

Power RF-DAC

The Design of a LDMOS
Class-E SMPA DRAC with
a CMOS Driver

R.J. Bootsman

Power RF-DAC

The Design of a LDMOS Class-E SMPA DRAC with a CMOS Driver

by

R.J. Bootsman

to obtain the degree of

Master of Science
in Electrical Engineering

at the Delft University of Technology,
to be defended publicly on Tuesday January 30, 2018 at 15:00

Student number: 4093844
Project duration: September 8, 2016 – November 1, 2017
Thesis committee: Prof. dr. ing. L.C.M. de Vreede, TU Delft, Chair ELCA, supervisor
Dr. ir. S.M. Alavi, TU Delft, ELCA
Prof. dr. J.E.J. Schmitz, TU Delft, Dean EEMCS
Dr. F. van Rijs, Ampleon, Innovation Manager and RF Architect
Ir. R.M. Heeres, Ampleon, Principal RF Engineer

This thesis is confidential and cannot be made public until December 31, 2021.

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.

Cover Photo Information: Radio Mast and Faculty by Sunrise

Camera NIKON D7200
Lens Sigma 105 mm *f*/2.8 EX DG OS HSM
Date Time Original 2018:01:07 09:02:08
Focal length 105 mm
Aperture *f*/5.6
Exposure time 1/200 s
ISO 200
Flash off, did not fire
Copyright ©2018 Robert Bootsman

Shown in foreground is a radio mast for telecommunications owned by T-Mobile, put into operation on the 24th of November 2017. It provides 4 types of communication: a microwave (μ W) transmission link, GSM1800, UMTS and LTE. The microwave transmission link is directed at 199,1° from north, operates at 32,771 GHz with an output power of 3,6 dBW. The main lobes of the antennae for the remaining types are directed at 10°, 120° and 240° from north.

Type	Direction	Frequency	Power
GSM1800	10°	1848,8 MHz	32,1 dBW
	120°	1872,2 MHz	32 dBW
	240°	1874,4 MHz	32 dBW
UMTS and LTE	10°, 120° and 240°	950 MHz	32,2 dBW
	10°, 120° and 240°	1860 MHz	32,1 dBW
	10°, 120° and 240°	2144,7 MHz	32,6 dBW

Shown in the background is high rise building of the faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS) of the Delft University of Technology (TU Delft), where the research for this thesis was conducted. This building is more commonly known as 'het EWI-gebouw', or simply EWI, and is situated along the Mekelpark. The high rise building is in use since 1969. With 23 floors it stands 90 m tall making it the second highest building of Delft, right after the Nieuwe Kerk. The building is designated as a municipal monument. As the radio mast is placed on a flat building lot it allows for a nice clear view to the faculty of EEMCS.

Summary

This document explains the design of a power RF-DAC, with as main focus to seek improvement in power efficiency for mobile Radio Frequency (RF) communication applications. For this reason a polar RF-DAC topology is selected. Due to limitations in current technology the power RF-DAC can not be integrated on one chip, therefore a combination of two chips is proposed to overcome these limitations. For the first chip a linear, high power analog LDMOS technology by Ampleon is proposed to implement a power Digital-to-RF-Amplitude Converter (DRAC) in and its design is the main focus of this thesis. The second chip will function as its driver and will be made in high integration, high speed digital 40 nm CMOS by TSMC, which has a maximum supply voltage for thick oxide devices of 2,5 V. The two chips will be connected using flip-chip bonding, allowing a maximum operation frequency of 3 GHz.

The LDMOS DRAC is designed to operate as non-linear Switch-Mode Power Amplifier (SMPA), allowing for high drain efficiencies. A Class-E matching network is used for the SMPA DRAC as it can incorporate the LDMOS output capacitance. The output voltage can be varied by modulating the Class-E transistor on-resistance, done by switching differently sized transistors. A LDMOS Total Gate Width (TGW) of 51 mm is used. When driven by a 2,0 V peak-to-peak square wave, post-layout simulations show a peak fundamental output power of $P_{out,fc} = 16,6$ W with a drain efficiency of $\eta_D = 64,4$ % at the maximum operation frequency of 3 GHz.

The DRAC has a dual line-up topology, primarily intended for push-pull operation to mitigate even orders of harmonic distortion. The TGW is divided over these two transistor banks. Within those banks the TGW is spread over 6 thermometer-coded segments and within these segments into a 6 bit binary-weighted fashion, creating an internal resolution of $N_b = 9$ bit which should allow for ACPR levels better than 50 dBc. Piecewise AM-AM linearization is applied on transistor level by applying a scaling factor to each of the segments. This results in a total of 2×43 individually sized LDMOS transistors. Static post-layout simulations show a peak Integral Nonlinearity (INL) error of $INL_{pk,LSB} = 10,7$ LSB. This error can be compensated for using Digital Pre-Distortion (DPD).

The CMOS driver operates in Class-D, which means the required drive power for the DRAC is proportional to fCV^2 . The peak input power for the LDMOS gate capacitances is $P_{in} = 0,91$ W for maximum operation frequency and DRAC output power. However, as there are two physically separate chip dies, both chips need to be provided with Electro-Static Discharge (ESD) protection which raises the required peak input power to $P_{in} = 2,13$ W, resulting in a peak total DRAC efficiency of $\eta_T = 59,5$ %. This required input power scales linearly with DRAC fundamental output voltage, or with the square root of the fundamental output power.

The CMOS driver output buffers consist of a chain of inverters increasing in size, which have an estimated DC power consumption 4,63 times the output power, resulting in a peak power consumption of 9,45 W. This factor remains constant for all possible values of output power (implying constant 'driver efficiency'), making it that driver power consumption scales linearly with required DRAC input power. This results in a lower average driver power consumption compared to an analog system where a Class-A or Class-AB predriver would be used. This also means that the RF-DAC system is frequency agile, implying that the operational frequency is only dependent on external output matching network components of the DRAC.

Including all other power consumptions of the CMOS driver, including two phase modulators and DSP, its expected peak total power consumption lands at 9,78 W. Using this an estimate for peak system efficiency can be made, giving $\eta_{T,system} = 46,7$ % at 3 GHz and peak DRAC output power. Using statistical data for a modulated signal with $PAPR = 8,0$ dB the average system efficiency is calculated to be $\overline{\eta_{T,system}} = 21,8$ % without the use of any energy efficiency enhancement techniques. As the DRAC design features a dual line-up topology, the push-pull operation could be switched out for e.g. Doherty operation to improve the average system efficiency, but this will something for future research.

Acknowledgments

This thesis is written in assignment of the Electronics Research Laboratory (ELCA) department at the faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS) of the Delft University of Technology (TU Delft). It is the final assessment for the completion of the master Electrical Engineering, track Micro Electronics.

Upon starting my master in the direction of Micro Electronics I did not yet know what specialization to pursue. At first I aimed for pure digital design, but at some point I wanted to do ‘something’ on the border between analog and digital. The course on digital RF was recommended to me with that idea in mind, which caused me to take the RF specialization courses in the subsequent semester. This resulted in me wanting to do a thesis project related to RF-DACs.

During the time working on this thesis project I have learned a lot, which I could not have done on my own. I take this opportunity to express my gratitude first towards my supervising professor, Leo de Vreede, for the time he was willing to make available for me, his support, enthusiasm and knowledge on this topic, the useful discussions and the feedback which all enabled me to get to this result.

Secondly I thank everyone in the ELCA group that shared their knowledge with me. It accelerated the design process for me as I could build upon previous research. In that sense it is truly a team effort. By name, I thank Earl McCune for interesting discussions and his ability to break complex matter down to extremely simple, yet surprisingly accurate, examples (of which the ‘variable resistor’ implementation of a DAC is a prime example), Morteza Alavi for his deep theoretical knowledge on RF-DACs, Mohsen Hashemi for providing me with good pointers for implementing the DRAC topology and transistor level AM-AM linearization, Atef Akhnoukh for his experience on the CMOS ESD protection and making the proposed CMOS layout possible and Yiyu Shen for his knowledge on CMOS design and according power consumptions which allowed me to make an estimation of system performance.

The research was done at the TU Delft, but with support from Ampleon. I thank Ampleon for their support during my thesis work and specifically Rob Heeres for his time on helping to creating the layout for the LDMOS DRAC and Fred van Rijs for supplying information on the LDMOS devices and the ‘customization’ options, such as the V_T -shift.

This thesis is the crown on a wonderful period as student at the TU Delft, for which I owe many thanks to my parents and friends. I hereby also thank my study association, the Electrotechnische Vereeniging (ETV), that facilitated connecting with a lot of people in Electrical Engineering and allowed me to quickly find new friends in Delft. Where the TU Delft taught me the essential academic and soft skills, the ETV and the committee work I have done there allowed me to further develop my soft skills and myself.

Lastly, I would like to thank the \LaTeX community for the upkeep of the software and related packages, such as `(Circuit)TikZ` and `PGFPLOTS`, that I could use to typeset this thesis and create most of the illustrations, and Stack Exchange for solving the many headaches related to \LaTeX encountered during the writing process of this thesis.

*R.J. Bootsman
Delft, January 2018*

Contents

Summary	v
Acknowledgments	vii
List of Figures	xi
List of Tables	xv
1 Introduction	1
1.1 Growing Required Wireless Data Capacity	2
1.1.1 Energy consumption	2
1.2 RF Transmitter Architectures	3
1.2.1 Analog RF TX Architectures	3
1.2.2 Digital-Intensive RF TX Architectures	4
1.3 Thesis Goals	6
1.4 Thesis Organization	7
I Theory	9
2 PA Basics	11
2.1 Definitions	11
2.1.1 PA Power	11
2.1.2 Gain	12
2.1.3 Efficiency	13
2.1.4 Linearity	13
2.1.5 Peak-to-Average Power Ratio	15
2.2 Amplifier Power Relations	15
2.3 Amplifier Classes	17
2.3.1 Device Model	17
2.3.2 Transconductance Amplifiers	20
2.3.3 Saturated Transconductance Amplifiers and Harmonic Tuning	23
2.3.4 Switching Amplifiers	26
2.4 Conclusion	31
3 RF-DAC Basics	33
3.1 DAC structures	33
3.1.1 Oversampled DACs	33
3.1.2 Nyquist rate DACs	34
3.1.3 DAC Performance Metrics	37
3.2 From DAC to DRAC	39
3.2.1 Frequency Domain Analysis	41
3.2.2 DRAC Performance Metrics	42
3.3 Towards a Power DRAC Implementation	44
3.4 Conclusion	46
II Implementation	47
4 System Design Considerations	49
4.1 Technology	49
4.1.1 Fraunhofer IAF GaN25	50
4.1.2 Ampleon LDMOS LM8	55
4.1.3 Technology Evaluation	56

4.2	Resolution and Dynamic Range	57
4.3	Driver	58
4.3.1	Driver Model	59
4.3.2	Inter-Chip Connection Methods	61
4.3.3	Driver Evaluation	64
4.4	Circuit Topology.	65
4.4.1	DRAC Structure	65
4.4.2	Single-Ended or Push-Pull Topology	67
4.5	Conclusions.	69
5	LDMOS DRAC Implementation	71
5.1	Technology Characterization.	71
5.1.1	Finding a Design Set	71
5.1.2	The AM-AM Curve	77
5.2	Implementing the DRAC	78
5.2.1	Increasing the Resolution and Individual Transistor Sizing.	79
5.2.2	Simulation Setup with Driver Model	79
5.3	Layout.	90
5.3.1	Layout Version 1	92
5.3.2	Layout Version 2	93
5.4	Conclusion	98
6	CMOS Driver	101
6.1	Driver Requirements	101
6.2	Output Buffer	102
6.2.1	Driver Model Revisited: Inverter Delays.	102
6.2.2	Designing the Inverter Chain.	105
6.2.3	Levelshifter	107
6.2.4	Complete Output Chain	108
6.3	Supply Decoupling	110
6.4	CMOS Driver Implementation Ideas.	111
6.5	System Performance Estimation.	112
6.5.1	Estimation of Future System Performance	113
6.6	Conclusion	115
7	Conclusion	117
7.1	Discussion	118
7.2	Recommendations for Further Work.	120
7.2.1	Short-Term Future Work	121
A	Appendix	123
A.1	Design Set Example Class-E	123
A.2	Transient DRAC Simulations.	124
A.3	Harmonic Distortion of Designed DRAC.	125
A.4	Mutual Inductance	125
A.5	OTA Transistor Widths	127
B	LDMOS Case Study: A Higher Drive Voltage	129
C	Design in GaN	133
	Bibliography	137
	Glossary	141
	Acronyms	143
	Symbols	147

List of Figures

1.1	Illustrating today's need for connectivity	2
1.2	Representation of $g(t)$ in the complex plane	3
1.3	Examples of analog modulation architectures	4
1.4	Digital polar TX architecture	5
1.5	Digital I/Q TX architecture	6
1.6	Combined RF-DAC and PA TX architecture	6
2.1	Power flows in a general PA	11
2.2	Different power gain measures	13
2.3	Generalized PA	15
2.4	Simple FET model	17
2.5	Switched FET model	19
2.6	FET $I-V$ curves	19
2.7	FET $I-V$ curves and example resistive load-line	20
2.8	Circuit topology of a transconductance PA	21
2.9	Class-A drain waveforms and load-line	21
2.10	Class-AB, B and C drain waveforms and load-lines	22
2.11	Performance of the different ideal transconductance amplifier classes	24
2.12	Saturated Class-A drain waveforms and load-line	24
2.13	Saturated Class-B drain waveforms and load-line	24
2.14	Circuit topology of a Class-F PA	25
2.15	Circuit topology of a Class-F PA with a transmission line	25
2.16	Class- F_3 , F_5 and F_∞ drain waveforms and load-line and Class-B dashed for reference	26
2.17	Class-D SMPA circuit topology and waveforms	27
2.18	Class-E SMPA circuit topology and switch model equivalent	27
2.19	Class-E SMPA waveform	29
2.20	Class-E SMPA waveform for non-zero m and corresponding load-line	30
2.21	Drain efficiencies for varying q and R_L	31
3.1	Simplified DAC circuit diagram	34
3.2	Switched circuit quantity DAC implementations	35
3.3	Origin of glitches	35
3.4	Different INL reference line definitions with corresponding INL curves and DNL curve	37
3.5	DAC settling time	38
3.6	Simplified variable switched resistance DRAC circuit diagrams	40
3.7	DRAC time domain waveform	40
3.8	Baseband ACW frequency spectrum with ZOH-effect	42
3.9	Upconverted bandpass RF absolute frequency power spectrum with ZOH-effect and harmonic replicas	42
3.10	DRAC settling time	43
3.11	Class-E SMPA DRAC	44
3.12	Ideal switch Class-E DRAC output curves	45
4.1	Technology comparison in terms of output power and speed [35]	49
4.2	GaN25: DC $I_{DS}-V_{GS}$ curves	51
4.3	GaN25: DC $I_{DS}-V_{DS}$ curves	52
4.4	S-parameter simulation circuit	52
4.5	GaN25: Drain to source resistance as function of V_{GS} to determine R_{on}	53
4.6	GaN25: Non-linear capacitance characterization	54

4.7	LDMOS LM8: DC I_{DS} - V_{GS} curves	55
4.8	LDMOS LM8: DC I_{DS} - V_{DS} curves	56
4.9	LDMOS LM8: Drain to source resistance as function of V_{GS} to determine R_{on}	57
4.10	LDMOS LM8: Non-linear capacitance characterization	58
4.11	Options for the DRAC driver	59
4.12	Class-D off-chip driver model	60
4.13	Simplified Class-D off-chip driver model	60
4.14	Inter-chip connection by bondwires	62
4.15	Bode plots for bondwire and flip-chip based driver models	63
4.16	Inter-chip connection using flip-chip bonding	64
4.17	Step responses of the driver model	64
4.18	Push-pull Class-E SMPA DRAC	68
5.1	Initial simulated time domain waveforms	72
5.2	Initial simulated load-line	73
5.3	Output power as function of q and K_C	74
5.4	Time domain waveforms for $q = 0,99$ and $K_X = 0$	74
5.5	Load-lines for $q = 0,99$ and $K_X = 0$	75
5.6	Output power as function of TGW	75
5.7	LDMOS output power and efficiency as function of rise and fall time	77
5.8	LDMOS baseline AM-AM curve and INL	78
5.9	Illustrating the individual transistor sizes	80
5.10	LDMOS output power and efficiency as function of τ_d	81
5.11	Activated TGW versus ACW illustrating non-linear segment sizing	82
5.12	Time domain V_{GS} waveforms for $\tau_d=17$ pF Ω	82
5.13	AM-AM linearity of implemented DRAC with $N_b = 9$ bit, using a simple driver model without series inductance	83
5.14	Power and efficiency of implemented DRAC with $N_b = 9$ bit at $f_c = 3$ GHz	84
5.15	EM model of flip-chip bond	85
5.16	Time domain V_{GS} waveforms using EM bond models	85
5.17	Time domain V_{GS} waveforms for $\tau_d=20$ pF Ω and $L_{bond}=24,8$ pH	86
5.18	AM-AM linearity of implemented DRAC with $N_b = 9$ bit, using $L_{bond} = 24,8$ pH	87
5.19	Effect of L_{bond} on V_{GS} , power and efficiency	87
5.20	Effect of glitches on AM-AM and AM-PM	88
5.21	AM-AM and AM-PM curves after applying multiphase RF clocking	90
5.22	Flip-chip setup illustrating inshing and distance requirements for wirebonding	91
5.23	LDMOS layout version 1	92
5.24	CMOS layout version 1	93
5.25	S-parameters from farthest input to output from 0 GHz to 48 GHz	94
5.26	Phase differences due to the layout with respect to the smallest transistor	94
5.27	LDMOS layout version 2	95
5.28	CMOS layout version 2	95
5.29	3D layouts of flip-chip connections, bump pads and required interconnect to ESD capacitances and CMOS drive transistors	96
5.30	Post-layout AM-AM linearity of implemented DRAC at $f_c = 3$ GHz	98
5.31	Post-layout power and efficiency of implemented DRAC at $f_c = 3$ GHz	99
6.1	Full RF-DAC	102
6.2	Class-D off-chip driver model with CMOS input and output capacitances	103
6.3	Simplified Class-D off-chip driver model with input and output capacitances	103
6.4	Thick oxide 270 nm CMOS propagation delays	104
6.5	The relationship between propagation delays and rise and fall times	105
6.6	An inverter chain of length N scaled by a factor f	106
6.7	Designed OTA with enhanced output swing	107
6.8	OTA DC output characteristics	108
6.9	OTA transient waveforms for $V_{DDh} = 2,0$ V	109

6.10 Complete output chain circuit diagram	109
6.11 Complete output chain transient simulation	109
6.12 Supply decoupling circuit diagram	110
6.13 Harmonic content of CMOS power supply voltage	111
6.14 DRAC output power and DC power, CMOS driver power and QAM256 output probability vs. ACW	113
A.1 Pre-layout, with sub-optimum AM-AM linearization	124
A.2 Post-layout harmonic distortion for $Q_L = 5$	125
A.3 Single-ended Class-E circuit diagram modified with mutual inductance	126
A.4 Effect of mutual inductance post-layout on output power and DC power consumption	126
B.1 LDMOS baseline AM-AM and AM-PM curves for $V_{in} = 6\text{ V}$	130
B.2 Power and efficiency corresponding to LDMOS baseline for $V_{in} = 6\text{ V}$ at $f_c = 3\text{ GHz}$, push-pull $TGW = 50\text{ mm}$, and the design set as shown in table B.1	130
C.1 GaN baseline AM-AM and AM-PM curves	134
C.2 Power and efficiency corresponding to GaN baseline for $V_{in} = 6\text{ V}$ at $f_c = 1\text{ GHz}$	134

List of Tables

2.1	Different subclasses of Class-E and their design set [26, 27]	31
4.1	Comparison between GaN25 and LDMOS LM8	57
4.2	Settling time of the driver models	65
4.3	Values of the digital input codes ACW_i for segments $d_{ACW,1}$ and $d_{ACW,2}$ for $k = 0, 1, \dots, 63$	67
5.1	Design sets and values for $TGW = 1$ mm	73
5.2	Design set and values for $TGW = 50,4$ mm (push-pull) and $f_c = 3$ GHz	76
5.3	Determining segment scaling, first iteration	79
5.4	Scaling factors	81
5.5	Sizes of smallest and largest transistors and their drive strength	82
5.6	Multiphase RF clocking delays	90
5.7	Simulated parasitic component values including interconnect	96
5.8	Design set and values for post-layout $TGW = 51,0$ mm (push-pull) and $f_c = 3$ GHz	97
5.9	Scaling factors post-layout	97
5.10	Multiphase RF clocking delays post-layout	97
6.1	System power consumption contributor and efficiency comparison at 3 GHz	114
7.1	Simulation based power efficiency comparison with measurements from literature	119
A.1	Ideal switch Class-E quasi load insensitive quasi parallel-circuit design set	123
A.2	OTA transistor widths and bias voltages	127
B.1	Design set and values when using $V_{in} = 6$ V, with $TGW = 50,0$ mm (push-pull) and $f_c = 3$ GHz	131
C.1	Design set and values with $TGW = 18,9$ mm and $f_c = 1$ GHz	135

1

Introduction

Sharing information is a basic human need. As means of direct sharing we use verbal communication and therefore we developed spoken language. The urge to share indirectly, even to unknown people, was satisfied by written communication. In prehistoric times this was done by cave murals, for anyone to be seen who might stumble upon such cave. A portable way of written communication was scratching pictographs in stone tablets. These pictographs eventually became standardized of which Egyptian hieroglyphs are a good example. Better portability was introduced by writing on papyrus. Written communication continued to evolve, the primal pictographs evolved into an alphabet, papyrus was replaced by paper and from around 1440 we had the possibility to use a printing press instead of having to write everything manually.

It was only in 1894 that Guglielmo Marconi laid the fundamentals for the next step in sharing information: wireless communication¹. Heinrich Hertz had previously proved the existence of the Electro-Magnetic (EM) waves theorized by James Clerk Maxwell. Marconi showed that information could be transferred by controlling Radio Frequency (RF) EM waves using a spark-gap transmitter. This transmitter was patented by Marconi in 1897 and later exploited commercially by “Marconi’s Wireless Telegraph Company”. These spark-gap transmitters used a large bandwidth, had low information content and a poor power efficiency [1].

Wireless communication has known incredible development since then. Important inventions that made this possible were the triode vacuum tube by Lee de Forest in 1906, the regenerative receiver in 1913, the super-heterodyne receiver during World War I and Frequency Modulation in 1933 by Edwin H. Armstrong, the invention of the transistor by William B. Shockley, John Bardeen and Walter H. Brattain in 1947, and the invention of the Integrated Circuit (IC) by Jack Kilby in 1958. In the early 1980s the first generation of ‘modern’ wireless telephone technology was launched, to be replaced in the 1990s by the digital second generation of mobile telecommunications, in Europe known as GSM and still in use today.

Of course developments did not stop there. Today we have access to high speed mobile networks of the fourth generation (4G) and we expect to be wirelessly connected at all times and places for our social interactions. I am no exception to that, as anecdotally illustrated in figure 1.1. This is a screenshot of my smartphone, connected to a 4G+ network² with a few recent unread messages of the popular instant messaging service WhatsApp and another application literally requesting me to share more.

¹It is debatable whether Guglielmo Marconi or Nikola Tesla is the father of wireless communication. Marconi’s patent was granted in England in 1897, while Tesla filed a patent in the U.S. that same year. In 1900 Marconi also filed for a U.S. patent, but the patent was granted to Tesla. However, this decision was reversed in 1904 by the U.S. Patent Office and Marconi won the Nobel Prize in 1911. In 1943—after the death of both Tesla and Marconi—the U.S. Supreme court again reversed the decision, possibly a politically biased verdict as the “Marconi Company” at that time was suing the U.S. Government for use of its patents during World War I.

²4G+ is Long Term Evolution (LTE)-Advanced with Carrier Aggregation. In this case an aggregation of a channel of 10 MHz on the 800 MHz carrier and a channel of 20 MHz on the 1800 MHz carrier, allowing for a total download rate of 225 Mbit s⁻¹.

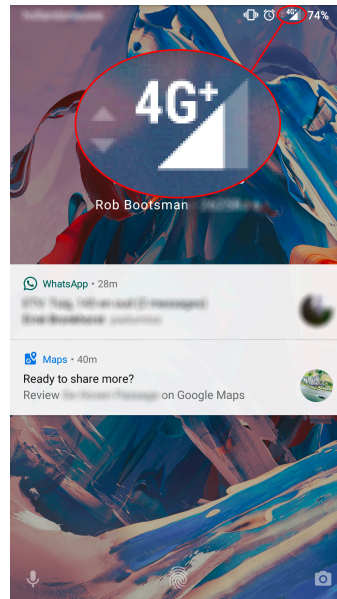


Figure 1.1: Smartphone lock-screen with push messages, illustrating today's need for connectivity

1.1. Growing Required Wireless Data Capacity

Technological development keeps continuing, allowing for more wireless capacity. Not without reason, data rate demands have been growing exponentially in the past and are expected to continue to do so [2]. As much as 1 GiB/day/user is expected by 2020 [3]. Also the amount of connected users keeps steadily increasing, of which the recent milestone of having 1 billion daily WhatsApp users is a prime example [4]. To cope with the increasing data demands the network operators have to increase their capacity. An example of this is that Vodafone has started providing peak data rates of 1 Gbit s^{-1} in some locations of the Netherlands by aggregating 4 carriers [5].

It will not stop there, as this will become normal in the fifth generation (5G) of mobile communication, which is expected by 2020. A peak data rate of 10 Gbit s^{-1} is aimed for in the specifications for 5G and a minimum data rate of 100 Mbit s^{-1} for all connected users at any given moment [6]. This implies a thousandfold increase in total wireless capacity.

1.1.1. Energy consumption

Increasing the wireless capacity by a factor thousand may not also increase the power consumption of the infrastructure by the same factor, this would be unsustainable. With current growth it is expected that by 2030 the entire ICT sector will be responsible for nearly 2% of the global greenhouse gas emissions, with a staggering amount of 1,25 Gigatonnes of carbon dioxide equivalents [7].

Not only the environmental aspect of the power consumption will be an issue, there is also an economical aspect involved. If continued at the current rate the total energy bill will increase to US\$343 billion [2]. The network operators are not too keen on paying for this out of their own pockets, for example, Orange is putting power consumption management at the heart of its corporate strategy to prevent its annual power bill reaching the US\$1 billion mark by 2020 [9]. During the nighttime there is a lot less wireless traffic, which was recognized by Tele2. They have admitted that they are powering down their 2600 MHz base stations during the nighttime, unless they observe increased data demand sustained for more than 2 minutes [8].

That the network operators are not willing to pay for it does not mean the average customer will. Not only in the economical sense, but also in terms of convenience. The base stations are not the only components in the network that will demand more energy with increased data speeds, the receivers will as well. As can also be seen in figure 1.1, right next to the 4G+ icon there is a battery indicator. Any increase in power consumption will lead to a shorter lasting battery. This is especially important for the expected advance of Internet of Things (IoT) devices, as many of these devices are expected to operate in the order of a few years on a single battery charge.

1.2. RF Transmitter Architectures

All information in electronic communication systems is in the form of binary data. This binary information is modulated by a transmitter (TX) to a *bandpass* RF signal, ready to be transmitted. This modulated bandpass RF signal $s(t)$ can be represented by [10]:

$$s(t) = \Re \{g(t)e^{j\omega_c t}\} \quad (1.1)$$

where $f_c = \omega_c/2\pi$ the carrier frequency and $g(t)$ the complex *baseband* envelope of the modulated signal. We can say that *the complex function $g(t)$ modulates the phasor $e^{j\omega_c t}$* . Because $g(t)$ is complex, its instantaneous value can be represented in the complex plane as shown in figure 1.2.

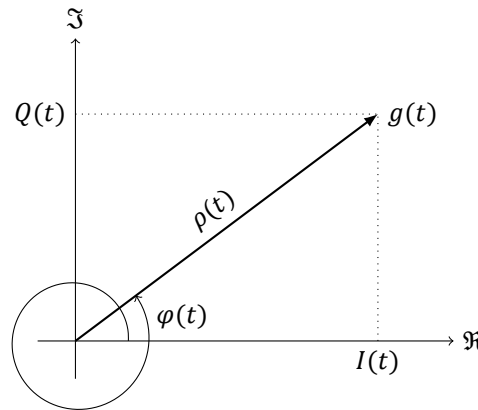


Figure 1.2: Representation of $g(t)$ in the complex plane

This instantaneous baseband value can be represented by Cartesian and polar coordinates:

$$g(t) = I(t) + jQ(t) \equiv \rho(t)e^{j\varphi(t)} \quad (1.2)$$

Using these representations, the modulated signal $s(t)$ can be rewritten as:

$$s(t) = \rho(t) \cos[\omega_c t + \varphi(t)] \quad (1.3)$$

$$= I(t) \cos(\omega_c t) - Q(t) \sin(\omega_c t) \quad (1.4)$$

where $\rho(t)$ represents the Amplitude Modulation (AM) and $\varphi(t)$ the Phase Modulation (PM) of the RF signal. Similarly, $I(t)$ represents the in-phase signal and $Q(t)$ the quadrature signal. The polar signals $\rho(t)$ and $\varphi(t)$ can be best used to describe what is happening at circuit level in terms of currents and voltages [11], while the Cartesian signals best correspond to the digital baseband information.

As there are two ways of describing the modulated RF signal, there are also two general transmitters: Cartesian or I/Q transmitters and polar transmitters. Traditionally these transmitters are implemented in an analog fashion. These types of transmitters are discussed in section 1.2.1, as well as the two modulation techniques. It is also possible to implement these transmitters digitally, as discussed in section 1.2.2.

1.2.1. Analog RF TX Architectures

The polar coordinates are defined and calculated as follows:

$$\rho(t) \triangleq |g(t)| = \sqrt{I^2(t) + Q^2(t)} \quad (1.5)$$

$$\varphi(t) \triangleq \angle g(t) = \tan^{-1} \left(\frac{Q(t)}{I(t)} \right) + 2n\pi \quad (1.6)$$

which result in the following reverse computations for the Cartesian coordinates:

$$I(t) = \Re \{g(t)\} \equiv \rho(t) \cos[\varphi(t)] \quad (1.7)$$

$$Q(t) = \Im \{g(t)\} \equiv \rho(t) \sin[\varphi(t)] \quad (1.8)$$

bandwidth
expansion

As can be seen in these equations, the conversion between I/Q and ρ/θ is highly non-linear. The I/Q notation for the modulated passband signal, equation (1.4), is fairly straightforward. As the signal summation is linear, the analysis of the frequency content of both the baseband and bandpass signal are relatively easy. More importantly, as the baseband conversion from Cartesian to polar is nonlinear it implies that the polar signals require larger bandwidth, called *bandwidth expansion* [12]. The baseband I/Q signals require the same bandwidth f_{BB} as the RF output signal.

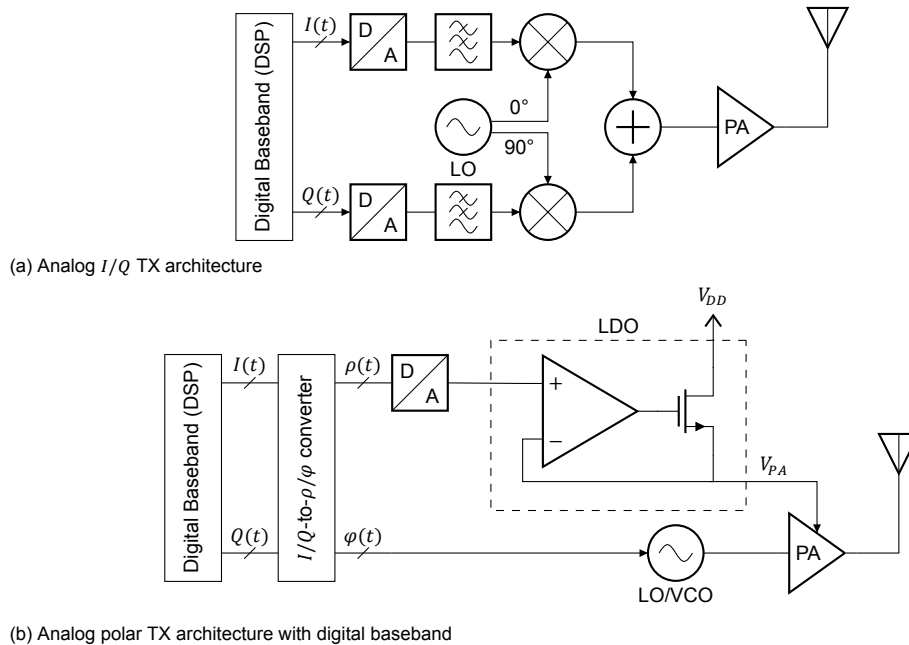


Figure 1.3: Examples of analog modulation architectures

Analog I/Q TX

Most likely due to their smaller required baseband bandwidth and the relatively easy frequency domain analysis, the I/Q transmitters have been the de-facto industry standard for transmitters requiring arbitrary amplitude and phase. These transmitters usually have an architecture similar to what is shown in figure 1.3a. First the $I(t)$ and $Q(t)$ signals are generated by a Digital-to-Analog Converter (DAC), then filtered if necessary and up-converted to RF. This up-conversion requires a Local Oscillator (LO) with a 0° and 90° offset. After up-conversion these signals are (linearly) summed and offered to the input of a Power Amplifier (PA), boosting the RF signal power up to the required transmit levels.

Analog Polar TX

This type of transmitter operates by modulating the amplitude and phase separately. Then this information is recombined in, for example, the PA. This architecture is shown in figure 1.3b. Here the LO functions as Phase Modulator (e.g. by a Voltage Controlled Oscillator (VCO)). Amplitude Modulation is achieved by modulating the PA's supply voltage, hence a non-linear PA can be used with excellent power efficiency. With the increasing awareness of energy consumption this is a big advantage. However, special attention must be paid to the timing alignment of the amplitude and phase signal paths to prevent distortion. Furthermore this architecture is unfit for wideband signals due to aforementioned bandwidth expansion.

1.2.2. Digital-Intensive RF TX Architectures

Silicon Complementary Metal-Oxide-Semiconductor (CMOS) technology underwent huge improvements and financial cost optimization over time. A general trend observed in 1965 by Gordon Moore was the optimum number of transistors in an IC for the lowest prize per transistor would roughly double every year [13]. Nowadays this observation, best known as Moore's Law, proved somewhat a self-fulfilling prophecy. Transistor gate lengths shrink by a factor $\sqrt{2}$ every two years, effectively halving

required device area and hence doubling the number of devices on a single IC. This so called *technology scaling* causes lower cost per device, lowering intrinsic delay (increased speed) and lowering supply voltages [14]. The lower supply voltages mainly benefits digital circuitry because of decreased power consumption, however this causes an increased power consumption for analog circuitry whilst keeping other performance similar [15].

technology scaling

This gave rise to the concept of *digital RF*: instead of conveying information in the signal amplitude, the information is transferred in the time-domain and in digital codewords. With technology scaling the time-domain resolution keeps improving. Meanwhile the voltage resolution keeps reducing due to lowering supply voltages, up to the point that it can be stated that the time-domain resolution of digital transitions, in combination with the additional information provided by digital codewords, in highly advanced silicon CMOS is superior to the voltage resolution of analog signals in the same process [16].

digital RF

Using this concept the analog components can be replaced by digital counterparts, allowing the modulators to be implemented in advanced CMOS in an energy efficient manner. An additional advantage of digital electronics is the high reconfigurability compared to analog electronics, allowing for, e.g., adjustable (digital) filters and Digital Pre-Distortion (DPD).

Digital Polar TX

A digital-intensive polar transmitter is quite similar to its analog counterpart. Amplitude and phase are modulated separately and then recombined as shown in figure 1.4. This recombining is not done in a PA, but in a Digital-to-RF-Amplitude Converter (DRAC). This is a Digital-to-Analog Conversion directly at the required RF output frequency, thus effectively a DAC and mixer combined. This DRAC acts like an amplitude modulator by supplying it with a so called Amplitude Code Word (ACW). The phase modulated signal functions as input clock of the DRAC, which is generated by for example a Digitally Controlled Oscillator (DCO). The phase modulator and DRAC together form a Radio Frequency Digital-to-Analog Converter (RF-DAC). Before offering the RF-DAC output to an external PA the signal is optionally filtered—if the spectral purity of the RF-DAC is not sufficient—by for example Surface Acoustic Wave (SAW)-filters [17].

This digital architecture suffers from the same bandwidth expansion as the analog polar architecture, as well as the required timing alignment between the amplitude and phase signal paths. A drawback is that it loses the energy efficiency advantage of analog polar transmitters in the final stage due to external PA. However the power consumption of the digital modulator can improve over time due to technology scaling and can be superior to that of an analog modulator. That improves overall system power efficiency at low required output powers, which can be beneficial in for example IoT-applications.

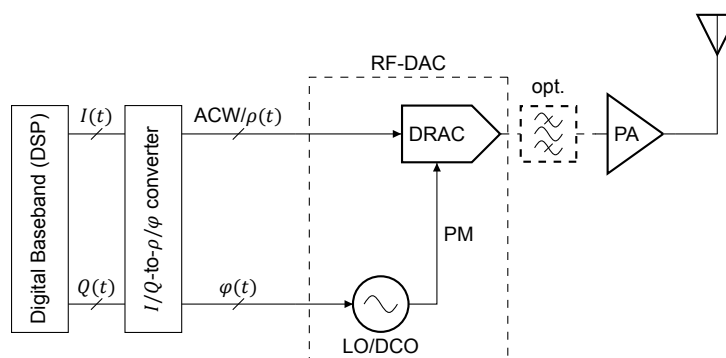


Figure 1.4: Digital polar TX architecture

Digital I/Q TX

It is also possible to make a digital intensive *I/Q* transmitter, which again is quite similar to its analog counterpart. Its general architecture is shown in figure 1.5. The $I(t)$ and $Q(t)$ signals are generated by two DRACs directly at the RF frequency. Just as with the digital polar transmitter the DRAC functions both as DAC and as mixer. It requires a LO with a 0° and 90° offset. The DRAC outputs are summed and then fed to an external PA. It does not suffer from the bandwidth expansion the digital polar architecture

suffers from, generally resulting in better linearity than the digital polar architecture. However, generally the power consumption is higher than that of a polar architecture.

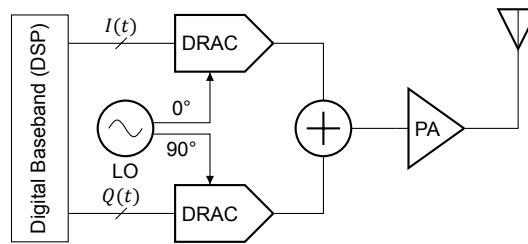


Figure 1.5: Digital I/Q TX architecture

1.3. Thesis Goals

As mentioned in section 1.2.2 in the digital polar transmitter, the transmitter loses efficiency due to the external PA which was not the case in the analog polar transmitter. But what if we would be able to combine the PA and RF-DAC? Would this improve overall power efficiency? This results in the following research question:

“Is it possible to design a power RF-DAC, in such a way that this RF-DAC also functions as the PA, with good power efficiency?”

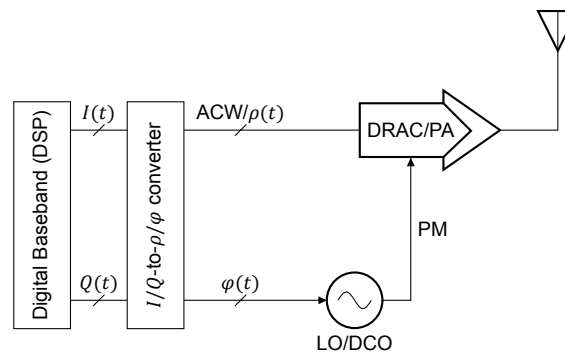


Figure 1.6: Combined RF-DAC and PA TX architecture

The goal of this thesis is to answer above question by designing a power DRAC as core component of the intended power RF-DAC. Its basic architecture is shown in figure 1.6. As the intention is to be able to connect the power RF-DAC directly to an antenna without additional filtering, its output signal should have good spectral purity, thus putting constraints on the linearity of the RF-DAC. The choice for a digital polar architecture is made as it will have better power efficiency, as is the case for the analog polar modulator.

Several technologies are available to make the components of the power RF-DAC. Two high power analog technologies are available, being Fraunhofer IAF Gallium Nitride (GaN) 250 nm and Ampleon Laterally Diffused Metal-Oxide-Semiconductor (LDMOS) LM8. These are most likely suitable for the power DRAC design. Also available is a high integration digital technology, TSMC CMOS 40 nm 1P7M4X1Z1U, which could be used to do the Digital Signal Processing (DSP) and include the phase modulator. Designs for these on CMOS are already available within the Electronics Research Laboratory (ELCA) research group and are therefore outside this thesis scope.

Taking in above considerations the following objectives are set for this thesis:

- The RF-DAC targets (5G) mobile cellular application, hence the operational frequency of the power RF-DAC should be in the Ultra High Frequency (UHF) range: 617 MHz to 2690 MHz and/or 3300 MHz to 5000 MHz [18]

- The RF-DAC's operational frequency should be agile and only dependent of external components, such as the applied matching network at the output
- Select semiconductor technologies to implement the power RF-DAC in
- Depending on the selected technology the targeted peak output power of the DRAC has to be determined

If Fraunhofer IAF GaN 250 nm is selected the targeted peak output power is 50 W to 100 W

If Ampleon LDMOS LM8 is selected the targeted peak output power is 10 W to 20 W

- Select a suitable circuit topology to implement the high power RF-DAC
- Design a power DRAC

The designed DRAC should have high drain efficiency

The designed RF-DAC should have—within the constraints of the available circuitry—a reasonable system power efficiency, preferably as high as possible

The designed RF-DAC's linearity should allow for ACPR levels better than 50 dBc

1.4. Thesis Organization

The thesis is structured as follows:

In part I the definitions and theoretical groundwork is laid out. This starts with discussing Power Amplifier theory, the used model assumptions and the operating principles of the different PA classes in chapter 2. In chapter 3 the operation of a DRAC is discussed, starting with the basics from a DAC. In its last section the theory of PAs and DRACs is combined to make the first step towards a power DRAC. Readers that have experience with these topics may skip these chapters, although it will help the understanding of the rest of the thesis as all relevant definitions are made within these chapters.

In part II the implementation of the RF-DAC design is discussed. Part II starts with chapter 4 where the different design choices during the design process are established, including choice of semiconductor technologies, models for the DRAC driver and resulting circuit topology. Due to the limitations of the technologies the choice will be made to implement the DRAC in LDMOS and a dedicated driver in CMOS. With the choices made the actual design steps and simulation results are discussed in chapter 5. At the end of this chapter a suitable layout for the DRAC is proposed.

In chapter 6 the first groundwork for a CMOS driver design is laid out, making it reasonable that a design compatible with the LDMOS DRAC is possible, but the full CMOS driver design lies outside this thesis scope. With the simulation results from the CMOS driver it is possible to make estimates regarding the system performance and these are discussed in section 6.5. The technology that is used to implement the presented power RF-DAC is not optimized for this purpose, therefore the resulting system performances can be improved by technology development. Estimates on how the technology could be used in a different way and resulting system performances are discussed as well.

Lastly, in chapter 7 the thesis conclusions are presented, together with a discussion on the results and recommendations for future work.



Theory

2

PA Basics

In this chapter the operation principles of PAs are discussed. First in section 2.1 the core performance metrics regarding PAs are defined. Then in section 2.2 the power flows at different frequencies are defined and relationships between them described. Finally section 2.3 starts by defining the model used for the active device used in a PA, after which the different classes of PA operation are described using the active device model.

2.1. Definitions

By its very nature, an amplifier is a three-port device that converts Direct Current (DC) power to an output with a reasonably similar power spectrum of its input signal [19, 20]. In the case of a RF PA, the output is typically connected to an antenna. The antenna can be characterized by a loading impedance Z_{ant} , which is designed to be resistive at the (bandpass) design frequencies. Hence, at these frequencies, the antenna can be represented by a single load resistor R_L . The power dissipated by R_L is, by definition, equal to the power of the EM wave radiated by the antenna [21]. This load resistance is usually 50Ω , but it can be converted to a higher or lower value to be seen by the active device of the PA. Note that the bandpass signal is assumed to be *narrowband*, meaning that the signals present in the circuit can be assumed to effectively consist of only sinusoids of the fundamental frequency and its harmonics [20]. In this thesis, the formal PA output quantity is assumed to be the voltage V_{out} , although assuming it to be a current delivered to an ohmic load would be just as correct.

narrowband

In the next sections the core performance metrics for a PA will be defined, which are power, gain, efficiency, linearity and the Peak-to-Average Power Ratio.

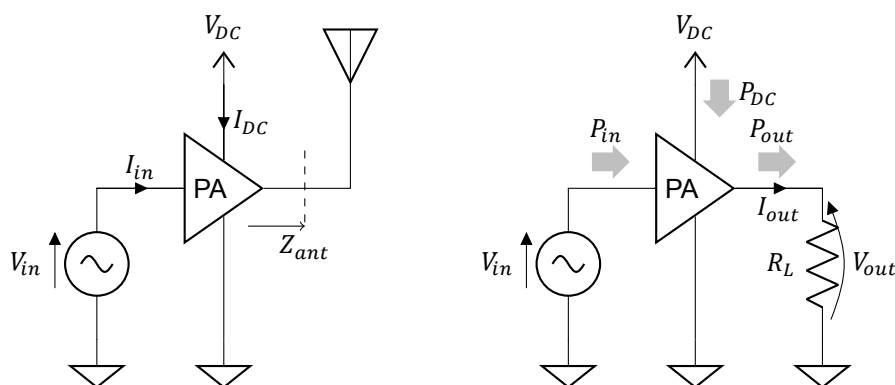


Figure 2.1: Power flows in a general PA

2.1.1. PA Power

There are three important power flows within a PA, these being the input power P_{in} , the output power P_{out} and the supply power P_{DC} . These power flows are illustrated in figure 2.1.

instantaneous
output power

The *output power* is defined as the power delivered by the PA and flowing into the load (antenna) and varies over time. The *instantaneous output power* is defined as

$$p_{out}(t) \triangleq V_{out}(t) \cdot I_{out}(t) \quad (2.1)$$

time average
output power

The *time average output power* is defined as

$$P_{out} \triangleq \langle p_{out}(t) \rangle \equiv \lim_{\tau \rightarrow \infty} \int_{-\tau/2}^{\tau/2} p_{out}(t) dt \quad (2.2)$$

where $\langle \cdot \rangle$ is the time average operator. In case of a periodic output with period T the equation simplifies to

$$P_{out} = \frac{1}{T} \int_{-\tau/2}^{\tau/2} p_{out}(t) dt \quad (2.3)$$

Assuming a resistive load this gives

$$P_{out} = \frac{\langle V_{out}^2(t) \rangle}{R_L} = \frac{V_{out,RMS}^2}{R_L} \quad (2.4)$$

RMS where the *Root Mean Square (RMS)* voltage is defined as

$$V_{out,RMS} \triangleq \sqrt{\langle V_{out}^2(t) \rangle} \quad (2.5)$$

input power
supply power

In a similar way, equations can be defined for *input power* and *supply power*

$$P_{in} \triangleq \langle p_{in}(t) \rangle = \langle V_{in}(t) \cdot I_{in}(t) \rangle \quad (2.6)$$

$$P_{DC} \triangleq \langle p_{DC}(t) \rangle = \langle V_{DC}(t) \cdot I_{DC}(t) \rangle \quad (2.7)$$

Since a PA converts power between different frequencies it might happen that power gets spread across the frequency spectrum. Typically only the power of the modulated bandpass around the carrier frequency f_c is useful and any additional harmonics should be filtered away. Therefore it is useful to define the *fundamental average output power*

fundamental
average output
power

$$P_{out,f_c} = \frac{V_{out,f_c}^2}{2R_L} = \frac{V_{out,f_c,RMS}^2}{R_L} \quad (2.8)$$

where V_{out,f_c} is the amplitude of the sinusoidal output voltage at f_c .

2.1.2. Gain

The gain of a PA can be defined as the dimensionless ratio between the useful output power, the fundamental average output power P_{out,f_c} , and the required input power P_{in} to drive the PA. This ratio is usually expressed in dB and is defined as

$$G_{P,dB} \triangleq 10 \log_{10} \left(\frac{P_{out,f_c}}{P_{in}} \right) \quad (2.9)$$

ratiometric and
slope power gain

where G_P is the gain, more specifically the *ratiometric power gain*.

Another metric of power gain is the *slope power gain* G_{SP} which is defined as

$$G_{SP,dB} \triangleq 10 \log_{10} \left(\frac{dP_{out,f_c}}{dP_{in}} \right) \quad (2.10)$$

Both gain definitions are large signal quantities. As long as the amplifier operates linearly these two power gain measures are equal to each other and the small signal gain [19]. Both gains are illustrated in figure 2.2

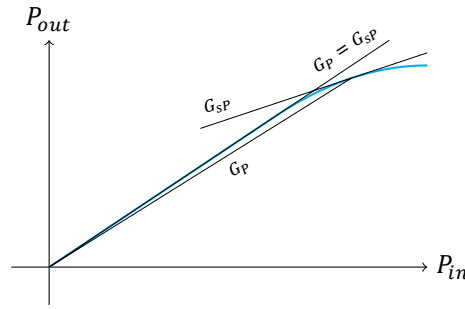


Figure 2.2: Different power gain measures

2.1.3. Efficiency

Conservation of power dictates that the amount of power that comes out of a system should always be less or equal to the amount of power that was put in, with the difference being dissipated in the form of heat or by undesired parasitic radiation of EM waves. An ideal amplifier would require negligible input power and would convert all DC supply power to useful output power, basically generating no heat. However, as practical PAs are not ideal, it is useful to define some measures of power efficiency.

First it is useful to see how well a PA output stage converts the DC supply power into useful output power. This is simply the ratio between the useful generated output power and the consumed DC supply power by the active device, called the *drain* or *collector efficiency*

drain efficiency

$$\eta_D \triangleq \frac{P_{out,fc}}{P_{DC}} \tag{2.11}$$

Perhaps a more honest way of expressing PA efficiency is by taking the ratio of the useful output power and all power put into the PA together, called the *total efficiency*

total efficiency

$$\eta_T \triangleq \frac{P_{out,fc}}{P_{DC} + P_{in}} \tag{2.12}$$

This can include also any amount of driver stages required to drive a PA and defines the *total chain* or *system efficiency*

total system efficiency

$$\eta_{T,system} \triangleq \frac{P_{out,fc}}{P_{DC} + \sum_{i=0}^n P_{DC,i} + P_{in}} \tag{2.13}$$

in which case P_{in} is the input power of the first driver stage and $P_{DC,i}$ the DC power consumption of the i^{th} stage in the chain.

Lastly, a commonly used power efficiency figure of merit in the RF and microwave field is the *Power Added Efficiency (PAE)*

PAE

$$PAE \triangleq \frac{P_{out,fc} - P_{in}}{P_{DC}} \tag{2.14}$$

This equation suggests that part of the output power is physically coming from the input power, which is not true for most, if not all, modern PAs.

It is common to display an efficiency measure as percentage.

2.1.4. Linearity

As the whole point of a PA in a TX system is to boost the power of a modulated bandpass signal to power levels required to transmit, it would be nice if the PA output signal is an exact linearly scaled replica of the input signal. The extent to which that is true is referred to as *linearity*. The accuracy in which the signal is reproduced can be split in to two types: amplitude linearity and phase linearity. Any *non-linearity* is referred to as *distortion*.

linearity

non-linearity distortion

In systems where only Phase Modulation (PM) is applied (constant envelope systems) the only linearity one needs to take care of is the Input Phase to Output Phase conversion (PM-PM) distortion. Similarly for Amplitude Modulation (AM) systems (constant phase systems) this is Input Amplitude to

Output Amplitude conversion (AM-AM) distortion. When both AM and PM are applied in one system also Input Amplitude to Output Phase conversion (AM-PM) distortion comes into play.

Modern communication systems require both amplitude and phase modulation. The two foremost distortion mechanisms in a polar modulator are AM-AM distortion and AM-PM distortion [21]. These distortion profiles are static approximations and are usually shown by plotting the AM-AM, AM-PM etc. profiles. The static approximation is attained by (periodic) steady-state values. Ideally the AM-AM curve should be a straight line through the origin (constant derivative, or constant power gain) and the AM-PM curve should be flat (zero derivative).

Both AM-AM and AM-PM curves can change in a dynamic setting due to environmental effects, such as temperature or supply voltage changes due to fluctuating bias current. An example could be that the PA had high output power causing a voltage (IR) drop and increased temperature, followed by a lower output power. Then it takes some time for the voltage to return to 'normal' and the amplifier to cool to nominal value, hence affecting the accuracy of the lower power signal. These effects are referred to as *memory effects*, since what happened in the past affects the signal in the future.

Multiple measures for quantifying linearity are used. Looking at a typical AM-AM curve initially the relation between input and output power is linear, indicating constant gains $G_P = G_{SP}$. When the PA starts to saturate non-linearity occurs, which causes the gains to diverge. The output power where G_P is 1 dB lower than its value in the linear region is called the *1 dB compression point* P_{-1dB} .

Perhaps more important are the non-linearities that become apparent in the frequency spectrum. The power outputted at integer multiples of the carrier frequency compared to the power at the carrier itself is referred to as *harmonic distortion* and is often expressed in dB

$$HD_{k,dB} \triangleq 10 \log_{10} \left(\frac{P_{out,k \cdot f_c}}{P_{out,f_c}} \right) = 10 \log_{10} \left(\frac{V_{out,k \cdot f_c}^2}{V_{out,f_c}^2} \right) \quad (2.15)$$

where $V_{out,k \cdot f_c}$ is the amplitude of the sinusoidal output voltage at $k \cdot f_c$. Most significant are usually *Second order Harmonic Distortion (HD2)* and *Third order Harmonic Distortion (HD3)*. The total power in the harmonics summed is referred to as *Total Harmonic Distortion (THD)* and is expressed as percentage

$$THD \triangleq \frac{\sum_{i=2}^{\infty} P_{out,i \cdot f_c}}{P_{out,f_c}} = \frac{\sqrt{\sum_{i=2}^{\infty} V_{out,i \cdot f_c}^2}}{V_{out,f_c}} \quad (2.16)$$

These harmonics are usually unwanted and filtered to suppress them at the output, for that reason HD2 and HD3 might give slightly optimistic values for how linear the PA actually is. Another way is by specifying the *Third order Intercept Point (IP3)* in a *two-tone test*. This is done by applying two frequencies at the input with equal amplitude, with tone spacing f_{ts} centered around the carrier frequency, and observing the *Third order InterModulation distortion (IM3)* at the output. With every 1 dB increase in signal amplitude the IM3 products increase by 3 dB. The virtual point where the output IM3 product power would be equal to the output power of the desired frequencies is defined as the IP3. It can be specified as the output power or as the input power to reach this point, being respectively referred to as OIP3 and IIP3. These measures should be more or less independent from the tone spacing f_{ts} and the carrier frequency f_c . A signature of memory effects is that the IM3 products actually do change with f_{ts} and f_c [22].

Complex modulated signals, however, are far from similar to a two-tone test. Hence in PA specifications for communication purposes other measures—related to the specifications of the targeted standard—are used. These measures are *Adjacent Channel Power Ratio (ACPR)* and *Error Vector Magnitude (EVM)*. ACPR, rather self explanatory, is the ratio of power outputted in an adjacent channel one signal bandwidth away and the power outputted in the main channel where both channel bandwidths are equal and is expressed in dBc. EVM is the accuracy in which symbols are transmitted within the main channel itself. As explained in section 1.2 and illustrated in figure 1.2 a symbol can be represented in the complex plane. All possible symbols together make up the *constellation diagram* where a single symbol is referred to as a *constellation point*. The error vector is defined as the difference between the actual transmitted constellation point $\hat{g}[k]$ and the ideal constellation point $g[k]$ of the k^{th} symbol

$$\varepsilon[k] = \hat{g}[k] - g[k] \quad (2.17)$$

memory effect

compression point

harmonic distortion

HD2, HD3
THD

intercept point

intermodulation

ACPR

EVM

constellation
diagram
constellation point

Note that the value of g represents a complex voltage. The EVM can be defined as

$$EVM[k] = \sqrt{\frac{|\varepsilon[k]|^2}{N^{-1} \sum_{i=1}^N |g[i]|^2}} \quad (2.18)$$

The RMS value of the EVM is defined as

$$EVM_{RMS} = \sqrt{\frac{\sum_{i=1}^N |\varepsilon[i]|^2}{\sum_{i=1}^N |g[i]|^2}} \quad (2.19)$$

Usually a RMS EVM and a peak EVM, $EVM_{pk} = \max_k (EVM[k])$, are specified.

2.1.5. Peak-to-Average Power Ratio

The *Peak-to-Average Power Ratio (PAPR)* is not a fundamental property of a PA, but more of a result of the used modulation scheme. In an amplitude modulated system the output power is—per definition—not constant. The PAPR gives a measure of how much AM is present and is defined as the ratio of the peak RF output power to the average RF output power PAPR

$$PAPR = \frac{P_{out,pk}}{P_{out,avg}} \quad (2.20)$$

The PAPR is often expressed in dB

$$PAPR_{dB} = 10 \log_{10} \left(\frac{P_{out,pk}}{P_{out,avg}} \right) \quad (2.21)$$

Note that the peak output power is not the same as the peak instantaneous output power. Peak output power is also referred to as peak envelope power. In other words, when the output signal is a single tone then $PAPR_{dB} = 0$ dB.

In a multi carrier system the peak power occurs when all carrier voltage waves add up constructively, therefore the PAPR increases with the number of carriers squared, if no special measures are taken to prevent this.

Generally, lower power levels have degraded energy efficiency. That means that a higher PAPR implies a lower PA average energy efficiency.

2.2. Amplifier Power Relations

As mentioned at the beginning of this chapter (section 2.1), a general PA is a three-port with an input, DC power supply and an output. The input drives some active device. This device is in turn connected to a passive network, tying the DC power supply, output and the active device together. This is illustrated by figure 2.3.

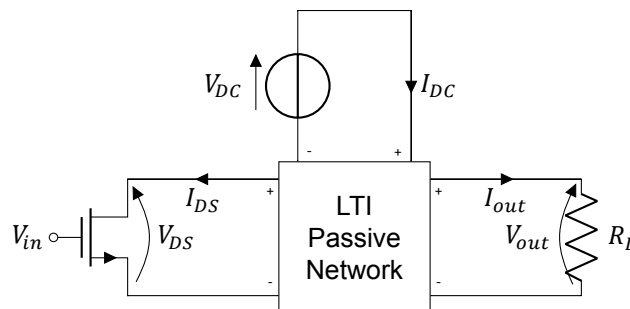


Figure 2.3: Generalized PA

The passive network is, obviously, passive, meaning that it cannot generate power. It is also assumed to be Linear Time-Invariant (LTI), which implies that it cannot convert power between frequencies. That means that the active device is the only component in the PA that can convert power from one

frequency to another. However it cannot generate power out of nothing, the physical law of conservation of energy *must* hold. That means that the output power will always be lower than, or equal to, the supplied DC power¹

$$P_{out} \leq -P_{DC} \quad (2.22)$$

Again, the law of conservation of energy dictates that the difference between these powers simply cannot disappear. This power is dissipated in the form of (unwanted) heat, P_{diss} , or in the form of unwanted EM radiation which is assumed to be negligible in this thesis. This dissipation can only occur in either the active device or in the LTI passive network, designated by respectively P_{DS} and P_{pasv} . In total the law of conservation of energy gives

$$P_{DC} + P_{DS} + P_{pasv} + P_{out} = 0 \quad (2.23)$$

All these powers have a frequency dependency. Rather self explanatory, the DC power supply can only deliver or dissipate power at DC (0 Hz), denoted by the subscript '0'. Since it is a power supply it definitely delivers power instead of dissipating it

$$P_{DC,0} < 0 \quad (2.24)$$

$$P_{DC,k \cdot f_c} = 0 \quad \forall k \geq 1 \quad (2.25)$$

where $P_{DC,0}$ is the power supplied at DC and $P_{DC,k \cdot f_c} \forall k \geq 1$ the power at any other nonzero frequency. A power without additional frequency dependent subscript is defined to be the sum of the powers at all frequencies. For that reason when P_{DC} is mentioned only $P_{DC,0}$ is implied since it is 0 for all other frequencies anyway

$$P_{DC} = \sum_{i=0}^{\infty} P_{DC,i \cdot f_c} = P_{DC,0} \quad (2.26)$$

As both the LTI passive network and output load are linear and passive

$$P_{pasv,k \cdot f_c} \geq 0 \quad \forall k \quad (2.27)$$

$$P_{out,k \cdot f_c} \geq 0 \quad \forall k \quad (2.28)$$

The device is non-linear and stable, and its input signal was assumed to be narrowband. It can therefore generate power at the fundamental and its harmonics. However, it cannot generate more power than it consumes at DC

$$P_{DS,0} \geq 0 \quad (2.29)$$

$$P_{DS,k \cdot f_c} \leq 0 \quad \forall k \geq 1 \quad (2.30)$$

$$\sum_{i=0}^{\infty} P_{DS,i \cdot f_c} \geq 0 \quad (2.31)$$

$$\therefore \sum_{i=1}^{\infty} P_{DS,i \cdot f_c} \geq -P_{DS,0} \quad (2.32)$$

Combining equations (2.22) and (2.23) gives

$$P_{diss} = -P_{DC} - P_{out} = P_{DS} + P_{pasv} \quad (2.33)$$

The fundamental average output power P_{out,f_c} was defined at equation (2.8) as the useful output power. All output power at other frequencies and dissipated powers over all frequencies will either cause useful output power P_{out,f_c} to lower or consumed DC power $-P_{DC}$ to increase, and therefore lowering power efficiency

$$P_{out,f_c} = -P_{DC} - \underbrace{\left(\sum_{i=0}^{\infty} P_{DS,i \cdot f_c} + P_{pasv,i \cdot f_c} \right)}_{P_{diss}} - \underbrace{\left(P_{out,0 \cdot f_c} + \sum_{i=2}^{\infty} P_{out,i \cdot f_c} \right)}_{P_{out} - P_{out,f_c}} \quad (2.34)$$

¹consumed power is positive, generated power negative

2.3. Amplifier Classes

The active device forms the core of a PA: without it there is no conversion of power between frequencies. By using different topologies of the LTI passive network and/or different biases for the active device it is possible to change the behavior of the PA. Therefore it is first important to understand how this device behaves under different circumstances and what its limitations are. After that different classes of PAs will be discussed divided in two groups, the group of transconductance amplifier classes and the group of switching amplifier classes.

