parallel capacitors. This might not always be desirable because one might for example need a series capacitor and since a series capacitor blocks DC, the center tap can no longer be used as part of the biasing network. Hence an explicit biasing network with an RF choke will be needed when a series capacitor blocks DC connection to the center tap and this will introduce new losses which gives rise to a decrease in efficiency. Consequently, a series capacitor in a matching network will most likely do more harm than good.

When a chip has to be designed, it is important to take the required surface area into account since this will have a serious impact on the cost of the component. To design transformers, inductors will be needed as well as capacitors. The latter is required for decent power matching or as a part of the transformer itself (in case of an LC lattice). When special attention is given to the desired surface area, one can easily see that the inductors will have a dominant impact on the total surface area of the device. In the case of an LC lattice this means that higher characteristic impedances of the transformation circuit (i.e. $\sqrt{R_1 R_2}$) require more space due to the fact that higher inductance values are needed. Consequently this 25-to-100 Ω LC structure takes in significantly more space than the 25-to-50 Ω variant as can be seen by comparing tables 5.2 and 5.4.

The 1:4 LC balun results in a lower bandwidth as compared to the 1:2 LC transformer as well as a larger required surface area. However, these problems are also present when a 1:4 LC transformer is compared to the 1:2 LC transformer. The biggest issue of using an LC lattice as a balun is the fact that both branches see a different circuit at frequencies different from the operating frequency when a load is connected to the secondary port. When one branch connects to the load via a low pass circuit, the other one will be connected via a high pass circuit (figure 5.2). Hence the trend of the impedances will be dual for the two primary terminals. The monolithic 1:4 transformer will also suffer from the balun operation (unless special geometries are used) because the geometric, magnetic and electric center no longer overlap resulting in a less efficient structure. Nevertheless, the impedances seen at both branches (P+ and P-) remain approximately identical to each other in the case of a monolithic transformer, even for frequencies that differ from the design frequency.

A final remark concerning the dependence of the impedance on the frequency is the fact that larger transformation ratios will inherently result in a more narrowband circuit. This will give rise to a steeper evolution of the impedances in function of the frequency as compared to the case of the 1:2 transformers discussed in an earlier section.

5.6.4   Final realisations

Combined device: transformer and RF choke

Last chapter, different ways to implement an RF choke were discussed. It was found that the eight-shaped RF choke results in the highest efficiency of the global system. Another component needed in the final stage is a 25:25 Ω balun. Since the RF choke as well as the balun induces losses, the separate implementation of both will only result in a sub-optimal solution. A better way to handle this problem is to use the center tap of the primary winding to connect to the power supply. This will not only result in less losses but will also save surface area.

The design process for the balun needed in the final stage is slightly different as compared to the design process discussed in the previous sections. The goal of this design consists of two parts. Firstly, a highly efficient transformer operation is needed. Secondly, the circuit around

the balun has a new purpose. Instead of power matching, the circuit should mimic an equivalent circuit at the operating frequency.

Equations 5.8 and 5.9 respectively indicate that L_p should approximate 270 pH while L_s should be approximately 330 pH. This is caused by the fact that a transformation is desired where the primary should look like 2 times 108 pH while the secondary needs to be designed to give an optimal transformation to 25 Ω. The constant mentioned in formula 5.8 will then take into account the nonidealities of the transformer resulting in the fact that L_p and L_s should be respectively 270 pH and 331.6 pH instead of 216 pH and 265.3 pH.

An implicit specification on the balun caused by the fact that power is coupled in via the center tap is the fact that the tracks should be wide enough to accommodate high DC currents. An extra advantage of using high track widths is that it results in a low DC resistance of the balun which effectively increases the efficiency of the final configuration. Earlier in this chapter, the concept of autoshielding was defined as the technique where the highest inductance lies on the upper most metal layers since the highest voltages will be induced in this inductor resulting in potentially higher parasitic effects. In this design it is however advantageous to implement the primary on the upper most metal layer. The reason for this is the fact that this primary will be wider, resulting in higher parasitic capacitances. Simulations prove that implementing the primary on the uppermost metal layer and the secondary on the metal layer just below it will indeed give rise to a more efficient balun than when the exact opposite is done (the maximum obtainable efficiency decreases from 79.3 % to 77.1 % by switching the primary and secondary).

An initial design was done in a straightforward way where 2 turns were used to implement each inductor. This however resulted in low Q-factors since there was a lot of metal present close to the center of the spiral, which is the place where the magnetic fields are highest. Hence in a next design iteration a one-turn spiral was used although the simulation model indicated that the desired inductance values are not feasible with one turn. However, due to the fact that the models are derived for a low ohmic substrate and we have chosen to use a high ohmic substrate, it becomes possible to realise these values with just one turn. Unfortunately, these one-turn inductors will take in a lot more surface area. As a consequence of the fact that the substrate losses decrease when using a high ohmic substrate, the optimal track width will be a lot larger resulting in lower metal losses. Since only one turn is needed, a high track width can be used without having an SRF that is too low for the application. Hence high Q-factors are possible which will result in more efficient baluns (figure 5.8). Only one major problem remains which is caused by the high track widths that are being used. While those wide tracks can potentially result in very efficient baluns, major overlap areas will quickly deteriorate the Q-factor of the inductors due to large interwinding capacitances. A solution proposed earlier in this chapter to overcome this problem is to shift the primary relative to the secondary to strongly decrease the interwinding capacitances. Since a shift results in a larger surface area and a decrease in the coupling factor of the transformer, it is important not to exaggerate with this shift. When shifting the primary with regard to the secondary, it is also essential to take into account that both branches should behave the same to get a well working differential circuit and hence the shift is done by moving the primary along the symmetry line.

The final layout and dimensions are respectively given in figure 5.14 and table 5.5. The biggest disadvantage of using a one turn spiral instead of a two turn spiral to implement this balun is the major increase in surface area. But since the specification for our amplifier is concerning efficiency rather than cost of the final chip, this one turn solution is to be preferred.

(a) Top view   (b) Tilted view
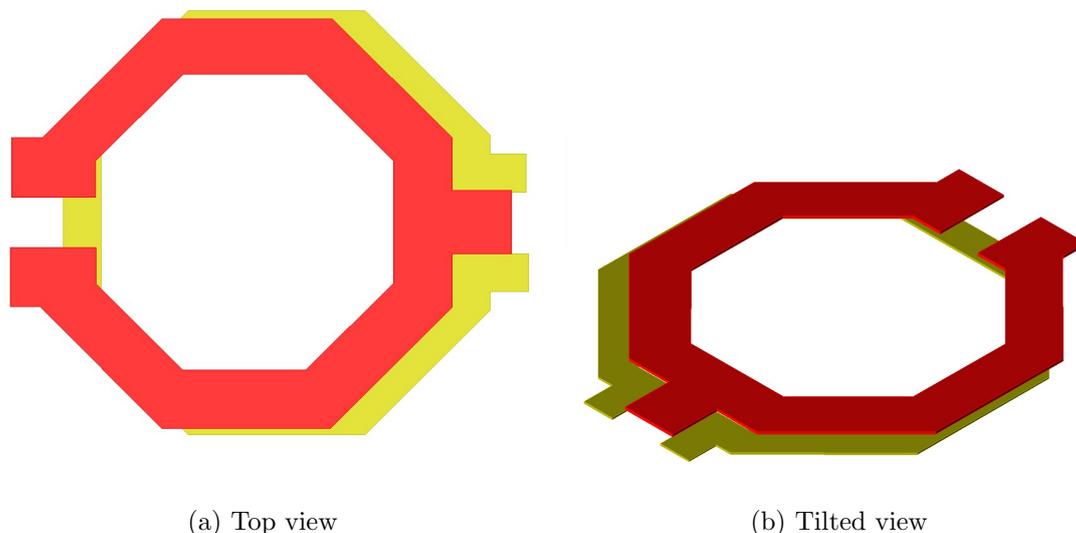
Figure 5.14: PA output balun: layout

	primary	secondary
center tap	yes	no
track width	27.2 μm	17.33 μm
inner radius	69.05 μm	81.47 μm
number of turns	1	1
surface area	41580 μm^2	

Table 5.5: PA output balun: dimensions

These track widths seem enormous but since only one turn is used for each inductor, the parasitic coupling to the substrate remains small and that is why a large track width can be used to obtain an optimal trade-off between metal losses and substrate losses at the operating frequency.

For the final layout, the effective inductance values are found to be $L_p = 282.2$ pH and $L_s = 327.3$ pH. Due to the fact that a single turn is used, combined with the shifting of the primary relative to the secondary, a combination of a relatively high coupling (k = 0.827) and high Q-factors ($Q_p = 12.296$ and $Q_s = 8.567$) is obtained. This gives rise to a very efficient balun, especially when compared to what is maximally achievable with the straightforward two-turn implementation. This final implementation results in a η_{max} of 79.3 % while the initial design had a much lower upper bound of the efficiency, namely 70 %.

When the center tap in a transformer or balun is used, it is important to have a low ohmic DC connection to the terminals of the device. For the final implementation, this DC resistance approximates 250 mΩ for each branch. It is important that this value is low since any losses in this DC connection will deteriorate the efficiency of the total system.

It has already been mentioned that the final design has lots of advantages as compared to the initial two turn design (the efficiency will be higher, the DC resistance will be lower, etc.). An additional advantage of this new implementation is the fact that it is way more symmetric which will result in a better approximation of the desired behaviour of the balun. This improvement in symmetry results from the absence of a cross-over part, which is inherently asymmetric. The

shift of the primary relative to secondary might induce asymmetry but by paying attention on how the shift is done, no problems concerning asymmetry arise from this shift.

However, this approximately ideal symmetry on the primary terminals disappears when balun operation is forced on the transformer. Due to an asymmetric voltage profile forced by the unbalanced operation, the parasitics for both halves of the primary will no longer be equivalent and thus the impedance seen at one terminal of the primary will be different from the impedance seen at the other terminal of the primary.

At the beginning of this section, it was stated that a certain equivalent circuit should be mimicked at 15 GHz instead of applying power matching to the balun. The circuit that is used is depicted in figure 5.15 together with the circuit that the final EF_2 stage of the amplifier needs to see between the primary terminals of the balun. In an initial design of the circuit, only C_1 was used to obtain the correct operation. However, since the balun operation makes the device asymmetric a better solution may be obtained when the desired capacitor is connected to the other secondary terminal. By using an Electromagnetic (EM) model for the balun, it is empirically found that the desired operation can be obtained when C_1 is replaced by a short and a capacitor of approximately 928 fF is used as C_2.

(a) Effective circuit at 15 GHz   (b) Apparent circuit at 15 GHz
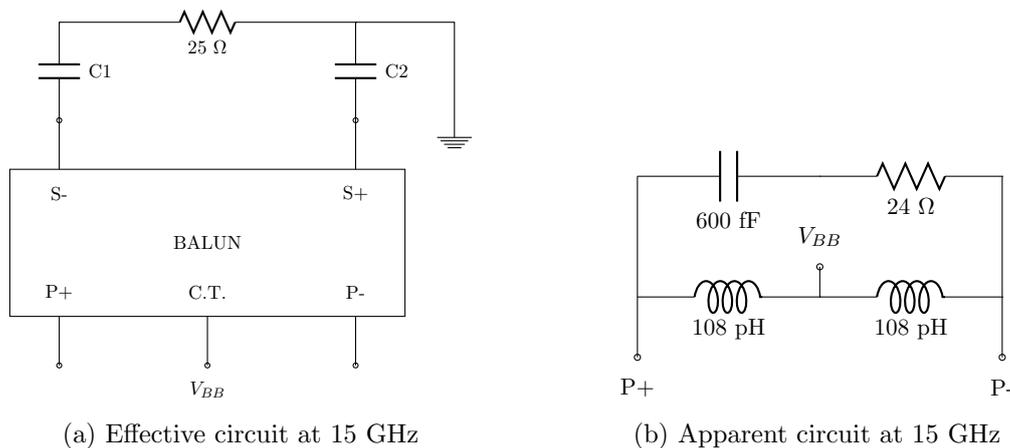
Figure 5.15: Desired output balun operation

When both circuits from figure 5.15 are simulated, the impedances can be checked to find out if the 15 GHz behaviour is indeed equivalent in both cases and if the impedances seen at both the primary terminals are approximately equal. In case of the apparent circuit the impedances at both the primary terminals are of course exactly the same, namely 8.528+j×9.229 Ω. The realised circuit will however result in small differences (caused by the balun operation) for the impedances seen at the P+ node (8.759+j×9.119 Ω) and the P- node (8.264+j×9.098 Ω). Nevertheless, they reasonably approximate the impedance of the apparent circuit.

Power combiner: Balun

The last component that will be discussed in this section is the power combiner at the output of the system. The purpose of this 50-to-50 Ω balun is to provide the conversion between the differential signals at the input of the balun to a single ended signal into an unbalanced load. While the implementation of the previous component was slightly different from what is normally done, the power combiner is designed in the usual way (i.e. power matching is required). When formulae 5.8 and 5.9 are used, the required inductance values are found. After two iterations, it is clear that the inductors will have a high coupling factor resulting in the fact that the prefactor

in formula 5.8 approximates 1.1. Hence, the L_p value should be made equal to 583.57 pH. This also holds true for the value of L_s since a 1:1 transformation is needed.

In contrast to the output balun of the EF_2 end stage, no wide tracks are needed for this transformer. This is caused by the fact that the AC current will have an amplitude of 63 mA if 20 dBm is injected into a 50 Ω load. Consequently the metal track should only have a width of 3.9375 μm.

The final layout and dimensions of the power combiner are mentioned in figure 5.16 and table 5.6. Simulations provide inductance values of 571.0 pH and 584.5 pH for L_p and L_s respectively. Since the primary was chosen to be implemented on top of the secondary, the Q-factor is higher in case of the primary ($Q_p = 6.362$ while $Q_s = 5.648$).

(a) Top view   (b) Tilted view
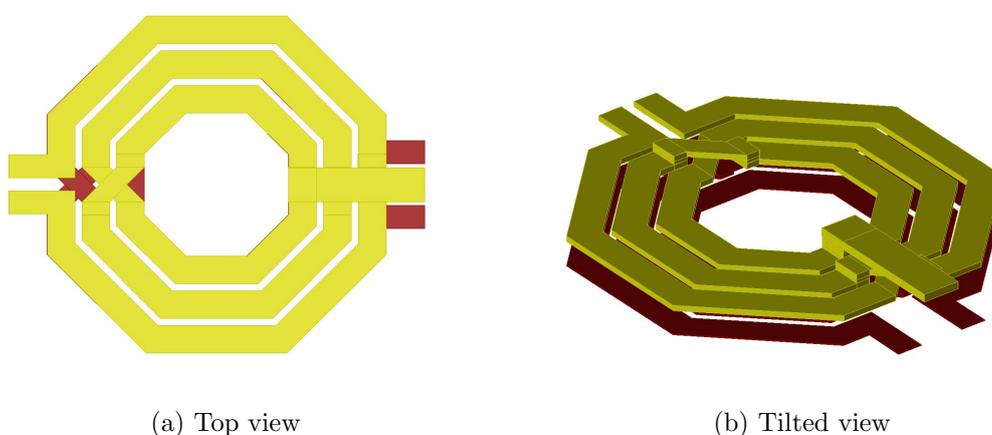
Figure 5.16: Power combiner: layout

	primary	secondary
center tap	yes	no
track width	5.35 μm	
spacing	1.5 μm	
inner radius	24 μm	
number of turns	3	
surface area	4309.9 μm^2	

Table 5.6: Power combiner: dimensions

According to literature [41], the efficiency of the matched balun will be upper bounded by 70.23% (since k equals 0.939). However, by using the circuit depicted in figure 5.7, an efficiency is found that is slightly higher (75.3 %). This can be caused by two important reasons. Firstly, a different matching circuit is used in the cited paper [41]. Secondly, the formulas are derived for a simple transformer model which might not suffice in this case.

To obtain 75.3 %, the parallel capacitors were chosen to be $C_p = 135.6$ fF and $C_s = 47.4$ fF. These values were found by matching the transformer while taking into account that the S+ terminal is to be connected to ground. When the matching is done for a regular transformer operation and the S+ terminal of the matched transformer is only afterwards connected to ground, an inferior matching circuit is obtained. For the power combiner derived in this section, an efficiency loss of 1.1 % will result when the balun operation is not considered during the matching process. When the values for the matching capacitors are compared, C_p is found to be

almost 3 times larger than C_s. In the matching of a regular transformer this would be odd since the values for L_p and L_s are comparable. This phenomenon is a direct result from including the balun operation in the design process of the matching circuit.

An interesting conclusion that can be drawn from the simulation results is the fact that the harmonics will be suppressed relative to the desired signal when present at the input of the power combiner. This can be seen by comparing the efficiency of the power combiner at the fundamental frequency (i.e. 75.3 %) with the efficiency at the harmonics ($\eta_{30G} = 45.3\%$, $\eta_{45G} = 37.8\%$, $\eta_{60G} = 35.4\%$, etc.). When the bandwidth is defined as the frequency range for which $|S_{11}| < $ -10 dB holds, a bandwidth of 10 GHz is found for this power combiner. While the bandwidth of the transformer is small enough to result in sufficient suppression of the harmonics, it is also large enough to provide a flat response to the relevant spectrum (i.e. the efficiency maximally decreases with 0.83 % in the frequency range starting from 14 GHz up to 16 GHz).

