Master's thesis

Ohjelmistotekniikka ja ICT

2024

Ossi Toivonen

Integrated CMOS22 Wi-Fi 6 RF Power Amplifier

OPINNÄYTETYÖ (YAMK) | TIIVISTELMÄ

TURUN AMMATTIKORKEAKOULU

Ohjelmistotekniikka ja ICT

2024 | 98 sivua

Ossi Toivonen

Integrated CMOS22 Wi-Fi 6 RF Power Amplifier

Wi-Fi 6 on yksi uusimmista standardeista, jonka allianssi esitteli vuonna 2019 parantaen verkon kokonaiskapasiteettia aiempiin standardeihin verrattuna. Suurtehovahvistimen toteutus langattomiin sovelluksiin on perinteisesti ollut hallitsevaa tietyillä puolijohdeteknologioilla kuten galliumarsenidi (GaAs) ja galliumnitridi (GaN). Tässä opinnäytetyössä tutkitaan mahdollisuutta integroida Wi-Fi 6 -standardiin yhteensopiva langaton tehovahvistin lähetinvastaanotinsiruun käyttäen CMOS22 -puolijohdeteknologiaa, jolloin voidaan välttää tarve ulkoisille PA moduuleille, jotka lisäävät kustannuksia.

Suunniteltu Wi-Fi -suurtehovahvistin on toteutettu käyttämällä differentiaalista common-source -vahvistintopologiaa, jonka differentiaalinen lähtösignaali muunnetaan yksipäiseksi. Vahvistimen fyysisen rakenteen parasiittiset vaikutukset otetaan huomioon käyttäen parasiittisten komponenttien mallinnustyökalua, sekä sähkömagneettista mallinnusta. Suorituskykyä simuloidaan piirisimulaattoreilla, ja lopuksi laboratoriomittaukset suoritetaan keskittyen 5 GHz:n toiminta-alueen suorituskykyyn.

Simuloitu tehovahvistus on 26 dB, ja maksimi saturoitunut lähtöteho on noin 27 dBm. Modulaation tarkkuusmittaukset MCS7:lle (64-QAM) osoittavat virhevektorin magnitudin (EVM) olevan alle −31 dB 16 dBm:n keskimääräisellä lähtöteholla.

Asiasanat:

RF power amplifier, integrated circuit, CMOS, Wi-Fi

MASTER'S THESIS | ABSTRACT

TURKU UNIVERSITY OF APPLIED SCIENCES

Ohjelmistotekniikka ja ICT

2024 | 98 pages

Ossi Toivonen

Integrated CMOS22 Wi-Fi 6 RF Power Amplifier

Wi-Fi 6 is one of the most recent standards introduced by the alliance in 2019 improving the overall capacity of the network over the previous standards. The implementation of high-power radio-frequency power amplifiers (PA) for wireless applications has traditionally been dominated by the semiconductor technologies such as GaAs and GaN. This thesis studies the feasibility to integrate a Wi-Fi 6 compatible wireless power amplifier (PA) into a transceiver chip using CMOS22 fabrication technology, to avoid the need for external front-end PA modules that increases the cost. The simulation results and the measurements are presented for the 5 GHz operating band.

The designed Wi-Fi PA employs a differential to single-end common-source cascode topology. Layout parasitics are considered using parasitics extractor and electromagnetic modeling. The performance is simulated using circuit simulator tools, and finally lab measurements are performed focusing on the 5 GHz operating band performance.

The simulated performance shows 26 dB power gain and a saturated output power of approximately 27 dBm. Modulation accuracy measurements for MCS7 (64-QAM) shows an error vector magnitude (EVM) of less than −31 dB at 16 dBm average output power.

Keywords:

RF power amplifier, integrated circuit, CMOS, Wi-Fi

CONTENT

LIST OF ABBREVIATIONS AND SYMBOLS 6

1 INTRODUCTION 10

2 TECHNOLOGIES FOR RF POWER AMPLIFIERS 12
2.1 Comparison of technologies 13
2.2 Advantage of silicon-based PAs 16
2.3 Semiconductor devices for amplifiers 17
2.3.1 MOSFET high-frequency model 24

3 PRINCIPLES OF WIRELESS RF POWER AMPLIFIERS 26
3.1 Introduction 26
3.2 Basics of linearity 27
3.2.1 Distortion 29
3.2.2 Gain compression 31
3.2.3 Linearity analysis 33
3.3 Intermodulation intercept point 36
3.4 Effects of nonlinearity in modulated system 39
3.5 Peak-to-average ratio 41
3.6 Figures of merit of Power Amplifiers 44
3.6.1 Output power and efficiency 44
3.6.2 Linearity performance 45
3.7 Stability 48
3.8 Amplifier categories 49
3.9 Power generation and power matching 53

4 DESIGN SPECIFICATION 57
4.1 Overall design 57
4.2 Modulation and radio spectrum specifications 58
4.3 Output power and linearity 60
4.4 Transmit spectral mask 61
4.5 Stability criteria 62

4.6 Summary of specifications for the PA 62

5 POWER AMPLIFIER DESIGN 64

5.1 Topology selection 64

5.2 Differential common-source 66

5.3 Amplifier core design 68

5.4 Bias circuit 72

5.5 Output matching network 74

5.6 Parasitic and electromagnetic modeling 76

5.7 Thermal considerations 77

6 RESULTS 79

6.1 Voltage gain 79

6.2 Power gain 81

6.3 Linearity and current consumption 82

6.4 Input/output impedance 85

6.5 Stability factor 86

6.6 Measurement results 87

7 CONCLUSIONS 92

REFERENCES 94

LIST OF ABBREVIATIONS AND SYMBOLS

A_{1dB}	Input amplitude at 1 dB compression point
AC	Alternating current
ACP	Adjacent channel power
A_{IIP3}	3rd order intercept point input amplitude
AM-AM	Amplitude to amplitude modulated distortion
AM-PM	Amplitude to phase modulated distortion
AP	Access Point
BG	Bandgap
BJT	Bipolar junction transistor
BT	Bluetooth
BW_{ch}	Channel bandwidth
C_{gd}	Gate to drain capacitance
C_{gs}	Gate to source capacitance
CMOS	Complementary Metal-Oxide-Semiconductor
C_{ox}	Oxide capacitance
CW	Carrier wave
DAC	Digital to analog converter
DC	Direct current
DQPSK	Differential quadrature phase shift keying
E	Electric field
EFF	Drain efficiency
EM	Electromagnetic
EVM	Error vector magnitude
f_{bb}	Baseband frequency
f_c	Carrier frequency
FEM	Front end module
f_{sc}	Subcarrier spacing
f_t	Unity-gain frequency
GaAs	Gallium Arsenide
GaN	Gallium Nitride
g_{ds}	Drain to source transconductance
GSM	Global System for Mobile Communications
IC	Integrated circuit
i_d	AC drain current
I_d	DC drain current
I_{DC_unit}	Unit transistor DC current

$i_{diff,max}$	Maximum differential current
I_{err}	In-phase error vector
i_g	Gate current
IL	Insertion loss in dB
IM_3	The relative intermodulation distortion
I_{rms}	Root-mean-square current
K	Rollet stability factor
L	Channel length
LNA	Low noise amplifier
LTE	Long term evolution
Mbps	Megabytes per second
MCM	Multi-chip module
MOSFET	Metal oxide semiconductor field effect transistor
MUX	Multiplexer
N	Transformer turn ratio
N_f	The number of MOSFET fingers
NMT	Nordic Mobile Telephony
N_p	Transformer number of primary turns
N_s	Transformer number of secondary turns
N_{sc}	Number of subcarriers
OFDMA	Orthogonal frequency-division multiple access
OIP3	3rd order intercept point output power
PA	Power Amplifier
P_{adj}	Adjacent channel power
PAE	Power added efficiency
$P_{ANT,MAX}$	Maximum antenna power
PAPR	Peak to average power ratio
PCB	Printed circuit board
P_{ch}	Transmission channel power
$P_{CORE,MAX}$	Maximum power delivered by the amplifier core
P_{DC}	DC power
P_{IN}	Input power
P_{OUT}	Output power
PPA	Pre-power amplifier
P_{sat}	Maximum saturated output power
PSD	Power spectral density
Q	Channel charge

QAM	Quadrature amplitude modulation
Q_{err}	Quadrature error vector
$Q_{L,UNIT}$	Channel charge per unit length
r_{ds}	Channel resistance
RF	Radio frequency
R_{gen}	Generator internal resistance
R_L	Load resistance
R_{LD}	Differential load resistance
R_{load}	Load resistance
RU	Resource unit
RX	Receiver
Si	Silicon
S_{ij}	S-parameter from *j*-port to *i*-port
SiO_2	Silicon dioxide
SOA	Safe operating area
t_{ox}	Oxide thickness
TX	Transmitter
u_n	Electron mobility
V_B	Bias voltage
v_d	Drain voltage
$V_{D,LOW}$	The lowest drain voltage
v_{dd}	Supply voltage
$v_{diff,max}$	Maximum differential voltage
v_{ds}	Drain to source voltage
v_e	Electron drift velocity
v_g	Gate voltage
v_{gs}	Gate to source voltage
V_i	Input signal
V_k	Knee voltage
V_o	Output signal
v_{ov}	Overdrive voltage
v_s	Signal voltage
v_t	Threshold voltage
W	Channel width
WBG	Wide bandgap
W_f	The width of a single MOSFET finger
Wi-Fi	Wireless network standard

Z_{gen}	Generator internal impedance
Z_{load}	Load impedance
ε_{ox}	Oxide permittivity
μ_L	Load plane mu stability factor
μ_S	Source plane mu stability factor

1 INTRODUCTION

Wireless local area network (Wi-Fi) is a common radio standard with more than 19 billion devices in use as of 2023. According to alliance 3.8 billion devices are forecast to ship in 2023 alone. The evolution of Wi-Fi standards, guided by the Institute of Electrical and Electronics Engineers (IEEE), traces back to 1997 and the first Wi-Fi standard 802.11 with a maximum data-rate of 2 Mbps. As the number of Wi-Fi devices and data traffic continue to grow globally the wireless standards have needed to evolve to keep pace with demands. With five generations in between the first standard 802.11 and Wi-Fi 6 which is also known as 802.11ax, Wi-Fi 6 is one of the most recent standards introduced by the alliance in 2019. Market has adopted Wi-Fi 6 devices faster than previous Wi-Fi generations due to the increased demand for high data-rates, low latency, and reliable connectivity in wireless devices such as phones, TVs, tablets, and access points (AP). Wi-Fi 6 is faster than its predecessor with a maximum theoretical throughput of 9.6 Gbps compared to the 3.5 Gbps of Wi-Fi 5. Wi-Fi 6 uses Orthogonal Frequency Division Multiple Access (OFDMA) and 1024 quadrature amplitude modulation (QAM) to improve multi-user capacity and spectral efficiency. Wi-Fi 6 products which can operate in the extended 6 GHz are known as Wi-Fi 6E. Due to the high data-rate capability of Wi-Fi 6, the standard sets challenging requirements for the transmitter system, especially in terms of the output power, linearity, and the frequency range in Wi-Fi 6E. Figure 1 shows a shipment forecast for the most recent Wi-Fi radio standards [1]

Each wireless communication system consists of several building blocks. What is common with all wireless communication systems, regardless of the standard or power class, is that signal fed to an antenna must be amplified to the correct power level while maintaining sufficient linearity. A power amplifier (PA) is usually one of the most critical parts of a wireless communication system as far as cost, power consumption, reliability, and system performance are concerned. The radio specifications for the entire transmitter system are defined by the applied radio standard. Many of these regulatory specifications define either

directly, or indirectly, the performance that is required from the PA. Most common specifications include design parameters such as output power, frequency range and spurious emissions.

This thesis studies the feasibility to integrate wireless local area network Wi-Fi power amplifier on a silicon transceiver chip using CMOS22 semiconductor technology. The transceiver is designed as a test chip for the client of this thesis to gain valuable information about the possible challenges and limitations for future Wi-Fi product development. Due to the electrical properties of silicon, the CMOS fabrication technology is not considered as the most optimal technology for a high output power PA [2]. The above-mentioned and the requirements by the Wi-Fi standard set a major challenge for the goal of this thesis. However, integrating a high output power PA on a single transceiver chip can serve as a great advantage for the client of this thesis in terms of simplicity and cost as the use of external PA modules could be avoided. This thesis focuses on achieving Wi-Fi 6 capable radio frequency (RF) PA using CMOS22 fabrication technology.

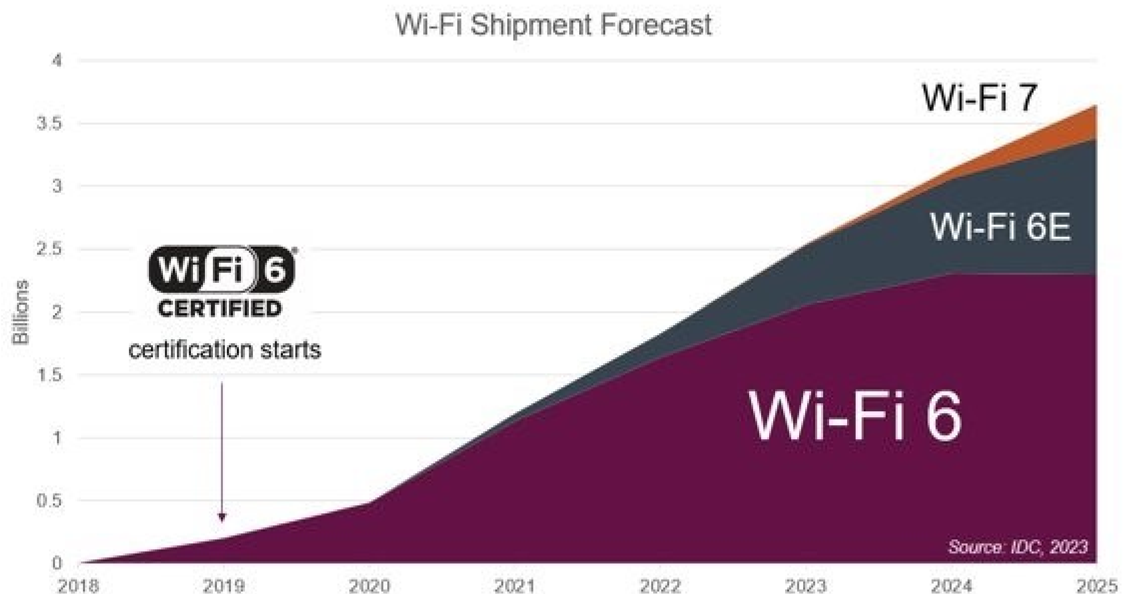

Figure 1. Wi-Fi shipment forecast for the most recent Wi-Fi radio standards [1].

2 TECHNOLOGIES FOR RF POWER AMPLIFIERS

The ever-growing amount of data transmission has pushed the technology vendors to research and investigate different semiconductor material options to improve the capacity and data-rates of integrated wireless electronics. There are several semiconductor technologies that are commonly used with wireless power amplifiers. Most common technologies that are used by the wireless semiconductor industry are Silicon (Si), Gallium Nitride (GaN) and Gallium Arsenide (GaAs). Electrical properties of the used technology usually have a major impact on the performance and the design of a PA. Several critical parameters are associated with technologies, including bandgap energy (BG) and carrier mobility.

Bandgap is the minimum amount of energy that is needed to excite an electron to break free from the valence band and move into the conduction band [3]. Semiconductor material with high BG value is called wide bandgap (WBG) semiconductor. High bandgap energy results in the high breakdown voltage for the device which is a great benefit for high-power amplifiers. High breakdown voltage allows the device to operate with high direct current (DC) supply, higher voltage swing and higher optimal load resistance for the amplifier. When the load value can be kept high, the amplifier currents are smaller, and this usually mitigates many possible problems such as electromigration and I^2R parasitic losses and self-heating in the chip. Additionally, the required transistor size scales down with the smaller required currents signal. Wide bandgap semiconductor can also maintain the electrical functionality at much higher temperature compared with typical BG semiconductor technologies [4]. This results in the fact that the use of wide bandgap technology is an advantage when dealing with an RF PA that typically operate at high power levels and at high temperatures.