2.3.1. Device Model

Today's active devices are implemented in solid-state materials (in contrast to earlier vacuum tubes) and generally referred to as *transistors*. For this thesis only Field-Effect Transistors (FETs) and their related device models are considered.

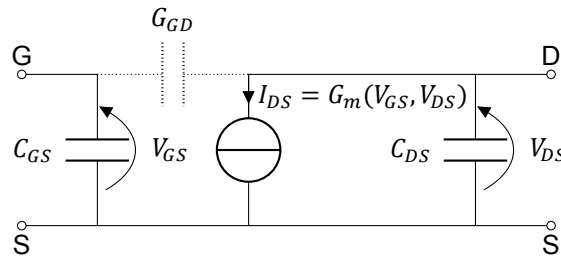


Figure 2.4: Simple FET model

In general, a transistor can be seen as a two-port, inside of which a voltage controlled current source can be found with surrounding parasitics due to the way it is implemented. In figure 2.4 a relatively simple FET model is shown with some common parasitics. The voltage V_{GS} between the gate (denoted by G) and the source (S) is the transistor input and acts as control voltage for the output current source. The output of the transistor is the current I_{DS} that can be found at the drain (D) of the transistor flowing to the source. The voltage between the drain and source is the transistor output voltage V_{DS} . The output current is a function of mainly the input voltage V_{GS} but also the output voltage V_{DS} . In general the the large signal output transconductance can be defined by

$$G_m \triangleq \frac{I_{DS}}{V_{GS}} \quad (2.35)$$

Multiple *operation regions* can be recognized. First is the *off-state*, when the input voltage V_{GS} is below the *threshold voltage* V_T . In the off-state negligible current will flow through the transistor: $I_{DS} \approx 0$ A. Next are the *resistive* or *triode region* when $V_{GS} - V_T > V_{DS}$, and the *saturation* or *transconductance region* when $V_{GS} - V_T < V_{DS}$. In these regions the output current is a function of V_{GS} and V_{DS} and can be given by [23]

$$I_{DS} = \frac{\mu_n C_{ox} W_g}{L_g} \left[(V_{GS} - V_T) V_{min} - \frac{V_{min}^2}{2} \right] \quad (2.36)$$

where the product $\mu_n C_{ox}$ is a process specific transconductance parameter, L_g and W_g the length and respectively the width of the transistor gate and

$$V_{min} = \min(V_{GS} - V_T, V_{DS}, V_{DSAT}) \quad (2.37)$$

where V_{DSAT} is the saturation drain voltage, the voltage where (velocity) saturation occurs. This is a first-order model that does not perfectly predict the output current, but effectively provides an understanding of what is going on in the transistor. Optionally equation (2.36) can be multiplied by a factor

$$1 + \lambda V_{DS}$$

representing channel length modulation, where λ is an empirically determined process specific parameter.

operation region
off-state
threshold voltage
triode region
transconductance
region

The three capacitors drawn are the most common present parasitics. The capacitance that can be found under the gate is split between a capacitance to the source and a capacitance to the drain, and is proportional to the gate length and width

$$C_{GS} + C_{GD} = C_{ox}W_gL_g \quad (2.38)$$

where C_{ox} is the gate capacitance per unit area. The exact capacitor values vary with the transistor operating point. Both capacitances are roughly equal when the transistor operates in the triode region since the channel exists under the entire gate. When the transistor operates in the transconductance region the channel pinches off before reaching the drain and

$$C_{GS} = \frac{2C_{ox}W_gL_g}{3} \quad (2.39)$$

$$C_{GD} = 0 \quad (2.40)$$

The capacitance C_{GD} is also called ‘feedback capacitance’ as it exists between the output and input of the transistor. In RF design this capacitance may lead to circuit instability. The physical output capacitance mainly consists of reverse biased PN-junctions from drain and source to the bulk and are mainly dependent on the area and perimeter of the drain and source. The output capacitance C_{DS} is a model of which can be said it is approximately proportional to the transistor width

$$C_{DS} \approx W_g \quad (2.41)$$

The exact values of these modeled parasitic capacitances are strongly influenced by the transistor build topology and applied metals to connect the transistor. Therefore these parameters are usually empirically characterized in transistor models rather than analytically calculated.

Breakdown

breakdown region The last region of “operation” would be the *breakdown region*. Two types of breakdown can be identified.

oxide breakdown First is *oxide breakdown*, which imposes a maximum on V_{GS} . When V_{GS} exceeds this maximum, the electric field in the gate oxide is so large that it no longer acts as insulator. This process is destructive and thus should be avoided. This type of breakdown is dependent of the oxide thickness, therefore gate length does not influence this maximum. Oxide breakdown can be caused by static electricity, hence Electro-Static Discharge (ESD) protection is applied to prevent this from happening.

junction breakdown The other is *junction breakdown*, which imposes a maximum on V_{DS} . This is when the electric field between the drain and source is large enough that charge carriers have enough energy to create new free electron-hole pairs, creating some sort of electron avalanche. If V_{DS} exceeds the breakdown voltage V_{bk} this avalanche breakdown occurs, causing I_{DS} to rapidly increase. This voltage reduces as the current I_{DS} increases, hence V_{bk} is defined as the voltage in which junction breakdown occurs when the device is turned off. This type of breakdown is not inherently destructive, but causes large amounts of energy to be dissipated in the transistor which might damage the transistor. Besides that, this behavior is generally unwanted and therefore this type of breakdown is also avoided.

Switch Model

In the transconductance region the transistor is best described as a current source largely independent of V_{DS} . In the triode region, when V_{DS} is low enough, the transistor can also be modeled as some sort of controlled resistor, meaning that there is a (mostly) linear relationship between the output voltage V_{DS} and output current I_{DS} . The transistor is said to be ‘fully on’, hence this resistance is referred to as the on-resistance R_{on} . Combining $V_{DS} = R_{on}I_{DS}$ and equation (2.36) for $V_{DS} < V_{GS} - V_T$ (triode region) gives the equation for R_{on} [21]

$$R_{on} = \frac{L_g}{\mu_n C_{ox} W_g (V_{GS} - V_T - v_{ds}/2)} \quad (2.42)$$

switched device This allows to simplify the model from figure 2.4 into a so called *switched device* model. This is shown in figure 2.5 and only holds if either the device is off or operates in the triode region.

The switch model is also used for simplifying the design of digital logic. Usually digital logic uses the binary voltages V_{DD} and ground which are applied to the gate(s) of the next logic cell. The gate

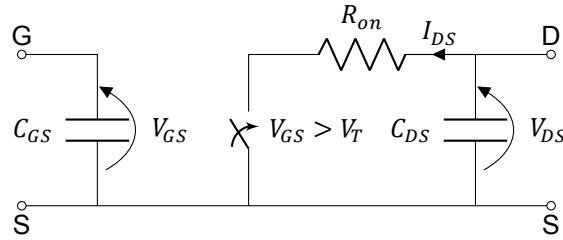


Figure 2.5: Switched FET model

is modeled by capacitance C_{GS} , so effectively it can be modeled as a capacitance is being charged through a resistance to V_{DD} by a PMOS transistor or to ground by an NMOS transistor. However, this resistance is not R_{on} because the transistor will start discharging in the transconductance region (since $V_{DS} = V_{DD}$) and not the triode region. An *Equivalent Series Resistance (ESR)* of the transistor R_{es} can be used that is the average value of the resistance over the operation region of interest [23]. This can be simplified to the average value of the resistances at the end-points of the transition. The commonly used transition is from V_{DD} to $1/2V_{DD}$, as that gives the definition of the *propagation delay*

equivalent series resistance
propagation delay

$$t_p = \ln(2)R_{es}C \tag{2.43}$$

where

$$R_{es} \approx \frac{3}{4} \frac{V_{DD}}{I_{DSAT}} \left(1 - \frac{7}{9} \lambda V_{DD} \right) \tag{2.44}$$

$$I_{DSAT} = \mu_n C_{ox} \frac{W_g}{L_g} \left[(V_{DD} - V_T)V_{DSAT} - \frac{V_{DSAT}^2}{2} \right] \tag{2.45}$$

$$C = C_{DS} + C_{load} \tag{2.46}$$

Device Curves

Usually the transistor output characteristics are characterized by plotting the output current I_{DS} against the output voltage V_{DS} with the input voltage V_{GS} as parameter. The transconductance is characterized by plotting output current against the input voltage V_{GS} . Idealized I_{DS} - V_{DS} and I_{DS} - V_{GS} curves are illustrated in figure 2.6. The operation regions are annotated in the curves. Besides the breakdown voltage

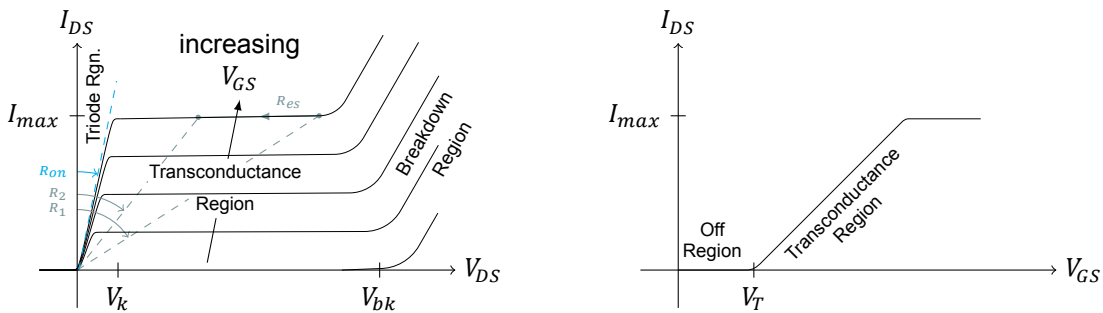


Figure 2.6: FET I-V curves

V_{bk} another voltage is indicated on the voltage axis, the *knee voltage* V_k . This voltage is the transition point between the triode and transconductance regions and is dependent on the input voltage V_{GS} . In this example the transition is illustrated relatively sharp to elucidate this transition, where in reality this transition is much more subtle. The different resistances that can be identified are illustrated by dashed lines: the on-resistance in the triode region is annotated with R_{on} and the ESR when transitioning from V_{DD} to $1/2V_{DD}$ can be estimated by $R_{es} \approx 1/2(R_1 + R_2)$ [23]. A transistor usually has a maximum output current, indicated by I_{max} . This maximum is determined by either the the point where increasing transistor V_{GS} does not increase I_{DS} , illustrated in the I_{DS} - V_{GS} curve, or by reliability limits. I_{max} is proportional to transistor width W_g by first order approximation. It can also be seen that the transistor

knee voltage

turns on at V_T and then has a perfectly linear relationship with V_{GS} until I_{max} is reached. Again, in reality these transitions are much more subtle.

These I - V curves can explain a lot how a transistor behaves in a circuit. For example a transistor in series with a resistor connected to a voltage source V_{DD} . The current through the resistor and the transistor will have to be equal and the voltage across both will remain constant. Increased current through the resistor means increased voltage across it, meaning a lower voltage across the transistor V_{DS} . Depending on the input voltage V_{GS} the operating point will change: higher I_{DS} means lower V_{DS} . This is illustrated in figure 2.7. Here the straight blue line is the relationship for I_{DS} and V_{DS} enforced by the resistor and is called a *load-line*. The intersections between the load-line and the I - V curves predict the different possible operation points. If the input voltage $V_{GS}(t)$ is a function of time, then the load-line predicts the trajectory of $I_{DS}(t)$ and $V_{DS}(t)$.

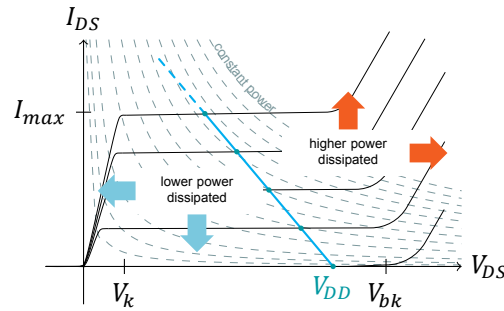


Figure 2.7: FET I - V curves and example resistive load-line

In a RF PA, $V_{GS}(t)$ is always a function of time. The load-line, or trajectory of $(I_{DS}(t), V_{DS}(t))$, will also predict some amplifier performance. A straight line indicates linear resistive operation, any kink in this straight line indicates possible non-linearities. If a phase shift between voltage and current is present, but the system is still linear, the load-line will take an elliptical shape. This indicates the presence of not only resistive elements, but also dynamic elements such as capacitors or inductors.

Also power efficiency can be partly predicted by the load-line. As stated in equation (2.1), instantaneous power is the product of voltage and current. The farther away the load-line is distanced from the I - V axes, the more DC power will be dissipated by the transistor without any power generated in return at the fundamental or its harmonics. Lines of constant power dissipation are illustrated by dashed lines. So the efficiency will generally improve if the load-line is closer to the I - V axes.

Comparing these two statements about the load-line it figures that a load-line as close as possible to the I - V axes will never be straight. This gives rise to another conclusion: a linear PA will not be power efficient and a power efficient PA will not be linear. Generally, the efficiency of a PA will improve with the generation of harmonic frequencies [20].

2.3.2. Transconductance Amplifiers

As the name implies, transconductance amplifiers operate with the transistor in the transconductance region. The transconductance amplifiers can be divided in Class-A, Class-AB, Class-B and Class-C amplifiers, which all share the same circuit topology shown in figure 2.8, but they differ in their input signal bias. The bandstop filter is open at and around the fundamental frequency f_c and a perfect short for any other frequency, making sure $V_{DS,k \cdot f_c} = 0$ for $k \geq 2$. Therefore no power is dissipated at the harmonic frequencies, thus not degrading the power efficiency: $P_{DS,k \cdot f_c} = 0$ for $k \geq 2$. The LTI passive network is assumed to be ideal, thus lossless: $P_{pasv} = 0$. A realistic implementation of the bandstop filter would be a tuned parallel LC-resonator.

Class-A

In a Class-A amplifier a DC bias is applied to V_{GS} in such a way that $V_{GS}(t) > V_T$ always holds with the applied input voltage swing. This means that the active device is always on and therefore conducting current at all times. That means the triode region is avoided and V_{DS} will therefore not fall below V_k . Peak V_{DS} should also not exceed V_{bk} to prevent the device from entering the breakdown region, just as peak I_{DS} should not exceed I_{max} . The Class-A waveforms and load-line are illustrated in figure 2.9. The waveform x-axis scale is in the angular time $\theta = \omega_c t$. A Class-A amplifier is perfectly linear, as

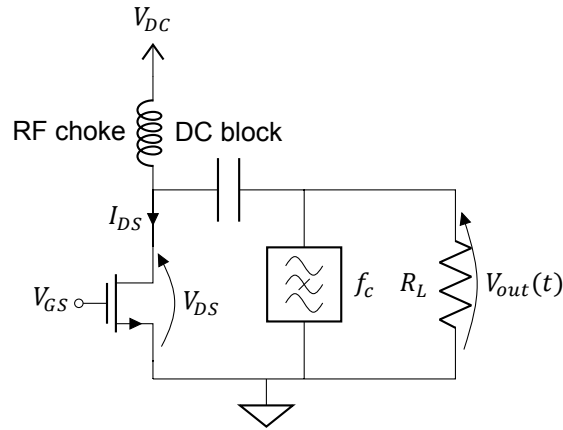


Figure 2.8: Circuit topology of a transconductance PA

the straight load-line suggests. This means that the device output current should have the same wave shape as the input voltage. If this input voltage does not contain any harmonics then the output current will not either, which makes the bandstop filter as shown in figure 2.8 unnecessary.

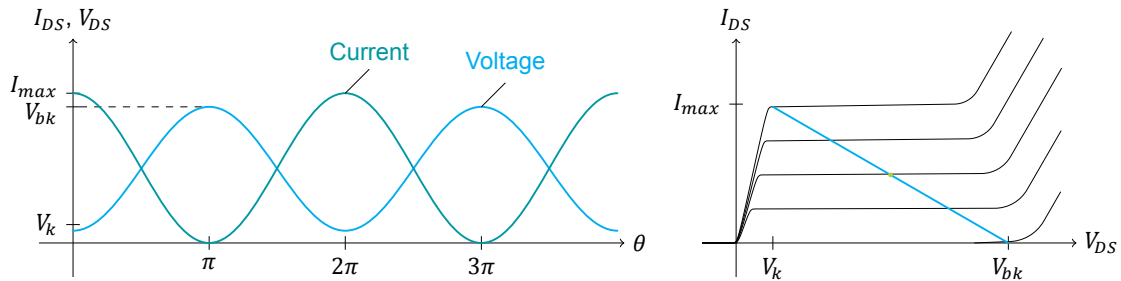


Figure 2.9: Class-A drain waveforms and load-line

For maximum output power and efficiency the output voltage should peak at V_k and V_{bk} , resulting in an output voltage amplitude

$$V_{out,fc,max} = \frac{V_{bk} - V_k}{2} \quad (2.47)$$

and a DC bias of

$$V_{DC} = \frac{V_{bk} + V_k}{2} \quad (2.48)$$

To make sure peak I_{DS} is I_{max} , the load resistance should be

$$R_L = \frac{2V_{out,fc}}{I_{max}} = \frac{V_{bk} - V_k}{I_{max}} \quad (2.49)$$

Filling in equation (2.8) results in a maximum output power of

$$P_{out,fc} = \frac{V_{out,fc}^2}{2R_L} = \frac{I_{max}(V_{bk} - V_k)}{8} \quad (2.50)$$

DC power consumed then is

$$P_{DC} = \langle V_{DS} \cdot I_{DS}(t) \rangle = V_{DC} \cdot \langle I_{DS}(t) \rangle = \frac{V_{bk} + V_k}{2} \cdot \frac{I_{max}}{2} \quad (2.51)$$

Using equations (2.50) and (2.51) the peak drain efficiency can be determined [20]

$$\eta_D = \frac{P_{out,fc}}{P_{DC}} = \frac{1}{2} \cdot \frac{1 - V_k/V_{bk}}{1 + V_k/V_{bk}} \quad (2.52)$$

Therefore the peak drain efficiency of a Class-A amplifier will at its best be

$$\eta_D|_{V_k=0} = 50\% \quad (2.53)$$

When reducing the output power of a Class-A amplifier P_{DC} will remain constant, therefore the drain efficiency will drop proportionally to the output power. In a transmitter where AM is applied the average drain efficiency can be calculated by using the PAPR

$$\overline{\eta_D} = \frac{\eta_D}{PAPR} \quad (2.54)$$

Of all the amplifier classes, Class-A has the worst power efficiency but the best linearity.

Class-AB, Class-B and Class-C

The efficiency of a PA can be improved by allowing non-sinusoidal waveforms. The output should remain a sinusoidal voltage, so it is the current that should assume a non-sinusoidal shape. A non-sinusoidal current implies that the current will consist of harmonics besides the fundamental. The bandstop filter, as shown in figure 2.8, makes sure these harmonic currents can flow freely without resulting in a voltage at that harmonic. This keeps the output power spectrum clean and, as mentioned at the begin of section 2.3.2, prevents power from being dissipated at these harmonics.

Enforcing a non-sinusoidal current shape is quite easy to obtain, all that needs to be done is turning the device off for some time. This can be achieved by simply lowering the DC bias of V_{GS} such that for a period of time V_{GS} will be lower than V_T . That way the input waveform can remain sinusoidal. The duration within one cycle that the device is on/conducting is called the conduction angle α . So in a Class-A scenario, when the device is conducting for the entire cycle, this would mean $\alpha = 2\pi$ rad.

Class-B operation is defined as the case that the device is on and off exactly half of a cycle, $\alpha = \pi$ rad. This can be done by putting the DC bias of V_{GS} exactly at V_T . When π rad $< \alpha < 2\pi$ rad the PA is somewhere between Class-A and Class-B, hence defined as Class-AB operation. Class-C operation is obtained by having the device on less than half a cycle: $\alpha < \pi$ rad. Example waveforms and load-lines are shown in figure 2.10.

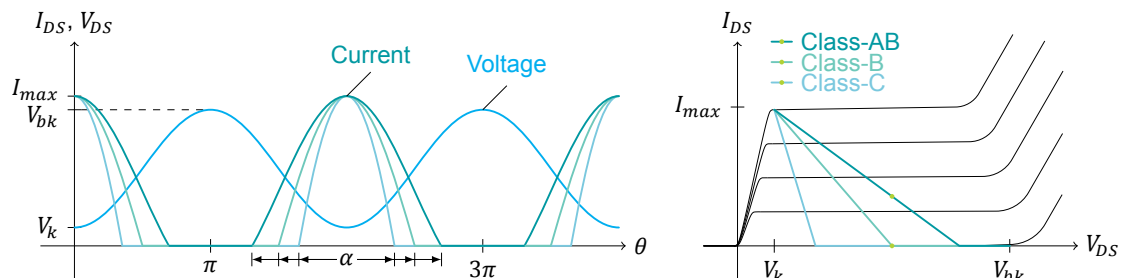


Figure 2.10: Class-AB, B and C drain waveforms and load-lines

The output voltage waveform and its requirements for maximum output power remain unchanged, so equations (2.47) and (2.48) still apply. The current for maximum output power in terms of angular time θ can be described for any transconductance amplifier of Class-A, AB, B, or C with [21]

$$I_{DS}(\theta, \alpha) = \begin{cases} I_{max} \frac{\cos(\theta) - \cos(\alpha/2)}{1 - \cos(\alpha/2)} & -\frac{\alpha}{2} \leq \theta + 2n\pi < \frac{\alpha}{2} \\ 0 & \text{otherwise} \end{cases} \quad (2.55)$$

By Fourier series expansion I_{DC} and $I_{DS,fc}$ can be calculated

$$I_{DC}(\alpha) = \frac{I_{max}}{2\pi} \frac{2 \sin(\alpha/2) - \alpha \cos(\alpha/2)}{1 - \cos(\alpha/2)} \quad (2.56)$$

$$I_{DS,fc}(\alpha) = \frac{I_{max}}{2\pi} \frac{\alpha - \sin(\alpha)}{1 - \cos(\alpha/2)} \quad (2.57)$$

Using equations (2.47) and (2.57) the maximum output power can be determined

$$P_{out,fc}(\alpha) = \frac{V_{out,fc} I_{DS,fc}(\alpha)}{2} = \frac{I_{max}(V_{bk} - V_k)}{8\pi} \cdot \frac{\alpha - \sin(\alpha)}{1 - \cos(\alpha/2)} \quad (2.58)$$

As equation (2.8) still holds, the optimum load resistance can be determined as

$$R_L(\alpha) = \frac{\pi(V_{bk} - V_k)}{I_{max}} \cdot \frac{1 - \cos(\alpha/2)}{\alpha - \sin(\alpha)} \quad (2.59)$$

Equations (2.48) and (2.56) give DC power consumed

$$P_{DC}(\alpha) = V_{DC} I_{DC}(\alpha) = \frac{V_{bk} + V_k}{2} \cdot \frac{I_{max}}{2\pi} \cdot \frac{2 \sin(\alpha/2) - \alpha \cos(\alpha/2)}{1 - \cos(\alpha/2)} \quad (2.60)$$

Using equations (2.58) and (2.60) the peak drain efficiency can be determined [20]

$$\eta_D(\alpha) = \frac{P_{out,fc}}{P_{DC}} = \frac{1}{2} \cdot \frac{1 - V_k/V_{bk}}{1 + V_k/V_{bk}} \cdot \frac{\alpha - \sin(\alpha)}{2 \sin(\alpha/2) - \alpha \cos(\alpha/2)} \quad (2.61)$$

It can now be shown that for a Class-B amplifier the peak drain efficiency is

$$\eta_D(\pi) \Big|_{V_k=0} = \frac{1}{2} \cdot \frac{\pi}{2} \approx 78,5 \% \quad (2.62)$$

which is quite an improvement from Class-A. The peak drain efficiency for a Class-C amplifier is in the limit

$$\lim_{\alpha \rightarrow 0} \eta_D(\alpha) \Big|_{V_k=0} = \frac{1}{2} \cdot \frac{2}{1} = 100 \% \quad (2.63)$$

however also

$$\lim_{\alpha \rightarrow 0} P_{out,fc}(\alpha) \Big|_{V_k=0} = 0 \text{ W} \quad (2.64)$$

Equations (2.58), (2.60) and (2.61) are plotted normalized in figure 2.11a. When aiming for high power efficiency with these kinds of amplifiers a trade-off should be made between power efficiency and output power. Maybe even more importantly, a trade-off between power efficiency and linearity. For now it was assumed that the bandstop filter was completely ideal. However ideality is not reality: linearity constraints restrict the choice of amplifier class.

Another important property of a PA is the drain efficiency in power back-off. As mentioned in equation (2.54) the drain efficiency of a Class-A amplifier is proportional to output power. It is therefore quadratically proportional to the output voltage. Besides having a higher peak drain efficiency, a Class-B amplifier also has a linearly proportional drain efficiency with output voltage. This is illustrated in figure 2.11b. That means that the average drain efficiency worsens with the root of the PAPR

$$\overline{\eta_D} = \frac{\eta_D}{\sqrt{PAPR}} \quad (2.65)$$

Something else to consider is that Class-B amplifiers require twice the voltage swing of V_{GS} than a Class-A amplifier to achieve the same peak output current. Their output powers are the same, but the amplifier power gain is reduced. In Class-C operation the DC bias of V_{GS} is (much) lower than V_T , so the input signal amplitude should be further increased and thus input power will increase. Therefore an optimum for total efficiency η_T with conduction angle α can be found, depending on device gain.

2.3.3. Saturated Transconductance Amplifiers and Harmonic Tuning

A significant restriction, up this point of this thesis, is the knee voltage V_k . It degrades the drain efficiency by a factor $\frac{1 - V_k/V_{bk}}{1 + V_k/V_{bk}}$. However, if it is allowed to distort the current waveform anyways, why wouldn't it be allowed to let the device enter the triode region? That is exactly what happens when an amplifier saturates. This is achieved by increasing the input signal amplitude. The device is then said to be *overdriven*. The voltage waveform is still not allowed to change. In figures 2.12 and 2.13 example waveforms and load-lines for a saturated Class-A and B amplifier are shown.

overdrive

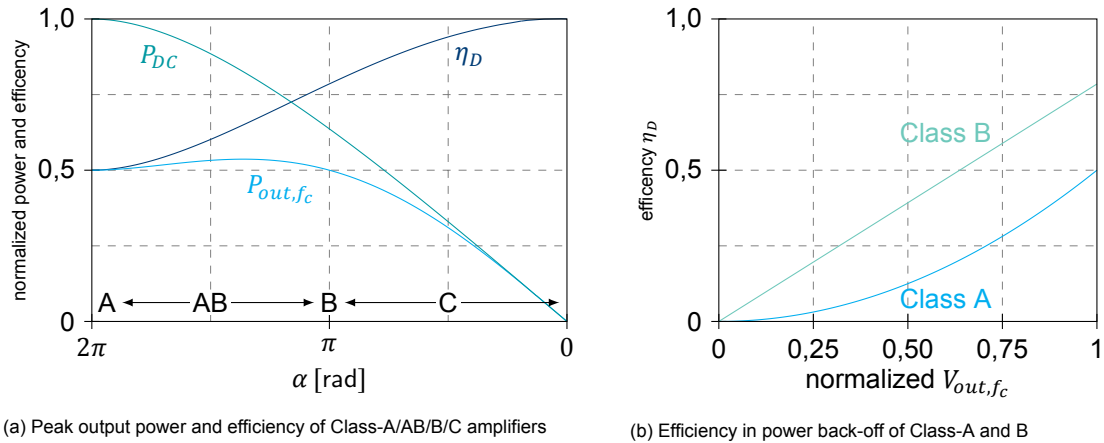


Figure 2.11: Performance of the different ideal transconductance amplifier classes

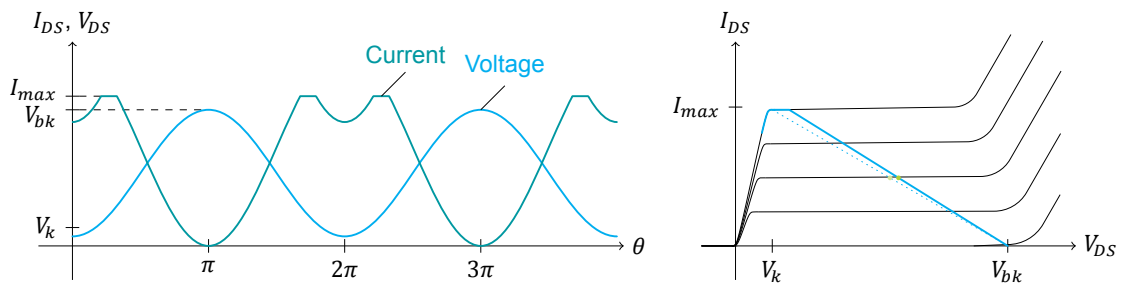


Figure 2.12: Saturated Class-A drain waveforms and load-line

In these figures it should especially be noted that the voltage waveform now extends below V_k , suggesting possible efficiency improvement due to possible lower losses at the point where the current is the highest. Also the output power is increased by the larger voltage amplitude, which might also improve efficiency. However, from the current waveform can be seen that the current now spends more time at a high level and at also times when the voltage is not at its lowest point, what may lead to an efficiency decrease. The combination of these effects are difficult to analyze analytically. It can be shown that in the limit, the maximally overdriven amplifier where $V_k = 0$ V, will have fully square waves in both voltage and current resulting in a theoretical efficiency of $\eta_D = 81,1\%$ [21]. In that case no power will be lost in the device ($P_{DS}=0$), but the remaining power can be found in the harmonics: $\sum_{i=2}^{\infty} P_{DS,i \cdot f_c}$ will contain the remaining 18,9% of the power. However, up to now it was always assumed the voltage waveform should remain sinusoidal, in which case the efficiency drops again.

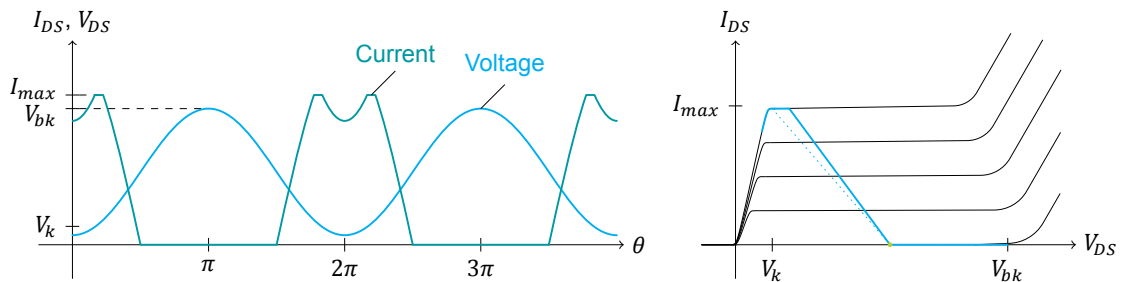


Figure 2.13: Saturated Class-B drain waveforms and load-line

Harmonic Tuning: Class-F

However, why shouldn't it be allowed to have a non-sinusoidal voltage? This does imply a change in circuit topology, since only the fundamental voltage may reach the load at the output of the PA. When it is allowed for the voltage waveform to take the shape of a square wave and the current to take the same half-sine shape as in Class-B no power should be lost in the device itself, leading to a theoretical perfect drain efficiency of 100%. This is what the Class-F amplifier aims to do.

The voltage is allowed to take a square waveform shape, which contains the fundamental and odd harmonics. The circuit topology should make sure that at the odd harmonics the device sees an open circuit, so no current can flow and therefore no power will be lost at the odd harmonics. A half-sine contains the fundamental and only even harmonics. Therefore the circuit topology should make sure that all the even harmonics are shorted, allowing the even harmonic currents to flow without resulting in a voltage and again no power will be lost at these harmonics. This results in a circuit topology as shown in figure 2.14. The odd harmonic bandstop filters allow the square wave voltage, while the fundamental bandstop filter only allows the fundamental current to pass to the load.

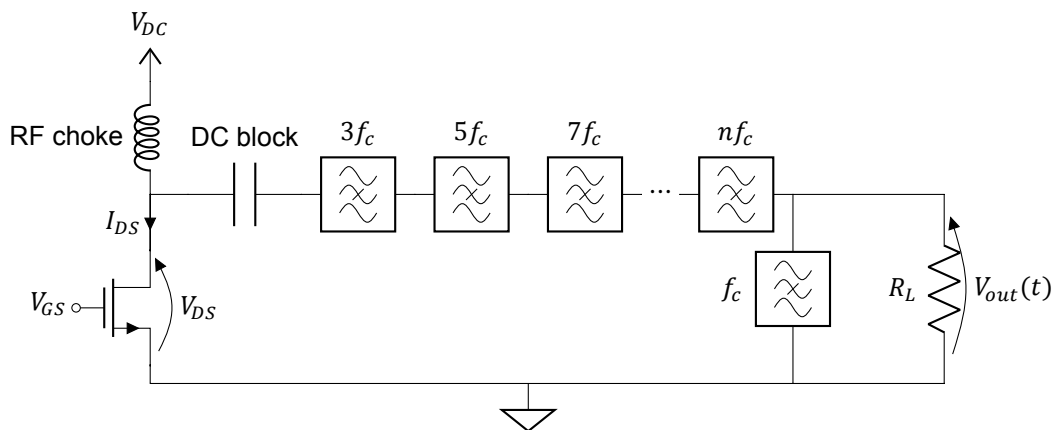


Figure 2.14: Circuit topology of a Class-F PA

For a perfect Class-F operation this would require control over infinite harmonics, so infinite harmonic resonators acting as bandstop filter. Perhaps an easier way of implementing is by using a $\lambda/4$ transmission line acting as impedance transformer. This circuit topology is shown in figure 2.15. The transmission line is directly connected to the power supply which acts as a signal ground. The impedance transformation results in that the transmission line acts as an open circuit for the fundamental and the odd harmonics. For DC and the even harmonics the transmission line will act as a short. The bandpass filter tuned at f_c will only pass the fundamental to the output load.

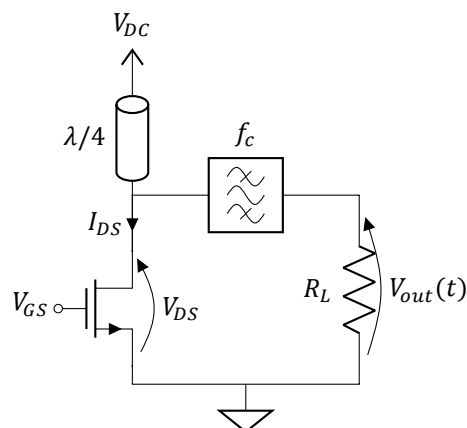


Figure 2.15: Circuit topology of a Class-F PA with a transmission line

Besides that it is physically not possible to construct an infinite amount of these resonators or that—depending on operation frequency—a $\lambda/4$ transmission line may be relatively large, referring back to

device model shown in figure 2.4, the device has a parasitic output capacitance C_{GS} which throws a spanner in the works. Therefore it is common that only the first few harmonics are tuned. The last harmonic that is tuned is denoted by a subscript in the Class notation. For example, when the circuit is tuned for two odd harmonics, being $3f_c$ and $5f_c$, the amplifier is said to be operating in Class-F₅. Note that Class-F₁ is equal to Class-B operation, hence resulting in a peak drain efficiency of $\eta_D = 78,5\%$. Only adding control over third harmonic, so Class-F₃ operation, increases the peak drain efficiency to $\eta_D = 88,4\%$ and in Class-F₅ operation the efficiency increases further to $\eta_D = 92,0\%$. For the ideal case of Class-F_∞ the efficiency becomes $\eta_D = 100\%$. Note that this is while neglecting the knee voltage V_k . Not only the efficiency increases, but so does the fundamental output power. As the voltage takes more and more a square shape, the normalized output power¹ increases from 0,50 to 0,63, 0,69 and 0,81 for respectively Class-B, F₃, F₅ and F_∞ [21].

The resulting load-lines and waveforms are shown in figure 2.16. Dashed in black is the Class-B case for reference. Depending on the exact harmonic content the voltage waveform may show more or less ripples. The shown dotted waveform is when the harmonic content would be exactly the Fourier decomposition of a square wave. The flattened waveform is achieved by slightly damping the harmonics.

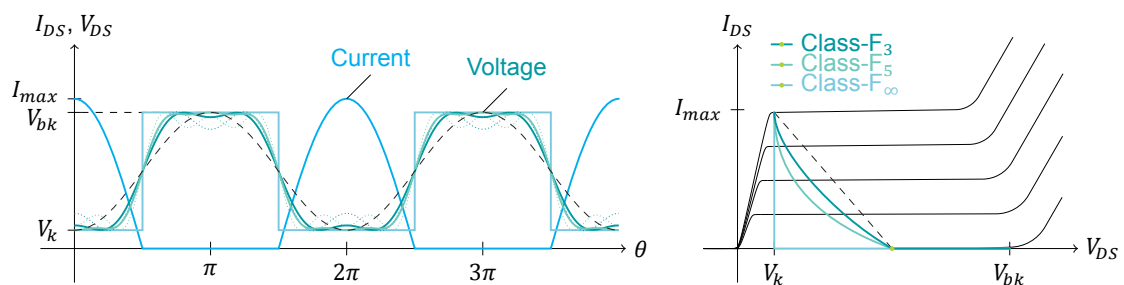


Figure 2.16: Class-F₃, F₅ and F_∞ drain waveforms and load-line and Class-B dashed for reference

2.3.4. Switching Amplifiers

The Class-F amplifier is sometimes called a switching amplifier. This is only true in the case of Class-F_∞ operation, then the device model may be simplified to the switched device model as shown in figure 2.5. Otherwise the drive requirements for a Class-F amplifier are the same as that of a Class-B amplifier.

The group of switching amplifier classes aim to implement the device as if it were a switch, what requires the device to either be off or to operate in the triode region. The active device will not conduct any current while off, when a large voltage across the device can be found. While on, the device is in the triode region. So when it is conducting current, the voltage across the device is low. This gives a major advantage in terms of power efficiency, since the product $V_{DS}(t) \cdot I_{DS}(t)$ then will be zero, if it wasn't for the on-resistance R_{on} .

SMPA The group of so called *Switch-Mode Power Amplifiers (SMPAs)* consists of Class-D and Class-E PAs.

Class-D

Class-D amplifiers are in principle the same as in Class-F_∞ operation, but are implemented using two switches in series. The circuit is shown in figure 2.17a. At a single moment in time only one of the switches is opened, therefore forcing the voltage at the common switch node to be a square wave. This can be seen in the waveforms shown in figure 2.17b. A bandpass filter makes sure the harmonics cannot pass to the load. This also ensures a half-sine current shape through each of the switches, which together result in a full-sine current shape through the load.

The Class-D SMPA, just as a Class-F_∞ PA, ideally has a peak drain efficiency of $\eta_D = 100\%$. This is assuming that R_{on} is $0\ \Omega$, but the knee voltage V_k does not pose a limiting factor for Class-D².

¹ $K_P = P_{out,fc} R_L / V_{DC}^2$, as will be defined later in equation (2.74)

²One could argue that this is possibly due to $V_k = 0\ \text{V}$ when no R_{on} is present

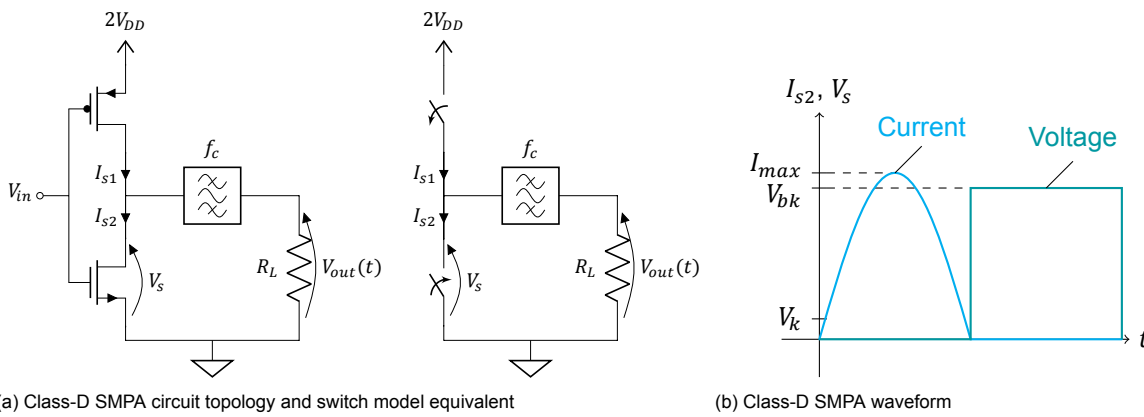


Figure 2.17: Class-D SMPA circuit topology and waveforms

Its topology however has some serious disadvantages. The topology for example does not take the devices' output capacitance into account. To make matters worse, the top switch should be implemented using a PMOS device, which generally has a lower carrier mobility than its NMOS counterpart, resulting in a device 2 to 3 times larger. That means both the input and output capacitance are up to four times larger than when only a single transistor is used. Therefore Class-D operation is restricted to relatively low operation frequencies where the output capacitance is not a limiting factor.

Class-E

The Class-E SMPA tries to compensate for the shortcomings of the Class-D amplifier, while maintaining the theoretical 100 % drain efficiency. Its circuit topology is shown in figure 2.20. Most notably it now incorporates a capacitance parallel to the switch C_S , in which the parasitic output capacitance of the device C_{DS} can be incorporated. It is implemented using only one active device. Besides reducing the input and output capacitance, this also avoids possible timing issues between multiple devices. To prevent possible harmonics from appearing at the output again a bandpass filter is placed, which can be implemented using a series LC-resonator. This bandpass filter also functions as the output DC block. Power to the circuit is fed through inductor L_{DC} , which may have a finite inductance. In the original Class-E implementation this inductor was a high reactance shunt-feed RF choke [24]. An additional reactance X is added to achieve proper Class-E operation.

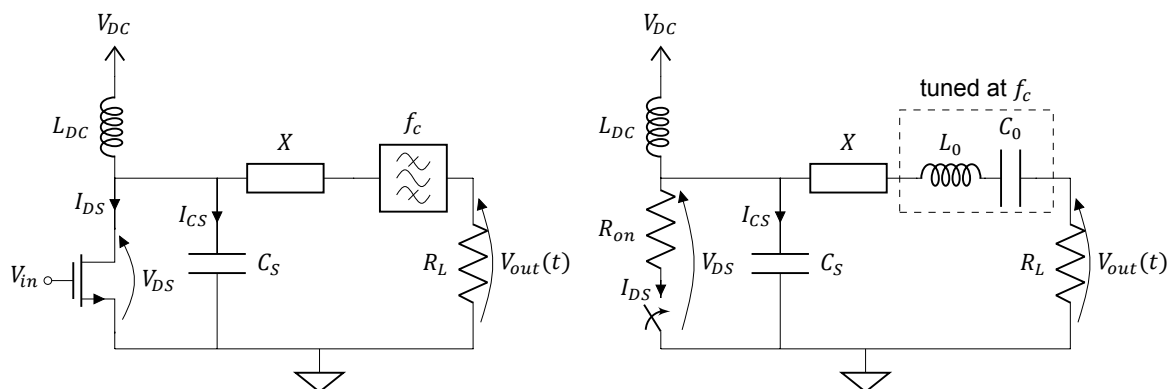


Figure 2.18: Class-E SMPA circuit topology and switch model equivalent

The primary aim of this circuit is, as in Class-D or F_{∞} , to make sure that at no point in time there is simultaneously a voltage across and current through the active device. Time domain analysis of the LTI passive network is used to design a proper network that achieves this rather than enforcing it by using two switches.

An important condition of optimum Class-E operation to ensure the 100 % theoretical drain efficiency is the *Zero-Voltage Switching (ZVS) condition*. ZVS means that the voltage across the switch is zero prior to closing the switch and is achieved by choosing the correct value for the additional reactance X

ZVS condition

in the circuit and is dependent of the chosen value for L_{DC} . If the ZVS condition is not met, a voltage across the switch capacitance C_S will be present when closing the switch. The capacitor will therefore be discharged through the switch and its energy will be dissipated by it, resulting in losses.

Of lesser importance, at least regarding the power efficiency, is the *Zero-Voltage slope Switching (ZdVS) condition*. This condition is not required for maximum drain efficiency, but minimizes sensitivity to variations of external component values and allows for maximum output power for a given V_{bk} [25].

Simplified Class-E Operation

In order to understand what happens in the Class-E circuit it is the easiest to review the original Class-E where L_{DC} is assumed to be a RF choke and therefore carries a constant DC supply current. The switch is assumed to be periodically open from $\pi \leq \theta + 2n\pi < 2\pi$ and closed for the rest of the time, so half of the time open and closed. No on-resistance is assumed. The output current can be assumed to be sinusoidal [20, 21, 25]:

$$I_{out}(\theta) = c_1 \cos(\theta) + c_2 \sin(\theta) \quad (2.66)$$

with c_1 and c_2 to be determined constants. The current through the switch and capacitor then have to be the DC current minus the output sinusoid

$$I_{DS} + I_{CS} = I_{DC} - c_1 \cos(\theta) - c_2 \sin(\theta) \quad (2.67)$$

During the time the switch is closed this current will flow only through the switch and when the switch is opened this current will flow through the capacitor. The voltage V_{DS} may be found by integrating I_{CS} during the time the switch is open

$$V_{DS}(\theta) = \frac{1}{\omega_c C_S} \int_{\pi}^{\theta} I_{DC} - c_1 \cos(\tau) - c_2 \sin(\tau) d\tau \quad (2.68)$$

$$= \frac{I_{DC}(\theta - \pi) - c_1 \sin(\theta) + c_2 [\cos(\theta) + 1]}{\omega_c C_S} \quad (2.69)$$

c_1 and c_2 can be determined by the ZVS and ZdVS conditions for the switch closing at $\theta = 2\pi$ [20]:

$$c_1 = I_{DC} \quad (2.70)$$

$$c_2 = -\frac{\pi}{2} I_{DC} \quad (2.71)$$

So I_{DS} can be given by

$$I_{DS}(\theta) = \begin{cases} I_{DC} \left[1 - \cos(\theta) + \frac{\pi}{2} \sin(\theta) \right] & 0 \leq \theta < \pi \\ 0 & \pi \leq \theta < 2\pi \end{cases} \quad (2.72)$$

and V_{DS} by

$$V_{DS}(\theta) = \begin{cases} 0 & 0 \leq \theta < \pi \\ \frac{I_{DC}}{\omega_c C_S} \left[\theta - \sin(\theta) - \frac{\pi}{2} \cos(\theta) - \frac{3\pi}{2} \right] & \pi \leq \theta < 2\pi \end{cases} \quad (2.73)$$

These equations have been plotted in figure 2.19.

The component values can be calculated analytically, but the equations quickly become quite involved. A design set $K : \{K_L, K_C, K_P, K_X\}$ for Class-E amplifiers can be specified from which the component values can be calculated as follows [26]

$$K : \begin{cases} K_L = \frac{\omega_c L_{DC}}{R_L} \\ K_C = \omega_c C_S R_L \\ K_P = \frac{P_{out,fc} R_L}{V_{DC}^2} \\ K_X = \frac{X}{R_L} \end{cases} \quad (2.74)$$

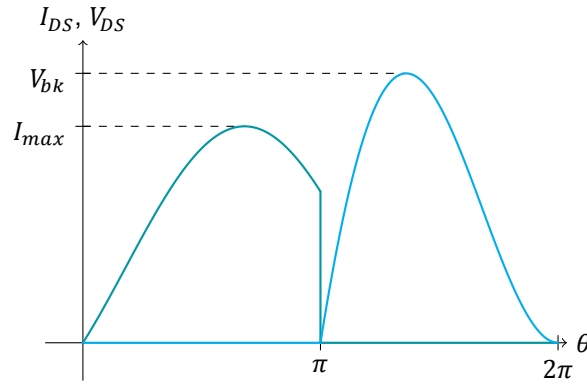


Figure 2.19: Class-E SMPA waveform

All design set values are real valued positive numbers, with exception of K_X which can also be negative. Positive K_X indicates additional inductance and negative K_X indicates additional capacitance.

For the original Class-E case this design set is [25, 26]

$$K : \begin{cases} K_L \rightarrow \infty \\ K_C = \frac{8}{\pi(\pi^2 + 4)} \approx 0,1836 \\ K_P = \frac{8}{\pi^2 + 4} \approx 0,5768 \\ K_X = \frac{\pi(\pi^2 - 4)}{16} \approx 1,1525 \end{cases} \quad (2.75)$$

A drawback of the Class-E SMPA is that the peak V_{DS} is quite high compared to the other amplifier classes [24, 25]. For maximum output power peak V_{DS} again should be equal to the breakdown voltage V_{bk}

$$V_{bk} = 2\pi \left[\frac{\pi}{2} - \tan^{-1} \left(\frac{\pi}{2} \right) \right] V_{DC} \approx 3,5620 \cdot V_{DC} \quad (2.76)$$

This results in a lower supply voltage than in the other amplifier classes, and a lower normalized output power than the high efficiency amplifier classes: For Class-E is $K_P \approx 0,58$, and if we recall they were for Class-F₃ $K_P \approx 0,63$, for Class-F₅ $K_P \approx 0,69$ up to Class-F_∞ or Class-D $K_P \approx 0,81$. Class-E performs better in this aspect than the group of transconductance amplifier classes. From figure 2.11a can be found that for Class-A and Class-B $K_P = 0,5$, the peak can be found in Class-AB with $K_P \approx 0,54$ for $\alpha \approx 1,362\pi$ and for Class-C $K_P < 0,5$.

Another drawback is that the output power of Class-E cannot be regulated by the magnitude of V_{GS} , making Class-E only suitable for constant envelope (phase modulated) signals. Applying amplitude modulation with Class-E requires special measures such as supply modulation or a Chireix (outphasing) design.

Generalized Class-E Operation

Quite a few assumptions have been made in the simplified Class-E operation. First, as mentioned earlier, L_{DC} does not necessarily have to be infinite. Finite L_{DC} can yield significant improvement of the Class-E operation. Secondly, the time the switch is opened or closed does not need to be exactly half of a cycle and lastly the on-resistance was neglected up to now. For that reason additional design parameters can be specified of which the design set K is a function: $K(q, d, m)$

$$d = \frac{\theta_{switch}}{\pi} \quad (2.77)$$

where θ_{switch} is the moment in the cycle where the switch opens. The duty-cycle is dimensionless and represented by $0 \leq d \leq 2$, a value of $d = 1$ represents a duty-cycle of 50%.

$$q = \frac{1}{\omega_c \sqrt{L_{DC} C_S}} \left(= \frac{1}{\sqrt{K_L K_C}} \right) \quad (2.78)$$

$$m = \omega_c R_{on} C_S \quad (2.79)$$

Both parameters q and m are positive real valued and dimensionless. If $q \rightarrow 0$ then $L_{DC} \rightarrow \infty$. Usually the value of q does not exceed 2. The parameter m most importantly gives an upper bound to the peak drain efficiency η_D and normalized output power K_P at a given operation frequency ω_c . Lower m results in higher drain efficiency and output power. It is therefore beneficial to have C_S as low as possible, with as minimum the parasitic output capacitance of the device C_{DS} . Recalling equations (2.41) and (2.42) $R_{on} \propto W_g^{-1}$ and $C_{DS} \propto W_g$, therefore

$$\min\left(\frac{m}{\omega_c}\right) = R_{on} C_{DS} \quad (2.80)$$

is a characteristic technology property independent of device sizing. For a given technology the performance in terms of power efficiency worsens with increasing frequency.

In figure 2.20 an example waveform and load-line for a non-zero m are illustrated. In the load-line the 'return trajectory' when the switch opens can spend a short time outside the triode region traveling to the off-state, but generally shouldn't impact the performance much as only a relatively short time is spent in this state.

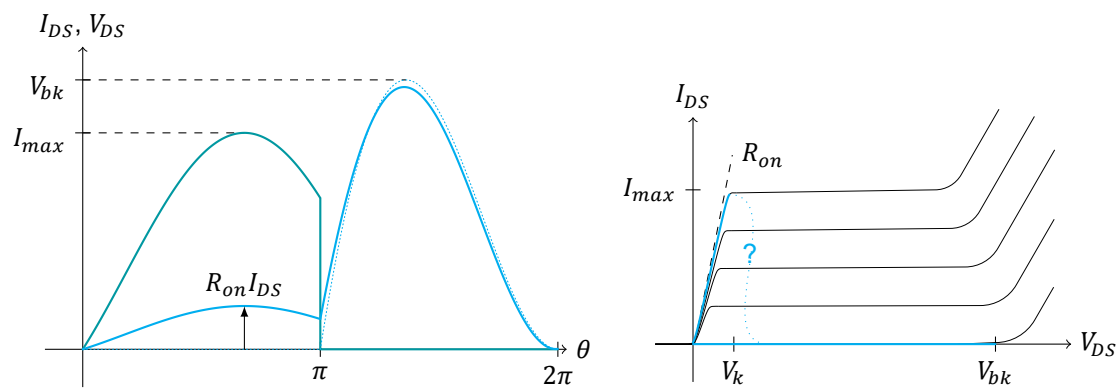


Figure 2.20: Class-E SMPA waveform for non-zero m and corresponding load-line

A few subclasses of Class-E operation have been identified, whereof the 'original Class-E' RF choke is one example that assumes $\{q, d, m\} = \{0, 1, 0\}$. Other known subclasses are 'parallel-circuit', 'second harmonic resonant' and 'load insensitive' Class-E with $q = 1,412$, $q = 2$ and $q \approx 1,3$ respectively while assuming $d = 1$. These may have significant advantages over the original Class-E. Parallel-circuit Class-E results in the highest output power and does not require additional reactance [26], second harmonic resonant Class-E (also called Class-E/ F_2^1) might have better efficiency for a given m [20, 27] and load insensitive Class-E keeps its efficiency constant irrespective of connected load R_L which makes it suitable for use in load-modulating efficiency enhancement techniques [28]. Design sets for some subclasses of Class-E are shown in table 2.1.

The effect of varying q to the efficiency can be seen in figure 2.21 for $m = 3,98 \cdot 10^{-3}$ and $d = 1$. Load insensitivity occurs in this case around $q \approx 1,25$, as can be seen by the drain efficiency being only a weak function of R_L . The drain efficiency stays within 10% of the peak drain efficiency at $R_L = 25 \Omega$, even though R_L changes an order of magnitude in both directions. The same simulation setup is used in section 3.3, the full overview of Class-E design set parameters and resulting circuit component values are given in the appendix in table A.1.

To capture all component values in equations only the bandpass filter components remain. Most obviously it should be tuned to be a bandpass filter for f_c , so

$$\omega_c = \frac{1}{\sqrt{L_0 C_0}} \quad (2.81)$$

¹The notation Class-E/ F_2 denotes that it is in Class-E operation with the second harmonic tuned as would be in inverse Class-F operation: second harmonic open [20]. Compared to non-inverted Classes, the inverse Class amplifiers have their current and voltage waveforms the other way around, as well as the harmonic tuning (open/short).

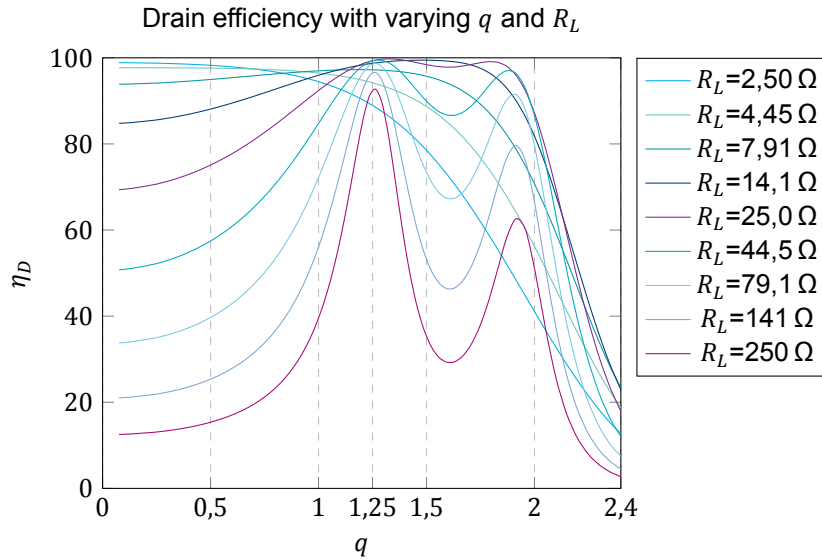


Figure 2.21: Drain efficiencies for varying q and R_L , where $m = 3,98 \cdot 10^{-3}$ and $d = 1$

All these subclasses assume that the output voltage is sinusoidal, so the quality factor of the bandpass filter should be sufficiently high to filter out unwanted harmonics. The quality factor is given by

$$Q_L = \frac{\omega_c L_0}{R_L} \quad (2.82)$$

Table 2.1: Different subclasses of Class-E and their design set [26, 27]

Quantity	Original RF choke	Parallel-circuit	Second harmonic resonant	Load insensitive
q	0	1,412	2	$\approx 1,3$
d	1	1	1	1
m	0	0	0	0
K_L	$\rightarrow \infty$	0,732	3,534	1,011
K_C	0,184	0,685	0,071	0,585
K_P	0,577	1,365	0,056	1,29
K_X	1,152	0	-4,903	0,26

2.4. Conclusion

Important performance metrics for a PA are the power flows to and inside a PA, the gain of a PA, its power efficiency and linearity. The active device is at the core of the PA, consuming DC power and (partly) converting the power consumed to another frequency. The primary goal is to convert the energy only to the frequency of interest, the operation frequency f_c . Linear PAs will be able to do that perfectly. A PA can be quite linear in Class-A operation, but it has poor power efficiency. By allowing more harmonics in the PA's current the power efficiency is seen to improve, such as in Class-B operation. At the extreme Class-C operation can be found with good power efficiency, but its output power drops as well. No harmonics in the voltage waveform are allowed up to that point to prevent unwanted power to be generated at the harmonics.

By also allowing the voltage waveform across the active device to have harmonics the efficiency can be further improved, such as in Class-F operation. The voltage harmonics are different than those of the current, so still no power is dissipated at the harmonics.

One step further is to use the active device as a switch to make sure that at no point in time there is simultaneously a voltage across and current through the active device. This results in a theoretical drain efficiency of 100% for these Switch-Mode Power Amplifiers (SMPAs). Class-D is such SMPA, but has serious shortcomings when taking the device parasitics into account which strongly limit the use of Class-D at higher frequencies.

The Class-E SMPA and its subclasses solve this issue by making sure no voltage across the device is present at the time of opening, called the Zero-Voltage Switching (ZVS) condition. A design set K that specifies Class-E component values is introduced and is dependent on the three parameters q , d and m : the “DC feed inductance L_{DC} with switch parallel capacitance C_S resonance frequency to ω_c ratio” as q , the duty-cycle d and “operation frequency ω_c , on-resistance R_{on} and switch parallel capacitance C_S product” the last parameter m .

The different Class-E subclasses can have advantages over each other, making a certain subclass more suitable for a certain design purpose. A drawback of Class-E is the relatively high peak V_{DS} voltage, therefore limiting its output power for a given breakdown voltage V_{bk} . Also its output power cannot be controlled by the magnitude of V_{GS} .

3

RF-DAC Basics

The concept of RF-DACs is perhaps the easiest to understand by first taking away the RF part and to focus on the DAC part first. In section 3.1 different DAC structures and DAC performance metrics are discussed. In section 3.2 the theory of a DAC is used to explain the operation of a DRAC. Combining the concept of a DRAC and the Class-E SMPA from section 2.3.4 the step towards a power DRAC can be made, as is explained in section 3.3.