(a) Differential signals at the balanced port   (b) Common mode signals at the balanced port
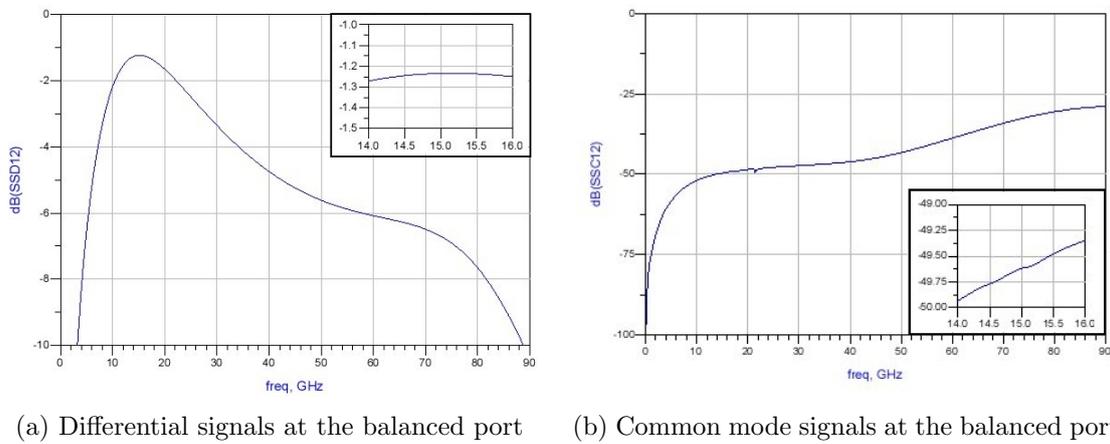
Figure 5.17: Mixed mode S-parameters: output combiner

From the mixed mode S-parameters shown in figure 5.17 similar conclusions can be drawn. The response at the unbalanced port resulting from a differential input at the balanced port will suffer least from attenuation when a 15 GHz signal is applied at the input. Additionally, it can be seen that the response is sufficiently flat for the frequency range of interest and that the harmonics are reasonably suppressed. An interesting observation that can be seen in figure 5.17b is the fact that common mode signals at the input will only have a negligible influence on the signal appearing at the unbalanced load.

		14 GHz	15 GHz	16 GHz
$\Re(Z_{P+})$		23.414	23.742	23.140
$\Re(Z_{P-})$		24.148	24.448	23.766
Desired value			25	

(a)   (b)
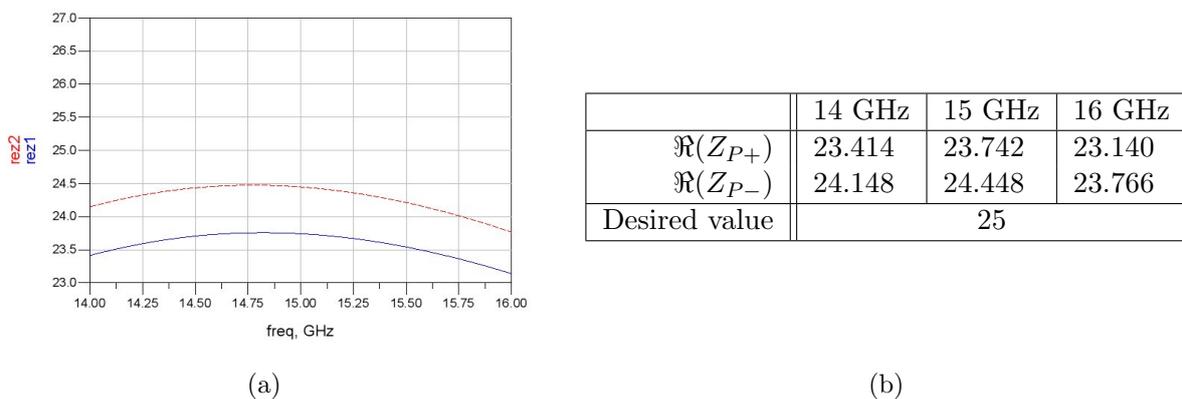
Figure 5.18: The real part of the impedances seen at the primary terminals of the power combiner
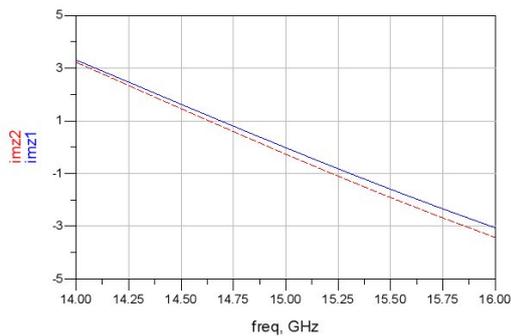
	14 GHz	15 GHz	16 GHz
$\Im(Z_{P+})$	3.300	-0.021	-3.055
$\Im(Z_{P-})$	3.213	-0.266	-3.424
Desired value		0	

(a)   (b)

Figure 5.19: The imaginary part of the impedances seen at the primary terminals of the power combiner

Due to the balun operation (i.e. asymmetric voltage profile) and the asymmetric parts of the transformer (i.e. the crossovers), a discrepancy will occur between the impedances of the primary terminals (depicted in figures 5.18 and 5.19). Since a 50:50 Ω balun is used together with a 50 Ω load resistor, the goal is to see approximately 25 Ω without a reactive part at each primary terminal (especially at the operating frequency). Since L_p and L_s do not exactly match and the prefactor is not entirely correct, obtaining this 25+j×0 Ω at each of the primary terminals is unfeasible. Nevertheless, for the given power combiner at 15 GHz, a resistive part is seen that is reasonably close to 25 Ω while the reactive part approximates 0.

Next to the 50:50 Ω balun needed at the output, a power splitter is required at the input with exactly the same specifications but for which the current requirements are even lower. Hence this power combiner is reused in our design to split the useful signal into differential components at the input of the outphasing circuitry.

Part IV

Power amplifiers

Chapter 6

Power amplifier

6.1 Ideal PA class overview

The main goal in the design of a power amplifier is to efficiently produce a "clean" output spectrum with sufficient power at the fundamental design frequency. The drain efficiency η_d defines how well DC power is converted into RF power at the fundamental frequency, and is independent of the input power and the PA gain ([1], p. 22). The overall efficiency takes the DC power of the drivers into account:

$$\eta_d = \frac{P_{out,fundamental}}{P_{DC,PA}}; \eta_{d,oa} = \frac{P_{out,fundamental}}{P_{DC,PA} + P_{DC,drivers}} \tag{6.1}$$

The power added efficiency (PAE) does take the input power into account. With G_P as the power gain:

$$PAE = \frac{P_{out,fundamental} - P_{in,total}}{P_{DC,PA}} = \eta_d \left(1 - \frac{1}{G_P}\right) \tag{6.2}$$

The power added efficiency (PAE) does take the input power into account. With G_P as the power gain:

$$PAE_{oa} = \frac{P_{out,fundamental} - P_{in,total}}{P_{DC,PA} + P_{DC,drivers}} = \eta_{d,oa} \left(1 - \frac{1}{G_P}\right) \tag{6.3}$$

If the input signal is not sinusoidal, $P_{in,total}$ also includes the power at the input harmonics. When G_P is high enough, e.g. 20 dB, the PAE is approximately equal to the drain efficiency: $PAE = 0.99\eta_d \approx \eta_d$. Power amplifiers are designed to deal with large signals. Small-signal parameters do not apply because the device is driven into a strongly non-linear working region and the input signals are large. This is not necessarily a disadvantage: to maximize the PA efficiency, power dissipation in the active components has to be minimized; overlap in time of current through and voltage over the active components is avoided. Naturally, this leads to the situation where either one will tend to be shaped into a square-wave; which ideally consists of an infinite number of tuned odd harmonics of the fundamental frequency. We consider three main causes of harmonic currents: a reduction of conduction angle, transconductance non-linearity and the transistor "knee region".

- Even with a perfectly linear transistor, harmonic currents are still generated if the conduction angle is smaller than 2π. The drain current becomes a clipped sine with a DC-offset.

Because the waveforms are periodic, Fourier integrals can be used to calculate the coefficients (amplitude and phase) of each harmonic. The result of this analysis is given in figure 6.2 (from [4], p. 40-42). In class B, at a conduction angle π, the amplitude of the fundamental is the same as in class A. In between class A and B, $|I_{f1}|$ increases, resulting in more output power. As the conduction angle decreases below π, both $|I_{f1}|$ and I_{DC} decrease and the current-voltage overlap decreases (cfr. figure 6.3, from [1], p.34), so the efficiency increases at the cost of a lower output power. If a larger output power is desired when the conduction angle is below 50%, the peak current I_{max} has to increase. Figure 6.3 demonstrates the efficiency increase: if α is reduced and $|I_{f,1}|$ is kept constant by changing I_{max} and $Z_{f_1} = R_L$ is kept constant, or if $|I_{f,1}|$ is to change but R_L is adjusted accordingly, then the voltage waveform will not change but the voltage-current overlap decreases, improving the efficiency.
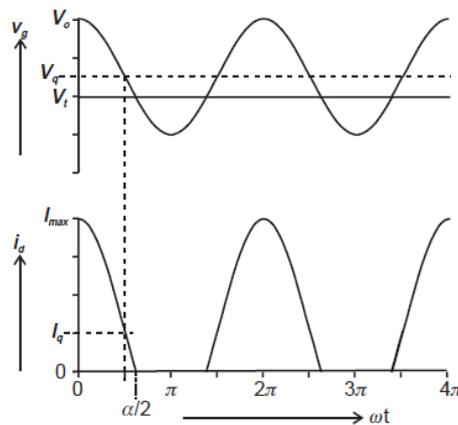
Figure 6.1: Reduction of conduction angle.
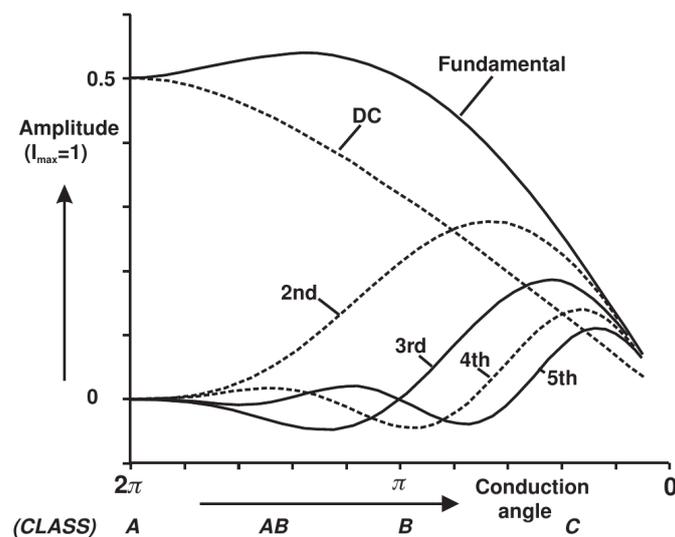
Figure 6.2: Amplitude of the harmonic current components as a function of the conduction angle, when I_{max} is constant.
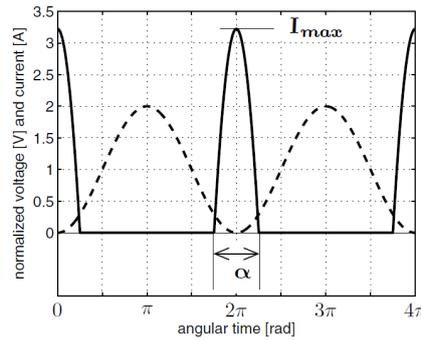
Figure 6.3: Current and voltage waveforms for a reduced conduction angle.

- Secondly, the transistor itself is not perfectly linear: $I = g_m V_{in}$ does not apply. If the non-linearity is not too strong, a power series could be used, and the harmonics of the terms in V_{in} appear directly: $I = g_{m,1} V_{in} + g_{m,2} V_{in}^2 + g_{m,3} V_{in}^3...$

- Another major cause of harmonics in the output current is the transistor "knee region": when the voltage drops below the saturation voltage V_{sat}, the transistor is no longer in the "current source region", where the output impedance is (relatively) large. I_{out} now strongly depends on the voltage over the transistor and will be forced low when the voltage is low. This distorts the clipped DC-offset sine of figure 6.1 and can introduce even more harmonics in the output current. This effect is especially strong in a MOSFET: when the V_{GS} is high, V_{DS} becomes low but $V_{DS,sat}$ is high, so we end up in the "knee region" quite often. This is not a disadvantage if we want to create harmonic currents on purpose. The resistive part of the small-signal output impedance is proportional to $r_0 \propto \frac{LV_E}{I_{DS}}$, with V_E equal to the Early voltage, which is technology-dependent. For the shortest-length transistors, the 45 nm - MOSFETs we intend to use, r_0 can be rather small, depending on $I_{DS,DC}$. In this case, I_{DS} will depend on V_{DS} even in the pentode region, which also contributes some non-linearity.
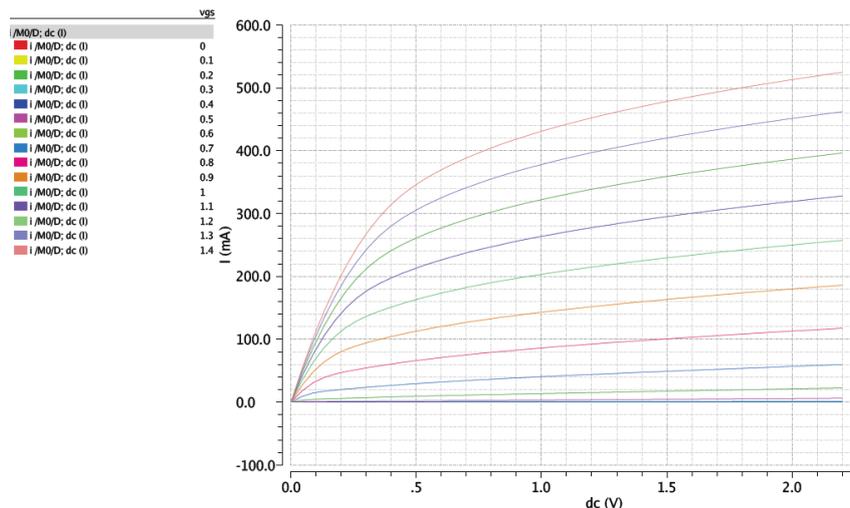
Figure 6.4: DC I-V characteristic of 1.1 V NMOS with $\frac{W}{L} = \frac{2\mu m}{45nm}$ and multiplier 300.

The desired amplitude and phase of the harmonics in the voltage over the device are obtained by terminating the harmonic currents with the correct impedance. This way, the essence of efficient PA (output stage) design is the shaping of the active device waveforms by introducing

and terminating harmonics correctly. A second aspect is filtering: the harmonics are necessary to shape the waveforms but they are very undesired at the output; where they are a distortion. All output power in the harmonics originally comes from the supply and is wasted in the load, which will decrease the efficiency.

Harmonic tuning does not only result in a larger efficiency, the output power can increase as well: if the voltage over the transistor is limited to V_{max} and no harmonic tuning is done, the maximal voltage amplitude at the first harmonic is $\frac{V_{max}}{2}$. This results in the maximal class A efficiency of 50 %. If the voltage over the device is shaped into a square wave of the same amplitude, the fundamental component of this sine is increased to $\frac{4}{\pi}\frac{V_{max}}{2}$, so $P_{out,fundamental}$ can potentially be increased by $\left(\frac{4}{\pi}\right)^2$.

The classical PA classes (A-... -F) are summarized in figures 6.5 and 6.6 (from [1], p. 58-59). For "reliable operation", it is assumed in [1] that hot carrier injection is the most dominant breakdown factor. Consequently, the maximal allowed voltage over the transistor whenever it conducts current, is limited to the nominal supply voltage of a given technology. When it does not conduct current, it is limited by a breakdown voltage: $|V_{gd}| < 2V_{dd,nom}$ and $V_{ds} < (2...3)V_{dd,nom}$. Four main types exist: the reduced-conduction angle amplifiers (A,B,C) with or without saturation, the tuned amplifiers (F, F^{-1}) and switching amplifiers (E,F) and combinations of tuning and switching (class-EF). The class E-PA combines 100% theoretical efficiency with relatively large output power, except when the main limitation is V_{DS} (or V_{GD}). The class F -PA delivers slightly less power at a lower theoretical efficiency.

Class	A	B	F3	Sat. A
peak efficiency	50.0%	78.5%	88.4%	81.0%
normalized output power [W] ($V_{DD} = 1V$ and $R_L = 1\Omega$)	0.5000	0.5000	0.6328	0.8106
maximum output power [W] ($v_{DS,max} = 1V$ and $R_L = 1\Omega$)	0.1250	0.1250	0.1582	0.2026
maximum output power [W] (reliable operation,$V_{DD,nom} = 1V$ and $R_L = 1\Omega$)	0.1250	0.5000	0.6328	0.2026

Figure 6.5: Overview of the typical PA classes, part 1.

Class	D	E
peak efficiency	100%	100%
normalized output power [W] ($V_{DD} = 1V$ and $R_L = 1\Omega$)	0.8106	0.5768
maximum output power [W] ($v_{DS,max} = 1V$ and $R_L = 1\Omega$)	0.2026	0.0455
maximum output power [W] (reliable operation,$V_{DD,nom} = 1V$ and $R_L = 1\Omega$)	0.2026	0.1820...0.4095

Figure 6.6: Overview of the typical PA classes, part 2.

included harmonic		3	5	∞
Class	B	F3	F5	D
peak efficiency	78.5%	88.4%	92.0%	100%
normalized output power [W] ($V_{DD} = 1V$ and $R_L = 1\Omega$)	0.5000	0.6328	0.6866	0.8106
maximum output power [W] ($v_{DS,max} = 1V$ and $R_L = 1\Omega$)	0.1250	0.1582	0.1717	0.2026

Figure 6.7: Effect of harmonic tuning on the class F performance.