Carrier mobility contributes to a speed at which the charge carriers (electrons or holes) move through the semiconductor channel when a voltage is applied. Higher carrier mobility leads to faster transistor operation and better overall

performance, including higher switching speeds and lower power consumption. Therefore, high carrier mobility contributes to higher feasible operating frequencies. [3]

2.1 Comparison of technologies

Silicon

Silicon based CMOS semiconductor technology has been a common choice for integrating electronics for a long time. The availability of *p*-type metal oxide semiconductor field effect transistors (MOSFET) makes it very suitable technology for a system on chip (SoC) that also includes digital design like processors, hardware state machined and memories etc. Integrated circuit (IC) can contain a lot of advanced analog and/or digital functionality, and the choice of used technology can be a compromise of many factors, including price, performance, and yield. [5]

Gallium Arsenide

Gallium Arsenide is a mixture of two elements that are gallium and arsenide that makes it as a compound semiconductor. Gallium Arsenide technology became common in power amplifiers primarily during the 1980s due to the advantages over other semiconductor materials in high power RF and microwave applications. GaAs based transistors are capable to operate up to 250 GHz frequencies due to the superior electron mobility. GaAs has a bandgap energy of 1.43 eV which is only slightly higher to that of Si. Typical applications using GaAs include solar cell, light emitting diodes (LEDs) and milli-meter wave integrated circuits (MMIC). GaAs has a thermal conductivity of only 50 $\mathrm{W/(m{\cdot}K)}$ which is less than half the thermal conductivity of GaN or Si. Low thermal conductivity can impose a challenge with applications generating a lot of heat. [6]

Gallium Nitride

Gallium Nitride is a relatively young semiconductor material in terms of its introduction. The Technology was introduced commercially in the 1990s but their impact on the commercially available amplifiers took place a little later [7]. Typical application with GaN is where high frequency operation and high-power capability is demanded. One of the main advantages of the GaN semiconductor material is its high electron mobility, high saturation velocity and high breakdown voltage making it a suitable technology for the development of high-frequency, high-power amplifiers for wireless communication systems, radar systems, and satellite communication. It also has a good operability at higher temperature levels. This makes GaN an attractive choice for the wireless PA design. The material also has superior tolerance against radiation which makes it a common choice for military and space applications [6]. High BG and electron velocity saturation allows GaN device to produce more power in less space compared to Si and GaAs resulting in high power densities. This makes it possible to produce high power devices in much less space if a way to remove the excess heat can be provided. GaN high-power amplifiers have been produced with more than 100 W of power over multioctave bandwidths and with power added efficiencies (PAE) of more than 60%. [7]

Table 1 summarizes the key parameters of Si, GaAs and GaN technologies. The first line lists that GaN has more than double the higher bandgap energy to that of Si or GaAs. The operational frequency limit of a transistor as well as the size required for a specific output current is fundamentally constrained by the transconductance characteristics of the device. Since the higher electron mobility contributes to a higher transconductance of a transistor depends on the electron mobility of its construction material, a high mobility is desired [8].

As mentioned earlier and indicated by the second line of the table, GaAs has more than six times higher electron mobility compared to Si and GaN. However, GaN's high electron velocity saturation can compensate its relatively small electron mobility in high-frequency and power applications [9]. Thermal conductivity is important factor for the heat dissipation. High linearity power

amplifiers usually have relatively weak efficiency which means that a certain amount of heat must be conducted away from the junction. GaN and Si seem to have comparable thermal conductivity performance. [7]

Table 1. Physical characteristics of different semiconductor materials. [9]

Parameter	**Si**	**GaAs**	**GaN**
Bandgap energy [eV]	1.12	1.43	3.40
Electron Mobility [$\mathrm{cm^2/Vs}$]	1350	8500	1200
Electron velocity saturation [$10^7\ \mathrm{cm/s}$]	1.0	1.3	2.5
Breakdown electric field [MV/m]	25	30	300
Thermal conductivity [$\mathrm{W/(m{\cdot}K)}$]	1.5	0.5	1.3

Based on the analysis with the table values, it can be determined that GaN is a very attractive choice of material for RF PA applications due to the high bandgap and high breakdown voltage. The significantly higher electron mobility of GaAs might suggest the selection of GaAs If the application operates at low power level while still high frequency operation is important.

Figure 2 shows the trending lines of the saturated output powers for the RF PAs using different semiconductor technologies. The semiconductor material study was concluded using data of more than 3800 scientific publications and papers since 2000. The trendline indicates the upper boundary for the maximum saturated output power P_{SAT} for each material. [10]

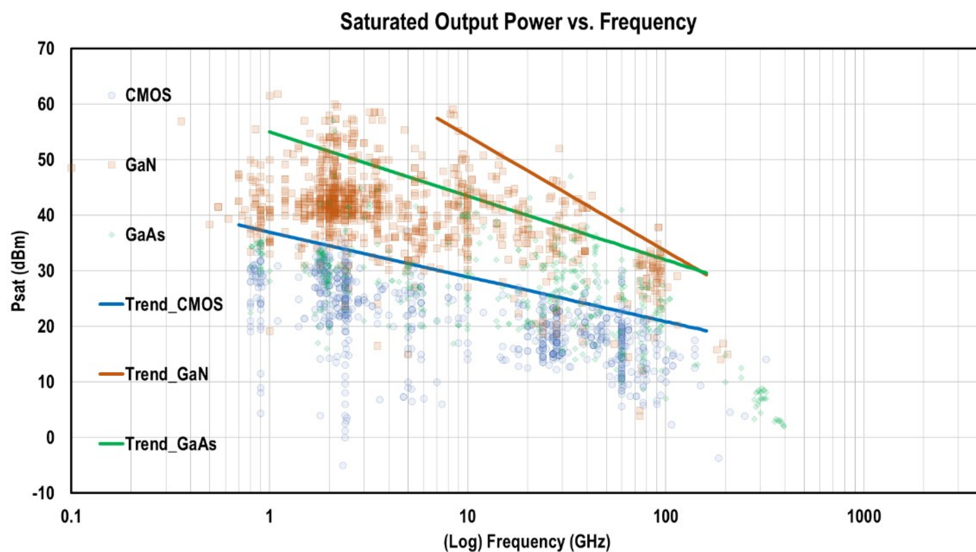

Figure 2. Trendlines for saturated output powers as a function of frequency for different semiconductor materials. [10]

2.2 Advantage of silicon-based PAs

The implementation of high-power RF PA for wireless applications, such as mobile handsets, has traditionally been dominated by III-V compound technologies like GaAs and GaN. This is due to their superior device characteristics when compared to silicon-based devices as discussed earlier in this chapter. Additionally, III-V compound processes provide low-loss substrates, high-quality metal layers, and through-substrate vias that support high-performance passive networks, high thermal conductivity, and minimized device grounding impedances.

Multi-chip module (MCM) is a common approach to assemble III-V compound PAs. For example, a GaAs die with the PA devices is combined with a CMOS IC as the PA controller chip, and passive networks implemented using either discrete components or integrated passive device chips. This solution offers an excellent trade-off between the overall cost and PA performance. However, the required footprint and package area of a such RF PA front-end can be higher than a fully integrated transceiver systems in a silicon die.

On the other hand, silicon-base CMOS PAs are becoming a competitive technology for the PA solutions of low-cost wireless mobile devices. The silicon CMOS process offers low-cost and high yield with mass-production capability. Furthermore, silicon-based PA naturally lends itself to a complete CMOS transceiver SoC solutions. With the capability of integrating billions of transistors on the same chip, silicon IC platform offers powerful on-chip computation and signal processing capabilities. [11]

2.3 Semiconductor devices for amplifiers

The main semiconductor device for amplifier application is a three terminal device that uses a principle of connecting a voltage between the two terminals to control current in the third terminal. This principle forms a basis for the switches, amplifiers, and signal amplification in general. There are basically two commonly used three terminal devices associated with amplifier design. These devices are MOSFET and bipolar junction transistor (BJT). The following section describes the fundamentals of MOSFETs like the physical structure and the operation principles. MOSFET is commonly used in the design of integrated RF transceivers and is also the main power component used in this thesis.

MOSFET transistor has become the most used component in electronic devices and especially with integrated circuits. The manufacturing process is relatively simple, and they can be made quite small compared to BJTs. Their power consumption is comparatively small due to the high impedance of the control terminal. Researchers and developers in the field have found many ways to utilize the use of MOSFETs in circuit topologies for digital and analog functions. Many functions can be implemented using almost entirely MOSFETs only or adding only a minimum number of other components like resistors. This makes it possible to pack huge number of MOSFETs on a single IC chip. A modern chip easily has multiple millions of transistors integrated. In 2022 Intel announced that it had identified materials and processes that pave the way for the aspirations of 1 trillion transistors on a package by 2030. [12] [3]

Triode/ohmic operating mode

Figure 3 shows the physical structure of an enhancement mode *n*-type MOSFET. The transistor is fabricated on a *p*-substrate that provides physical support for the circuit. Two heavily doped *n*-type regions are created in the substrate which are known as source and drain terminals respectively. The area between the source and drain regions forms the channel. Above the channel is the third terminal known as gate. Gate terminal is used to control the current in the channel and is electrically isolated from the channel by thin (typically 1…10 nm) layer of silicon dioxide (SiO_2). Even though the MOSFET is usually regarded as a three-terminal device, the substrate/body must also be tied to some potential and hence forms a fourth terminal.

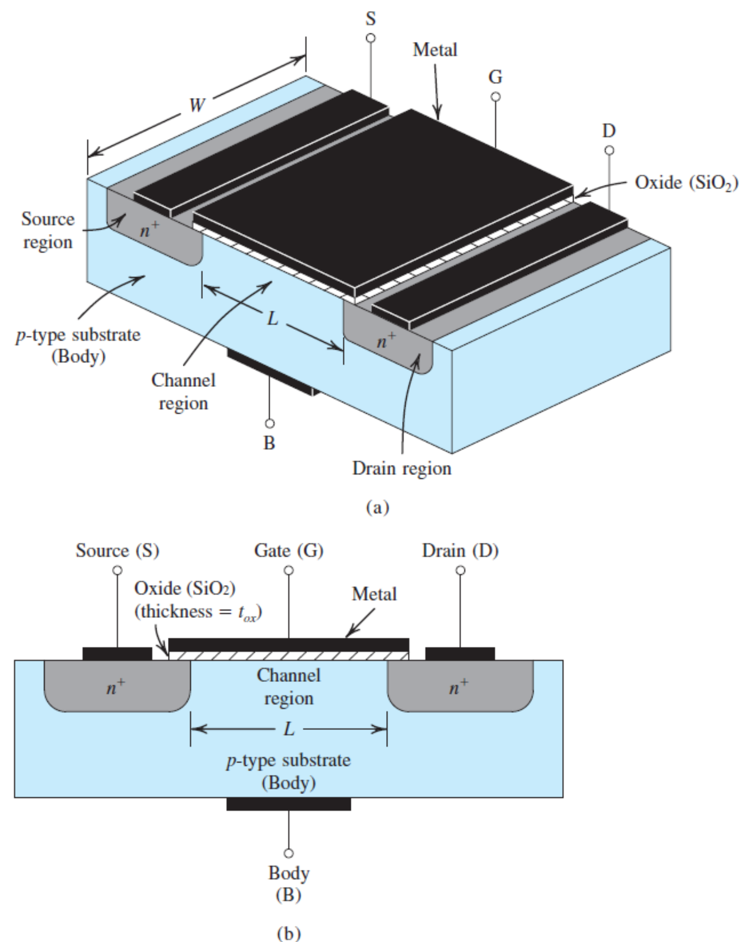

Figure 3. Physical structure of a MOSFET. (a) perspective view and (b) cross sectional view. [3]

When there is zero voltage applied between the gate and the source terminal, there are two back-to-back connected diodes between source and drain terminals that are formed by the *p*-type substrate and heavily doped *n*-type regions at the source and drain. The current in the channel is effectively blocked by these diodes making the resistance between the source and drain very high (in the order of $10^9\ \Omega$). Figure 4 depicts an example when the channel is formed. All the other terminals except the gate are grounded. When a positive voltage is applied between the gate and the source, the holes in the substrate are being repelled and pushed downward into the substrate as the minority charge carriers (electrons), and the majority charge carriers (also electrons) in the source and drain regions are attracted by the positive gate voltage and the channel is formed beneath the gate.

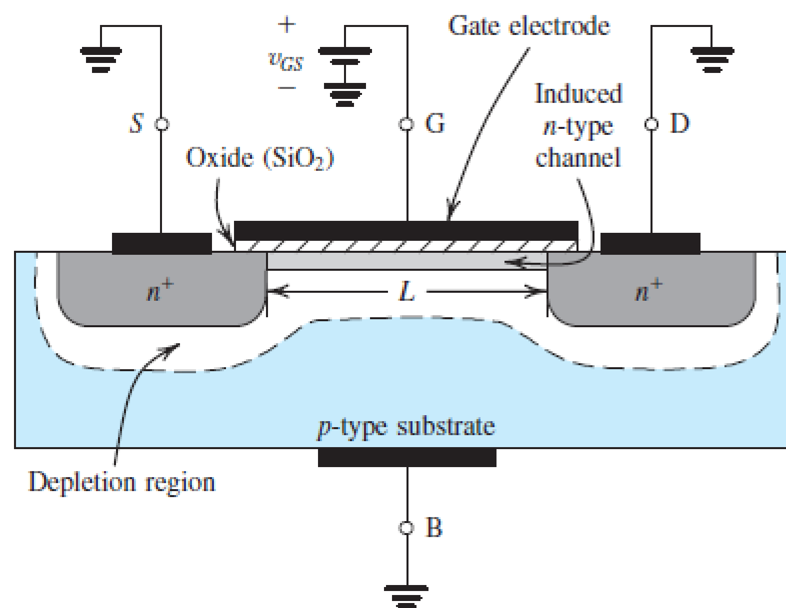

Figure 4. MOSFET with an *n*-channel formed below the gate. [3]

Phenomenon is often regarded as inversion when the *p*-type substrate below the gate inverts into an *n*-type channel. The gate terminal and the channel form a capacitor and the charge in the channel is formed by the electric field of this capacitor. Magnitude of the electron charge is expressed by the following equation

$$|Q| = c_{ox}WL(v_{gs} - v_t), \qquad (2.1)$$

where Q is charge in the channel, c_{ox} is the oxide capacitance, W is the channel width, L is the channel length, v_{gs} is the voltage between the gate and source, and v_t is the threshold voltage. The threshold voltage v_t is defined by the physical properties of a MOSFET and it is the minimum required gate to source voltage to induce a conducting channel between the source and drain terminals.

From Equation 2.1 it comes clear that the charge in the channel is proportional to $(v_{gs} - v_t)$. Capacitance C_{ox} is defined according to:

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}, \qquad (2.2)$$

Where, ε_{ox} is permittivity of the oxide, and t_{ox} is thickness of the oxide.

Oxide thickness is defined by the fabrication technology and is usually in the order on nanometers with current technologies. When the drain to source voltage, v_{ds} is assumed small enough the device is considered to operate in triode or ohmic mode and the voltage difference between the gate and the various points along the channel can also be assumed to remain nearly constants and the same as the voltage difference at the source end v_{gs}. When calculating the drain current, i_d it is interesting to define the charge per unit channel length. That can easily be obtained from equation 2.1 by simply dividing the total charge Q by the channel length *L*.

$$Q_{L,UNIT} = c_{ox}W(v_{gs} - v_t), \qquad (2.3)$$

where $Q_{L,UNIT}$ is charge in the channel per unit length.

Electric field E across the channel is simply v_{ds} over the length L. This field causes the electrons to drift from the source to drain with the following drift velocity v_e.

$$E = \frac{v_{ds}}{L}, \qquad (2.4)$$

$$v_e = u_n E = u_n \frac{v_{ds}}{L}, \quad (2.5)$$

where E is the electric field in the channel, v_e is the electron drift velocity, u_n is the electron mobility.

Finally, the drain current i_d can be obtained by multiplying the parameters $Q_{L,UNIT}$ and v_e. as follows:

$$i_d = [u_n c_{ox} \frac{W}{L}(v_{gs} - v_t)]v_{ds} \quad (2.6)$$

$$g_{ds} = \frac{i_d}{v_{ds}} = u_n c_{ox} \frac{W}{L}(v_{gs} - v_t) \quad (2.7)$$

$$r_{ds} = \frac{1}{g_{ds}} = \frac{1}{u_n c_{ox} \frac{W}{L}(v_{gs} - v_t)}, \quad (2.8)$$

where, i_d is the drain current, g_{ds} is the transconductance between the source and drain, r_{ds} the channel resistance between the source and drain.

To further investigate the equation for the transconductance g_{ds} it can be determined that it is a product of three factors. First part $u_n c_{ox}$, is defined by physical properties of the fabrication technology. Second part $\frac{W}{L}$, is simply the geometrics of the MOSFET and is up for the designer to choose. The last factor $(v_{gs} - v_t)$ is often called as an overdrive voltage v_{ov}, which is simply the amount of v_{gs}that exceeds the threshold voltage v_t.

There is, however, a minimum value for the channel length L that is defined by the technology. The number in the naming convention of a CMOS technology typically denotes the minimum channel length. For example, this thesis studies a PA that has been designed and fabricated using CMOS22 process meaning that a minimum possible value for L is 22 nm. Equation 2.6 shows that i_d is linearly dependent on the drain voltage v_{ds}, so it can be determined that in the triode/ohmic operation mode the device simply operates as a resistor whose value can be adjusted by the gate voltage. It is undesirable in amplifier design since the ideal goal is to control i_d by the changes of gate voltage only.

It must be, however, noted that Equation 2.6 is only true for small values of v_{ds}. As v_{ds} is increased the charge along the channel cannot be considered remaining constant and the same as at the source end of the channel. When v_{ds} is increased, i_d is obtained by:

$$i_d = u_n c_{ox} \frac{W}{L} [(v_{gs} - v_t) v_{ds} - \frac{1}{2} v_{ds}^2] \quad (2.9)$$

$$g_d = u_n c_{ox} \frac{W}{L} [(v_{gs} - v_t) v_{ds} - \frac{1}{2} v_{ds}^2$$

The linear operation mode voltage current characteristics are illustrated in Figure 5. The current i_d is plotted as a function of drain voltage v_{ds} for different v_{gs} values. Now the slope of each plot equals the transconductance g_{ds}.