3.1. DAC structures

In general, a DAC converts a digital binary code to an analog voltage or current. The number of possible analog output levels corresponding with the digital input words can be defined as the *resolution* of the DAC. This resolution is commonly expressed in bit, so a N_b bit resolution implies 2^{N_b} possible analog output levels. The resolution commonly refers to the number of digital input bits and isn't necessarily equal to the effective output resolution (see section 3.1.3). The digital input B_{in} can be represented by the N_b bit binary word [29]:

$$B_{in} = b_1 2^{-1} + b_2 2^{-2} + \dots + b_N 2^{-N_b} = \sum_{i=1}^{N_b} \frac{1}{2^i} b_i \quad (3.1)$$

Here b_1 forms the *Most Significant Bit (MSB)* and b_N the *Least Significant Bit (LSB)*. Ideally this creates an output voltage that has a linear relationship with the digital input word and some reference voltage

$$V_{out} = V_{ref} B_{in} \quad (3.2)$$

Any deviation from that is referred to as non-linearity.

The family of DACs can be roughly split into two main types: Nyquist rate DACs and oversampled DACs. A suitable DAC structure can be chosen while keeping in mind that the application is a power RF-DAC that has to be implemented using transistors and maybe resistors.

3.1.1. Oversampled DACs

Oversampled DACs operate at a significantly higher sampling frequency than the *Nyquist rate*, twice the bandwidth of the to be reproduced analog signal. By doing so the internal resolution of the DAC can be lowered while maintaining the required output resolution. An important mechanism in these kinds of DACs when using Delta-Sigma modulation is *noise shaping*, the designable property of shifting most of the quantization noise to frequencies higher than the frequency of interest. This noise can then be filtered away by a relatively relaxed analog anti-alias filter.

Often when implementing an oversampled DAC an internal resolution of only 1 bit is chosen because it results in an inherently linear output characteristic. This, however, quickly requires a sampling rate at least two orders of magnitude higher than the Nyquist rate in order to get a reasonable output resolution.

Oversampling is unfeasible in this design, as the targeted application is a RF-DAC operating in the UHF range. Perhaps an oversampling ratio of $2 \times$ to $4 \times$ can be achieved depending on the choice of

technology, but certainly not several orders of magnitude. Hence the main topology of the RF-DAC should be a Nyquist rate DAC.

3.1.2. Nyquist rate DACs

Nyquist rate DACs generate a single output for each single input sample and can again be split into four types: decoder-based, binary-weighted, thermometer-code and hybrid [29]. Whichever type is chosen, the main building block is always some switched element controlled by the input digital binary codes. These elements can either be a switched resistor, a switched current source or a switched capacitor circuit. The switch will be implemented by a transistor. As explained in section 2.3.1 a transistor can also function as a current source or as some sort of controlled resistor. So by controlling the transistors properly they can function as both the switch and the switched element itself, with exception of the switched capacitor¹. The switched capacitor implementation is not very suited for the intended high RF output power purpose since a switched capacitor implementation requires high levels of integration. Due to technology scaling these levels of integration can be found in advanced CMOS which has low supply voltage, resulting in increased size of, and losses in, the impedance transformation network [30]. Furthermore, a large total capacitance would be required to store enough charge for the intended output power, the charge times associated with it would significantly limit maximum operation frequency. That leaves switched resistor or switched current implementations for the intended high power application.

Decoder-Based DAC

The first type of Nyquist rate DAC, the decoder-based DAC, operates by creating as much reference signals as possible desired outputs and selecting one by turning on appropriate switches selected by the digital input code. The reference signals are usually a voltage, created by a chain of 2^{N_b} resistors. The created reference voltage is fed to a buffer as the resistors should not be loaded by another circuit. Reflecting upon the intended application it means that the DAC and PA are still separate and the decoder-based DAC is therefore not a suitable topology for a power DRAC.

Binary-Weighted and Thermometer-Coded DACs

The binary-weighted and thermometer-coded DACs are in their core principle the same: they combine circuit elements (currents or resistors) to create a varying effective circuit quantity, as illustrated in figure 3.1. In regular DAC applications these quantities are also buffered to make the DAC insensitive to the connected load impedance. However, the aim is to combine the DAC and PA into one, hence the connected impedance R_L at the output is well-known and should be designed towards.

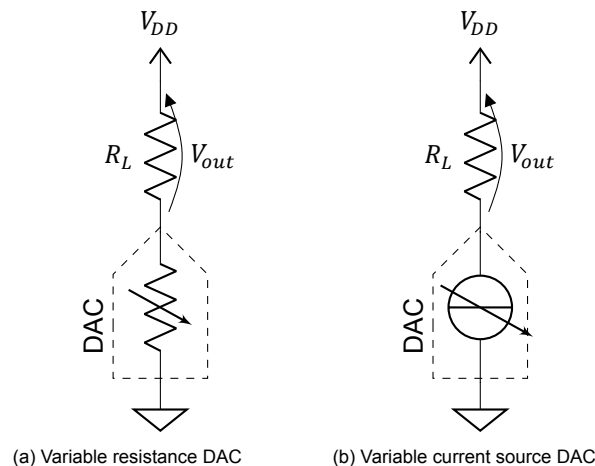


Figure 3.1: Simplified DAC circuit diagram

Their difference lies in how the variable circuit quantities are implemented. The possible implementations are shown in figure 3.2.

binary-weighted

In the *binary-weighted* type circuit quantities with magnitudes of a power of 2 are combined to create the required output voltage. In a DAC with N_b bit resolution this requires N_b circuit elements to realize

¹strictly speaking a transistor could also function as (non-linear) capacitor, but is undesirable in this case anyway

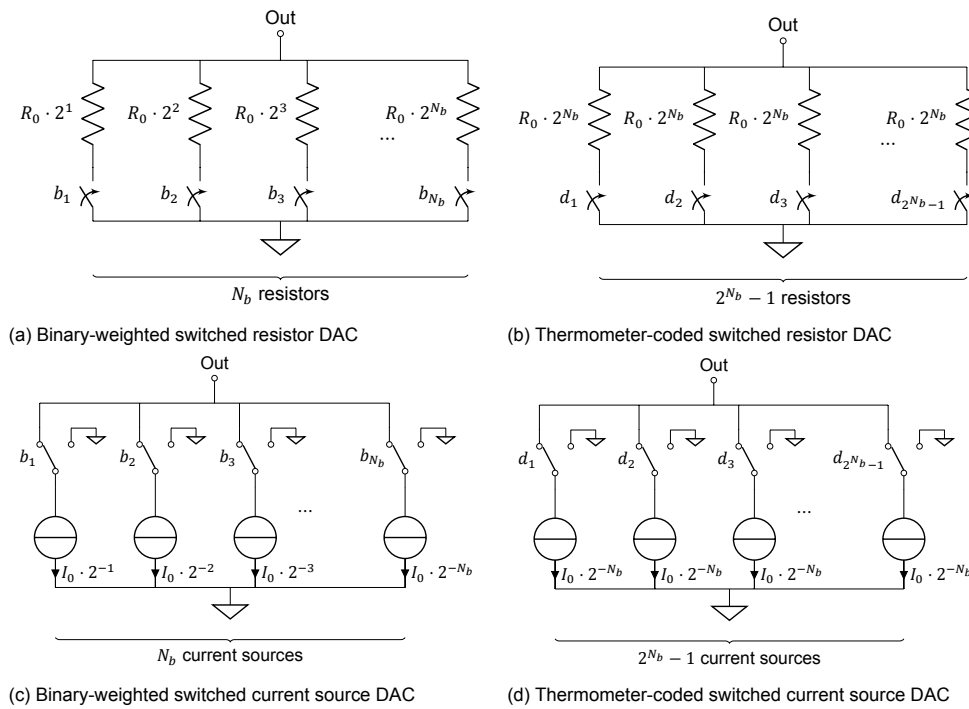


Figure 3.2: Switched circuit quantity DAC implementations

all possible outputs, hence being very hardware efficient. It is also very simple to control as the switched quantities are directly controlled by the digital input word B_{in} . The hardware efficiency does come at a price however. The different circuit components may be mismatched, thus creating non-linearity. Also, if any timing mismatch is present between the LSBs and MSBs this can result in *glitches* as illustrated in figure 3.3. In this case the LSBs (represented by I_1) turn off slightly before the MSBs (represented by I_2) turns on, causing a glitch to $I_{out} = 0$ A. Of course it is also possible to have a glitch that causes I_{out} to have an unintended peak.

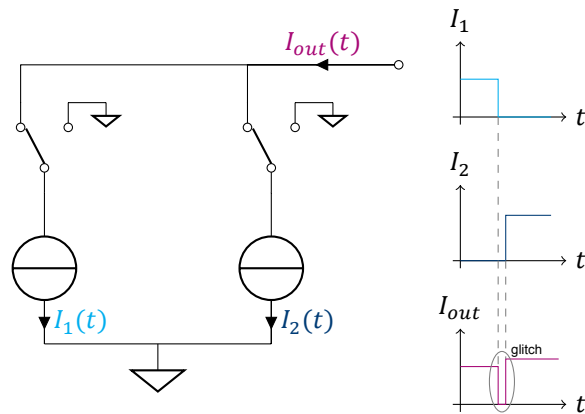


Figure 3.3: Origin of glitches

The *thermometer-coded* topology does not have this problem, but requires more hardware. It combines circuit quantities with the same magnitude to create the required output voltage, hence requiring $2^{Nb} - 1$ circuit elements. They are not controlled by the binary word $b_1 b_2 \dots b_N$, but by *thermometer-code* $d_1 d_2 \dots d_{2^{Nb}-1}$:

$$B_{in} = \frac{1}{2^{Nb}} \sum_{i=1}^{2^{Nb}-1} d_i \tag{3.3}$$

thermometer-coded
thermometer-code

Because all circuit elements are exactly the same they ought to be well matched, improving the linearity.

Linearity of the Binary-Weighted and Thermometer-Coded DACs

There is a significant difference between the resistor and current source implementations in this application. In a regular DAC the output signal is provided by some buffer, for example an opamp. This opamp keeps the voltage across the switched resistor bank equal, effectively converting them to currents. Parallel currents can be added together nicely, creating a linear relation between the digital input and the output voltage:

$$V_{out} = B_{in} I_0 R_L = \begin{cases} R_L \sum_{i=1}^{N_b} \frac{I_0}{2^i} b_i & \text{binary-weighted} \\ \frac{I_0 R_L}{2^{N_b}} \sum_{i=1}^{2^{N_b}-1} d_i & \text{thermometer-coded} \end{cases} \quad (3.4)$$

This is not the case when the resistors are directly connected to the output impedance

$$\begin{aligned} V_{out} &= \frac{V_{DD} R_L}{R_L + R_{eq}} \\ &= \frac{V_{DD} R_L}{R_L + \frac{R_0}{B_{in}}} = \begin{cases} \frac{V_{DD} R_L}{R_L + R_0 \left(\sum_{i=1}^{N_b} 2^{-i} b_i \right)^{-1}} & \text{binary-weighted} \\ \frac{V_{DD} R_L}{R_L + 2^{N_b} R_0 \left(\sum_{i=1}^{2^{N_b}-1} d_i \right)^{-1}} & \text{thermometer-coded} \end{cases} \\ &= \frac{V_{DD} R_L B_{in}}{B_{in} R_L + R_0} \end{aligned} \quad (3.5)$$

where R_{eq} is the equivalent resistance of the switched resistor bank. This equation clearly indicates a non-linear behavior with the digital input. The thermometer-coded resistor variant can be linearized by scaling the resistors in such way that the relationship between the input code and the output voltage becomes linear. The equivalent resistance of the switched resistor bank should then be

$$R_{eq}[k] = \frac{2^{N_b} R_L}{k} - R_L = \frac{R_L (2^{N_b} - k)}{k} \quad (3.6)$$

where $k = \sum_{i=1}^{2^{N_b}-1} d_i = 2^{N_b} B_{in}$, an integer number being equal to the k^{th} resistor that is connected to the output. The value of the k^{th} resistor R_k then should be

$$R_{eq}[k] = R_{eq}[k-1] \parallel R_k \quad (3.7)$$

$$R_k^{-1} = R_{eq}^{-1}[k] - R_{eq}^{-1}[k-1] \quad (3.8)$$

$$= \frac{k}{R_L (2^{N_b} - k)} - \frac{k-1}{R_L (2^{N_b} - k + 1)} \quad (3.9)$$

Simplifying gives

$$R_k = \frac{R_L (2^{N_b} - k) (2^{N_b} - k + 1)}{2^{N_b}} \quad (3.10)$$

The smallest resistance in the bank $R_{2^{N_b}-1}$ then should be

$$R_{2^{N_b}-1} = \frac{2R_L}{2^{N_b}} \quad (3.11)$$

which implies that the maximum attainable resolution is limited by the ratio of the load impedance and the minimum implementable resistance. This linearization technique is not possible in the binary-weighted switched resistor DAC.

Hybrid DACs

The last type is the *hybrid DAC*, which combines the previously explained structures. This combines the advantages while minimizing the disadvantages of these structures. An often used concept is the *segmented DAC*, where the top few MSBs are implemented in thermometer-code and the LSBs are implemented in a binary-weighted fashion. This improves the matching, thus linearity, of the DAC while saving hardware resources.

3.1.3. DAC Performance Metrics

Besides the resolution of a DAC, there are more metrics that indicate the performance of a DAC. As indicated by equation (3.2) the DAC output ideally is linear with the digital input word, any deviation is a non-linearity. There are several ways to measure possible non-linearities. Other performance metrics include settling time and dynamic range.

DAC Linearity

Some of the linearity metrics specified for PAs also are applicable to DACs. These are harmonic distortion (equation (2.16)) and intermodulation distortion as explained in section 2.1.4.

Specific for a DAC is the *Integral Nonlinearity (INL) curve*, defined as the difference between a straight reference line and the DAC output. Just as the AM-AM curve of a PA the INL curve is a static approximation of accuracy. The reference line can in principle be any line given by some linear function $f_{ref}[k]$. Three common lines that can be defined are: endpoint, best-fit and best fit trough origin. Best-fit methods can either be least squares or minimized extreme values. These lines are illustrated in figure 3.4 using an example (binary-weighted) switched resistor DAC as described in equation (3.5) with $R_0 = R_L$, $N_b = 4$, a least squares best-fit and endpoint through origin. These curves are often represented in terms of 1 LSB. The definition of 1 LSB is fixed by equations (3.1)–(3.2) as

$$V_{LSB} \triangleq \frac{V_{ref}}{2^{N_b}} \tag{3.12}$$

The INL definition then is

$$INL_{LSB}[k] = \frac{V_{out}[k] - f_{ref}[k]}{V_{LSB}} \tag{3.13}$$

In this example

$$V_{LSB} = \frac{V_{DD}R_L}{2^{N_b}(R_L + R_0)} = \frac{0,5V_{DD}}{2^4} = V_{DD} \cdot 0,03125 \text{ V}$$

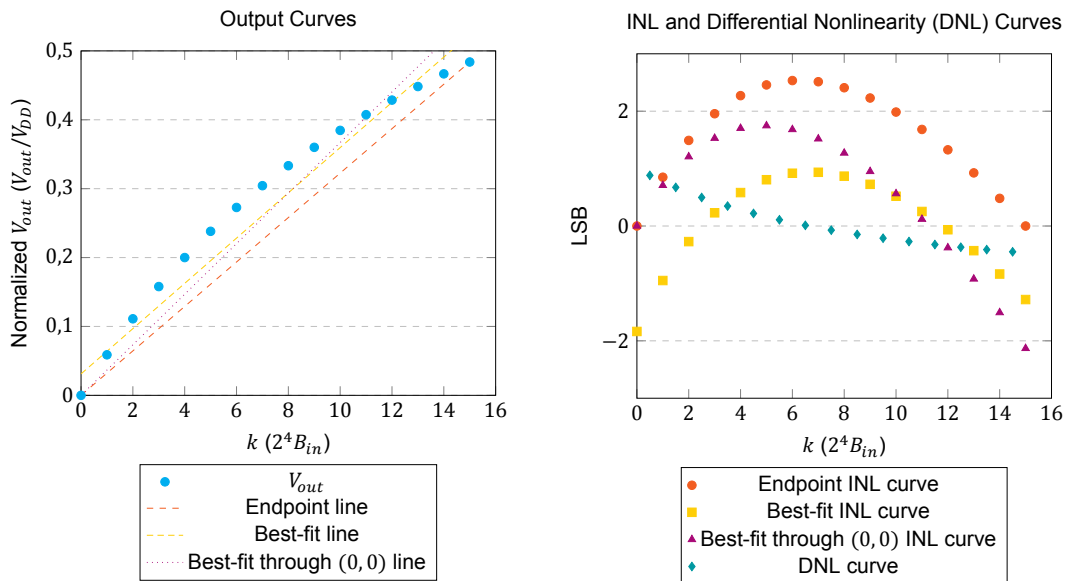


Figure 3.4: Different INL reference line definitions with corresponding INL curves and DNL curve

DNL curve Next the *DNL curve* can be defined by the difference between two consecutive analog outputs and 1 LSB

$$DNL_{LSB}[k-0,5] = \frac{V_{out}[k] - V_{out}[k-1]}{V_{LSB}} - 1 \quad (3.14)$$

Just as INL, DNL is a static approximation of accuracy, acquired by DC or low frequency Alternating Current (AC) simulation or measurement [29]. That means that neither INL nor DNL incorporate memory effects. For both INL and DNL a peak value can be defined

$$INL_{pk,LSB} = \max_k |INL_{LSB}[k]| \quad (3.15)$$

$$DNL_{pk,LSB} = \max_k |DNL_{LSB}[k-0,5]| \quad (3.16)$$

Note that peak INL is highly dependent of the chosen line f_{ref} .

monotonicity A related DAC property is *monotonicity*. A DAC is said to be monotonous if V_{out} always increases with increasing V_{in} . To be monotonous a sufficient and/or necessary conditions can be defined

$$DNL_{LSB}[k-0,5] > -1 \text{ LSB} \quad \forall k \stackrel{\Delta}{\Leftrightarrow} \text{DAC is monotonous} \quad (3.17)$$

$$DNL_{pk,LSB} < 1 \text{ LSB} \Rightarrow DNL_{LSB}[k-0,5] > -1 \text{ LSB} \quad \forall k \quad (3.18)$$

$$\therefore DNL_{pk,LSB} < 1 \text{ LSB} \Rightarrow \text{DAC is monotonous} \quad (3.19)$$

$$INL_{pk,LSB} < 0,5 \text{ LSB} \Rightarrow DNL_{pk,LSB} < 1 \text{ LSB} \quad (3.20)$$

$$\therefore INL_{pk,LSB} < 0,5 \text{ LSB} \Rightarrow \text{DAC is monotonous} \quad (3.21)$$

DAC Settling Time

settling time The *settling time* is a dynamic measure, defined as the time duration the DAC requires to settle its output voltage within a certain margin of the final value and is illustrated in figure 3.5. A common margin is chosen within $\pm 0,5$ LSB of the final value. Typically, the settling time should be lower than the sampling time.

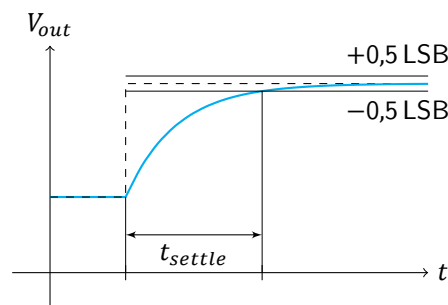


Figure 3.5: DAC settling time

DAC Noise and Dynamic Range

dynamic range In a DAC application the *Dynamic Range (DR)* is defined as the ratio of the maximum to minimum signal power it can meaningfully generate [29]. A signal is not meaningful anymore if it is indistinguishable from noise, hence the Dynamic Range can also be defined as the maximum attainable Signal-to-Noise Ratio (SNR).

quantization noise A major noise source in a DAC is the *quantization noise*, caused by the (ideal) DAC only being able to reproduce the values $V_{ref}B_{in}$ and not those in between. This causes an error V_{qn} in the reproduced voltage with a magnitude of $-0,5$ LSB to $0,5$ LSB, which is referred to as quantization noise. Generally it is assumed that the quantization noise has a uniform random distribution resulting in a RMS quantization noise voltage of

$$V_{qn,RMS} = \frac{V_{LSB}}{\sqrt{12}} \equiv \frac{V_{ref}}{2^{N_b} \sqrt{12}} \quad (3.22)$$

The maximum meaningful signal would be a sinusoid between 0 V and V_{ref} , thus an amplitude of $1/2 V_{ref}$. This gives an RMS voltage of $1/2\sqrt{2} V_{ref}$. This way the maximum attainable SNR can be defined as a Signal-to-Quantization-Noise Ratio (SQNR) and is often expressed in dB

$$SQNR_{dB} \triangleq 20 \log_{10} \left(\frac{V_{out,RMS}}{V_{qn,RMS}} \right) = 20 \log_{10} \left(\frac{V_{ref}/2\sqrt{2}}{V_{ref}/2^{N_b}\sqrt{12}} \right) \quad (3.23)$$

$$= 20 \log_{10} \left(2^{N_b} \sqrt{3/2} \right) \approx 6,021 N_b + 1,761 \text{ dB} \quad (3.24)$$

DAC Dynamic Range Influenced by Linearity

What a meaningful signal is can also be influenced by the linearity of the DAC. This gives rise to another dynamic range, defined by the ratio of the highest unwanted signal (such as intermodulation or harmonic distortion) and the wanted signal [31]. This is referred to as *Spurious-Free Dynamic Range* (SFDR). Just as with PAs the distortion amount is a function of signal amplitude (see section 2.1.4). Therefore it is common to specify SFDR as the SNR where the highest unwanted signal is equal to the (quantization) noise floor [29].

SFDR

DAC Resolution

As mentioned in the beginning of section 3.1 there is a difference between the resolution in the sense of the number of digital input bits and the effective output resolution. To define this *Effective Number Of Bits* (ENOB) we need to specify a measure that incorporates all unwanted components. This measure is the Signal-to-Noise and Distortion Ratio (SNDR), the ratio of the signal power to the combined power of all noise and distortion components up to the Nyquist frequency.

ENOB

$$SNDR_{dB} = 10 \log_{10} \left(\frac{V_{out,fc}^2}{V_{qn}^2 + \sum_{i=2}^{\infty} V_{out,i,fc}^2} \right) \quad (3.25)$$

By substituting the SQNR by the SNDR in equation (3.24) and solving for N_b gives the ENOB

$$ENOB = \frac{SNDR_{dB} - 1,761}{6,021} \quad (3.26)$$

3.2. From DAC to DRAC

Now the DAC part of the RF-DAC has been established it is time to add the RF part. As a DAC consists of only switched circuit elements, we can choose to periodically disconnect and reconnect all elements with a RF clock. The principle is shown in the circuit in figure 3.6a. Disconnecting causes the output to go to 0 V and reconnecting them again making the output go back to the value dictated by the digital input. The digital input does not specify the output signal voltage anymore as was the case in a DAC. Rather it specifies the output signal envelope or amplitude. To avoid confusion, the digital input in context of a DRAC will be defined as the Amplitude Code Word (ACW). Similar to equation (3.1)

$$ACW = ACW_1 2^{-1} + ACW_2 2^{-2} + \dots + ACW_N 2^{-N_b} = \sum_{i=1}^{N_b} \frac{1}{2^i} ACW_i \quad (3.27)$$

Again ACW_1 forms the MSB and ACW_N the LSB. We can also define a thermometer-coded variant d_{ACW}

$$ACW = \frac{1}{2^{N_b}} \sum_{i=1}^{2^{N_b}-1} d_{ACW,i} \quad (3.28)$$

and k still indicating the k^{th} possible ACW value

$$k = \sum_{i=1}^{2^{N_b}-1} d_{ACW,i} = 2^{N_b} ACW \quad (3.29)$$

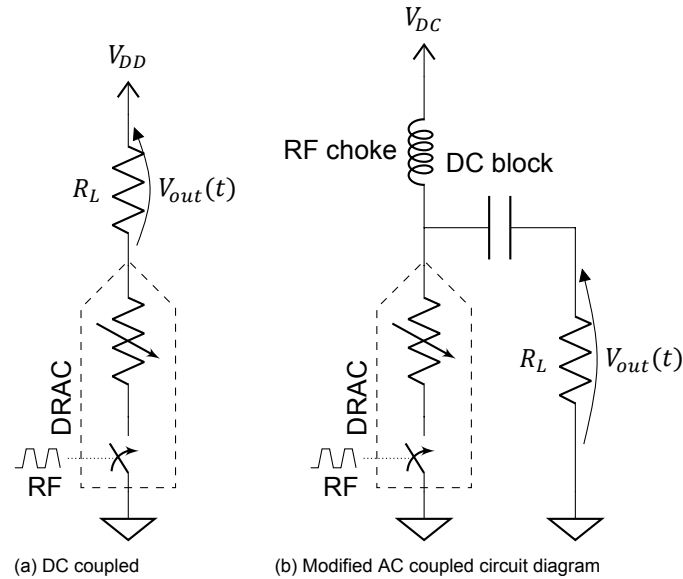


Figure 3.6: Simplified variable switched resistance DRAC circuit diagrams

The DRAC is intended to generate a bandpass signal whereas the DAC generates a baseband signal. Therefore it makes no sense to have DC coupling between the DRAC and the load. The adjusted AC coupled circuit is shown in figure 3.6b and its corresponding waveform is illustrated in figure 3.7. As the DRAC output is a bandpass signal the frequency of interest is around the RF carrier frequency f_c . With that reasoning it makes sense to redefine the input-output relationship

$$V_{out,f_c} = \alpha_V V_{DC} ACW \quad (3.30)$$

similar to equation (2.8), V_{out,f_c} is the amplitude of the sinusoidal output voltage at f_c . Furthermore α_V is some constant related to the waveshape of $V_{out}(t)$ and can be defined by

$$\alpha_V \triangleq \frac{V_{pk}}{V_{DC}} \quad (3.31)$$

where V_{pk} is the peak voltage of the waveform. In the case of a square wave $\alpha_V = 4/\pi$ since the fundamental amplitude is $4/\pi$ times the amplitude of a square wave.

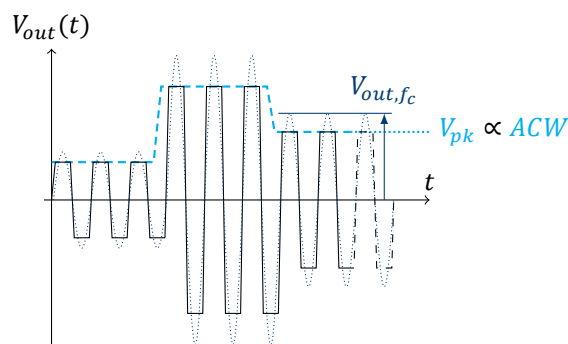


Figure 3.7: DRAC time domain waveform

In figure 3.6 the DRAC is shown with a switched resistor implementation. However, similarly to the DAC, the current source implementation is possible as well. Also all the DAC structures (binary-weighted, thermometer-coded or hybrid) are applicable to the DRAC.

With the AC coupling the DAC equations change. The DRAC output node connected to the RF choke and DC block will have an average voltage of V_{DC} due to the RF choke. The DC current through

the RF choke will be constant at I_{avg} . With a current source implementation that means that half of the time there is no current flowing through the DRAC, so I_{avg} has to flow through R_L . During the other half of the time a current I_{avg} in the opposite direction has to flow out of R_L due to the AC coupling, hence a current of $2I_{avg}$ will flow through the DRAC. That means for the switched current source implementation $2I_{avg} = I_0 ACW$, and the modified version of equation (3.4) becomes

$$V_{out,fc} = \alpha_V ACW I_0 R_L = \begin{cases} \alpha_V R_L \sum_{i=1}^{N_b} \frac{I_0}{2^i} ACW_i & \text{binary-weighted} \\ \frac{\alpha_V I_0 R_L}{2^{N_b}} \sum_{i=1}^{2^{N_b}-1} d_{ACW,i} & \text{thermometer-coded} \end{cases} \quad (3.32)$$

Similarly in the switched resistor implementation the DRAC output node voltage gets pulled down by the output voltage, hence $2I_{avg} = \frac{V_{DC}-V_{out}}{R_{eq}}$ and equation (3.5) becomes

$$\begin{aligned} V_{out,fc} &= \alpha_V \frac{V_{DC} R_L}{R_L + 2R_{eq}} \\ &= \alpha_V \frac{V_{DC} R_L}{R_L + 2 \frac{R_0}{ACW}} = \begin{cases} \frac{\alpha_V V_{DC} R_L}{R_L + 2R_0 \left(\sum_{i=1}^{N_b} 2^{-i} ACW_i \right)^{-1}} & \text{binary-weighted} \\ \frac{\alpha_V V_{DC} R_L}{R_L + 2^{N_b+1} R_0 \left(\sum_{i=1}^{2^{N_b}-1} d_{ACW,i} \right)^{-1}} & \text{thermometer-coded} \end{cases} \\ &= \alpha_V \frac{V_{DC} R_L ACW}{ACW R_L + 2R_0} \end{aligned} \quad (3.33)$$

which again indicates a similar non-linearity as the switched resistor DAC albeit slightly different with a factor 2 before R_{eq} in the denominator. This can be linearized similar to equation (3.6)

$$R_{eq}[k] = \frac{1}{2} \left(\frac{2^{N_b} R_L}{k} - R_L \right) = \frac{R_L (2^{N_b} - k)}{2k} \quad (3.34)$$

and equation (3.10)

$$R_k = \frac{R_L (2^{N_b} - k)(2^{N_b} - k + 1)}{2^{N_b+1}} \quad (3.35)$$

3.2.1. Frequency Domain Analysis

Described switching is effectively multiplying the targeted output voltage with a square wave in the time domain, effectively upconverting the baseband signal like a switching mixer. To analyze the frequency domain effects it first is important to analyze the baseband frequency spectrum.

The baseband data consists of ACW symbols, which dictate the RF amplitude. The baseband bandwidth is designated as f_{BB} . The frequency spectrum is periodic with the applied sampling frequency, which might result in problems regarding the required spectral purity of the output signal. Usually this baseband bandwidth is resampled to some higher frequency f_{r_s} , to prevent this issue. In practical applications this upsampling clock frequency f_{r_s} is lower than the carrier frequency f_c . Therefore, in the time domain, the value of ACW is kept constant until the next sample as can be seen in figure 3.7. This is equivalent to a time domain convolution with rectangle function $\Pi(f_{r_s} t)$ and is referred to as *Zero-Order Hold (ZOH)*. In the frequency domain the Fourier transform specifies that this is equivalent to multiplying the spectrum with a sinc function $\mathcal{F}\{\Pi(\omega_{r_s} t)\} = \text{sinc}(\omega/\omega_{r_s})$. The resulting frequency spectrum is shown in figure 3.8, illustrating the attenuation of the periodic repetitions of the frequency spectrum. This attenuation is referred to as the ZOH-effect.

The established baseband spectrum is upconverted by the switching action, effectively multiplying the time domain signal with a square wave. Ideal mixing would be multiplying with a sinewave as this only consists of one frequency. However Fourier expansion of a square wave tells us it consists of a

ZOH

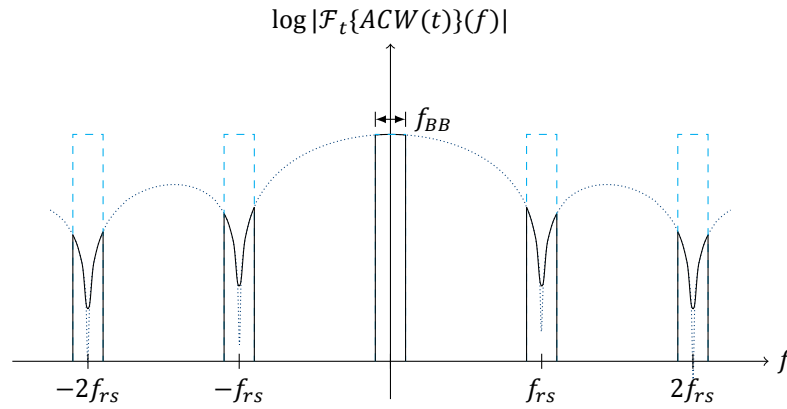


Figure 3.8: Baseband ACW frequency spectrum with ZOH-effect

fundamental with an infinite amount of odd harmonics.

$$\mathcal{F}_t\{\sin(\omega_c t)\}(\omega) = \frac{j}{2} [\delta(\omega - \omega_c) + \delta(\omega + \omega_c)] \quad (3.36)$$

$$\mathcal{F}_t\{\text{square}(\omega_c t)\}(\omega) = \frac{2j}{\pi} \sum_{i=1}^{\infty} \frac{\delta[\omega - (2i-1)\omega_c] + \delta[\omega + (2i-1)\omega_c]}{2i-1} \quad (3.37)$$

where $\delta(\omega)$ is the Dirac delta function. So multiplying with a square wave in the time domain results in a convolution in the frequency domain with the fundamental and the odd harmonics, effectively upconverting a baseband signal to a bandpass signal around f_c with harmonic replicas around the odd harmonics. The resulting frequency spectrum is shown in figure 3.9, together with the spectral mask of the ZOH-effect.

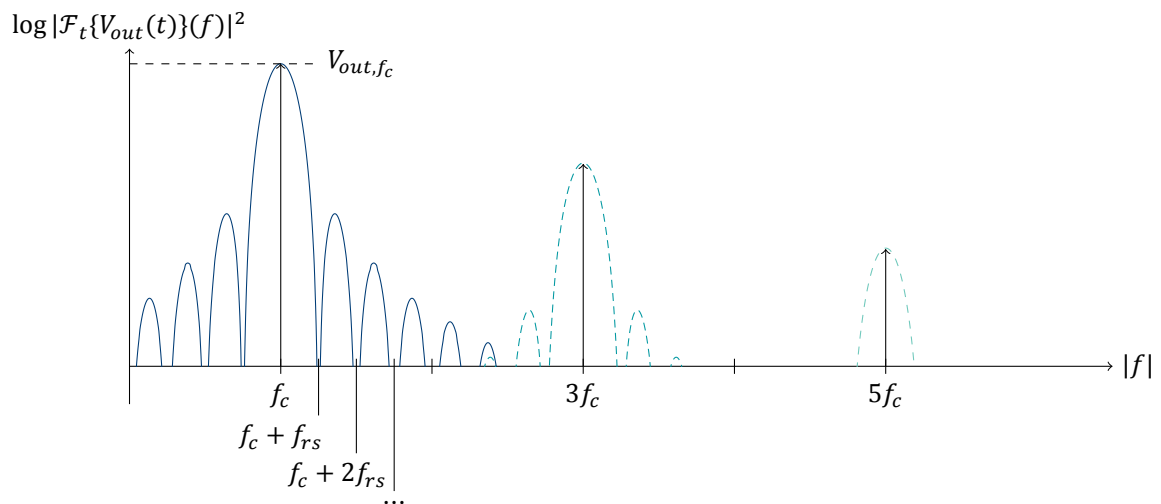


Figure 3.9: Upconverted bandpass RF absolute frequency power spectrum with ZOH-effect and harmonic replicas

This illustrates the need for filtering at the output to attenuate the harmonic replicas, or use some harmonic rejection technique. Also, the higher f_{rs} , the easier it is to filter out the periodic repetitions of the baseband frequency spectrum. The need for higher f_{rs} increases with increased channel bandwidth f_{BB} . In the limit, it would be possible to have f_{rs} equal to the RF carrier f_c .

3.2.2. DRAC Performance Metrics

The performance metrics of a DAC can be reused in context of a DRAC, but need some redefinition since we now have a bandpass signal. Basically it involves changing V_{out} to V_{out,f_c} in all equations. It

was chosen to use V_{out,f_c} in this thesis, while $V_{out,f_c,RMS}$ would also be correct with only a factor $1/\sqrt{2}$ difference.

DRAC Linearity

With a DRAC the linearity measures are more similar to those of a PA. The linearity that is similar to that of a DAC is the AM-AM curve and INL and DNL can be defined for a DRAC. As this linearity is often represented in LSB, the LSB has to be defined in context of the DRAC and can be defined using equations (3.27) and (3.30)

$$V_{LSB} \triangleq \frac{\alpha_V V_{DC}}{2^{N_b}} \quad (3.38)$$

That makes INL

$$INL_{LSB}[k] = \frac{V_{out,f_c}[k] - f_{ref}[k]}{V_{LSB}} \quad (3.39)$$

and DNL

$$DNL_{LSB}[k-0,5] = \frac{V_{out,f_c}[k] - V_{out,f_c}[k-1]}{V_{LSB}} - 1 \quad (3.40)$$

DRAC Settling Time

Settling time in context of a DRAC refers to the time duration the DRAC requires to settle its output voltage at the fundamental V_{out,f_c} within a certain margin of a periodic steady state and is illustrated in figure 3.10. The peak values of the sinusoid can be used to determine whether V_{out,f_c} is within the specified range and settling time can be expressed in a number of RF cycles.

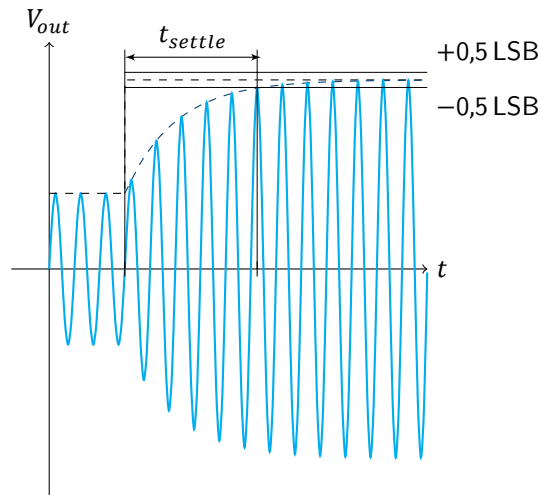


Figure 3.10: DRAC settling time

DRAC Dynamic Range

With a DRAC, quantization noise is a bit different from a DAC since its output is some sinusoid. This means that $V_{out}(t)$ does contain all analog values between $-V_{pk}$ and V_{pk} . However the available output levels of V_{out,f_c} are quantized. That means a limitation to how accurately a constellation point g can be transmitted, thus a relation to the distortion hence influencing the out-of-band spectrum. The power level in the out-of-band can be seen as some “noise floor” and is still very similar to a DAC as in equation (3.24), but due to the ZOH-effect the DRAC Dynamic Range $DR(f)$ is now frequency dependent and can be defined for a single-tone baseband signal in the band from 0 Hz to the Nyquist frequency $f_{rs}/2$ by [32]

$$DR_{ST,dB}(f)|_{f_{rs}/2} = 6,021N_b + 1,761 + 20 \log_{10} \left[\text{sinc} \left(\frac{f - f_c}{f_{rs}} \right) \right] \quad (3.41)$$

this leads to a Dynamic Range for a single-tone baseband signal in a 1 Hz bandwidth

$$DR_{ST,dB}(f)|_{\text{Hz}} = DR_{dB}(f)|_{f_{rs}/2} + 10 \log_{10} \left(\frac{f_{rs}}{2} \right) \quad (3.42)$$

which implies that the DRAC noise floor lowers and the Dynamic Range increases by 3 dB when doubling f_{rs} , allowing for an effective resolution increase of 0,5 bit.

The Dynamic Range is degraded as the average output power drops, which is the case in any modulation scheme involving AM. Therefore the PAPR should be subtracted from the single-tone Dynamic Range to define a multi-tone Dynamic Range

$$DR_{MT,dB}(f)|_{Hz} = DR_{ST,dB}(f)|_{Hz} - 10 \log_{10}(PAPR) \quad (3.43)$$

The noise floor gets higher when a higher baseband signal bandwidth f_{BB} is used, giving a general expression for the Dynamic Range of a RF-DAC

$$DR_{MT,dB}(f)|_{f_{BB}} = DR_{MT,dB}(f)|_{Hz} - 10 \log_{10}(f_{BB}) \quad (3.44)$$

3.3. Towards a Power DRAC Implementation

Looking at figures 2.18 and 3.6b it becomes clear that the AC coupled DRAC is implementation wise quite similar to a Class-E SMPA and therefore it is well possible to combine their concepts. All that is required to transform the SMPA into a DRAC is to incorporate a digitally controllable transistor R_{on} . This digitally controllable R_{on} can be obtained by the one of the D(R)AC structures as explained in section 3.1.2 with transistors itself as both the switch and the switched circuit elements.

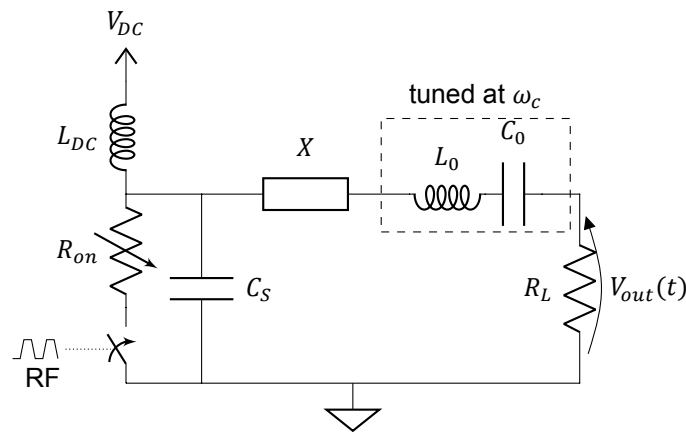


Figure 3.11: Class-E SMPA DRAC

In section 2.3.4 Class-E operation was discussed. The parameter m is a function of R_{on} and will influence the output power. This way the issue that V_{GS} cannot control the output power of Class-E is overcome. It is, however, quite difficult to analyze analytically, but R_{on} should have values quite similar to what is predicted by equations (3.6) and (3.34):

$$R_{on}[k] = \frac{1}{\chi} \frac{R_L(2^{N_b} - k)}{k} \quad (3.45)$$

where χ is some constant.

This was verified using an arbitrary simulation of an Class-E SMPA with the circuit diagram shown in figure 3.11, using an ideal switch, a digital input resolution of 6 bit and empirically determined $\chi = 2,76$. The same simulation setup is used in section 2.3.4, the full overview of Class-E design set parameters and resulting circuit component values are given in the appendix in table A.1. The choice was made to design it between load insensitivity and maximum output power ($q = 1,33$). The values of $R_{on}[k]$, corresponding output curve and linearity curves are shown in figure 3.12. From the linearity curves can be seen that the empirical value for χ results in good linearity: $INL_{pk,LSB} = 0,259$ LSB and $DNL_{pk,LSB} = 0,031$ LSB.

When aiming to implement, for example, a thermometer-coded DRAC by using transistor R_{on} with actual transistors one should target every transistor to have a R_{on} of

$$R_{on,k} = \frac{1}{\chi} \frac{R_L(2^{N_b} - k)(2^{N_b} - k + 1)}{2^{N_b}} \quad (3.46)$$

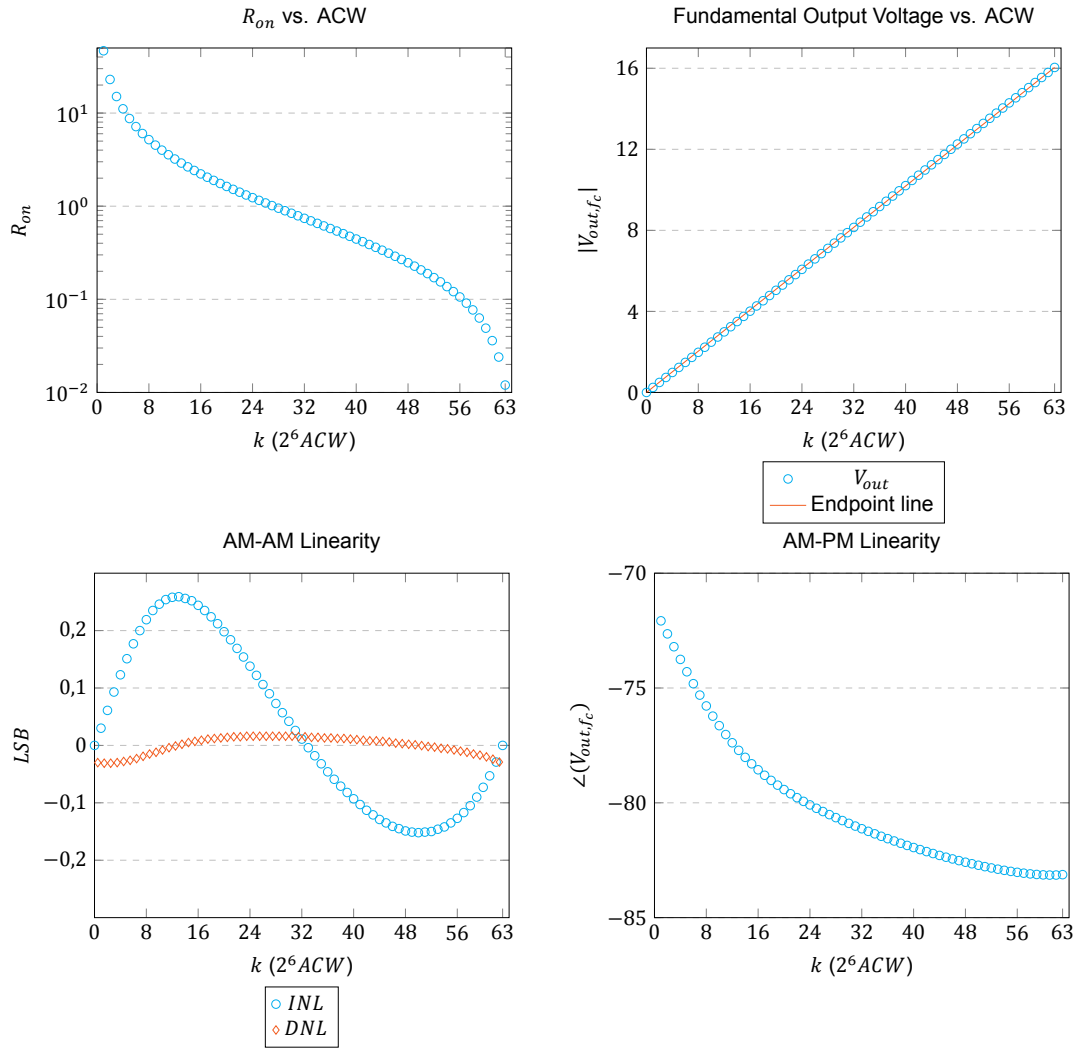


Figure 3.12: Ideal switch Class-E DRAC output curves with implemented thermometer-coded $R_{on}[k]$ using $\chi = 2,76$

similar to equation (3.35), but with added χ , where $R_{on,k}$ is the R_{on} of the k^{th} transistor being turned on. Remembering that R_{on} is inversely proportional to gate width W_g , the different transistor R_{on} values can be obtained.

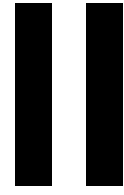
However, transistors have the tendency not to be an ideal switch with a resistor in series and some capacitance in parallel, seen from the drain to source. Another, more pragmatic, method can be used to solve this issue, especially in the already difficult to analytically analyze situation of a Class-E SMPA. By first simulating with transistors that all have the same size $R_{on,k} = R_{on}$, one can analyze the resulting AM-AM curve, invert this curve and translate that to non-linearly sized transistors, as done in [33, 34]. In these papers it was chosen to divide the thermometer-coded DRAC into 8 segments where all transistors in one segment have the same size and therefore piecewise linearizing the AM-AM curve. This was done to limit complexity and improve the predictability and matching of the individual transistors. It is expected that 8 segments will generally result in good enough linearity of the AM-AM curve.

Also, as can be seen in the fourth graph of figure 3.12, even in this idealized case there is significant AM-PM distortion present in the Class-E SMPA DRACs. In the paper this issue was overcome by applying multiphase RF clocking: applying different time delays to the phase modulated RF clocks for each segment to straighten the phase response of the DRAC.

3.4. Conclusion

A DRAC can be seen as a Nyquist rate DAC with its output signal multiplied by a square wave at the carrier frequency. Its building blocks are switched resistors or switched current sources, which can be arranged in a binary-weighted, thermometer-coded or hybrid fashion. Binary-weighting results in less circuit elements but poor matching between these elements, resulting in non-linearity and possible glitches. The thermometer-coded topology does not have this matching problem, but results in a large amount of circuit elements. The hybrid solution tries to balance the advantages and disadvantages of either solution.

The AC coupled DRAC concept can be combined with the Class-E SMPA using R_{on} to modulate the output voltage, which overcomes the Class-E limiting factor of uncontrollable output power by V_{GS} . R_{on} can be digitally controlled by switching differently sized transistors. The transistors ideally function as a switched resistance in this Class-E scenario. There is a non-linear relationship between R_{on} and the output voltage, but this can be linearized in a thermometer-coded topology by sizing the transistors according to equation (3.45). As transistors are neither a resistor nor ideal, the more pragmatic solution is by simulating the AM-AM curve with transistors of the same size, invert this curve and translate that to non-linearly sized transistors.



Implementation

4

System Design Considerations

In this chapter the different design choices made during the design process are established. First in section 4.1 a choice is made between the available technologies for the power RF-DAC. In section 4.2 the minimum required resolution is determined. By constraints in technology the RF-DAC will need an off-chip driver, in section 4.3 the options for this driver are examined which also result in constraints on the maximum operation frequency. Finally in section 4.4 a suitable structure for the DRAC and topology are chosen, while assuming a limited amount of connections from the off-chip driver.

4.1. Technology

Semiconductor technologies keep improving over time. The most known example is the technology scaling of silicon CMOS, as already mentioned in section 1.2.2. With technology scaling the speed and level of integration of silicon CMOS is ever increasing, making it a very interesting technology for something like a RF-DAC. However, with supply voltages close to 1 V to avoid breakdown it is easy to see that silicon CMOS isn't very suited to efficiently generate higher powers.

For generating higher powers other technologies are available. A selection of power RF technologies is shown in figure 4.1. In this figure silicon does not refer to silicon CMOS but to specialized high power silicon technologies, such as LDMOS which has a much higher breakdown voltage.

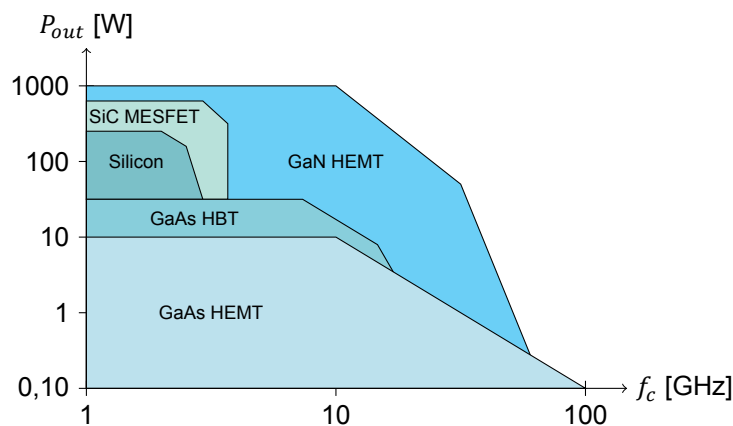


Figure 4.1: Technology comparison in terms of output power and speed [35]

The basic idea is to implement a power DRAC on such high power technology. It will be driven by some off-chip driver. This driver may be existing general purpose RF equipment. Or a special CMOS driver chip could be designed to combine the benefit of the level of integration of CMOS and the power of a specialized technology into one power RF-DAC. For a possible CMOS driver implementation the 40 nm 1P7M technology by Taiwan Semiconductor Manufacturing Company (TSMC) is available.

For this project there are two available power technologies that may be considered for the design of the DRAC. First is a GaN Monolithic Microwave Integrated Circuit (MMIC) process by Fraunhofer-Institut für Angewandte Festkörperphysik (IAF) with $L_g=0,25\ \mu\text{m}$ on a $100\ \mu\text{m}$ thick SiC substrate. Secondly a silicon LDMOS process by Ampleon on a $50\ \mu\text{m}$ thick substrate in the LM8 process is considered. In the next sections these technologies will be characterized in context of use in a Class-E power DRAC.

4.1.1. Fraunhofer IAF GaN25

The Fraunhofer IAF GaN25 is a AlGaIn/GaN heterojunction High-Electron-Mobility Transistor (HEMT) process, designed for high power RF applications [36]. The high power is consistent with the position of GaN HEMTs in figure 4.1. It is a MMIC process that allows the integration of components on-chip, such as microstrip transmission lines, Metal-Insulator-Metal (MIM) capacitors and NiCr thin film resistors. The backside of the substrate is a ground plane to which vias can be created to have access to high quality, low series inductance ground at the top of the MMIC. The high power transistors can have vias for their source directly connecting to the ground at the back of the substrate, providing for low local inductive feedback, but is not required. This ground connection is included in the internal transistor models. Only N-channel devices are possible in GaN. The GaN devices can operate at a temperature of maximally $250\ ^\circ\text{C}$, have a I_{max} of at least $800\ \text{mA mm}^{-1}$, a nominal supply voltage up to $40\ \text{V}$ and a breakdown voltage V_{bk} of at least $80\ \text{V}$. It has been reported that when the devices are off the breakdown voltage exceeds $150\ \text{V}$ [37].

DC Characteristics

For the technology characterization the device model of figure 2.4 is assumed. A first step is by characterizing the non-linear intrinsic current source by plotting the device I-V curves. Figure 4.2 shows the DC $I_{DS}-V_{GS}$ curves for a $12\times 150\ \mu\text{m}=1,8\ \text{mm}$ transistor with individual source vias. The GaN devices are depletion mode HEMTs, meaning their threshold voltage V_T is negative. Depletion mode devices are also called “normally on” as the device is on for a zero gate voltage. From the figure it can be estimated that $V_T \approx -2,5\ \text{V}$.

Something interesting happens to I_{DS} when $V_{GS} > 1\ \text{V}$. For low V_{DS} the I_{DS} drops for higher V_{GS} . The solid line in figure 4.2 is the current that flows into the drain, while the dashed line is the current that flows through the intrinsic current source I_{DS} . No change in current is seen there. This is because the GaN HEMT does have a Schottky diode as gate instead of a Metal-Oxide-Semiconductor (MOS) gate. For higher V_{GS} it becomes forward biased and current can flow from the drain to the gate, which does not contribute to the current through the intrinsic I_{DS} . This is unwanted behavior that also invalidates the assumed FET model and therefore V_{GS} should not exceed $1\ \text{V}$. For a $1,8\ \text{mm}$ device the I_{max} should be $1,44\ \text{A}$, which is consistent with what can be seen in figure 4.2. Only for $V_{DS} = 5\ \text{V}$ the current is more than, even more than at the higher voltages. This can also be seen in the $I_{DS}-V_{DS}$ curves.

The $I_{DS}-V_{DS}$ are shown in figure 4.3. The line for $V_{GS} = 1,5\ \text{V}$ is dotted since it was established that V_{GS} should not exceed $1\ \text{V}$, but is included for completeness. A peak I_{DS} can be seen for $V_{DS} \approx 5\ \text{V}$. Why this happens is not fully clear, but it is suspected that this might be due to thermal effects included in the model. Neither the model nor its documentation is very clear about the thermal effects other than that they are included. In this DC simulation it might be assumed that the device will operate continuously at that operating point, dissipating $V_{DS} \cdot I_{DS}$ as heat in the device and assume settling at some operation temperature. Changing device model parameters regarding temperature do yield different DC curves, though some form of this I_{DS} -peak is always visible. The curves shown in figure 4.3 are for the default parameters. Also R_{on} is illustrated, which will be determined in the next section.

Class-E Relevant Parameters

For Class-E operation it can be recalled from section 2.3.4, specifically equation (2.80), that the product of the on-resistance R_{on} and parasitic output capacitance C_{DS} partly predict how well a technology can perform regarding power efficiency. Equally important is the parasitic input capacitance C_{GS} , to predict the amount of required input power to drive the device. In order to determine these parameters a scattering parameter (S-parameter) simulation is used with the circuit shown in figure 4.4. The drawn capacitors are ideal DC block capacitors and the inductors are ideal DC feed inductors/RF chokes. All parameters will be normalized to $1\ \text{mm}$ transistor Total Gate Width (TGW).

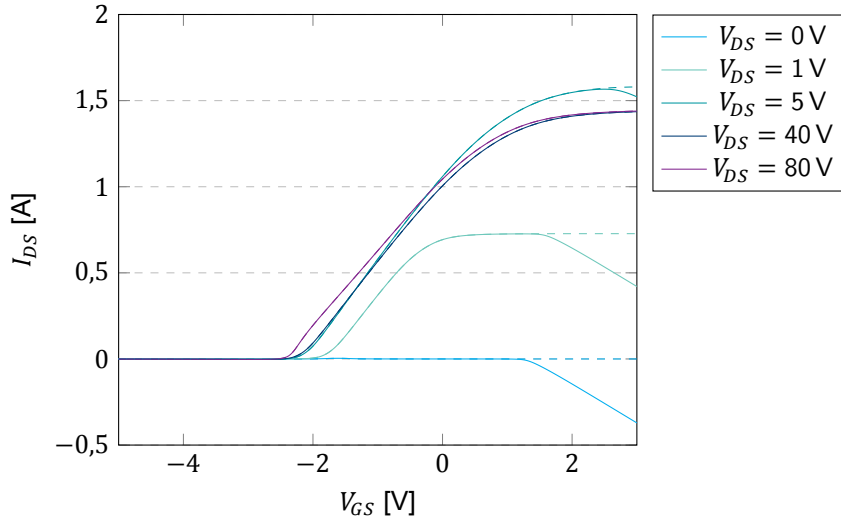
Fraunhofer IAF GaN 0,25 μm , CS_LS_ISV_enc 12x150 μm DC curve

Figure 4.2: DC I_{DS} - V_{GS} curves for Fraunhofer IAF GaN, Common Source, Large Signal model, Individual Source Via, 12 fingers \times 150 μm , dashed: internal I_{DS} node

Assuming the switched device model from figure 2.5 the admittance parameters (Y-parameters) calculated from the S-parameters can be used to determine the model parameters that are searched for. The Y-parameters of interest are Y_{11} and Y_{22} :

$$Y_{11} = j\omega C_{GS} \quad (4.1)$$

$$Y_{22} = \frac{1}{R_{on}} + j\omega C_{DS} \quad (4.2)$$

and therefore

$$R_{on} = \frac{1}{\Re(Y_{22})} \quad (4.3)$$

$$C_{DS} = \frac{\Im(Y_{22})}{\omega} \quad (4.4)$$

$$C_{GS} = \frac{\Im(Y_{11})}{\omega} \quad (4.5)$$

$$(4.6)$$

In figure 4.5 the reciprocal of real part of Y_{22} is plotted for a normalized GaN device with a TGW of 1 mm. The simulation was done at 1 GHz for a large and a small device (1,8 mm and 0,54 mm respectively). The drain to source resistance per unit gate width appears to be independent of transistor TGW, indicating that normalizing R_{on} to a per unit gate width is representative for a range of transistor sizes. R_{on} is simulated to be 2,17 Ω mm for $V_{GS} > V_T$.

Up to now the switch was assumed to be perfect in the switch model. However, the off resistance is also finite and of importance to the efficiency when operating in Class-E. Assuming a peak voltage of $V_{bk} = 150$ V and a peak instantaneous power dissipated in the device of $V_{bk}^2/R_{off} < 1$ W, the off-resistance R_{off} has to be at least 22,5 k Ω mm. Larger R_{off} is preferred to turn the transistor 'off enough' to prevent it from dissipating significant power while off, therefore $V_{GS} \leq -4$ V in the part of the RF cycle that is responsible to turn the transistor.

Finding the output capacitance C_{DS} per unit gate width is slightly more complicated. The imaginary part of Y_{22} divided by ω is plotted in figure 4.6a, again for a normalized GaN device with $TGW = 1$ mm at 1 GHz for a large and small device. C_{DS} per unit gate width appears to be a function of V_{DS} , V_{GS} as well as the transistor TGW.