6.2   Ideal class F

Looking at figure 6.2, we see that there are no odd harmonic currents in pure class-B operation. In the ideal class-F theory, the class B PA is taken as a starting point and this problem is solved by presenting these "zero" currents with an "infinite" impedance, resulting in a finite harmonic voltage component, which should shape the drain voltage into a square wave. This is not the case in practice: odd harmonic currents will exist, but might be quite weak, depending on the transistor linearity and "knee region". The impedance that needs to be presented to these non-zero currents is therefore also not infinite, but this is not an issue because the quality factors of practical resonators are finite.
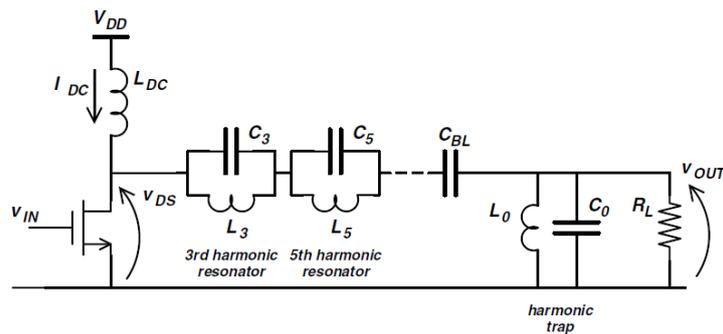
Figure 6.8: Theoretical class F with a finite number of resonators.
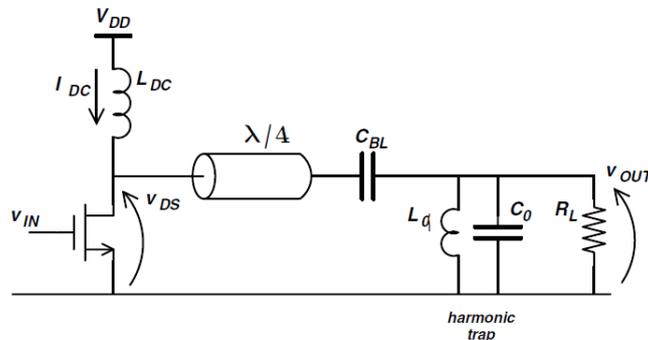
Figure 6.9: Theoretical class F with a $\frac{\lambda}{4}$-TL.

Two classical ideal class-F designs are given in figures 6.8 and 6.9 ([1], p. 44-47). The DC-current is provided through an "infinite" RF-choke, which is an "open" in AC. In figure 6.8, a finite number of harmonic resonators is added to terminate some of the lower odd harmonics with an "open". In figure 6.9, all odd harmonics are terminated with an open: the parallel LC-filter at the output shorts all odd harmonics except f_1, the $\frac{\lambda}{4}$-TL converts this short into an open at the drain of the transistor. The even harmonics are shorted as well at the output, and an ideal $\frac{\lambda}{4}$-line does nothing at the even harmonics, were its length is a multiple of $\frac{\lambda}{2}$, so all even harmonics are shorted. In theory, V_{DS} over the transistor could be a square-wave and I_{DS} is a positive-half sine because the conduction angle is 50 %. A real transmission line will suffer from losses, so the even harmonics will not be perfectly shorted and the odd harmonics are certainly not be presented with an "open". The main disadvantage with this theoretical design is the output capacitance of the transistor, which is not integrated in the network that shapes the drain voltage. This capacitance partially shorts the harmonic currents and reduces

the efficiency considerably. Therefore, even the class F design with a $\frac{\lambda}{4}$-TL will not deliver 100 % efficiency in practice, even though an infinite number of harmonics have been terminated.

The infinite RF choke is also unrealisable, and certainly in CMOS, where inductors are very lossy and large and inductances are limited. The RF-choke can be replaced with a $\frac{\lambda}{4}$-TL to the supply. An "infinite" impedance is presented at the fundamental and at the odd harmonics, and all even harmonics are shorted directly. A filter is still needed in parallel with the load to prevent the odd harmonics from reaching the output, and a DC block needs to be integrated as well.

6.3 Ideal class E
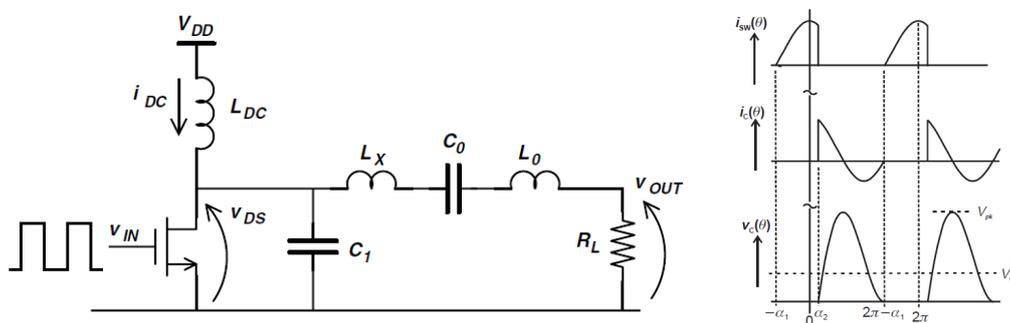
Figure 6.10: Theoretical class-E.

The ideal class-E schematic ([1], p. 51), with a simplified analysis ([4], p. 187) is given in figure 6.10. It is assumed that the drive signal is sufficiently strong to switch the transistor completely. The classical analysis of the Zero-Voltage Switching (ZVS)-class E assumes that only current at the fundamental frequency can flow through the ideal LC-filter to the output load. Only DC-current can flow through the infinite DC-feed. Consequently, the sum of the current through the switch and the current through the parallel capacitance must be a DC-offset sine. When the switch is off, no current can flow through it: the DC-offset sine is integrated on the parallel capacitance and a V_{DS}-peak arises. With the correct capacitance and the correct phase shift in the fundamental current, V_{DS} will return to zero before the switch is switched on. V_{DS} must be continuous, so it is zero when the switch is switched off as well. I_{DS} and V_{DS} do not overlap in time, and 100 % efficiency is obtained. The transistor is simplified as an ideal switch with an R_{on} in series and a parasitic parallel capacitance. Another condition is imposed: $\frac{dV_{DS}}{dt}(t_{ON}) = 0$; this makes the solution less sensitive to component tolerances: no current flows through the total parallel capacitance at $t = t_{ON}$, the current through the transistor cannot increase instantaneously. The Zero-Current Switching (ZCS)-class E is a dual variant, where the roles of current and voltage are interchanged.

6.4 PA distortion: AM-AM, AM-PM, PM-AM, PM-PM

The PA behaviour is summarized in 4 characteristics:

- AM-AM: this characteristic describes the amplitude linearity of the power amplifier. At high input levels, the gain will reduce, leading to e.g. an 1 dB-compression point. The class A PA is clearly the most linear type, but the linearity comes at a very high cost:

the theoretical maximal efficiency is only 50%. The class AB-PA's combine a higher gain with reasonable linearity. Some non-linearity occurs because the effective conduction angle depends on the input drive level (cfr. figure 6.3); and figure 6.2 clearly demonstrates that the fundamental current component is a non-linear function of the conduction angle.

- AM-PM describes the undesired phase deviations on the output caused by amplitude variations at the input. Again, the best AM-PM performance is expected from a class A PA, as demonstrated in figure 6.11 (from [4], p. 253). In theory, the AM-PM characteristic is not important here because the amplitude of the outphasing signals should be constant. In practice, the amplitude might fluctuate slightly due to bandwidth limitations, which suppress the high-frequency components needed to keep the amplitude constant, as described in section 2.4.5. If this case, the output phase might be affected. PM-AM describes the dual effect: output amplitude modulation caused by input phase modulation. AM-PM also suffers from memory effects, such as e.g. temperature dependence, which are difficult to measure because they are not captured by a DC-sweep. Feedforward from the driver stage through the gate-drain capacitance of the output stage MOSFETS is the dominant contribution in AM-AM and AM-PM ([1], p. 150-151).
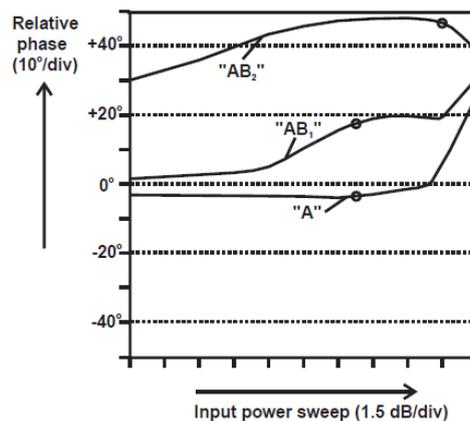
Figure 6.11: Example of measured AM-PM data.

- PM-PM describes the input-dependent phase shift that is introduced at the PA output. Because the outphasing signals contain only phase modulation, this characteristic is very important. An example of a possible PM-PM-characteristic was given in section 2.7. On the other hand, for systems with a small bandwidth compared to the carrier frequency, the distortion due to PM-PM might be less pronounced because of the broadening introduced by the finite quality factors of the resonators; resulting in an approximately linear phase characteristic, constant group delay and good transmission of phase modulation ([1], p. 150).

6.5 Selection of a PA class

The constant envelope bandpass signals contain only phase modulation. This allows the use of a power amplifier in which the relation of the amplitude of the (e.g. non-sinusoidal) input signal and the output sine wave is non-linear. Two classical PA types seem suited:

- The class-F PA is the most attractive reduced-conduction angle PA: high power is combined with high efficiency, at the cost of increased complexity because of the necessity to

terminate multiple harmonics correctly. In the theoretical class-F design, the parasitic C_{ds} is not included into this harmonic termination network, but has to be done in practice to obtain a higher efficiency.

- The class-D PA is a digital inverter with an analog output filter. Due to the presence of a PMOS, the total area is large and the total parasitic capacitance at the drain node is very large. Therefore, we did not consider this type of amplifier as an output stage. When the invertor is biased in the active transition region, it can be used as a simple non-linear amplifier at the start of a driving stage, cfr. section 6.10.

- The class-E PA also has 100% theoretical efficiency, and a major advantage compared to the class D PA: the total drain capacitance is no longer entirely parasitic, but it is directly integrated into the network of passive components that shape the waveforms. This makes the class-E PA very suited for integration in CMOS. A major disadvantage is the large V_{DS}-peak that occurs when the transistor is off, and can rise to $\approx 3.56 V_{DD}$. In a practical, non-ideal design, his peak magnitude will be reduced, but this still remains a problem, especially with decreasing channel lengths and breakdown voltages. Cascoded class-E PA's are a popular solution: because of the increased voltage tolerance, an inherently lossy impedance reduction is not necessary to obtain a given output power without exceeding the breakdown voltages, resulting in a higher efficiency ([24]).

- The class-EF family of PA's aims to combine the most desirable features of both class-E and class-F, and is further discussed in section sec:idealclassEF.

6.6   Designed class-F

6.6.1   DC I-V characteristic

Based on the DC I-V characteristic as in figure 6.4, we can approximate the optimal R_L and the optimal $V_{DS,bias}$ to extract maximal power from the MOSFET. When $R_L = R_{L,opt}$, a given current sweep leads to a voltage sweep $R_L I$, which should ideally be equal to $[0, V_{ds,breakdown}[$ to obtain maximal power. Plotting I_{DS} vs. V_{DS} for an entire conduction period results in a dynamic load-line. For the class-A PA, the entire load-line is in the plot. For a class-B PA, the voltage swing above $V_{DS,bias}$ occurs for "zero" current. The loadline becomes more distorted in a class-F PA with a square V_{DS}. The calculation of $R_{L,opt}$ based on the DC I-V-curve was done initially but the results are not accurate, especially in a MOSFET, due to the behavior in the large "knee region". Demanding that the transistor is continuously in the pentode region results in a very inefficient design. The loadline for the designed class F-PA with a 1.8V NMOS, a harmonic termination network based on a 25 Ω-load and with ideal inductors is plotted in figure 6.12.
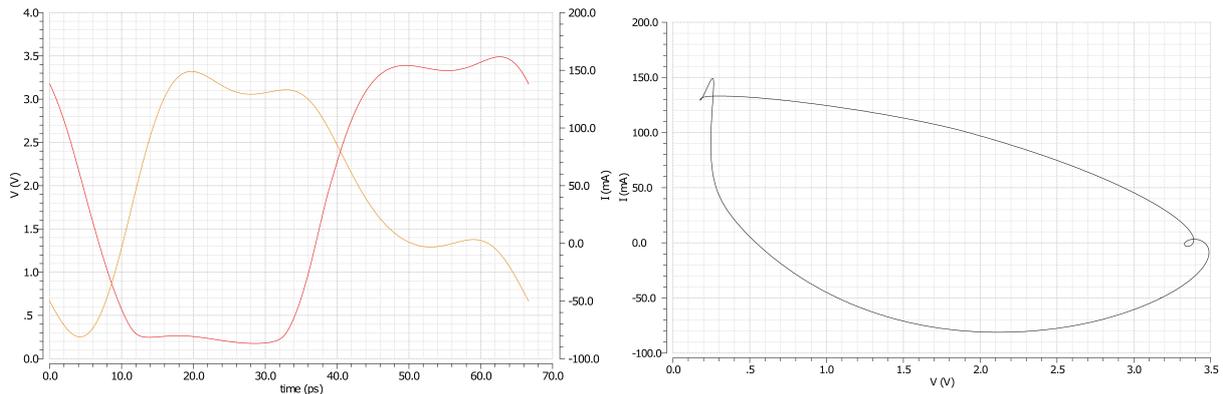
Figure 6.12: I_{DS}, V_{DS} and the dynamic loadline for the ideal class F.

In reality, the transistor cannot be separated from its parasitic drain capacitance (to ground) C_{dd}. For discrete transistors, parasitic bondwire and interconnection inductance needs to be brought into account as well. The optimal impedance which can be presented externally to the device is then no longer resistive. If C_{par} and L_{par} are known, the optimal load should be such to present $R_{L,opt}$ to the idealized transistor, which is viewed as an ideal current generator, separated from C_{par} and its output impedance r_0. $Z_{L,opt,external}$ can be determined with a load-pull measurement: based on an estimate $Z_{L,opt,external,estimate}$, several impedances corresponding with adjacent reflection coefficients Γ_L are swept, and the output power is measured, resulting in load-pull contours, of which the shape can be predicted based on the estimates of C_{par} and L_{par}, as in figure 6.13 ([4], p. 30).
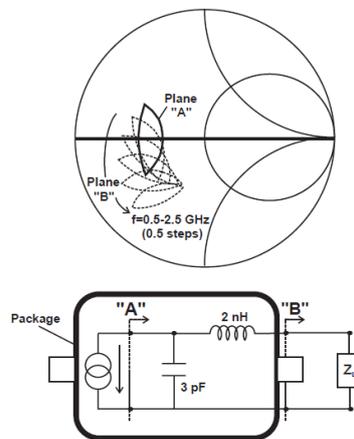
Figure 6.13: Load pull-contour example.

When C_{par} and L_{par} are known, they are included in the harmonic termination network ([46]). Initially, the load is ignored: L_1, L_{par} and C_1 short the second harmonic; and C_{par}, L_1, L_{par} and C_1 form an open at the third harmonic. The load is added through an LC-lowpass matching network, which simultaneously provides filtering an a high-impedance at the third harmonic to prevent that the "open" that was just created is partially shorted.

$$C_1 = \frac{5}{4}C_{par} \tag{6.4}$$

$$L_1 = \frac{1}{5\omega^2 C_{par}} - L_{par} \tag{6.5}$$
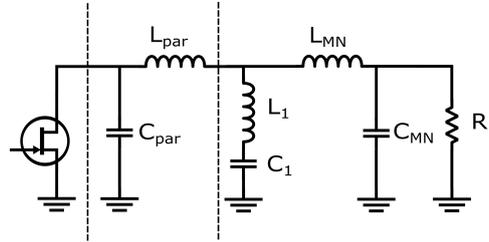
Figure 6.14: Harmonic termination network.

6.6.2 Results of the designed class-F

In our case, $L_{par} = 0$, but we can integrate a DC-block instead. C_{par} ($C_{dd,tot} \approx 260fF$) was determined first by simulation for a fixed DC-point first and checked with an S-parameter measurement and a load pull simulation in Cadence. The formulas now provide the initial values, the impedance optimisation with the optimizer in ADS, resulting in the schematic of figure 6.15. We can add a series capacitor as a DC-block by demanding that $\frac{C_{series}C_1}{C_{series}+C_1} = \frac{5}{4}C_{par}$. A $\frac{\lambda}{4}$-TL is added to replace the RF-choke and to provide an additional path to short the even harmonics.
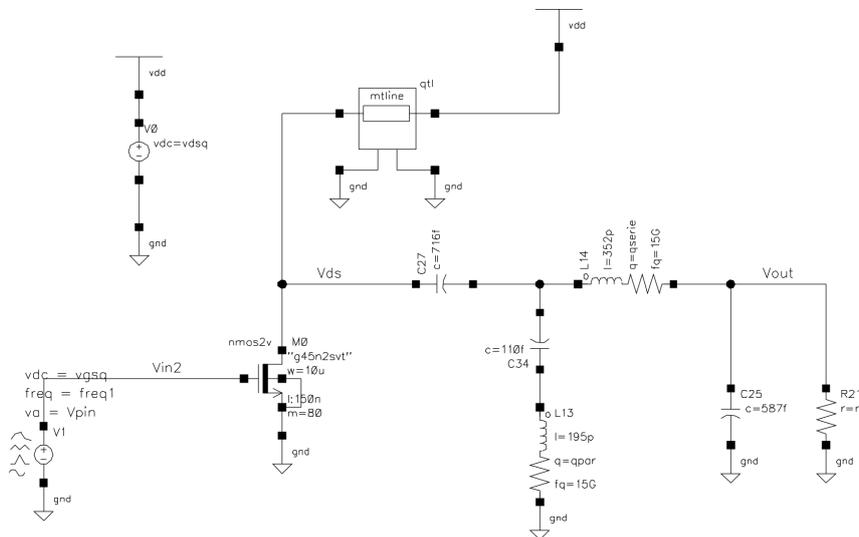
Figure 6.15: Schematic of the designed class-F PA, with $\frac{\lambda}{4}$-model.
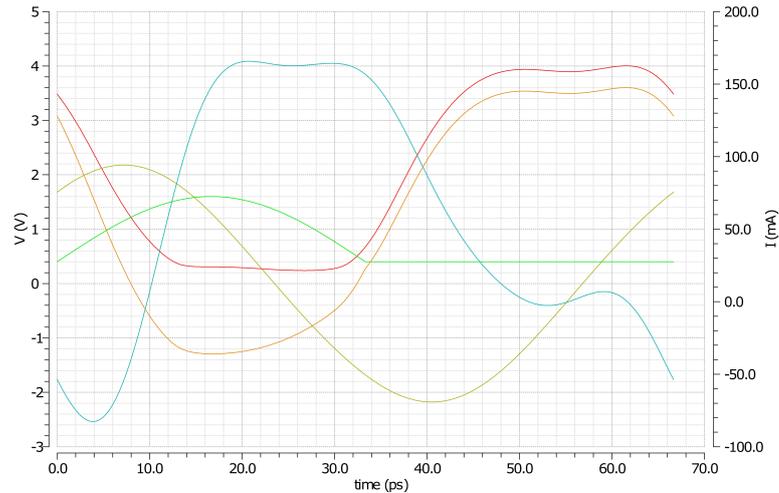
Figure 6.16: Waveforms in the designed class-F PA, with ideal inductors, a $\frac{\lambda}{4}$-TL-model and a clipped input signal.