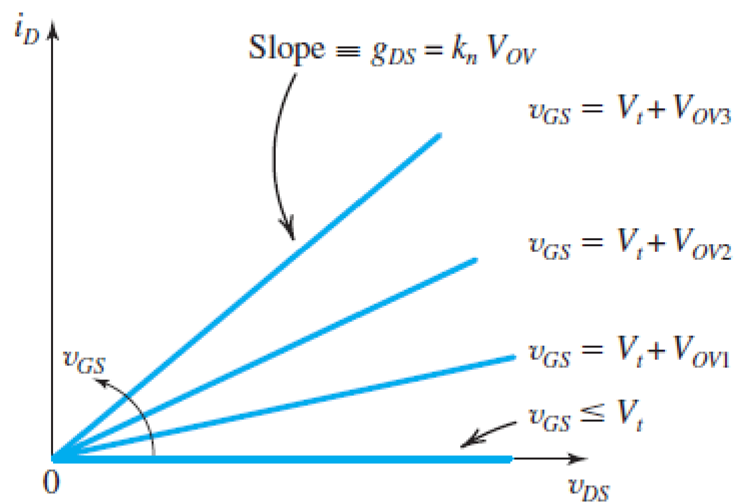

Figure 5. Drain current id is plotted as a function of drain to source voltage v_{ds} for different v_{gs} voltages.

Saturation operating mode

When v_{ds} is increased to a significant level, the voltage at the different points under the gate along the channel cannot be considered constant anymore. Now the voltage along the channel goes from v_{gs} at the source to $(v_{gs} - v_{ds})$ at the drain. Since the depth of the channel depends on the amount this voltage exceeds v_t, the channel shape is no longer uniform. The depth varies along the

channel being deeper at the source and shallower at the drain. This is illustrated by Figure 6 (a). The depth of the channel is proportional to the v_{ov}. As the voltage along the channel increases, the v_{ov} gets smaller and hence the depth of the channel becomes tapered. When v_{ds} reaches v_{ov}, the gate to drain voltage v_{gd} equals threshold voltage v_t and in this condition the channel depth ate the drain is reduced to zero. This term is often referred as channel pinch-off. This is illustrated in Figure 6 (b). Any additional increase in v_{ds} has no effect in the channel charge or i_d and the current is controlled by the gate voltage only. It is desirable in PA design to keep the transistor in saturation mode when i_d is dependent on the gate voltage only. This forms a basis for a load-line theory that aims to define a load value that keeps the amplifier in saturation mode during the entire input/output voltage swing. Equations for the drain current i_d and transconductance g_m is defined as follows:

$$i_d = \frac{1}{2} u_n c_{ox} \frac{W}{L} \left(v_{gs} - v_t\right)^2 = K\left(v_{gs} - v_t\right)^2, \tag{2.10}$$

$$g_m = \frac{\partial i_d}{\partial v_{gs}} = u_n c_{ox} \frac{W}{L} \left(v_{gs} - v_t\right) = 2K\left(v_{gs} - v_t\right) = 2\sqrt{K i_d} \tag{2.11}$$

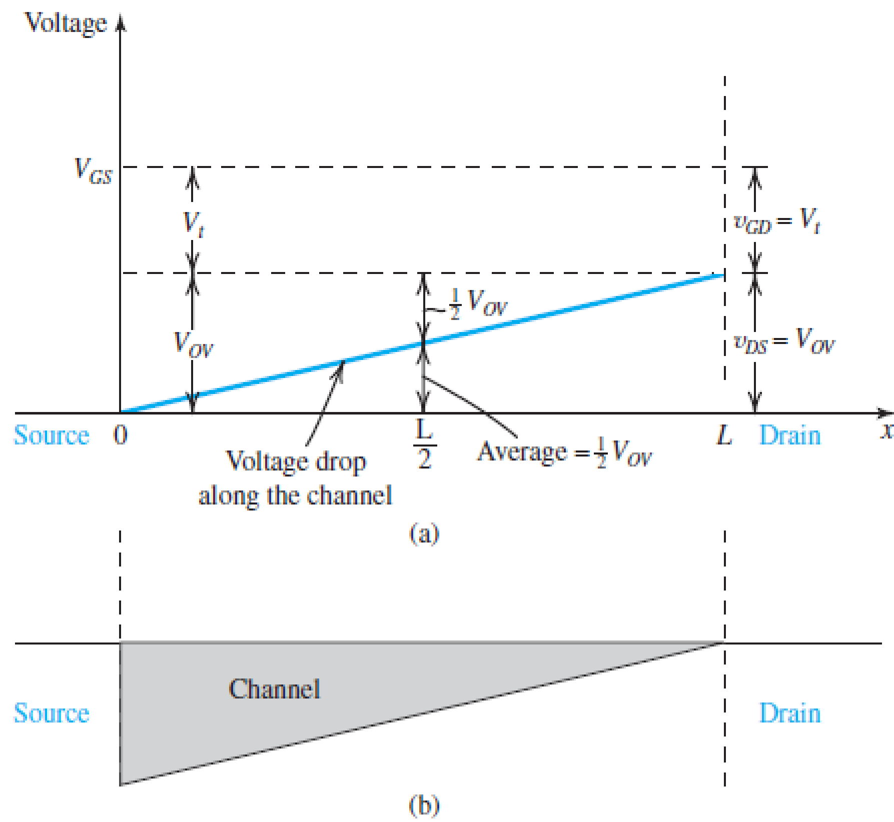

Figure 6. (a) Voltage drop along the channel varies linearly when going from source towards the drain. In the middle point the voltage drop is $0.5v_{ds}$. Channel becomes pinched-off when reaching the drain since the $(v_{gs} - v_{ds})$ equals v_t. (b) The shape of the channel corresponding the situation in (a). The depth of the channel is proportional to v_{ov}. [3]

2.3.1 MOSFET high-frequency model

A high frequency small-signal model of a MOSFET is depicted in Figure 7. The capacitances C_{gs} and C_{gd}represent gate-to-source capacitance, and gate-to-drain capacitance respectively. If the drain is short-circuited and the gate is driven with a current source i_g, the drain current is

$$i_d(s) = g_m v_{gs}(s) - sC_{gd}v_{gs} \approx g_m v_{gs}(s), \tag{2.12}$$

and the gate current using v_{gs} is

$$i_g(s) = s(C_{gs} + C_{gd})v_{gs}(s) \tag{2.13}$$

Using the equations for $i_d(s)$ and $i_g(s)$, and setting the complex laplace variable $s = jw$ the current gain A_i for the MOSFET is

$$A_i(jw) = \frac{g_m}{s(C_{gs} + C_{gd})}. \tag{2.14}$$

The unity-gain frequency f_t at which the magnitude of the short-circuit gain is equal to 1

$$f_t = \frac{g_m}{2\pi(C_{gs} + C_{gd})} = \frac{\sqrt{\frac{1}{2} u_n c_{ox} \frac{W}{L} I_d}}{\pi(C_{gs} + C_{gd})}, \tag{2.15}$$

where I_d is the DC operating current. The unity-gain frequency f_t is a figure-of-merit of high frequency operations of a MOSFET. However, the weakness of f_t is the neglected gate access resistance that effects the high-frequency gain. [13]

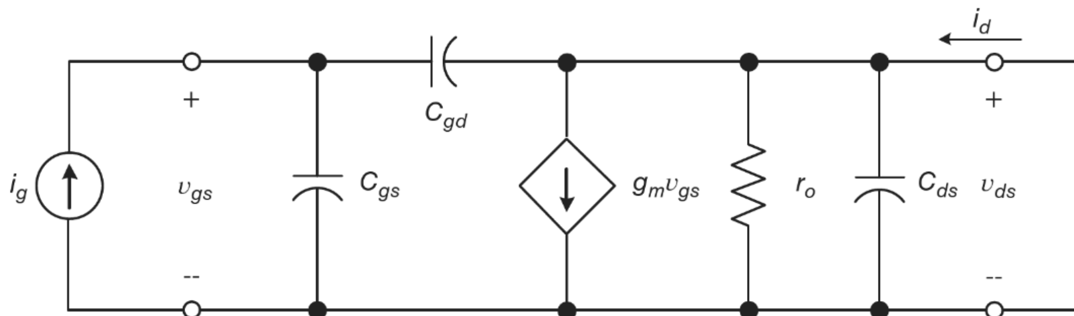

Figure 7. A high frequency model of a MOSFET. [13]

3 PRINCIPLES OF WIRELESS RF POWER AMPLIFIERS

3.1 Introduction

The amount of mobile data being transmitted is continuously increasing as the number of mobile devices and users continue to increase. To keep up with this demand, manufacturers are forced to develop products that are both energy-efficient and capable of high data throughput. However, due to limited radio frequencies, they must strive for maximum data throughput per hertz. There are various technologies available, including the Long-Term Evolution (LTE), Wi-Fi, and Bluetooth (BT), each with their own methods of modulating and demodulating signals.

As the amount of wireless data traffic continues to grow, the already-limited wireless frequency spectrum becomes increasingly congested. In the past, frequency modulation was a popular method of wireless communication (e.g., radio, Nordic Mobile Telephone (NMT), and Global System for Mobile Communications (GSM)), due to its ability to withstand noise, interference, and nonlinearities [14]. However, the nonlinear operation results in the spectral regrowth that combined with a relatively small data rate of frequency modulation, results in poor spectral efficiency (i.e., data rate vs. bandwidth). [15]

Variable amplitude modulation methods are becoming increasingly popular in modern telecommunications systems and standards due to their ability to provide better spectral efficiency compared to frequency and phase modulation methods [16]. These methods typically use both amplitude and phase to transmit the message and are usually implemented through baseband quadrature signals. This technique significantly increases the data-carrying capacity per signal bandwidth. Variable amplitude modulation uses a technique of storing information in the amplitude of the signal. Because of that, these methods are highly susceptible to disruptions that affect the amplitude of the signal, such as nonlinear amplification. These disruptions not only cause data errors but also disrupt the adjacent channels by spreading the signal spectrum. This imposes a problem when multiple devices transmit signals in the adjacent frequency channels.

Wi-Fi and many other communication systems such as LTE and TV broadcasting, are utilizing variable amplitude modulation in which the signal comprises multiple subcarriers that are N-QAM-modulated. A higher modulation order result in higher spectral efficiency but also results in a higher linearity requirement that is needed from the power amplifier. Even a constant amplitude modulation scheme becomes variating in amplitude when pulse shaping filtering is applied to limit the bandwidth. Variating amplitude always means spectral growth when fed through a nonlinear amplifier. [17]

The primary function of an RF power amplifier is to linearly amplify radio signals for transmission. Meeting the required radio standard specifications involves achieving crucial performance parameters such as peak power, efficiency, gain, error vector magnitude (EVM) and adjacent channel power (ACP). Achieving high linearity, high output power and stability design of linear, sets a challenge and finding a balance between the most critical performance parameters is important in PA design.

Typically, the best efficiency is achieved at peak power, while the worst linearity is also seen at peak power. The high linearity is trade-off with efficiency. Achieving overall power efficiency can be challenging in multicarrier amplitude modulated systems since the signal remains below peak power most of the time due to the high peak to average power ratio (PAPR).

3.2 Basics of linearity

To calculate and analyze how a nonlinear amplifier effects the system, it must first be modelled somehow. The best method of course is to use transistor-level spice models by the semiconductor technology providers. However, they are not always available for commercial power amplifiers and simulations with transistor-level models can be heavy and time-consuming. In addition, theoretical calculations with transistor-level models can be tricky and unintuitive.

Non-linear effects in amplifiers are usually expressed as a black box device with an assigned power series transfer function. Figure 8 shows a case when input signal $V_i(t)$ is fed to a nonlinear device. [18]

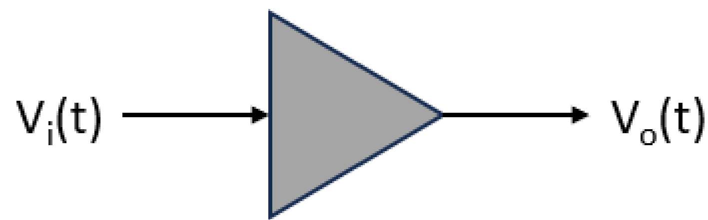

Figure 8. Nonlinear device with input and output signals V_i and V_o respectively.

Power amplifier models can be divided in two different categories called memoryless models and models with memory. Memoryless models use assumption that the previous input values cannot affect the present or future output values. The memoryless modelling divides the distortion in amplitude-to-amplitude (AM-AM) distortion and amplitude-to-phase distortion (AM-PM) components. The AM-AM describes how the output amplitude envelope expand/compress and the AM-PM describes the phase shift of the signal as a function of the input signal envelope [17].

Models with memory take into account the rate of change of the modulating signal. Memory effect can be considered as a bandwidth dependent distortion. Typical single-tone AM-AM and AM-PM laboratory measurement has a zero bandwidth that fails to model the bandwidth dependent distortion [19]. However, a simple way to discover memory effect is to analyze the dependence of the intermodulation products on the spacing of a two-tone signal tones [20].

Distortion components around the harmonics of the RF carrier can easily be attenuated and filtered away so modelling the behavior around the carrier is usually enough. With a passband model approach the modelling is done using real-valued signal that operates at carrier frequency. When also baseband frequencies are present in the signal it can be a time-consuming simulation, and therefore, a typical method is to use complex-valued baseband equivalent model that significantly reduces the frequency range and simplifies calculations. [17]

3.2.1 Distortion

From theoretical perspective a system:

$$V_o(t) = aV_i(t), \tag{3.1}$$

where $V_o(t)$ is the output signal and $V_i(t)$ is the input signal. The output signal $V_o(t)$ is linear if a is a constant [21]. In practice it means that the output signal is simply a scaled version of the input signal. Unfortunately, when working with physical components like MOSFETs this is not the case. A simple way to model a nonlinear system is using power series polynomial expansion. The output $V_o(t)$ presents an infinite series of nonlinear products as shown by the following equation.

$$V_o(t) = a_1V_i(t) + a_2V_i^2(t) + a_3V_i^3(t) + \cdots (etc.) \tag{3.2}$$

This power series is generally used to define nonlinear behavior even though it has some limitations. The most significant limitation is the absence of timing/phase information that would require the use of a more accurate equation with complex power series coefficients. Another limitation is a sensitivity to any changes of the input/output impedance and bias point. However, the power series in (3.2) is still useful in the small operating zone around the DC bias point.

The solid line in Figure 9 shows an example of strongly nonlinear characteristics of a MOSFET. The drain current is cut-off below the threshold point and exhibits linear behavior until the channel is fully open with no changes in the drain current by any further change of gate voltage. The characteristics is highly idealized and presents sharp transitions in both limiting regions. In addition, the signal is infinitely linear in the middle. More realistic characteristics can be modelled with the following 3rd order equation and is shown by the dashed plot in Figure 9:

$$i_d = 3v_g^2 - 2v_g^3 \tag{3.3}$$

It must be noted that this is not a full power series model, but it aims to model the behavior between the cut-off and the saturation point. With 3rd order model in (3.3) the drain current i_d is not physical to a MOSFET with v_g values outside 0…1, and thus higher order terms would be required to account for the area outside the range. A combination of both curves would again result in the more realistic model, but it would require a power series of very high order to mathematically express that and the calculation for the basic theory of nonlinearity would become complex and unpractical.

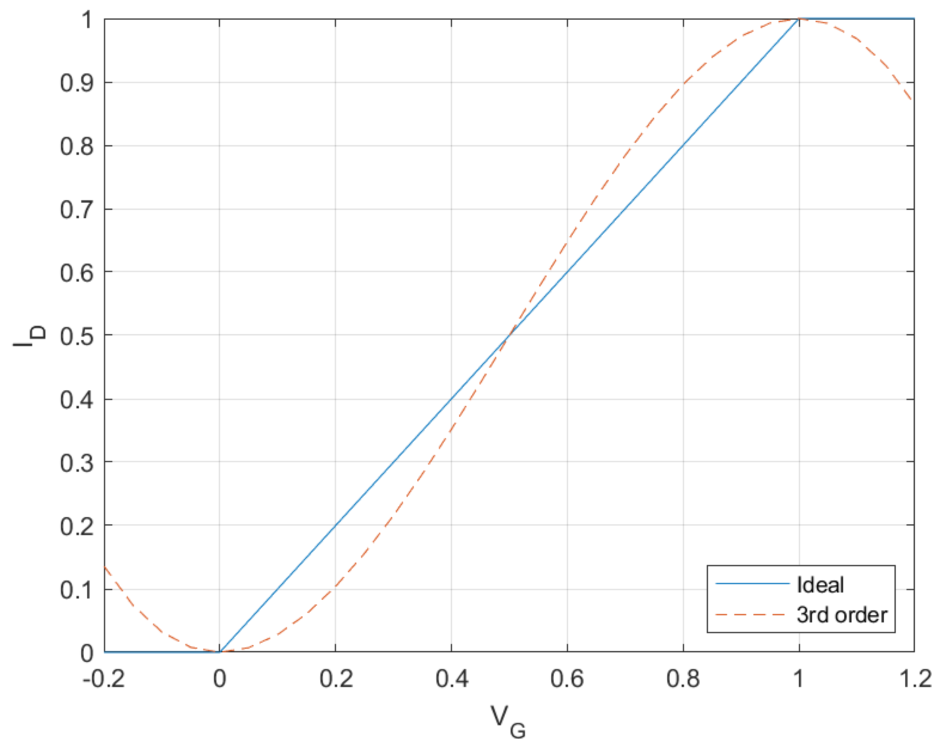

Figure 9. Ideal strongly nonlinear model and more realistic weakly nonlinear model.