The only time C_{DS} is important for the Class-E operation is when the device is off: when it is on C_{DS} is shorted to ground. At this point the aim is to find the technology limits, so the best performance can be

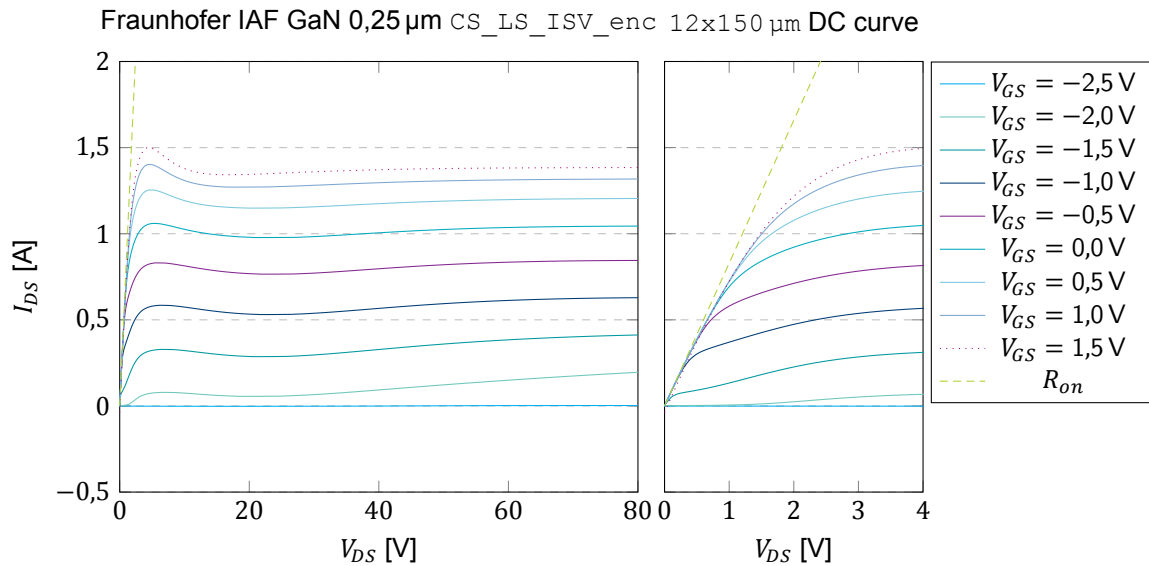


Figure 4.3: DC I_{DS} - V_{DS} curves for Fraunhofer IAF GaN, Common Source, Large Signal model, Individual Source Via, 12 fingers \times 150 μm , right: zoomed in on the triode region

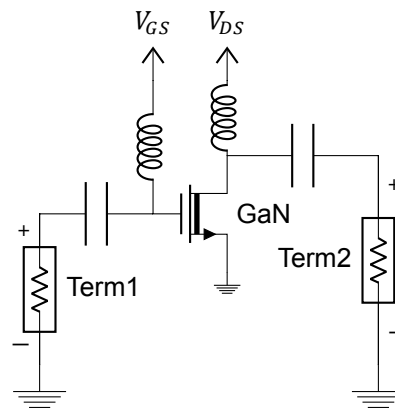


Figure 4.4: S-parameter simulation circuit

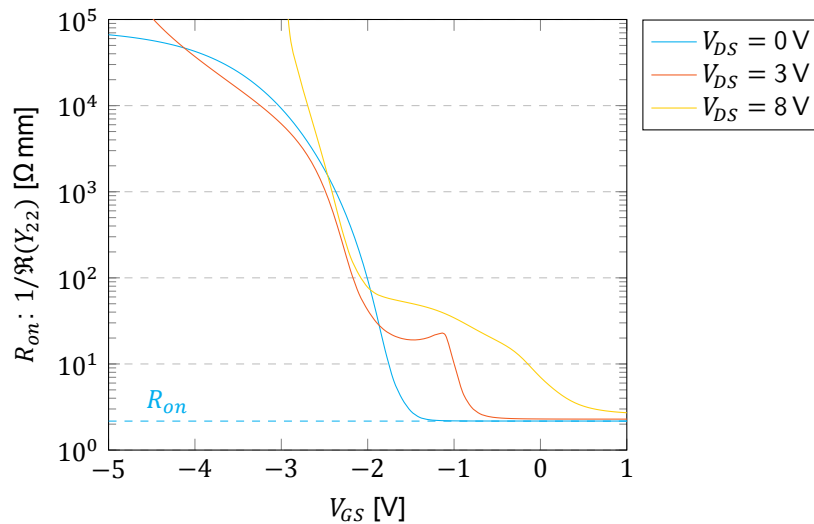
expected for the largest device as it has a lower C_{DS} per unit gate width. That leaves the dependency on V_{DS} : the effective C_{DS} is expected to be somewhere between $0,26 \text{ pF mm}^{-1}$ to $0,62 \text{ pF mm}^{-1}$. In ideal Class-E operation V_{DS} will start increasing from 0V the moment the switch closes to V_{bk} and return to 0V since the ZVS condition will be satisfied. The effective C_{DS} can then be estimated by finding the charge q_{DS} in C_{DS} divided by the voltage V_{DS} by definition

$$C \triangleq \frac{q}{V} \quad (4.7)$$

The charge in C_{DS} can be found by applying a step voltage from 0V to V_{bk} in a transient simulation and then integrating the current that flows into the drain. Again assuming $V_{bk} = 150 \text{ V}$ simulating this gives $C_{DS} = 0,70 \text{ pF mm}^{-1}$. The other way around, a step voltage from V_{bk} to 0V, gives $C_{DS} = 0,60 \text{ pF mm}^{-1}$. Assuming the effective C_{DS} is the average of these the value becomes $C_{DS} = 0,65 \text{ pF mm}^{-1}$. However, if it is assumed that the starting voltage of V_{DS} to be $R_{on}I_{DS}$ where it is assumed that $I_{DS} = I_{max}$ then $V_{DS} = 2,17 \cdot 1,44 \approx 3 \text{ V}$. Applying a step voltage from 3V to V_{bk} yields $C_{DS} = 0,56 \text{ pF mm}^{-1}$, resulting in an effective C_{DS} of $0,58 \text{ pF mm}^{-1}$.

It may be clear that this is a relatively unreliable method, but it may give a decent starting point for the Class-E SMPA DRAC design. At least it can be concluded that using multiple smaller transistors will yield a worse performance than using one big transistor that has the same TGW.

In figure 4.6b the imaginary part of Y_{11} divided by ω is plotted, with the same conditions as previous

Fraunhofer IAF GaN 0,25 μm , CS_LS_ISV_enc, normalized $TGW = 1 \text{ mm}$ Figure 4.5: GaN25: Drain to source resistance as function of V_{GS} to determine R_{on}

two plots. Finding the input capacitance C_{GS} per unit gate width is slightly easier than it was with the output capacitance. First it is independent from TGW , as implied by only one line style in the plot. Secondly, only the transition from off to on needs to be considered to predict required drive power, as that is the energy that has to be delivered by the driver's power supply. The transition from on to off will dissipate the charge from C_{GS} to ground. Assuming the ZVS condition is met, V_{DS} will be 0 V when turning on the transistor.

Applying the same integrating technique to find the charge in C_{GS} , but now with a step voltage from -4 V to 0 V , gives $C_{GS} = 1,50 \text{ pF mm}^{-1}$. If the ZVS condition is not met and it is assumed that, e.g., $V_{DS} = 20 \text{ V}$ the simulated integration gives $C_{GS} = 1,86 \text{ pF mm}^{-1}$. This may seem initially counter intuitive as C_{GS} should decrease from the triode region to the transconductance region, as explained in section 2.3.1. However, this was assuming a MOS device and not a GaN HEMT with a Schottky junction at its gate. Looking at figure 4.6b it might be explained by C_{GS} rising at a lower V_{GS} .

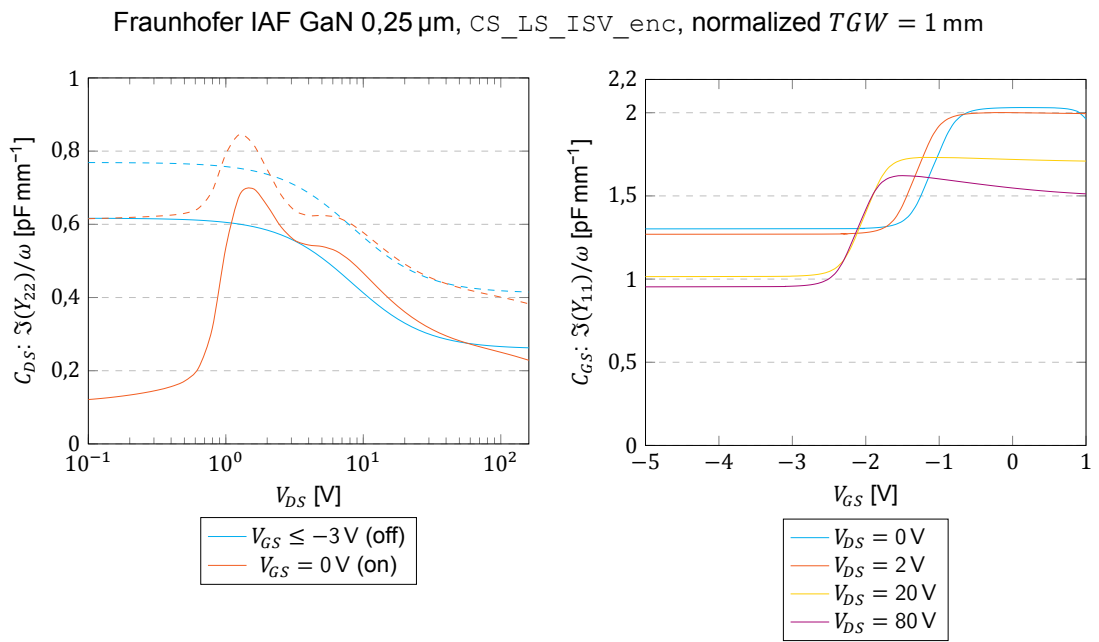


Figure 4.6: GaN25: Non-linear capacitance characterization

4.1.2. Ampleon LDMOS LM8

The Ampleon LDMOS LM8 is a silicon MOS process, designed for power RF communication applications requiring good linearity. From figure 4.1 it can be seen that GaN HEMT is expected to have a higher power density than silicon (LDMOS). Also a lower possible operating frequency for the LDMOS is expected. The backside of the substrate is a ground plane, the transistor source connections can be diffused through the substrate directly to the backside to have access to high quality, low series inductance ground, providing for low local inductive feedback, but is not required. Only high performance NMOS devices are available in Ampleon's LDMOS LM8. In this process MIM capacitors are also available. The nominal supply voltage is 28 V. Other specifications are not documented, but V_{bk} will have to be larger than 56 V to be of use for any transconductance amplifier class using a 28 V supply. The device models are scalable and based on measurement of a device with a Total Gate Width (TGW) of 2,3 mm.

DC Characteristics

For the technology characterization again the device model of figure 2.4 is assumed. The non-linear intrinsic current source is characterized first by plotting the device I-V curves, as shown in figures 4.7 and 4.8. A device with 2 cells with each 2 gate fingers are used, each gate finger has a length of 250 μm , giving a TGW of 1 mm. From the figure it can be estimated that $V_T \approx 2,1 \text{ V}$.

From figure 4.7 can be seen that increasing V_{GS} beyond 6 V does not increase I_{DS} much, which is largely supported by figure 4.8. From this can be estimated that $I_{max} \approx 0,25 \text{ A mm}^{-1}$. The breakdown voltage is clearly visible and occurs at $V_{bk} \approx 64 \text{ V}$. The device models can take thermal effects into account, but its influence can be scaled freely. For these DC simulations the influence of thermal effects are disabled.

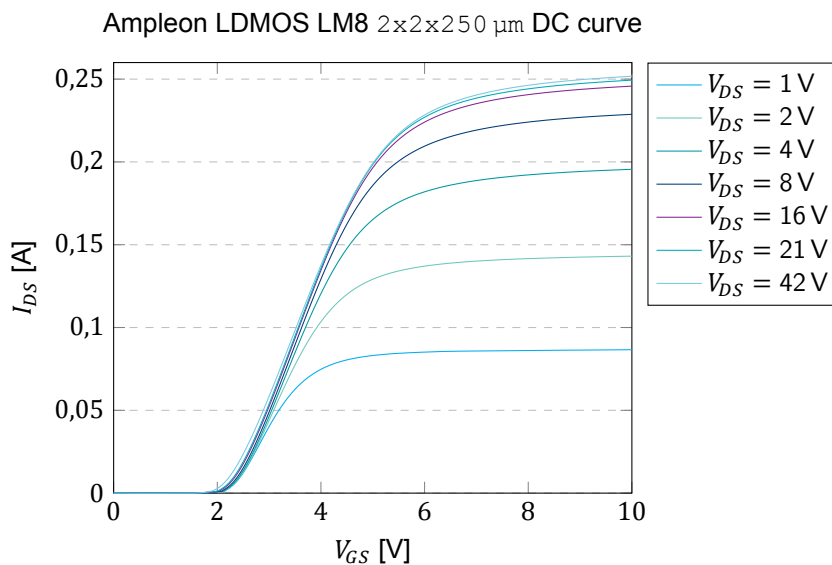


Figure 4.7: DC I_{DS} - V_{GS} curves for Ampleon LDMOS, 2 cells x 2 fingers x 250 μm

Class-E Relevant Parameters

For Class-E operation R_{on} , C_{DS} and C_{GS} are of importance. These are determined similarly as done for the GaN25 technology: a S-parameter simulation is used with the circuit shown in figure 4.4, but then the GaN transistor is replaced with a LDMOS transistor. All parameters will be normalized to 1 mm transistor TGW. In the LDMOS model changing parasitics due to different TGW is not included, therefore the parameters per unit width are independent of TGW.

In figure 4.9 the reciprocal of real part of Y_{22} is plotted for a normalized LDMOS device with a TGW of 1 mm. The simulation was done at 3 GHz, R_{on} is simulated to be 10,5 $\Omega \text{ mm}$ for $V_{GS} > V_T$.

The imaginary part of Y_{22} divided by ω is plotted in figure 4.10a to find the output capacitance C_{DS} per unit gate width, again for a normalized LDMOS device with $TGW = 1 \text{ mm}$ at 3 GHz. When the

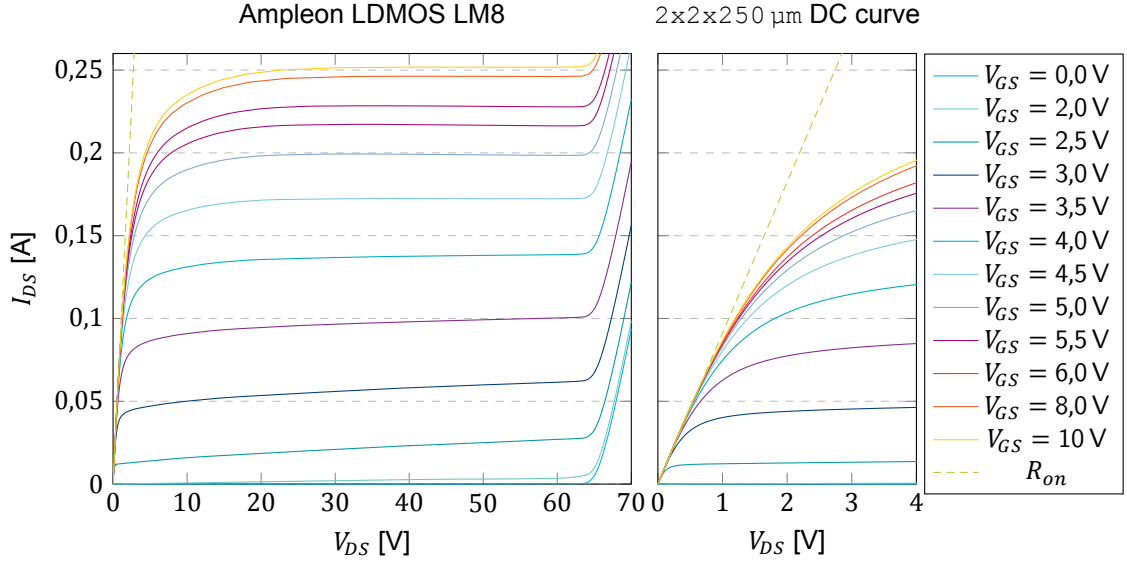


Figure 4.8: DC I_{DS} - V_{DS} curves for Ampleon LDMOS, 2 cells \times 2 fingers \times 250 μm , right: zoomed in on the triode region

transistor turns on $\Im(Y_{22})$ becomes negative for lower V_{DS} . This is most likely due to the device being conductive, hence being more inductive than capacitive.

C_{DS} has a value from 0,287 pF mm $^{-1}$ to 0,348 pF mm $^{-1}$ when the device is off, much closer together compared to the GaN technology. To estimate an effective value the same technique as in the GaN case can be used: applying a step voltage from 0 V to V_{bk} to C_{DS} in a transient simulation and integrating the current that flows into the drain to find the charge q_{DS} in C_{DS} . Assuming $V_{bk} = 56$ V and $V_{GS} = 1,2$ V gives $C_{DS} = 0,298$ pF mm $^{-1}$. The other way around, a step voltage from V_{bk} to 0 V, gives the same result. Assuming the starting voltage of V_{DS} to be $R_{on} \cdot I_{DS}$ where it is assumed that $I_{DS} = 0,25$ A then $V_{DS} = 10,50 \cdot 0,25 \approx 3$ V. Applying a step voltage from 3 V to V_{bk} yields $C_{DS} = 0,296$ pF mm $^{-1}$. Therefore it is assumed from now on that the effective C_{DS} is 0,297 pF mm $^{-1}$.

In figure 4.10b the imaginary part of Y_{11} divided by ω is plotted, with the same conditions as previous two plots. To find the effective C_{GS} the charge integration technique is used again. Assuming the ZVS condition is met, V_{DS} will be 0 V when turning on the transistor. Simulation then gives the effective $C_{GS} = 0,860$ pF mm $^{-1}$. Increasing V_{DS} will slightly lower the effective C_{GS} which is consistent with the expectation from section 2.3.1: for e.g. $V_{DS} = 7$ V simulation gives effective C_{GS} then to be 0,835 pF mm $^{-1}$ down to 0,828 pF mm $^{-1}$ for $V_{DS} = 28$ V.

4.1.3. Technology Evaluation

In table 4.1 a comparison of the two technologies in terms of Class-E relevant parameters is given. GaN25 is expected to outperform LDMOS LM8 in terms of power efficiency for a given operation frequency, due to its much lower R_{on} even though C_{DS} is higher, as can be concluded by the lower value for m (see ‘‘Generalized Class-E Operation’’ of section 2.3.4). Also its output power is expected to be higher than that of LDMOS since I_{max} and V_{bk} are higher.

However, the input capacitance of LDMOS is lower than that of GaN, therefore requiring less drive power for the same TGW and input signal swing. That may be compensated though by higher expected power gain from the GaN technology.

Both technologies require an external driver, since both technologies are targeted at power applications and therefore only have large NMOS transistors suitable for high voltages. They do not have complementary transistors available that can do high speed digital logic in an energy efficient manner. This digital logic will be required to decode the Amplitude Code Word (ACW) to the DRAC which will require multiple transistors to be switched separately, as can be taken from section 3.3. The GaN technology will need an input signal from -4 V to 0 V. There is no out-of-the-box solution available that is capable of supplying multiple channels with this voltage range, so a specialized driver will have to be designed.

When a specialized DRAC driver has to be designed anyway other functionality can be implemented

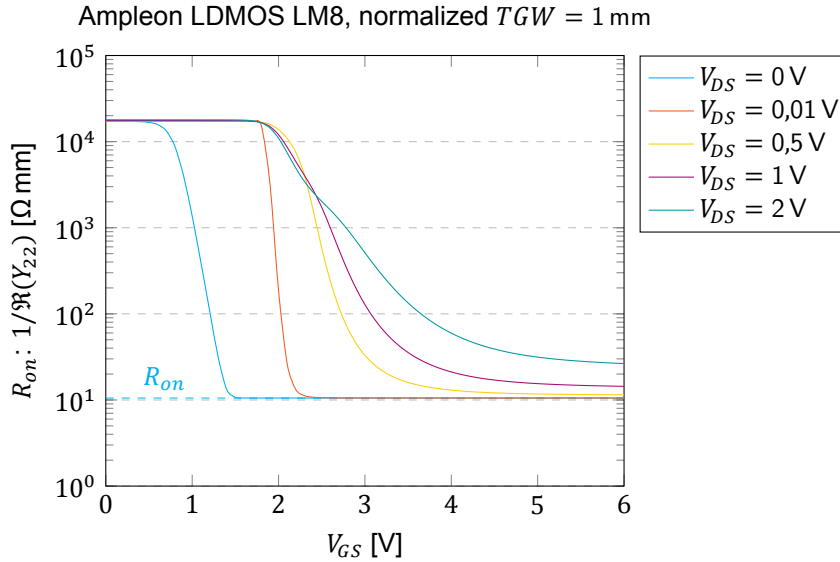
Figure 4.9: LDMOS LM8: Drain to source resistance as function of V_{GS} to determine R_{on}

Table 4.1: Comparison between GaN25 and LDMOS LM8

Quantity	GaN25		LDMOS LM8	
	min	max	min	max
R_{on} [$\Omega \text{ mm}$]		2,17		10,5
C_{DS} [pF mm^{-1}]	0,58	0,65		0,297
C_{GS} [pF mm^{-1}]	1,50	1,86	0,828	0,860
I_{max} [A mm^{-1}]		0,80		0,25
V_{bk} [V]	>80	>150 ¹	>56	$\approx 64^2$
m/ω_c [$\text{pF } \Omega$]	1,3	1,4		3,12

on this driver as well, such as including a phase modulator to make it a full-fledged RF-DAC. The LDMOS has the upper hand in this case. The technology allows for small batches with variations in production process, so for example a V_T -shift can be applied. From Ampleon the information was supplied that the V_T can be lowered by 1,2 V. This comes at a cost of an increased C_{GS} to $1,50 \text{ pF mm}^{-1}$. That does mean that the voltages required to drive the LDMOS transistors is now within reach of the voltage swing of TSMC 40 nm CMOS process, which is for the thick oxide devices a maximum of 2,5 V. To allow for some Process, Voltage and Temperature (PVT) variations and calibration afterwards the targeted drive voltage is decided to be 2 V.

This power RF-DAC project started by making a DRAC design in GaN, but with advancing insights after approximately a half year the choice was made to switch to LDMOS technology for above reasons. The rest of this thesis will mainly focus on the DRAC design in LDMOS with applied V_T -shift. This V_T shift is modeled by driving the LDMOS devices with a DC offset of +1,2 V and multiplying C_{DS} by 1,7 what can be done by setting transistor model parameter $MCGS = 1,7$. For the design work done on GaN please see appendix C.

4.2. Resolution and Dynamic Range

The Class-E SMPA DRAC will require a certain resolution N_b . From the thesis goals can be remembered that an ACPR of better than 50 dBc is needed. That directly means that the system's Dynamic Range should be better than 50 dB. A minimum number of bits N_b can be calculated using equati-

¹When transistor is off [37]

²From DC curves

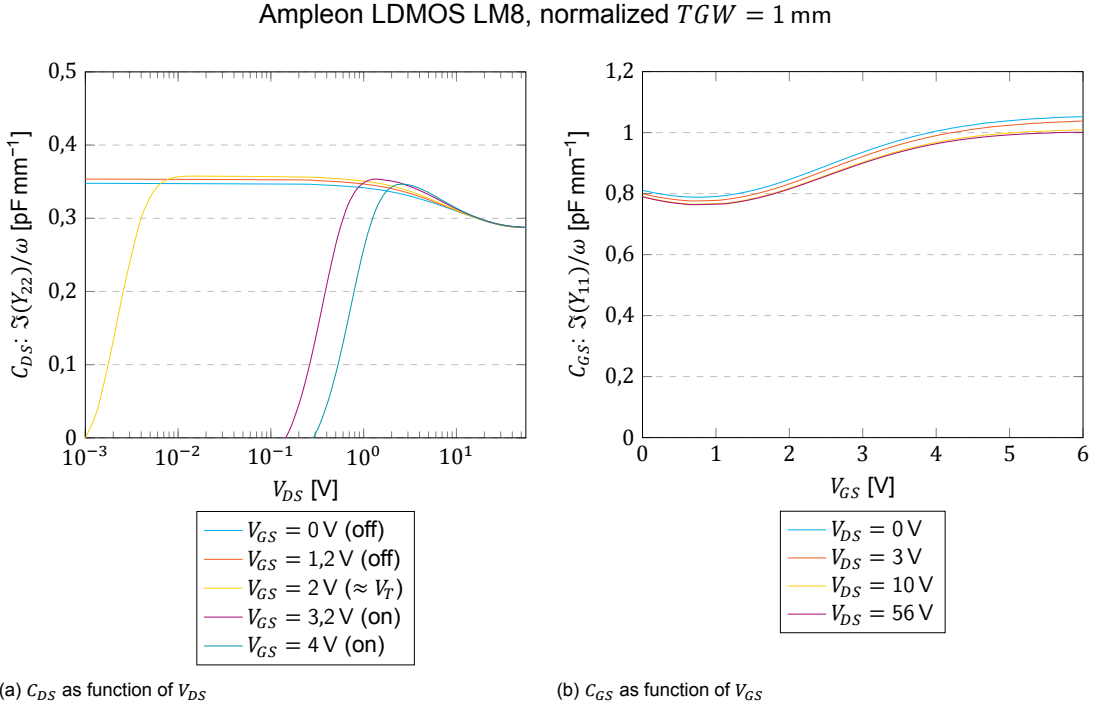


Figure 4.10: LDMOS LM8: Non-linear capacitance characterization

ons (3.41)–(3.44).

This does require some additional assumptions. First it is assumed that

$$f - f_c \ll f_{rs} \quad (4.8)$$

where f is the frequency of the adjacent channel and therefore

$$\text{sinc}\left(\frac{f - f_c}{f_{rs}}\right) \approx 1 \quad (4.9)$$

It is further assumed that $f_{rs} = 625$ MHz [33], $f_{BB} = 40$ MHz and QAM256 modulation is used that has a PAPR of 8,2 dB. That gives

$$50 \text{ dB} < 6,021N_b + 1,761 + 0 + 10 \log_{10}\left(\frac{625 \cdot 10^6}{2}\right) - 8,2 - 10 \log_{10}(40 \cdot 10^6) \quad (4.10)$$

$$< 6,021N_b + 2,489 \quad (4.11)$$

Giving $N_b > 7,90$. So, from now on a minimum resolution of $N_b = 8$ bit is assumed.

4.3. Driver

The LDMOS transistors of the Class-E SMPA DRAC will need to be individually driven by some off-chip driver, as mentioned in section 4.1.3. The driver must be capable of driving a capacitive load, as the LDMOS transistor gate can be primarily modeled by a capacitor. The DRAC linearization technique that is described in chapter 3 will most definitely result in a range of differently sized LDMOS transistors, and therefore also a range of differently sized load capacitances. The goal is to have a universal design strategy for the driver.

Even if the digital logic cannot be integrated on the LDMOS DRAC, it might be possible to have some final driver stage embedded on the LDMOS chip that in turn will be driven by an external CMOS chip. This possible final driver stage will need to be directly connected to the input of the DRAC transistors, as no external impedance matching or harmonic tuning can be applied. The only option here would be the simplest possible Class-A configuration that consists of a resistor connected to a positive supply with

in series a transistor to ground as shown in figure 4.11a. Its input and output will have some sinusoidal waveform. To maintain proper switch-like behavior from the DRAC transistor it would require a quite high voltage amplitude to have a steep waveform around the threshold voltage to prevent slow switching and with it switching losses. This type of driver will continuously consume power even when the DRAC transistor is supposed to be stay off. Also, the bias current required to result in switching behavior of the DRAC transistor requires a drive transistor that is (at least) just as large as the transistor that is to be switched and will therefore only uselessly waste (static) power.

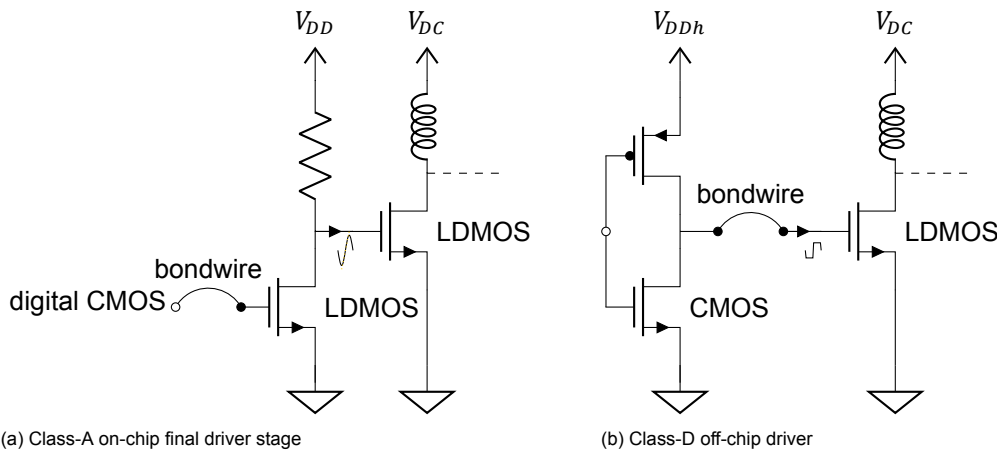


Figure 4.11: Options for the DRAC driver

The transistors in a SMPA will only have two states: off or on; ‘open’ or ‘closed’; ‘0’ or ‘1’. That means it can be driven with a ‘digital’ driver like an inverter as shown in figure 4.11b. In principle that is the same as a Class-D driver, resulting in a square voltage waveform. This will require a much lower voltage swing to have a steep switching edge and therefore much lower power consumption compared to a Class-A driver stage. The power consumption of this Class-D driver can be simply given by

$$P_{drive} = f_c \cdot C_{tot} \cdot V_{DDh}^2 \quad (4.12)$$

where f_c the switching frequency, C_{tot} the total capacitive load that needs to be driven, mainly including C_{GS} of the LDMOS transistor, and V_{DDh} the supply voltage of the thick oxide transistors in the 40 nm-CMOS process. It should be emphasized that all DRAC transistors need to be driven individually because not all transistors should be on when not the full output power is required. The drivers for the transistors that remain in the off state will not switch and therefore not use any power. Therefore input power scales with input ACW, hence also with output voltage and power. This is a big advantage over currently used analog Class-A/AB drivers for Class-E PAs.

The Class-D off-chip driver is the best choice available with the given technology options, both in terms of feasibility and of power efficiency.

4.3.1. Driver Model

The chosen Class-D driver needs to be off-chip (figure 4.11b) due to the lack of available PMOS devices. This implies that some connection from the 40 nm-CMOS chip to the LDMOS chip is required. Some sort of bondwire is required to create this connection, which will have parasitics. It is important to model these parasitics to ensure that the waveform at the gate of the LDMOS stays by approximation digital (a square wave) with the correct moment of switching. As mentioned before in section 2.3.3, a square wave contains the fundamental and odd harmonics. All parasitics combined should allow for the fundamental and at least the third harmonic to pass with negligible difference in amplitude and phase to retain a squarish shape. More harmonics that meet this criterion are preferred.

Because two different physical chip dies will be used they need to be protected against ESD. This is implemented using some protection diodes at the output of the CMOS chip as well at the input of the LDMOS chip. These will significantly add capacitance at either side of the inter-chip connection. The connection itself is modeled by an inductor with series resistance. The driver’s CMOS transistors are modeled as switch with an Equivalent Series Resistance (ESR) R_{es} , as introduced in section 2.3.1. The total circuit model is shown in figure 4.12.

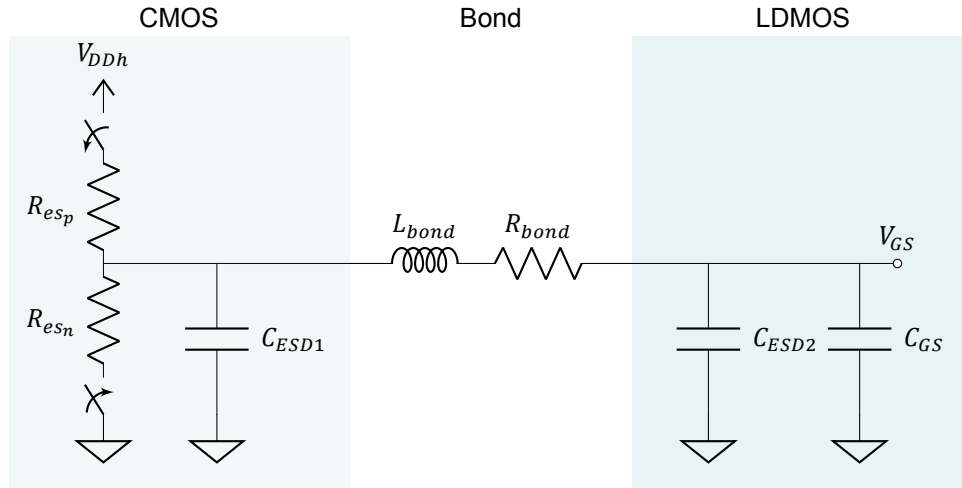


Figure 4.12: Class-D off-chip driver model

The driver can be simplified to a square wave voltage source with an ESR if it is assumed that $R_{es_n} = R_{es_p} = R_{es}$, also leading to equal rise and fall times for V_{GS} ¹. This simplification is shown in figure 4.13. Assuming the frequency is low enough the inter-chip bond inductance is negligible. Also

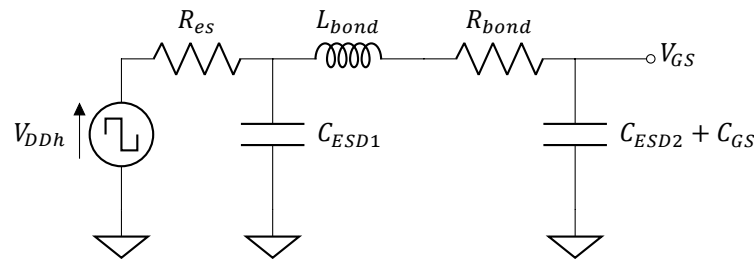


Figure 4.13: Simplified Class-D off-chip driver model

assuming R_{bond} can be neglected, then only a first order RC-lowpass filter remains with

$$\tau_d = R_{es, simple}(C_{ESD1} + C_{ESD2} + C_{GS}) \quad (4.13)$$

where τ_d is the driver's time constant. This time constant can be seen as a time delay between the driver and the gate of the LDMOS transistor. For correct Class-E operation of the DRAC this delay should be constant for all differently sized transistors to prevent both the possibility of glitches and the additional creation of unwanted AM-PM non-linearities. When specifying τ_d as a design constant then the value for R_{es} can be calculated as a function of the TGW of the LDMOS transistor using equation (4.13). Taking R_{bond} also into account the total delay of this chain can be estimated by the Elmore delay since the drive signal is assumed to be digital [23]

$$\tau_{d, tot} = R_{es}C_{ESD1} + (R_{es} + R_{bond})(C_{ESD2} + C_{GS}) \quad (4.14)$$

Combining equations (4.13) and (4.14) then gives

$$R_{es} = \frac{\tau_d}{C_{ESD1} + C_{ESD2} + C_{GS}} - R_{bond} \frac{C_{ESD2} + C_{GS}}{C_{ESD1} + C_{ESD2} + C_{GS}} \quad (4.15)$$

$$= R_{es, simple} - R_{bond} \frac{C_{ESD2} + C_{GS}}{C_{ESD1} + C_{ESD2} + C_{GS}} \quad (4.16)$$

The lower the calculated R_{es} , the larger the transistors in the Class-D driver will have to be. Therefore the value R_{es} can be seen as some measure of 'drive strength' and will be done throughout the rest of this thesis, e.g. a stronger driver means a lower R_{es} .

¹This is assuming no non-linearity of C_{GS} or C_{ESD} .

So far it seems that a lower R_{es} will be better, because it will mean a lower delay and a more square shape of the drive signal. However, that is by approximating the driver as a simple first order RC-lowpass filter. Also a series second order RLC-lowpass filter can be identified consisting of $R_{es} + R_{bond}$, L_{bond} and $C_{ESD2} + C_{GS}$. The damping factor can be specified by definition as

$$\zeta \triangleq \frac{R}{2} \sqrt{\frac{C}{L}} \quad (4.17)$$

$$= \frac{R_{es} + R_{bond}}{2} \sqrt{\frac{C_{ESD2} + C_{GS}}{L_{bond}}} \quad (4.18)$$

A larger $R_{es} + R_{bond}$ will mean more damping of this second order filter, resulting in lower overshoot. With a preferred damping factor an appropriate value for τ_d can be calculated by substituting equation (4.15) in equation (4.18) that is solved for R .

$$\frac{\tau_d}{C_{ESD1} + C_{ESD2} + C_{GS}} - R_{bond} \frac{C_{ESD2} + C_{GS}}{C_{ESD1} + C_{ESD2} + C_{GS}} + R_{bond} = 2\zeta \sqrt{\frac{L_{bond}}{C_{ESD2} + C_{GS}}} \quad (4.19)$$

solving for τ_d gives

$$\tau_d = (C_{ESD1} + C_{ESD2} + C_{GS}) \left(2\zeta \sqrt{\frac{L_{bond}}{C_{ESD2} + C_{GS}}} + R_{bond} \frac{C_{ESD2} + C_{GS}}{C_{ESD1} + C_{ESD2} + C_{GS}} - R_{bond} \right) \quad (4.20)$$

This is all a coarse estimation, neglecting any mutual inductance, but provides for a well-founded starting point for the design.

The driver's CMOS transistors will also have parasitic output capacitances but are disregarded for now. These parasitic output capacitances will be proportional to the transistor width, inversely proportional to the ESR and with that approximately proportional to the total capacitive load, especially when R_{bond} is low. These output capacitances will be taken into account upon designing the CMOS driver itself.

The ESD protection capacitance values per connection are 600 fF for 40 nm-CMOS by TSMC [38] and 500 fF for LDMOS by Ampleon. The value of τ_d will have to be determined during the design process itself. L_{bond} and R_{bond} will be dependent of the inter-chip connection method, which will be described in the next section.

4.3.2. Inter-Chip Connection Methods

The chips will be connected to each other using some sort of wire. The longer the connection, the larger L_{bond} will be. As a rule of thumb a partial inductance of 1 nH mm^{-1} can be assumed for bondwires, although the actual value is dependent on wire thickness, length, distance from the substrate and mutual inductance between other wires.

Two methods are possible to connect the two chips. First is the 'traditional' bondwire, done by placing the two chips beside each other and connecting them with wires. The other option is using a flip-chip connection, where the two chips are soldered on top of each other. This method has a much smaller distance, thus L_{bond} , but is financially more expensive. Next these methods will be evaluated using the previously described driver model.

Bondwires

In figure 4.14 the principle of bondwires is illustrated. Two offset rows of bondpads are shown, known as staggered bondpads. This is done to effectively halve the pitch between the bondpads, doubling the amount of individual wires for a given chip width. The distance between a set of bondpads is something like $600 \mu\text{m}$, which means by rule of thumb $L_{bond} \approx 0,6 \text{ nH}$. But the bondwire is only part of the connection. From the bondpad some interconnect is needed to connect the bondpad to wherever the connection needs to go. For the LDMOS chip that is from the bondpad to the gate and for the CMOS driver that is from the buffer output to the bondpad.

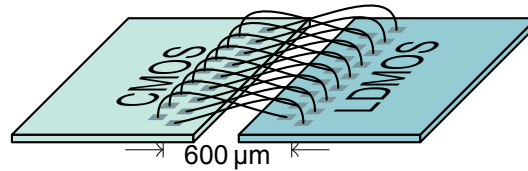
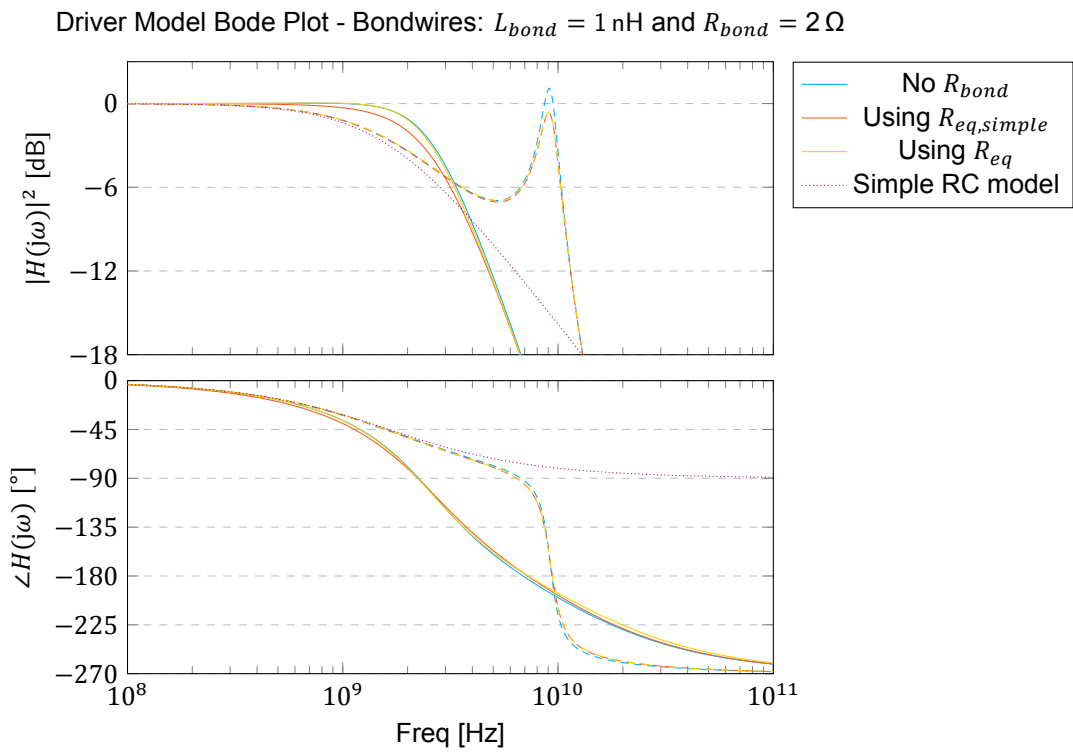


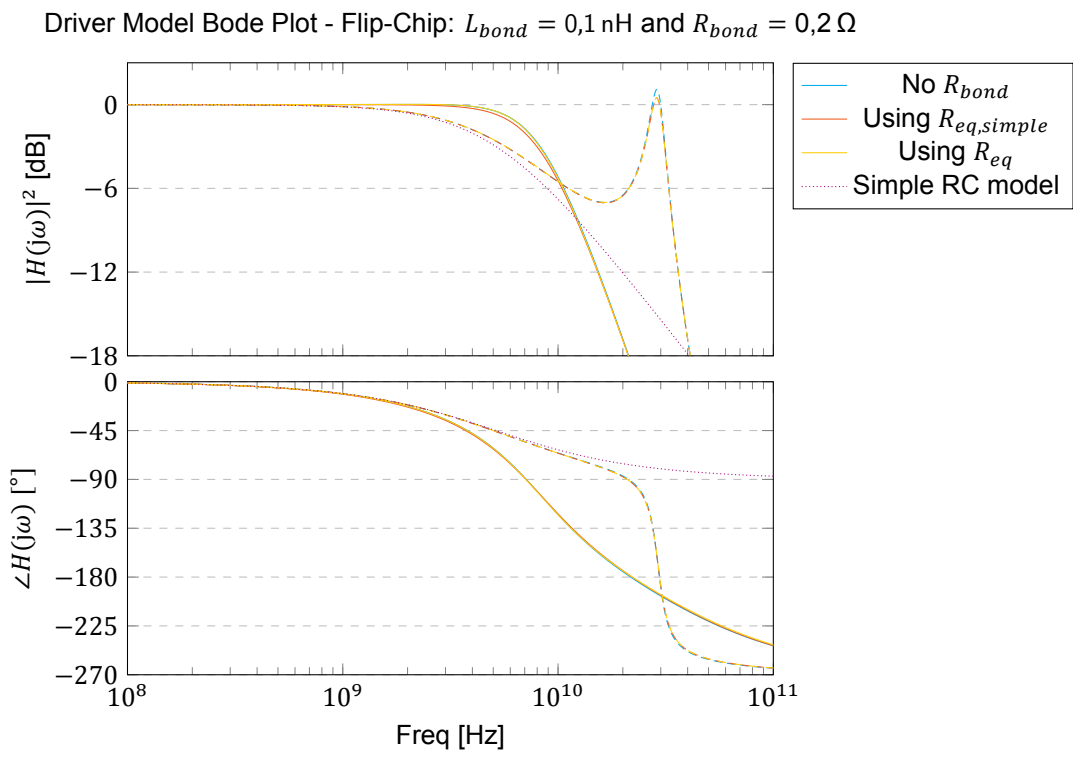
Figure 4.14: Inter-chip connection by bondwires

For now let's assume a worst case scenario of a total $L_{bond} = 1$ nH and something like $R_{bond} = 2 \Omega$. Further $C_{ESD1} = 600$ fF and $C_{ESD2} = 500$ fF. From later steps in the design process we can borrow the approximate values for the largest and smallest LDMOS transistor, having a TGW of 3,0 mm and 40 μ m respectively. This leads to a C_{GS} of 4,5 pF maximum and 0,06 pF minimum. Assuming the second order RLC-lowpass filter to be critically damped, $\zeta = 1$, the time constant can be calculated as $\tau_d = 157$ pF Ω for the largest transistor and $\tau_d = 97$ pF Ω for the smallest. The value for the smallest transistor is taken to minimize general delay. This allows to determine the entire transfer function of the driver model of figure 4.13.

The bode plot of the entire transfer function can be seen in figure 4.15a for both the large and small LDMOS transistor. The first order RC-lowpass transfer is also shown as reference. For the large transistor the driver compensated for R_{bond} is nearly indistinguishable from the situation that R_{bond} is not present at all. This is different for the small transistor, there the compensated driver is nearly indistinguishable from the uncompensated driver. This is to be expected, as the compensation is dependent on the ratio $\frac{C_{ESD2} + C_{GS}}{C_{ESD1} + C_{ESD2} + C_{GS}}$, which is larger for the big transistor than that for small one.



(a) Bode plot for bondwire based driver model with $\tau_d = 97 \text{ pF } \Omega$, solid: large transistor $C_{GS} = 4,5 \text{ pF}$; dashed: small transistor $C_{GS} = 0,06 \text{ pF}$



(b) Bode plot for flip-chip based driver model with $\tau_d = 31 \text{ pF } \Omega$, solid: large transistor $C_{GS} = 4,5 \text{ pF}$; dashed: small transistor $C_{GS} = 0,06 \text{ pF}$

Figure 4.15: Bode plots for bondwire and flip-chip based driver models

Flip-Chip

In figure 4.16 the principle of flip-chip is illustrated. The name flip-chip comes from the fact that

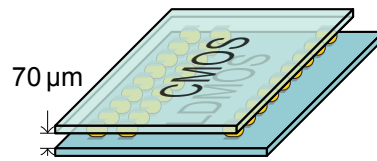


Figure 4.16: Inter-chip connection using flip-chip bonding, bonds (e.g. solder balls) in yellow

the top chip is flipped upside down, facing its bump pads down towards the bottom chip. The exact position of a connection does not matter much, it only needs aligned bump pads on either side of the flip-chip connection. For example a ball of solder can be used to form the connection between the chips, displayed in yellow. This allows for much more connections between the chips and smaller bonding distance. It also allows more effective placement of the bump pads, therefore requiring less interconnect on the chip itself. The distance between the chips will be 70 μm, which means by rule of thumb $L_{bond} \approx 0,07$ nH.

For now lets assume a worst case scenario of a total $L_{bond} = 0,1$ nH and something like $R_{bond} = 0,2 \Omega$. Just as in the bondwire case it is assumed that the bond is critically damped, so $\zeta = 1$. The time constant can then be calculated as $\tau_d = 50$ pF Ω for the largest transistor and $\tau_d = 31$ pF Ω for the smallest. Again the value for the smallest transistor is taken to determine the transfer function of the entire driver model.

The bode plot of the entire transfer function can be seen in figure 4.15b for both the large and small LDMOS transistor. The response is very similar to the bondwire case, but has a higher cutoff frequency.

4.3.3. Driver Evaluation

The bode plots from figure 4.15 clearly indicate that the flip-chip solution allows for a higher operation frequency of the RF-DAC than the bondwire solution. What the maximum operation frequency would be, can be best determined in the time domain by looking at the settling time of the step response. If the waveform is settled before applying the next step the waveform approximates a square wave.

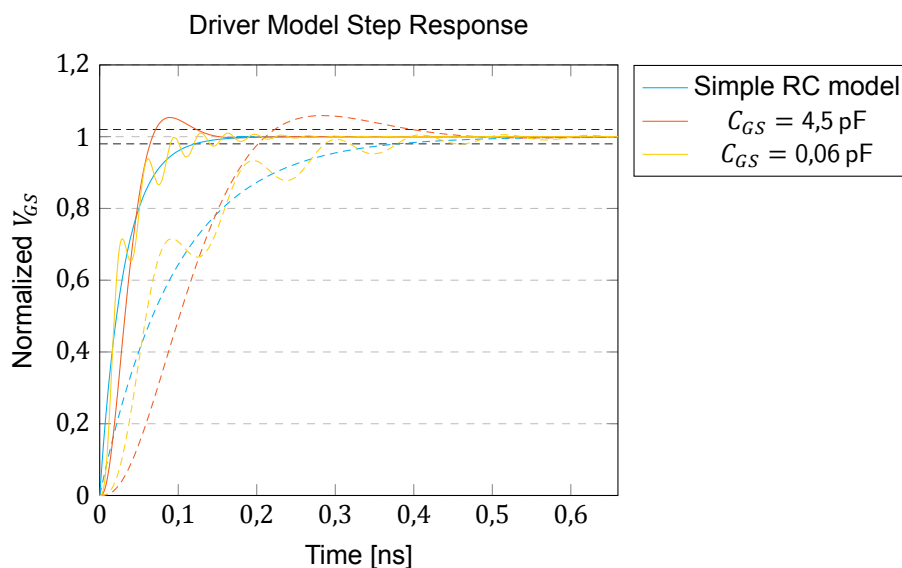


Figure 4.17: Step responses of the driver models with indicated settling time margin, solid: flip-chip; dashed: bondwire

The step response for both the driver options are plotted in figure 4.17, with the flip-chip solution plotted with a solid line and the bondwire solution in a dashed line. As expected the flip-chip solution is faster than the bondwire solution. In table 4.2 the calculated settling times are shown. If the maximum operating frequency is defined by having a period of twice the worst case settling time of

a solution $1/(2t_{settle})$, then the maximum operation frequency is 1,081 GHz for the bondwire solution and 3,334 GHz for the flip-chip solution.

Table 4.2: Settling time of the driver models

Driver Model	Settling Time
Bondwire: $C_{GS} = 4,5$ pF	0,398 ns
Bondwire: $C_{GS} = 0,06$ pF	0,463 ns
Flip-Chip: $C_{GS} = 4,5$ pF	0,123 ns
Flip-Chip: $C_{GS} = 0,06$ pF	0,150 ns

The flip-chip solution is the preferred option, as it allows for more than three times higher operation frequency than the bondwire solution with the assumptions made so far. From equation (4.12) it can be seen that with a three times higher operation frequency the drive power also increases by a factor three. It will also require larger CMOS drive transistors as the drive strength needs to be higher, which also means a higher parasitic output capacitance of the CMOS drive transistors and thus again higher driver power consumption. But the three times higher maximum operation frequency makes the system much more usable for current cellular communication standards, and therefore outweighs the disadvantage of requiring larger drive transistors. When the system operates at a lower frequency than the maximum the drive power will decrease proportionally, but possibly with a better drain efficiency due to faster switching than compared to using bondwires.

In conclusion, the flip-chip solution is chosen. To have some margin for e.g. PVT variations, a maximum operation frequency of $f_c = 3$ GHz is chosen. This directly eliminates the possibility to apply oversampling noise shaping techniques as mentioned in section 3.1.1, as the flip-chip connection will not reliably allow higher frequencies.

4.4. Circuit Topology

Now there is more information on the available technologies and implementation a choice has to be made on how to build the targeted power DRAC. An implementation has to be chosen for the digitally controllable R_{on} . Also whether the Class-E SMPA will be build in a single-ended or push-pull fashion. So far only single-ended has been considered, but a push-pull implementation may have significant benefits.

4.4.1. DRAC Structure

Three possible structures have been identified in section 3.1.2 for implementing the in section 3.3 described digitally controllable transistor R_{on} : binary-weighted, thermometer-coded or hybrid. In terms of linearity the thermometer-coded structure is suspected to perform best and the binary-weighted structure to perform worst. Thermometer-coded may therefore seem like the best choice. However, it has been established by now that the driver will need to be off-chip (see section 4.3), limiting the available connections and controllable transistors.

From section 4.2, a minimum resolution N_b of 8 bit is required, implying at least $2^8 - 1 = 255$ individual LDMOS transistors, CMOS to LDMOS connections and CMOS driver channels for a thermometer-coded structure. Besides the point that 255 CMOS to LDMOS connections may not even be possible in a reasonable chip area or resulting in significant difficulty in making a chip layout, this will require 255 pairs of ESD protection with $C_{ESD} = 1,1$ pF each. From equation (4.12) the required drive power for just the ESD protection in this case will be $3 \text{ GHz} \cdot 255 \cdot 1,1 \text{ pF} \cdot (2 \text{ V})^2 = 3,4 \text{ W}$. This is undesirable, which leads to the conclusion that thermometer-coded will not be an option.

The linearity of a binary-weighted structure will most likely be insufficient. More importantly, the linearity will not be accurately predictable with the available models. Therefore a binary-weighted structure is also not an option. That leaves the hybrid option, where a combination between thermometer-coded and binary-weighted is made.

As mentioned in section 3.3 it is expected that 8 thermometer-coded segments will result in good enough linearity. The segments will be scaled in such way that the resulting AM-AM curve will be linear, the actual scaling factors will be determined later. Since the to be designed DRAC cannot be

thermometer-coded, the transistors within the segment will have to be binary-weighted, in contrast to the implementation described in section 3.3. The 8 thermometer-coded segments will form the MSBs with $\log_2 8 = 3$ bit of resolution. That leaves 5 bit to be implemented in the binary-weighted LSBs. To clarify, with binary-weighted transistors is meant that the gate width of each transistor differs a factor 2 from the previous transistor within a segment, such that the LSB transistor is the smallest (has highest R_{on}).

At first glance this may seem like 8 segments with 5 binary-weighted transistors each will result in a total of 40 transistors, which is not true. This can be seen mathematically by defining the digital input code ACW using the notation from section 3.2 in terms of the 8 thermometer-coded MSB segment binary control signals $d_{ACW,1}d_{ACW,2} \cdots d_{ACW,8}$ and the 5 binary-weighted LSB binary control signals $ACW_4ACW_5 \cdots ACW_8$:

$$ACW = \underbrace{\frac{1}{2^3} \left(-1 + \sum_{i=1}^8 d_{ACW,i} \right)}_{MSB} + \underbrace{\left(\sum_{i=4}^8 \frac{ACW_i}{2^i} \right)}_{LSB} \quad (4.21)$$

This equation is only correct when assuming that the LSB sum can equal 2^{-3} , after which a next segment is switched on. This is not the case: the 5 LSBs allow for $11111_2 = 31_{10}$ output codes each. Evaluating equation (4.21) for the maximum input code, when the entire digital input code is 1, gives

$$ACW = 8 \cdot \underbrace{\frac{2^5 - 1}{2^8}}_{LSB} = \frac{8 \cdot 31}{256} = \frac{248}{256} \quad (4.22)$$

meaning that $255 - 248 = 7$ possible output codes are missing. The desired behavior of the input codes is shown in table 4.3 for the first 2 segments. It can be seen that there is a duplicate code when switching on a next segment, in this case for $k = 31$ and $k = 32$. This can be solved by adding an extra transistor with the same size as the transistor that would be turned on by the ACW_8 transistor of segment $d_{ACW,2}$, controlled by binary input $ACW_{8'}$. The same principle holds for all higher segments. It is not required for the first segment, hence resulting in an additional 7 transistors to complete the entire 8 bit range: a total of 47 transistors. With this addition equation (4.21) becomes

$$ACW = \frac{1}{2^3} \left(-1 + \sum_{i=1}^8 d_{ACW,i} \right) + \left(\frac{ACW_{8'}}{2^8} + \sum_{i=4}^8 \frac{ACW_i}{2^i} \right) \quad (4.23)$$

This means that the segment's normalized TGW will now be equivalent to $\frac{1}{2^8} + \frac{2^5-1}{2^8} = \frac{1}{8}$ of the entire TGW of the DRAC¹.

There are also 47 input signals. Notation-wise it is easiest to combine the thermometer-coded segment notation $d_{ACW,i}$ and the binary-weighted LSB notation ACW_n . This combination becomes

$$ACW_{i,n} = d_{ACW,i} \wedge ACW_n \quad \text{for } i = 1, 2, \dots, 8 \text{ and } n = 4, 5, \dots, 8 \quad (4.24)$$

where \wedge is the boolean AND operation and

$$d_{ACW,i} = \begin{cases} 0 & \text{if } k < 32(i-1) \\ 1 & \text{if } k \geq 32(i-1) \end{cases} \quad \text{for } i = 1, 2, \dots, 8 \quad (4.25)$$

where $k = 2^8 ACW$. The control signals for the 7 extra transistors to complete the entire 8 bit range are

$$ACW_{i,8'} = d_{ACW,i} \quad \text{for } i = 2, 3, \dots, 8 \quad (4.26)$$

¹The normalized TGW of the DRAC is defined as 1. Due to the "missing" transistor for $ACW_{1,8'}$ the actual sum of the gate widths will be $\frac{2^{N_b}-1}{2^{N_b}} TGW$. To have the definition of TGW as simple as possible this is neglected for now, but this 'issue' makes its return in equation (5.10)

Table 4.3: Values of the digital input codes ACW_i for segments $d_{ACW,1}$ and $d_{ACW,2}$ for $k = 0, 1, \dots, 63$

Input		Segment $d_{ACW,1}$					Segment $d_{ACW,2}$					
ACW	k	ACW_4	ACW_5	ACW_6	ACW_7	ACW_8	ACW_4	ACW_5	ACW_6	ACW_7	ACW_8	$ACW_{8'}$
0/256	0	0	0	0	0	0	0	0	0	0	0	0
1/256	1	0	0	0	0	1	0	0	0	0	0	0
2/256	2	0	0	0	1	0	0	0	0	0	0	0
3/256	3	0	0	0	1	1	0	0	0	0	0	0
4/256	4	0	0	1	0	0	0	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
27/256	27	1	1	0	1	1	0	0	0	0	0	0
28/256	28	1	1	1	0	0	0	0	0	0	0	0
29/256	29	1	1	1	0	1	0	0	0	0	0	0
30/256	30	1	1	1	1	0	0	0	0	0	0	0
31/256	31	1	1	1	1	1	0	0	0	0	0	0
32/256	32	1	1	1	1	1	0	0	0	0	0	1
33/256	33	1	1	1	1	1	0	0	0	0	1	1
34/256	34	1	1	1	1	1	0	0	0	1	0	1
35/256	35	1	1	1	1	1	0	0	0	1	1	1
36/256	36	1	1	1	1	1	0	0	1	0	0	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
59/256	59	1	1	1	1	1	1	1	0	1	1	1
60/256	60	1	1	1	1	1	1	1	1	0	0	1
61/256	61	1	1	1	1	1	1	1	1	0	1	1
62/256	62	1	1	1	1	1	1	1	1	1	0	1
63/256	63	1	1	1	1	1	1	1	1	1	1	1

4.4.2. Single-Ended or Push-Pull Topology

So far all amplifiers have been considered to be single-ended, meaning that they have a single output. Push-pull amplifiers have two outputs that are opposite in sign for the fundamental frequency. Using a balun these outputs can be converted to a single-ended output by subtracting the outputs from each other so the fundamental adds up constructively. The key benefit of having a push-pull topology is that the even harmonics have the same sign and will therefore not appear at the output. This can be seen with relative ease by approximating the system's output characteristics by a time domain Taylor series expansion¹

$$V_{out}(t) = \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t) + \alpha_4 x^4(t) + \dots \quad (4.27)$$

Assuming the positive side of the push-pull topology has a $V_{out+}(t)$ with $x_+(t) = A \cos(\omega_c t)$ and the negative side $V_{out-}(t)$ with $x_-(t) = -A \cos(\omega_c t)$ then

$$V_{out+}(t) = \alpha_1 A \cos(\omega_c t) + \frac{\alpha_2 A^2 [1 + \cos(2\omega_c t)]}{2} + \frac{\alpha_3 A^3 [3 \cos(\omega_c t) + \cos(3\omega_c t)]}{4} + \frac{\alpha_4 A^4 [3 + 4 \cos(2\omega_c t) + \cos(4\omega_c t)]}{8} + \dots \quad (4.28)$$

$$V_{out-}(t) = -\alpha_1 A \cos(\omega_c t) + \frac{\alpha_2 A^2 [1 + \cos(2\omega_c t)]}{2} - \frac{\alpha_3 A^3 [3 \cos(\omega_c t) + \cos(3\omega_c t)]}{4} + \frac{\alpha_4 A^4 [3 + 4 \cos(2\omega_c t) + \cos(4\omega_c t)]}{8} - \dots \quad (4.29)$$

The single ended output then becomes

$$V_{out}(t) = V_{out+}(t) - V_{out-}(t) \quad (4.30)$$

$$= 2\alpha_1 A \cos(\omega_c t) + \frac{\alpha_3 A^3 [3 \cos(\omega_c t) + \cos(3\omega_c t)]}{2} + \dots \quad (4.31)$$

¹only valid in memoryless systems, but a similar reasoning using the Volterra series could be used albeit more expansive

which only contain the fundamental and odd orders of harmonic distortion. Even in the case there is a mismatch between the positive and negative side of the push-pull topology $\alpha_{2+} - \alpha_{2-}$ the HD2 is reduced by a factor $\frac{1}{\alpha_{2+} - \alpha_{2-}}$ compared to the single-ended topology.

Implementing a Class-E SMPA DRAC in a push-pull manner is simply a matter of duplicating the circuit and transistor banks, and connecting the outputs to a balun, as illustrated in figure 4.18. If the intended output power and voltage needs to stay the same it is a matter of halving the transistor gate widths at each of the transistor banks (keeping the TGW the same) and scaling the other components proportionally so the design set from equation (2.74) stays the same.