With an ADS-model of a $\frac{\lambda}{4}$-TL on the chip substrate, ideal inductors, and an input sine with 1.2V amplitude and a DC bias of 0.4V, we obtain $P_{out,f_1} = 18.58$ dBm at $\eta_d = 84.26\%$ and the waveforms of figure 6.12. Both current and voltage resemble a square wave; but there is a problem: the ideal driver brings V_{gs} far too low and V_{DS} is simultaneously high, so $V_{dg,breakdown} = 3.6V$ is exceeded. When an ideal half-sine drive signal with the same amplitude is used, we obtain figure 6.16, and the efficiency decreases but the power increases:

Specification	Value
P_{out,f_1}	19.67 dBm
P_{dc}	120 mW
η_d	77.1 %
HD_{21}	-100.7 dB
HD_{31}	-36.53 dB

Table 6.1: Ideal class-F, with a half-sine drive signal.

With the simulated $\frac{\lambda}{4}$-TL and all quality factors set to 13, η_d decreases to 57.98 % and P_{out} is 18.36 dBm. Now, a real quarter-wavelength transmission line on the chip substrate was designed and folded up. With ideal inductors, we obtain $\eta_d = 64.78\%$ and $P_{out} = 18.87$ dBm. With all inductor quality factors set to 13, we obtain figure 6.18. Even though V_{DS} still resembles a square-wave, η_d decreases to only 47.21 % and P_{out,f_1} is only 17.77 dBm: a low quality factor in L_1 can be tolerated, but the key problem is the loss due to the low Q in the series inductor L_{MN}, which decreases the output power and efficiency, and the losses and relatively low impedances of the real $\frac{\lambda}{4}$-TL at the odd harmonics. It was attempted to replace L_{MN} by a $\frac{\lambda}{4}$-TL at 45 GHz or a short transmission line stub (similar to the inductor equivalent in a stepped-impedance lowpass filter), to transform the low impedance formed by C_{MN} and the load into an open at 45 GHz. This did not improve the results: the $\frac{\lambda}{4}$-TL is too lossy. No other topologies were found without a transmission line or inductor in series with the load. The efficiency of the individual PA with ideal drivers is already lower than the specification on η_d for the entire outphasing PA with real drivers, so we cannot proceed with this PA type.
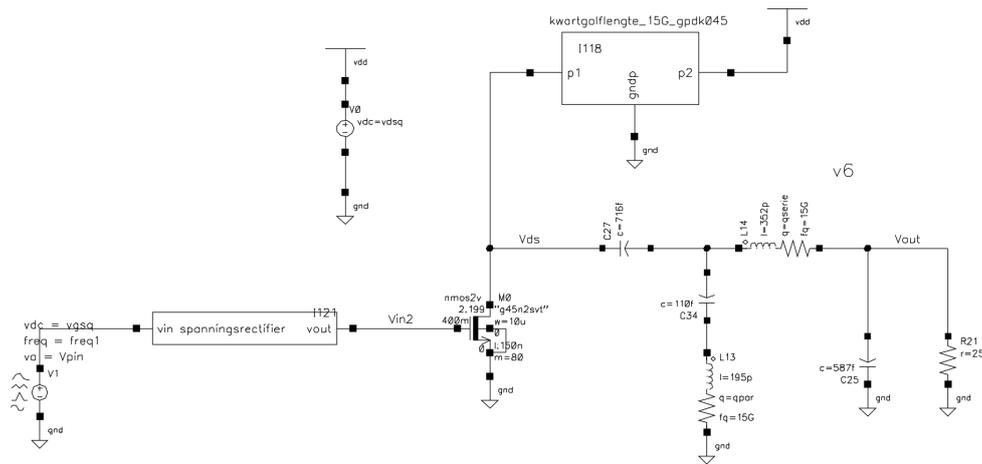
Figure 6.17: Schematic of the designed class-F PA , with the real $\frac{\lambda}{4}$-line with and a clipped input signal.
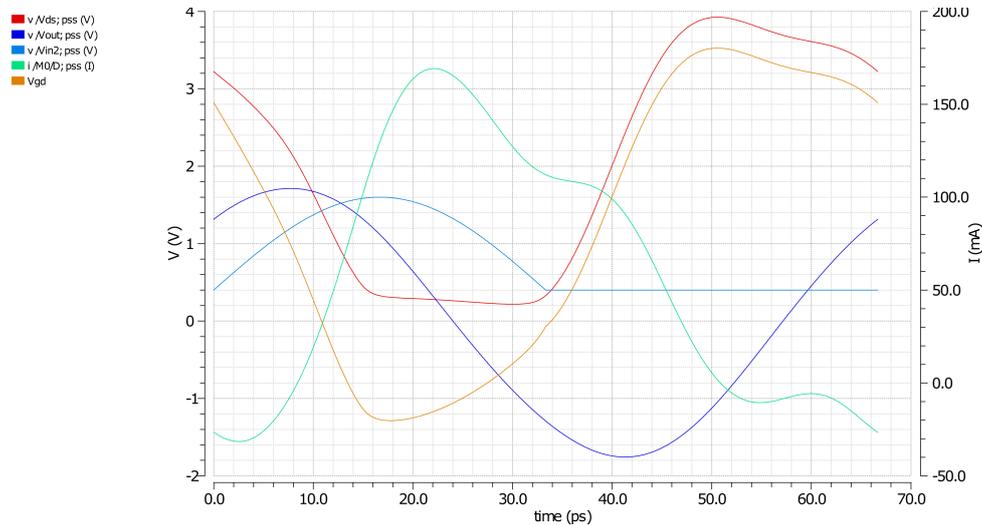
Figure 6.18: Waveforms in the designed class-F PA, with the real $\frac{\lambda}{4}$-line, all inductor quality factors set to 13, and a clipped input signal.

6.7 Designed class-E

The infinite RF-choke cannot be realised in practice, certainly not on-chip in CMOS. In [1] (p. 69), an analysis with a finite DC-feed and switch R_{on} is presented, based on a state-space model. The solution is obtained numerically and iteratively: starting from initial values, the parameters have to be tuned until a solution with sufficiently low $V_{DS}(t_{ON})$ and $\frac{dV_{DS}}{dt}(t_{ON})$ is found. In [24] (p. 31), the differential equations are solved, resulting in a convenient solution as a function of one variable: $q = \frac{1}{\omega\sqrt{L_d C_{shunt}}}$. In [24] (p.38), we find:

$$K_{ld} = \frac{\omega L_d}{R_L}, K_{C,shunt} = \omega C_{shunt} R_L \tag{6.6}$$

$$K_X = \frac{X}{R_L}, K_P = \frac{P_{out} R_L}{V_{dd}^2} \tag{6.7}$$

The most ideal class-E PA would need only a small inductance L_d (small K_{ld}) to realise a high P_{out} (high K_P) at the highest possible frequency and shunt capacitance (high $K_{C,shunt}$). A negative series reactance ($K_X < 0$) is useful because this is a negative inductor, which can be implemented by just reducing L_0. Starting from the q for maximal frequency $q = 1.468$ and by choosing $C_0 = 280fF$, the theoretical values are: $L_d = 175pH$, $C_{shunt} = 300fF$, $L_x = -42pH$, $L_{series} = 402pH + L_x \approx 360pH$. After some tuning, we obtain figure 6.22.
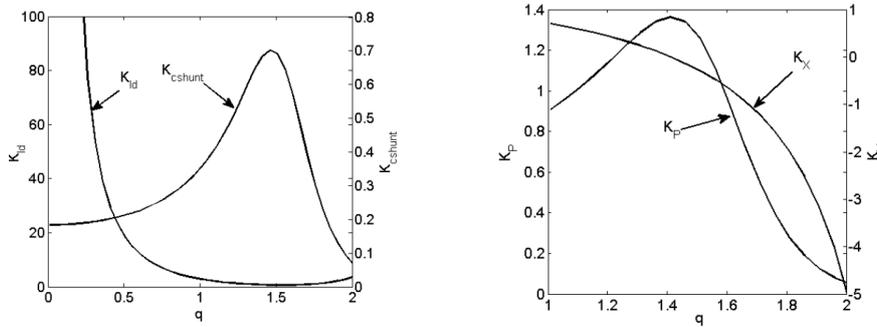
Figure 6.19: K_{ld}, $K_{C,shunt}$, K_{ld} and $K_{C,shunt}$ as a function of q.

q	K_{ld}	K_{cshunt}	K_X	K_P
0	∞	0.1836	1.152	0.5768
1.412	0.732	0.685	0	1.365
1.468	0.661	0.702	-0.159	1.331

Figure 6.20: Performance measures for two q-values: maximal power at $q = 1.412$, and maximal frequency at $q = 1.468$.
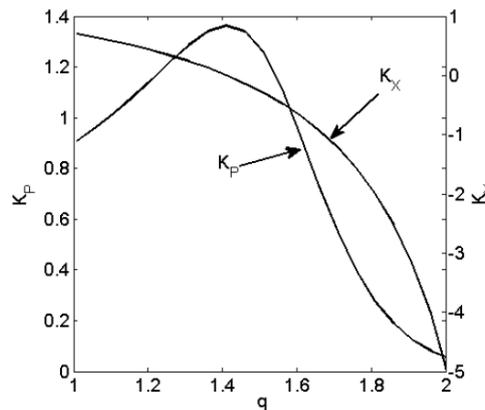
Figure 6.21: Summary of the ideal class-E PA with a finite DC-feed.
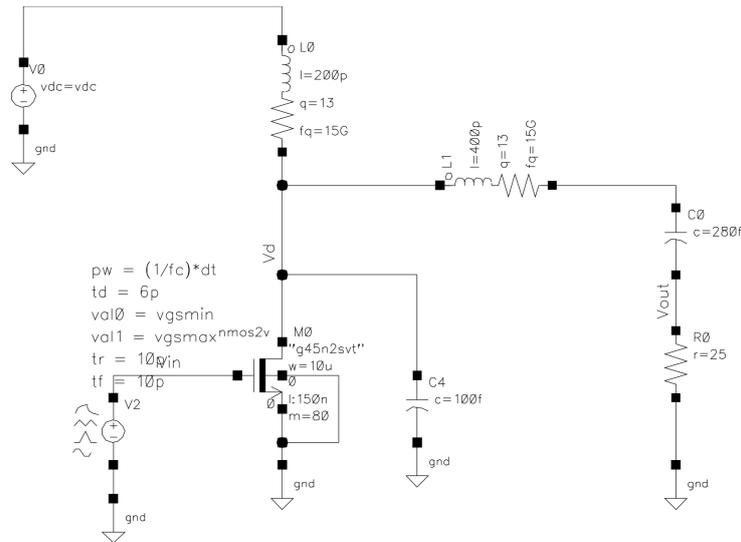
Figure 6.22: Schematic of the designed class-E PA.
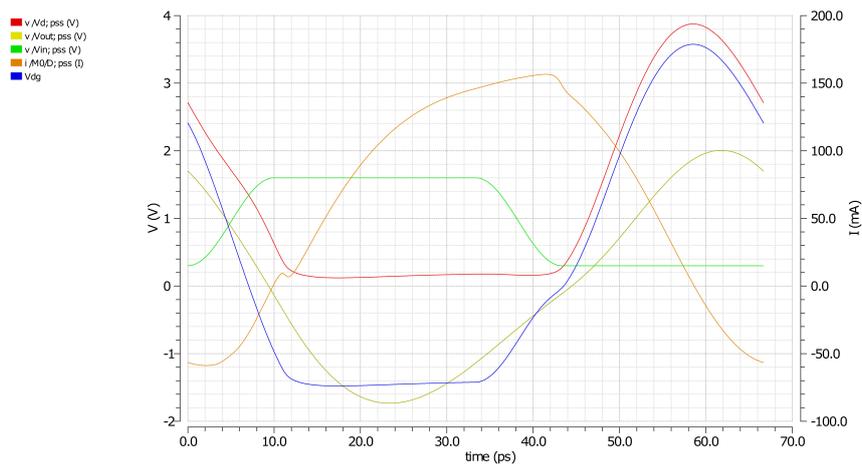
Figure 6.23: Waveforms in the designed class-E PA, with ideal inductors. With non-ideal inductors, V_{ds} and V_{gd} are decreased due to losses in the DC-feed inductor, and the supply can be increased, but the efficiency stays lower.

Specification	Value
P_{out,f_1}	17.9936 dBm
P_{dc}	81.5176 mW
η_d	77.72 %
HD_{21}	-15.42 dB
HD_{31}	-33.5147 dB

Table 6.2: Results of the ideal class-E PA.

Decreasing the quality factor of the filter to 13 reduces η_d to 69.65 % and P_{out} to 17.2 dBm. If only the quality factor of the DC-feed is set to 13, the supply can be increased, but still η_d decreases to 64.44 % and P_{out} is 17.8393. When both inductors have $Q = 13$, we η_d to 57.83 %

and P_{out} to 17.0861 dBm, and $HD_{21} = -14.8067$, which is quite high. A different design was now made:
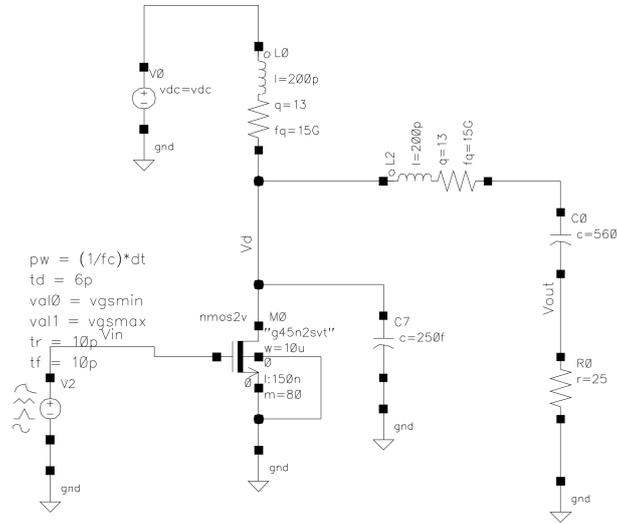
Figure 6.24: Schematic of the designed class-E PA.
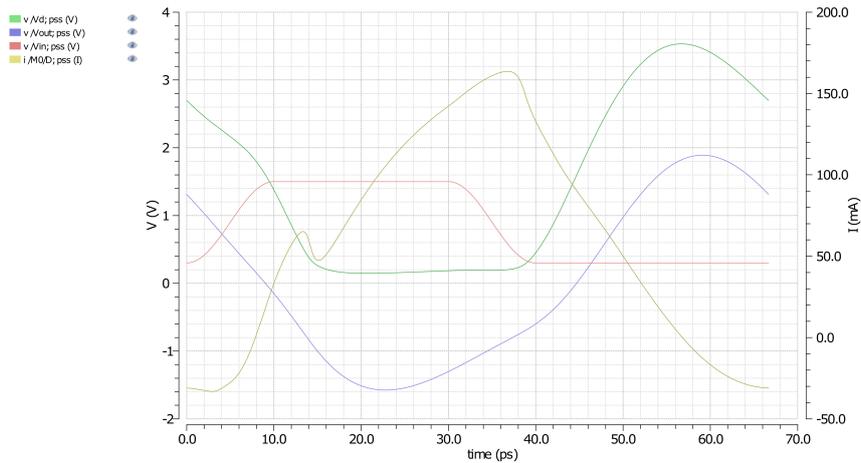
Figure 6.25: Waveforms of the designed class-E PA.

Specification	Value
P_{out,f_1}	17.49 dBm
P_{dc}	95.56 mW
η_d	58.69 %
HD_{21}	-18.37 dB
HD_{31}	-23.89 dB

Table 6.3: Results of the non-ideal class E-PA.

The efficiency is as low as the class-F PA, but some undesirable features remain: the second harmonic is still strong, the output power is not high, the input is a very strong and fast signal ($[0.3V, 1.6V]$, with a rise and fall time of only 6 ps) , so a lot of efficiency can be lost in the

drivers in reality. No real inductors models have yet been simulated, and we have not used the 1.1V-transistors because of the high voltage stress of the class-E PA. It was therefore decided to advance with the class-EF-PA.

6.8 Ideal class-EF

In [47], an new family of ZVS PA's is proposed: the class E/F-PA's, intended to combine the advantages of class E and class F. The parasitic switch capacitance should be included in the harmonic termination network, as in class-E, but the large V_{DS}- and V_{GD}-peak is a disadvantage and should be avoided. In a class-F PA, the odd harmonics are presented with an infinite impedance and the even harmonics are shorted; I_{DS} is a positive half-sine, as in class B; and V_{DS} is shaped into a square wave. In the inverse class-F PA (" F^{-1} "), the roles of current and voltage are interchanged: the odd harmonics are shorted, the even harmonics are "open", V_{DS} is shaped into a positive half-sine and I_{DS} becomes a square wave. The implementation is simplified in a differential topology, as demonstrated in figure 6.26, because the even harmonic impedances do not depend on the odd harmonic impedances. Due to the differential inputs and the symmetry, the even harmonics appear as common-mode voltages, so all differentially connected loads are seen as "open". The odd harmonics appear as differential voltages, so they will differentially see the sum of the even-harmonic impedances of each branch as well, but this is not a disadvantage, because they have to be shorted.