Input voltage is a combination of gate bias voltage V_B and signal $v_s(t)$. Substituting $v_g = V_B + v_s(t)$ in (3.3) yields:

$$i_d = (3V_B - 2V_B) + (6V_B - 6V_B^2)v_s(t) + (3 - 6V_B^2)v_s^2(t) - 2v_s^3(t) \qquad (3.4)$$

If the result is compared with power series equation in (3.2) it can be noted that the coefficients for a_n can be extracted and are the following:

$$a_0 = (3V_B - 2V_B)$$

$$a_1 = (6V_B - 6V_B^2)$$

$$a_2 = (3 - 6V_B)$$

$$a_3 = (-2)$$

It can be observed that the small signal gain a_1 is dependent on the DC bias point. This is intuitive considering the derivative of the curve that clearly changes along the v_g. Small signal gain peaks in the middle and fades near the cutoff or saturation. Second-order nonlinearity coefficient a_2 is also dependent on the bias point and has a null point exactly at $V_B = 0.5$. Third order nonlinearity coefficient a_3 is constant –2. This is due to the selection of 3rd order power series only. In practice there are higher order terms that would affect when approaching and reaching the hard saturation area. Third order model is simple and useful to describe the behavior when operating at backed off input levels away from hard saturation. This makes it feasible to use in linearity analysis. [18]

3.2.2 Gain compression

When $v_s(t)$ in the power series model (3.2) is replaced with a sinusoidal tone $A\cos(wt)$ the output voltage $v_o(t)$ becomes:

$$v_o(t) = a_0 + a_1 A\cos(wt) + a_2 A^2 \cos(wt)^2 + a_3 A^3 \cos(wt)^3$$

$$= a_0 + \frac{a_2 A}{2} + (a_1 A + \frac{3a_3 A^3}{4})\cos(wt) + \frac{a_2 A^2}{2}\cos(2wt) + \frac{a_3 A^3}{4}\cos(3wt) \quad (3.5)$$

As can be seen the output signal $v_o(t)$ contains the harmonic components of the fundamental tone w. In the example the polynomial contains coefficient up to 3rd degree term, but in reality, the terms would continue up to infinity. However, 3rd order polynomial is often useful and enough to describe nonlinearity up to certain point. Fundamental tone amplitude becomes $(a_1 A + 3a_3 A^3/4)$ and dividing that with input amplitude A results in the gain of the

fundamental component $(a_1 + 3a_3A^3/4)$. An interesting note can be made when examining the fundamental gain. If coefficients a_1 and a_3 are of different sign (e.g $a_1 > 0$ and $a_3 < 0$), the magnitude of the gain decreases as the magnitude A of the fundamental tone increases. This is called gain compression and a typical figure of merit for linearity is the input or output power at which the gain has reduced 1 dB. Input voltage level A_{1dB} for 1 dB gain compression can be found using:

$$a_1A_{1dB}10^{-0.05} = a_1A_{1dB} + \frac{3a_3{A_{1dB}}^3}{4}$$

$$A_{1dB} = \sqrt{\frac{4a_1(1 - 10^{-0.05})}{3a_3}} \quad (3.6)$$

(a) The phenomenon is further illustrated in

Figure 10. In (a) the input/output power is plotted on a logarithmic scale. The difference in the output power between the linear and non-linear case is exactly 1 dB at the compression point. The same is depicted in (b) for the gain of the linear and non-linear case.

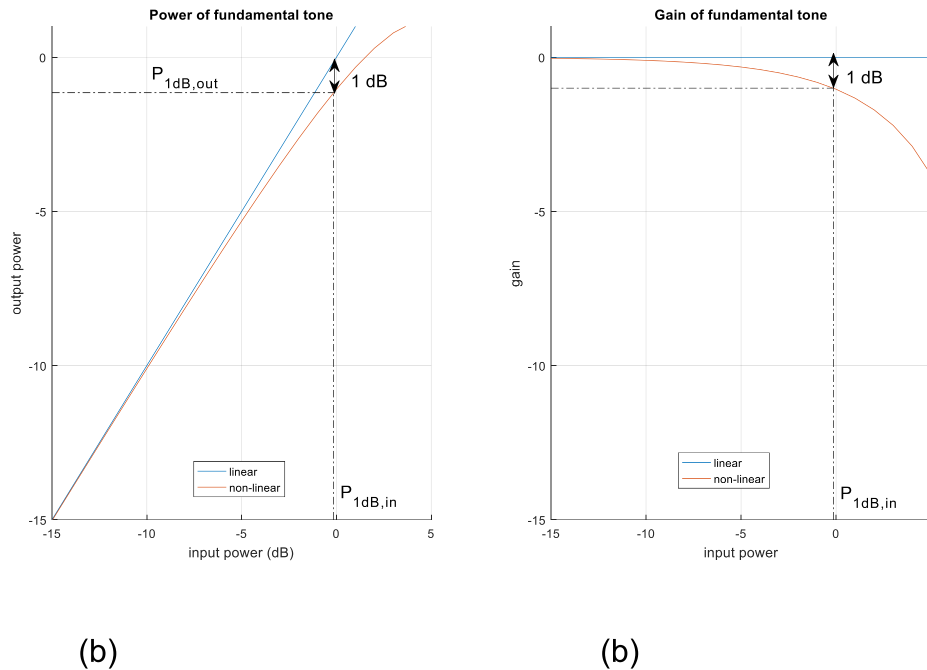

Figure 10. Dashed marker indicates 1dB compression point with respect to input/output power (a) and gain (b).

3.2.3 Linearity analysis

As discussed in Chapter 3.1 the efficiency and effects of nonlinearity is significantly dependent on the used modulation and PAPR of the signal. Traditional method to analyze nonlinear characteristics and to obtain some figure of merit for it is to perform two-tone analysis. It's an act of applying two different frequency tones to a nonlinear system. This way the output of the system does not only contain the harmonics of the input tones but also additional intermodulation components that are results of raising the sum of two signal to a power greater than one. Input signal

$$v_i(t) = A\cos(w_1 t) + A\cos(w_2 t) = 2A\cos\left(\frac{w_1 - w_2}{2}t\right)\cos\left(\frac{w_1 + w_2}{2}t\right) \qquad (3.7)$$

can be considered as an amplitude modulated signal with a fixed modulating frequency. If the angular frequency of the carrier tone is $w_{rf} = (w_1 + w_2)/2$ and the angular frequency of the modulating fixed baseband tone is $w_{bb} = (w_1 - w_2)/2$, Equation (3.7) becomes

$$v_i(t) = 2A\cos(w_{bb}t)\cos(w_{rf}t), \quad (3.8)$$

envelope of the signal reaching $2A$. Using the power series model, the output $v_o(t)$ becomes the following:

$$\begin{aligned} v_o(t) = {} & a_1A(\cos(w_1t) + \cos(w_2t)) + \\ & a_2A(\cos(w_1t) + \cos(w_2t))^2 + \\ & a_3A(\cos(w_1t) + \cos(w_2t))^3 + \\ & a_4A(\cos(w_1t) + \cos(w_2t))^4 + \\ & a_5A(\cos(w_1t) + \cos(w_2t))^5 + \cdots \end{aligned} \quad (3.9)$$

As discussed in Chapter 3.2 the set of coefficients $a_1 \ldots a_n$ are valid for some single operating point and frequency only. Once again, the 3rd order model is usually enough to describe the nonlinear behavior when operating well behind the saturating, or above cutoff regions of the voltage-current curve. Therefore, for the sake of clarity, trigonometric relationships are used to evaluate the distortion components caused by the power series terms up to the 3rd order only. Each degree of distortion generates number of additional distortion components which have either the same or lower order. For example, 3rd degree term generates components at $2w_1 \pm w_2$, $2w_2 \pm w_1$, $3w_1$ and $3w_2$. Expanding (3.9) results in the additional spectrum components. The following represents a list of generated components by a 3rd order power series model.

DC: a_2A^2

$\mathrm{w_1, w_2}$: $(a_1A + \frac{9a_3A^3}{4})\cos(w_1t) + (a_1A + \frac{9a_3A^3}{4})\cos(w_2t)$

$\mathrm{2w_1, 2w_2}$: $\frac{a_2A^2}{2}\cos(2w_1t) + \frac{a_2A^2}{2}\cos(2w_2t)$

$w_1 \pm w_2$: $$a_2 A^2 \cos((w_1 + w_2)t) + a_2 A^2 \cos((w_1 - w_2)t)$$

$2w_1 \pm w_2$: $$\frac{3a_3 A^3}{4}\cos((2w_1 + w_2)t) + \frac{3a_3 A^3}{4}\cos((2w_1 - w_2)t)$$

$2w_2 \pm w_1$: $$\frac{3a_3 A^3}{4}\cos((2w_2 + w_1)t) + \frac{3a_3 A^3}{4}\cos((2w_2 - w_1)t)$$

$3w_1, 3w_2$: $$\frac{a_3 A^3}{4}\cos(3w_1 t) + \frac{a_3 A^3}{4}\cos(3w_2 t)$$

The fundamental tones w_1, w_2 are often called the “wanted” components that were also present in the input signal. The second order nonlinearity components at $2w_1$,$2w_2$ and $w_1 \pm w_2$. Usually, the even order nonlinearities land out of band and don’t present a major concern for the adjacent channels.

Of particular interest are the intermodulation components $2w_1 \pm w_2$ and $2w_2 \pm w_1$. The intermodulation components land in the immediate vicinity of the fundamental tones w_1, w_2. Since the radio spectrum is tightly packed the intermodulation is troublesome in wireless systems: the neighboring channels will experience interference by these intermodulation components that corrupt the information carrying frequency components. Additionally, there is also an in-band distortion component

$$\frac{9a_3 A^3}{4},$$

that interfere the wanted components w_1, w_2. The sign of a_3 is usually negative so the in-band distortion component can be seen as compressing the gain of the wanted fundamental components in a similar way as with single-tone analysis in Chapter 3.2.2. Other components that are further away from the fundamental tones are usually attenuated by natural frequency filtering of matching network circuits or even implementing analog filters for those harmonic frequencies are feasible. Third order intermodulation figure of merit is usually called IM_3 and it is defined as the ratio of wanted fundamental component and unwanted intermodulation component in decibels.

$$IM_3 = 20\log\left(\frac{\left|\frac{3}{4}a_3A^3\right|}{\left|a_1A + \frac{9}{4}a_3A^3\right|}\right)\text{dBc} \quad (3.10)$$

When operating with input levels that are small and away from the hard saturation area, then $\left|a_1\frac{A}{2}\right| \gg \left|\frac{9}{32}a_3A^3\right|$ the equation becomes:

$$IM_3 \approx 20\log\left(\frac{3}{4}\left|\frac{a_3}{a_1}\right|A^2\right)\text{dBc}, \quad (3.11)$$

Equation shows that the distortion depends on the square of the amplitude as well as the distortion coefficient a_3. An interesting observation that can be made is that IM_3 always increases twice the increase of A in decibels: If A increases 3 dB, IM_3 increases 6 dB.

3.3 Intermodulation intercept point

Another performance indicator that is often associated with amplifiers is the input level A_{IP3}, when the fundamental and the third order intermodulation product are equal (i.e. $IM_3 = 0\ \text{dBc}$):

$$|a_1|A_{IIP3} = \frac{3}{4}|a_3|A_{IIP}^3 \quad (3.12)$$

$$A_{IIP3} = \sqrt{\frac{4}{3}\left|\frac{a_1}{a_2}\right|} \quad (3.13)$$

In practice when the input level is increased the gain of the fundamental component begins to compress and higher order terms that exist in practice become significant. In practice, the intermodulation component would never reach the fundamental tone in magnitude. It is often practical to define the output power level P_{OIP3} for this condition, which is simply the output power level when $A = A_{IP3}$. Practical method for defining the output power intercept point, $OIP3$ is to measure/simulate the output power level P_{OUT} and corresponding

IM_3at small signal level and utilize interpolation on a logarithmic scale or using equation:

$$OIP3 = P_{OUT} - \frac{IM_3}{2} \tag{3.14}$$

Combining (3.6) and (3.13) gives

$$\frac{A_{IIP3}}{A_{1dB}} = \sqrt{\frac{1}{1 - 10^{-0.55}}}, \tag{3.15}$$

Which results in the ratio of 9.6 dB. So, the A_{IIP3} is about 10 dB higher than the single-tone compression point. This is equivalent to 13 dB difference if the total power is used as the reference in the two-tone case. [18] [21]

Figure 11 shows an FFT of a two-tone analysis when signal is corrupted by a nonlinear component that can be a power amplifier or any other nonlinear block of a radio system. Wanted spectrum components are at $w_{rf} \pm w_{bb}$. In addition to wanted components there are unwanted intermodulation products at $w_{rf} \pm 3w_{bb}$. Figure 12 is an example AM-AM characteristics of a PA that has a linear gain of 25 dB. The orange dashed curve presents the increasing 3rd order intermodulation component. The Intercept points as well as 1 dB compression point and the saturated output power level are marked in the figure.

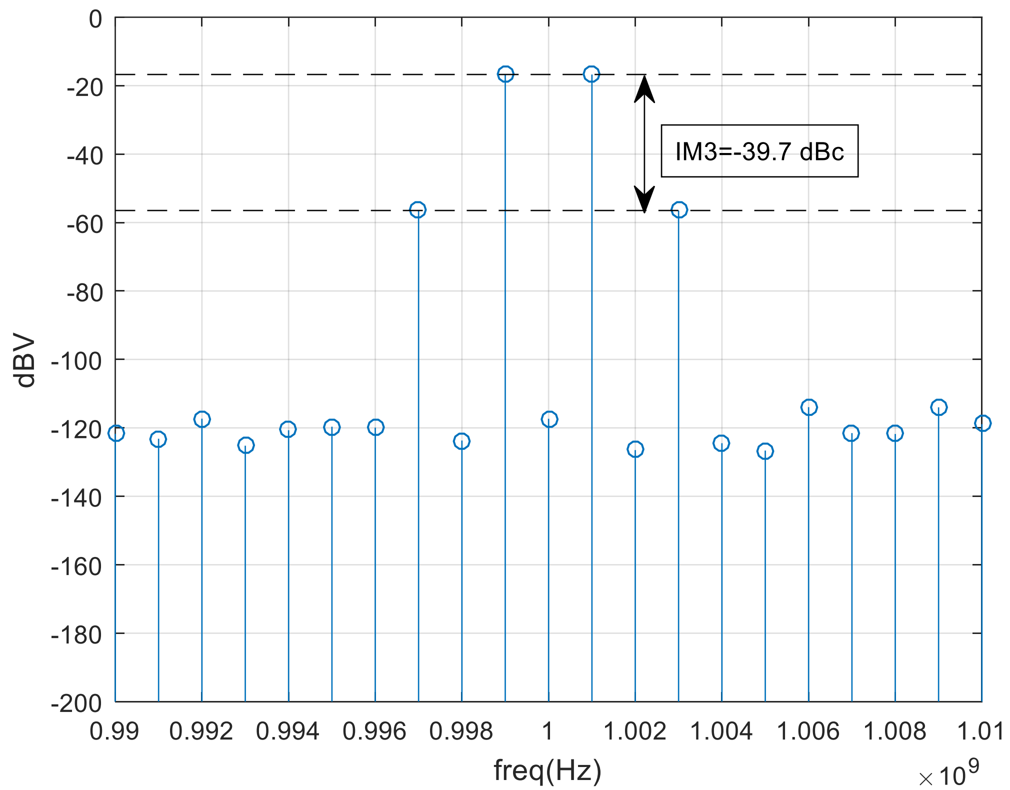

Figure 11. An FFT of a two-tone signal. Now amplitude $A = 0.1$ and polynomial coefficients $a_1 = 1.5, a_3 = -2$. Additional spectrum components are formed at frequencies $2w_1 \pm w_2$ and $2w_2 \pm w_1$.