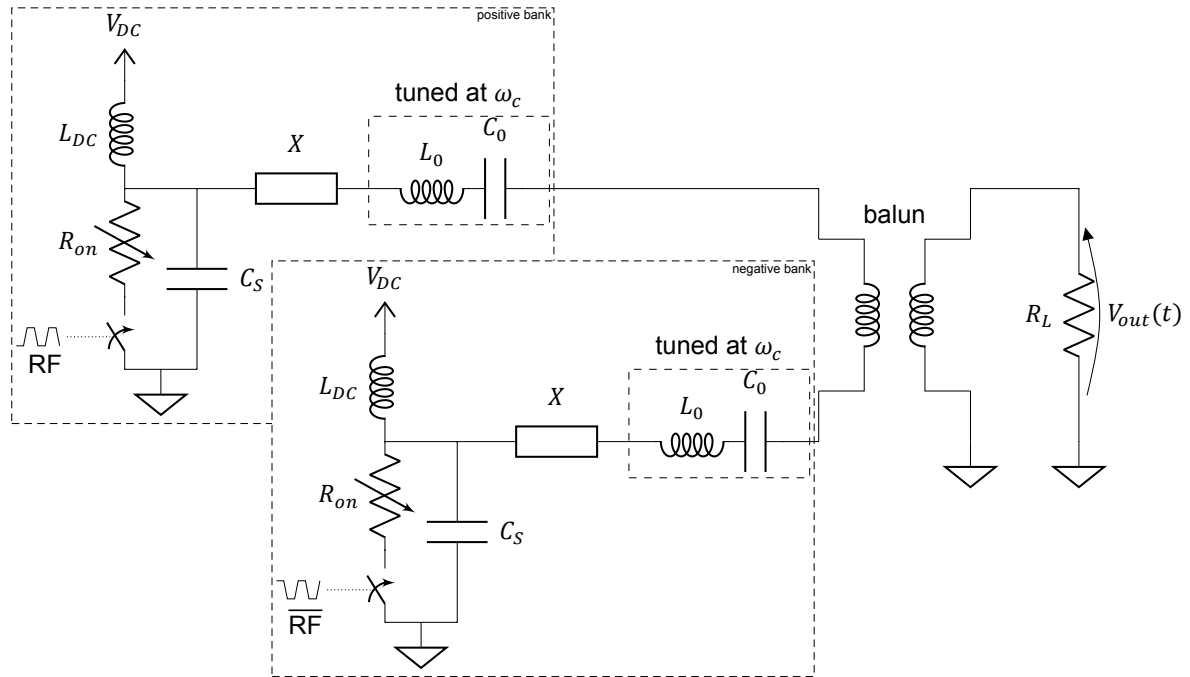


Figure 4.18: Push-pull Class-E SMPA DRAC

The price for making the DRAC push-pull is that it doubles the required amount of individual transistors, each with their own flip-chip connection. That would make for 94 total connections, which is no problem for a flip-chip solution. It does require the same number of ESD protection pairs, effectively doubling the drive power requirements due to the ESD protection capacitances compared to the single-ended topology.

The push-pull topology is preferred to cancel the even order harmonics. The drive signals allow the DRAC to operate in push-pull, but also identical signals can be applied to both transistor banks to let it operate single-endedly by shorting the two outputs and connecting it to the output matching network. By changing the matching network also other modes of operation can be allowed, such as a Doherty PA where one output is used as the main PA and the other as peak PA or for use at different carrier frequencies to allow for carrier aggregation.

4.5. Conclusions

From the available technologies, Fraunhofer IAF GaN25 and Ampleon LDMOS LM8, the choice is made to construct the DRAC in the LDMOS technology. Although GaN promises better Class-E performance, it requires a large input signal swing which cannot be attained by CMOS. Both technologies are focused on high power analog applications and lack the possibility to integrate high-speed digital logic to make it a full RF-DAC. This high-speed digital logic can be found in advanced CMOS technology nodes. Therefore an external CMOS driver is required, for which the TSMC 40 nm 1P7M technology is available, which has thick oxide transistors available that can handle 2,5V maximum. A V_T -shift can be applied to the LDMOS devices so they can be driven by CMOS voltage levels.

The off-chip CMOS driver will have some connection to the LDMOS DRAC, which can be modeled as an inductor with series resistance and a capacitance at either end of the connection due to ESD protection. The differently sized LDMOS transistors can be modeled by a capacitance at the LDMOS side of the connection. The driver operates in Class-D and is realized by an CMOS inverter, which can be modeled by an Equivalent Series Resistance (ESR). Together this system can be approximated by a first order RC-lowpass filter with time constant τ_d (delay) and a second order RLC-lowpass filter. Assuming the delay to all transistor sizes needs to be constant and critical damping for maximum speed a value for τ_d can be determined. From this value a required 'drive strength' (ESR R_{es}) for some given LDMOS transistor size can be calculated. The drivers for LDMOS transistors that need to remain off will not use any power, so input power scales with output voltage. A flip-chip connection between the CMOS and LDMOS is chosen due to the lower inductance and more possible connections. This flip-chip solution allows for a maximum operation frequency of 3 GHz, directly eliminating the possibility to apply oversampling noise shaping techniques.

From the ACPR requirement better than 50 dBc can be determined that a minimum resolution of $N_b = 8$ bit is required. Using a push-pull topology for the Class-E SMPA DRAC with 8 thermometer-coded MSB segments and 5 binary-weighted LSB transistors, results in 2×47 individual LDMOS transistors and flip-chip connections to the driver. The thermometer-coded segments will have different sizing to apply the piecewise AM-AM linearization technique. The push-pull topology allows for push-pull operation, canceling even orders of harmonic distortion. With other output matching networks and drive signals also other operation modes can be created, such as single-ended, Doherty and carrier aggregation.

5

LDMOS DRAC Implementation

The design steps and simulation results of the LDMOS DRAC design are described in this chapter. In section 5.1 the behavior of the technology in context of a Class-E SMPA DRAC is characterized. First in section 5.1.1 by finding a suitable design set and scaling it to meet required output power. Then in section 5.1.2 the resulting AM-AM curve is analyzed and the method to select appropriate transistor scaling for a piecewise linear output curve is explained.

In section 5.2 the total required gate width is divided across the thermometer-coded segments and within the segments into binary-weighted bits. The choice will be made to reduce the number of thermometer-coded segments to 6 and to increase the DRAC resolution to 9 bit. The effects of the driver model and possible glitches are investigated, resulting in a requirement for programmable delay-lines. With these delay-lines also AM-PM linearization can be applied.

Lastly in section 5.3 the relevant design rules for a layout are presented. Two versions for a layout are discussed together with their performance.

5.1. Technology Characterization

As discussed in chapter 4, the DRAC will be implemented using Ampleon's LDMOS LM8 technology with a maximum operation frequency of $f_c = 3$ GHz. It was determined that a peak fundamental output power of 20 W would be a realistic value with a use case in Multi-Input Multi-Output system (MIMO) systems.

5.1.1. Finding a Design Set

The first step towards a design is finding a working Class-E design set using the LDMOS device simulation model. For simplicity the single-ended Class-E topology from figure 2.18 is used with a 1 mm device, so $TGW = 1$ mm. The device size can be scaled later without loss of generality by keeping the design set the same. That way an appropriate TGW can be found in order to reach the targeted 20 W output power.

By taking $C_S = C_{DS} = 0,297$ pF the value of m is minimized to get the best possible drain efficiency. From equation (2.79) and table 4.1

$$m = \omega_c R_{on} C_{DS} = 6\pi \cdot 10^9 \cdot 10,5 \cdot 0,297 \cdot 10^{-12} = 58,8 \cdot 10^{-3} \quad (5.1)$$

Furthermore a duty-cycle of 50 % is assumed, so $d = 1$. The last parameter to be determined for the design set K is q and is taken as variable. The design set for this value of m is not analytically known. To ease the characterization parallel-circuit Class-E is assumed for now ($q = 1,412$ when $m = 0$) because it does not require additional reactance, so $K_X = 0$ for any value of m . Parallel-circuit Class-E is easily recognizable since it has maximum output power. The design set for parallel-circuit Class-E with $m = 0$ is taken as starting point for the component values. The supply voltage V_{DC} is assumed at 25 V. Filling

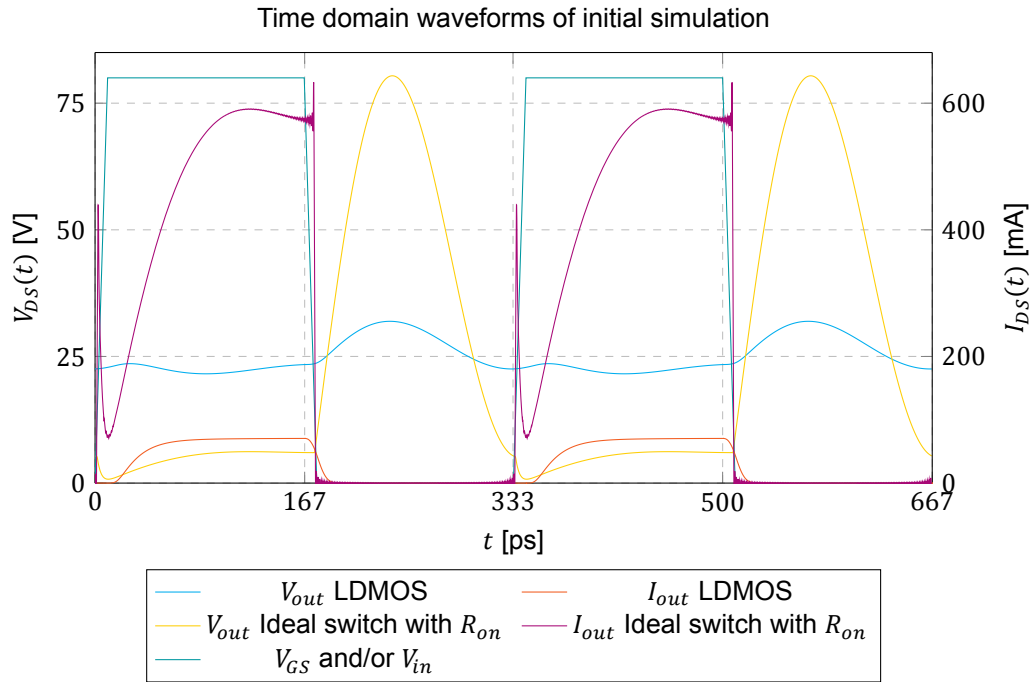


Figure 5.1: Initial simulated time domain waveforms for $TGW = 1$ mm using the design set for $q = 1,412$ when $m = 0$ at $f_c = 3$ GHz

in equation (2.74)

$$K : \left\{ \begin{array}{l} K_L = \frac{\omega_c L_{DC}}{R_L} = 0,732 \\ K_C = \omega_c C_S R_L = 0,685 \\ K_P = \frac{P_{out,fc} R_L}{V_{DC}^2} = 1,365 \\ K_X = \frac{X}{R_L} = 0 \end{array} \right\} \begin{array}{l} L_{DC} = 4,75 \text{ nH} \\ R_L = 122 \ \Omega \\ P_{out,fc} = 6,99 \text{ W} \\ X = 0 \end{array} \quad (5.2)$$

With increased m it is expected that the value for q where maximum output power occurs will slightly increase, K_C will increase and K_L and K_P will decrease [27]. The only remaining input variable to be able to calculate all component values is the quality factor Q_L and is for now assumed to be 5: $L_0 = 32,5$ nH and $C_0 = 86,7$ fF.

The LDMOS transistor is driven with an ideal square voltage source switching between 1,2V and 3,2V to mimic the behavior of a V_T -shifted LDMOS device driven by a driver with a peak-to-peak voltage of 2,0V. The simulation result of this starting point is shown in figures 5.1 and 5.2 in the form of time-domain waveforms and load-lines respectively. In the time domain waveform the driver voltage is scaled by a factor 40 so it is visible on the plot. It can be concluded that this design set does not result in correct Class-E behavior for the LDMOS device.

For reference an ideal switch with the same expected R_{on} and C_S is simulated as well, which does show a waveform and load-line that is expected from Class-E operation. Only the ZVS criterion is not met, as can be seen by the load-line not passing through the origin. From that can be concluded that the initial design set is not too far off, but somehow does not work for the LDMOS transistor. The explanation therefor is that the voltage swing is simply not great enough for the transistor to enter the triode region: it behaves as a current source rather than a resistor, as can be seen from the reference LDMOS DC curve in the load-line for $V_{GS} = 3,2$ V.

The load resistance has to be increased in order to have a voltage swing large enough for the transistor to behave more as a resistor. Therefore K_C needs to be increased and a corresponding value for q has to be found where maximum power occurs. Figure 5.3 shows that maximum power happens for $q \approx 1$ and $K_C \approx 4$, a surprising result since maximum power was expected at a higher

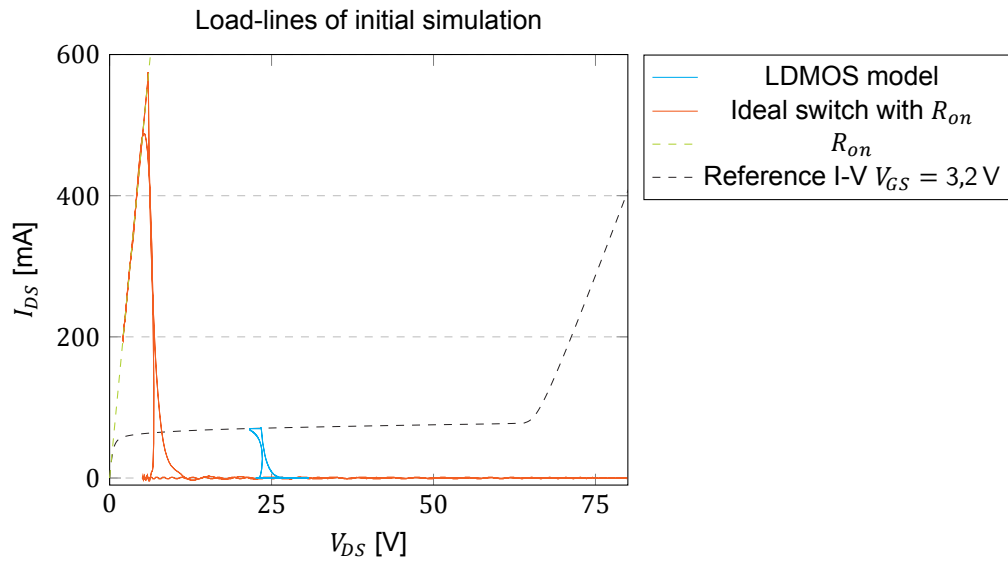


Figure 5.2: Initial simulated load-line for $TGW = 1$ mm using the design set for $q = 1,412$ when $m = 0$ at $f_c = 3$ GHz

Table 5.1: Design sets and values for $TGW = 1$ mm

Quantity	Value	Value
q	0,99	0,99
K_C	3,9	4,5
K_X	0	0
K_L	0,262	0,227
K_P	0,497	0,544
L_{DC}	9,67 nH	9,67 nH
R_L	697 Ω	804 Ω
P_{out,f_c}	0,446 W	0,423 W
η_D	66,4 %	67,5 %

value for q .

Further inspection shows that the point where maximum power occurs is for $q = 0,99$, $K_C = 3,9$ and $f_c = 3$ GHz: $P_{out,f_c} = 0,455$ W and $\eta_D = 66,5$ %. The time domain waveforms and load-line for this design set are shown in figures 5.4 and 5.5 respectively, the transistor now edges the triode region. Further increasing K_C allows the transistor to operate more in the triode region, up to $K_C = 4,5$ where maximum drain efficiency occurs: $P_{out,f_c} = 0,423$ W and $\eta_D = 67,5$ %. V_{DS} does not exceed 52 V, so the operation is within the breakdown limits of the LDMOS. The full design sets are given in table 5.1. The ZVS criterion is still not met, the load-lines resemble more those of saturated Class-B operation (figure 2.13) although a Class-E matching network is used and the transistor is driven by a square wave signal. This phenomenon could be explained by that C_S is larger than what would result in actual Class-E operation at a frequency of 3 GHz.

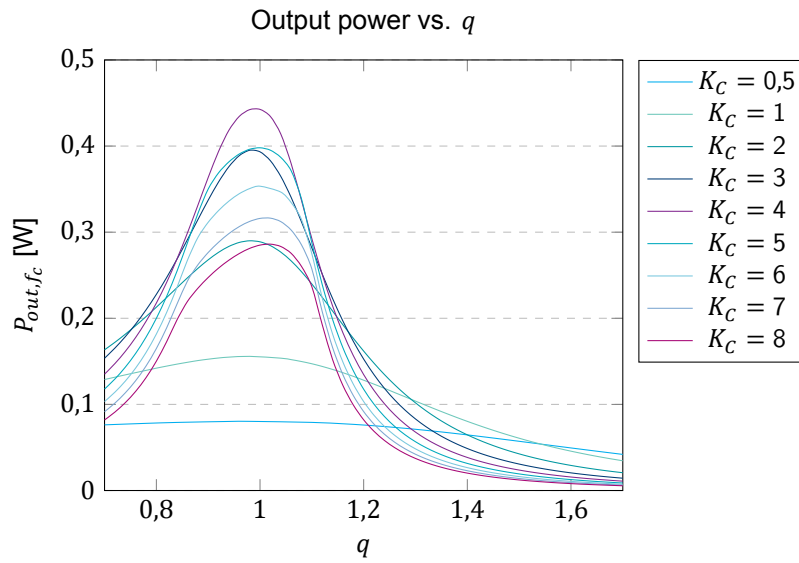


Figure 5.3: Output power as function of q and K_C to find a design set for $TGW = 1$ mm while $K_X = 0$

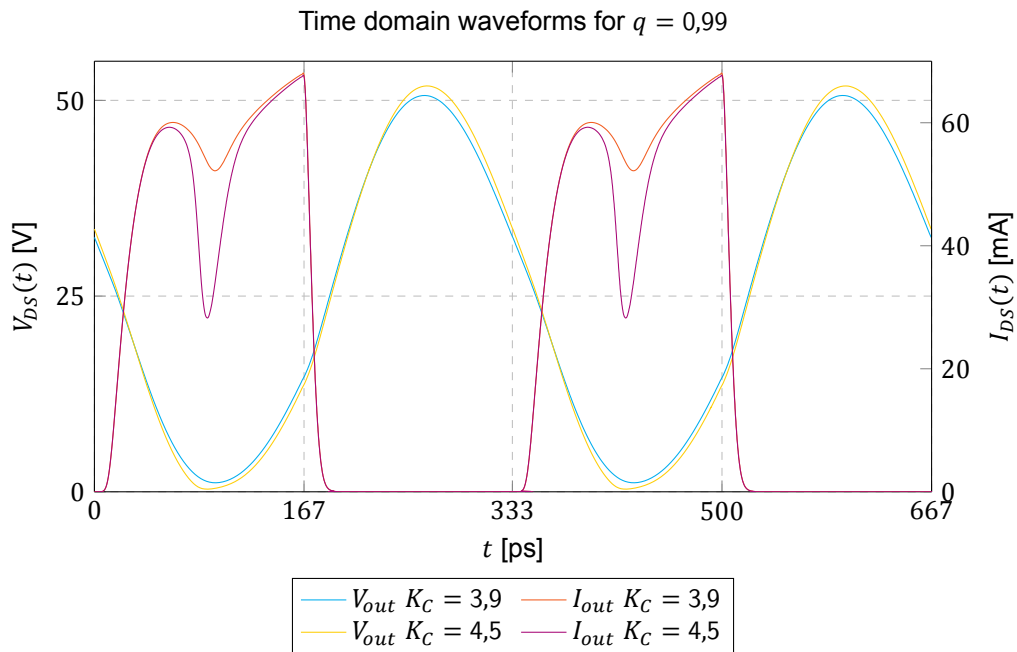


Figure 5.4: Time domain waveforms for $q = 0,99$ and $K_X = 0$ at $f_c = 3$ GHz, $K_C = 3,9$ for maximum power and $K_C = 4,5$ for maximum drain efficiency

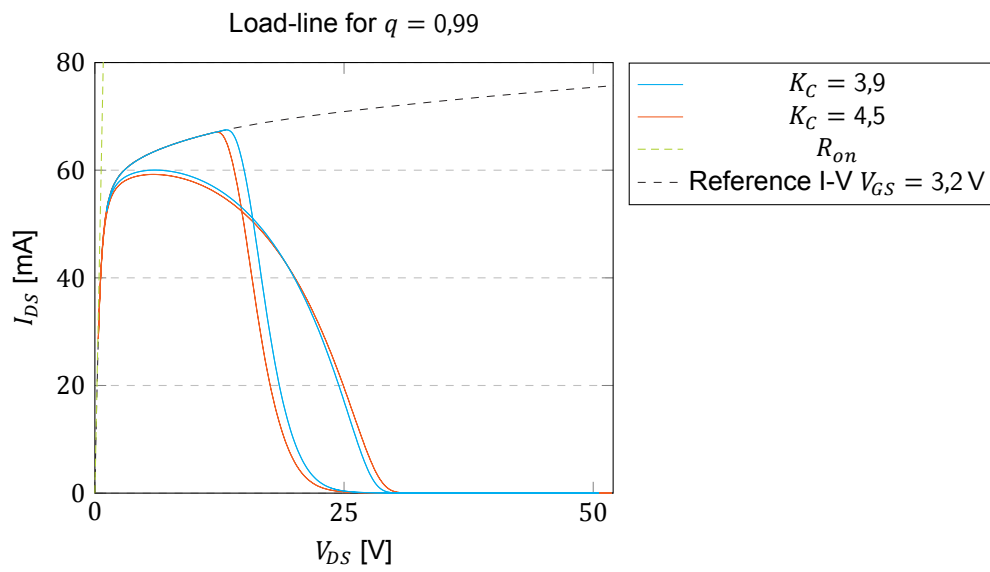


Figure 5.5: Load-lines for $q = 0,99$ and $K_X = 0$, $K_C = 3,9$ for maximum power and $K_C = 4,5$ for maximum drain efficiency

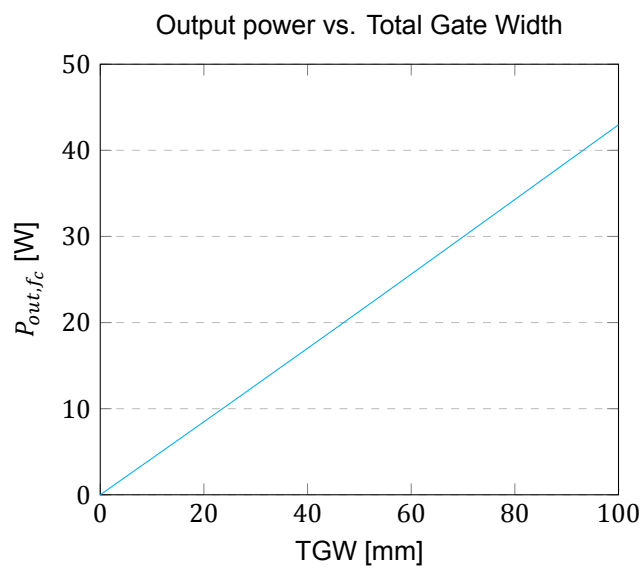


Figure 5.6: Output power as function of TGW for $q = 0,99$, $K_X = 0$ and $K_C = 4,5$

Table 5.2: Design set and values for $TGW = 50,4$ mm (push-pull) and $f_c = 3$ GHz

Quantity	Value
q	0,99
d	1,000
m	58,8 10^{-3}
K_L	0,227
K_C	4,500
K_P	0,546
K_X	0,000
Q_L	5,000
C_{DS}	7,48 pF
R_{on}	0,417 Ω
L_{DC}	0,384 nH
$2 \cdot R_L$	63,79 Ω
C_0	333 fF
L_0	8,461 nH
$2 \cdot P_{out,fc}$	21,38 W
η_D	67,5 %

Scaling the design set

Next is finding the required TGW in order to reach 20 W output power. By keeping the design set equal (scaling all components proportionally) the output power should scale linearly with TGW, which is verified by figure 5.6. Using the design set for maximum efficiency this gives

$$TGW = \frac{20 \text{ W}}{0,423 \text{ W}} \cdot 1 \text{ mm} = 47,3 \text{ mm} \quad (5.3)$$

To leave some room for future layout losses it should be slightly over-dimensioned, therefore $TGW \approx 50$ mm is assumed.

The efficiency drops slightly with increasing TGW due to how the simulation is configured: the amount of gate fingers is kept the same in the simulation. When the gate finger width is increased the drain resistance also increases slightly resulting in some losses. This can be easily resolved by e.g. doubling the amount of fingers and halving the transistor width. This effect should also be taken in to account when making the design and layout by preventing very wide gate fingers.

Making it push-pull with drive signal rise and fall time

In order to make the design push-pull (as shown in figure 4.18) the two transistor banks will each take care of half of the TGW. The design set again is kept the same for each of the banks, as well as the resistive load seen by each bank. Depending on the used balun the output load can be chosen at 50 Ω , but for now an ideal 1:1 transformer is assumed. To make each bank see the correct load R_L should be twice that of what each individual bank should see, which is equal to what the load would have been if the circuit was single-ended with the same TGW. The resulting component values are shown in table 5.2 with the design set again for completeness. These component values seem realizable in a 'physical world' design.

So far the drive signal was assumed to be an ideal square wave, but in reality it will have some rise and fall time as discussed in section 4.3.1. The rise and fall time will be assumed equal and will be proportional to τ_d , implying a stronger driver will have lower rise and fall times and vice versa. For now a linear edge is used. The impact of rise and fall time is shown in figure 5.7. It can be seen that increasing rise and fall time lowers the output power, but an optimum for the drain efficiency can be found at $t_{rise} = t_{fall} = 0,078$ ns where $\eta_D = 71,3$ %, although $P_{out,fc}$ drops to 19,65 W. In a more realistic implementation the edge will not be linear, but this simple test scenario shows that the driver's waveform will have impact on the output power and efficiency of the power DRAC.

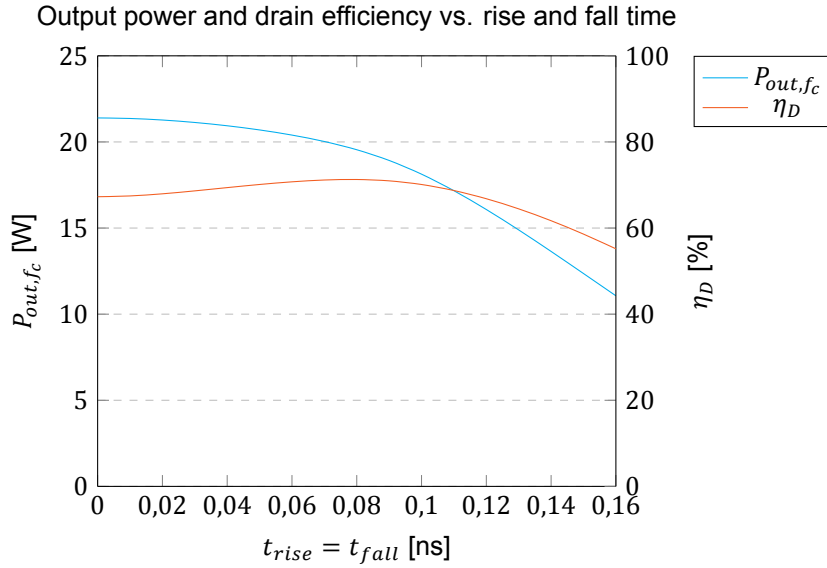


Figure 5.7: LDMOS output power and drain efficiency as function of rise and fall time for push-pull $TGW = 50,4$ mm, $q = 0,99$, $K_X = 0$ and $K_C = 4,5$

5.1.2. The AM-AM Curve

In order to apply the piecewise AM-AM correction as described in section 3.3 to the segments first the 'baseline' AM-AM curve should be known. This baseline AM-AM curve is obtained by implementing a DRAC simulation with thermometer coded equal sized transistors. For sake of simplicity $N_b = 6$ bit is assumed, therefore the $TGW = 50,4$ mm will be divided over $2^6 - 1 = 63$ transistors per push-pull bank, resulting in $400 \mu\text{m}$ per transistor. The resulting baseline AM-AM curve is shown in figure 5.8.

For reference the ideal linear output is shown dashed. This ideal linear output can be recognized as the endpoint INL reference line (see section 3.1.3) and will be used as such for the remainder of this thesis. The baseline AM-AM curve is fairly close to the ideal curve, indicating the DRAC is intrinsically quite linear. For the lower values of k the transistor behaves mostly as current source. For values of k from 50 to 55 the transistors start to enter the triode region and thus behave more as a resistor. From section 3.2 it is known that switched current sources operate more linearly than switched resistors. The high (peak) drain efficiency is reached when the transistor operates as resistor. The part of this curve where the transistor operates more as current source can be designated as the *current limited* region of the DRAC and where the transistor operates as resistor the *voltage limited* region. The transition between these regions is not as obvious in this case due to the LDMOS only being driven by a square wave with 2 V amplitude, therefore spending a relatively large portion in the current limited region. This results in potentially a lower drain efficiency but a better linearity. For a case study using a higher LDMOS drive amplitude where the transition is more obvious see appendix B. The transition was also more clear in the GaN design, see appendix C.

Linearizing the AM-AM Curve

It is expected that dividing the DRAC into 8 non-linearly sized segments will provide good enough linearity, as also stated in section 4.4.1. In figure 5.8 the resulting INL is shown if the DRAC was sized linearly. By constructing a piecewise linear output voltage the AM-AM curve can be characterized. By comparing the slope of the linear pieces to the ideal slope the segment sizing can be determined. This sample case is done for the simple 6 bit DRAC, but the segment sizing can also be applied on the to be designed DRAC.

First the values of $V_{out,fc}$ are taken when all transistors in a segment are turned on, just before enabling the next segment. This happens for the i^{th} segment for

$$k[i] = \max\left(0, \frac{2^{N_b}}{N_{seg}} \cdot i - 1\right) \quad \text{for } i = 0, 1, 2, \dots, N_{seg} \quad (5.4)$$

where N_{seg} is the number of segments and, as introduced in section 3.2, $k = 2^{N_b} ACW$. The voltage

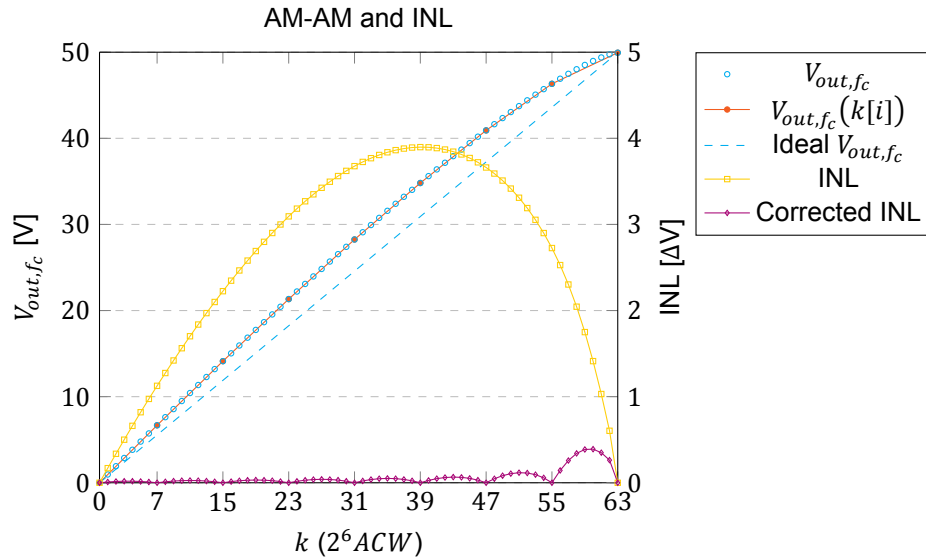


Figure 5.8: LDMOS baseline AM-AM curve and INL for $N_b = 6$, push-pull $TGW = 50,4$ mm, $q = 0,99$, $K_X = 0$, $K_C = 4,5$ and $t_{rise} = t_{fall} = 0,078$ ns

slope of that segment can then be given by

$$\Delta V_{out,fc}[i] = \frac{V_{out,fc}(k[i]) - V_{out,fc}(k[i-1])}{k[i] - k[i-1]} \quad \text{for } i = 1, 2, \dots, N_{seg} \quad (5.5)$$

In figure 5.8 the values for $V_{out,fc}(k[i])$ are also plotted. The wanted or average voltage slope is

$$\overline{\Delta V_{out,fc}} = \frac{V_{out,fc}(k[N_{seg}])}{k[N_{seg}]} \quad (5.6)$$

The relative segment size, or segment scaling, for the i^{th} segment S_i can then be determined by the ratio between the wanted slope and the actual slope of that segment. This will balance more gate width to the last segments. The TGW needs to stay the same, but the average of the calculated scaling factors is generally not equal to 1. Therefore additional normalization of S_i is required and can be given by

$$S_i = \frac{\overline{\Delta V_{out,fc}}}{\Delta V_{out,fc}[i]} \cdot \frac{N_{seg}}{\sum_{n=1}^{N_b} \Delta V_{out,fc} / \Delta V_{out,fc}[n]} \quad \text{for } i = 1, 2, \dots, N_{seg} \quad (5.7)$$

slope correction
normalization

This method works by first order approximation, so it does not result in the wanted linearized AM-AM curve right away. It slightly overcompensates the AM-AM curve. By iterating this process the AM-AM curve converges to the wanted version, usually within 4 to 5 iterations. The calculation of the first iteration for the example is shown in table 5.3. In this example $\overline{\Delta V_{out,fc}} = 0,7927$ V. The implementation of these scaling factors will be discussed in the next main section, section 5.2. After implementing this linearization method the corrected INL curve as shown in figure 5.8 can ideally be obtained. The corrected $INL_{pk} = 0,390$ V while uncorrected $INL_{pk} = 3,898$ V. Note that this is a calculated curve to demonstrate the concept of the linearization.

5.2. Implementing the DRAC

The found Total Gate Width (TGW) needs to be divided over the differently sized transistors. As discussed in section 4.4.1, the DRAC structure consists of 8 thermometer-coded segments with each 5 binary-weighted transistors to reach the resolution of 8 bit. First the sizing of the different transistors will be determined. Then this sizing is applied to a simulation setup with the simple driver model.

Table 5.3: Determining segment scaling, first iteration

i	$k[i]$	$V_{out,fc}(k[i])$	$\Delta V_{out,fc}[i]$	$\frac{\overline{\Delta V_{out,fc}}}{\Delta V_{out,fc}[i]}$	S_i
0	0	0,00			
1	7	6,67	0,9536	0,8314	0,7908
2	15	14,12	0,9301	0,8523	0,8107
3	23	21,33	0,9013	0,8795	0,8366
4	31	28,25	0,8656	0,9158	0,8711
5	39	34,81	0,8204	0,9663	0,9191
6	47	40,92	0,7632	1,0387	0,9880
7	55	46,33	0,6758	1,1730	1,1158
8	63	49,94	0,4520	1,7537	1,6680
$\Sigma =$				8,4107	8

5.2.1. Increasing the Resolution and Individual Transistor Sizing

From table 5.3 can be seen that the scaling for the first segments do not differ very significantly from each other. Therefore the choice was made to merge segment $d_{ACW,2}$ with $d_{ACW,1}$ and $d_{ACW,4}$ with $d_{ACW,3}$. This can be done with relative ease by incorporating the total segment gate width of segment $d_{ACW,2}$ into a single transistor for segment $d_{ACW,1}$. This transistor and its control signal will be referred to as $ACW_{1,3}$ conform the notation used in section 4.4.1. Similarly, the functionality of segment $d_{ACW,4}$ can be captured into transistor $ACW_{3,3}$. With this action for both transistor banks a total of $2 \cdot 2 \cdot 6$ transistors are replaced with only 4 transistors, therefore saving 20 transistors and flip-chip connections, with negligible loss in performance. This is illustrated by the second and fourth rows of transistors in figure 5.9.

As can be seen from the calculated corrected INL in figure 5.8 it will be fundamentally impossible with this implementation to get a perfectly straight AM-AM curve. To reach the linearity requirements it was calculated in section 4.2 that a resolution of 7,9 bit was required. This is the value for the required ENOB (see also section 3.1.3). With an implementation resolution of $N_b = 8$ bit and this INL behavior the linearity requirements will most definitely not be met. Therefore the choice was made to increase the implementation resolution to $N_b = 9$ bit, which is done by adding transistors and control signals

$$ACW_{i,9} \quad \text{for } i = 1, 3, 5, 6, 7 \text{ and } 8 \quad (5.8)$$

The transistors that were added to complete the entire range will therefore also be renamed with from a $ACW_{i,8'}$ to $ACW_{i,9'}$, implying they will be half the width as they were before. The total amount of transistors now is 2×43 , as can be seen in figure 5.9.

The gate width of every transistor can be expressed in the form $W_{i,n}$ denoting the gate width of the transistor responsible for the n^{th} bit in the i^{th} segment. This results in the following expression

$$W_{i,n} = \frac{TGW}{2} \frac{S_i}{2^n} \quad (5.9)$$

where the 2 in the first denominator originates from the push-pull topology. Due to the lack of transistor $ACW_{1,9'}$, the actual sum of the gate widths does not entirely equal TGW

$$2 \left(\sum_{i=1,3} \sum_{n=3}^9 W_{i,n} + \sum_{i=5}^8 \sum_{n=4}^9 W_{i,n} + \sum_{i=3,5,6,7,8} W_{i,9'} \right) = \frac{2^9 - 1}{2^9} TGW \approx 0,998 \cdot TGW \quad (5.10)$$

This error is so small that its effect is negligible, therefore there is no need to redefine the transistor gate width or the TGW. The current definition of TGW allows equation (5.9) to be as straightforward as it is right now.

5.2.2. Simulation Setup with Driver Model

A simulation setup with individual transistor sizes given by equation (5.9) and table 5.3 is created with a driver model as shown in figure 4.13. For starters no (flip-chip) bond inductance or resistance is

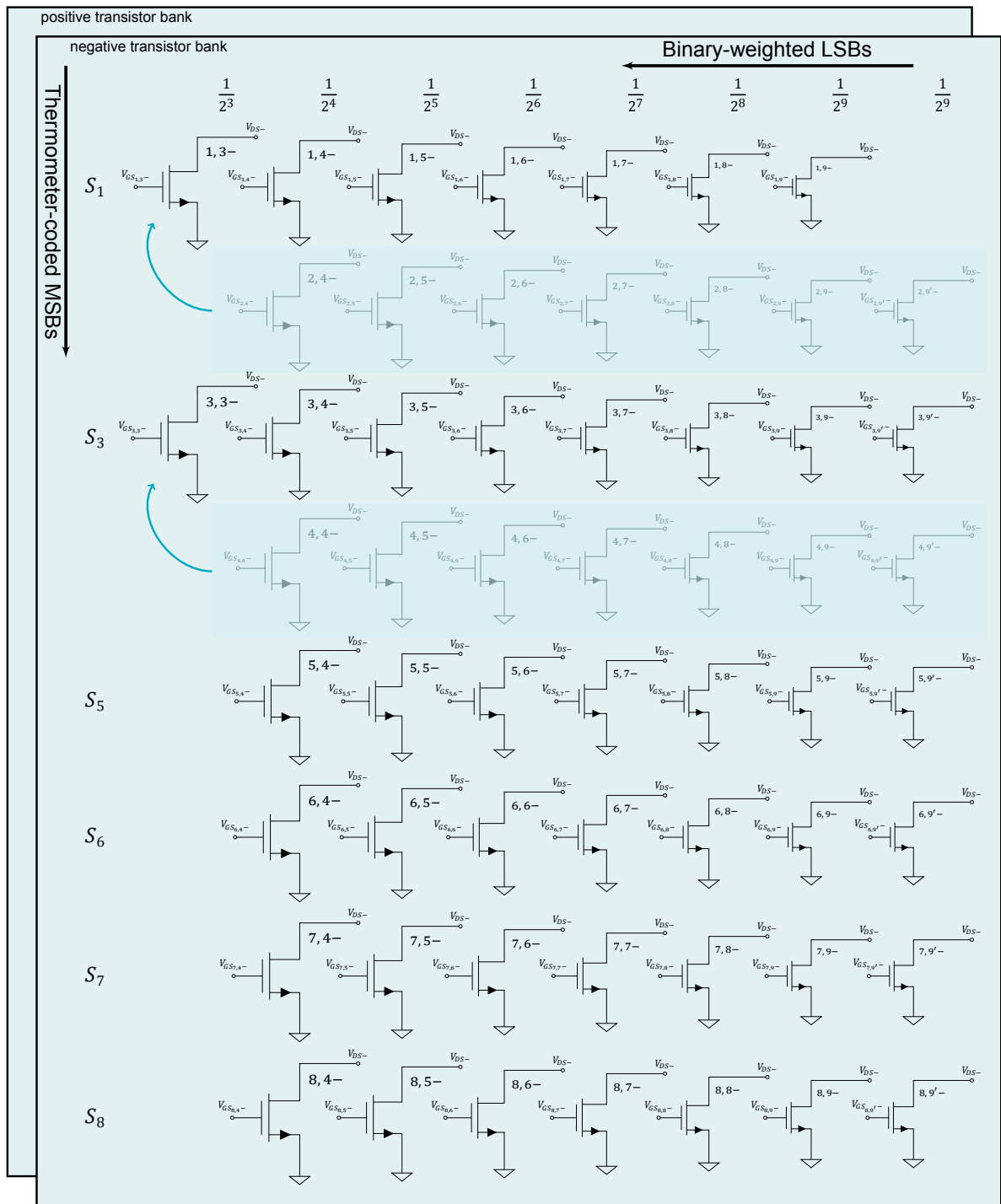


Figure 5.9: Illustrating the individual transistor sizes

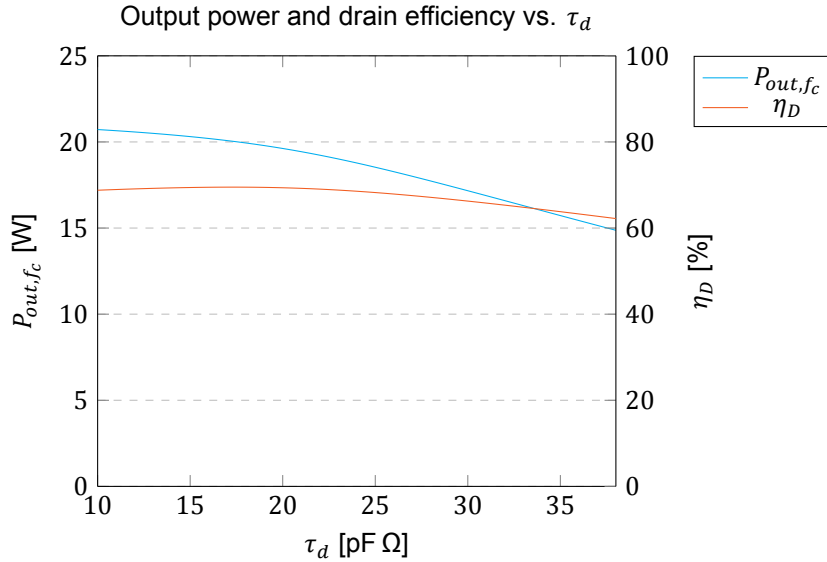


Figure 5.10: LDMOS output power and drain efficiency as function of τ_d for the design set from table 5.2 but with $TGW = 50,6$ mm

Table 5.4: Scaling factors

S_1	S_3	S_5	S_6	S_7	S_8	TGW
0,7955	0,8388	0,8891	0,9443	1,0377	1,4742	50,586 mm

assumed, only a certain drive strength R_{es} which is given by equation (4.15). Every transistor will require its own driver with corresponding drive strength, largely dependent of τ_d and individual $C_{GS_{i,n}}$ where

$$C_{GS_{i,n}} = W_{i,n} \cdot 1,50 \text{ pF mm}^{-1} \quad (5.11)$$

With this information the required value for τ_d can be determined for best efficiency by repeating the simulation shown in figure 5.7, but now including individually sized transistors with their own simple driver model. The design set given in table 5.2 is used, only with a slightly increased TGW to 50,6 mm. The result is shown in figure 5.10. It can be seen that optimum τ_d for best drain efficiency is achieved at 17 pF Ω , while the output power remains enough. Increasing τ_d lowers both drain efficiency and output power, which is undesired.

Applying the linearization algorithm from section 5.1.2 results in segment scaling factors given in table 5.4. S_2 and S_4 are not required anymore due to the merging of the first segments. Comparing these values to those in table 5.3 it can be confirmed that the algorithm slightly overcompensated the AM-AM non-linearity in the first iteration, but was fairly close. In figure 5.11 the resulting TGW that is activated versus ACW is shown, illustrating the inverted AM-AM curve by segment sizing.

The smallest transistors can be found in segment $d_{ACW,1}$, specifically the transistors for its LSB: $W_{1,9} = 0,0393$ mm. There are two of these due to the push-pull topology. Since the first segments are merged, their MSBs are twice the size than for the others. Because $2S_3 > S_8$ the largest transistors can be found in segment $d_{ACW,3}$: $W_{3,3} = 2,652$ mm. Resulting gate capacitance and drive strengths are shown in table 5.5. Note that ESD capacitance (1,1 pF) is also taken into account when calculating required drive strength. In figure 5.12 the time domain V_{GS} waveforms are shown for these transistors. It can be seen that by choosing equal τ_d for these transistors indeed results in the same delay from the driver to the LDMOS gate. Note that the gate voltage is shifted 1,2 V to mimic the V_T -shift.

Using the segment scaling factors from table 5.4 the 9 bit DRAC has an AM-AM linearity and AM-AM curve as shown in figures 5.13a and 5.13b respectively. Here $LSB = 99,20$ mV. Peak values are $INL_{pk} = 608,7$ mV ($INL_{pk,LSB} = 6,136$) and $DNL_{pk} = 45,3$ mV ($DNL_{pk,LSB} = 0,456$). From that can be concluded that the DRAC is monotonous. Comparing the corrected INL from figure 5.8 and the simulated INL from figure 5.13a it can be seen that the first one is perfectly smooth while the second has some irregularities. This is a side-effect of the binary-weighted nature due to mismatches in e.g.

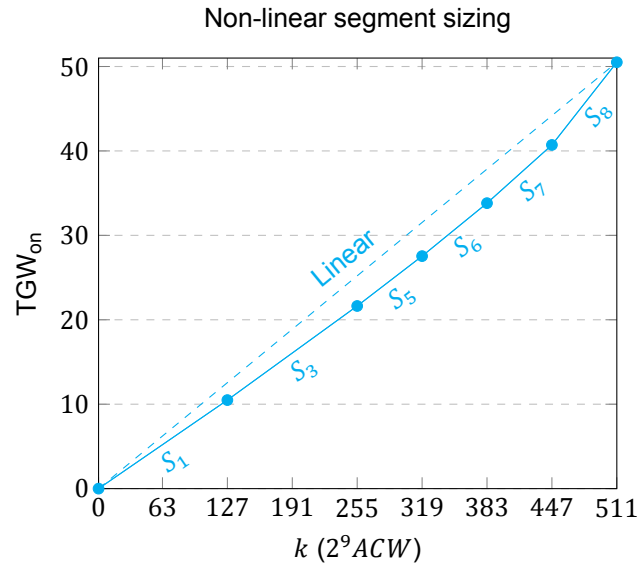
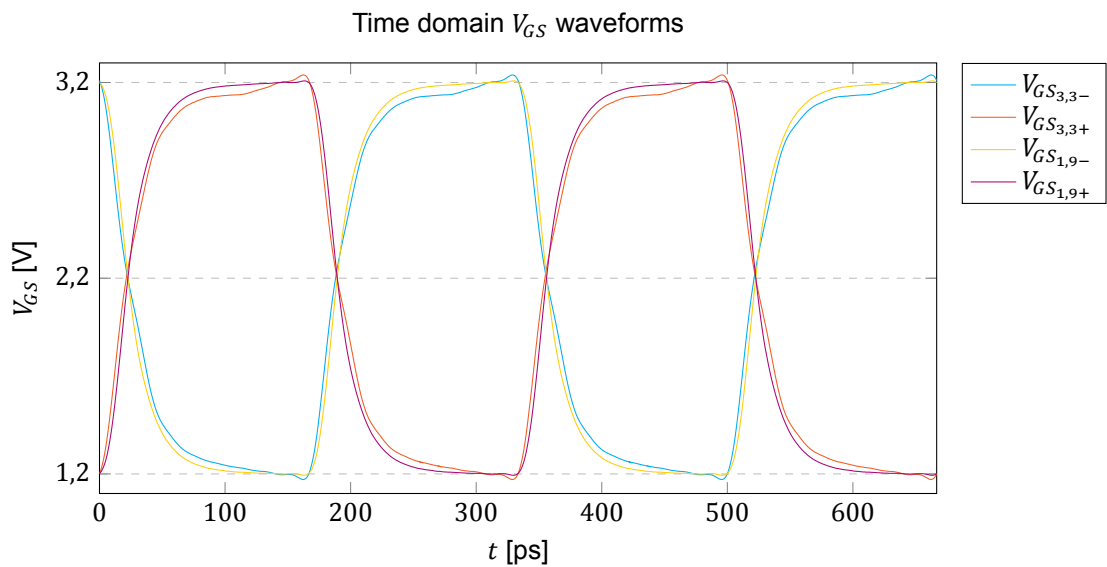
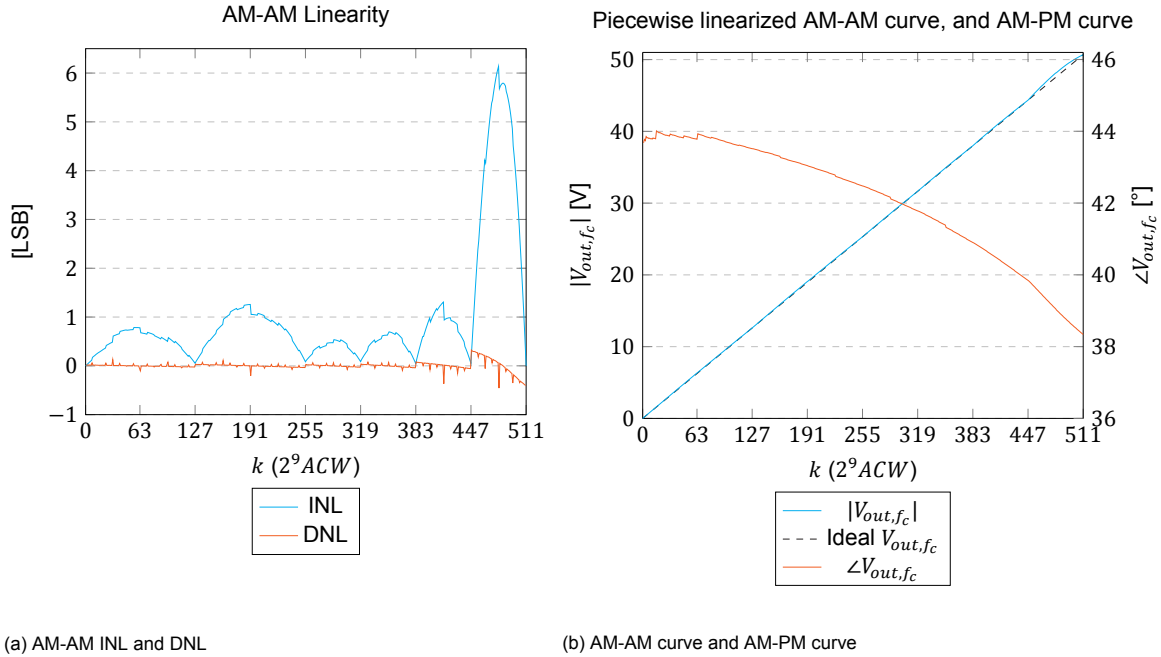


Figure 5.11: Activated TGW versus ACW illustrating non-linear segment sizing

Table 5.5: Sizes of smallest and largest transistors and their drive strength

$\tau_d = 17 \text{ pF } \Omega$	$W_{i,n}$	$C_{GS_{i,n}}$	$R_{es_{i,n}}$
$i, n = 1, 9$	0,0393 mm	0,0590 pF	14,66 Ω
$i, n = 3, 3$	2,652 mm	3,979 pF	3,075 Ω

Figure 5.12: Time domain V_{GS} waveforms for the largest and smallest transistors when $\tau_d = 17 \text{ pF } \Omega$



(a) AM-AM INL and DNL

(b) AM-AM curve and AM-PM curve

Figure 5.13: AM-AM linearity of implemented DRAC with $N_b = 9$ bit, using a simple driver model without series inductance

transistor parasitics. The deviations are also visible in the DNL plot in the form of little spikes. The fabricated version will most likely exhibit similar behavior but more pronounced due to actual fabrication mismatches.

In figure 5.14 the power and efficiency of the DRAC are shown. At full power and $f_c = 3$ GHz it requires $P_{DC} = 29,0$ W for $P_{out,fc} = 20,2$ W resulting in $\eta_D = 69,5\%$. The DC power consumed is linearly proportional to the output voltage, so the drain efficiency as well, similar to Class-B operation. The simulation based $P_{in} = 2,13$ W at full power. P_{in} is proportional to ACW, as predicted in section 4.3. The P_{in} curve is irregular, mainly due to the contribution of the ESD protection capacitances. Theoretical P_{in} can be calculated using equation (4.12). The total capacitance that needs to be driven at full power is

$$\begin{aligned}
 C_{tot} &= C_{GS} + 2 \cdot 43 \cdot (C_{ESD1} + C_{ESD2}) \\
 &= 1,50 \text{ pF mm}^{-1} \cdot \frac{2^9 - 1}{2^9} \cdot 50,586 \text{ mm} + 86 \cdot (0,6 \text{ pF} + 0,5 \text{ pF}) \\
 &= 75,73 \text{ pF} + 94,6 \text{ pF} = 170,3 \text{ pF}
 \end{aligned} \tag{5.12}$$

and results in $P_{in} = 2,04$ W at $f_c = 3$ GHz. It is slightly off from the simulation based P_{in} , although it can be that the method of calculating P_{in} from the simulation is not entirely accurate. Since the driver is modeled as a square wave voltage source the current will flow both from and into the voltage source, so an average of 0 A. The calculation of P_{in} is therefore done by time domain integration of instantaneous power, which can be sensitive to ringing artifacts resulting in higher calculated power consumption. With the relatively clean voltage waveforms as shown in figure 5.12 this simulation based P_{in} is likely the most accurate estimate of input power that can be obtained by simulation using this method. With increasingly complex waveforms further on in the thesis the simulation based P_{in} starts to deviate significantly and is henceforth omitted. Nonetheless it can be concluded from equation (5.12) that the largest portion of P_{in} is spent on driving the ESD capacitances. The calculated P_{in} reduces to 0,909 W when only the LDMOS gate capacitance is used.

Flip-Chip Bond Characterization

In section 4.3.2 driver models were assumed based on rules of thumb. For actual design work this will not suffice, so a better estimate is required. A 2.5D EM simulation of a flip-chip bond is done to get a founded estimate of the bond inductance. Due to signed Non-Disclosure Agreements (NDAs) the EM

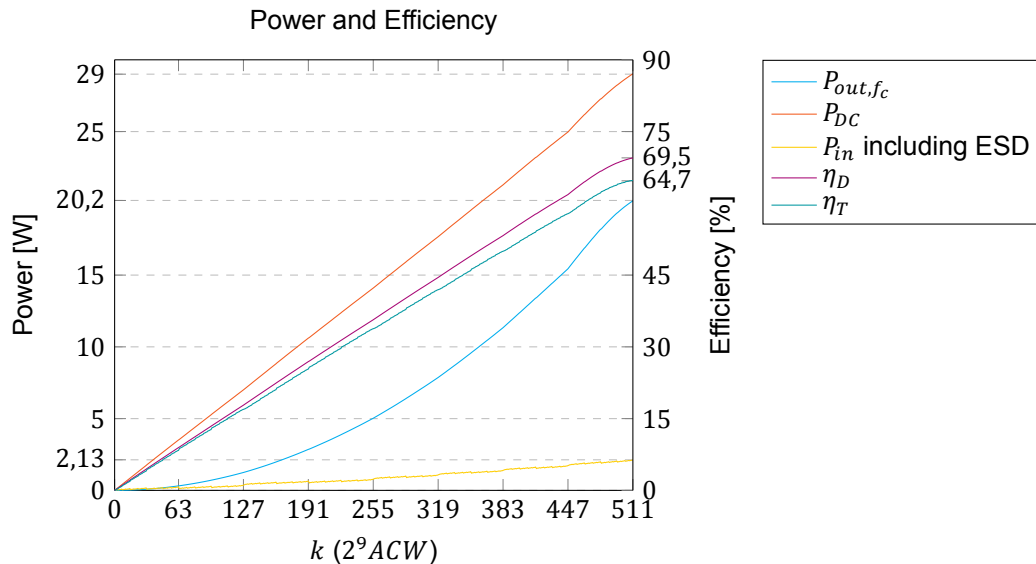


Figure 5.14: Power and efficiency of implemented DRAC with $N_b = 9$ bit at $f_c = 3$ GHz

substrate info can not be provided, but includes TSMC's CMOS substrate and all metal layers, and Ampleon's LDMOS substrate and top metal layer. A 3D preview of the flip-chip bond layout can be seen in figure 5.15. The actual flip-chip bond is modeled as a 70 μm long copper rod.

The flip-chip bond is simulated using a Momentum Method of Moments (MoM) microwave 2.5D EM¹ simulation. This results in a S-parameter model for a range of frequencies, from that an equivalent inductance and series resistance value can be determined. This can be done using Y-parameters similarly to equations (4.1) and (4.2), but then this is a series connection. Impedance parameters (Z-parameters) cannot be applied here since the opposing terminal should be shorted in order to determine the component values. This way any parallel capacitance is shorted as well and gives

$$Y_{12}(\omega) = Y_{21}(\omega) = \frac{1}{R_{bond} + j\omega L_{bond}} = \frac{R_{bond} - j\omega L_{bond}}{R_{bond}^2 + \omega^2 L_{bond}^2} \quad (5.13)$$

which complicates things a bit. Unambiguous values for L_{bond} and R_{bond} can be found by changing the simulation frequency

$$R_{bond} = \lim_{\omega \rightarrow 0} \frac{1}{\Re\{Y_{12}\}} \quad (5.14)$$

$$L_{bond} = \lim_{\omega \rightarrow \infty} \frac{1}{-\omega \cdot \Im\{Y_{12}\}} \quad (5.15)$$

Using an infinite simulation frequency is not practical, but a high simulation frequency such as 40 GHz should do the trick. This results in a $R_{bond} = 44,5 \text{ m}\Omega$ and $L_{bond} = 24,8 \text{ pH}$. This is verified by comparing the gate waveforms resulting from using these component values to using the flip-chip bond S-parameter model based on the 2.5D EM simulation. These waveforms are shown in figure 5.16 for both the smallest and largest transistor. It can be seen that they overlap nicely, with a bit more ringing when the lumped model is applied in the case of the small transistor. The S-parameter models show some ringing artifacts just before a transition, which could be explained by the S-parameters being extrapolated to higher frequencies. In conclusion the assumed RL-series model accurately describes the flip-chip bond behavior. It was initially overestimated at an inductance of 0,1 nH, although possible additional interconnect might be required when constructing the layout which adds inductance.

Simulation Result when Including Bond Inductance

The same DRAC design is used for simulation, but now including 24,8 pH of flip-chip bond inductance. Repeating the simulation from figure 5.10 now gives optimum efficiency at $\tau_d = 20 \text{ pF } \Omega$. Again the time

¹2.5D refers to all dielectric media only having arbitrary shapes in the X and Y direction and a fixed thickness in the Z direction. EM fields are calculated in all directions. This type of simulation is sometimes also called 3D planar EM simulation.