At the first harmonics, each branch works as a normal class-E PA with a load $R_L + jX_L$. The load is connected differentially. A parallel LC-filter is added in between the drains of the transistors to block the first harmonic and force it into the load, the capacitance of this filter shorts the odd harmonics. There is a trade-off in the LC-values, because the quality factor of the inductor, and consequently of the filter, is finite: the impedance at the first harmonic increases as the finite-Q inductor increases, but this reduces the capacitor, which increases the impedance at the third harmonic and causes more third harmonic to appear in the output. The parasitic capacitance $C_{dd,tot}$ is resonated into an open at the second harmonic by reusing the finite inductance that was already present in a class E-design that was made before. We did not add special terminations for the fourth harmonic yet.

Some type of power combination is necessary to obtain enough output power from the small 1.1 V NMOS transistors which only allow $V_{dg,max} = 2.2V$. Power combination is never lossless, so in that case, we might as well use this differential class-EF PA and benefit from the simplified harmonic tuning. The balun that is needed to enable the differential load will be lossy, but it is integrated with the supply inductors by using the inductances on the primary side and feeding the supply through a center tap. This does have a disadvantage: asymmetry in the inductances or in any element on one side with respect to the other, will lead to an asymmetric supply current. This current is time-dependent because the inductance of the primary is not "infinite". Current imbalance is equivalent to a differential current and can no longer be distinguished from a differential signal current, so it will appear directly in the output.
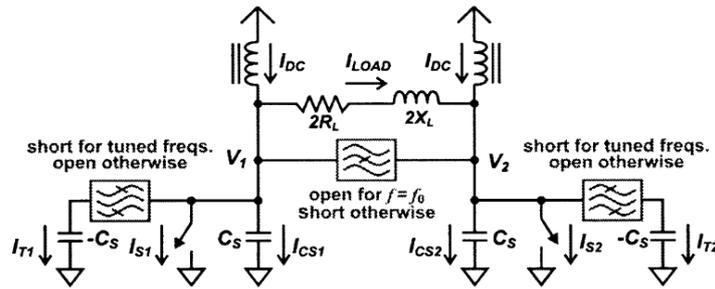
Figure 6.26: Concept of a class-$EF_{2,odd}$-PA.

6.9 Designed class-EF

The large 1.1V NMOS with $\frac{W}{L} = \frac{2\mu m}{45nm}$ (1 finger) and a multiplier 300 was used (multiplier 250 in the first versions). With this sizing of the transistor, the total $C_{dd,tot}$-capacitance at 1.1V is approximately $260fF$, which is quite large. Initially, we did not increase the number of fingers because this reduces $C_{dd,tot}$ and increases the inductance that is needed to resonate this capacitance away at 30 GHz: with $C_{dd,tot} = 260fF$, an inductance $L_d = 108pH$ is needed. L_d is small, but this is not a disadvantage: in the previous section, a class-E PA was made with a small DC-feed. The inductance of the primaries of the balun is also about 108 pH, and it cannot be changed easily because it is determined by matching requirements.

In the first version, separate inductors were used. In the schematic of figure 6.27, we set: $L_d = 150pH$, $L_0 = 150pH$, $C_0 = 300fF$, $C_e = 500fF$, $C_{ds,extra} = 1fF$ (it can be omitted), $R_L = 23\Omega$. The ideal driver outputs vary between $[0.4V, 1.2V]$ with a rise and fall time of 10 ps. The waveforms are plotted in figure 6.28. The current through the transistors is not exactly a square wave, but this would only be the case if all even harmonics were open-circuited. We have to take into account that we cannot distinguish the current through the transistor from the current through the parasitic capacitance of the transistor.

Specification	Value
P_{out,f_1}	18.3 dBm
$P_{in,tot}$	0.4575 dBm
Gain	17.84 dB
P_{dc}	83.03 mW
η_d	81.48 %
PAE	80.14 %
HD_{21}	-110.4 dB
HD_{31}	-20.97 dB
HD_{51}	-31.34 dB

Table 6.4: Performance of the ideal class EF-PA, with $V_{gd,max} = 2.18$ V, assuming ideal drivers with $V_{out,min} = 0.4V$, $V_{out,max} = 1.2V$ and a rise/fall time of 10 ps with a smoothed transition; and assuming ideal inductors ($V_{dd} = 0.9V$).
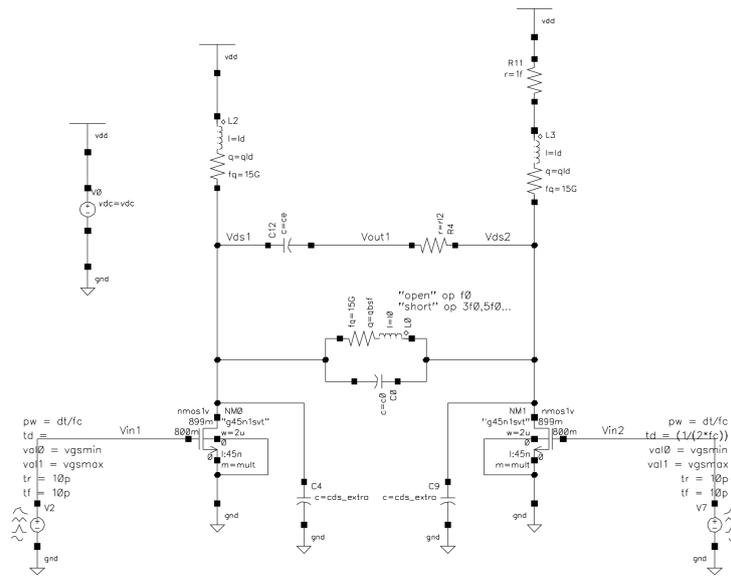
Figure 6.27: Schematic of the ideal class-EF PA with ideal drivers.
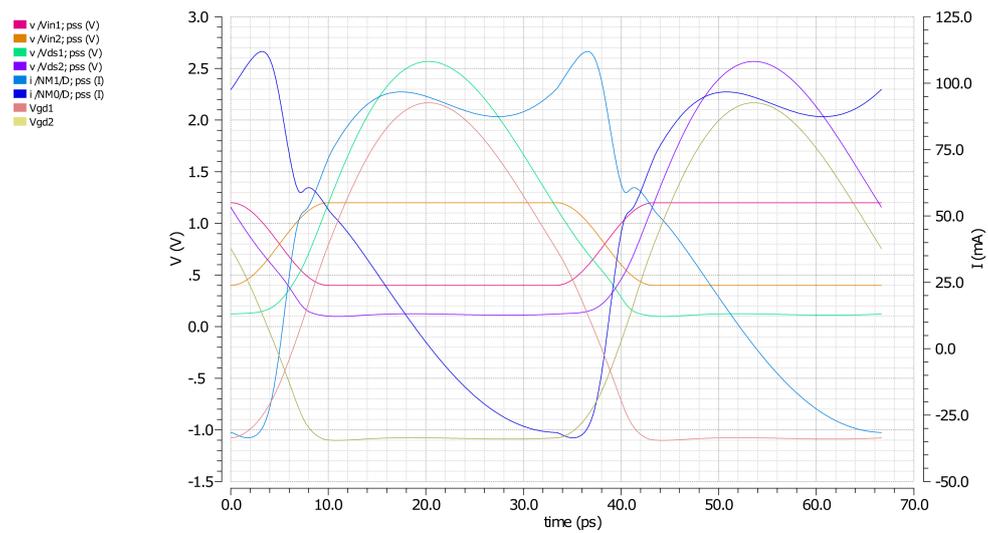
Figure 6.28: Waveforms of the ideal class-EF PA with ideal drivers.

These results seem promising at first, but again the efficiency drops quickly by decreasing the quality factor of the inductors to 13. V_{dd} is increased to 1V.

Specification	Value
P_{out,f_1}	18.26 dBm
$P_{in,tot}$	0.704 dBm
Gain	17.56 dB
P_{dc}	110.2 mW
η_d	60.83 %
PAE	59.77 %
HD_{21}	-110.6 dB
HD_{31}	-19.43 dB
HD_{51}	-29.12 dB

Table 6.5: Performance of the ideal class EF-PA, assuming ideal drivers and inductors with $Q = 13$.

An attempt was made to improve the efficiency by adding a T-network of $\frac{\lambda}{4}$-TL's (figure 6.29). At 15 GHz, the T-connection is at virtual ground and the $\frac{\lambda}{4}$-TL's of each path transform it into an open. At 30 GHz, the top $\frac{\lambda}{4}$-line creates an open and both paths are at the same voltage and see an open as well. However, the results did not improve, due to the high transmission line losses, and because we no longer have a pure class-E PA at the fundamental frequency.
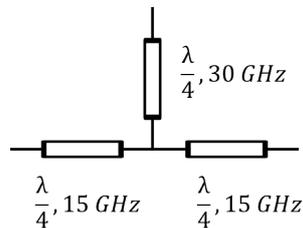
Figure 6.29: T-network of transmission lines.

We now replace the separate inductors with the "RFC-balun", which serves as "RF-choke" and balun. In the earlier versions of the RFC-balun, there was some asymmetry in the impedances seen by each branch, which was solved by adding a 50 pH slab-inductor from the center tap of the primary to the supply. In the ideal case of a perfectly symmetrical primary, this should not have any influence because the center tap becomes a virtual ground. In reality, this modification suppresses the relative inductance error of the two inductances of the primary to the center tap. In the version of figure 6.30, this modification was no longer necessary. The 50 pH slab-inductor will be reused in section 6.10.2. The capacitance in series with the load has now been moved to the secondary, otherwise, it would block the DC-current.
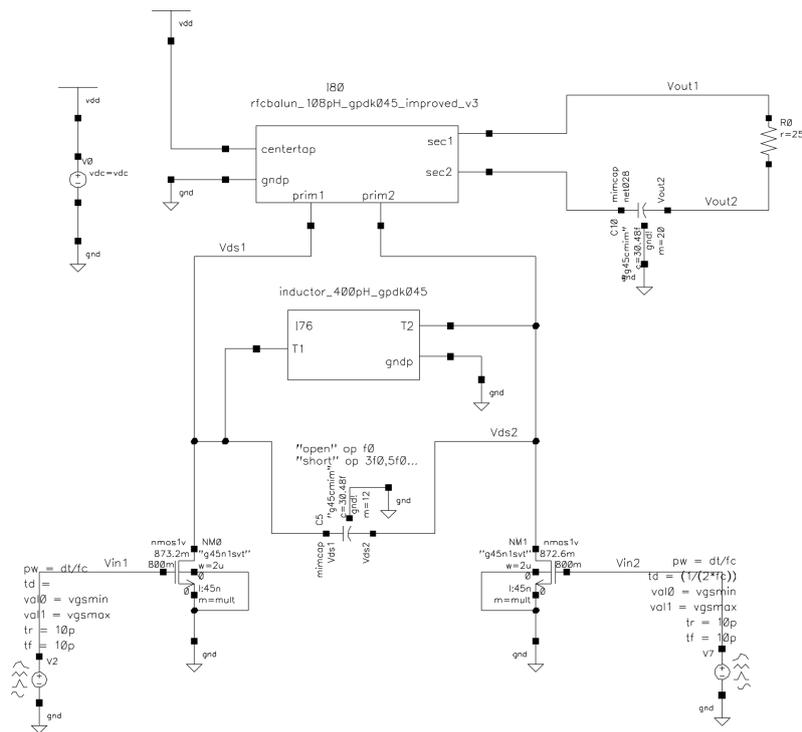
Figure 6.30: Schematic of the class-EF PA with ideal drivers.
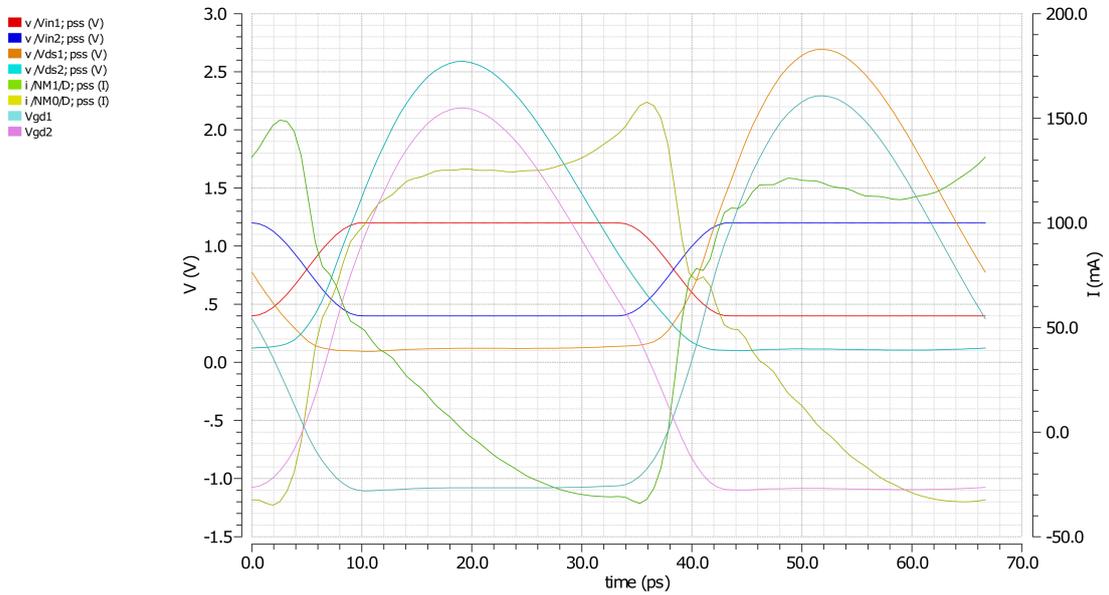
Figure 6.31: Waveforms of the class-EF PA with ideal drivers.

Specification	Value
P_{out,f_1}	18.21 dBm
$P_{in,tot}$	0.4717 dBm
Gain	17.74 dB
P_{dc}	109.1 mW
η_d	60.71 %
PAE	59.69 %
HD_{21}	-30.01 dB
HD_{31}	-20.89 dB
HD_{41}	-48.47 dB
HD_{51}	-33.78 dB

Table 6.6: Performance of the class EF-PA with RFC-balun, assuming ideal drivers.

6.10 Drivers

The class EF-output stage of the PA needs a strong input signal to switch completely and efficiently, but the power of the two outphasing signals is only 0 dBm, so driver stages have to be added. The DC power consumption of these stages is added to the total DC power consumption, so they have to be very efficient to avoid a low total efficiency. The drivers need to preserve the phase modulation, but amplitude linearity is not necessary or desired.

6.10.1 Invertors

As a first driver stage, a digital invertor is used because of the relatively high gain and saturation, which is desired in order to produce a square-wave-like signal. This invertor is biased in its "transition region" via resistive feedback. The DC-operating point of is set implicitly: in DC, the drain currents have to be equal because the DC-input impedance of the NMOS and PMOS is "infinite". Because of the DC-feedback resistor: $V_{GS,DC} = V_{DS,DC}$. By choosing $\frac{(\frac{W}{L}),PMOS}{(\frac{W}{L}),NMOS} = \frac{I_{0,NMOS}}{I_{0,PMOS}}$ this will be the case when $V_{GS,DC} = V_{DS,DC} \approx \frac{V_{DD,invertor}}{2}$. The input then has to be AC-coupled.

This invertor does not behave linearly at all, the sizing was done based on an approximation of the worst-case time constant at the output node. The $\frac{W}{L}$ of the PMOS is now fixed as a function of the $\frac{W}{L}$ of the NMOS, so we now express the equations as a function of the NMOS only. For example, the NMOS determines the speed of the falling edges: $R_{on,NMOS} \propto \frac{\alpha}{\frac{W}{L}}$, $C_{dd} \propto \beta \frac{W}{L}$ ($L = L_{min}$), C_L is independent of $\frac{W}{L}$. A dual equation exists for the rising edge, which is performed by the PMOS, but the same tradeoff holds there. We obtain:

$$\tau_{ON} \approx R_{on}(C_{dd} + C_L) \approx \alpha\beta + \frac{\alpha}{\frac{W}{L}}C_L \tag{6.8}$$

Equation 6.8 shows that we cannot obtain a time constant $\tau_{ON} < \alpha\beta$. Increasing $\frac{W}{L}$ will increase the steepness of the rising and falling edges and suppress the contribution of C_L on τ_{ON}, but $I_{DS,DC}$ increases with $\frac{W}{L}$ as well. To maximize the overall efficiency, the driver power

consumption needs to be small. The optimal sizing depends strongly on C_L, which is the total input capacitance of the next stage.

By not AC-coupling the output, this same bias point is used for the CS-stage as well. With $V_{dd,invertors} = 1.2V$ and $\frac{\frac{W}{L},PMOS}{\frac{W}{L},PMOS} = \frac{I_{0,NMOS}I_{0,PMOS}}{\approx}2$, we obtain $V_{DC,output} \approx \frac{V_{dd,invertors}}{2} \approx 0.6V$. This is only about 200 mV above V_{th}, so the CS-stage is biased in class-AB, as intended. The output stage transistors are intended to switch completely: their gate bias is equal to the supply of the LC-driver stage, which is quite low, but this is less important because they are driven very strongly.