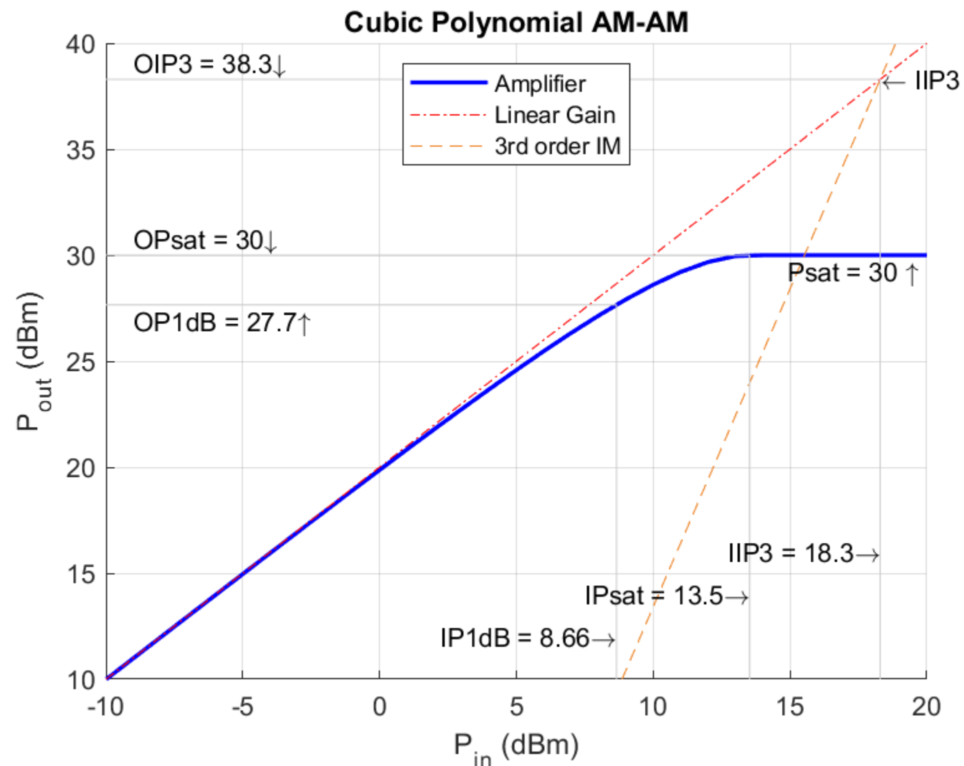

Figure 12. An example AM-AM characteristics of a PA that has a linear gain of 25 dB. The input and output power levels are marked for the intercept point, 1 dB compression point, and for the saturated power level.

3.4 Effects of nonlinearity in modulated system

Shape/size of the spectrum is defined by the used modulation and bandwidth, but third order nonlinearity products cause spurious emissions in the adjacent channel reducing ACP performance as well as degradation of EVM since the transmitted original signal is now modified/corrupted by the nonlinearities of the system. As pointed by equation 3.11, IM performance is a function of the envelope amplitude. The previous two-tone example has an envelope amplitude of a sinusoidal form. However, variable amplitude modulation methods, while sensitive to nonlinearities, are common in modern communication systems. Examples of such methods include N-QAM, which is utilized for example in digital TV broadcasting, LTE and Wi-Fi. In addition to these methods, constant amplitude modulation schemes also translate into variable amplitude signals when filtered to limit the bandwidth. An example of this is a DQPSK-modulated

signal that is filtered with a root-raised cosine filter. These methods do not suffer from data corruption due to distortion but are still susceptible to spectral spreading, which can interfere with signals on adjacent channels. [15]

The RF signal of variable amplitude modulation can be described with

$$v(t) = A(t)\cos\left(2\pi f_c t + \varphi(t)\right) \quad (3.16)$$

, where the information is carried by the variable amplitude $A(t)$ and the variable phase $\varphi(t)$. Using the Euler's formula [22]:

$$e^{ix} = \cos(x) + i * \sin(x)\,, \quad (3.17)$$

it is possible to represent the carrier signal in equation (3.16) as follows:

$$v(t) = A(t)\cos\left(2\pi f_c t + \varphi(t)\right) = \mathrm{Re}\left\{A(t)e^{i\left(2\pi f_c t + \varphi(t)\right)}\right\} =$$

$$\mathrm{Re}\left\{A(t)e^{i(2\pi f_c t)}e^{i(\varphi(t))}\right\} = \mathrm{Re}\left\{B(t)e^{i(2\pi f_c t)}\right\}, \quad (3.18)$$

Where $B(t) = A(t)e^{i(\varphi(t))}$ is the modulating baseband signal. The baseband signal can be considered as a product of two phasors that manipulate the carrier signal: $A(t)$ is real and modulates the amplitude of the carrier and $e^{i(\varphi(t))}$ is a complex unit vector with phase $\varphi(t)$ modulating the phase of the carrier. Sometimes it is practical to represent the baseband signal $B(t)$ as cartesian in-phase (I), and quadrature (Q) components:

$$B(t) = I(t) + jQ(t). \quad (3.19)$$

The modulating baseband signal points to a constellation point that indicates a certain bit pattern. Figure 13 shows an example of 16-QAM constellation. Each cross in the picture represents a symbol that can decode four bits. Orange original symbols are being transmitted. Blue symbols exhibit situation when distortion is introduced and the crosses on the constellation spread and might fall outside the decision region resulting in bit error. Gain compression or AM-AM distortion causes the blue crosses to fall short on the target points as the modulating signal gets larger and the distance from the origin increases.

Similarly, AM-PM distortion corrupt the symbol as the phase $\varphi(t)$ of the carrier signal changes as a function of amplitude $A(t)$. On the constellation it is seen as a fluctuation of phase. This is one of the reasons variable amplitude modulations are pose a challenge in nonlinear systems. [15]

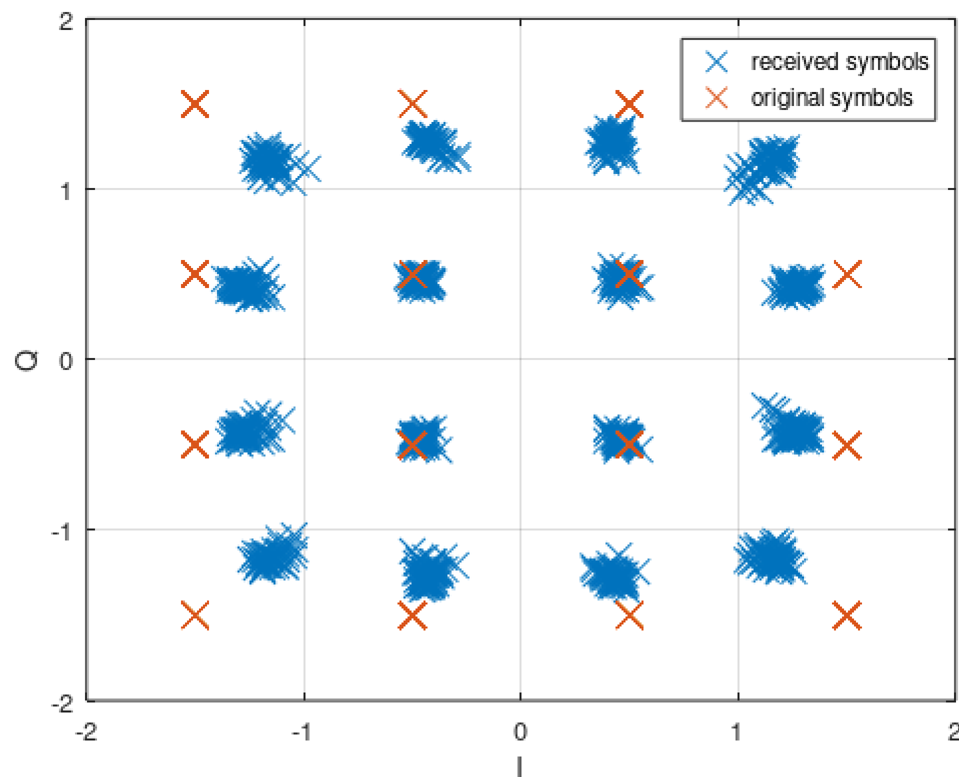

Figure 13. Constellation of 16QAM. Original symbols are transmitted through a nonlinear system that results in the spreading of the received symbols.

3.5 Peak-to-average ratio

As pointed out in this chapter the amplitude of the signal affects distortion level of a nonlinear system. Different signal types have different amplitude profiles that depends on the modulation scheme. If for some specific signal type, the probability of amplitude to change is small then the distortion of such signal is smaller compared to that of signal type with large and frequent variations. A figure commonly used with communication systems describing the change of amplitude is called peak to average power ratio (PAPR)

$$PAPR = \frac{\max(s(t))}{\frac{1}{t}\int |s(t)|^2 dt}, \quad (3.20)$$

where $s(t)$ is the analysed signal. High PAPR value indicates that the signal has smaller amplitude values most of the time but has random higher peaks. If two signals are normalized to have same average power, the signal with higher PAPR has higher amplitude spikes and hence introduces more distortion when driven through 3rd order nonlinearity. Another problem with signals of high PAPR is the low efficiency operation of the amplifier because one must operate at back-off power levels to meet the linearity requirements. High PAPR is a common challenge with radio systems like Wi-Fi utilizing orthogonal frequency-division multiple access (OFDMA), that uses a technique of transmitting data on parallel equally spaced orthogonal frequency bins that are often called subcarriers. Figure 14 illustrates the OFDMA spectrum usage of Wi-Fi 6. A single 20 MHz channel can have a total of 256 subcarrier tones at a spacing of 78.125 kHz. [23]

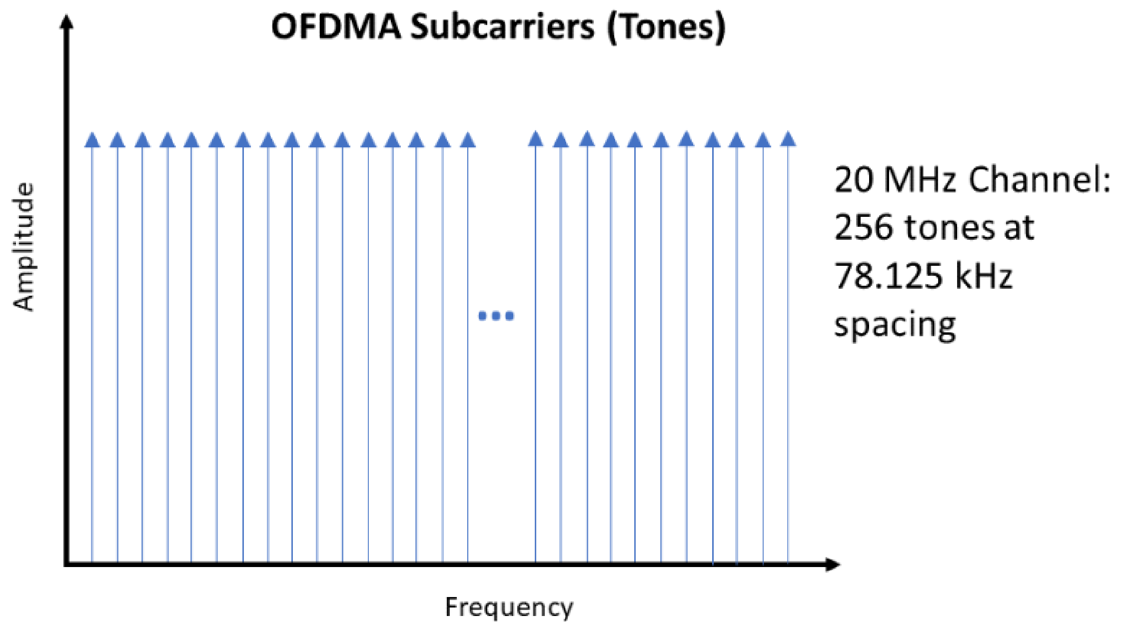

Figure 14. OFDMA in Wi-Fi 6. 20 MHz channel can have a total of 256 subcarriers at a spacing of 78.125 kHz. [24]

The reason to use high PAPR signals is the need to improve the spectral efficiency to transmit as many bits as possible within a certain bandwidth. Spectral efficiency is increased at the cost of increased PAR. Table 2 shows spectral efficiencies with PAPRs for several different modulations. With N-QAM signals the PAPR increases with modulation index N. It is worth noting that variating amplitude/power is not limited to variable amplitude modulations only. Phase modulated signals have constant amplitude only before filtering.

Table 2. Spectral efficiency and PAR for different modulations. RRC filter with roll-off factor of 0.5 is applied to signals. [25]

Method (RRC, roll off=0.5)	**QPSK**	**16-QAM**	**64-QAM**	**256-QAM**
Spectral efficiency $\frac{bit}{s \cdot Hz}$	1.3	2.7	4	5.3
PAPR	3.1	5.2	5.9	6.2

3.6 Figures of merit of Power Amplifiers

The following section briefly discusses the most common parameters that are used to describe the performance of an RF amplifier.

3.6.1 Output power and efficiency

Most common performance indicator for a power amplifier is usually output power. However, as discussed in 3.2 the linearity is significantly dependent on the output power. It means that stating an output power value often refers to a power level resulting in some specific linearity performance. When comparing output power performance of amplifiers, one must take the fact into account. More explicit parameter is to use saturated power P_{sat} that is the power level when the amplifier is in deep compression and any increase in the input power results on no change in the output power. This of course means that the output spectrum presents significant harmonic content.

Another important parameter for high power amplifiers is efficiency. It simply describes the amount of DC power that is converted into RF power. The efficiency can be expressed using the equation:

$$EFF = \frac{P_{OUT}}{P_{DC}}, \tag{3.21}$$

where P_{OUT} is the power delivered to the load and P_{DC} is the total power taken from the DC supply. With power amplifiers it is also common to talk about the power added efficiency (PAE) that takes the input signal power P_{IN} into account:

$$PAE = \frac{P_{OUT} - P_{IN}}{P_{DC}}, \tag{3.22}$$

Typical mobile device can operate at a transmitting power level between 20…30 dBm, hoverer, the building blocks responsible for modulating or performing the conversion of the signal can often only operate at power levels less than 1 mW. So, it is obvious that a significantly high-power gain is required from the

following blocks that can be made up of multiple cascaded amplifiers. Total efficiency of cascaded amplifiers can be calculated as:

$$EFF_{tot} = \frac{1}{\frac{1}{EFF_1 * G_2} + \frac{1}{EFF_2}}, \qquad (3.23)$$

Where EFF_1 and EFF_2 are the efficiencies of the of the first and the second stages and G_2 is the gain of the second amplifier stage. The equation reveals that the total efficiency of a system is dominated by the last amplifier stage. If both EFF_1 and EFF_2 are 30% and G_2 is 15 dB, the overall efficiency is 29 %. If EFF_1 is increased to 50 %, the overall efficiency is increased less than 0.5 %. This proves that to improve the total power efficiency of a system, most of the design efforts should be put into the last stage and especially into the trade-off between the efficiency and linearity of the stage. [19]

3.6.2 Linearity performance

Intermodulation test using a two-tone input is a simple way to analyze the linearity performance of a device. Parameters such as IM_3 and P_{OIP3} studied in chapter 3.2 were used when designing and analyzing the linearity of the PA designed in this thesis. A Two-tone simulation is a simple and relatively fast way to obtain the parameters with circuit simulator tools.

However, simple two-tone test doesn’t model the performance as it would be with a real modulated signal that has a spectrum characteristic like white noise. The unwanted effect of intermodulation is the power leakage or interference at the adjacent channels. This effect is characterized by the adjacent power ratio (ACPR or ACP):

$$ACPR = 10\log\frac{P_{adj}}{P_{ch}} \text{ dBc}, \qquad (3.24)$$

Where P_{ch} is the power of the transmission channel and P_{adj} is the power of adjacent channel at some specific offset from the transmission channel. The

measurement is outlined in Figure 15. The actual measurement bandwidths and offsets from the carrier frequency is defined by the radio standard specifications. Figure 16 depicts an intermodulation spectrum of a band-limited digitally modulated signal showing that spurious spectrum components extend to three times the bandwidth of a modulating signal with 3rd order coefficients and five times the bandwidth of a modulating signal with 5th order coefficients. [26] [18].