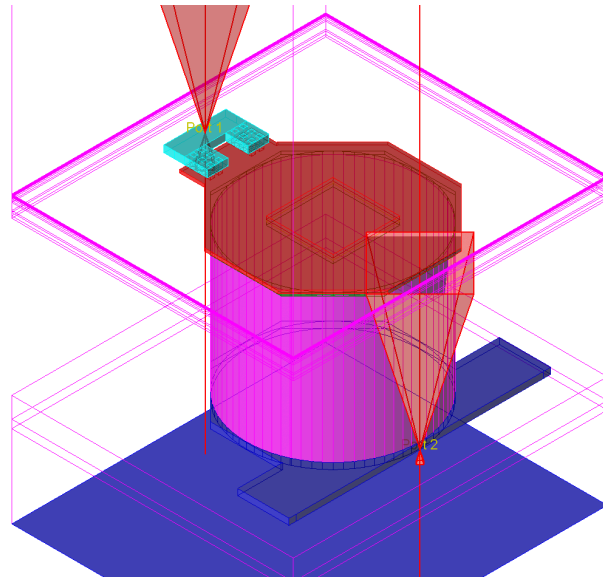
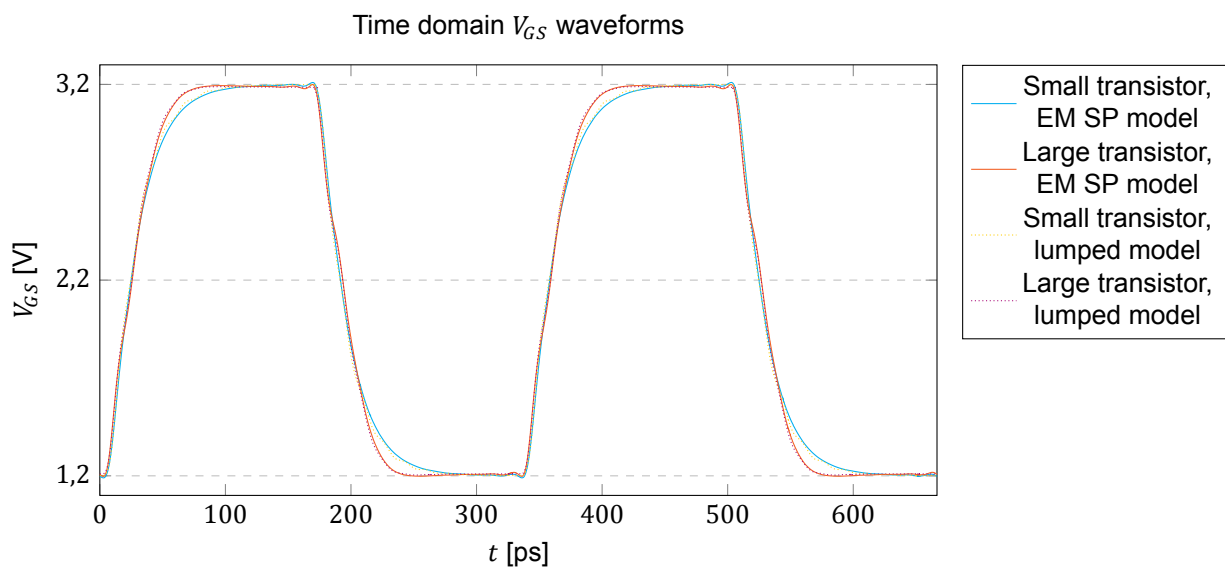


Figure 5.15: EM model of flip-chip bond

Figure 5.16: Time domain V_{GS} waveforms for the largest and smallest transistors when $\tau_d = 17 \text{ pF} \Omega$, solid: using EM bond models; dashed: using lumped component model

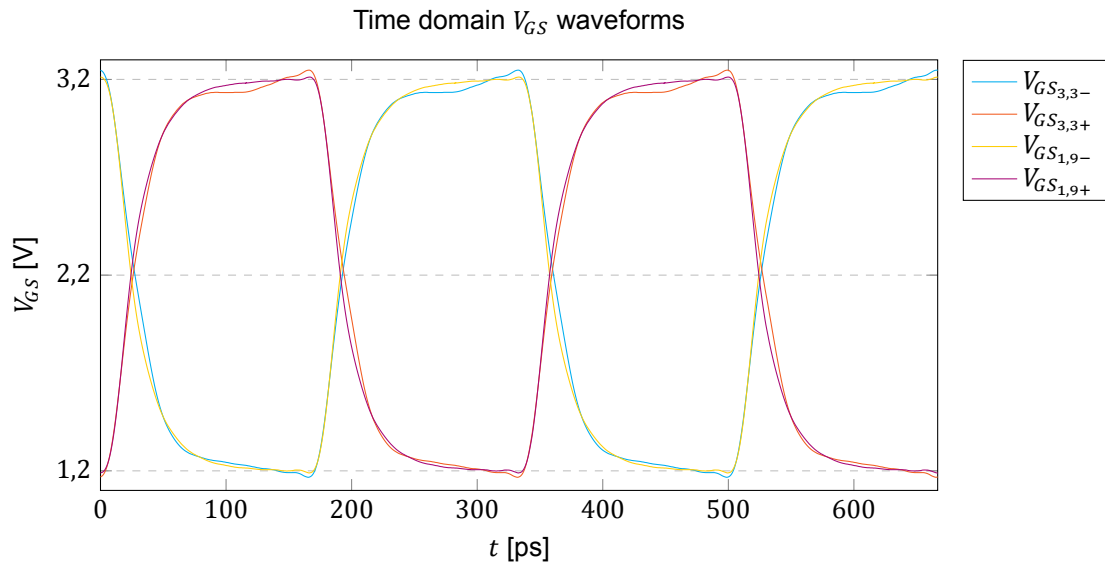


Figure 5.17: Time domain V_{GS} waveforms for the largest and smallest transistors when $\tau_d = 20 \text{ pF } \Omega$ and $L_{bond} = 24,8 \text{ pH}$

domain V_{GS} waveforms can be plotted of the smallest and largest transistors, as shown in figure 5.17. The delays from the driver to the LDMOS gates are slightly different, although the same τ_d is used for the transistors, but at this point nothing too worrisome. If it does start to become a problem this may be resolved by multiphase RF clocking (from [33] and mentioned in section 3.3) not only between segments, but also for the largest transistors. For the larger transistors V_{GS} now shows a slightly higher voltage amplitude short after a transition.

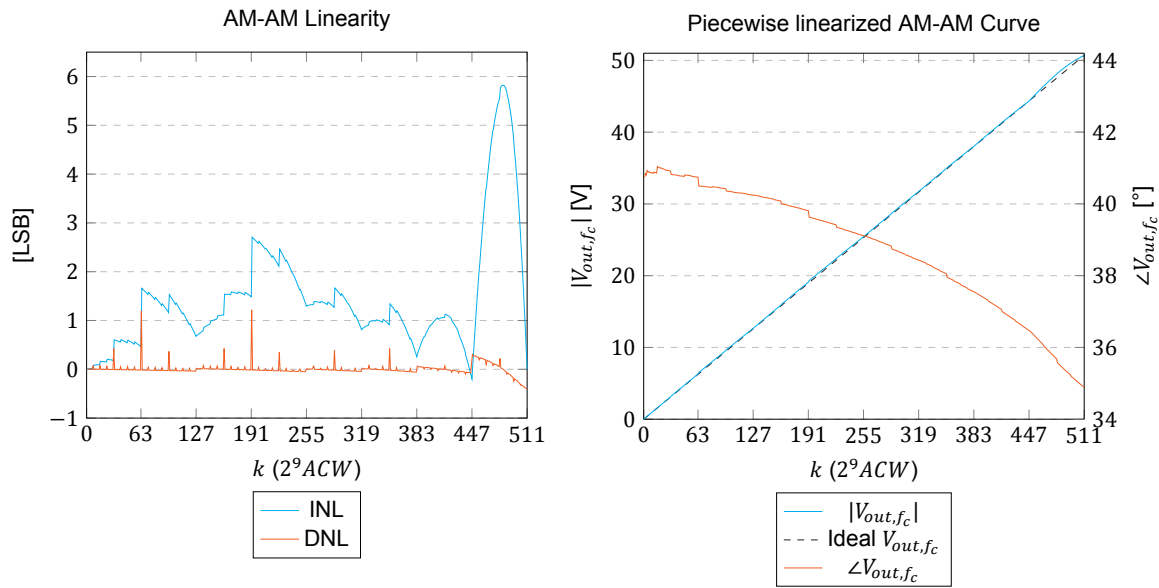
The AM-AM linearity is shown in figure 5.18a and the AM-AM curve in figure 5.18b. The segment sizes are not adjusted, as can be deduced from the INL curve not touching 0 at the end of each segment. The spikes in the DNL curve are now more pronounced, especially for the largest transistors that are controlled by $ACW_{1,3}$ and $ACW_{3,3}$. Using equation (4.21) the k -value these transistors turn on can be calculated, which are $k = 2^9 ACW = 2^9 \cdot ([1-1]2^{-3} + 2^{-3}) = 64$ and $k = 2^9 \cdot ([3-1]2^{-3} + 2^{-3}) = 192$ respectively. This behavior can be explained by the slightly higher V_{GS} amplitude short after a transition. The transistor controlled by $ACW_{8,4}$ is also one of the larger transistors, but the effect is not as noticeable at $k = 2^9 \cdot ([8-1]2^{-3} + 2^{-4}) = 480$. The effect is relatively smaller due to the DRAC operating more into the voltage limited region, and is partially masked by the fact that in figure 5.13a it caused the DNL to have a negative spike. Similarly for $ACW_{7,4}$ ($k = 2^9 \cdot ([7-1]2^{-3} + 2^{-4}) = 416$).

The AM-PM curve has decreased by approximately a constant 3° , which can be explained by slightly more delay from the buffer to the LDMOS gates. This is not an issue, the relative differences in the AM-PM curve are important. The general curve has not changed, other than that it is now slightly more irregular.

Sensitivity to Bond Inductance

It may be clear that L_{bond} will influence the performance of the DRAC due to the differently sized gate capacitances C_{GS} . Unfortunately L_{bond} is more of a parasitic than a designable parameter, so it would make sense to investigate how much influence this has on the DRAC performance.

In figure 5.19 two graphs are shown that can help predict performance. In the first graph the V_{GS} amplitude is shown of the smallest and largest transistor for the fundamental, 3rd and 5th harmonic. For reference the value $4/\pi$ is shown as a dotted line, which is the amplitude of the fundamental in an ideal square wave. For larger values of L_{bond} the fundamental voltage for the larger transistor increases significantly, while that does not happen for the smaller one. Although in the end it is the time domain waveform that is important, it is safe to conclude that with increasing L_{bond} the observed spikes in the DNL curve will worsen. This can be resolved by making the large transistors slightly smaller or by reducing its driver's drive strength. Also the harmonics decrease more rapidly for the large transistor with increasing L_{bond} , meaning that it's waveform will resemble more a sine-wave than a square-wave. This will lead to lowered drain efficiency due to switching losses. This can also be seen as general



(a) AM-AM INL and DNL

(b) AM-AM curve and AM-PM curve

Figure 5.18: AM-AM linearity of implemented DRAC with $N_b = 9$ bit, using $L_{bond} = 24,8$ pH

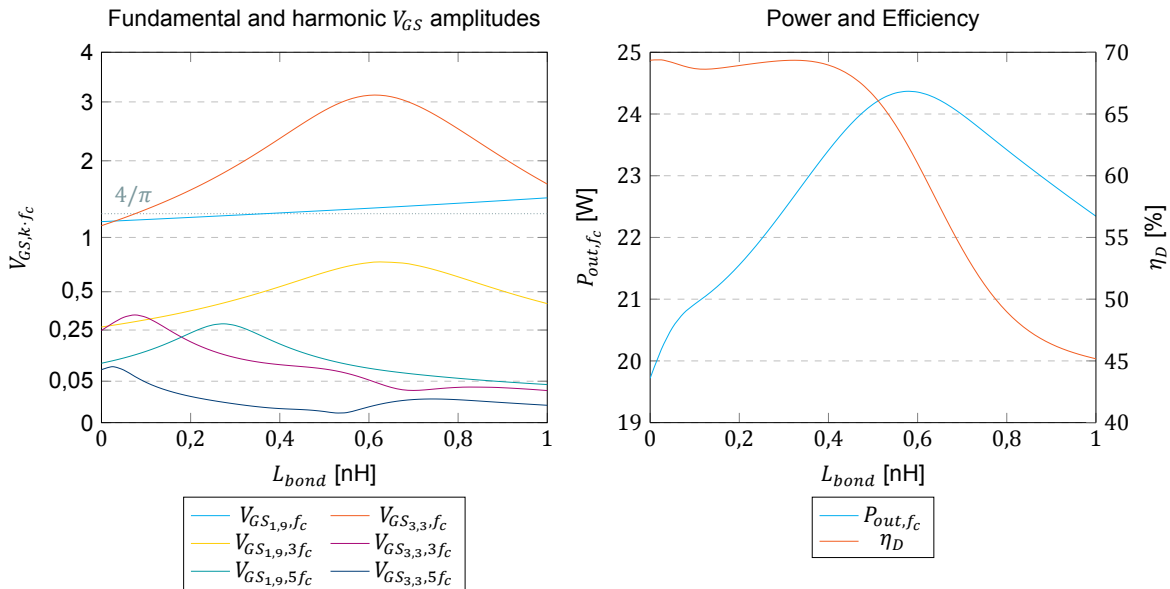


Figure 5.19: Effect of L_{bond} on V_{GS} , power and efficiency

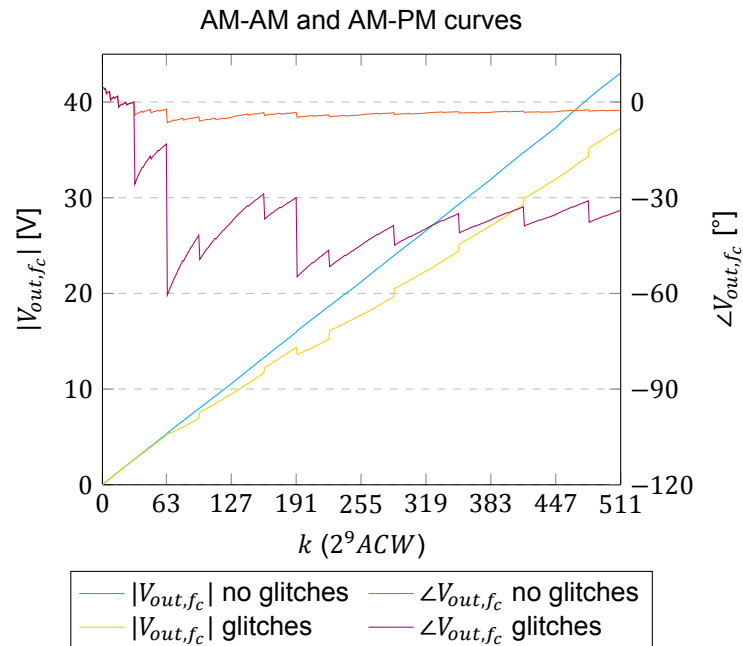


Figure 5.20: Effect of glitches on AM-AM and AM-PM

trend in the second graph. It does show a peak in both power and efficiency, which can be explained by the increased drive voltage.

Sensitivity to Glitches

The AM-AM curves shown so far are all determined in a static environment, but the DRAC will be used in a dynamic setting. With other words, the values of $V_{out,fc}$ so far are determined when fully settled for a given input ACW. The full RF-DAC will have changing phase and input ACW, where the amplitude information is not necessarily synchronous with the rising or falling edges of the phase modulated RF clock. This may cause problems in terms of glitches as the DRAC is partially binary-weighted. A glitch is in principle caused by timing mismatches between the LSBs and MSBs, as explained in section 3.1.2.

The timing mismatch can be caused by different mechanisms. First the flip-chip bond itself can cause different delays for the small and large transistors, which was already observed slightly in figure 5.17. The delay difference here is 1,64 ps, equivalent to a phase difference of 1,77° at 3,00 GHz. That is due to the flip-chip bond only, not including any other layout related effects.

Secondly, the driver can cause difference in delays. The CMOS driver will consist of multiple driver stages in order to be able to drive the total capacitive load. The number of required stages and their sizing will differ between the large and the small transistors in order to have the same τ_d for all LDMOS transistors. Most likely this will contribute more timing mismatch than the flip-chip bond.

What the actual delays and their differences will be is shrouded with uncertainty. Partly due to an unknown layout at this point, but mainly due to PVT variations in the CMOS process. To investigate what the influence of timing mismatch could be—without actually knowing what the delays are going to be—a simulation is done with some introduced driver delay for the largest MSB transistors of each segment ($n = 4$). The applied delay increases slightly with each segment, as the size of each segment increases as well. The first two segments have an additional transistor for which $n = 3$, whose driver has double delay applied compared to the delay for $n = 4$. This is because those transistors will also be twice the width. The delays are chosen such way that for $ACW_{3,3}$ the delay is slightly more than, but comparable to, the delay applied to $ACW_{7,4}$.

In figure 5.20 the AM-AM and AM-PM curves are shown in the presence of these timing delays that introduce glitches, for reference the case without glitches is shown. The average applied delay for the MSB transistor drivers is 42 ps (46°), with a maximum of 60 ps (65°) for the largest drivers. From the plot can be seen there is both significant AM-AM and AM-PM distortion. Some output voltages are not present at all or the output is non-monotonous.

It may be clear that these timing mismatches should be prevented. Making sure that the delay is robust by design is inherently impossible due to the binary-weighted nature of the devices, too many uncertainties and production variations. Therefore the delay should be post-production tunable for the differently sized devices, which can be done by e.g. a programmable delay-line. This is independent from the required timing alignment of amplitude and phase signals for polar architectures in general. Therefore it should be applied after the combination of the two, before the chain of driver stages. The exact implementation requirements for such delay-line is mostly dependent on the rest of the CMOS driver design, but it is expected that it is feasible to implement. More exact details on the requirements of these programmable delay-lines is given in section 6.2.4.

AM-PM Linearization

From figure 5.18b can be seen that the AM-AM curve is reasonably linear, assuming no glitches. The AM-PM curve, however, still requires some correction. This can be done either by using DPD for the phase or by multiphase RF clocking: applying a phase offset for the different segments (section 3.3). This should have no further effects on the CMOS driver design, as it is now expected to have programmable delay-lines anyway. It might require a certain precision of the delay-line.

In order to straighten the output phase response more delay is required for the lower segments. The last segment ($i = 8$) doesn't require any delay. These delays can be found by doing something similar to the AM-AM linearization using the points specified by equation (5.4). Every smaller segment requires the same or more delay than the next segment

$$t_{d,i} = \sum_{k=i}^7 t_{\delta,k} \quad \text{for } i = 1, 3, 5, 6, 7 \quad (5.16)$$

where $i = 2$ and $i = 4$ are skipped due to the merging of these segments, $t_{d,i}$ the applied delay to the i^{th} segment and $t_{\delta,i}$ the delay difference between the segments $d_{ACW,i}$ and $d_{ACW,i+1}$. First $t_{\delta,1}$ is determined, then $t_{\delta,3}$ and so on. It is done in that order since the delays of the first segments will influence the output phase of the later segments, but not the other way around. This may sound counter intuitive as the delays of the later segments will be added to that of the first segments. However, it is not the absolute output phase that counts, but rather the change in output phase between the different input ACWs. the delay is not straightforward to determine due to the influence of the next segment, but an estimate can be made using

$$t_{\delta,i} = \begin{cases} \frac{\angle V_{out,fc}(k[i]) - \angle V_{out,fc}(k[i+1])}{360^\circ} \cdot \frac{c_i}{f_c} & \text{for } i = 4, 5, 6, 7 \\ \frac{\angle V_{out,fc}(k[i]) - \angle V_{out,fc}(k[i+2])}{360^\circ} \cdot \frac{c_i}{f_c} & \text{for } i = 2 \end{cases} \quad (5.17)$$

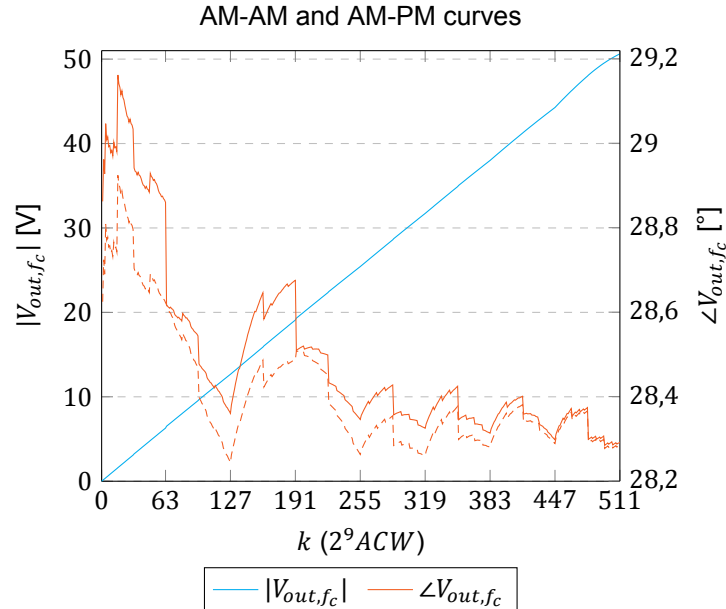
where c_i an empirical value, approximately between 2 and 3. Implicitly this means that $t_{\delta,1} = t_{\delta,3} = 0$. Since $d_{ACW,2}$ and $d_{ACW,4}$ do not exist anymore the delays $t_{\delta,2}$ and $t_{\delta,4}$ are solely used to influence $t_{d,1}$ and $t_{d,3}$.

In table 5.6 the expected delays ($c_i = 1$) are shown with the found delays (precise to 0,1 ps) that result in the output phases $\angle V_{out,fc}(k[i])$ being approximately equal. The resulting AM-AM and AM-PM curves are shown in figure 5.21, proving that this method can linearize the AM-PM response. The peak phase deviation is 0,88°, mainly caused by a phase jump due to $ACW_{1,3}$ turning on. If the delay $t_{d,2}$ would be applied only for this transistor and $t_{d,1}$ for the rest of the segment the phase deviation would be even less, as shown dashed. With the presence of individual delay-lines this would certainly be possible to implement. It can be concluded that a delay-line resolution of 0,1 ps will result in good AM-PM linearity. A resolution of 1 ps will most likely result in good enough AM-PM linearity as other components (such as the phase modulator) could introduce larger phase errors than this.

The required amount of delay is relatively small as the phase deviation was not much to start with, a total of 5,8°. This is due to the same reason as the relatively linear baseline AM-AM curve (see section 5.1.2): the DRAC spends most of the ACWs in the current limited region. When a higher drive voltage is used the total phase deviation is 24,4° as can be seen in appendix B. These delays affect the drain efficiency of the DRAC slightly. Without them $\eta_D = 69,4\%$ and with $\eta_D = 69,0\%$. This is due to the intentional 'misfiring' of the segments, resulting in the segments slightly opposing one another.

Table 5.6: Multiphase RF clocking delays

	$t_{\delta,2}$	$t_{\delta,4}$	$t_{\delta,5}$	$t_{\delta,6}$	$t_{\delta,7}$
expected	1,04 ps	0,63 ps	0,82 ps	1,00 ps	1,47 ps
applied	2,0 ps	1,9 ps	2,1 ps	2,2 ps	2,8 ps

Figure 5.21: AM-AM and AM-PM curves after applying multiphase RF clocking, dashed phase: additional phase correction for $ACW_{1,3}$ and $ACW_{3,3}$

It is expected that the drain efficiency will suffer more from multiphase RF clocking when there is more AM-PM distortion present.

5.3. Layout

From previous simulations it is known that the design is fairly sensitive to series inductance from the driver to the LDMOS gate. Therefore it is important to have as little as possible interconnect from the driver to the DRAC to ensure predictable performance. It is expected that the *drain bar*, the interconnect at the drain side of the LDMOS devices to the output, will have less influence on the performance.

The drain bar will add parallel capacitance at the drain side and series resistance and inductance to the output. The parallel capacitance will increase the design set parameter m and is therefore expected to decrease efficiency. Also the series resistance will introduce losses. It is expected that the series inductance can be incorporated into the Class-E output matching network, for example in the additional reactance X . These drain bar parasitics of distributed nature, in contrast to being a lumped component. It is therefore suitable to see them as lossy transmission line, which is likely to introduce different delays from each of the transistor drains to the output. It is expected that this effect can be compensated by having a delay-line for each individual transistor, aligning the phases of the transistors with respect to the output. These delay-lines would be present anyway to prevent glitches and apply AM-PM linearization.

Design Rules

The layout will need to be compatible with three sets of design rules: those enforced by Ampleon's LDMOS LM8, TSMC's 40 nm CMOS and the requirements for flip-chip bonding.

Important constraints for the LDMOS include that all gates should point in the same direction. The minimum pitch between gate fingers is 37 μm . There are 5 metal layers available on the chip, all of which are used in the transistors. The top metal layer M5 can be used for the drain bar, but may obviously not run over the transistors since it is also in use there. Flip-chip connections can be made arbitrarily

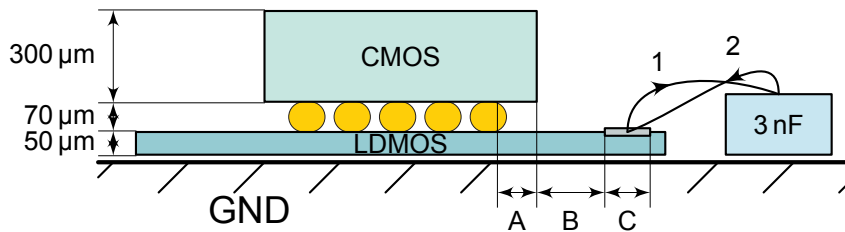


Figure 5.22: Flip-chip setup illustrating inshing and distance requirements for wirebonding

to M5. Further, ESD protection is required for each individual transistor. Its core is $118\ \mu\text{m} \times 69\ \mu\text{m}$ and has to point in the same direction as the LDMOS gates. It requires some additional clearance from other components, making it $162\ \mu\text{m} \times 113\ \mu\text{m}$ in total. That gives $22\ \mu\text{m}$ of possible overlap between the individual ESD protections. They can be placed under patches of (at least) the top metal layer M5, which is important to know for the flip-chip connections.

For CMOS there is an abundance of design rules which are not relevant at this point. The most constraining one for the LDMOS layout is that the ESD protection can only be placed in straight lines and 90° corners. All corners must turn in the same direction, resulting in a maximum of 4 corners, essentially forming an ESD ring. This ring has an 'outside' and 'inside' area. At the outside of the ESD protection no active devices may be placed, only at the inside. Flip-chip connections are done to the topmost metal layer AP, under which no upper metal layers may be used. In the ESD protection all metals up to M5 are used, leaving M6, M7 and AP available for routing over the ESD ring. The flip-chip connections can be made to both the inside or the outside of the ring.

The CMOS design rules also dictate most of the flip-chip restrictions. The minimum *pitch* of the flip-chip *bump pads*, the minimum distance between their centers, is $162\ \mu\text{m}$. The minimum sized patch of AP used for these bump pads is $84,6\ \mu\text{m}$ and has an octagonal shape. There are also larger ones available, roughly increasing in steps of $9\ \mu\text{m}$ in size and pitch. On the LDMOS an additional $5\ \mu\text{m}$ of M5 at all sides of the bump pad is required for passivation. There should be some alignment marker somewhere for placement orientation. For correct flip-chip bonding the flip-chip bonds should be placed in such way that the connections are rotation asymmetric. After that the flip-chip bonding is self aligning. All connections to the CMOS die will need to come from the LDMOS die by means of the flip-chip connections.

All connections from 'the outside world' to the LDMOS are done by bondwires. The minimum pitch for straight wirebonding is $130\ \mu\text{m}$. These bondwires will lead to a need of supply decoupling for both the CMOS and LDMOS. This is to ensure the supply is 'clean', meaning that it only contains a DC voltage and no voltage at RF frequencies. The CMOS supply decoupling can be done by MIM capacitors available in the LDMOS process. These have capacitive density of $240\ \text{pF mm}^{-2}$ to $250\ \text{pF mm}^{-2}$. On the CMOS itself it is possible to combine MOS and Metal-Oxide-Metal (MOM) capacitors to get a capacitive density¹ of $3,6\ \text{fF } \mu\text{m}^{-2}$ to $5,0\ \text{fF } \mu\text{m}^{-2}$. See section 6.3 for more on the CMOS supply decoupling. Although the LDMOS supply has to be fed to the drain bar through an inductor, the required inductance still is quite small and should be well-defined. This inductor can be implemented using a bondwire, but needs a decoupled supply. This can be done using *inshing*, placing a capacitor chip as decoupling next to the LDMOS die.

In figure 5.22 the concept of inshing is shown. For bonding the drain bar should be $167\ \mu\text{m}$ wide, denoted with 'C' in the figure. There are two options to thread the wondwires: 1) from the LDMOS to the decoupling capacitor or 2) the other way around. This influences the shape of the bondwire and the clearance 'B' needed for the bonding machine to the CMOS die, $B \approx 100\ \mu\text{m}$ in case 1 and $B = 340\ \mu\text{m}$ in case 2. The distance 'A' is the distance from a bump pad to the edge of the CMOS die. Assuming the bump pad is the outermost instance of the layout (not the ESD ring) a total of $A = 220\ \mu\text{m}$ is required by the CMOS design rules. First a seal ring must be placed around the entire CMOS chip spaced at least $50\ \mu\text{m}$ from any other layout instance and is $20\ \mu\text{m}$ wide. This seal ring also serves as crack stop in the corners. Then $100\ \mu\text{m}$ from there the saw line starts. This saw line generally is not perfectly straight, but has a certain roughness. This can be up to $50\ \mu\text{m}$ from the start of the saw line.

¹ $1\ \text{fF m}^{-2} = 1000\ \text{pF mm}^{-2}$. The capacitive density of the MIM capacitors on LDMOS is lower as they allow a much higher voltage.

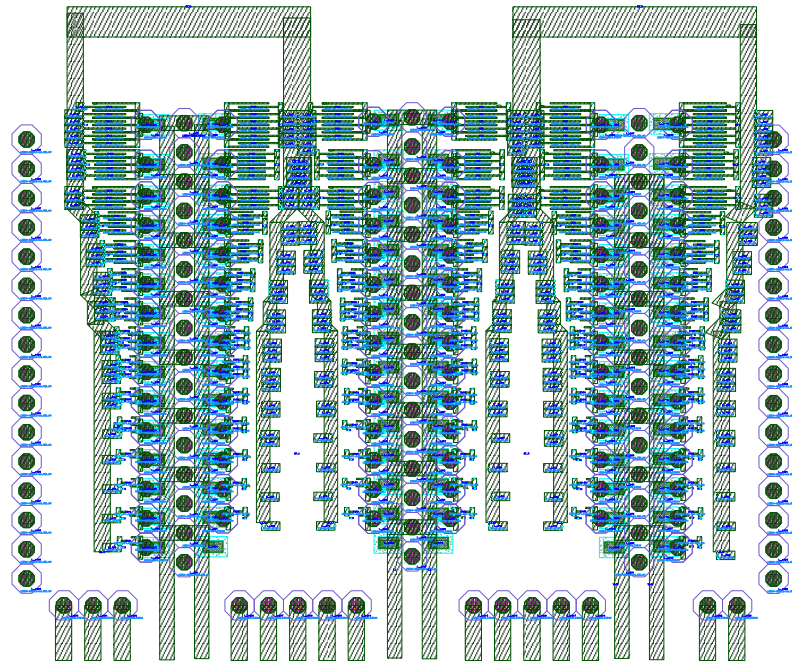


Figure 5.23: LDMOS layout version 1

5.3.1. Layout Version 1

Taking into account all the design rules a first layout version was created, as shown in figure 5.23. The M5 layer used for the drain bar is shown in dark green. The drain bar output is at the top and is shaped as a fork with two teeth. One forks again into two other teeth, effectively resulting in three teeth. Along one tooth two segments are situated, sorted from the largest transistors at the top to the smallest at the bottom. The largest transistors are placed closest to the output to minimize losses as they will draw the largest currents. All flip-chip bump pads are placed on a line and can be recognized by the purple octagonals indicating the minimum pitch of $162\ \mu\text{m}$ to another bump pad. The bump pads are directly connected to the LDMOS gates. At the sides and bottom of the layout some bump pads are placed that may be used for digital interfacing to the CMOS or for connecting DC supply power. The $V_{DDh} = 2,0\text{V}$ supply for the drivers are shown as 3 pairs of M5 running from the bottom to almost the top, connected to bump pads in the middle and are alternated with ground pads. Under these lines of M5 MIM capacitors can be placed for supply decoupling. The empty space at the top is required for the mentioned $220\ \mu\text{m}$ space to the CMOS edge and then additional $340\ \mu\text{m}$ clearance for wire bonding. The total length from the farthest point of the drain bar to the output is $3016\ \mu\text{m}$.

The matching CMOS layout is shown in figure 5.24. Only the corners of the ESD protection is shown, creating a total of four ESD rings. Dotted in white sheets of some top metal are shown that functions as a ground plane, shielding against capacitive and inductive coupling from the LDMOS.

Layout Performance

A S-parameter model of the layout has been created using a Momentum MoM RF 2.5D EM simulation. EM ports have been added for the output and every transistor drain. The flip-chip bonds have been removed to simplify the layout, otherwise the EM simulation would take up incredible amounts of memory and take an unfeasibly long time. The CMOS substrate with the shield metal is included in the layout simulation.

The layout can be characterized by connecting port impedance terminations for S-parameter simulation at the two outputs (ports 1 and 2) and at the farthest ends of the three layout teeth (ports 3 and 4). No transistor models are connected to the layout for this simulation. By converting the S-parameters to Z-parameters at a low frequency the capacitance C_{layout} added by one drain bar can be investigated. Looking into the two outputs gives very similar results as the layout is symmetrical

$$C_{layout} = \lim_{\omega \rightarrow 0} \frac{1}{-\omega \cdot \Im\{Z_{11}\}} \approx \lim_{\omega \rightarrow 0} \frac{1}{-\omega \cdot \Im\{Z_{22}\}} \quad (5.18)$$

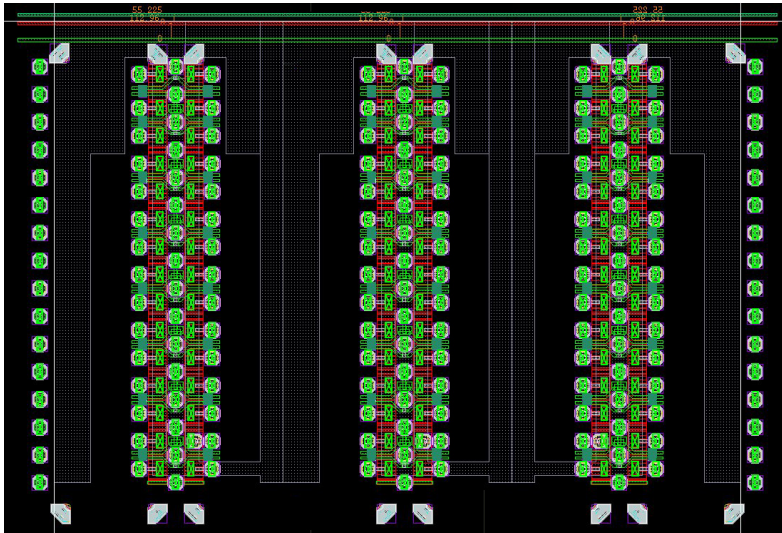


Figure 5.24: CMOS layout version 1

Evaluating the Z-parameters at 100 MHz gives $C_{layout} \approx 4,75$ pF. This results in

$$C_S = C_{DS} + 2C_{layout} = 14,99 \text{ pF} + 2 \cdot 4,75 \text{ pF} = 24,50 \text{ pF} \quad (5.19)$$

$$m = \omega_c R_{on} C_S = 6\pi \cdot 10^9 \cdot 0,2080 \cdot 24,49 = 96,0 \cdot 10^{-3} \quad (5.20)$$

whereas it used to be $m = 58,8 \cdot 10^{-3}$ without the layout, which will have influence on the design set and the drain efficiency.

The layout losses can be determined by calculating $|S_{13}^2| + |S_{33}^2| \approx |S_{24}^2| + |S_{44}^2|$ at 3 GHz, which equates to $-0,11$ dB, approximately 2,5 % loss.

Another expected layout effect was that the drain bar would act as a transmission line, introducing delays from the transistor drains to the output. The same S-parameter simulation setup is used to verify this hypothesis. The simulated S-parameters are shown in figure 5.25, which show circular patters that can be associated with a lossy transmission line. At $f_c = 3$ GHz the full layout shows 37,7 ps delay, equivalent with $40,8^\circ$ phase shift. This delay is applicable to the smallest transistors. The delay for the larger transistors will be lower.

This phase shift was expected to be compensable by applying more delay in the driver for the largest transistors and none for the smallest. To investigate how much delay would be required the transistor models were connected to the layout and the phase difference with respect to the smallest transistor, the transistor farthest from the output, was determined. The result is shown in figure 5.26. The approximate linear relationship of the phase difference with frequency indicates that the delay will remain constant with operation frequency, so frequency agility would be preserved after initial delay calibration is performed.

However it turned out it was not possible to match all transistors to the output. A working class-E design set could be found, but then only the largest transistors were correctly matched. The smaller devices could not be correctly matched and were primarily generating unwanted harmonics. The hypothesis that the delay caused by the layout could be compensated is hereby rejected and this version of the layout therefore abandoned.

5.3.2. Layout Version 2

From the first version of the layout it became clear that the design is more sensitive to the layout effects of the drain bar than initially thought. Therefore the design should be made even more compact. This led to the LDMOS layout as shown in figure 5.27 and the CMOS layout as shown in figure 5.28. Again the drain bar output can be found at the top with the largest transistors connected closest to the output. The transistors used are long enough to fill up the required $560 \mu\text{m}$ space to the CMOS edge and the clearance for wire bonding. The flip-chip bump pads are placed on a line and are directly connected to the gates of the largest LDMOS transistors. A gap is left open to the next row of bump pads. This

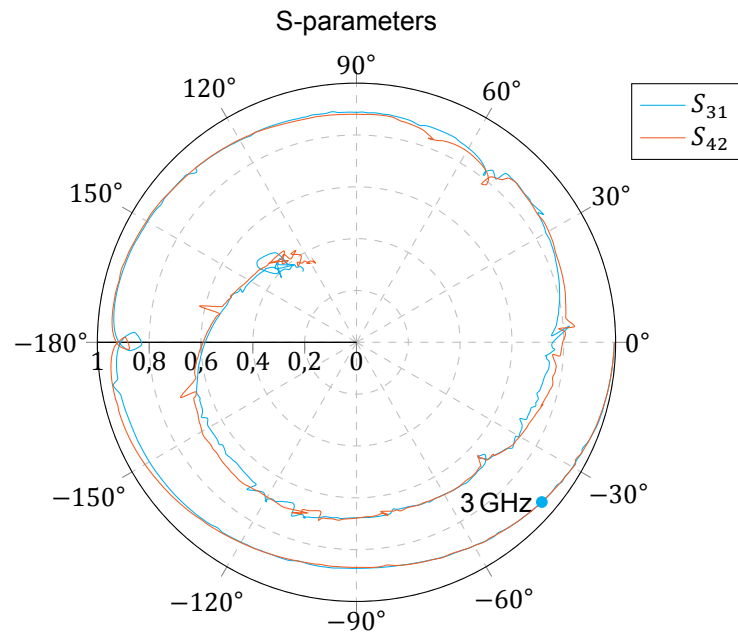


Figure 5.25: S-parameters from farthest input to output from 0 GHz to 48 GHz

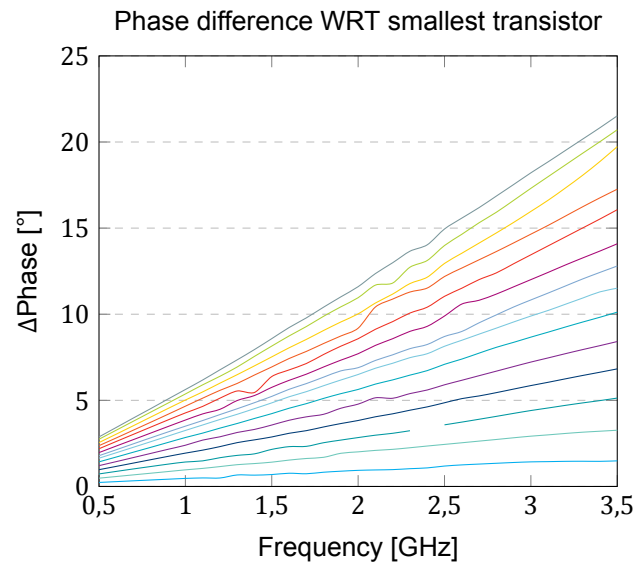


Figure 5.26: Phase differences due to the layout with respect to the smallest transistor; each line represents a transistor along a layout tooth, increasing phase difference means a transistor closer to the output

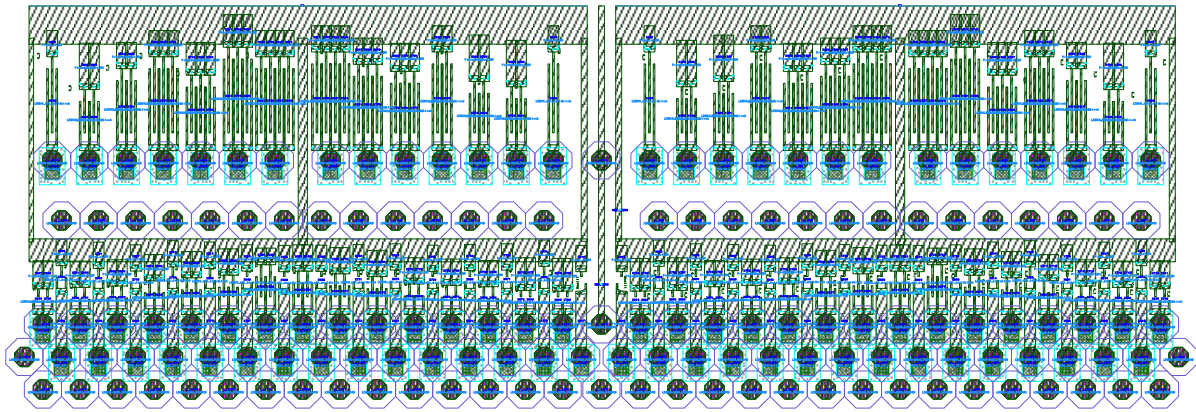


Figure 5.27: LDMOS layout version 2

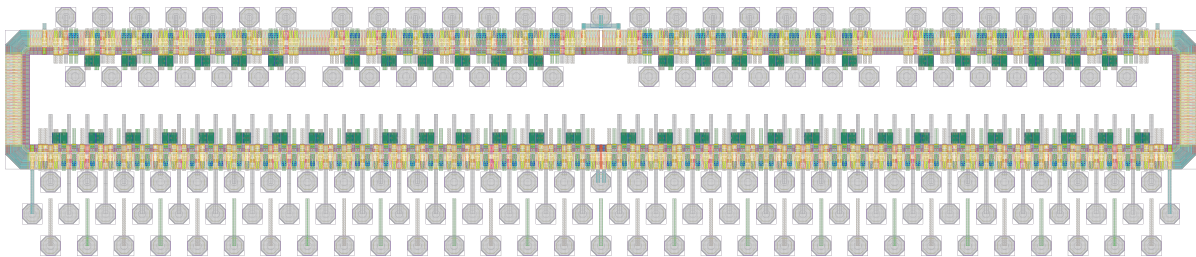


Figure 5.28: CMOS layout version 2

allows the placement of the ESD ring of the CMOS chip, which is clearly visible in the CMOS layout. These bump pads are used for the CMOS V_{DDh} supply and ground. On the LDMOS these can be routed using M4 or lower, as the drain bar is made using M5.

After that the drain bar continues with a second row of connected LDMOS transistors. These are the smaller transistors of the design. The total length from the second row of the drain bar to the output is now 1116,9 μm , roughly a third of layout version 1. Three rows of bump pads can be identified below that. The first two will be used to connect to the LDMOS gates. The second of the two is used for the very smallest transistors as their gate capacitance is the lowest. The ESD protection is the more significant capacitance in this case, which can be placed directly under the bump pad. This ESD protection can be identified by the cyan colored borders. The inductance caused by the additional interconnect is suspected to not cause additional problems since only a very small gate capacitance can be found at the end of it. The bottom row of bump pads is again used to provide the CMOS chip of V_{DDh} supply and ground. The left and rightmost bump pads (to be found in the second bottom row) are additional ground pads required for the CMOS ESD protection. They can be recognized by the lack of placed ESD protection under the bump pads or the lack of any connected M5 to them. Two additional of these bump pads can be found in the very middle of the design. These have a piece of M5 connected which acts as a shield between the two push-pull sides, preventing capacitive coupling between the two.

Lastly, the CMOS ESD ring has been verified to comply with all CMOS design rules. The area within that ring will be used to place the 'high-voltage' ($V_{DDh} \leq 2,5\text{V}$) drivers for the DRAC. The remaining digital logic and phase modulators will be included in some other part of the layout operating at a lower voltage. It is still to be designed and therefore not shown in figure 5.28.

Layout Performance

A S-parameter model of the layout has been created using a Momentum MoM RF 2.5D EM simulation. Just as for version 1 of the layout the EM ports have been added for the output and every transistor drain. The flip-chip bonds have again been removed to simplify the layout.

The layout can be characterized by connecting port impedance terminations for S-parameter simulation at the two outputs (ports 1 and 2) and at the second row of the drain bar (ports 3 and 4). No transistor models are connected to the layout for this simulation. Again the S-parameters are converted

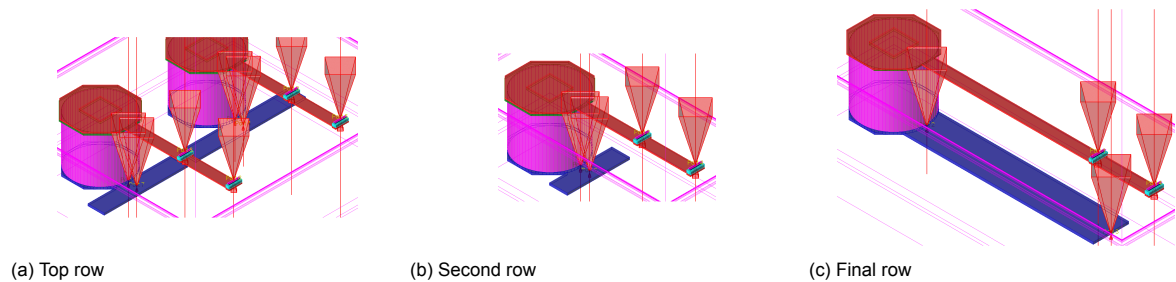


Figure 5.29: 3D layouts of flip-chip connections, bump pads and required interconnect to ESD capacitances and CMOS drive transistors

Table 5.7: Simulated parasitic component values including interconnect

Quantity	L_{bond}	R_{bond}
Top row	0,11 nH	0,28 Ω
Second row	0,11 nH	0,28 Ω
Final row	0,25 nH	0,50 Ω

to Z-parameters to determine the capacitance C_{layout} added by one drain bar. At 100 MHz it is found that $C_{layout} \approx 5,06$ pF, resulting in $C_S = 25,11$ pF and $m = 98,5 \cdot 10^{-3}$. This is slightly larger than the first layout version. However, the layout losses are found to be only 0,033 dB at 3 GHz, approximately 0,76 % loss.

The ‘post-layout’ flip-chip bond parasitics are estimated by applying the same method as in section 5.2.2 (‘Flip-Chip Bond Characterization’). The 3D EM previews of the three different rows of bonds are shown in figure 5.29. The resulting parasitic component values are shown in table 5.7. For the top row also the mutual inductance between two flip-chip bonds is investigated as they will carry the largest currents, since they are connected to the largest transistors. This mutual inductance M_{bond} is low with approximately 3 pH and therefore does not require special attention. The final row has the largest inductance—as expected—due to additional interconnect. The largest portion of the connected capacitance can be found closer to the drive transistors, effectively resulting in a lower series inductance value. The capacitance added by the flip-chip bond is relatively small, approximately 80 fF for the top and second rows and approximately 130 fF for the final row.

The parasitics added by the layout require adjustments in the design set in order to operate most efficient. The resulting design set and component values are shown in table 5.8. Most notably is the increase of m due to increased C_S and with it the increase of K_C . This results in a slightly lowered K_P and drain efficiency η_D . The results of the AM-AM and AM-PM linearization process are shown in tables 5.9 and 5.10.

Using these values result in the AM-AM and AM-PM curves shown in figure 5.30b. In figure 5.30a the resulting linearity is shown. In figure 5.31 the resulting powers and drain efficiencies are shown for $f_c = 3$ GHz, having peak $P_{out,f_c} = 16,56$ W and $\eta_D = 64,4$ %. Two interesting layout effects can be seen. First is the effect of the bond inductance L_{bond} , best observable in the DNL error being close to 1 LSB for values $k = \{63,5; 191,5; 479,5\}$. This is similar to what was observed in figure 5.18a, which is caused by transistors $ACW_{1,3}$, $ACW_{3,3}$ and $ACW_{8,4}$ having a large C_{GS} . This results in more inductive peaking of V_{GS} , driving these transistors with a slightly higher voltage.

However at $k = \{15,5; 31,5; 47,5; 79,5; 95,5; 111,5; 143,5; \dots; 463,5\}$ the opposite can be observed: the DNL error is -1 LSB or worse (thus resulting in non-monotonicity). This is due to the second layout effect. The transistors responsible for this— $ACW_{i,4}$ and $ACW_{i,5}$ for all available i —are located closest to the output of the drain bar. This results in a matching and/or phase difference. It could be solved for only one specific value of k by incorporating a compensating delay for these transistors. This was arbitrarily tested for $k = 367,5$ (transistor $ACW_{6,5}$), resulting in a layout delay of 46 ps. This however worsened the DNL errors for the other values, resulted in large AM-PM jumps of approximately -8° each and lowered drain efficiency to 57 %. This is similar to the observations from the first layout version, but since this layout is much more compact it’s impact is much smaller. Simply put, it seems like a bad

Table 5.8: Design set and values for post-layout $TGW = 51,0$ mm (push-pull) and $f_c = 3$ GHz

Quantity	Value	
q	1,03	
d	1,000	
m	97,0	10^{-3}
K_L	0,12	
K_C	8,00	
K_P	0,45	
K_X	0,11	
Q_L	5,39	
τ_d	31	pF Ω
C_{DS}	7,57	pF
C_S	12,63	pF
R_{on}	0,412	Ω
L_{DC}	0,215	nH
$2 \cdot R_L$	64	Ω
X	0,200	nH
C_0	293	fF
L_0	9,583	nH
$2 \cdot P_{out,f_c}$	16,56	W
η_D	64,4	%

Table 5.9: Scaling factors post-layout

S_1	S_3	S_5	S_6	S_7	S_8	TGW
0,7864	0,8450	0,9181	0,9877	1,1002	1,7433	51,046 mm

Table 5.10: Multiphase RF clocking delays post-layout

$t_{\delta,1}$	$t_{\delta,2}$	$t_{\delta,3}$	$t_{\delta,4}$	$t_{\delta,5}$	$t_{\delta,6}$	$t_{\delta,7}$
1,2 ps	2,2 ps	2,0 ps	0,8 ps	2,8 ps	3,3 ps	5,7 ps

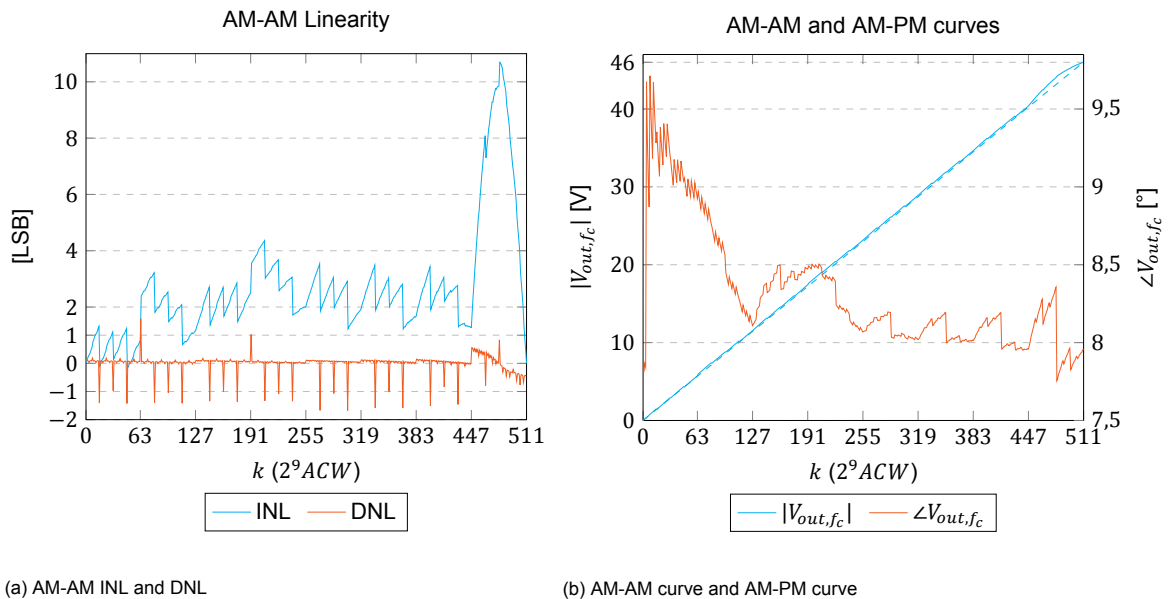


Figure 5.30: Post-layout AM-AM linearity of implemented DRAC at $f_c = 3$ GHz

idea to compensate this second layout effect.

5.4. Conclusion

A Class-E design set is found for which the selected technology (Ampleon LDMOS LM8) can operate at $f_c = 3$ GHz. The LDMOS technology is not yet optimized for this application, so the design set requires a relatively high load resistance R_L in order to get the transistor into the triode region, which can be explained by the relatively large value of m and the low drive voltage of 2 V. This results in a 'saturated Class-B'-like load-line, due to the transistor mostly operating as a current source (the transconductance region) when on. This results also in a fairly straight baseline AM-AM curve.

To reach a peak fundamental output power of 20 W a TGW of approximately 50 mm is required. Since the baseline AM-AM curve is already fairly linear, this transistor width is divided across 6 thermometer-coded segments instead of the expected 8. The internal DRAC resolution is increased to 9 bit to have enough flexibility to compensate the resulting INL errors, primarily for the last segment. This results in a total of 2×43 individual transistors and flip-chip connections.

Using the driver model without (flip-chip) bond inductance the basic design has a linearity according to expectation from theory. It results in a peak drain efficiency of $\eta_D = 69,5\%$ for $f_c = 3$ GHz and peak $P_{out,fc} = 20,2$ W. A total peak drive power of $P_{in} = 2,13$ W is required to drive the SMPA DRAC, resulting in a total efficiency of $\eta_T = 64,7\%$.

It is important to notice, however, that the largest portion of P_{in} is spent on driving the required ESD capacitances. The calculated P_{in} reduces to 0,909 W when only the LDMOS gate capacitance is used. Also, the required input power scales with ACW. This means when there is no output power, there is also no input power present.

When including bond inductance L_{bond} an increase in DNL and INL errors can be observed. This is caused by the largest transistors having a (slightly) larger V_{GS} amplitude due to L_{bond} . The bond inductance also results in a timing mismatch between the small and large transistors, although it is a small one. It is expected that the CMOS driver will introduce additional mismatches, giving rise to possible glitches. Applying an average glitch of 42 ps results in notable changes in both the amplitude and phase of $V_{out,fc}$. To compensate this possible mismatch programmable delay-lines should be implemented in the CMOS driver design. These delay-lines can also be used to apply AM-PM linearization at the cost of drain efficiency. As there is not much AM-PM distortion present there is also low price to pay in terms of drain efficiency.

In the layout the drain bar introduces additional parallel capacitance and series resistance and inductance. These parasitics are of distributed nature, acting as a lossy transmission line. A first layout

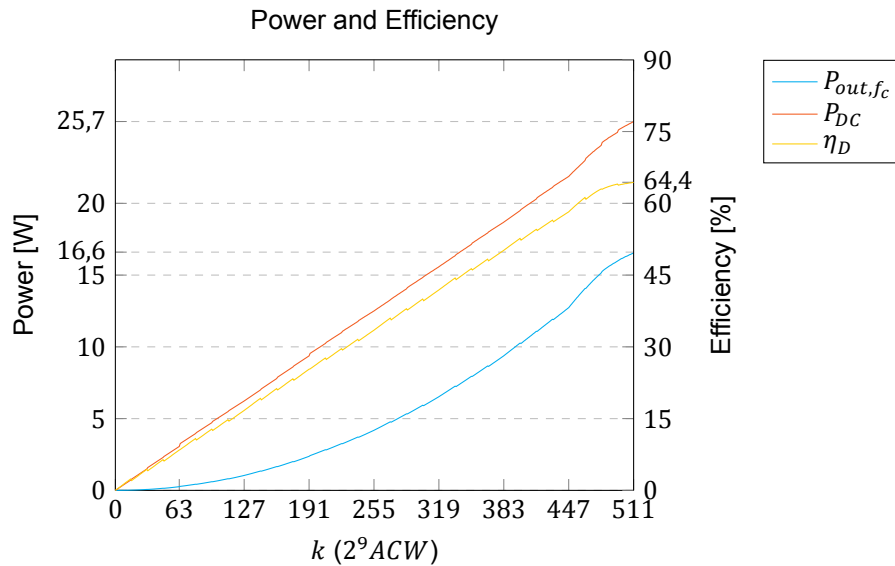


Figure 5.31: Post-layout power and efficiency of implemented DRAC at $f_c = 3$ GHz

was designed with the idea that the transmission line effects could be compensated with driver delays. This turned out not to be true, hence the first layout version was abandoned. A second, more compact layout is introduced. The added parallel capacitance by the drain bar is incorporated in C_S , increasing design set parameter m . This results in a higher K_C and a lower output power and drain efficiency. Post-layout peak fundamental output power $P_{out,fc} = 16,56$ W and drain efficiency $\eta_D = 64,4$ %.

The layout also adds interconnect from the CMOS driver to the LDMOS gates, resulting in an increased L_{bond} . The same increase in V_{GS} amplitude from the pre-layout simulation is observed, resulting in a similar (positive) increase in DNL and INL error. An additional, seemingly opposite, effect is observed which is due to the two ‘rows’ of LDMOS transistors. The different distance to the output creates mismatches, similar to what was observed in the first layout version. However, the second layout is much more compact, so this effect is much smaller.

6

CMOS Driver

The full CMOS driver design lies outside the scope of this thesis. This chapter functions to make a reasonable case for a working CMOS driver and to make an estimate of system performance. TSMC 40 nm 1P7M4X1Z1U CMOS technology is available to create a driver design.

First in section 6.1 the minimum requirements regarding the CMOS driver design are discussed and a block diagram for the full RF-DAC is presented. In section 6.2 the previously proposed model for the driver is expanded with CMOS input and output capacitances. Requirements for the CMOS driver output chain are specified according to this model, which also allows to make a reasonable estimate of driver power requirements. The feasibility of power supply decoupling is investigated in section 6.3. Additional ideas to implement on the CMOS driver are briefly annotated in section 6.4.

Lastly in section 6.5 the complete power RF-DAC system performance is estimated in terms of power and energy efficiency. As the flip-chip design knows some limitations also estimates are made if these limitations become more relaxed by future technological development, which are presented in section 6.5.1.

6.1. Driver Requirements

From chapter 4 it is known that an external driver for the power DRAC is required in order to make it a full RF-DAC. This is necessary since the technology selected for the power DRAC is optimized for high power analog applications and does not have the required high-speed digital logic. The power DRAC is designed to operate using a 2 V peak-to-peak 'square' wave drive signal. This directly results in the first requirements for the CMOS driver: the output buffers have to be constructed with the available 270 nm thick oxide devices that can handle $V_{DDh} \leq 2,5 \text{ V}$. The core voltage of the 40 nm thin oxide devices is $V_{DDl} = 1,1 \text{ V}$, so a (voltage) levelshift is required before an output buffer.

In chapter 4 also the choice is made to give the DRAC a push-pull topology to mitigate even orders of harmonic distortion. This does not necessarily have to operate in push-pull mode, but with other matching networks other modes of operation can be created, such as single-ended, Doherty or carrier aggregation. This requires two separate driver banks which have their own baseband data, DSP (upsampling) and phase modulator, creating a 'dual line-up topology'. Since the DRAC design evaluated so far is push-pull oriented, the two driver banks will be called the 'positive bank' and the 'negative bank'.

In order to prevent glitches and apply AM-PM linearization, in chapter 5 it is concluded that individual programmable delay-lines will be required to compensate (post-production) mismatches between the larger and smaller devices and their output buffers. Depending on the required delay-line range and resolution a course/fine topology is most likely preferred. From chapter 5 it is known that a resolution of 0,1 ps can result in an AM-PM INL less than 1°. Earlier designs of programmable delay-lines are available within the ELCA research group. However these have a resolution of 5 ps.

It is known that the conversion from Cartesian coordinates to polar coordinates is non-linear and hence results in bandwidth expansion. Therefore it could be interesting to experiment with direct polar modulation: upsampling baseband polar data instead of baseband I/Q data. Including any real-time DPD onto the driver is possible but is considered as a later step, especially since the required DPD algorithm is largely unknown prior to production. Supplying the driver real-time with upsampled base-

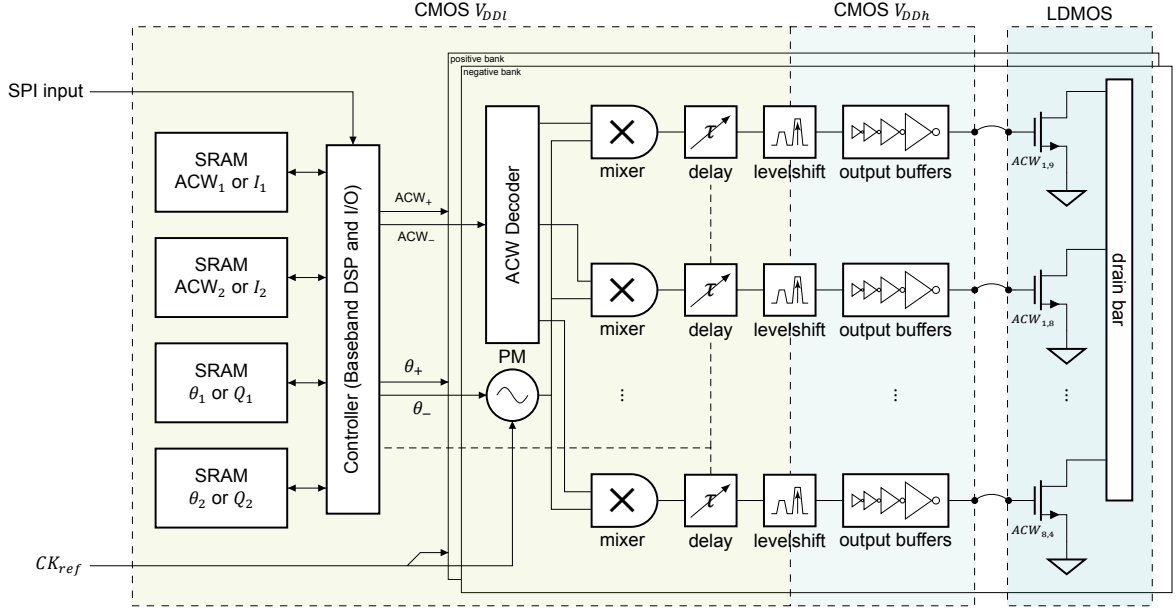


Figure 6.1: Full RF-DAC

band data is feasible but requires interface Intellectual Property (IP) blocks which are not available at this point, so on-chip Static Random-Access Memory (SRAM) cells will be used to store the baseband data for each of the dual line-up banks. The resulting block schematic for the complete aimed RF-DAC is shown in figure 6.1.