The class-EF PA is differential, so we need differential input signals. These are obtained from a single-ended input by the 50Ω/50Ω-balun which was originally intended for use at the output, before the choice was made to use the more efficient common mode combiner. The input of the total PA needs to be matched to 50Ω. Narrowband matching is sufficient. A large inductance is needed if we use a single LC-matching network, but this will be provided by the bondwire inductance. With the rule-of-thumb of 1 nH per mm, we will need 0.85 mm of bondwire. The two inputs of the complete outphasing PA are now well matched: $S_{11}, S_{22} < -25dB$ at 15 GHz, and $S_{11}, S_{22} < -10dB$ from 14.23 GHz to 16 GHz. This is sufficient for our application, and allows for some deviation of the bondwire inductance, which is not perfectly known or fixed. In this case, input matching is important but not extremely critical, because of the large gain in the driver stages, especially the LC-driver stage.
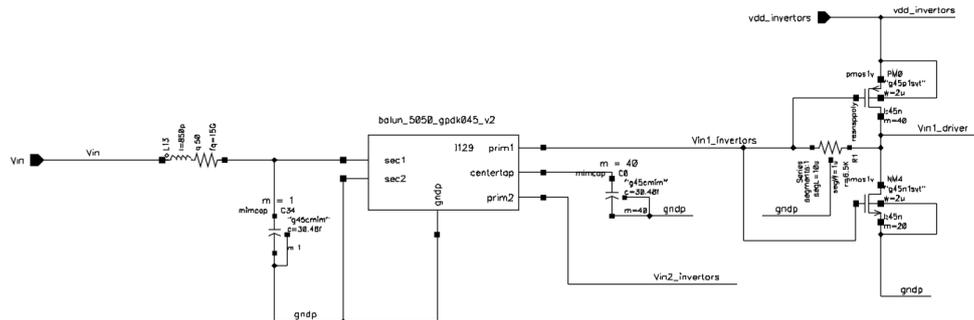
Figure 6.32: Schematic of the invertor drivers, with input balun and the LC-input matching for the entire PA. An identical invertor is added in between the nets "$Vin2\_invertors$" and "$Vin2\_driver$".

Because the CS-stage with LC-load is quite small, the load capacitance for the invertor stage is quite small, and the invertors succeed in transforming the small input sine into a larger square-wave, as demonstrated in figure 6.42 (e.g. from "$/PA1/Vin2\_invertors$" to "$/PA1/Vin2\_driver$"). Increasing the width further improves this "squaring" but only slightly improves the performance, while the DC-current is increased. The width of the invertors needs to be minimized without compromising the performance: $W_{NMOS} = 40\mu m$, $W_{NMOS} = 80\mu m$.

6.10.2 CS-stage with LC-load
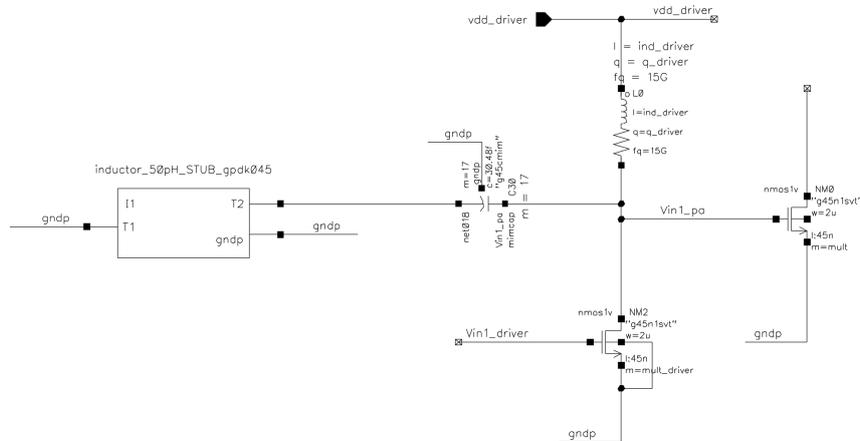
Figure 6.33: Schematic of the one of the CS-drivers with a modified LC-load.
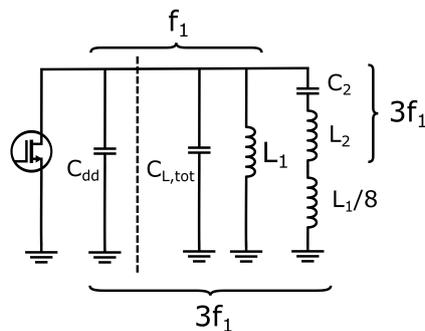
Figure 6.34: Initial concept of the load network for the LC-driver, to obtain an "open" at the fundamental frequency and the third harmonic. Although not exact, it is a good enough starting point: the equations can be solved to obtain zero admittance at the first and third harmonic, for a given $C_{dd,tot,eff}$.

The output stage transistors are the largest transistors: $W = 2\mu m$, with a multiplier of 300, so $W_{effective} = 600\mu m$. This implies a large $C_{gg} \approx 500fF$. The voltage swing at the drain is large and $C_{gd} \approx 85fF$, so some contribution due to Miller effect is expected as well, but the voltage transfer function from V_{GS} to V_{DS} is strongly time-dependent: it is almost zero when the transistor is on, and can become large when the transistor is off. This implies that the Miller contribution to the input capacitance is time-dependent and not well known, but certainly present. The resulting total input capacitance is very large: $C_{in,eff} \approx 600fF$. If this large capacitance were to be driven directly by an invertor stage, the DC power consumption of the drivers would become unacceptably large and multiple cascaded invertor stages, with a tapered down width when moving towards the input, would be necessary.

A more power-efficient solution is found by using a class-AB-biased NMOS with an inductive load, which resonates with $C_{in,eff}$ at the fundamental frequency. An ideal driver would produce a square-wave-like drive signal for the output stage, so we want to include a contribution of the third harmonic as well. The conceptual schematic is given in figure 6.34. This concept is a decent starting point: we now calculate the admittance and substitute $C_1 = C_{in,eff}$ and we want $L_2 = 50pH$, because we already have a $50pH$ slab inductor. With $C_{in,eff}$ and $L_2 = 50pH$, we need $L_1 = 88pH$ and $C_2 = 546fF$. This addition of the third harmonic seems to works quite

well, as demonstrated in figure 6.40, where the input voltages on one side of the PA output stage are clearly resemble a square-wave. Without load modulation, this is achieved for all output stage transistors. Figure 6.40 also shows that load modulation can effect the drivers, because the driver outputs have become asymmetric, as described in section 6.13.

6.11 Outphasing PA performance, with real drivers, at peak output power

The schematic of the outphasing PA, with 50 Ω test sources and the $Z_0 = 40\Omega$-transmission lines, is given in figure 6.35. Only 45 nm - 1.1V NMOS-transistors have been used. Each invertor-NMOS has $W_{tot} = 40\mu m$, the invertor-PMOS has a total width of $80\mu m$. The LC-driver-NMOS has $W_{tot} = 60\mu m$, each output stage NMOS has a total width of $W_{tot} = 600\mu m$. This significant size difference is possible thanks to the inductive load, which resonates with the large $C_{gg,tot,eff}$.
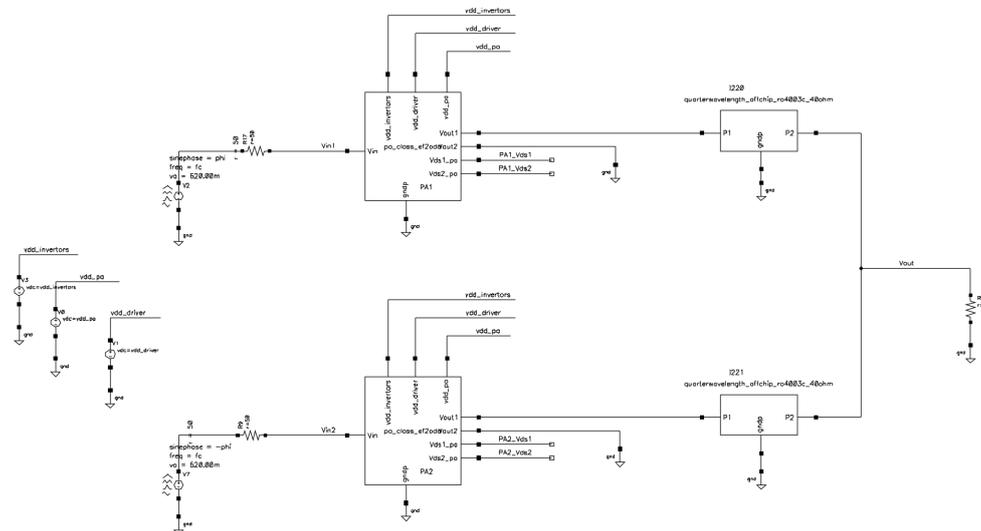
Figure 6.35: Schematic of the class-EF outphasing PA, with test sources and a common-mode combiner.
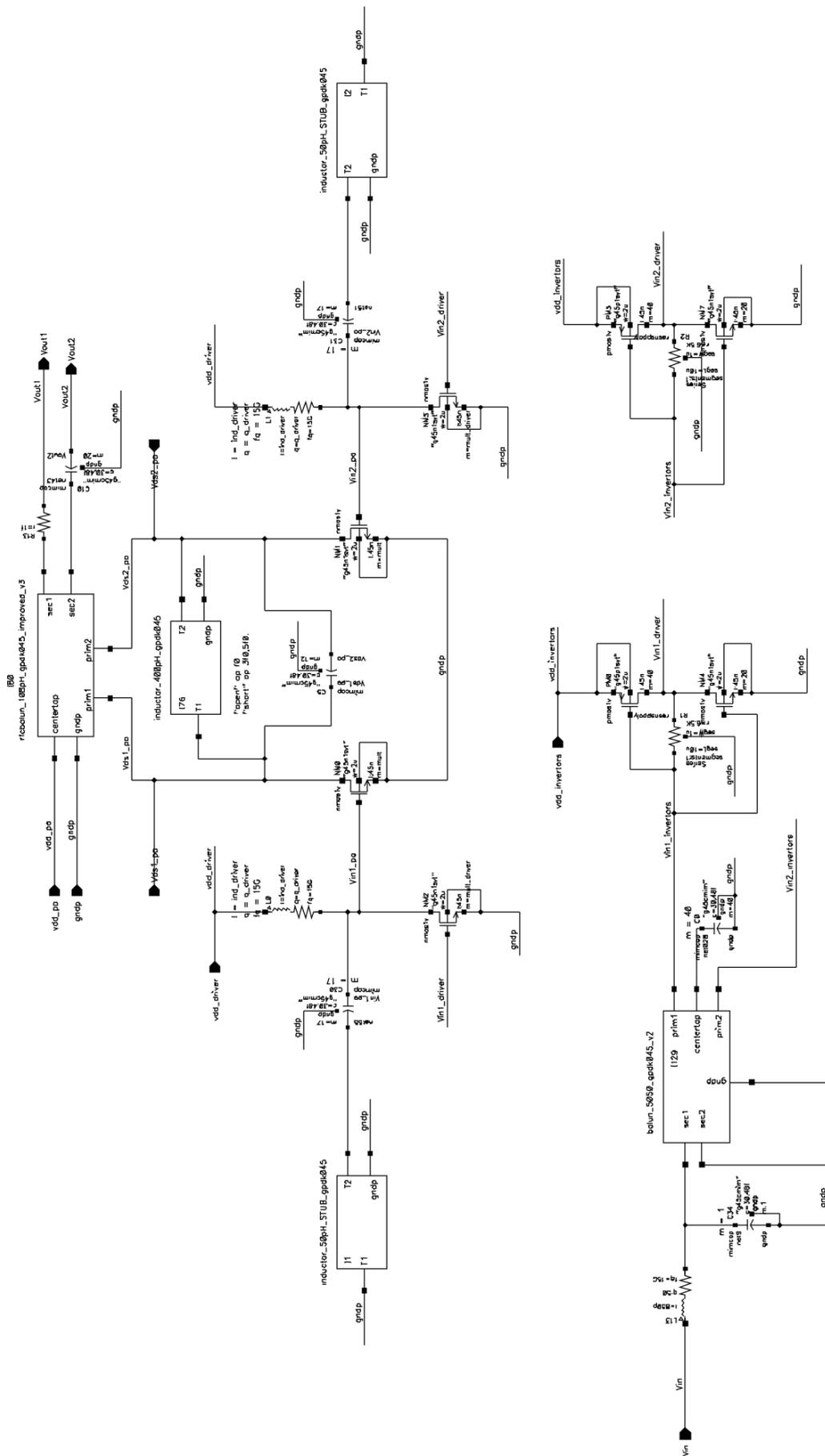
Figure 6.36: Total schematic of a single differential class-EF PA, with the invertors and LC-drivers.

With the addition of the real drivers, table 6.7 is obtained: a peak output power of 21.15 dBm is reached at an overall efficiency of 49.55 %. The invertors consume only 9.5 mW, the LC-drivers consume 22.64 mW.

Specification	Value
P_{out,f_1}	21.15 dBm
$P_{in,tot}$	2.81 dBm
Gain	18.34 dB
P_{dc}	263.2 mW
η_d	49.55 %
HD_{21}	-46.96 dB
HD_{31}	-24.01 dB
HD_{41}	-46.49 dB
HD_{51}	-43.6 dB

Table 6.7: Performance of the outphasing class EF-PA with load modulation, at peak power.

6.12 Outphasing with load modulation and ideal drivers
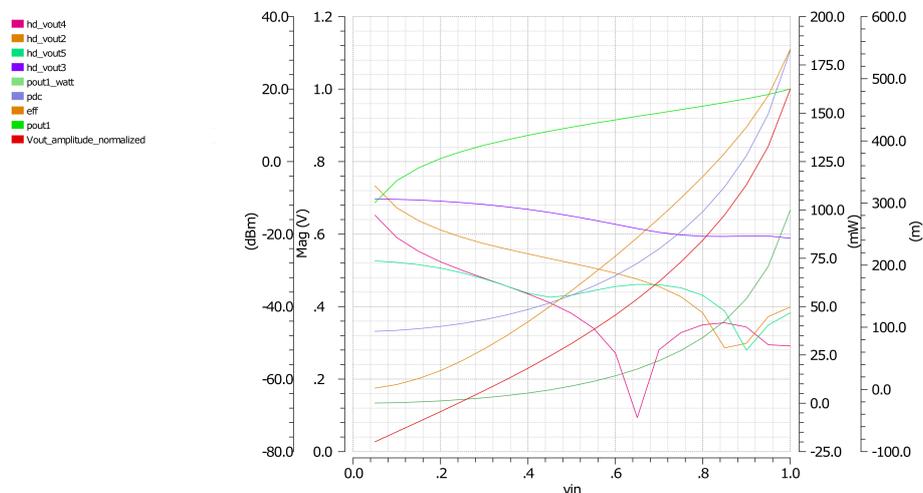
Figure 6.37: Waveforms for the 1-LINC class-EF outphasing PA, with ideal drivers and a common-mode combiner.
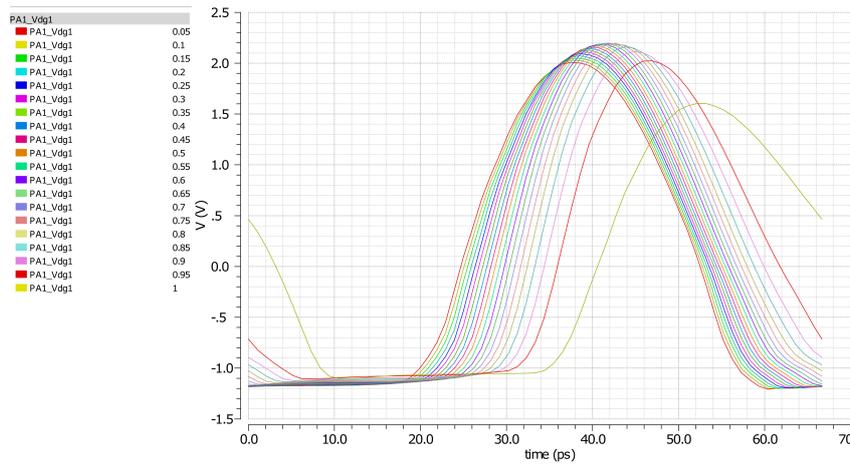
Figure 6.38: V_{dg}-waveforms for the 1-LINC class-EF outphasing PA, with ideal drivers and a common-mode combiner.
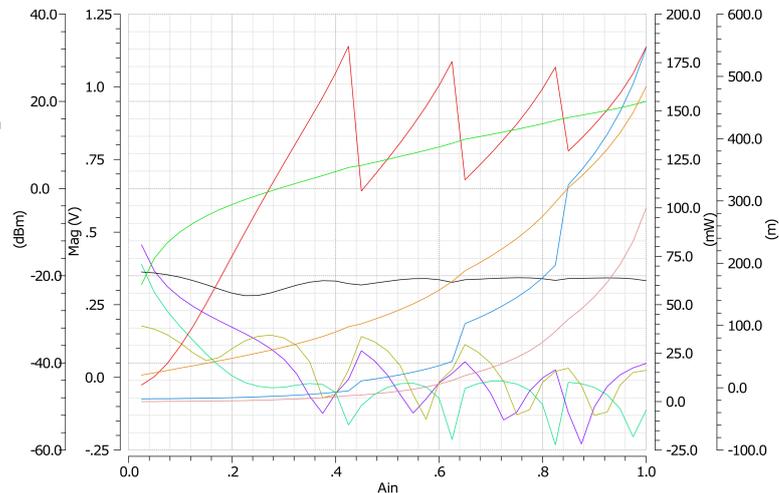
Figure 6.39: Waveforms for the 4-LINC class-EF outphasing PA, with ideal drivers and a common-mode combiner.

The entire design of a PA is based on a fixed load impedance; but with outphasing, the load is strongly dependent on the outphasing angle, e.g. equivalent parallel resistance is in both the common and differential combining cases inversely proportional to $cos(\phi)^2$, and increases to infinity when ϕ goes to $\frac{\pi}{2}$. During the design, we focussed on the load at $\phi = 0$. Problems were expected, and found, when ϕ is increased: as the load increases, the V_{gd}-breakdown voltage of the small 1.1V NMOS-transistors was exceeded, in both the common mode and the differential combiner.