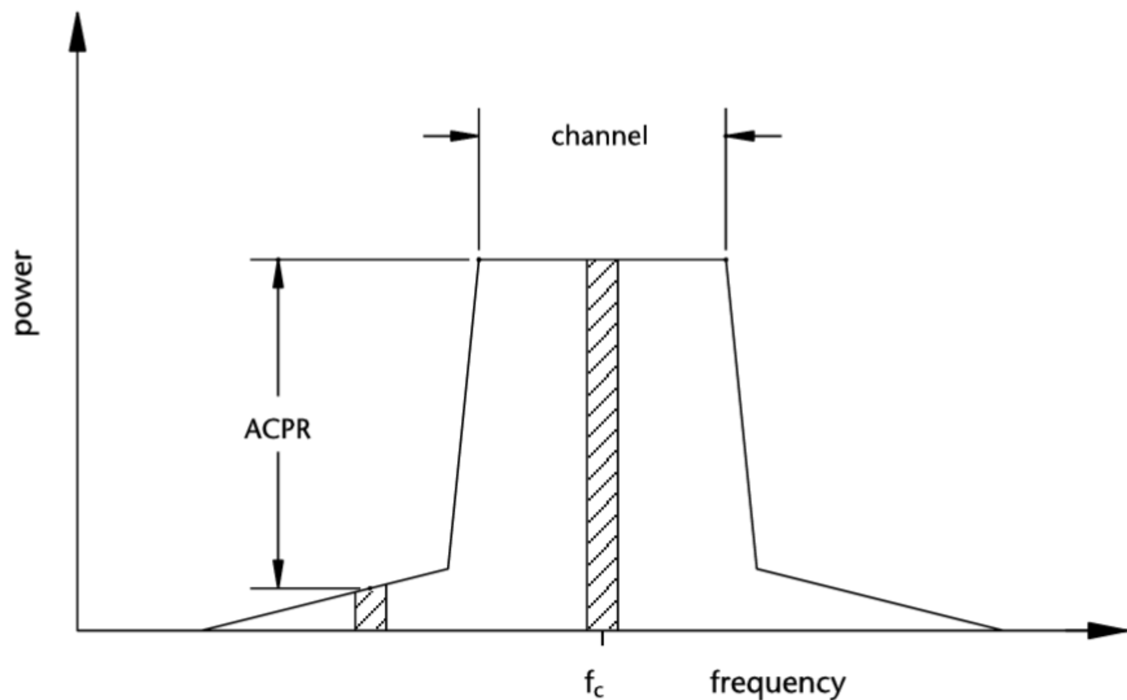

Figure 15. ACPR measurement. The ratio of power in a certain bandwidth centered at a specific offset from the carrier (f_c) to the power in a certain bandwidth centered at the f_c. [26]

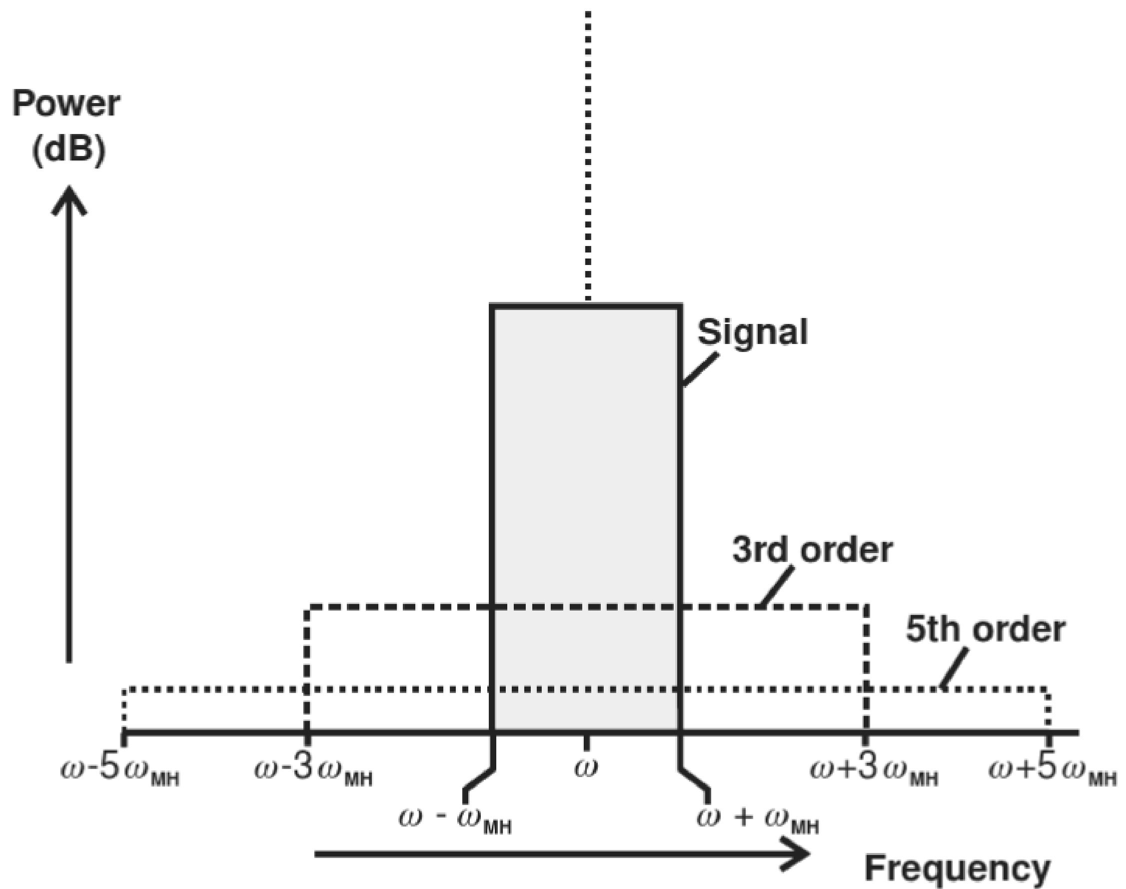

Figure 16. Intermodulation spectrum of a band-limited digitally modulated signal. [18]

Error Vector Magnitude (EVM) is a crucial metric used to quantify the accuracy of transmitted data. It measures the difference between the ideal constellation points (expected signal) and the actual received constellation points (received signal). In simpler terms, EVM assesses how much the transmitted signal deviates from its intended state due to noise, interference, distortion, or other impairments in the communication channel. A lower EVM indicates higher signal quality and better performance of the wireless system. EVM is often analyzed as an RMS value over multiple symbols and subcarriers and is expressed either in percentages or decibels. [27]

$$EVM_{\%} = \frac{\sqrt{\frac{1}{N}\sum_{n=0}^{N-1} I_{err}(n)^2 + Q_{err}(n)^2}}{REF_{rms}} \cdot 100\ \%, \qquad (3.25)$$

where $EVM_{\%}$ is the measured error vector magnitude, N is the total number of measured symbols, n is the symbol index, I_{err} and Q_{err} are the in-phase and quadrature components of the error vector, and REF_{rms} is the RMS value of the reference symbols.

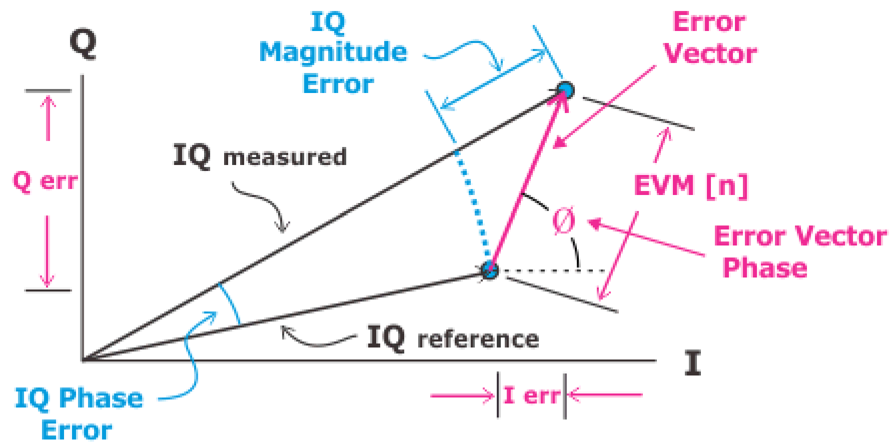

Figure 17. Illustration of an EVM measurement. [27]

3.7 Stability

It is common for the PA circuits to experience instability which can be detrimental to their performance. Therefore, it is important to analyze the stability factor of each stage and the overall cascaded stages in both differential mode and common mode. The presence of parasitic inductances in the bias, supply, and ground path can introduce additional poles which can lead to an unstable circuit. To avoid this, the design should aim to reduce the impact of these parasitic effects and maintain a sufficient stability margin. [11]

Defining the stability of a two-port amplifier network is crucial in PA electrical engineering. The K factor, introduced by Rollett, is commonly used as a stability criterion in designing microwave and RF amplifiers and is defined with the help of scattering parameters as follows:

$$K = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2 * |S_{21}S_{12}|}, \quad (3.26)$$

$$\Delta = S_{11}S_{22} - S_{12}S_{21}, \quad (3.27)$$

Where K, is the Rollett stability factor, S_{11} and S_{22} are the input and output port reflection coefficients respectively, $S_{21}$and S_{12} are the forward and reverse transmission coefficients respectively, and Δ is the determinant of the 2-port scattering parameter matrix. [26]

The *μ* factor is an alternative method to check the stability information across range of frequencies. It evaluates stability differently by concentrating on the circuit's input or output plane stability. The *μ* factor is particularly beneficial for circuits where unilateral assumptions are not valid. The *μ* factor can be calculated for source and load plane based on the S-parameters of a two-port network, like the Rollett stability factor K. [28] [29]

$$\mu_L = \frac{1 - |S_{11}|^2}{|S_{22} - S_{11}^*\Delta| + |S_{21}S_{12}|}, \quad (3.28)$$

$$\mu_S = \frac{1 - |S_{22}|^2}{|S_{11} - S_{22}^*\Delta| + |S_{21}S_{12}|}, \quad (3.29)$$

where μ_L is the load plane stability that indicates the distance from the center of smith chart to the nearest unstable region. The same applies for the source plane μ_S. Values $\mu_L, \mu_S > 1$ indicates that the nearest unstable region is outside smith chart meaning that no passive impedance termination can cause oscillation in the device under test.

3.8 Amplifier categories

Different types of PAs can be categorized in terms of their linearity and efficiency. Class A, B, and AB amplifiers are examples of categories that are fairly linear, but they only have low or moderate efficiencies. These amplifiers

conduct current for at least half of the input signal cycle. Class C amplifiers sacrifice linearity for efficiency by reducing conducting time. Class D amplifiers use switching to rapidly turn the output signal between some predefined voltage levels. This way it produces square waveforms of current and/or voltage. These PAs can achieve very high efficiency but at the cost of linearity. The thesis focuses on A, and AB PAs, which are commonly used in the field of RF design and with high-order modulation schemes due to their relatively high linearity. [18]

It is a common belief to consider class A amplifier as a linear device, especially when it comes to RF power amplifiers. However, this is not always the case. Class A amplifiers are not always linear, and highly linear amplifiers are not always Class A. The classical definition of Class A is easy to understand from Figure 18. The solid line describes ideal and perfectly linear voltage-current characteristics. If the bias point is selected in a way that the input signal never reaches beyond the saturation limits (0…1), then the output signal is perfectly linear version on input signal. This is the classical definition of class A amplifier that in ideal case the amplifier is linear for this limited range. In practice, of course, the weak nonlinearities exist, and more realistic voltage-current characteristics is presented by the dashed line. The output current will present harmonic spectrum content as the drive level is increased towards the hard saturation region. Output is usually coupled with a matching network to transform the impedance from 50 Ω termination into a correct level to meet the linearity requirements. The matching network made up of reactive components usually has some filtering and the harmonic components are hence attenuated. However spectral spreading problem arises when an amplitude modulated signal is applied and intermodulation products are generated in the output spectrum. Intermodulation products are discussed in chapter 3.2. However, class A amplifiers are generally considered cleaner spectrum than class AB amplifiers that are utilizing strongly non-linear regions of the voltage-current characteristics to improve efficiency. Class A amplifier has lowest efficiency, and the DC current consumption remains the same regardless of the signal size. Maximum theoretical efficiency for class A amplifier is 50 %. Class B

amplifier is biased at the cutoff point and the DC current with zero input signal is virtually zero. Only the positive half cycle of the input signal swing results in output current. Class B amplifier exhibits significant even-mode harmonic distortion. Maximum theoretical efficiency for class B amplifier is 78.5 %. [18]

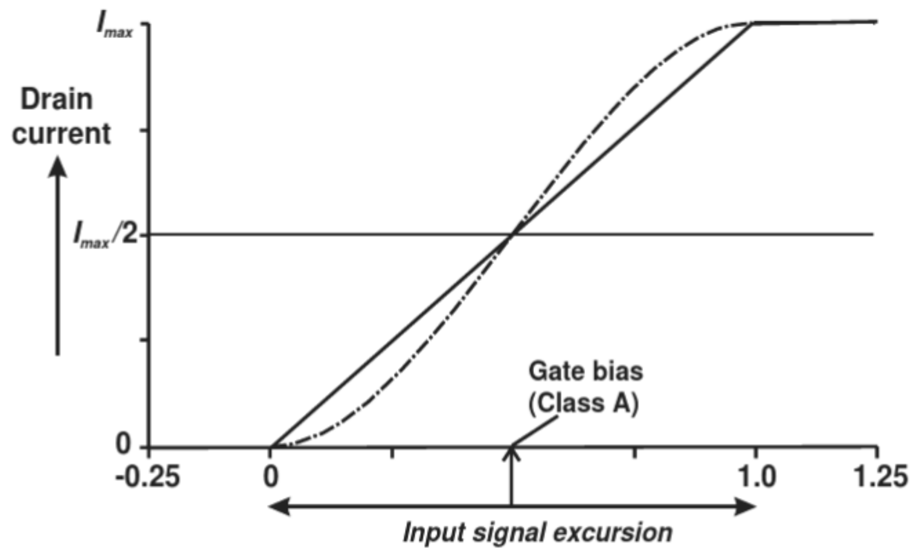

Figure 18. Voltage-current characteristics and bias point of a class A amplifier. [18]

Figure 19 shows a load line and current waveform representation for different amplifier categories and Table 3 shows the PA classification according to conduction angle and the trend of few key parameters such as efficiency, gain, and linearity for different amplifier classes.

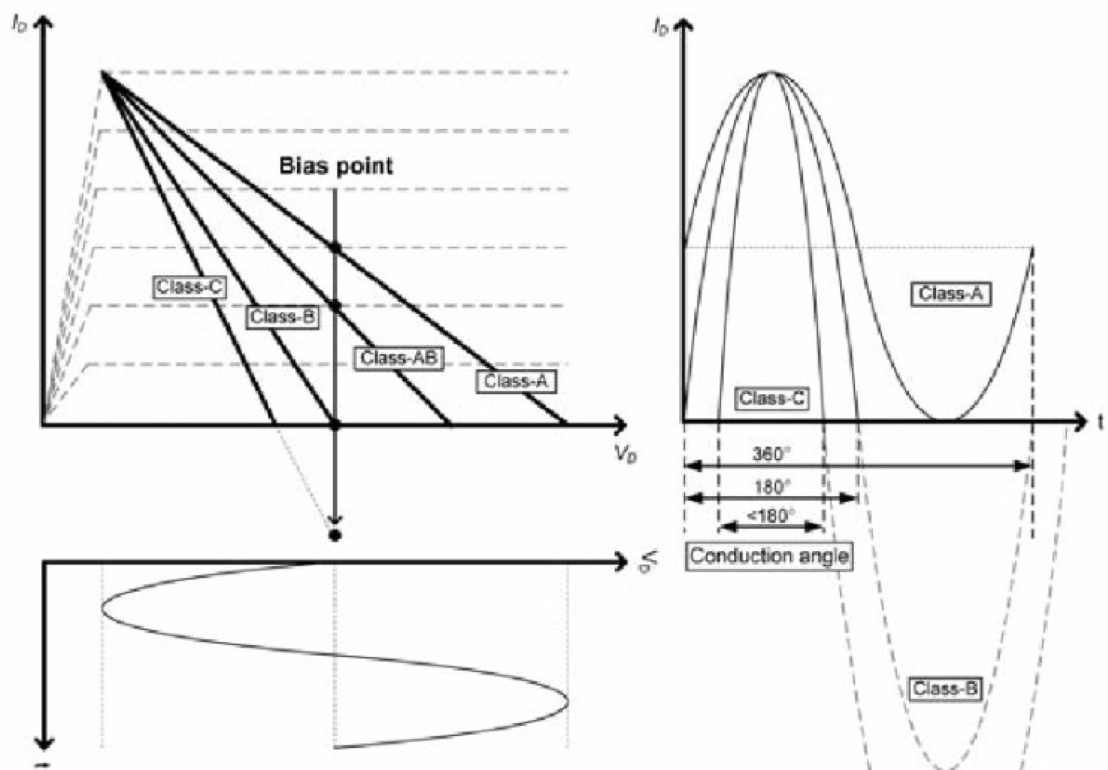

Figure 19. Load lines and conduction angle waveforms for different amplifier classes. [11]

Table 3. PA classification according to conduction angle

Class	Conduction Angle	Bias Point	Efficiency (Max Theoretical)	Gain	Linearity	Output Power (Normalized)
A	360°	Higher than V_T	50%	Largest	Good	1
AB	360°–180°		50–78.5%	↓ Lowest	↓ Bad	Larger than 1 (max 1.15 at 240°)
B	180°	V_T	78.5%			1
C	180°–0°	Lower than V_T	78.5–100%			1 at 180° 0 at 0°

3.9 Power generation and power matching

The concept of gain match and power match refers to the design of output matching network seen by the amplifier. Figure 20 shows an example of a generator with an internal impedance of R_{gen}. The generator can deliver the maximum power to a load impedance of $R_{load} = R_{gen}$. If a generator impedance Z_{gen} presents reactive part, the load impedance Z_{load} must be a complex conjugate of Z_{gen} ($Z_{load} = Z_{gen}*$).