Besides the mentioned available design for a programmable delay-line, also several designs for phase modulators are available within the ELCA group. Earlier designs also include DSP and upsampling, hence these items will assumed to be available and not be discussed here.

6.2. Output Buffer

In section 4.3 it was determined that the DRAC driver will operate in Class-D, which is simply a large inverter. However, a significant simplification was made in its model. In section 4.3.1 the driver was modeled as a switch with an Equivalent Series Resistance (ESR) R_{es} , with the ESR as introduced in section 2.3.1. This neglects the parasitic output capacitance of the inverter. It does not invalidate previously obtained results, but it will increase the total capacitance that needs to be driven. Additionally, an inverter will also have an input capacitance which has to be driven as well. This requires a chain of inverters, each increasing in size. This will further increase the total capacitance that needs to be driven, effectively increasing the total input power and decreasing the total system efficiency $\eta_{T,system}$ (see equation (2.13)).

6.2.1. Driver Model Revisited: Inverter Delays

The driver model extended with input and output capacitances is shown in figure 6.2. The ESR and capacitances are dependent on transistor width (see section 2.3.1)

$$R_{esn} \propto W_{gn}^{-1} \qquad R_{esp} \propto W_{gp}^{-1} \qquad (6.1)$$

$$C_{GSn} \propto W_{gn} \qquad C_{GSp} \propto W_{gp} \qquad (6.2)$$

$$C_{DSn} \propto W_{gn} \qquad C_{DSp} \propto W_{gp} \qquad (6.3)$$

As mentioned earlier in section 2.3.4, a PMOS device generally has a lower carrier mobility than its NMOS counterpart, resulting in a device 2 to 3 times wider to have the same R_{es} . The ratio between PMOS and NMOS widths can be defined as [23]

$$\beta = \frac{W_{gp}}{W_{gn}} \qquad (6.4)$$

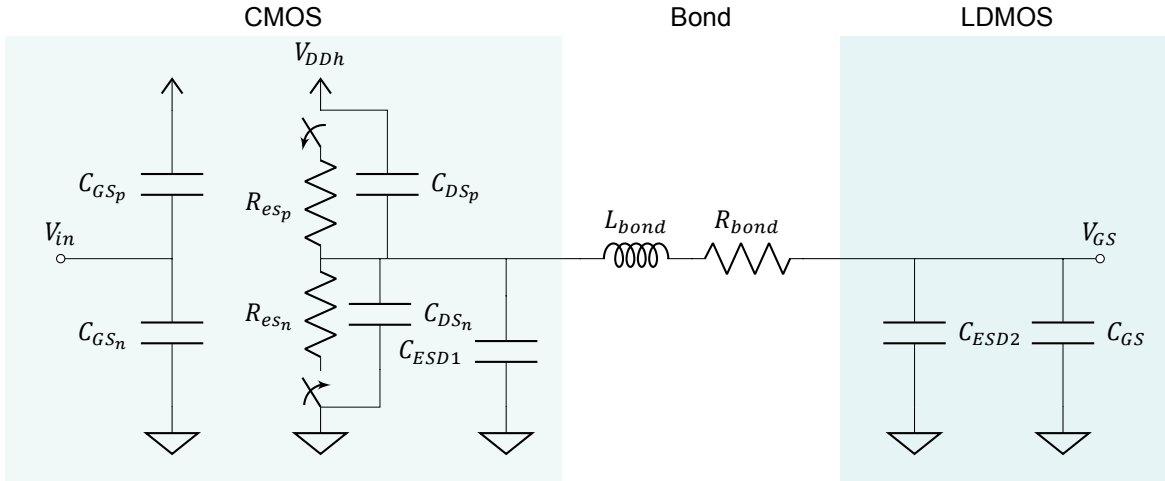


Figure 6.2: Class-D off-chip driver model with CMOS input and output capacitances

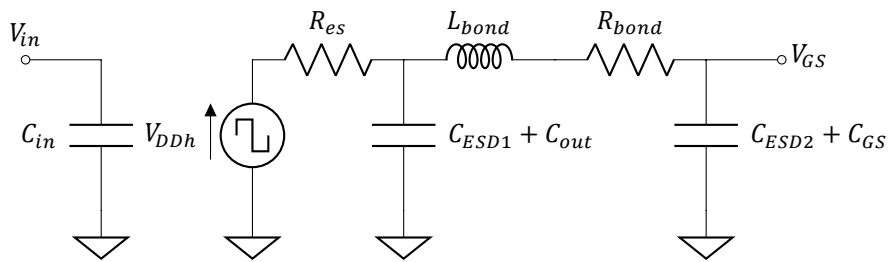


Figure 6.3: Simplified Class-D off-chip driver model with input and output capacitances

Just as in section 4.3.1 this model can be simplified by assuming

$$R_{es_n} = R_{es_p} = R_{es} \quad (6.5)$$

$$C_{in} = C_{GS_n} + C_{GS_p} \quad (6.6)$$

$$C_{out} = C_{DS_n} + C_{DS_p} \quad (6.7)$$

The resulting simplified circuit is shown in figure 6.3. There is a value for β for which equation (6.5) holds. Finding it, however, requires some additional definitions as R_{es} , C_{in} and C_{out} are non-linear.

Recalling the propagation delay from equation (2.43) as being the transition time to 50% of the final value, two propagation delays are applicable for the inverter. The first, $t_{p,H \rightarrow L}$, is when pulling the output down from logic 'high' (or 1) to logic 'low' (or 0), for which the NMOS transistor is responsible. The second, $t_{p,L \rightarrow H}$, is when pulling the output up from low to high, which is the responsibility of the PMOS transistor.

$$t_{p,H \rightarrow L} = \ln(2)R_{es_n}C \quad (6.8)$$

$$t_{p,L \rightarrow H} = \ln(2)R_{es_p}C \quad (6.9)$$

Where C is the total capacitance that can the inverter is loaded with, including its own intrinsic non-linear output capacitance C_{out} . The propagation delay can now be refined for the inverter as the average of the two propagation delays

$$t_p = \frac{t_{p,H \rightarrow L} + t_{p,L \rightarrow H}}{2} \quad (6.10)$$

Increasing β will lower $t_{p,L \rightarrow H}$, but will also increase C_{out} and thus increase $t_{p,H \rightarrow L}$. In order to let equation (6.5) be true it must be that $t_p = t_{p,H \rightarrow L} = t_{p,L \rightarrow H}$. The value of β for which this is the case will be highly dependent on the value of C . Furthermore, in a chain of inverters the propagation delay of the $i-1$ th inverter will increase the propagation delay of the i th inverter

$$t_p^i = t_{step}^i + \Lambda t_p^{i-1} \quad (6.11)$$

where t_{step}^i is the propagation delay of the i^{th} inverter if an ideal step would be applied to its input and Λ an empirical constant.

To demonstrate these concepts for the 270 nm thick oxide CMOS devices, first the *intrinsic inverter delay* t_{p0} is investigated. This is with an unloaded inverter, so $C = C_{out}$, and an (nearly) instantaneous input signal transition. The result is plotted in figure 6.4a, resulting in $\beta|_{t_{p,H \rightarrow L} = t_{p,L \rightarrow H}} = 1,97$ for $V_{DDh} = 2,0\text{V}$.

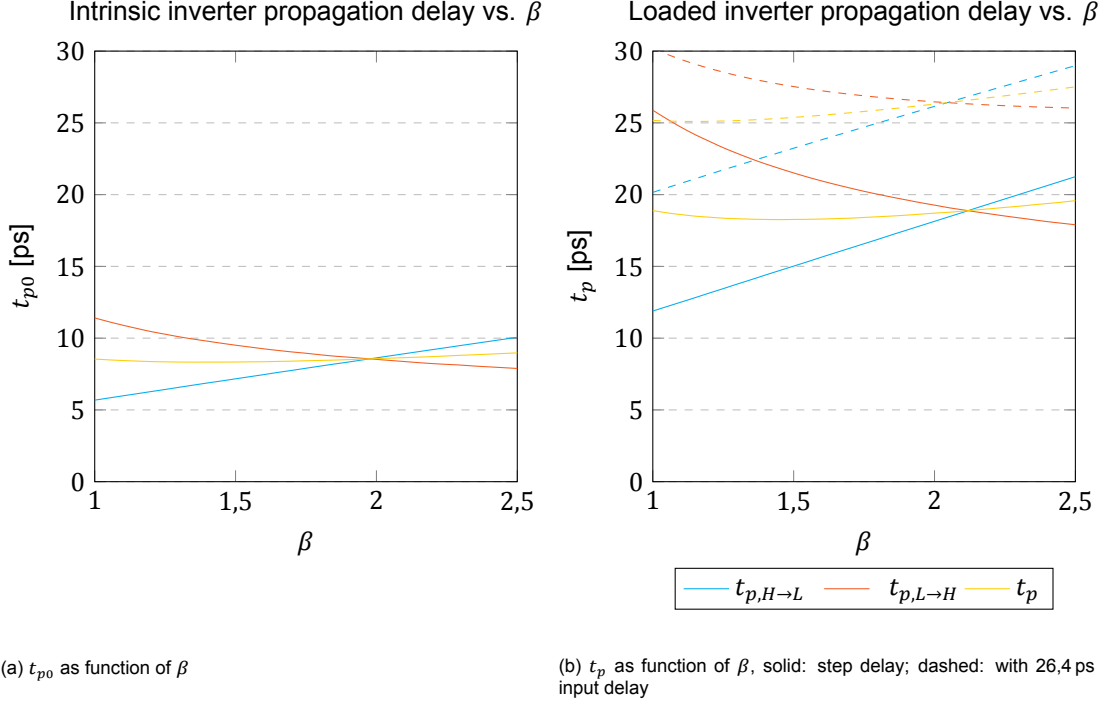


Figure 6.4: Thick oxide 270 nm CMOS propagation delays

For two chained equal sized inverters the output load of the first will be

$$C = C_{out} + C_{in} \quad (6.12)$$

$$= C_{out} (1 + \gamma^{-1}) \quad (6.13)$$

where γ the *self-loading factor*

$$\gamma = \frac{C_{out}}{C_{in}} \quad (6.14)$$

This allows to rewrite the propagation delay in terms of the intrinsic propagation delay

$$t_p = \ln(2)R_{es}(C_{out} + C_{in}) \quad (6.15)$$

$$= \ln(2)R_{es}C_{out}(1 + \gamma^{-1}) \quad (6.16)$$

$$= t_{p0}(1 + \gamma^{-1}) \quad (6.17)$$

The result is plotted in figure 6.4b, giving $\beta|_{t_{p,H \rightarrow L} = t_{p,L \rightarrow H}} = 2,12$ for $V_{DDh} = 2,0\text{V}$. The value for β changes due to the non-linearity of the capacitances and resistor. Adding an input propagation delay and setting as goal $t_p^i = t_p^{i-1}$ this resulted in a β of 2,05 and $t_p = 26,4$ ps. Then $t_{p0}|_{\beta=2,05} = 8,60$ ps and $t_{step}|_{\beta=2,05} = 18,8$ ps. This allows to determine the unknowns

$$\Lambda = \frac{t_p^i - t_{step}^i}{t_p^i} = \frac{26,4 - 18,8}{26,4} = 0,288 \quad (6.18)$$

$$\gamma = \left(\frac{t_p}{t_{p0}} - 1 \right)^{-1} = \left(\frac{18,8}{8,60} - 1 \right)^{-1} = 0,843 \quad (6.19)$$

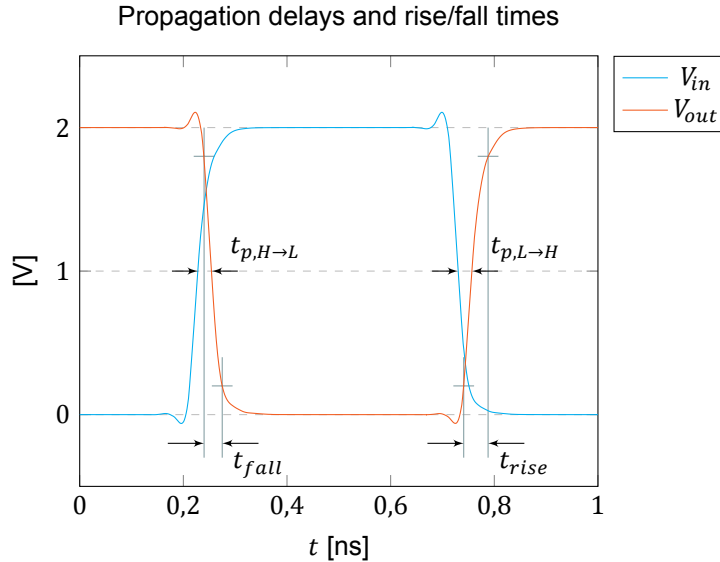


Figure 6.5: The relationship between propagation delays and rise and fall times

6.2.2. Designing the Inverter Chain

Most important is the design constant τ_d , the time constant of the RC network (see section 4.3.1). Its purpose in the design has mainly to do with the rise and fall times of the signal. The propagation delay as a metric itself is more important in digital designs. What its value exactly is does not matter here, as long as the two propagation delays are equal so the duty cycle d is constant throughout the inverter chain. Nonetheless, the propagation delay is closely related to the rise and fall times and τ_d .

$$\tau_d = R_{es}C \quad (6.20)$$

$$= R_{es}(C_{out} + C_{ESD1} + C_{ESD2} + C_{GS}) \quad (6.21)$$

This could lead to the conclusion that $t_p = \ln(2)\tau_d$, however R_{es} is very non-linear. For this application the rise and fall times are more important. The rise and fall times are defined here as the time required for the signal to transition from 10% to 90% of the final value. This gives

$$t_{rise/fall} = \ln(9)\tau_d \quad (6.22)$$

In figure 6.5 the transient simulation waveforms of the inverter are shown for the case where $V_{DDh} = 2,0\text{ V}$, $t_p^i = t_p^{i-1} = 26,4\text{ ps}$ and $\beta|_{t_{p,H \rightarrow L} = t_{p,L \rightarrow H}} = 2,05$. The propagation delays and rise and fall times are annotated in it. It results in $t_{fall} = 35,6\text{ ps}$ and $t_{rise} = 47,7\text{ ps}$. The rise time is most important for the Class-E operation, since it determines the speed of closing the switch. When opening the switch the voltage across it should be fairly close to 0 V anyway. Therefore calculation is continued with only the rise time. The ratio between the propagation delay and rise time is

$$\frac{t_{rise}}{t_p} = \frac{47,7}{26,4} = 1,81 \quad (6.23)$$

From table 5.8 it is known that τ_d should be 31 pF Ω , resulting in a wanted rise time and propagation delay of

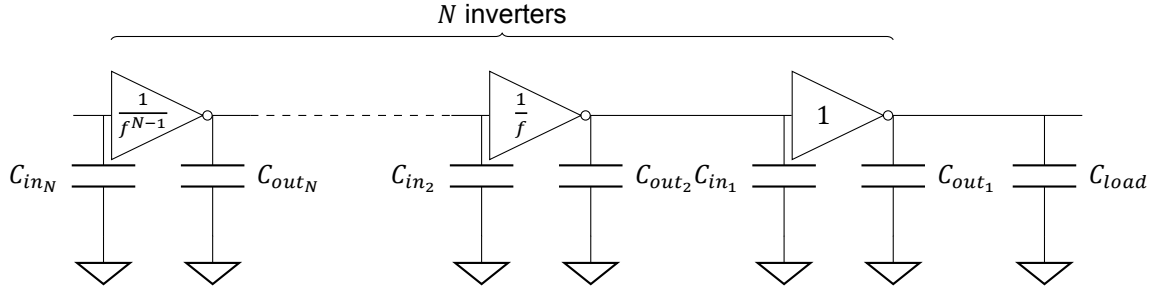
$$t_{rise,wanted} = \ln(9) \cdot 31 = 68,1\text{ ps} \quad (6.24)$$

$$t_{p,wanted} = \frac{t_{rise,wanted}}{1,81} = \frac{68,1}{1,81} = 37,7\text{ ps} \quad (6.25)$$

So an inverter chain should be designed with $t_p = 37,7\text{ ps}$.

The point of having an inverter chain is have every inverter connected to a larger inverter at its output, step by step increasing the capacitive load it can successfully drive. The *effective fanout* f is defined as the ratio between the capacitive load to the C_{in} of that inverter, meaning that

$$C = C_{out} + fC_{in} \quad (6.26)$$

Figure 6.6: An inverter chain of length N scaled by a factor f

which allows to rewrite equation (6.17) to

$$t_p = t_{p0}(1 + f\gamma^{-1}) \quad (6.27)$$

Including equation (6.11) and $t_p^i = t_p^{i-1}$ gives

$$t_p = t_{p0} \frac{1 + f\gamma^{-1}}{1 - \Lambda} \quad (6.28)$$

solving for f

$$f = \gamma \left(\frac{t_p}{t_{p0}} [1 - \Lambda] - 1 \right) \quad (6.29)$$

$$= 0,843 \left(\frac{37,7}{8,60} [1 - 0,288] - 1 \right) = 1,79 \quad (6.30)$$

that means that for one stage

$$\frac{C_{out}}{C_{load}} = \frac{1}{1,79 \cdot 0,843^{-1}} = 0,471 \quad (6.31)$$

C_{out} is 47 % of its connected load, increasing the power consumption with the same amount.

The number of required stages N is completely dependent of the load that needs to be driven and the size of the first inverter in the chain. The ratio between the input capacitance of the first inverter to the load that needs to be driven is called the *overall effective fanout* F . The largest load will be LDMOS transistor $ACW_{3,3}$ that has $C_{GS,3,3}$ 4,04 pF. Combined with the ESD capacitance this gives $C_{load} = 5,14$ pF. By integrating charge (similar to as done before in section 4.1) it is found that $C_{out} = 4,45$ fF for an inverter with W_{gn} is $1 \mu\text{m}$ and W_{gp} is $\beta \mu\text{m}$. Starting from a small sized inverter $W_{gn} = 480 \text{ nm}$ this means that

$$C_{in_N} = 0,48 \cdot \frac{4,45}{\gamma} = 2,53 \text{ fF} \quad (6.32)$$

so

$$F = \frac{C_{load}}{C_{in_N}} = \frac{5,14 \cdot 10^{-12}}{2,53 \cdot 10^{-15}} = 2,03 \cdot 10^3 \quad (6.33)$$

$$N = \log_f(F) = \log_{1,79}(2,03 \cdot 10^3) \approx 13 \quad (6.34)$$

meaning that 13 inverter stages are required with effective fanout $f = 1,79$. That means that the total capacitance that has to be charged and discharged throughout the entire chain is the geometric series

$$C_{tot} = C_{load} + \sum_{i=1}^N (C_{in_i} + C_{out_i}) \quad (6.35)$$

$$= C_{load} + (C_{in_1} + C_{out_1}) \sum_{i=0}^{N-1} \frac{1}{f^i} \quad (6.36)$$

$$= C_{load} + (C_{in_1} + C_{out_1}) \cdot \frac{1 - f^{-N}}{1 - f^{-1}} \quad (6.37)$$

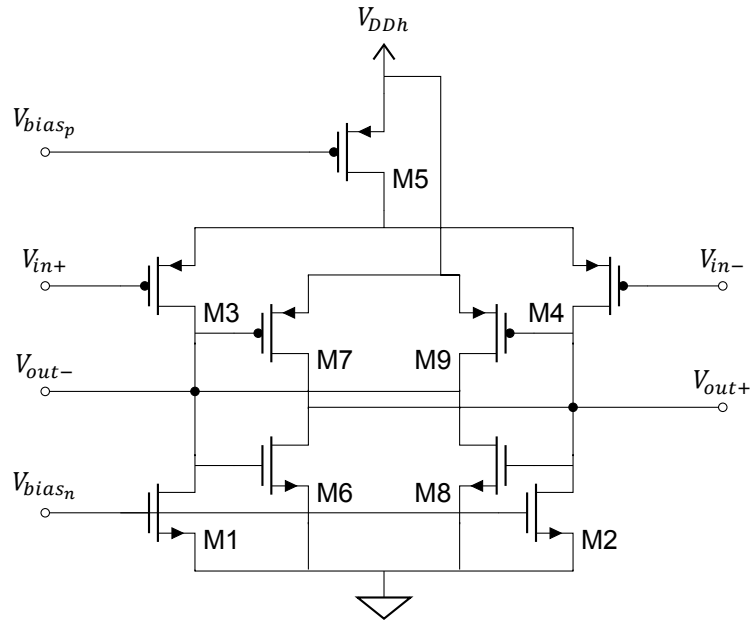


Figure 6.7: Designed OTA with enhanced output swing

The capacitance

$$C_{in_1} + C_{out_1} = \left(\frac{C_{out}}{\gamma C_{load}} + \frac{C_{out}}{C_{load}} \right) C_{load} \quad (6.38)$$

$$= (0,559 + 0,471) C_{load} = 1,03 C_{load} \quad (6.39)$$

and so

$$C_{tot} = C_{load} \left(1 + 1,03 \frac{1 - 1,79^{-13}}{1 - 1,79^{-1}} \right) \quad (6.40)$$

$$= C_{load} (1 + 1,03 \cdot 2,26) = 3,33 C_{load} \quad (6.41)$$

substituting this result in equation (4.12) gives

$$P_{drive} = 3,33 \cdot f_c \cdot C_{load} \cdot V_{DDh}^2 \quad (6.42)$$

combining that with equation (5.12) ($P_{in} = 2,04 \text{ W}$) gives peak $P_{drive} = 6,79 \text{ W}$. Using the information from table 5.8 a peak total system efficiency can be estimated by

$$\eta_{T,system} = \frac{P_{out,f_c}}{\frac{P_{out,f_c}}{\eta_D} + \sum_{i=0}^n P_{DC,i} + P_{in}} = \frac{16,6}{25,7 + 6,79} = 51,0\% \quad (6.43)$$

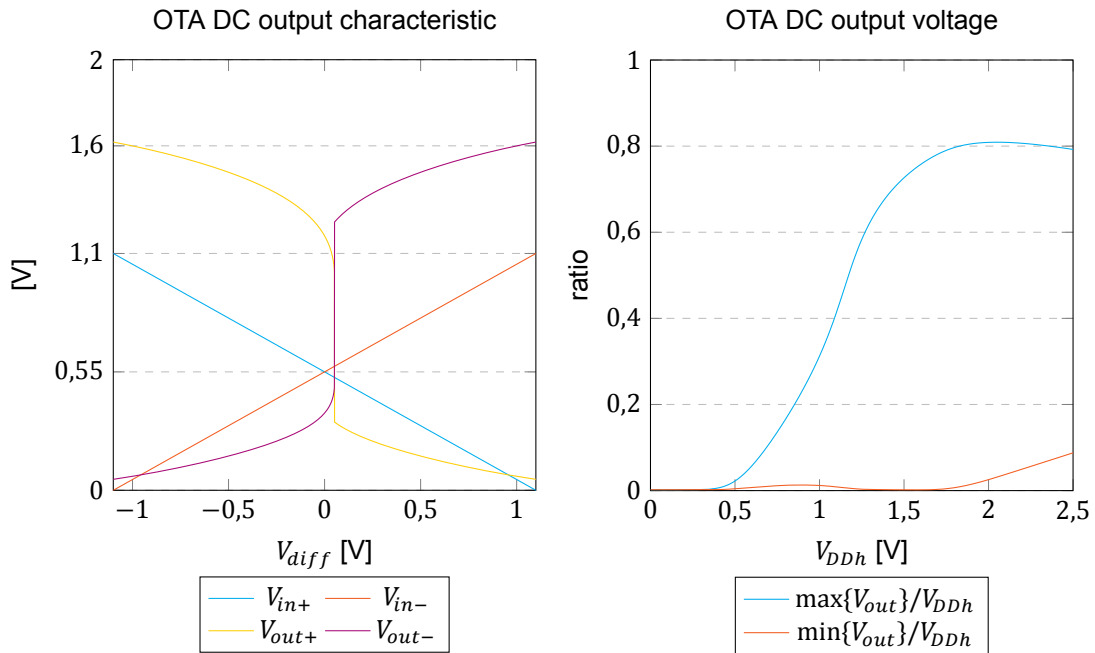
6.2.3. Levelshifter

The levelshifter is required to shift the digital logic levels from $V_{DDl} = 1,1 \text{ V}$ to $V_{DDh} = 2,0 \text{ V}$. V_{DDh} should remain tunable after production to calibrate the DRAC AM-AM curve, implying that a voltage gain of roughly 2 is required. Simply connecting the lower voltage levels to a high voltage inverter does not work, a higher voltage than $1,1 \text{ V}$ is required to exceed the switching threshold, especially if $V_{DDh} = 2,5 \text{ V}$. The high voltage inverter switching threshold is approximately $0,6 V_{DDh}$.

This calls for some biased circuit, which will be connected to an inverter: a capacitive load. Therefore the output of the biased circuit can be a current, so an Operational Transconductance Amplifier (OTA) is chosen for the task. By default an OTA has a differential input and may have a differential output. Differential input and outputs can be helpful to maintain a 50 % duty-cycle and minimize phase noise by adding phase aligners. The OTA should be able to operate with a square wave input at 3 GHz and therefore should be as simple as possible. It does not have to be linear.

The designed OTA is shown in figure 6.7. In transistors M1 to M5 a fairly standard PMOS differential input OTA can be recognized. As the outputs are connected through two PMOS transistors, its highest output voltage will be limited. Therefore transistors M7 and M9 are added to enhance the output swing, this however resulted in a positive bias of the output current. To compensate this M6 and M8 are added, forming two cross-coupled inverters to the output. The different transistor widths are given in the appendix in table A.2.

The DC characteristic as shown in figure 6.8a is obtained by connecting 1 M Ω resistors from the outputs to ground and V_{DDh} in order to convert the output current to a voltage. Connecting it to capacitors wouldn't make sense for a DC simulation, as this would effectively be an open circuit. There is an offset of 50 mV present at the input, which does not matter since it will normally operate only at the extremes of V_{diff} . These extremes as ratio to V_{DDh} are plotted in figure 6.8b. As the switching threshold is approximately $0,6V_{DDh}$, the levelshifter is expected to work for $1,3\text{ V} < V_{DDh} \leq 2,5\text{ V}$.



(a) OTA DC output characteristic for $V_{DDh} = 2,0\text{ V}$

(b) OTA DC output voltage extremes as ratio to V_{DDh}

Figure 6.8: OTA DC output characteristics

For verification the levelshifter is simulated using a transient simulation. The input is an ideal differential square wave, the outputs are loaded with inverters 1,2 times minimum size. The inverter outputs are connected to minimum size cross-coupled inverters, creating a phase aligner. In figure 6.9 these signals are shown in the time domain. As expected, the OTA output is not rail-to-rail, but enough to reach the inverter's switching threshold at the targeted maximum operation frequency of 3 GHz. After the phase aligner the output is rail-to-rail. The average supply current required for the OTA is 255 μA .

6.2.4. Complete Output Chain

The designed inverter chain is connected to the phase aligner after the levelshifter. The levelshifter is connected to a low voltage phase aligner to create a differential signal that can drive the OTA. The circuit diagram is shown in figure 6.10.

In figure 6.11 the transient simulation result is shown for the complete output chain. Examining the time V_{load} is high results in 168,6 ps, which means the duty-cycle $d = 0,977$. The propagation delay of the different components can be determined, which is 125 ps for the levelshifter and 508 ps for the inverter chain. This gives a total output chain delay of 633 ps, almost 2 RF cycles. Also the average consumed current is examined, giving $I_{DDl} = 74,7 \cdot 10^{-3}\text{ mA}$ and $I_{DDh} = 147\text{ mA}$. The peak current consumption occurs when V_{load} is rising and is $I_{DDh,pk} = 367\text{ mA}$.

If $V_{DDh} = 2,0\text{ V}$ it means a power consumption of 294 mW to drive 5,3 pF. Assuming this scales

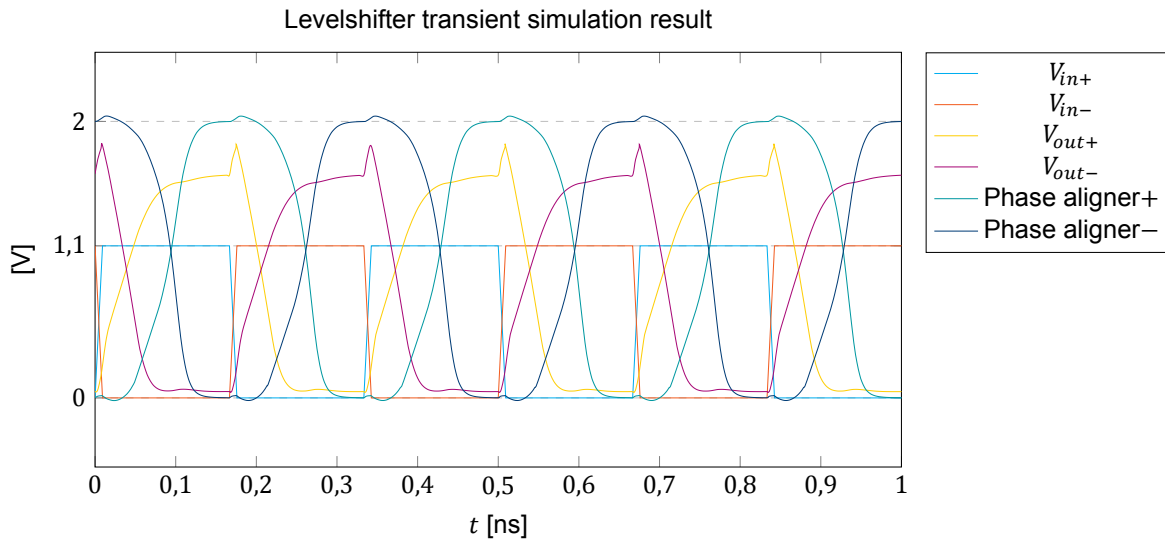


Figure 6.9: OTA transient waveforms for $V_{DDh} = 2,0$ V

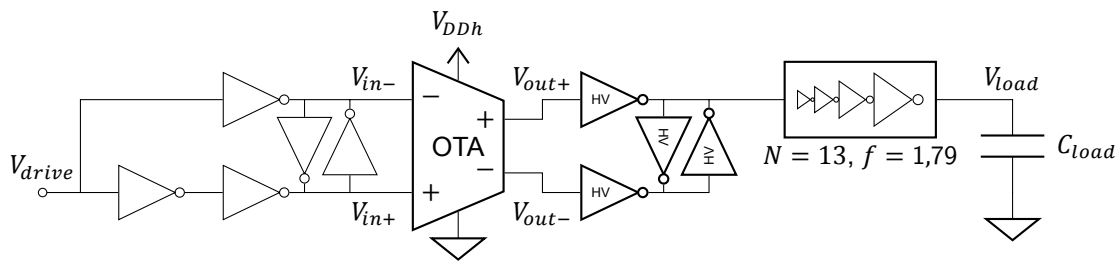


Figure 6.10: Complete output chain circuit diagram

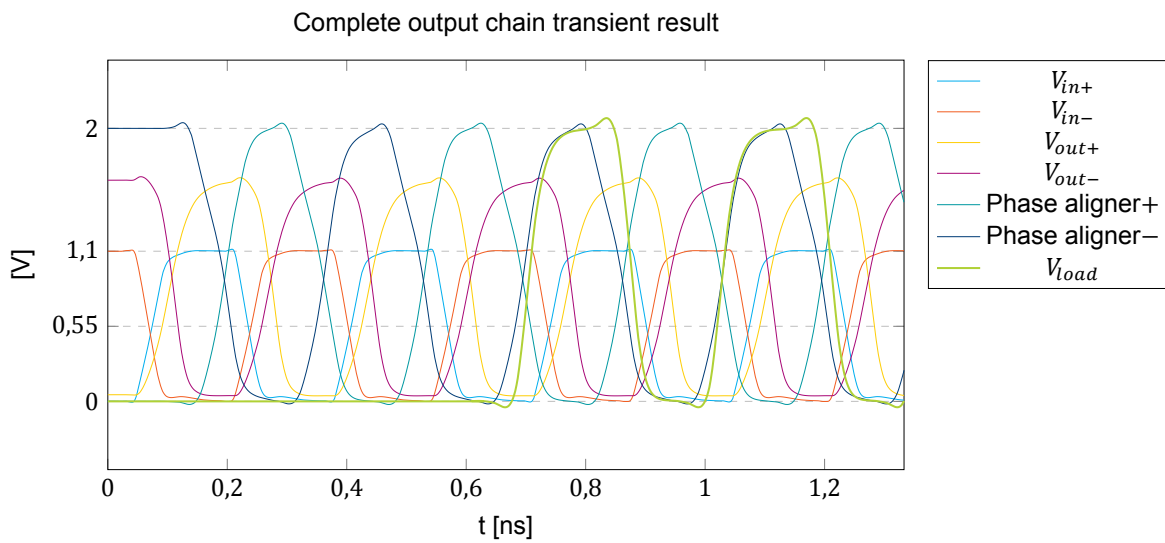


Figure 6.11: Complete output chain transient simulation

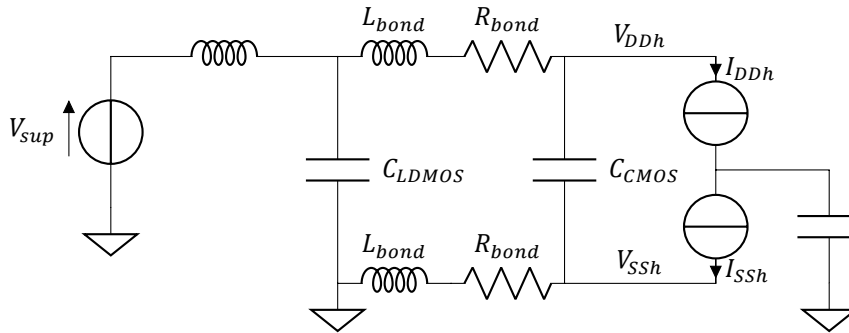


Figure 6.12: Simplified supply decoupling circuit diagram

linearly with the capacitance to be driven and that it is known from equation (5.12) that the total capacitance to be driven is 170,3 pF, the total required drive power P_{drive} lands at 9,45 W, a factor 4,63 times the theoretical minimum for driving the LDMOS gates and ESD capacitances. This is more than the predicted factor 3,33 from section 6.2.2. First, this could be explained by the largest transistors having relatively more parasitic capacitance from interconnect and additional resistive losses, requiring them to be slightly larger and have a lower effective fanout f than initially predicted. This was taken into account with the actual simulated design by using the CMOS RF transistor models that incorporate more layout parasitics for the final 5 inverter stages, instead of the default inverter from the transistor model library. Secondly, 9,45 W could be an overestimate as the assumption that the drive power scales linearly with the capacitance to be driven may be an overestimate itself. The smaller transistors will have insignificant capacitance compared to the ESD capacitance, making $C_{load} \approx 1,1$ pF. An inverter chain capable of driving that will require 2 to 3 stages less.

The simulation results give an idea what the requirements for the programmable delay-line will be. From table 5.10 it is estimated that for multiphase RF clocking a total delay of 18 ps is required. More than that is required for compensating the difference in inverter chain propagation delay. Assuming 3 inverter stages less are required for the smallest values of C_{load} compared to the largest, combined with t_p designed to be 37,7 ps, it gives a total delay requirement of 131,1 ps. This is quite a lot, therefore it is suggested to have every inverter chain the same length by adding extra (minimum sized) inverters at the beginning of the chain having only $f = 1$. From that is known that $t_p = 26,4$ ps, reducing the estimated maximum required delay to 51,9 ps.

6.3. Supply Decoupling

It is important to know whether the power supply for the CMOS can keep up with the required drive power. That is why the requirements for supply decoupling are investigated. From transient simulations done in section 6.2.4 it is known that the peak supply current of one output chain is approximately 2,5 times the average supply current. The currents for all output chains will constructively add together, any multiphase RF clocking will have negligible influence on that. Therefore it is safe to assume the overall peak-to-average CMOS supply current will also be 2,5.

In the layout there are 29 flip-chip connections dedicated for V_{DDh} and 30 for the ground of the same power supply domain. Each connection will therefore carry an average current of 163 mA and thus a peak current of 400 mA. The current peak will occur periodically with a frequency equal to f_c , which is 3 GHz maximum. Fourier analysis shows the second harmonic is a dominant component in the supply current, which is caused by the current peak.

Figure 6.12 shows the simplified circuit schematic which is used to investigate the requirements on the supply decoupling. In the simulation the S-parameter model is used for the flip-chip bond, not the lumped equivalents L_{bond} and R_{bond} . The CMOS power supply, as well as its ground, have to come from flip-chip connections. The current I_{DDh} is given the same harmonic content as the simulated CMOS output chain current. The current I_{SSh} is similar to I_{DDh} , but time shifted half a period, since it is associated with the discharging of earlier charged capacitances.

A goal is set to have a voltage amplitude $< 1\%$ of the DC value on any frequency. When the inductance from the supply to the LDMOS is 2 nH a value for $C_{LDMOS} \geq 0,03$ nF per pair of flip-chip connections helps reducing the amplitude of the fundamental and third harmonic. Higher values have

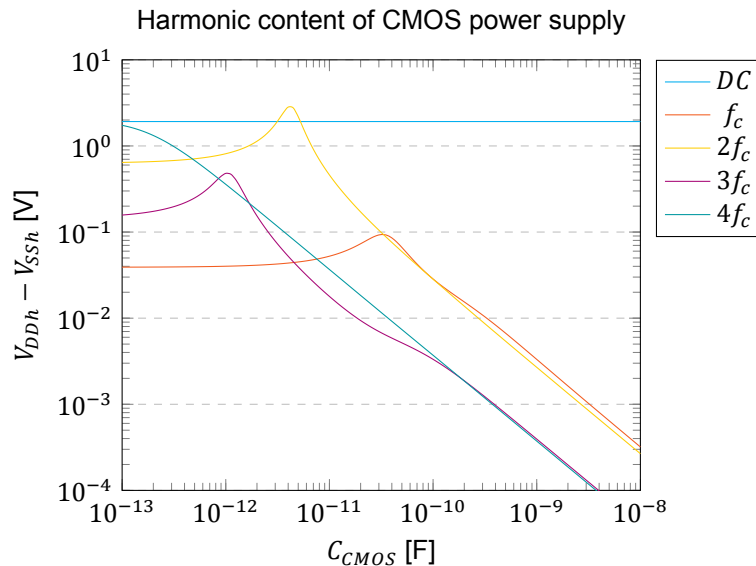


Figure 6.13: Harmonic content of CMOS power supply voltage vs. C_{CMOS} for $C_{LDMOS} = 0,03$ nF

negligible further impact, lower values can be compensated by more capacitance on the LDMOS. In figure 6.13 the harmonic content of the CMOS voltage supply is shown as function of C_{CMOS} per pair of flip-chip connections for $C_{LDMOS} = 0,03$ nF. An IR drop of 8 mV can be observed for the DC value, which is uncompensable with decoupling capacitance. When $C_{CMOS} \geq 0,154$ nF per pair of flip-chip connections the stated 1% condition is met.

On the CMOS layout there is an area of $1,45 \cdot 10^6 \mu\text{m}^2$ inside the ESD ring available to place active components. For the output buffers it is estimated that in the worst case $0,42 \cdot 10^6 \mu\text{m}^2$ is required. This leaves $1,03 \cdot 10^6 \mu\text{m}^2$ available to place MOS capacitors. For it to have a total capacitance of $29C_{CMOS} \geq 4,47$ nF a capacitive density of at least $4,34$ fF μm^{-2} is required. As mentioned in section 5.3 the capacitive density for a combination of MOS and MOM capacitors is $3,6$ fF μm^{-2} to $5,0$ fF μm^{-2} . This range is mostly due to the non-linearity of the MOS capacitor with applied voltage bias, for example its capacitive density increases approximately fourfold when increasing V_{GS} from 0V to 1V. Increasing it further increases the capacitive density only slightly, so it is safe to assume it will operate fairly close to 5 fF μm^{-2} . Therefore there is enough area available to place the required supply decoupling capacitance.

6.4. CMOS Driver Implementation Ideas

These ideas and their effects are not worked out in detail, but are presented here anyhow for future reference.

The “extra” transistors $ACW_{i,9'}$ require relatively simple control logic to generate its control signals (see section 4.4.1 and equation (4.26)). They are sized according to the scaling factors of the different segments, which could allow these transistors to function as some sort of ‘repair bits’ that can compensate post-production INL errors. For example, some of the negative DNL spikes that were observed in figure 5.30a could be compensated by instead having e.g. $ACW_{3,9'}$ on, it can be swapped out for the slightly larger $ACW_{4,9'}$. There is space in the layout for an additional transistor $ACW_{1,10}$ that can be added, depending on how small the LDMOS devices can be fabricated. This particular transistor would have a width $W_{1,10}$ of $10 \mu\text{m}$. It would serve as $\frac{1}{2}$ LSB that could help improve the accuracy of the linearity compensation. It is not intended to make it a 10 bit DRAC, rather to improve the Effective Number Of Bits (ENOB). This could be coupled to some programmable Look-Up Table (LUT) that specifies which of these 6 transistors should be activated for what values of ACW, effectively improving the AM-AM linearity.

There are two push-pull banks available, that will have their own phase modulator and baseband data. By shifting the upsampled baseband amplitude clocks f_{rs} by half a period this would result in a ‘free’ 2-fold version of L-fold linear interpolation of the baseband data [39]. This will effectively add half

a bit of resolution, but it has to be paid for in a higher HD2 since it will disturb the push-pull operation of the DRAC. However, this would make it possible to increase the limit of the upsampling clock frequency f_{rs} to twice the RF carrier $2f_c$, with linear interpolation.

A last idea is to synchronize the ACW data to the phase modulated clock, by replacing the AND-gate mixer by a D flip-flop. This will introduce an intentional synchronization error with an average related to the phase modulation index, which will cause non-linearity. However, this will mitigate possible problems with glitches. The non-linearity caused by the AM-PM data desynchronization might actually be smaller than non-linearities caused by possible switching glitches within a RF cycle.

6.5. System Performance Estimation

The estimated total peak power for the complete output chain is 9,45 W, excluding the delay-lines. From design experience in the ELCA group it is known that five programmable delay-lines require approximately 7 mW at 2,5 GHz. This means 0,56 mW GHz⁻¹ per delay-line. For this design there are 86 delay-lines required, at 3 GHz this gives an estimated peak power consumption of 144 mW for the delay lines. This gets the total estimated peak power consumption of the complete output chain at 9,59 W. It purposely is referred to as *peak* power consumption, as at lower required output voltages of the DRAC some delay-lines and inverter chains are not required to do anything as some of the LDMOS transistors are supposed to stay off.

Aside from power consumptions that scale with DRAC output voltage, there are also components that will continuously consume the same power. This includes the memories, the DSP (COordinate Rotation Digital Computer (CORDIC) and upsampling) and the phase modulators. From another design it is known that the DSP consumes 60 mW. This other design is a Doherty topology, so it also includes two transistor banks. The phase modulator in this design requires 50 mW at $4 \times 2,5$ GHz = 10 GHz. As in this design two separate phase modulators at $4 \times 3,0$ GHz = 12 GHz will be required this gives 120 mW. This makes the estimated non-scaling power consumption to be 180 mW.

The total peak power consumption of the entire CMOS driver can now be estimated at $P_{drive} = 9,77$ W. With the DRAC DC power consumption $P_{DC} = 25,7$ W and output power $P_{out,fc} = 16,6$ W this gives an estimated peak system efficiency of

$$\eta_{T,system} = \frac{16,6}{25,7 + 9,77} = 46,7 \% \quad (6.44)$$

This also allows to estimate average system efficiency by assuming a probability (density) function for the amplitude of a modulated signal. In section 4.2 QAM256 modulation was assumed with a PAPR of 8,2 dB. The probability function for the amplitude of a random QAM256 modulated signal is plotted in figure 6.14, together with the simulated DRAC powers and estimated total CMOS powers. By applying the probability function as window to the different powers and integrating over all possible values of k the average values can be obtained

$$\begin{aligned} \overline{P_{out,fc}} &= 2,61 \text{ W} \\ \overline{P_{DC}} &= 9,07 \text{ W} \\ \overline{P_{in}} &= 0,581 \text{ W} \\ \overline{P_{drive}} &= 2,91 \text{ W} \\ \overline{PAPR} &= 8,03 \text{ dB} \end{aligned} \quad \begin{aligned} \overline{\eta_D} &= 28,8 \% \\ \overline{\eta_T} &= 27,0 \% \\ \overline{\eta_{T,system}} &= 21,8 \% \end{aligned} \quad (6.45)$$

where P_{in} is the required DRAC input power including ESD protection and P_{drive} the complete CMOS driver power consumption. The PAPR is slightly lower since a random QAM256 signal of finite length was used, which also causes the non-smoothness of the probability curve in figure 6.14.

A special point in figure 6.14 is for $k = 0$, which has up to now received little special attention. Almost all powers become equal to 0 W at that point, with exception of the CMOS driver which will consume the estimated 180 mW. As mentioned in the introduction base stations are mostly idle during the night or in rural areas. The idle power consumption of the designed power RF-DAC is low compared to a 'traditional' analog intensive approach. Assuming an ideal Class-B PA with the same output power as this system, a gain of $G_{P,dB} = 15$ dB and an ideal Class-A driver, this driver will always consume 1,05 W, no matter the output power of the system.

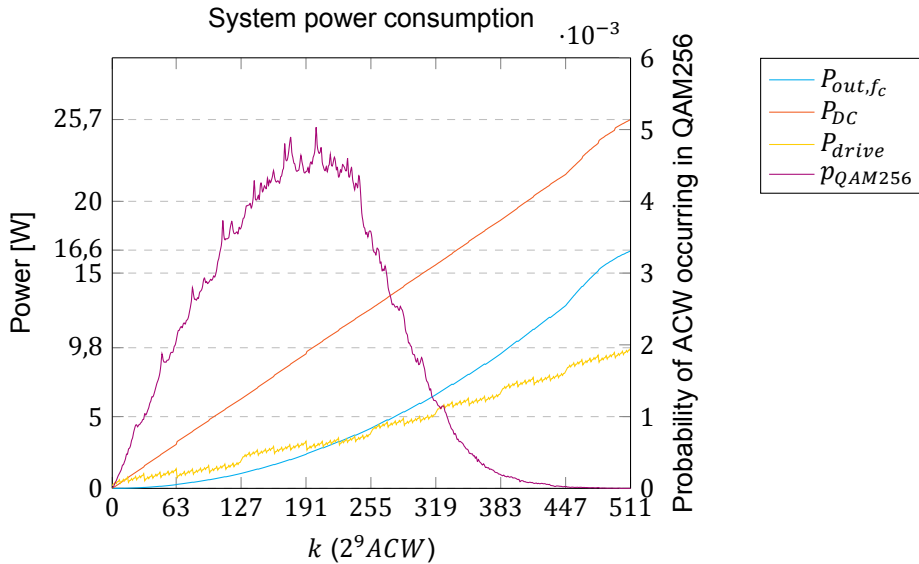


Figure 6.14: DRAC output power and DC power, CMOS driver power and QAM256 output probability vs. ACW

Assuming the same QAM256 probability function, the ideal Class-B with Class-A driver system would have an average drain efficiency of $\overline{\eta_D} = 33,6\%$ and an average total efficiency of $\overline{\eta_T} = 29,5\%$. When the PA is idle 80% of the time the designed power RF-DAC will outperform the ideal analog approach: the RF-DAC will then have an average total system efficiency of $\overline{\eta_{T,system}} = 20,5\%$ whereas the ideal analog case would have an average total efficiency $\overline{\eta_T} = 19,9\%$.

When the final analog stage is not ideal or a Class-AB PA would be used, the average efficiency will quickly drop. Usually in base station applications the average efficiency is much higher since efficiency enhancement techniques are employed, such as Doherty topologies. But has been demonstrated that it is also possible to apply a Doherty topology to a (low power) RF-DAC operating in Class-E, as done in [40]. Therefore it is expected that this can also be applied to the power RF-DAC.

6.5.1. Estimation of Future System Performance

When it comes to complete system performance the designed power RF-DAC has a major disadvantage: it needs an external CMOS driver. First of all, this requires ESD protection, which already takes up more than half of the required input power. Secondly, the LDMOS devices are underutilized in terms of their power handling capabilities as they are only driven with a 2V square wave. For that also the V_T -shift is required, from which was known it would increase C_{GS} to $1,50\text{ pF mm}^{-1}$. However, whilst this thesis was being written Ampleon indicated that it also possible to create the same V_T -shift *without* increasing C_{GS} , which then is $0,860\text{ pF mm}^{-1}$. Just incorporating this while keeping all other scaling factors the same the peak required CMOS driver power lowers to 7,99 W.

A next step would be to get rid of the ESD capacitances somehow, preferably by integrating high-speed CMOS transistors in the LDMOS process. When assuming these CMOS transistors will have similar properties as the 270 nm thick oxide devices as available in the TSMC 40 nm process, then again all scaling factors remain the same. Only 1,1 pF of ESD capacitance is removed per connection. Still assuming no C_{GS} increase due to V_T -shift, this then results in a peak required CMOS driver power of 2,73 W. This will not increase the peak nor the average drain efficiencies, but this definitely increases the total (chain) efficiencies. The required peak input power P_{in} just to drive the LDMOS gates is now only 0,520 W. If it is possible to reduce the found multiplier of 4,63 in section 6.2.4 to the theoretical value of 3,33, while still including the delay-lines and idle consumption, this yields a peak total CMOS driver power of only 2,05 W.

If the CMOS drive transistors can be integrated in the LDMOS it would be preferred if they have a breakdown voltage V_{bk} higher than the 2,5V of the TSMC 40 nm process. If $V_{bk} > 6\text{ V}$ is supported the DRAC output power will significantly increase, as well as its drain efficiency. In appendix B DRAC simulations are shown using an input voltage V_{in} of 6V (square wave, peak-to-peak). It results in $P_{out,fc} = 55,9\text{ W}$ with $\eta_D = 79,0\%$. The required input power P_{in} grows quadratically with voltage (see

Table 6.1: System power consumption contributor and efficiency comparison at 3 GHz

		C_{GS}	$V_{in} = 2\text{ V}$		$V_{in} = 6\text{ V}^*$	
			1,50 pF mm ⁻¹	0,860 pF mm ⁻¹	0,860 pF mm ⁻¹	
Power	DRAC	$P_{out,fc}$	16,56 W		55,90 W	
		P_{DC}	25,70 W		70,77 W	
	Intrinsic LDMOS Gate		0,907 W	0,520 W	4,64 W	
	2×43 ESD Capacitance (1,1 pF)		1,135 W		—	
	CMOS Output Chain	×3,33		6,780 W	5,511 W	15,46 W
		×4,63		9,453 W	7,662 W	21,50 W
	2×43 Delay-lines 12 GHz PM + DSP			0,144 W		0,180 W
Total	CMOS ×3,33	w/o ESD	3,34 W	2,05 W	15,8 W	
		w ESD	7,12 W	5,83 W	—	
	CMOS ×4,63	w/o ESD	4,52 W	2,73 W	21,8 W	
		w ESD	9,78 W	7,99 W	—	
Peak efficiency	η_D		64,4 %		79,0 %	
	η_T	w/o ESD	62,2 %	63,2 %	74,1 %	
		w ESD	59,7 %	60,5 %	—	
	$\eta_{T,system} (\times 4,63)$	w/o ESD	54,8 %	58,2 %	60,4 %	
		w ESD	46,7 %	49,2 %	—	
Avg efficiency	$\overline{\eta_D}$		28,8 %		33,8 %	
	$\overline{\eta_T}$	w/o ESD	27,9 %	28,3 %	31,9 %	
		w ESD	27,0 %	27,4 %	—	
	$\overline{\eta_{T,system} (\times 4,63)}$	w/o ESD	24,6 %	26,0 %	26,6 %	
		w ESD	21,8 %	22,8 %	—	

*Based on pre-layout simulations

equation (4.12)), resulting in a required peak input power P_{in} for the LDMOS gates of 4,64 W. Assuming similar values are kept for Λ , γ and t_{p0} —so f stays the same—the same multiplication factors can be used to estimate total drive power P_{drive} . Using the factor 4,63 it results in a peak total system efficiency of 60,4 % and an average total system efficiency of 26,6 %. Using the factor 3,33 this even becomes 64,6 % and 28,2 % respectively.

All these values are also presented in table 6.1.

6.6. Conclusion

Requirements for the CMOS driver are established that result in a design for the output buffer. This design consists of a levelshifter that is able to change the logic levels from $V_{DDl} = 1,1 \text{ V}$ to $1,3 \text{ V} \leq V_{DDh} \leq 2,5 \text{ V}$. To drive a load of approximately 5,3 pF—with quick enough rise time to let the LDMOS DRAC operate efficiently at $f_c = 3 \text{ GHz}$ —a 13-stage inverter chain is required with each an effective fanout of $f \approx 1,8$. A conservative estimate is made that all output chains together at $f_c = 3 \text{ GHz}$ result in a power consumption of $P_{drive} = 9,45 \text{ W}$. Assuming this power consumption it was shown that the CMOS layout provides for enough available area to place an appropriate amount of decoupling capacitance to have a voltage amplitude $< 1 \%$ of the DC value on any frequency of the CMOS voltage supply.

Adding estimates for the power consumption of the programmable delay-lines, phase modulators and DSP (CORDIC and upsampling) the total peak power consumption at $f_c = 3 \text{ GHz}$ of the entire CMOS driver is estimated at 9,77 W, giving a peak system efficiency of $\eta_{T,system} = 46,7 \%$. By applying a probability curve based on QAM256 modulation with $PAPR = 8,03 \text{ dB}$ an average system efficiency is estimated at $\overline{\eta_{T,system}} = 21,8 \%$. An advantage of the designed power RF-DAC is the relatively low idle power consumption. This makes it more efficient than an ideal Class-B PA with Class-A driver if the PA would be idle for 80 % of the time, assuming no special energy enhancement techniques such as Doherty topologies.

Finally estimates are made what the different system efficiencies could be if the technology limitations, that now result in an external CMOS driver for the DRAC, are relieved by future technological development. Keeping the conservative scaling factors this would result in a peak system efficiency of $\eta_{T,system} = 60,4 \%$, and an average system efficiency of $\overline{\eta_{T,system}} = 26,6 \%$ with the PAPR of 8,0 dB without the energy enhancement techniques.

7

Conclusion

In this thesis the design for a power RF-DAC was elaborated. The key point was power efficiency, therefore the choice was made to design a polar RF-DAC. Since the RF-DAC also has to function as a PA, first the fundamentals of PAs were discussed and their operation principles explained. It was seen that high power efficiency inherently requires a non-linear PA. The group of Switch-Mode Power Amplifier (SMPA) classes have a theoretical peak drain efficiency of 100 %, making them the ideal candidate for the power RF-DAC. In particular Class-E SMPAs are suited for RF designs as this class includes the parasitic output capacitance of the active device.

A RF-DAC operates in principle very similar to any other Nyquist rate DAC, but has its output signal multiplied by a square wave at the carrier frequency. The topologies for DACs are also applicable to RF-DACs. The most linear topology would have been a thermometer-coded topology, but that would require too much circuit elements. The least amount of circuit elements would have been used in a binary-weighted topology, however that would result in poor linearity and possibly glitches. Therefore a hybrid solution was preferred.

The core component of the power RF-DAC is the power DRAC. In order to make it a *power* DRAC, a semiconductor technology has to be used that can provide these higher powers. Two possible semiconductor technologies were evaluated, Fraunhofer IAF GaN25 and Ampleon LDMOS LM8. Both technologies are specialized in high power analog applications, lacking the possibility of integrating high-speed digital logic. However, this high-speed digital logic could be found in advanced CMOS technology, for which TSMC 40 nm CMOS technology is available. The RF-DAC principle requires high-speed digital logic, so in order to design a power RF-DAC the combination of two chips was proposed to combine the benefits of high power analog electronics and high-speed digital electronics: a LDMOS Class-E SMPA DRAC with a CMOS driver.

The combination of two chips did imply some limitations. For the power DRAC the choice of semiconductor technology got restricted to the LDMOS technology from Ampleon as only it can be made compatible to the voltage levels that the CMOS technology can provide (being $V_{DDh} \leq 2,5 \text{ V}$) by means of a V_T -shift. Also an interface between the LDMOS and CMOS chip is required. For this interface a flip-chip solution was chosen since it provides minimum negative side-effects, primarily having the lowest inductance between the two chips of the possible chip interfaces and providing the highest number of connections between the chips. This flip-chip connection limits the maximum operational frequency to roughly 3 GHz. Both chips require ESD protection diodes, causing 1,1 pF of additional capacitance for each flip-chip connection. This more than doubled the total capacitance to be driven by the CMOS. This meant also more than doubling the required input power, as the required input power is directly proportional to the total capacitance.

To comply with the requirement of ACPR levels better than 50 dBc, it was determined that an effective resolution of $N_b = 7,9$ bit was required. It was chosen to implement the DRAC with a resolution of 9 bit and a push-pull topology to cancel even orders of harmonic distortion. A LDMOS transistor Total Gate Width of 51,0 mm was divided over the two push-pull banks, each with 8 thermometer-coded segments as MSB and 6 bit binary-weighted LSBs within the segments. By scaling the segments the AM-AM curve could be linearized in a piecewise fashion. The found Class-E design set for maximum drain efficiency required a relatively high load resistance in order for the LDMOS transistors to enter

the triode region. The transistors operated mostly in the current limited region, resulting in a relatively linear baseline AM-AM curve. Therefore the choice was made to merge the first segments to save flip-chip connections, now resulting in 6 thermometer-coded segments. This amounts to a total of 43 individual transistors and flip-chip connections for each push-pull bank.

A LDMOS DRAC design was proposed that is compatible with the design rules regarding the ESD protection requirements of the CMOS driver. A post-layout simulation of the DRAC was performed, which required a slightly larger load resistance due to increased parallel capacitance caused by the drain bar. This resulted at a frequency of 3 GHz in a peak fundamental output power $P_{out,fc} = 16,56$ W and drain efficiency $\eta_D = 64,4$ %. The required peak input power is $P_{in} = 2,13$ W, making the peak total efficiency $\eta_T = 59,5$ %. The last segment caused the peak INL to be $INL_{pk,LSB} = 10,7$ LSB, the other segments stayed within $INL_{LSB} = 4,35$ LSB. This behavior of the last segment is according to theoretical expectation. It was proposed to include programmable delay lines on the CMOS driver to prevent glitches and to use it for multiphase RF clocking. Depending on the resolution of the programmable delay lines the AM-PM deviation can be within 2° .

Lastly the viability of a CMOS driver design is evaluated by designing a complete output buffer chain for the largest LDMOS transistor present in the DRAC design. This output buffer chain consists of a levelshifter that converts the signal from $V_{DDl} = 1,1$ V to $1,3$ V $\leq V_{DDh} \leq 2,5$ V and 13 stage inverter chain. Extrapolating this result and including power consumption estimates for the programmable delay-lines, two phase modulators and DSP (CORDIC and upsampling) the total peak power consumption of the CMOS driver is estimated at 9,77 W when operating at 3 GHz, giving a peak system efficiency of $\eta_{T,system} = 46,7$ %. Applying a probability curve based on QAM256 modulation with $PAPR = 8,03$ dB an average system efficiency is estimated at $\overline{\eta_{T,system}} = 21,8$ %. An advantage of the designed power RF-DAC is the relatively low idle power consumption.

7.1. Discussion

With the simulation results from chapters 5 and 6 an estimate of system performance is made. It does assume an ideal output matching network, aside from the DRAC layout effects. Nonetheless it enables a comparison in terms of power efficiency with previous work, which is shown in table 7.1. The table includes four other RF-DAC designs in CMOS and two PA designs in GaN. For the sake of comparison also the estimated future system performance is included, as elaborated on in section 6.5.1. All designs operate in Class-E, with exception of the last which operates in Class-D⁻¹. The first to notice is that this work has much higher output power than the other RF-DAC designs, which was one of the primary goals of this thesis project. The estimated power efficiencies from this design are close to the reported RF-DAC efficiencies from literature, indicating that the estimate is not ridiculously far off from what can be expected based on other Class-E designs.