Load modulation with a floating load is the most efficient option: the secondary of the RFC-balun does not have to be grounded, which improves its efficiency and symmetry slightly; and no power is lost in the power combiner. If the load is e.g. an (on-chip) antenna, this is the most efficient option. On the other hand: a differential load is not convenient for measurements.

If a single-ended output is needed, the common-mode combination is the most efficient option; but due to the size and loss of the $\frac{\lambda}{4}$-lines, this combination will have to be done off-chip. The load is referenced to ground, and the $\frac{\lambda}{4}$-TL's transform this impedance, but it is still referenced to ground, so compensation reactances have to be placed to ground and the secondary of the

RFC-balun has to be grounded as well. This reduces the efficiency and impedance of the RFC-balun slightly, but is still more efficient than differential combining with a very efficient balun (e.g. 80%). An EM-model of the $\frac{\lambda}{4}$-TL's was made on a Rogers RO4003C substrate, because of the low losses (low $tan(\delta)$=0.0027 at 10 GHz, $\epsilon_{r,design} = 3.55$).

Even in this case, the breakdown voltage of the 1.1 V NMOS was exceeded at rather moderate outphasing angles. Because there was excess output power, the supply was reduced to reduce V_{DD}. Still, with $Z_0 = 50\Omega$, V_{gd} limits V_{DD} and P_{out}, which was still too low. We now exploited the extra degree of freedom of the common mode combination: $Y_{in,1,CM} = \frac{2R_L}{Z_0^2}[cos(\phi)^2 - j\frac{sin(2\phi)}{2}]$, so the the (real part of) the load can be reduced by decreasing Z_0. With $Z_0 = 40\Omega$, 19.99 dBm output power is reached at a peak efficiency of approximately 54.74 %. The V_{dg} of all transistors stays below the maximal 2.2 V.

This impedance reduction should also result in more realisable compensation reactances. These can be be placed in between the drains of the transistors in a single PA, or at the secondary. In both cases, these turned out to be unrealisable: C_c is very low (e.g. 50 fF) and L_c is very high (e.g. 4 nH), regardless of the location. If we place only C_c, the efficiency stays higher for a higher range of outphasing angles, but the output power does not decrease fast enough with increasing ϕ and P_{out} at high ϕ is far too high (e.g. 6 dBm instead of -12.5 dBm, so an unacceptable amount of distortion is introduced.

The ML-LINC system will simultaneously solve both the issue of large V_{gd} at high ϕ and increase the efficiency, because the effective ϕ does not exceed a certain threshold. It also cleans up the output spectrum (especially the "leak through" of the third harmonic), which also degrades as ϕ increases. ϕ and $V_{DD,eff}$ have to be recalculated based on the original ϕ or on the relative input amplitude. This calculation was done by an ideal "block", it was included in the simulation setup itself. In reality, the location of the thresholds in ϕ_{orig} or $A_{in,relative}$ at which ϕ is reset have to be optimized to maximize the total efficiency for a given amplitude probability density. All performance have improved significantly by increasing the number of levels from 1 to 4.

6.13   Outphasing with load modulation and real drivers
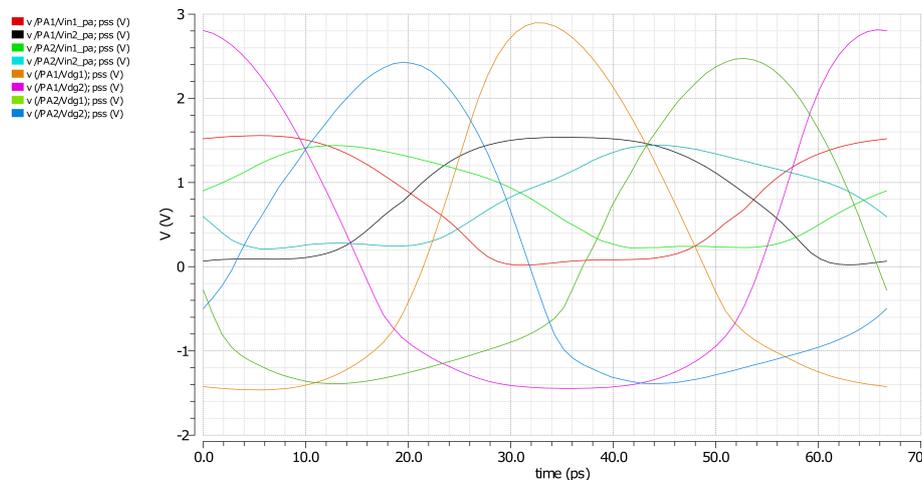
Figure 6.40: Asymmetric waveforms for the class-EF outphasing PA, with real drivers and a common-mode combiner, at a $A_{in} = 0.9$.

The specifications could not be not met without exceeding $V_{gd,breakdown}$ at a high P_{out} smaller than $P_{out,max}$. The effect of the load modulation is felt even into the LC-driver stage, resulting in asymmetric inputs at the output stage and a higher V_{gd}-peak (cfr. figure 6.40).

The PA with drivers can be used with a lower supply voltage, but the specifications will no longer be met, because $P_{out,max}$ will not exceed 20 dBm in this case. The origin of the problem can be found in the LC-drivers: when the loads are modulated, this modulation is partially transferred to the output of the LC-drivers because of the C_{gd} ($\approx 85fF$) of the output stage transistors. The effective loads seen by these drivers become strongly asymmetric in both paths, resulting in asymmetric waveforms at the input of the output stage. With ideal drivers, this problem was not found and only the effect of the load modulation on the output stage was present. With the real drivers, the asymmetry is much worse because the output stage is driven asymmetrically as well, and this results in a higher V_{gd}-peak. This V_{gd}-breakdown problem at low outphasing angles with the highest supply setting should be solved, e.g. by using different drivers or by preventing the load modulation from affecting the load seen by the drivers.

To solve this problem, we briefly experimented with a cascoded output stage, which is slower but allows for higher voltage peaks: the supply voltage can be increased and $P_{out,max}$ increases. Still, the efficiency significantly because the output voltage swing is limited on the lower side: to allow current through the top and bottom NMOS, each V_{DS} has to sufficiently high and power is dissipated in the transistors. This is also because the gate of the top transistor was connected to a DC-voltage, which forces more current through the top and thus also the bottom transistor and thus increases the minimal V_{DS} that can be reached by each transistor. This could be avoided by steering the gate of the top transistor as well. The current does become very similar to a square wave. Thanks to the differential topology and 40Ω $\frac{\lambda}{4}$-transmission lines, the real part of the load seen by each NMOS is already quite low, e.g. 8 or 9 Ω. Therefore, a cascoded PA, which is primarily useful to obtain a larger output power in a larger load, becomes unnecessary. Another option could be to use the class-EF version with the 1.8V 150 nm-transistors, but because of the longer channel length, the width of the output stage transistors is large, and the load capacitance which has to be driven, becomes very large as well.

6.14 Outphasing without load modulation, with real drivers

At this point, load modulation was abandoned in favour of the certainty of a fixed output load at each PA. The efficiency at output power back-off is improved by using 4-level-LINC system. An off-chip Wilkinson combiner, consisting of the same 40Ω - $\frac{\lambda}{4}$-transmission lines as in the CM-combiner and an extra 32 Ω-resistor, is used to present each PA with the same load as in the case with load modulation, while keeping the single-ended output load at 50 Ω. Because the transmission lines are identical, the maximal output power is also 21.15 dBm at an overall efficiency of 49.55 %; but the supply does not have to be reduced to prevent problems at output power back-off. The efficiency at power backoff is kept relatively high by using 4-level LINC, and the $A_{out}(A_{in})$-relation is very linear (slope 0.9964, offset -0.0079), because the only cause of amplitude distortion is a phase shift difference introduced by the PA's. With load modulation, a voltage division factor resulted for the unknown and non-linear PA output impedance and the variable load, which introduced amplitude distortion.
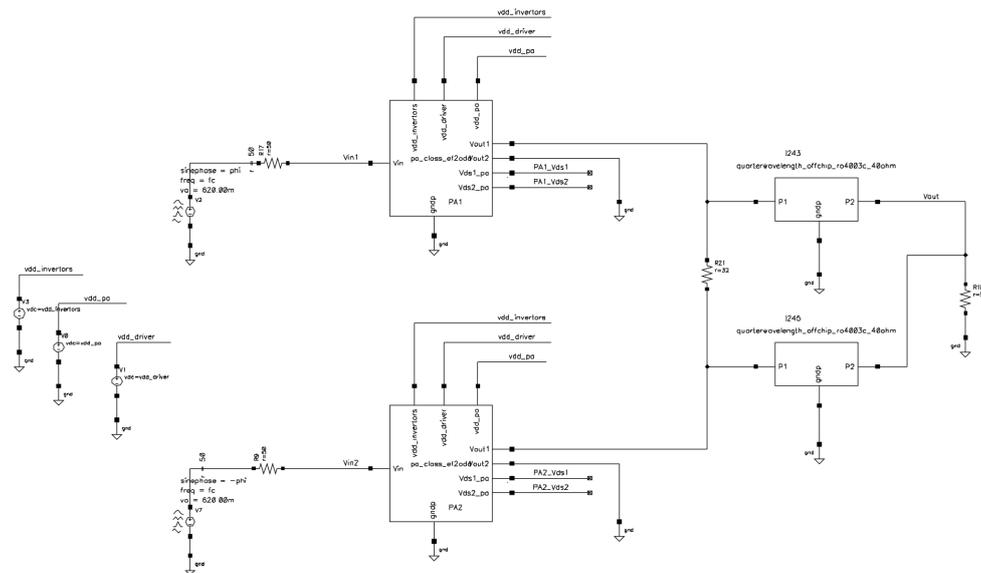
Figure 6.41: Class EF-outphasing PA, with real drivers and an isolating Wilkinson combiner (no load modulation).
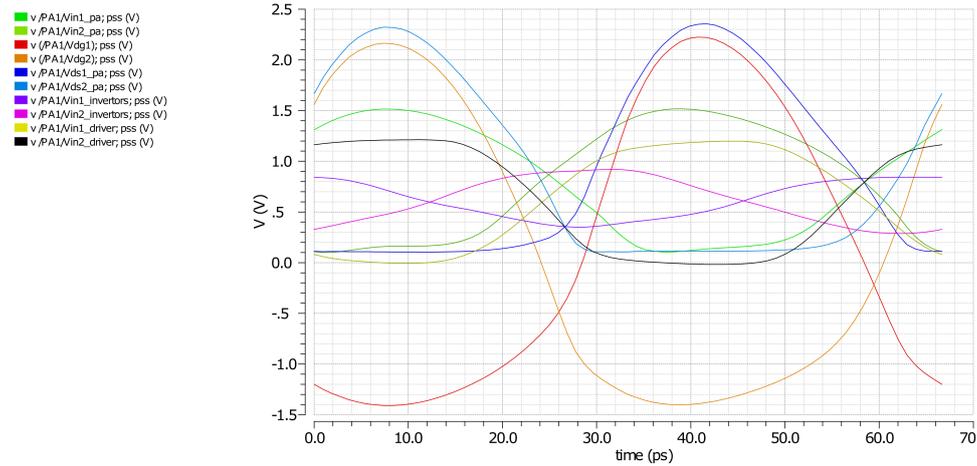
Figure 6.42: Waveforms for the class EF-outphasing PA, with real drivers and an isolating Wilkinson combiner (no load modulation).
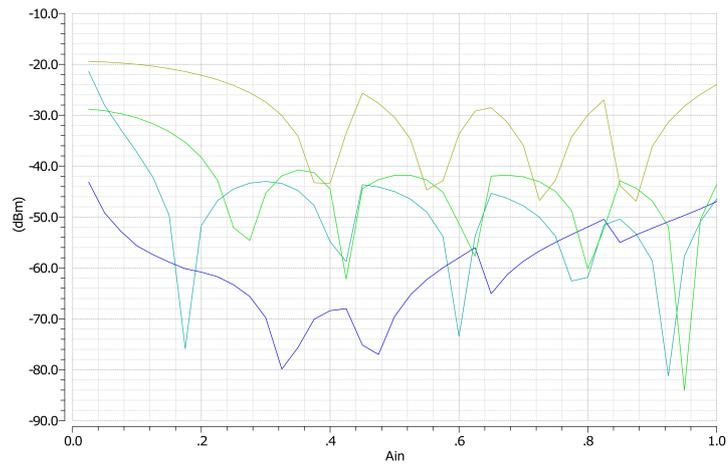
Figure 6.43: Waveforms for the class EF-outphasing PA, with real drivers and an isolating Wilkinson combiner (no load modulation).
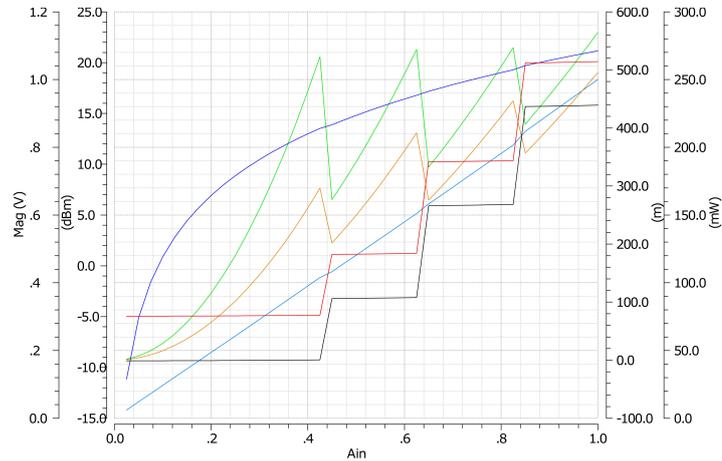
Figure 6.44: 4-LINC without load modulation, with real drivers, without driver supply scaling.
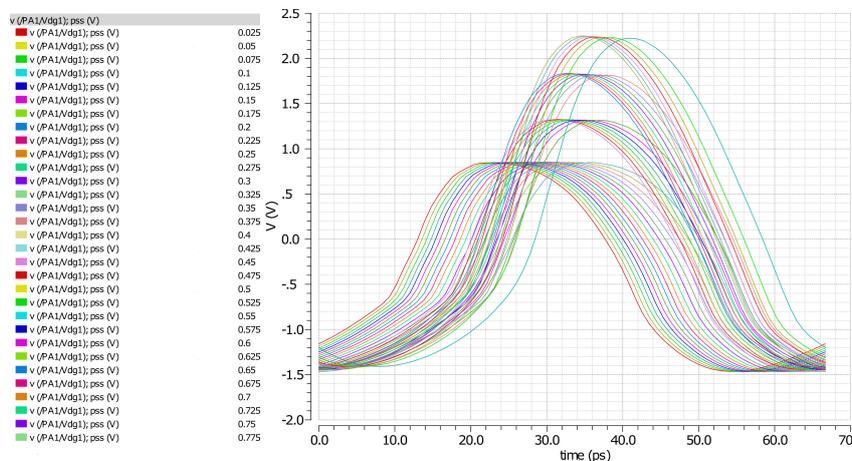
Figure 6.45: V_{dg}-waveforms for the class EF-outphasing PA, with real drivers and an isolating Wilkinson combiner (no load modulation).

Because the transmission lines are identical as in the CM-combiner, the maximal output power

is also 21.15 dBm at an overall efficiency of 49.55 %; but the supply does not have to be reduced to prevent problems at output power back-off. The supply reductions are clearly visible as steps in the DC power consumption. The overall linearity now only depends on the static difference of the phase shift introduced by PA's, which seems to be small. No dynamic effects are captured in this simulation: each A_{in} is a separate simulation iteration. This is clearly another advantage of not using load modulation: the amplitude at the output of each PA will not deform, regardless of A_{in}. $A_{out}(A_{in})$-is very linear, with the exception of the transition regions where the supply is reduced: to avoid "steps" in $A_{out}(A_{in})$, the power supply control has to be quite accurate. The power in the harmonics, which generally became much larger for increasing ϕ, is also improved for all output amplitudes: the third harmonic stays below -20 dB for almost all A_{in}.

6.15 Comparison

For convenient comparison, the results of the designed outphasing PA's are summarized here, allowing some conclusions to be drawn. Unfortunately, no solution with real drivers and load modulation was found yet. To still be able to somewhat compare the results, the efficiency of the PA output stage is also plotted separately in full lines in figure 6.46.
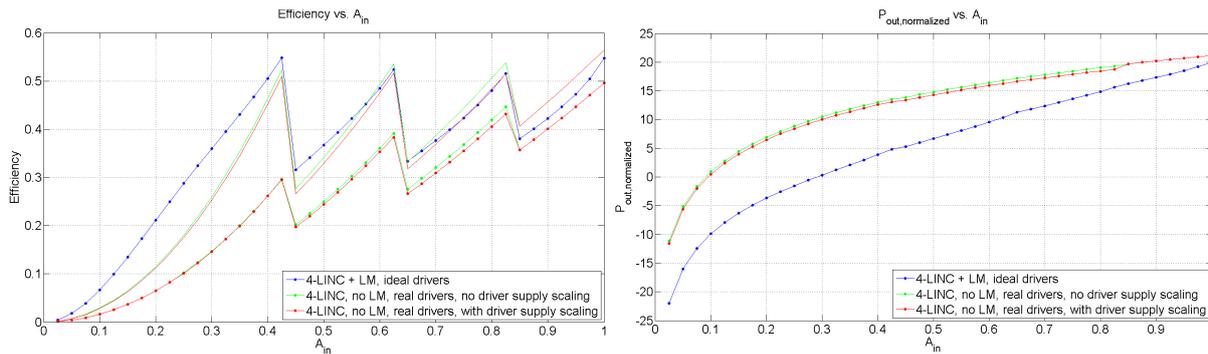
Figure 6.46: Efficiency and output power comparison.
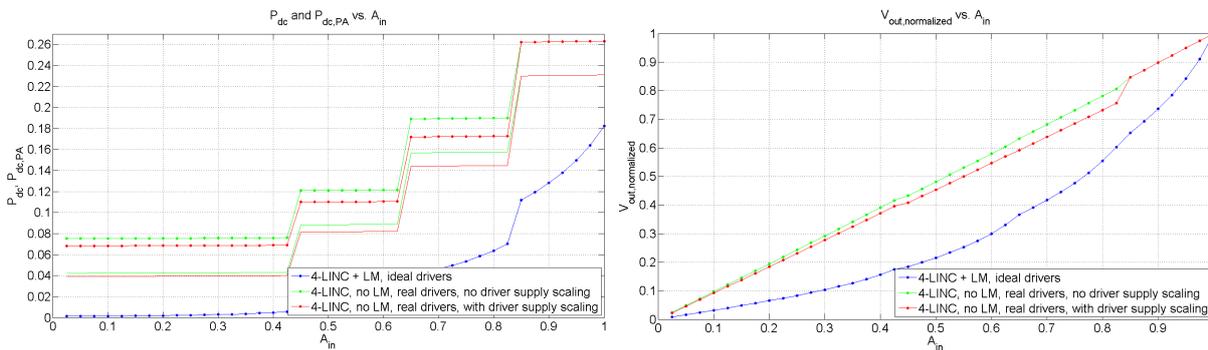
Figure 6.47: DC power and output amplitude comparison.