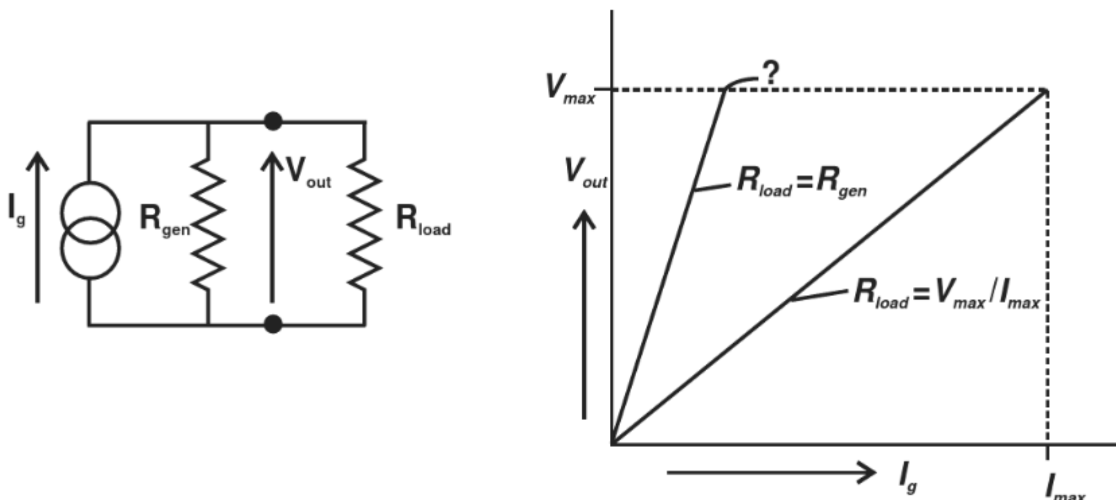

Figure 20. If the current generator in the picture is terminated with $R_{load} = R_{gen}$, the voltage limit V_{max} in its output terminal is reached much before the maximum current supply capacity I_{max} reached and in this case the power capacity of the source is wasted. [18]

This theory can be applied when working with amplifiers as the amplifier in its ideal form can be considered as a voltage controlled current source. However, when working with real physical devices there is a limitation to the theory. If, as an example, a generator with $R_{gen} = 100\ \Omega$, that can supply a maximum current of 1 A is connected to a load impedance of 100 Ω, then the maximum power transfer is achieved but the generator output voltage is as high as 50 V. It is very unlikely that a real amplifier structure could deliver such high voltage at its drain terminal as the voltage is strictly limited by the DC supply for the transistor. In this case the output current would show a limiting behavior much

before the rated maximum current I_{max}. The performance of the amplifier is not being used to its full capacity which is undesirable. A lower value for R_{load} would be required to maximize the power capacity. Figure 21 shows the characteristics for a load value of R_L that aims for maximizing the power generation capability.

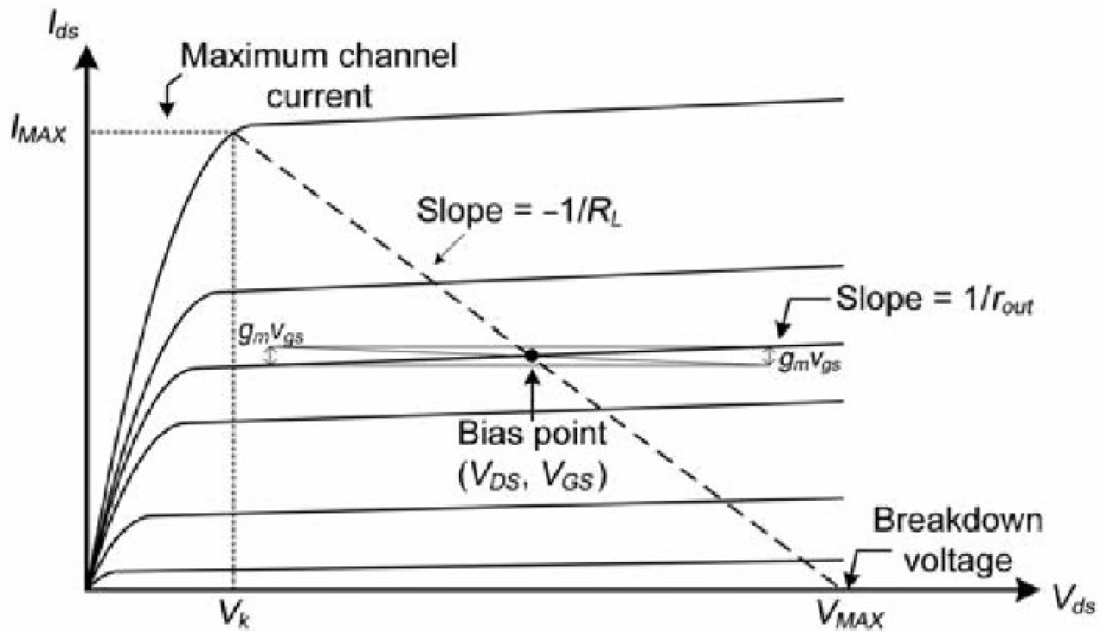

Figure 21. $I_{ds} - V_{ds}$ curve and power generation of output matching. [11]

In this case the correct value for R_L that is commonly referred as a loadline match can be obtained using

$$R_L = \frac{V_{max} - V_k}{I_{max}}, \tag{3.30}$$

where V_{max} is the maximum voltage swing, I_{max} is the maximum channel current, and V_k is the minimum drain-to-source voltage to keep the transistor in the saturation operating mode (knee voltage). The assumption is that $R_{gen} \gg R_L$. When the output impedance R_{gen} of the amplifier/generator is considered, the equation becomes

$$\frac{R_{gen} * R_L}{R_{gen} + R_L} = \frac{V_{max} - V_k}{I_{max}} \tag{3.31}$$

It is common to encounter the two different theories conjugate match and loadline match when working with RF amplifiers. This might appear contradictory, but one must keep in mind that the conjugate match can only be applied in the case of a generator or amplifier that doesn't have any physical limitations in its terminal voltages. This is not a realistic case. For example, with silicon technologies the breakdown electric field sets a limitation for the DC supply voltage, and hence, for the usable voltage swings. Selecting correct value for the load seen by the amplifier plays key role in maximizing the linearity of the amplifier.

Figure 22 is an example of the power characteristics when a class A amplifier has been terminated with conjugate match (S22) or power match (loadline match). The solid line shows a response when the amplifier is terminated to a conjugately matched load which results in the lower maximum power levels. The points A is the is the power level when nonlinear behavior can initially be detected and point B is the 1 dB compression point. It is common in the field to use 1 dB compression point as a measure of power, and linearity capability of an amplifier. An amplifier that is terminated with conjugate match usually results in the compression point that is significantly lower compared to a case when termination is carefully chosen to maximize the power capacity. The dashed line in the Figure 22 is a power matched case and as can be seen both points A and B deliver approximately 2 dB higher power compared to the solid line (conjugate match). The curves in the Figure 20 and Figure 22 relate and they have the same root cause: limited voltage and/or current range in the output terminal. [18]

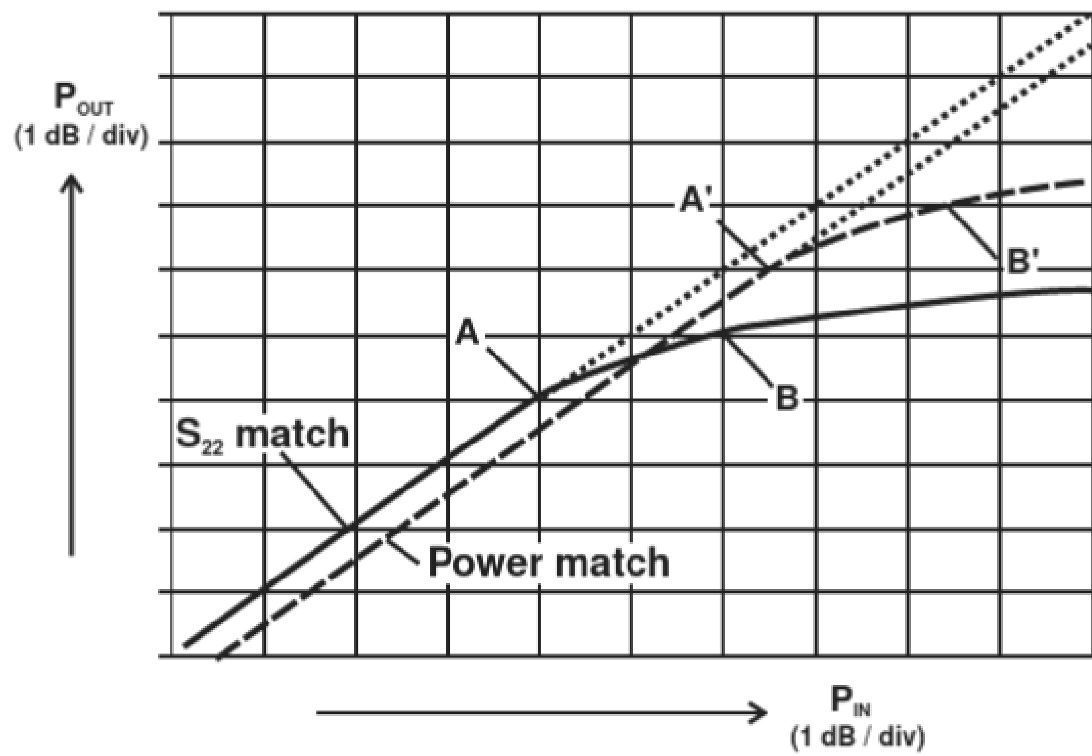

Figure 22. Conjugate match (S22) and power match. [18]

4 DESIGN SPECIFICATION

This wireless PA is integrated on a silicon die using CMOS 22 nm semiconductor technology. The designed PA must be Wi-Fi 6E capable so it must comply with Wi-Fi 6 or IEEE 802.11ax radio standard. Supported radio bands in Wi-Fi 6E are 2.4 GHz, 5 GHZ and 6 GHz. The main challenge of this work is the integration of PA delivering the specified output power/linearity inside Wi-Fi transceiver IC. Typical approach is to use an external front-end moule (FEM) component as a last block before antenna that is responsible for handling and delivering the output power. Such an external front end component usually integrates only PA, LNA and RF switches. It means it can be fabricated using process technology that is optimized for high power operation. However, a radio frequency integrated circuit (RFIC), such as transceiver IC usually integrates analog and digital design so optimizing for PA design only is not possible. Integrating the PA and antenna switch capability inside the chip serves an advantage as additional FEM components results in additional cost and complexity outside the RFIC.

4.1 Overall design

The block diagram of the transmitter path is shown in Figure 23. Only the parts of the transmitter that are directly interacting with PA or effecting the performance are displayed. As indicated by the dashed rectangle all the blocks are built inside the integrated circuit. As the Wi-Fi 6E contains three different frequency bands there is a dedicated transmitter path for each band. This makes it possible to optimize PA, LNA and matching networks for each band separately.

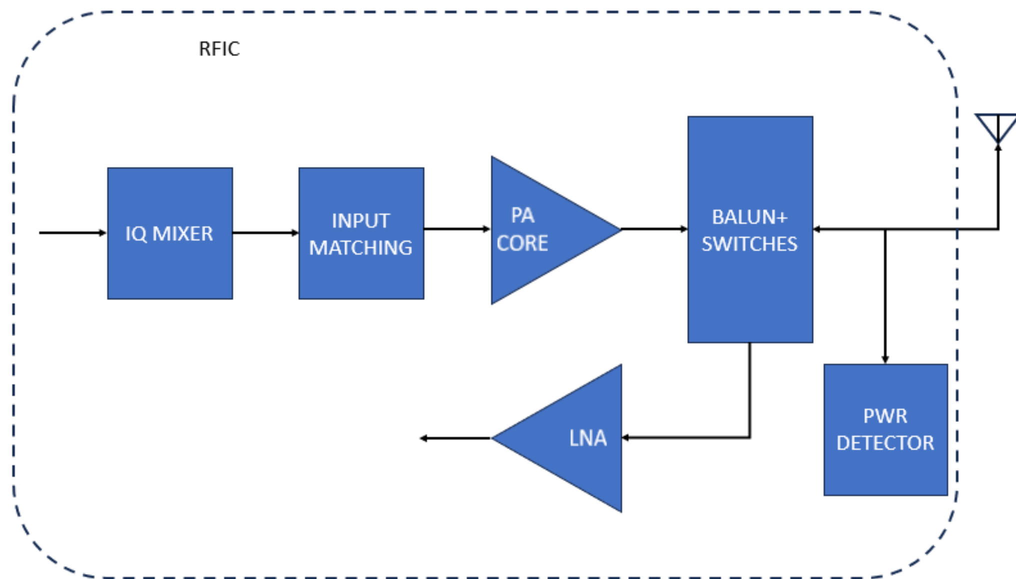

Figure 23. Block diagram of the transmitter. Each of the Wi-Fi 6 frequency band has a dedicated transmitter path.

The additional challenge is to implement the switching capability between the antenna and TX/RX path as the instantaneous maximum voltage spikes at the antenna node are relatively high due to the high PAPR of multicarrier OFDMA Wi-Fi 6E signal.

4.2 Modulation and radio spectrum specifications

The three radio bands of Wi-Fi 6 consist of multiple numbered sub-channels that are 20 MHz in bandwidth. Most sub-channels are spaced 5 MHz from each meaning that the channels are overlapping each other on the spectrum. In Wi-Fi standard, one 20 MHz channel is split into smaller information carrying tones often called as subcarriers. A subcarrier spacing in Wi-Fi 6 is 78.125 kHz which is one quarter compared to that of in previous Wi-Fi 5 standard. Number of subcarriers N_{sc} can be obtained simply:

$$N_{sc} = \frac{BW_{ch}}{f_{sc}} = \frac{20 \cdot 10^6\ Hz}{78.125 \cdot 10^3\ Hz} = 256, \qquad (4.1)$$

where BW_{ch} is the channel bandwidth and f_{sc} is the subcarrier spacing in hertz respectively. Not all 256 subcarriers are used for data transmission. Out of 256 subcarriers within 20 MHz channel, 234 are used for data transmission, seven DC tones in the middle for identification and four pilot tones for channel estimation. In Wi-Fi 6 a subcarrier can be modulated with either BPSK, QPSK, or 16/64/256/1024-QAM. The used modulation and error coding are defined by the modulation and coding scheme (MCs) index in Table 4. [24] [30]

Table 4. Modulation and coding schemes used in Wi-Fi 6.

MCS Index	**Modulation**	**Coding rate**
MCS0	BPSK	1/2
MCS1	QPSK	1/2
MCS2	QPSK	3/4
MCS3	16-QAM	1/2
MCS4	16-QAM	3/4
MCS5	64-QAM	2/3
MCS6	64-QAM	3/4
MCS7	64-QAM	5/6
MCS8	256-QAM	3/4
MCS9	256-QAM	5/6
MCS10	1024-QAM	3/4
MCS11	1024-QAM	5/6

Another definition regarding the OFDMA-based scheduling in Wi-Fi 6 is a resource unit (RU). Seven types of RUs are defined by the standard based on the number of subcarriers assigned. The smallest RU consists of 26 subcarriers. Figure 24 shows an example of RU setup for a 20 MHz Wi-Fi 6 channel. RU allocation for different users is made dynamically by the AP. One 20 MHz channel can serve 9 users when 26-tone RUs are allocated for each. The largest RU that a 20 MHz channel can serve is 242-tone RU when the entire BW is dedicated for a single user. This results in the highest data-rate. [31]

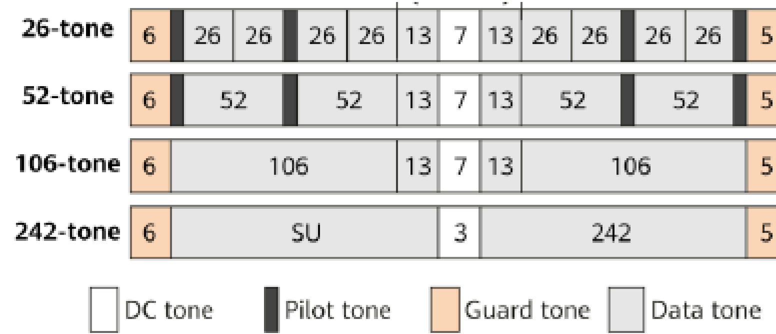

Figure 24. RUs supported by 20 MHz Wi-Fi 6E channel. The $2 \cdot 13$ subcarriers around the midpoint form one RU.

The BW in Wi-Fi 6 is not limited to 20 MHz only. In addition, 40 MHz, 80 MHz, and 160 MHz BWs can be used. The following table lists the complete frequency range that the device is targeted to support. As can be seen in Table 5 the supported frequency range in Wi-Fi 6E, which has an “extended” frequency range compared to Wi-Fi 6, includes 2.4 GHz band and stretches from 5150 MHz to 7125 MHz resulting in the overall bandwidth of approximately 2 GHz. This sets a major challenge for the design.

Table 5. The complete frequency range of different Wi-Fi 6E bands.