When comparing it to the GaN PAs the simulated drain efficiency is lower. From the technology characterization done in section 4.1 it was determined that GaN technology would most likely result in better Class-E performance, so it is no surprise that the reported GaN designs perform better in terms of power efficiency, even though a different GaN technology was used than the one characterized in this thesis. The PAs, however, do not include a RF-DAC and require a higher drive voltage up to $V_{DD} = 6$ V which is currently not possible with the given CMOS 40 nm technology.

It is interesting, however, to take the information from these papers on the required drive power when using 6 V. A 9,6 mm wide GaN device is used in [37] with $C_{DS} = 3,6$ pF. From the GaN technology used is known that the input capacitance is approximately 3,5 times the output capacitance, so the input capacitance can be estimated at $C_{GS} \approx 12,5$ pF. From the efficiency figures can be calculated that a drive power of 3,9 W is required at 2,14 GHz. The LDMOS design has 50 mm of gate, resulting in $C_{GS} = 43$ pF, extrapolating the driver power consumption gives a drive power of 18,9 W for the LDMOS at 3 GHz, which is not too dissimilar from the estimated 21,8 W from section 6.5.1 and table 6.1. In fact, this might indicate that this value is even an overestimate of what could be achievable. It is important to note that for the presented power RF-DAC design the required amount of input power required will scale linearly with wanted output voltage $V_{out,fc}$ and that the CMOS driver efficiency will be constant. This is not the case for [28, 37] where required input power will remain constant, even when the wanted output power is low.

The different papers also mention some linearity specifications, such as ACPR and EVM. Unfortunately, the performance of power RF-DAC is a prediction based on simulations. It was not possible to

Table 7.1: Simulation based power efficiency comparison with measurements from literature

architecture RF-DAC technology	this work simulated		future work simulated		[33] measured		[34] measured		[37] measured		[28] measured		[40] measured		[41] measured	
	Polar Yes	LDMOS + 40 nm CMOS ^a	Polar	LDMOS + CMOS ^b	Polar Yes	40 nm CMOS	Polar Yes	40 nm CMOS	SMPA No	GaN + 65 nm CMOS ^c	Outphasing SMPA No	GaN + 65 nm CMOS ^c	Polar Doherty Yes	40 nm CMOS	Polar Yes	65 nm CMOS
f_c [GHz]	3,0	3,0	3,0	3,0	2,0	2,0	2,6	2,6	2,14	2,14	1,95	1,95	2,5	2,5	2,25	2,25
V_{DC} (+ V_{DD}) [V]	25,0 + 2,0	25,0 + 6,0 ^b	25,0 + 6,0 ^b	25,0 + 6,0 ^b	0,5	0,5	0,7	0,7	32,0 + 6,0	32,0 + 6,0	28,0 + 5,0	28,0 + 5,0	0,7 + 1,1	0,7 + 1,1	1,0	1,0
peak P_{out,f_c} [W]	16,6	56 ^b	28,8 · 10 ⁻³	56 ^b	28,8 · 10 ⁻³	28,8 · 10 ⁻³	52,5 · 10 ⁻³	52,5 · 10 ⁻³	28	28	19	19	138 · 10 ⁻³	138 · 10 ⁻³	158 · 10 ⁻³	158 · 10 ⁻³
peak η_D [%]	64,4	79 ^b	43,8	79 ^b	43,8	43,8	67	67	85	85	73 ^e	73 ^e	49,4	49,4	44	44
peak η_T [%]	59,5	74 ^b	38,1 ^d	74 ^b	38,1 ^d	38,1 ^d	54,9 ^d	54,9 ^d	76 ^d	76 ^d	68 ^e	68 ^e	n.a.	n.a.	37 ^e	37 ^e
peak $\eta_{T,system}$ [%]	47,1	60 ^b	38,1 ^d	60 ^b	38,1 ^d	38,1 ^d	54,9 ^d	54,9 ^d	76 ^d	76 ^d	68 ^e	68 ^e	n.a.	n.a.	37 ^e	37 ^e
PAPR [dB]	8,0 ^f	8,0 ^f	8,5	8,0 ^f	8,5	8,5	6,0	6,0	—	—	7,5	7,5	7,5	7,5	8,0	8,0
$\overline{\eta_D}$ [%]	25,8 ^{f,g}	34 ^{b,f,g}	17,5 ^g	34 ^{b,f,g}	17,5 ^g	17,5 ^g	n.a.	n.a.	—	—	65,1	65,1	24	24	18 ^{g,h}	18 ^{g,h}
$\overline{\eta_T}$ [%]	24,4 ^{f,g}	32 ^{b,f,g}	17,0 ^{d,g}	32 ^{b,f,g}	17,0 ^{d,g}	17,0 ^{d,g}	n.a.	n.a.	—	—	51,6	51,6	n.a.	n.a.	—	—
$\overline{\eta_{T,system}}$ [%]	20,2 ^{f,g}	27 ^{b,f,g}	17,0 ^{d,g}	27 ^{b,f,g}	17,0 ^{d,g}	17,0 ^{d,g}	n.a.	n.a.	—	—	41,9	41,9	n.a.	n.a.	—	—

^aThick oxide 270 nm CMOS^bCMOS embedded in LDMOS technology to be developed, results based on pre-layout simulation and calculation^cThin oxide 65 nm Extended-Drain MOS (EDMOS)^dConverted from PAE^eEstimated from graph image^fEstimate by QAM256 probability function^gNo energy efficiency enhancement techniques applied^hUnclear whether drain or total efficiency

simulate these linearity metrics with simulations for the power RF-DAC design. Therefore these metrics are left out of table 7.1. It would require post-layout transient simulations, but these were impossible to perform on the S-parameter model. This was due to immense computational requirement to convert the S-parameter model to a model suitable for transient simulations. The simulated linearity metrics INL and DNL are difficult to relate to ACPR and EVM. Even if this could be investigated, the results would be of limited use due to additional non-idealities post-production. Secondly, the used LDMOS transistor models are primarily targeted at large transistors. Not much is currently known about the validity of the transistor models for the very small LDMOS transistors. Also, post-fabrication it is possible to apply DPD which is not done in the simulations.

The design for the power RF-DAC presented in this thesis has AM-AM linearization applied on transistor level. However, the design uses a low drive voltage from the LDMOS perspective, resulting in already a quite linear AM-AM baseline curve. This results in the largest segment being only a factor 2,2 larger than the smallest segment. It could be debatable whether it would be better not to apply this linearization and solve the non-linearity by DPD, especially since DPD will most likely be used anyway.

Lastly, it is good to reflect upon the initially set thesis goals. First is the operation frequency and the frequency agility. The design has a maximum operational frequency of $f_c = 3$ GHz, and therefore complies with the lower 5G range of 617 MHz to 2690 MHz. The frequency agility has not been discussed in this thesis, however a Class-E design set has been found that works on 500 MHz, with $\eta_D = 86,8\%$ and $P_{out,f_c} = 20,20$ W. The design set is completely dependent of the external components in the matching network and not on the RF-DAC design itself.

All in all this allows to answer the research question, “Is it possible to design a power RF-DAC, in such a way that this RF-DAC also functions as the PA, with good power efficiency?” The answer will be a “yes”. It should be stressed however that the available technology is currently not optimized for this purpose. It required the use of a separate high-speed digital CMOS driver flip-chipped on a high power analog LDMOS DRAC chip. When CMOS logic can be integrated with high power analog technology, it is expected to see improved performance.

7.2. Recommendations for Further Work

As mentioned in the discussion, the choice to apply AM-AM linearization on transistor level could be replaced by DPD. That would change the used DRAC topology, since the segments would not be required anymore. More bits could be implemented in a thermometer-coded fashion, which should result in a simpler design with improved linearity. However, a major restricting factor will be the design rules regarding the flip-chip connections. Having more bits in thermometer-coded means that most of the transistors will have the same size, but connections are required to be at least $162\ \mu\text{m}$ apart. Staggering the flip-chip connections would result in different inductance L_{bond} to the LDMOS devices, again resulting in non-linearity. Using two rows of LDMOS transistors, similar to what is done in the proposed LDMOS DRAC layout, will result in large devices located relatively far from the drain bar output. This will cause mismatching between the devices ‘in the front’ and ‘in the rear’. In current design this is already present, but much less influential in the linearity as it only involves relatively small LDMOS transistors.

However, when it becomes possible to integrate CMOS logic in the LDMOS process this issue can be avoided entirely. If these CMOS transistors will have a similar breakdown voltage as the used thick oxide devices in TSMC 40 nm technology, i.e. in the order of 2 V, it will definitely make sense to toss the transistor level AM-AM linearization and shift the linearization to the digital domain by DPD.

Given the same LDMOS transistor specifications, then preferably the integrated CMOS transistors would allow for a higher breakdown voltage, in the order of 6 V, as it will improve SMPA performance. This will cause the DRAC to operate more in voltage limited region, causing more non-linearity in the baseline AM-AM curve. Then the transistor level AM-AM linearization will make sense again.

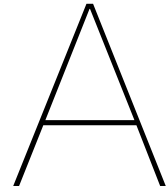
Currently the LDMOS transistors are optimized for linear analog operation. This may not be the optimum for SMPAs that are inherently non-linear. Generally the power DRAC performance benefits from higher drive voltages, but input power scales quadratically with input voltage when using the Class-D driver. It should therefore be possible to find an optimum for system efficiency.

7.2.1. Short-Term Future Work

It should be aimed for to actually fabricate the in this thesis proposed power RF-DAC system. This will involve creating the complete CMOS driver design. Investigating the performance of the DRAC in other modes of operation than push-pull, such as Doherty, should be investigated as well. The duty-cycle d has been fixed at 1 the entirety of this thesis, however decreasing the duty-cycle could mean that the Class-E ZVS criterion can be met. This will improve drain efficiency, at the price of lower output power. The required drive power will stay the same, so therefore it might not be beneficial for system efficiency. Also, it might prove difficult to create a reliable method to control precise values of duty-cycle on the CMOS driver when driving many differently sized capacitive loads.

As indicated in the discussion, it was not possible to perform transient simulations post-layout. If it is possible to create a more simple, lumped component, version of the layout that still accurately models the layout effects, it might prove possible to perform transient simulations. Then it will be interesting to see what the DRAC performance will be when provided with provided with input data created from modulated signals.

Lastly, different Electronic Design Automation (EDA) software suites are used to make the LDMOS DRAC and the CMOS driver designs. This is mostly due to the transistor models being only supplied for one of them. Therefore the DRAC has only been simulated using an ideal resistor that models the CMOS driver. The other way around the CMOS driver is only simulated with an ideal capacitor as load. Also nothing is known about possible non-linearity of the ESD capacitances. The effect of the non-linearity of the LDMOS gate and ESD capacitances on the CMOS, or the non-linearity of the CMOS transistor on the DRAC performance, has been left uninvestigated. The transistor models should have some degree of compatibility between the two EDA software suites and mutual performance should be investigated.



Appendix

A.1. Design Set Example Class-E

Table A.1: Ideal switch Class-E design set, quasi load insensitive without additional reactance (quasi parallel-circuit) for the example simulations in sections 2.3.4 and 3.3

Quantity	Value	
f_c	1	GHz
V_{DC}	10	V
R_L	25	Ω
q	1,33	
d	1	
$m[63]$	3,98	10^{-3}
K_C	0,685	
K_L	0,829	
K_P	1,326	
K_X	0	
Q_L	2,00	
χ	2,76	
C_S	4,36	pF
L_0	7,96	nH
C_0	3,18	pF
L_{DC}	3,30	nH
X	-	
$R_{on}[63]$	0,145	Ω

A.2. Transient DRAC Simulations

Transient simulation was possible to perform on the design before the S-parameter models for the layout were integrated. As the results from these transient simulations are of limited use as they will not represent final performance they are omitted from the thesis body. Nonetheless, in figure A.1 a successful transient simulation result is shown in order to not withhold these results. The DRAC input ACW has a sinusoidal envelope and no phase modulation. It was sampled at $f_{rs} = f_c$, therefore no spectral replicas can be observed, only harmonic distortion. As it is a push-pull topology the even orders of harmonic distortion are suppressed, leaving only the odd harmonics. The intention was to perform a two-tone test with $f_{ts} = 20$ MHz, but as no phase modulation was present this failed and the carrier at $f_c = 3$ GHz is clearly visible in the output spectrum. Therefore the power present at 2,98 GHz and 3,02 GHz is not a pure IM3 product as other frequency components are mixed on top of it. Nonetheless it could be found that there is a difference of 43,5 dBc between these ‘supposed to be IM3 products’ and the carrier, showing some degree of system linearity. This transient simulation was performed on a DRAC sizing that did not yet have good piecewise AM-AM linearization applied. This transient simulation itself and the loading of the result were already very resource intensive and time consuming, therefore no further attempts were made to do the simulation over.

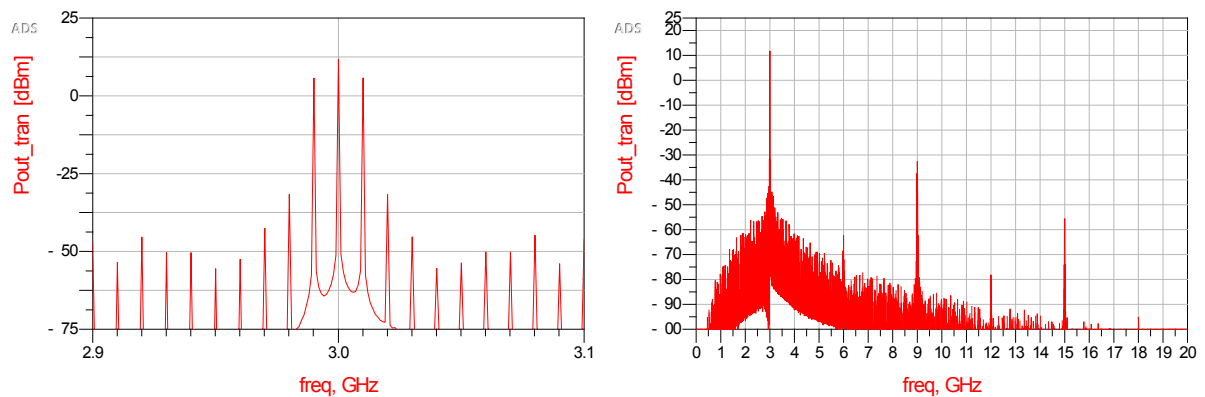


Figure A.1: Pre-layout, with sub-optimum AM-AM linearization

A.3. Harmonic Distortion of Designed DRAC

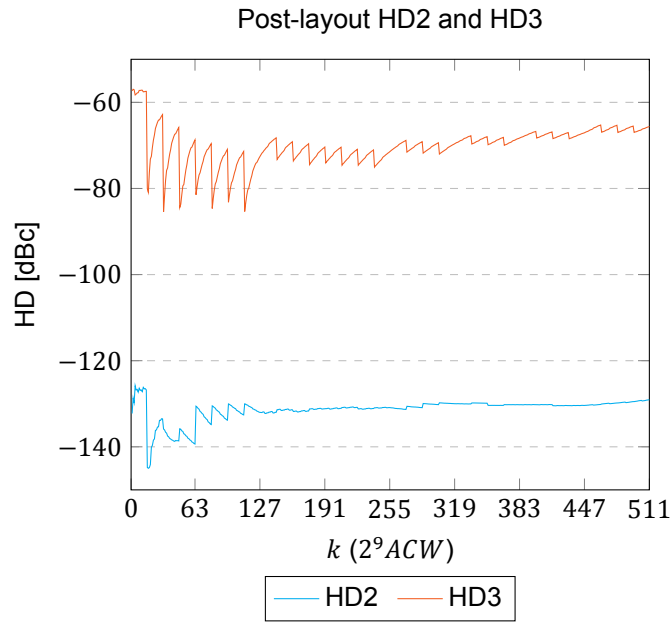


Figure A.2: Post-layout harmonic distortion for $Q_L = 5$

A.4. Mutual Inductance

Not much is known at the time of publishing this thesis on how much mutual inductance could be present between the inductors required for Class-E L_{DC} and L_0 . Therefore it is omitted from the main thesis body, but was investigated nonetheless. The circuit diagram for a Class-E SMPA DRAC with the T-equivalent of the mutual inductance between L_{DC} and L_0 is shown in figure A.3. For simplicity the single-ended version of the Class-E SMPA DRAC is shown. The mutual inductance M can be specified by

$$M = k\sqrt{L_{DC}L_0} \quad (\text{A.1})$$

$$= k\sqrt{0,215 \cdot 10^{-9} \cdot 9,583 \cdot 10^{-9}} = 1,435 \cdot 10^{-9} \cdot k \quad (\text{A.2})$$

where k is the coupling coefficient between the two inductors. The resulting output and DC power and efficiency post-layout is shown in figure A.4. When solving k for $M = L_{DC} = 0,215$ nH this results in $k = 0,1498$, which is the point where the output power and drain efficiency become 0. This makes sense, as the node between M and $L_0 - M$ is now shorted through $L_{DC} - M = 0$ H to V_{DC} . For all values $k > 0,15$ the inductance $L_{DC} - M$ is actually negative, which is not a realistic scenario. Most likely this issue can be resolved by optimizing the value of L_{DC} and L_0 for a given value of k . This is not investigated as there is at this point nothing known about what possible values of k could be. However, it is most likely that minimized mutual inductance will be beneficial for the drain efficiency.

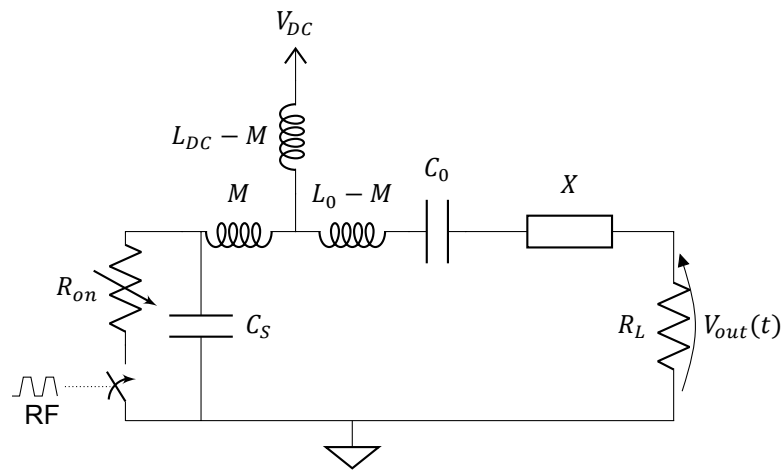


Figure A.3: Single-ended Class-E circuit diagram modified with mutual inductance

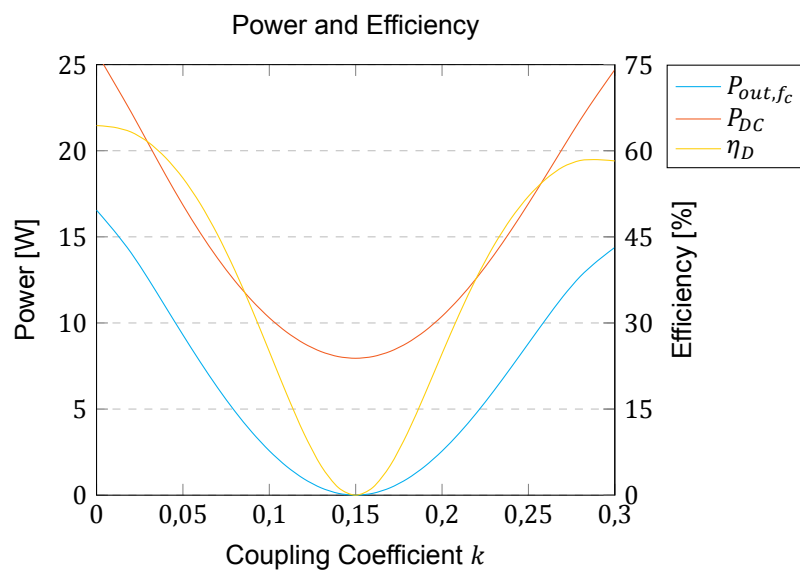


Figure A.4: Effect of mutual inductance post-layout on output power and DC power consumption

A.5. OTA Transistor Widths

All transistors are minimum length (270 nm). Circuit diagram is shown in figure 6.7.

Table A.2: OTA transistor widths and bias voltages

Transistor	Width
M1 & M2	870 nm
M3 & M4	3000 nm
M5	6000 nm
M6 & M8	320 nm
M7 & M9	320 nm
V_{bias_p}	1210 mV
V_{bias_n}	$V_{DDh} - 1066$ mV

B

LDMOS Case Study: A Higher Drive Voltage

The power DRAC design presented in the thesis body is limited to a drive voltage of 2 V, due to the breakdown voltage limitations of the CMOS driver. When driving the LDMOS transistors with a higher voltage this results in a higher output power, as well as a higher drain efficiency. For this example a Total Gate Width (TGW) of 50 mm in push-pull configuration is taken, which is similar to the TGW of the designed power DRAC. The DRAC supply voltage is $V_{DS} = 25$ V.

In figure B.1 the baseline AM-AM and AM-PM curves are shown. The AM-AM non-linearity is much more pronounced, compared to the 2 V drive voltage scenario. This is also true for the AM-PM curve, now showing a phase deviation of $24,4^\circ$ over the range of input codes. Around $k \approx 25$ the transition from the current limited region to the voltage limited region can be seen. It can be identified by the AM-AM curve straying away from its initially linear path.

The transition between the two regions can also be seen in figure B.2 in the fundamental output power. It has a quadratic relationship to ACW in the current limited region, which it loses when entering the voltage limited region. From the drain efficiency curve can be seen that the DRAC operates most efficient when in the voltage limited region, which is how Class-E is intended.

The complete design set and component values are shown in table B.1. They result in a peak fundamental output power of $P_{out,f_c} = 55,9$ W with a drain efficiency of $\eta_D = 79,0\%$ at $f_c = 3$ GHz. Another design set could be found that was able to have peak $P_{out,f_c} = 70,2$ W, but had a lower drain efficiency.

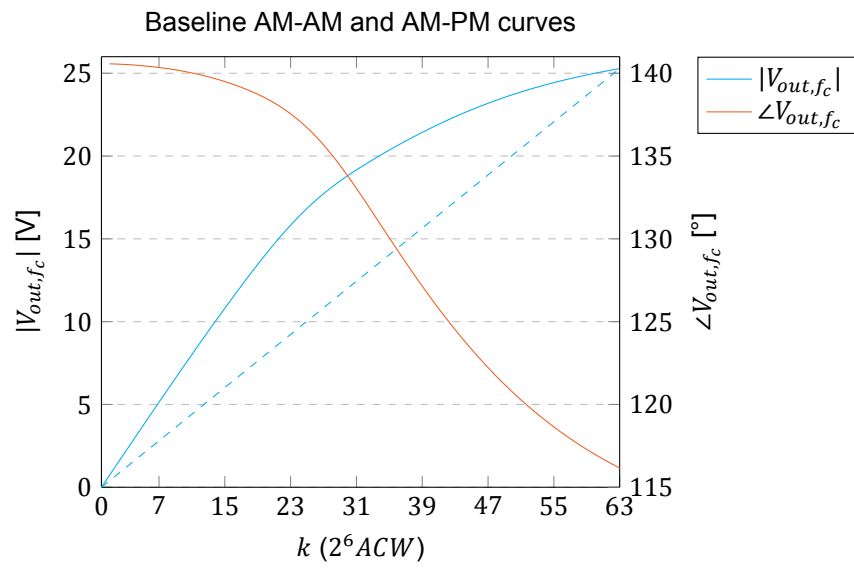


Figure B.1: LDMOS baseline AM-AM and AM-PM curves for $V_{in} = 6$ V, $N_b = 6$ at $f_c = 3$ GHz, push-pull $TGW = 50$ mm, and the design set as shown in table B.1

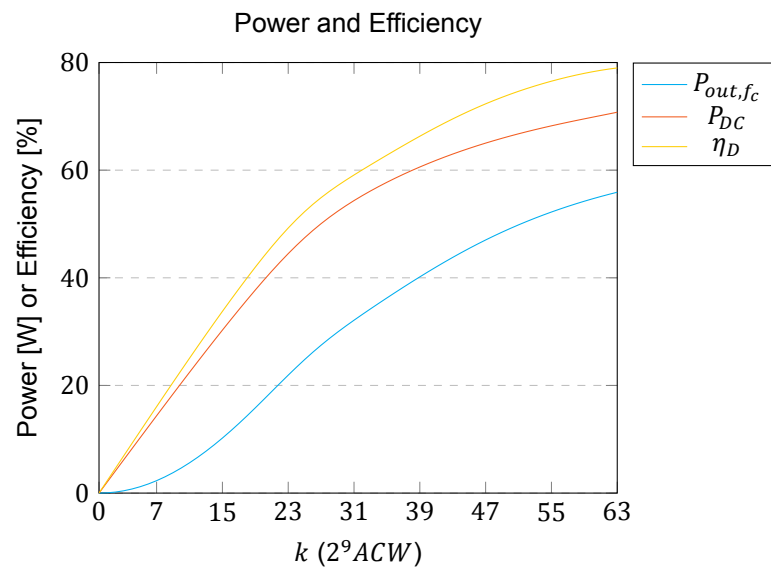
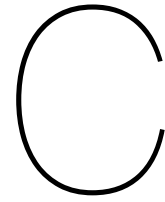


Figure B.2: Power and efficiency corresponding to LDMOS baseline for $V_{in} = 6$ V at $f_c = 3$ GHz, push-pull $TGW = 50$ mm, and the design set as shown in table B.1

Table B.1: Design set and values when using $V_{in} = 6$ V, with $TGW = 50,0$ mm (push-pull) and $f_c = 3$ GHz

Quantity	Value	
q	1,42	
d	1,000	
m	97,0	10^{-3}
K_L	1,23	
K_C	0,40	
K_P	0,51	
K_X	-2,38	
Q_L	5,00	
$t_{rise/fall}$	42	ps
C_S	7,43	pF
R_{on}	0,42	Ω
L_{DC}	0,758	nH
$2 \cdot R_L$	5,72	Ω
X	7,804	pF
C_0	3,712	pF
L_0	0,758	nH
$2 \cdot P_{out,f_c}$	55,90	W
η_D	79,0	%



Design in GaN

This thesis project started with trying to make a DRAC design using GaN technology. For this example a Total Gate Width (TGW) of 18,9 mm in single-ended configuration is taken. The DRAC supply voltage is $V_{DS} = 40$ V. The operation frequency is set at $f_c = 1$ GHz, as this is early thesis work that still assumed the use of bondwires to connect to the GaN transistor gates.

In figure C.1 the baseline AM-AM and AM-PM curves are shown. Compared with the LDMOS $V_{in} = 6$ V scenario the AM-AM non-linearity is similarly pronounced. The AM-PM curve shows here a phase deviation of $42,2^\circ$ over the range of input codes. Around $k \approx 21$ the transition from the current limited region to the voltage limited region can be seen. It can be identified by a relatively sharp curve in the AM-AM and AM-PM paths.

The transition between the two regions can also be seen in figure C.2 in the DC supply power. After entering the voltage limited region the DC supply power stops growing, while the fundamental output power keeps increasing.

The complete design set and component values are shown in table C.1. They result in a peak fundamental output power of $P_{out,f_c} = 58,9$ W with a drain efficiency of $\eta_D = 84,1\%$ at $f_c = 1$ GHz.

Finding working design sets proved to be difficult, as harmonic balance simulations often failed to converge. Early on in the thesis project a design set was found for a TGW of 36 mm, which was not documented and could not be reproduced. The resulting performance from the design set was documented, and had a peak fundamental output power of $P_{out,f_c} = 103$ W with a drain efficiency of $\eta_D = 91,1\%$ at $f_c = 1$ GHz. It used larger GaN devices which have lower output capacitance per unit gate width, what explains the better performance.

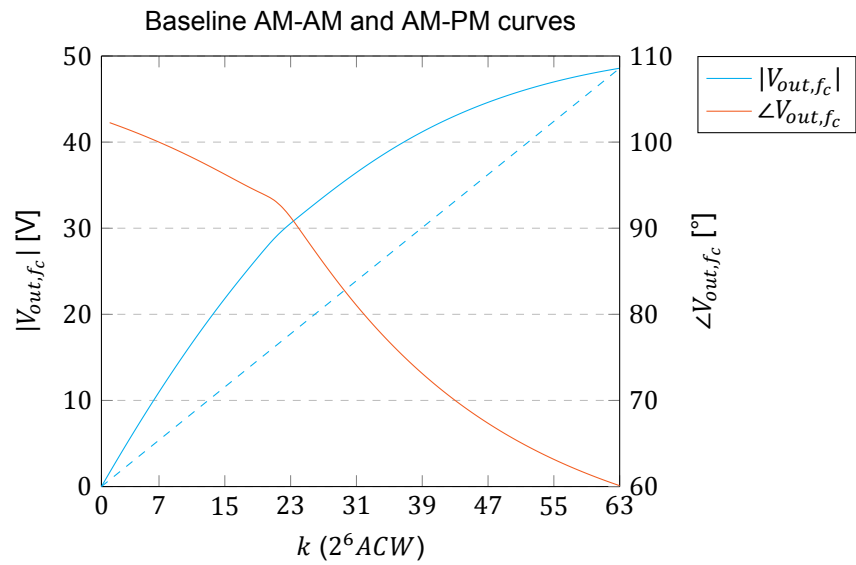


Figure C.1: GaN baseline AM-AM and AM-PM curves for $N_b = 6$ at $f_c = 1$ GHz, $TGW = 18,9$ mm, and the design set as shown in table C.1

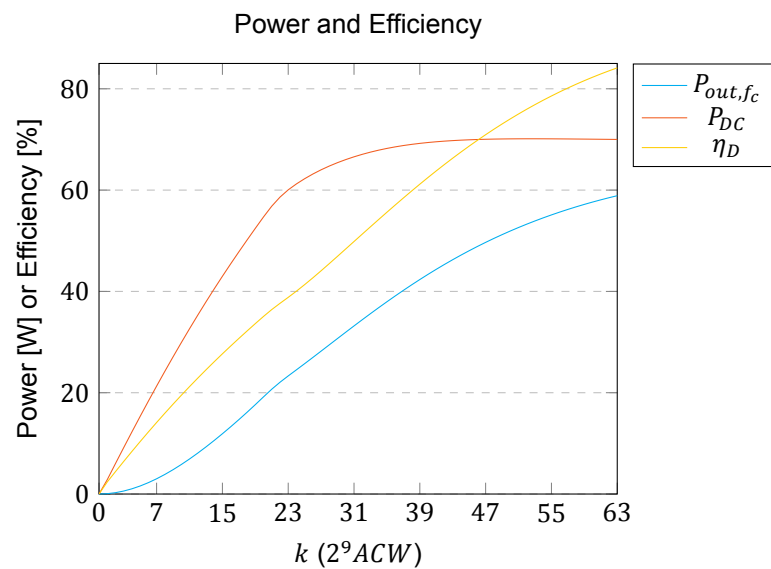


Figure C.2: Power and efficiency corresponding to GaN baseline for $V_{in} = 6$ V at $f_c = 1$ GHz

Table C.1: Design set and values with $TGW = 18,9$ mm and $f_c = 1$ GHz

Quantity	Value	
q	0,94	
d	1,00	
m	9,8	10^{-3}
K_L	0,49	
K_C	1,78	
K_P	0,74	
K_X	0,63	
Q_L		
$t_{rise/fall}$	10	ps
C_S	14,2	pF
R_{on}	0,11	Ω
L_{DC}	1,56	nH
R_L	20,0	Ω
X	2,0	nH
C_0	8,80	pF
L_0	2,88	nH
P_{out,f_c}	58,9	W
η_D	84,1	%

Bibliography

- [1] J. F. Ramsay, “Microwave antenna and waveguide techniques before 1900,” *Proceedings of the IRE*, vol. 46, no. 2, pp. 405–415, Feb 1958.
- [2] Alcatel Lucent Bell Labs research, *The Future X Network*. CRC Press, Taylor & Francis Group, 2016.
- [3] L. C. N. de Vreede, “Next generation wireless technologies, “a search for the perfect match!”,” Sept 2016.
- [4] WhatsApp, “Connecting one billion users every day,” Online (<https://blog.whatsapp.com/10000631/>), July 2017.
- [5] Tweakers.net, “Vodafone begint met aanbieden 1 Gbit s⁻¹-4G en gebruikt ziggo-frequenties,” Online (<https://tweakers.net/nieuws/130377/>), October 2017.
- [6] “<https://5g-ppp.eu/>,” Online, June 2017.
- [7] Global e-Sustainability Initiative (GeSI), “#smarter2030 ict solutions for 21st century challenges,” 2015.
- [8] Tweakers.net, “Tele2 schakelt 2600 MHz op masten ’s nachts uit om stroom te besparen,” Online (<https://tweakers.net/nieuws/127269/>), July 2017.
- [9] R. Le Maistre, “Energy bill shocks orange into action,” Online (<http://www.lightreading.com/optical/100g/energy-bill-shocks-orange-into-action/d/d-id/711038>), September 2014.
- [10] L. W. Couch, M. Kulkarni, and U. S. Acharya, *Digital and analog communication systems*. Prentice Hall, 1997, vol. 7.
- [11] E. McCune, “A technical foundation for RF CMOS Power Amplifiers: Part 4: Misunderstandings in PA design,” *IEEE Solid-State Circuits Magazine*, vol. 8, no. 2, pp. 75–82, Spring 2016.
- [12] M. S. Alavi, R. B. Staszewski, L. C. N. de Vreede, A. Visweswaran, and J. R. Long, “All-digital RF I/Q modulator,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 11, pp. 3513–3526, Nov 2012.
- [13] G. E. Moore, “Cramming more components onto Integrated Circuits, reprinted from electronics, volume 38, number 8, april 19, 1965, pp.114 ff.” *IEEE Solid-State Circuits Society Newsletter*, vol. 11, no. 5, pp. 33–35, Sept 2006.
- [14] R. H. Dennard, F. H. Gaensslen, V. L. Rideout, E. Bassous, and A. R. LeBlanc, “Design of ion-implanted MOSFET’s with very small physical dimensions,” *IEEE Journal of Solid-State Circuits*, vol. 9, no. 5, pp. 256–268, Oct 1974.
- [15] K. Bult, “The effect of technology scaling on power dissipation in analog CMOS circuits,” *ISSCC*, 2009.
- [16] R. B. Staszewski, “Digital deep-submicron CMOS frequency synthesis for RF wireless applications,” Ph.D. dissertation, University of Texas at Dallas, 2002.
- [17] K. Konanur, “CMOS RF transceiver chip tackles multiband 3.5G radio system,” *RF DESIGN*, vol. 29, no. 4, p. 22, 2006.
- [18] 3rd Generation Partnership Project, “3GPP TS 38.101-1 V1.0.0,” Technical Specification, Dec 2017, 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; NR; User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone (Release 15).

- [19] E. McCune, "A technical foundation for RF CMOS Power Amplifiers: Part 3: Power Amplifier 3-port characteristics," *IEEE Solid-State Circuits Magazine*, vol. 8, no. 1, pp. 44–50, winter 2016.
- [20] S. D. Kee, "The class E/F family of harmonic-tuned switching Power Amplifiers," Ph.D. dissertation, California Institute of Technology, 2002.
- [21] P. Reynaert and M. Steyaert, *RF Power Amplifiers for mobile communications*. Springer Science & Business Media, 2006.
- [22] J. H. K. Vuolevi, T. Rahkonen, and J. P. A. Manninen, "Measurement technique for characterizing memory effects in rf power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, no. 8, pp. 1383–1389, Aug 2001.
- [23] J. M. Rabaey, A. P. Chandrakasan, and B. Nikolic, *Digital integrated circuits*. Prentice hall Englewood Cliffs, 2002, vol. 2.
- [24] N. O. Sokal and A. D. Sokal, "Class e—a new class of high-efficiency tuned single-ended switching power amplifiers," *IEEE Journal of Solid-State Circuits*, vol. 10, no. 3, pp. 168–176, Jun 1975.
- [25] F. Raab, "Idealized operation of the class e tuned power amplifier," *IEEE Transactions on Circuits and Systems*, vol. 24, no. 12, pp. 725–735, December 1977.
- [26] M. Acar, A. J. Annema, and B. Nauta, "Analytical design equations for class-e power amplifiers," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 54, no. 12, pp. 2706–2717, Dec 2007.
- [27] —, "Analytical design equations for class-e power amplifiers with finite dc-feed inductance and switch on-resistance," in *2007 IEEE International Symposium on Circuits and Systems*, May 2007, pp. 2818–2821.
- [28] M. P. van der Heijden, M. Acar, J. S. Vromans, and D. A. Calvillo-Cortes, "A 19 W high-efficiency wide-band CMOS-GaN class-e chireix RF outphasing Power Amplifier," in *2011 IEEE MTT-S International Microwave Symposium*, June 2011, pp. 1–4.
- [29] T. C. Carusone, D. A. Johns, and K. W. Martin, *Analog Integrated Circuit design*. John Wiley & Sons, 2012, vol. 2.
- [30] S. M. Yoo, J. S. Walling, E. C. Woo, B. Jann, and D. J. Allstot, "A switched-capacitor RF Power Amplifier," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 12, pp. 2977–2987, Dec 2011.
- [31] M. J. Pelgrom, "ET4369 Nyquist-Rate Data Converters lecture slides," 2015.
- [32] S. M. Alavi, "All-digital I/Q RF-DAC," Ph.D. dissertation, Delft University of Technology, 2014.
- [33] M. Hashemi, Y. Shen, M. Mehrpoo, M. Acar, R. van Leuken, M. S. Alavi, and L. de Vreede, "17.5 an intrinsically linear wideband digital polar PA featuring AM-AM and AM-PM corrections through nonlinear sizing, overdrive-voltage control, and multiphase RF clocking," in *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, Feb 2017, pp. 300–301.
- [34] M. Hashemi, L. Zhou, Y. Shen, M. Mehrpoo, and L. de Vreede, "Highly efficient and linear class-e cmos digital power amplifier using a compensated marchand balun and circuit-level linearization achieving 67aclr without dpd," in *2017 IEEE MTT-S International Microwave Symposium (IMS)*, June 2017, pp. 2025–2028.
- [35] S. Oliver, "Optimize a power scheme for these transient times," *ElectronicDesign*, Sept 2014.
- [36] *IAFpdk_GE_MSL_A200_enc, version 2.00*, Fraunhofer IAF, November 2013, description of active AlGaN/GaN HEMT microstrip components on SiC substrate for the IAF GaN25 technology.
- [37] M. P. van der Heijden, M. Acar, and S. Maroldt, "A package-integrated 50 W high-efficiency RF CMOS-GaN class-e Power Amplifier," in *2013 IEEE MTT-S International Microwave Symposium Digest (MTT)*, June 2013, pp. 1–3.

- [38] TSMC *Universal Analog I/O Library General Application Note, version 2.00*, TSMC, April 2008.
- [39] A. Kavousian, D. K. Su, M. Hekmat, A. Shirvani, and B. A. Wooley, "A digitally modulated polar CMOS Power Amplifier with a 20 MHz channel bandwidth," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 10, pp. 2251–2258, Oct 2008.
- [40] Y. Shen, M. Mehrpoo, M. Hashemi, M. Polushkin, L. Zhou, M. Acar, R. van Leuken, M. S. Alavi, and L. de Vreede, "A fully-integrated digital-intensive polar doherty transmitter," in *2017 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, June 2017, pp. 196–199.
- [41] D. Chowdhury, L. Ye, E. Alon, and A. M. Niknejad, "An efficient mixed-signal 2.4-ghz polar power amplifier in 65-nm cmos technology," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 8, pp. 1796–1809, Aug 2011.

Glossary

- fourth generation** the fourth generation of Wireless mobile/cellular Communication, mainly LTE
- fifth generation** the fifth generation of Wireless mobile/cellular Communication, currently under development
- bandpass** a bandpass waveform has a spectral magnitude that is nonzero for frequencies in some band concentrated about a round the carrier frequency $f_c \gg 0$ and is negligible elsewhere [10] xi, 3, 4, 11–13, 25–27, 30, 31, 40, 42
- bandwidth expansion** the phenomenon that polar signals require larger bandwidth due to nonlinear conversion between Cartesian and polar coordinates 4, 5, 101
- baseband** a baseband waveform has a spectral magnitude that is nonzero for frequencies in the vicinity of the origin (i.e., $f = 0$) and negligible elsewhere [10]; has a sample frequency (bandwidth) of f_{BB} xi, 3, 4, 40–43, 147–149
- compression point** the output power level where the ratiometric power gain drops X dB from the small signal gain 14
- CORDIC** a simple and efficient algorithm to calculate hyperbolic and trigonometric functions, requiring no hardware multipliers
- digital RF** the use of digital electronics to convey information in discrete amplitude and continuous time rather than sampled analog which is continuous amplitude and discrete time vii, 5, 137
- drain bar** the interconnect at the drain side of the LDMOS devices to the output 90–93, 95, 96, 98, 99, 102, 118, 120, 147
- dynamic range** the ratio of the maximum to minimum signal power it can meaningfully generate [29] 37–39
- glitch** timing glitch due to different timing delays when switching different signals, mainly present in binary-weighted systems xi, xii, 35, 71, 88–90, 98, 101, 112, 117, 118
- inshing** placing a decoupling capacitor close to a RFIC as signal ground and then using bondwires as (shunt) inductor; comes from INductor SHunt xii, 91
- Internet of Things** the inter-networking of physical objects which enable these objects to collect and exchange data
- monotonicity** a DAC property that the output always increases when the input increases 38, 81, 88, 96
- narrowband** a narrowband waveform can be assumed to only consist of sinusoids of the fundamental frequency and its harmonics [20] 4, 11, 16
- Nyquist rate** the minimum sampling rate to prevent aliasing, equal to twice the Nyquist frequency or twice bandwidth of the to be sampled signal ix, 33, 34, 39, 43, 46, 117
- pitch** minimum center-to-center distance between two (of the same) components or devices (on chip) 61, 91

resolution resolving power: the number of (meaningful) output levels, usually expressed in bit [33](#), [34](#), [66](#)

Root Mean Square the root ($\sqrt{\cdot}$) of the squared average, usually a time average of a voltage; $RMS = \sqrt{\langle \cdot^2 \rangle}$

technology scaling the decreasing feature size of transistors over time due to technology improvements [5](#), [34](#), [49](#), [69](#)

two-tone test a linearity test by applying two frequencies at the input, usually centered around the carrier frequency f_c , and observing the frequency components at the output [14](#), [43](#), [124](#), [148](#)

Acronyms

- 4G** fourth generation 1, 2, 137, *Glossary*: fourth generation
- 5G** fifth generation 2, 6, 120, *Glossary*: fifth generation
- AC** Alternating Current 38, 40, 41, 44, 46
- ACPR** Adjacent Channel Power Ratio v, 7, 14, 57, 69, 117, 118, 120
- ACW** Amplitude Code Word xi–xiii, xv, 5, 6, 39–42, 45, 56, 59, 66, 67, 77–79, 81–84, 86–90, 96, 98, 99, 102, 106, 112, 113, 124, 125, 129, 130, 134, 148
- AM** Amplitude Modulation 3–5, 13–15, 22, 29, 44, 143
- AM-AM** Input Amplitude to Output Amplitude conversion v, vii, x, xii, xiii, 13, 14, 37, 43, 45, 46, 65, 69, 71, 77–79, 81, 83, 86–90, 96, 98, 107, 112, 117, 118, 120, 124, 129, 130, 133, 134, 138
- AM-PM** Input Amplitude to Output Phase conversion xii, xiii, 14, 45, 60, 71, 83, 86–90, 96, 98, 101, 112, 118, 129, 130, 133, 134, 138
- CMOS** Complementary Metal-Oxide-Semiconductor i, iii, v, vii, x, xii, xiii, 4–7, 19, 27, 34, 49, 57–59, 61, 62, 64, 65, 69, 84, 88–93, 95, 96, 98, 99, 101–104, 106, 108, 110–115, 117–121, 129, 137–139, 147–149
- CORDIC** COordinate Rotation Digital Computer 112, 115, 118, *Glossary*: CORDIC
- DAC** Digital-to-Analog Converter vii, ix, xi, 4, 5, 7, 33–43, 46, 117, 141, 145, 149
- DC** Direct Current v, xii, xiii, 11–13, 15, 16, 20–25, 27–32, 38, 40, 41, 43, 44, 50, 55, 57, 59, 68, 71–73, 76, 83, 84, 91, 92, 97, 99, 107, 108, 111–115, 119, 123, 125, 126, 130, 131, 133–135, 147–150
- DCO** Digitally Controlled Oscillator 5, 6
- DNL** Differential Nonlinearity xi, 37, 38, 43–45, 81, 83, 86, 87, 96, 98, 99, 112, 120
- DPD** Digital Pre-Distortion v, 5, 89, 101, 120
- DR** Dynamic Range 38, 43, 44, 57, *Glossary*: dynamic range
- DRAC** Digital-to-RF-Amplitude Converter i, iii, v, vii, ix–xiii, 5–7, 33, 34, 39–46, 49, 50, 52, 56–60, 65, 66, 68, 69, 71, 72, 74–90, 92, 94–96, 98, 99, 101, 102, 107, 112–115, 117, 118, 120, 121, 124, 125, 129, 133, 147–149
- DSP** Digital Signal Processing v, 4–6, 101, 102, 112, 114, 115, 118
- EDA** Electronic Design Automation 121
- EEMCS** Electrical Engineering, Mathematics and Computer Science iii, iv, vii,
- ELCA** Electronics Research Laboratory iii, vii, 6, 101, 102, 112
- EM** Electro-Magnetic xii, 1, 11, 13, 16, 83–85, 92, 95, 96
- ENOB** Effective Number Of Bits 39, 79, 112
- ESD** Electro-Static Discharge v, vii, xii, 18, 59–62, 65, 68, 69, 81, 83, 84, 91, 92, 95, 96, 98, 103, 105, 106, 110, 111, 113, 114, 117, 118, 121

- ESR** Equivalent Series Resistance 19, 59–61, 69, 102, 149
- ETV** Electrotechnische Vereeniging vii,
- EVM** Error Vector Magnitude 14, 15, 118, 120
- FET** Field-Effect Transistor xi, 17, 19, 20, 50, 150
- GaN** Gallium Nitride ix–xi, xiii, xv, 6, 7, 50–57, 69, 77, 117–119, 133, 134, 138
- HD2** Second order Harmonic Distortion v, xiii, 14, 37, 68, 69, 101, 112, 117, 125
- HD3** Third order Harmonic Distortion 14, 68, 125
- HEMT** High-Electron-Mobility Transistor 49, 50, 53, 55
- IC** Integrated Circuit 1, 4, 5, 50, 137, 138, 141, 144
- IM3** Third order InterModulation distortion 14, 37, 39, 124
- INL** Integral Nonlinearity v, xi, xii, 37, 38, 43–45, 77–79, 81, 83, 86, 87, 98, 99, 101, 112, 118, 120
- IoT** Internet of Things 2, 5, *Glossary*: Internet of Things
- IP** Intellectual Property 102
- IP3** Third order Intercept Point 14
- LDMOS** Laterally Diffused Metal-Oxide-Semiconductor i, iii, v, vii, ix, x, xii, xiii, xv, 6, 7, 49, 50, 55–62, 64, 65, 69, 71–74, 76–78, 80–84, 86, 88, 90–96, 98, 99, 106, 110–112, 114, 115, 117–121, 129, 130, 133, 141
- LDO** Low DropOut regulator 4,
- LO** Local Oscillator 4–6
- LSB** Least Significant Bit v, 33, 35, 37–39, 43–45, 66, 69, 80, 81, 83, 87, 88, 96, 98, 112, 117, 118
- LTE** Long Term Evolution iv, 1, 141
- LTI** Linear Time-Invariant 15–17, 20, 27, 149
- LUT** Look-Up Table 112
- MIM** Metal-Insulator-Metal 50, 55, 91, 92
- MIMO** Multi-Input Multi-Output system 71
- MMIC** Monolithic Microwave Integrated Circuit 50
- MOM** Metal-Oxide-Metal 91, 111
- MoM** Method of Moments 84, 92, 95
- MOS** Metal-Oxide-Semiconductor 4, 6, 50, 53, 55, 56, 59, 91, 111, 119, 143, 144, 150
- MSB** Most Significant Bit 33, 35, 37, 39, 66, 69, 80, 81, 88, 117
- NDA** Non-Disclosure Agreement 83
- OTA** Operational Transconductance Amplifier x, xii, xv, 107–109, 127

- PA** Power Amplifier v, ix, xi, 4–7, 11–18, 20–26, 28, 30–32, 34, 37, 39, 43–45, 59, 67, 68, 113, 115, 117, 118, 120, 137–139, 145, 147, 149
- PAE** Power Added Efficiency 13, 119
- PAPR** Peak-to-Average Power Ratio v, ix, 11, 15, 22, 23, 44, 58, 112, 113, 115, 118, 119
- PM** Phase Modulation 3–6, 13, 14, 29, 45, 57, 101, 102, 112, 114, 115, 143, 145
- PM-PM** Input Phase to Output Phase conversion 13
- PVT** Process, Voltage and Temperature 57, 65, 88, 89
- RF** Radio Frequency iii, v, vii, ix, xi, xii, xv, 1, 3–5, 11, 13, 15, 18, 20, 27, 28, 30, 33, 34, 39–45, 49–51, 55, 68, 86, 88–92, 95, 97, 108, 110, 112, 117, 118, 126, 137, 138, 141, 143, 145
- RF-DAC** Radio Frequency Digital-to-Analog Converter i, iii, v, vii, ix, xi, xii, 5–7, 33, 34, 36, 38–40, 42, 44, 46, 49, 57, 64, 69, 88, 101, 102, 113, 115, 117–121, 138
- RMS** Root Mean Square 12, 15, 38, 39, 43, 142, *Glossary*: Root Mean Square
- S-parameter** scattering parameter xi, xii, 50–52, 55, 84, 85, 92–95, 111, 120, 124
- SAW** Surface Acoustic Wave 5
- SFDR** Spurious-Free Dynamic Range 39
- SMPA** Switch-Mode Power Amplifier i, iii, v, xi, xii, 26, 27, 29, 30, 32, 33, 44–46, 52, 57–59, 65, 68, 69, 71, 98, 117, 119, 120, 125
- SNDR** Signal-to-Noise and Distortion Ratio 39
- SNR** Signal-to-Noise Ratio 38, 39
- SQNR** Signal-to-Quantization-Noise Ratio 39
- SRAM** Static Random-Access Memory 102
- TGW** Total Gate Width v, xii, xiii, xv, 50–58, 60, 62, 66, 68, 71–79, 81, 82, 97, 98, 117, 129–131, 133–135
- THD** Total Harmonic Distortion 14, 39
- TU Delft** Delft University of Technology iii, iv, vii,
- TX** transmitter ix, xi, 3–6, 13
- UHF** Ultra High Frequency 6, 33
- VCO** Voltage Controlled Oscillator 4
- Y-parameter** admittance parameter 51, 52, 55, 56, 84
- Z-parameter** impedance parameter 84, 92, 93, 96
- ZdVS** Zero-Voltage slope Switching 28
- ZOH** Zero-Order Hold xi, 41–43
- ZVS** Zero-Voltage Switching 27, 28, 32, 52, 53, 56, 72, 73, 121

Symbols

- α transconductance amplifier conduction angle 22–24, 29
- $\rho(t)$ (baseband) amplitude signal 3–6
- α_V waveform related ratio between peak voltage and DC supply voltage 40, 41, 43
- β PMOS to NMOS width ratio in a CMOS logic gate 102–106
- C_{in} effective input capacitance of a CMOS logic gate 103–107
- C_{layout} parasitic parallel capacitance of one drain bar 92, 93, 96
- C_{out} effective output capacitance of a CMOS logic gate 19, 103–107
- C_S Class-E total parallel switch capacitance 27–30, 32, 44, 68, 71–73, 93, 96, 97, 99, 126, 131, 135, 148, 149
- C_{DS} device model parasitic output capacitance 18, 19, 27, 30, 50–52, 54–58, 71, 76, 93, 97, 102, 103, 118
- C_{GS} device model parasitic input capacitance 18, 19, 26, 50, 51, 53–65, 81–83, 86, 96, 102, 103, 105, 106, 114, 118
- d thermometer-coded digital input, always provided with a subscript indicating the ‘position’ of the thermometer-code; also used to refer to the thermometer-coded DRAC segments xv, 35, 36, 39, 41, 66, 67, 79, 81, 89
- d Class-E duty-cycle 29–32, 71, 76, 97, 105, 108, 121, 123, 131, 135
- η_D drain efficiency v, 13, 21–24, 26, 30, 31, 73, 75–77, 81, 83, 84, 87, 89, 96–99, 107, 113, 114, 118–120, 126, 129–131, 133–135
- $\eta_{T,system}$ total chain efficiency (PA and all driver stages) or system efficiency v, 13, 102, 107, 112–115, 118, 119
- $\varepsilon[k]$ error vector 14, 15
- η_T total efficiency (PA only) v, 13, 23, 84, 98, 113, 114, 118, 119
- F CMOS inverter chain overall effective fanout 106
- f CMOS logic gate effective fanout xii, 105–107, 109, 110, 114, 115
- \mathcal{F} Fourier transform: $F(\omega) = \mathcal{F}_t\{f(t)\} = \int_{-\infty}^{\infty} f(t)e^{-j\omega t} dt$ 41, 42, 110
- f_c carrier frequency (also center frequency) v, xii, xiii, xv, 3, 12–14, 16, 20–28, 30, 31, 39–43, 45, 46, 49, 58, 59, 65, 71–79, 81, 83, 84, 87–90, 93, 96–99, 107, 110–115, 117–120, 123, 124, 126, 129–131, 133–135, 141, 142, 149, 150
- ω_c angular carrier frequency (also angular center frequency) 3, 20, 28–32, 42, 44, 51, 52, 54–58, 67, 68, 71, 72, 84, 92, 93, 148, 149
- f_s sampling frequency , 148
- f_{BB} baseband sampling frequency, channel bandwidth 4, 41, 42, 44, 58, 141

- f_{rs} sampling frequency of the upsampled baseband data 41–44, 58, 112, 124
- f_{ts} two-tone spacing 14, 124
- f_T Transit Frequency, point where small signal short circuit current gain of an intrinsic MOS transistor is unity
- γ CMOS logic gate self-loading factor 104, 106, 107, 114
- $g[k]$ constellation point 14, 15
- $\hat{g}[k]$ transmitted constellation point 14
- G_m large signal transconductance 17
- $g(t)$ complex baseband envelope xi, 3
- i discrete (summation) index, DRAC segment number xv, 13–16, 24, 33, 35, 36, 39, 41, 42, 66, 67, 77–79, 81, 82, 89, 96, 103–107, 112
- I_{DC} DC supply current 11, 12, 15, 22, 23, 28
- I_{DS} drain to source current xi, xii, 15, 17–30, 50–52, 55, 56, 72–74
- I_{in} input current 11, 12
- \Im Imaginary part of complex number or signal 3, 51, 54, 56, 58, 84, 92
- j the imaginary unit 3, 42, 51, 63, 84, 147
- I_{max} maximum transistor output current 56, 57
- $I(t)$ (baseband) in-phase signal 3–6
- I_{out} DC or instantaneous output current 11, 12, 15, 28, 72, 74
- K Class-E design set, $K : \{K_L, K_C, K_P, K_X\}$ 28, 29, 32, 71, 72
- k discrete index, in the case of discrete time index: $t = k/f_s$; in the case of DRAC input code $k = 2^{N_b} ACW$ xv, 14–16, 20, 36–39, 41, 43–45, 66, 67, 77–79, 82–84, 86–90, 96, 98, 99, 113, 125, 129, 130, 133, 134, 148
- K_C Class-E design set parallel capacitance parameter xii, 28, 29, 31, 72–78, 96, 97, 99, 123, 131, 135, 148
- K_L Class-E design set feed inductor parameter 28, 29, 31, 72, 73, 76, 97, 123, 131, 135, 148
- K_X Class-E design set additional reactance parameter xii, 28, 29, 31, 71–78, 97, 123, 131, 135, 148
- K_P normalized output power, Class-E design set output power parameter 26, 28–31, 72, 73, 76, 96, 97, 123, 131, 135, 148
- L (transistor) length 17–19,
- L_{bond} inter-chip connection equivalent inductance xii, 60–62, 64, 84, 86, 87, 96, 98, 99, 103, 111, 120
- L_{DC} finite DC feed inductor 27–30, 32, 44, 68, 72, 73, 76, 97, 125, 126, 131, 135, 149
- m Class-E switch R_{on} and C_S product with ω_c xi, 29–32, 44, 56, 57, 71–73, 76, 90, 93, 96–99, 123, 131, 135
- n discrete (summation) index, DRAC bit number 3, 66, 78, 79, 81, 82, 88

- N_b digital input bitwidth/resolution of a DAC/DRAC v, xii, 33–39, 41, 43, 44, 57, 58, 65, 66, 69, 77–79, 83, 84, 87, 117, 130, 134, 148
- $\varphi(t)$ (baseband) phase signal 3–6
- P_{DC} DC input power for a PA 11–13, 16, 21–24, 83, 84, 99, 107, 112–114, 126, 130, 134
- P_{diss} total power dissipated in the PA 16
- P_{DS} power dissipated in the transistor 16, 20, 24
- P_{in} input power v, 11–13, 59, 83, 84, 98, 107, 110, 112–115, 118
- P_{out} time averaged output power 11–13, 15, 16, 49
- $p_{out}(t)$ instantaneous output power 12
- $P_{out,fc}$ fundamental time averaged output power v, 12–14, 16, 21, 23, 24, 26, 28, 72–77, 81, 83, 84, 87, 96–99, 107, 112–114, 118–120, 126, 129–131, 133–135
- P_{pasv} power dissipated in the LTI passive network 16, 20
- q Class-E L_{DC} and C_S resonance ratio to ω_c xi, xii, 29–32, 44, 71–78, 97, 123, 131, 135
- $Q(t)$ (baseband) quadrature signal 3–6
- R_{bond} inter-chip connection Equivalent Series Resistance 60–62, 64, 84, 96, 103, 111
- \Re Real part of complex number or signal 3, 51, 53, 57, 84
- R_{eq} equivalent resistance 36, 41
- R_{es} Equivalent Series Resistance of CMOS driver, or ‘drive strength’ 19, 59–61, 69, 81, 82, 102–105
- R_L load resistance xi, 11, 12, 15, 21, 23, 25–28, 30, 31, 34, 36, 37, 40, 41, 44, 68, 72, 73, 75, 76, 97, 98, 123, 126, 131, 135
- R_{on} transistor on-resistance xi, xii, 18, 19, 26, 27, 30, 32, 44–46, 50–53, 55–57, 65, 66, 68, 71–73, 76, 93, 97, 123, 126, 131, 135, 148
- S relative segment size or segment scaling 78–81, 97
- τ_d driver first order RC-lowpass approximation time constant xii, 60–64, 69, 75, 81, 82, 84–86, 88, 97, 105
- t time xi, 3, 12, 19–21, 25–27, 38, 40–44, 67, 68, 72, 74, 75, 77, 78, 82, 85, 86, 103–106, 109, 110, 114, 126, 147–149
- θ angular time, $\theta = \omega t$ 20–22, 24, 26, 28–30
- t_{p0} intrinsic CMOS logic gate propagation delay 104, 106, 114
- t_p average CMOS logic gate propagation delay 19, 103–106, 110
- $t_{p,H \rightarrow L}$ CMOS logic gate propagation delay when output transitions from high to low 103–105
- $t_{p,L \rightarrow H}$ CMOS logic gate propagation delay when output transitions from low to high 103–105
- V_{bk} transistor junction breakdown voltage 18–24, 26–30, 32, 50–52, 55–57, 114
- V_{DC} DC supply voltage for a PA, usually with a feed inductor 11, 12, 15, 21, 23, 25–29, 40, 41, 43, 44, 59, 68, 71, 72, 119, 123, 125, 126

- V_{DD} positive supply voltage xii, 18–20, 27, 34, 36, 37, 40, 59, 60, 95, 101–105, 107–111, 115, 117–119, 127
- V_{DS} drain to source voltage xi, xii, 15, 17–22, 24–30, 32, 50–58, 72–74, 129, 133
- V_{GS} gate to source voltage xi, xii, 17–23, 25, 29, 32, 44, 46, 50–58, 60, 64, 72, 73, 81, 82, 85–87, 96, 98, 99, 103, 111
- V_{in} input voltage xiii, xv, 11, 12, 15, 27, 72, 114, 130, 131, 133, 134
- V_k transistor knee voltage, transition between triode and transconductance regions 19–24, 26, 27, 30
- V_{out} DC or instantaneous output voltage 11, 12, 15, 21, 25, 27, 33, 34, 36–45, 67, 68, 72, 74, 126
- $V_{out,fc}$ amplitude of the sinusoidal output voltage at f_c , unless specifically used as vector quantity with a phase 12, 14, 21, 23, 24, 39–43, 45, 77–79, 83, 87–90, 98, 118, 130, 134
- V_{pk} output waveform peak voltage 40, 43
- V_T MOSFET threshold voltage vii, 17–20, 22, 23, 50, 51, 55, 57, 58, 69, 72, 81, 114, 117
- W (transistor) width 17–19, 30, 79, 81, 82, 102, 106, 112
- X additional Class-E reactance 27, 28, 44, 68, 72, 90, 97, 126, 131, 135