Some features become apparent:

- The real drivers deliver a larger drive signal to the output stage than the assumed ideal drivers but the rise and fall time are increased. Still, the efficiency peaks of the output stages are comparable as long as the outphasing angle is not too large. With load

> modulation, the efficiency drops less steeply with increasing outphasing angle because the PC-power decreases gradually since the real part of the load is increasing.

- With the drivers included, the overall efficiency drops significantly below the output stage efficiency, and this difference increases as the outphasing angle increases and $V_{dd,PA}$ is scaled down. Even though e.g. the $C_{dd,tot}$ is voltage-dependent, the output stage efficiency does not decrease with a decreasing supply voltage at first. An attempt was made to scale down the supply of the drivers, but this degrades the output stage efficiency and thus the overall efficiency, and the output linearity as well.

- The PA without load modulation is much more linear because the voltage division factor resulting for the unknown and non-linear PA output impedance and the variable load is not present. The only cause of amplitude distortion is a phase shift difference introduced by the PA's.

The design without load modulation offers more power and a more linear output. This power difference is partially due to the fact that the real drivers drive the gate of the output stage higher than the ideal drivers, resulting in a lower R_{on}, but they do so more slowly. They also bring the gate lower, which is not an advantage: if the transistor is already sufficiently below threshold, bringing V_{gs} lower only increases V_{dg}, which has to be compensated for by decreasing the supply and thus the output power and efficiency. It was attempted to avoid the drivers from bringing V_{gs} too low, but no good solution was found. E.g. diodes could be added to turn "on" when V_{gs} is already low enough, to "steal" the current provided by the drivers, preventing it from pulling the gate even lower, but these are very temperature-dependent and were not added. Therefore, the solution without load modulation is not yet optimal. Still, the difference in performance with and without load modulation is quite significant, considering that the output stages are almost identical.

When combined with ML-LINC, load modulation seems to lead to a sub-optimal design, even with ideal drivers: if no compensation reactances can be added, the supply has to be reduced to avoid V_{dg}-breakdown at moderate outphasing angles and the highest supply setting, reducing the peak output power and efficiency without a very substantial benefit in the efficiency at power back-off when compared to the ML-LINC system without load modulation. E.g. with the ideal drivers, V_{dg} at $P_{out,max}$ is only 1.6 V instead of the 2.2 V which is allowed without load modulation.

Giving up 27 % of the voltage margin in a rather low-voltage system, an output power reduction from 21.15 dBm to 19.99 dBm and a reduced linearity seems a rather large cost in exchange for a slower efficiency decrease and more efficiency at very low amplitudes; although this might become a significant advantage if the probability of these amplitudes is high enough. The DC power decreases gradually, and less DC-power is better for thermal reasons: the cooling requirements are reduced and the reliability and lifespan of the PA are most likely improved.

The optimal choice will depend on the number of realisable levels and the probability density of the amplitude of the signals. In general, the ML-LINC system without load modulation seems the most attractive option, certainly when combined with AMO (and unbalanced phase calibration).

Part V

Conclusion

Chapter 7

Conclusion and future work

During the course of this master thesis, a power amplifier with drivers was designed in 45 nm CMOS, as part of a 15 GHz outphasing transmitter. This amplifier is not the only important building block of the transmitter. Firstly, a 50 Ω unbalanced load was assumed in this thesis but this load should eventually be replaced by a transmitting antenna. Hence a 1-port antenna with an input impedance of 50 Ω has to be designed next. Secondly, a signal separator is needed to produce the two constant-amplitude phase-modulated outphasing signals starting from the modulated input signal in the baseband.

By theoretical simulations, it was determined that a good output quality can only be reached when both paths are very tightly matched, e.g. within 1° or 2° degrees of phase difference and $+/-1\%$ gain difference. In reality, this cannot be guaranteed under all circumstances without adding a continuous and adaptive mismatch calibration block and algorithm. Some option were found in literature and compared. AMO could be implemented and integrated with the SCS and (e.g. unbalanced phase) calibration block.

When combined with ML-LINC, load modulation seems to lead to a sub-optimal design, even with ideal drivers: if no compensation reactances can be added, the supply has to be reduced to avoid V_{dg}-breakdown at moderate outphasing angles and the highest supply setting, reducing the peak output power and efficiency without a very substantial benefit in the efficiency at power back-off when compared to the ML-LINC system without load modulation. In general, the ML-LINC system without load modulation seems the most attractive option, certainly when combined with AMO (and unbalanced phase calibration) (cfr. section 6.15).

Lastly, it became clear that the quality of the passives, the inductors and transformers, has a major influence on the efficiency of the PA and the total system. One way to improve the efficiency is by choosing the most strongly doped substrate to minimize the amount of substrate losses and thereby increasing the Q-factor of the inductors and consequently the efficiency of the transformers.

Bibliography

[1] P. Reynaert; M. Steyaert. *RF Power Amplifiers for Mobile Communications*. Springer, first edition, 2006.

[2] Mouser Electronics. Rf power amplifier efficiency: Big challenges for designers. `"http://www.mouser.com/publicrelations_techarticle_rf_power_amp_efficiency_ 2015final/"`. Accessed: 2015-04-14.

[3] Eta Devices. Eta devices website: Our technology. `"http://etadevices.com/ our-technology/"`. Accessed: 2015-05-20.

[4] Steve C. Cripps. *RF Power Amplifiers for Wireless Communications*. Artech House, second edition, 2006.

[5] SungWon Chung ; Godoy P.A. ; Barton T.W. ; Huang E.W. ; Perreault D.J. ; Dawson J.L. Asymmetric multilevel outphasing architecture for multi-standard transmitters. *Radio Frequency Integrated Circuits Symposium, 2009. RFIC 2009. IEEE*, pages 237 – 240, June 2009.

[6] Philip A. Godoy ; David J. Perreault; Joel L. Dawson. Outphasing Energy Recovery Amplifier with resistance compression for improved efficiency. *IEEE Transactions on Microwave Theory and Techniques, (Volume:57 , Issue: 12)*, pages 2895 – 2906, December 2009.

[7] Dixian Zhao ; Kulkarni S. ; Reynaert P. A 60-GHz Outphasing Transmitter in 40-nm CMOS. *IEEE Journal of Solid-State Circuits, (Volume:47 , Issue: 12)*, 47(12):3172 – 3183, October 2012.

[8] Tsan-Wen Chen ; Ping-Yuan Tsai ; Jui-Yuan Yu; Chen-Yi Lee. A sub-mw all-digital signal component separator with branch mismatch compensation for OFDM LINC transmitters. *Solid State Circuits Conference (A-SSCC), 2010 IEEE Asian*, pages 1 – 4, 8-10 Nov. 2010.

[9] Ying Tian ; Hammi O. ; Boumaiza S.; Ghannouchi F.M. Design and optimization of digital signal components separator of LINC transmitters using FPGA processors. *IEEE International Conference on Signal Processing and Communications, 2007. ICSPC 2007.*, pages 836 – 839, 24-27 Nov. 2007.

[10] Panseri L. ; Romano L. ; Levantino S. ; Samori C. ; Lacaita A.L. Low-power signal component separator for a 64-QAM 802.11 LINC transmitter. *IEEE Journal of Solid-State Circuits, (Volume:43 , Issue: 5)*, pages 1274 – 1286, May 2008.

[11] Yan Li ; Zhipeng Li ; Uyar O. ; Avniel Y. ; Megretski A. ; Stojanovic V. High-throughput signal component separator for Asymmetric Multi-Level Outphasing power amplifiers. *IEEE Journal of Solid-State Circuits, (Volume:48 , Issue: 2)*, pages 369 – 380, February 2013.

[12] Raab F.H. Efficiency of Outphasing RF Power-Amplifier Systems. *IEEE Transactions on Communications, (Volume:33 , Issue: 10)*, 33(10):1094 – 1099, October 1985.

[13] MA (US) Inventors: Joel L. Dawson; David J. Perreault; SungWon Chung; Philip Godoy; Everest Huang. Assignee: MIT, Cambridge. Asymmetric Multilevel Outphasing Architecture for RF Amplifiers. *US patent No. 8,026,763 B2*, pages 1–23, Sep. 27, 2011.

[14] Patentdocs. Eta devices, inc. patent applications. `"http://www.faqs.org/patents/assignee/eta-devices-inc/"`. Accessed: 2015-05-20.

[15] MIT Technology Review. Efficiency Breakthrough Promises Smartphones that Use Half the Power. `"http://www.technologyreview.com/news/506491/efficiency-breakthrough-promises-smartphones-that-use-half-the-power/"`. Accessed: 2015-05-20.

[16] Nagareda R. ; Fukawa K. ; Suzuki H. An MMSE based calibration of LINC transmitter. *Vehicular Technology Conference, 2002. VTC Spring 2002. IEEE 55th (Volume:2)*, pages 625 – 629 vol.2, 2002.

[17] Zhiwen Zhu ; Xinping Huang. An iterative calibration technique for LINC transmitter. *2013 IEEE 56th International Midwest Symposium on Circuits and Systems (MWSCAS)*, pages 384 – 387, 4-7 Aug. 2013.

[18] Joonhoi Hur ; Hyoungsoo Kim; Ockgoo Lee; Kwan-Woo Kim; Kyutae Lim; Bien F. An amplitude and phase mismatches calibration technique for the LINC transmitter with unbalanced phase control. *IEEE Transactions on Vehicular Technology (Volume:60 , Issue: 9)*, pages 4184 – 4193, 10 Oct. 2011.

[19] Chandrasekaran R. ; Gandhi R.; Kolanek J.C. ; Shynk J.J. ; Thomas A.L. Adaptive algorithms for calibrating a LINC amplifier. *IEEE Radio and Wireless Conference, 2001. RAWCON 2001.*, pages 214 – 244, 19 Aug 2001-22 Aug 2001.

[20] Robin Wesson; Mark van der Heijden. *Switch-mode RF PAs using Chireix outphasing (Simplified theory and practical application notes)*. NXP, May 2013.

[21] Binkley D.M. Tradeoffs and Optimization in Analog CMOS Design. *14th International Conference on Mixed Design of Integrated Circuits and Systems, 2007. MIXDES '07.*, pages 47 – 60, 21-23 June 2007.

[22] Binkley D.M. *Tradeoffs and Optimization in Analog CMOS Design.* John Wiley & Sons Ltd, first edition, 2008.

[23] P. Rombouts. *Advanced analog design, course at Ghent University.* 2015.

[24] Zisheng Li. *Analysis and Design of Highly Efficient Class-E Amplifiers for Indoor Ranging.* 2012-2013.

[25] Razavi B. *RF Microelectronics.* Pearson Education, 2nd edition, 2011.

[26] Voinigescu S. *High-Frequency Integrated Circuits.* Cambridge University Press, 2013.

[27] Jan Craninckx; Michiel S. J. Steyaert. A 1.8-GHz Low-Phase-Noise CMOS VCO Using Optimized Hollow Spiral Inductors. *IEEE journal of solid-state circuits*, 32(5):736–744, 1997.

[28] P.E. Allen. CMOS Analog Circuit Design: Lecture 070 – Resistors and Inductors. `http://www.aicdesign.org/SCNOTES/2010notes/Lect2UP070_(100419).pdf`, 2010.

[29] William B. Kuhn; Noureddin M. Ibrahim. Analysis of Current Crowding Effects in Multi-turn Spiral Inductors. *IEEE Transactions on Microwave Theory and Techniques*, 49(1):31–38, 2001.

[30] Ji Chen; Juin J. Liou. On-Chip Spiral Inductors for RF Applications: An Overview. *journal of semiconductor technology and science*, 4(3):149–167, 2004.

[31] Göran Jerke; Jens Lienig. Hierarchical Current-Density Verification in Arbitrarily Shaped Metallization Patterns of Analog Circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits And Systems*, 23(1):80–90, 2004.

[32] Ji Chen; Juin J. Liou. On-Chip Spiral Inductors for RF Applications: An Overview. *Journal Of Semiconductor Technology and Science*, 4(3):149–167, 2004.

[33] Yu Cao; Robert A. Groves; Xuejue Huang; Noah D. Zamdmer; Jean-Olivier Plouchart; Richard A. Wachnik; Tsu-Jae King; Chenming Hu. Frequency-Independent Equivalent-Circuit Model for On-Chip Spiral Inductors. *IEEE journal of solid-state circuits*, 38(3):419–426, 2003.

[34] Sunderarajan S. Mohan; Msaria del Mar Hershenson; Stephen P. Boyd; Thomas H. Lee. Simple Accurate Expressions for Planar Spiral Inductances. *IEEE journal of solid-state circuits*, 34(10):1419–1424, 1999.

[35] Jan Vandewege. Course: High Speed Electronics (University of Ghent). 2013-2014.

[36] Winfried Bakalski; Werner Simbiirger; Herbert Knapp; Hans-Dieter Wohlmuth; Arpad L. Scholtz. Lumped and Distributed Lattice-type LC-Baluns . *Microwave Symposium Digest, 2002 IEEE MTT-S International*, 1, 2002.

[37] Tadashi Kawai; Yoshihiro Kokubo; Isao Ohta. Broadband Lumped-element 180-degree Hybrids Utilizing Lattice Circuits . *Microwave Symposium Digest, 2001 IEEE MTT-S International*, 1, 2001.

[38] L. Besser and R. Gilmore. *Practical RF Circuit Design for Modern Wireless Systems*, volume 1 of *Artech House microwave library*. Artech House, Incorporated, 2002.

[39] Tak Shun Dickson Cheung; John R. Long. A 2126-GHz SiGe Bipolar Power Amplifier MMIC. *IEEE journal of solid-state circuits*, 40(12):2583–2597, 2005.

[40] Sunderarajan S. Mohan. *The design, modeling and optimization of on-chip inductor and transformer circuits*. PhD thesis, Stanford University, 1999.

[41] Ichiro Aoki; Scott D. Kee; David B. Rutledge; Ali Hajimiri. Distributed Active TransformerA New Power-Combining and Impedance-Transformation Technique. *IEEE transactions on microwave theory and techniques*, 50(1):316–331, 2002.

[42] Pozar D. *Microwave Engineering*. Wiley, 4rd edition, 2012.

[43] W. Fan; Albert Lu; L. L. Wai; B. K. Lok. Mixed-Mode S-Parameter Characterization of Differential Structures. *Electronics Packaging Technology, 2003 5th Conference (EPTC 2003)*, pages 533–537, 2003.

[44] Jerry Sevick. A Simplified Analysis of the Broadband Transmission Line Transformer. High Frequency Electronics, February 2004.

[45] Anaren. Measurements techniques for baluns. `"https://www.anaren.com/sites/default/files/uploads/File/BalunTesting_0.pdf"`.

[46] Kopp W.S.; Pritchett S. High efficiency power amplification for microwave and millimeter frequencies. *IEEE MTT-S International Microwave Symposium Digest, 1989.*, 3(1):857–858, 1989.

[47] Scott D. Kee; Ichiro Aoki; Ali Hajimiri; David B. Rutledge;. The Class-E/F Family of ZVS Switching Amplifiers. *IEEE Transactions on Microwave Theory and Techniques*, 51(6):1677–1690, 2003.

[48] Ali M. Niknejad. *ASITIC*. `http://rfic.eecs.berkeley.edu/~niknejad/asitic.html`.

[49] Ali M. Niknejad. *Analysis, Simulation, and Applications of Passive Devices on Conductive Substrates*. PhD thesis, Berkeley University, 2000.

ASITIC

To simulate the passive components and transformers, ADS was used. An alternative would have been ASITIC[48]. A technology file needs to be written before using ASITIC. This technology file defines the substrate, metal layers, ... Subsequently, several commands are entered in the command line to draw the desired layout. Finally, the design can be run to find out how the component behaves (e.g. ASITIC generates the narrowband equivalent system for an inductor at a given frequency).
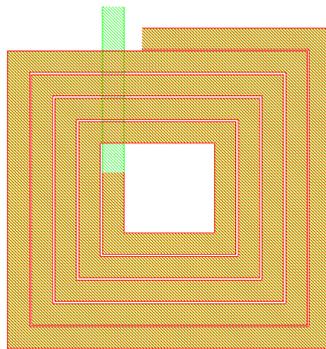
Figure 1: Example of an ASITIC layout

More information on how ASITIC works and on the physics used in the EM solver can be found in the PHD report written by Ali M. Niknejad[49].

The main reason to consider using ASITIC is that it is faster than ADS. A major drawback concerning ASITIC is that the files containing the substrate info are quite large (in our case approximately 300 MB per simulated frequency). But the occupied memory strongly depends on the pixel size used for the simulation. Hence increasing the accuracy will increase the simulation time and the required memory size to store the result.

The main advantages of ADS are that it is more accurate and user friendly than ASITIC. However, to make an initial guess on the dimensions, the accuracy of ASITIC might be sufficient.