2.4G	**5G**	**6G**
2412…2484 MHz	5150…5895 GHz	5925…7125 MHz

4.3 Output power and linearity

The maximum output power for wireless applications has been defined by ETSI (European Telecommunications Standards Institute). The maximum mean EIRP (Equivalent Isotropic Radiated Power) is 23 dBm for the frequency range (5150…5350) MHz and as high as 30 dBm for the frequency range between (5470…5725) MHz [32]. The maximum mean output power target for the wireless PA in this work is 16 dBm. With PAPR of approximately 10 dB the instantaneous voltage swings at the output reach ~26 dBm power. This must be

considered in the safe operating area (SOA) design of the transistor stages. The most challenging linearity requirement in Wi-Fi 6 is in MCS11 when 1024-QAM is being used. In this modulation scheme the transmitter must comply with the EVM requirement of less than −35 dB. As discussed in 3.2.2 the gain compression/distortion is trade-off between power. It is therefore possible to use higher output power with lower order modulations if the transistor voltages stay within SOA limits. [33]

4.4 Transmit spectral mask

The transmit spectrum mask defines the power contained in a specific frequency bandwidth at a certain offset from the carrier frequency. The purpose of the test is to ensure that multiple Wi-Fi devices operating within a range do not interfere each other. This measurement is typically expressed as a ratio of power spectral densities (PSD) between the carrier and a specified offset frequency. The standard is using dBr for expressing dB relative to the maximum spectral density of the signal. There are four different spectrum masks defined depending on the bandwidth allocation for the device under test. Figure 25 shows an example of 20 MHz spectral mask specification. The spectral mask example defines 0 dBr for the bandwidth of 19.5 MHz (± 9.75 MHz) centered at the carrier frequency, −20 dBr at 10.5 MHz offset, −28 dBr at 20 MHz offset and −40 dBr at 30 MHz offset. The rest of the frequency points falling between the specified values are obtained by interpolating in dB domain. Measurements are performed using 100 kHz resolution bandwidth. The rest of the spectrum masks for different bandwidths are defined in a similar manner. [34]

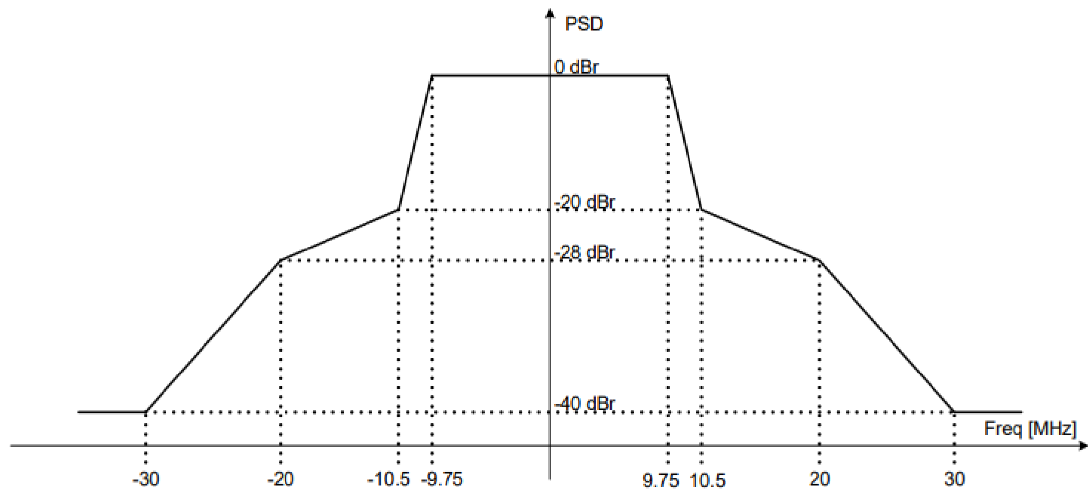

Figure 25. Example of transmit spectral mask for 20 MHz bandwidth.

4.5 Stability criteria

As discussed in 3.7 it is important to analyze the stability of an amplifier for possible oscillations. This work uses μ_L, μ_S stability factors for the stability analysis and the analyzed frequency range is 1 MHz…100 GHz. The specification for stability is $\mu_L, \mu_S > 1$ indicating unconditionally stable operation.

4.6 Summary of specifications for the PA

The total performance of the transmitter is affected by multiple individual building blocks such as Digital to analog converter (DAC), mixer and pre-power-amplifier (PPA) stages. Each cascaded nonlinear block contributes to the overall linearity. However, the contribution of PA is dominant for certain parameters such as current consumption. The PPA and prior blocks are specified to have roughly 10 dB better IM3 linearity than PA to relax the linearity specification for the PA as much as possible. Table 6 summarizes the target specifications for the PA block of this thesis. The ultimate EVM target for the entire transmitter system is the MCS11 requirement ≤ –35 dB [34].

Table 6. Target specification for the PA.

Specification	Limit
Voltage Gain	24 dB
Power Gain	30 dB
Current consumption	< 600 mA
Stability factors (μ_L, μ_S) (1)	≥ 1
Linearity (IM3) @ 16 dBm PWR	≤ −44 dBc
OIP3	38 dBm
EVM (2)	≤ –35 dB

(1) Analysis range is 1 MHz…100 GHz

(2) Target for the entire transmitter

5 POWER AMPLIFIER DESIGN

The design and selection for the topology started with analyzing the voltage ratings of the available devices in the used process technology. Three device types were available in the used 22 nm CMOS process as shown in Table 7.

Table 7. Specifications of available devices in the used process technology.

Device type	Oxide type	V_{ds}	V_{gs}	V_{dg}	L_{min} (nm)
Core	thin	1.0	1.0	1.0	22
Thick oxide	thick	2.80	2.80	2.80	270
HMOS	thick	5.55	2.55	5.55	550

Values in Table 7 are not the absolute breakdown voltage limits but they indicate the maximum operating values that results in some specific lifetime assured by the technology vendor. More specifically, the variation of electrical characteristics of a device within certain time period is less than defined percentage value. Values are often referred as the safe operating area limits. In terms of voltage ratings, the core devices have the weakest robustness. Thick oxide devices have higher voltage ratings, but they also have higher minimum channel length value. Higher oxide thickness means lower oxide capacitance c_{ox} and that together with higher channel length results in the lower value for transconductance g_m according to Equation 2.11. It becomes obvious that the core devices provide the highest transconductance per width unit.

5.1 Topology selection

The main specifications for the PA are the saturated output power P_{sat} of 28 dBm and the linearity requirement is defined as an $IM_3 \leq -44$ dBc at 16 dBm output power. The specification for the voltage gain is $G_v \geq 24$ dB. Traditional class-A/AB differential common source cascode circuit was selected for the design topology. The advantage of Class-A is its high linearity when

operating in the linear region of the i_d, v_{gs} plot, and the minimized common-mode signal components. On the other hand, class-AB provides higher efficiency depending on the bias point. Common-source amplifier topology was selected because it provides the highest power gain. Class-A amplifier results in the lowest efficiency that impose a challenge especially combined with the high linearity requirement and high PAPR of the modulating signal. It is therefore important to utilize as high voltage swing as possible without exceeding the safe operating area of the transistor as defined in Table 7. With cascode structure it is possible to have high-voltage HMOS as an output device that can sustain higher output voltage swing while thin oxide core transistor can serve as an input stage providing higher transconductance. The use of cascode transistor also improves the isolation between input and output that improves the stability of the circuit and reduces the miller effect. [35] [21]

Figure 26 shows an example of common-source cascode amplifier. The voltage swing V_{d1} at the drain of the input transistor M1 can be mitigated by a careful design of the cascode transistor M2. The impedance seen by the input transistor M1 is the inverse of the transconductance, g_{m2} of the cascode transistor. The idea is to have $1/g_{m2}$ significantly lower than the load seen by the cascode transistor. This results in much lower voltage swing, V_{d1}, compared to V_{d2} that is the voltage swing at the drain of the output transistor M2.

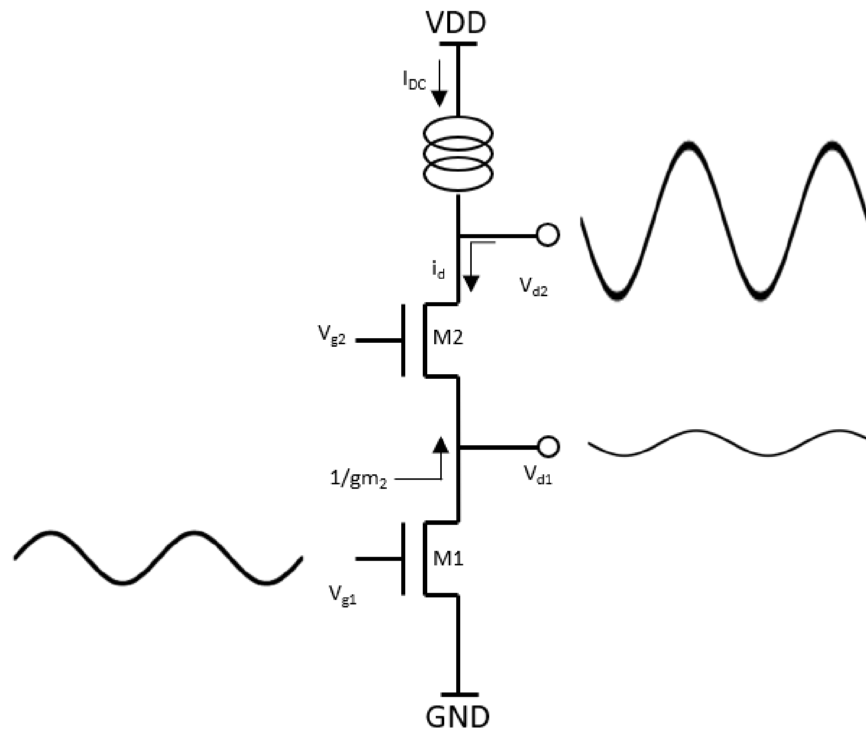

Figure 26. Single-end cascode structure.

5.2 Differential common-source

Many studies on power amplifiers utilize a differential circuit structure. However, there are few trade-offs when comparing using a single-ended circuit or a differential one. The most obvious advantage of a differential PA is a larger voltage swing with a given DC supply voltage. If the knee voltage is assumed zero, a single-end structure can provide a voltage amplitude swing of $V_{dd},$ when the DC feed is provided with an inductor connected to the supply voltage V_{dd}. With differential structure both drains can swing the same amount resulting in the differential amplitude of $2V_{dd}$. This results in the 3 dB increase in power for a given signal current. Additionally, differential structure when biased in A-class does not have any signal currents in the supply or ground network that mitigates the possible supply modulation effects. Therefore, the significance of any parasitics in the supply networks is reduced and a degree of modeling for the supply network is relaxed. Single-end structure naturally pulls all the signal

current from supply networks. Single-end and differential configurations are shown in Figure 27.

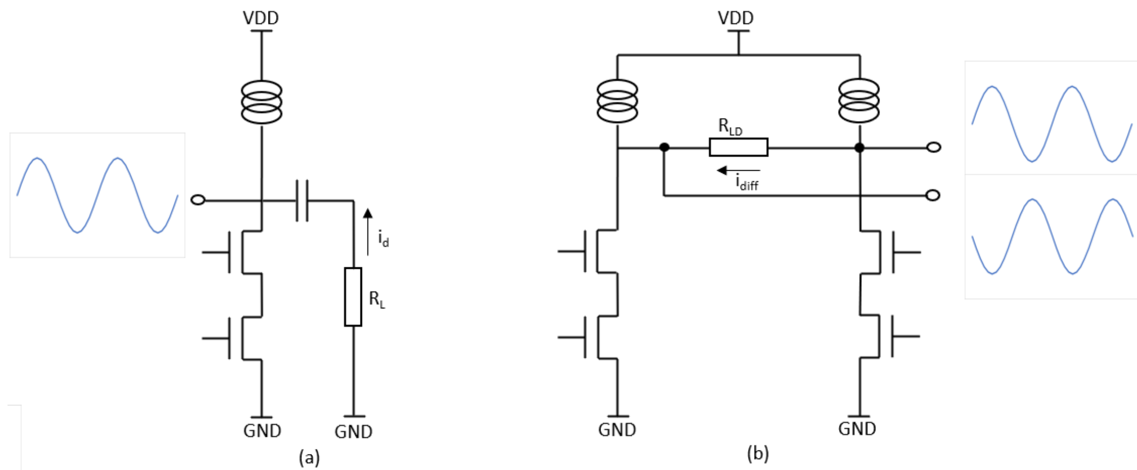

Figure 27. Single-end (a) and differential (b) configuration.

To achieve the differential amplitude of $2V_{dd}$ the differential load value R_{LD} must be twice the single-end load value R_L so that $R_{LD} = 2R_L$. A larger output resistance makes it easier to implement the output matching network by enhancing bandwidth, reducing the number of lumped elements, or lowering power losses. Additionally, a larger output voltage reduces the output current needed to deliver a given power to the load, which has numerous benefits, such as minimizing electromigration problems, reducing self-heating, decreasing the size of power transistors, and lowering losses in the passive elements connected at the PA output. Differential structure requires a transformer to convert the differential signal to a single-end load that is typically a 50 Ω antenna. Integrated transformer is often occupying large chip area and introduces additional losses that can be counted as a drawback of differential structure. However, the use of transformer can provide harmonic filtering for the common-mode voltage components that further increase the voltage swing at the drain resulting in the clipping of the transistor output voltage. The common-mode impedance seen by the amplifier is determined by the common mode properties of the transformer and supply line impedance that connects to the

transformer center tap. It is possible, within certain limits, to tweak the common-mode impedance without effecting the differential impedance. [18] [21]

If the layout design is done carefully the differential current can be enclosed in the circuit core area only, and no significant signal currents exist in the supply or ground routing. This reduces parasitic and magnetic coupling effects that requires complex and extensive modelling/analysis to debug.

5.3 Amplifier core design

This section describes the design of amplifier parameters such as transistor dimensions, DC operating point and load impedance.

The supply voltage for the amplifier core is 3.3 V. First step was to calculate the current needed to generate 28 dBm output power. A HMOS was used as a cascode transistor due to its capability to handle as high as 5.55 V drain to source voltage. A voltage of 5.5 V is used in calculations. With these parameters given it is possible to define the signal current needed to generate the 28 dBm antenna power. If matching network is assumed to have 1 dB loss the power needed from the core is simply:

$$P_{CORE,MAX} = P_{ANT,MAX} + IL = 28\ \mathrm{dBm} + 1\ \mathrm{dB} = 29\ \mathrm{dBm}, \qquad (5.1)$$

where $P_{ANT,MAX}$ is the absolute maximum antenna power in dBm, $P_{CORE,MAX}$ is the absolute maximum power generated by the amplifier core in dBm, and IL is the loss by the matching network in dB. The maximum differential signal amplitude $v_{diff,max}$ can be obtained by:

$$v_{diff,max} = 2 \cdot (5.5\ \mathrm{V} - 3.3\ \mathrm{V}) = 4.4\ \mathrm{V} \qquad (5.2)$$

This means that the lowest voltage $V_{D,LOW}$ at the drain of the cascode is

$$V_{D,LOW} = 3.3\ \mathrm{V} - \frac{v_{diff,max}}{2} = 1.1\ \mathrm{V}. \qquad (5.3)$$

To keep the transistor in the saturation mode, the drain voltage must follow $V_{D,LOW} \geq V_g - V_{th}$. Assuming that the cascode transistor gate is biased at 1.6 V and the threshold voltage for HVMOS is 0.6 V this should be enough to keep the transistor in saturation. The power generated by the core is

$$P_{CORE,MAX} = \frac{i_{diff,max} v_{diff,max}}{2} = 29 \text{ dBm}. \quad (5.4)$$

Therefore, the current i_{diff}, and the load value R_{LD} is obtained by:

$$i_{diff} = \frac{2P_{CORE,MAX}}{v_{diff,max}} = \frac{2 \cdot 1 \cdot 10^{-3} \cdot 10^{\left(\frac{29}{10}\right)}}{4.4\ V} \approx 360 \text{ mA}, \quad (5.5)$$

$$R_{LD} = \frac{v_{diff}}{i_{diff}} = \frac{4.4 \text{ V}}{360 \cdot 10^{-3}} \approx 12\ \Omega. \quad (5.6)$$

Next step is to define the bias current and the size of the transistors. Analysis is started with a unit transistor with the following parameters:

$$W_f = 2\ \mu m, \qquad L = 30\ nm, \qquad N_f = 32,$$

where W_f is the finger width, L is the channel length, and N_f the number of parallel fingers. The effective total size of the transistor is simply:

$$\left(\frac{W}{L}\right)_{TOT} = \frac{N_f W_f}{L}. \quad (5.7)$$

Common-source circuit in Figure 27 (a) is used to get a preliminary, and coarse estimation of the output current linearity. Load value $R_L = 1$ is selected to minimize the voltage swing at the drain for the preliminary analysis. Figure 28 shows DC current and two-tone linearity as a function of gate DC bias voltage V_G. It can be observed that $V_G = 0.45$ V and $I_{DC} \approx 5$ mA gives an IM3 linearity of about -46 dBc that seems to be a local optimum and gives a good balance between the IM3 performance and DC current consumption.

The specification for IM3 performance is ≤ -44 dBc at 16 dBm power, but some margin is needed as the used simulation setup can only be used to give a coarse estimation of performance. For example, there will be voltage swing at

the drain of the input transistor when the cascode transistor with finite transconductance is connected at the drain. Another important point to note is that lower bias point shows larger relative even-mode harmonic content as shown in Figure 29, which in turn results in the additional common-mode voltage swing that is strongly dependent on the common-mode impedance of the output resonator. It is therefore accepted that the most optimum bias point cannot be determined using the circuit in Figure 27 (a) and final adjustments are needed later with more complete design. The used root-mean-square (RMS) input voltage in the analysis is $v_{in,rms} \approx 70\ \mathrm{mV}$.

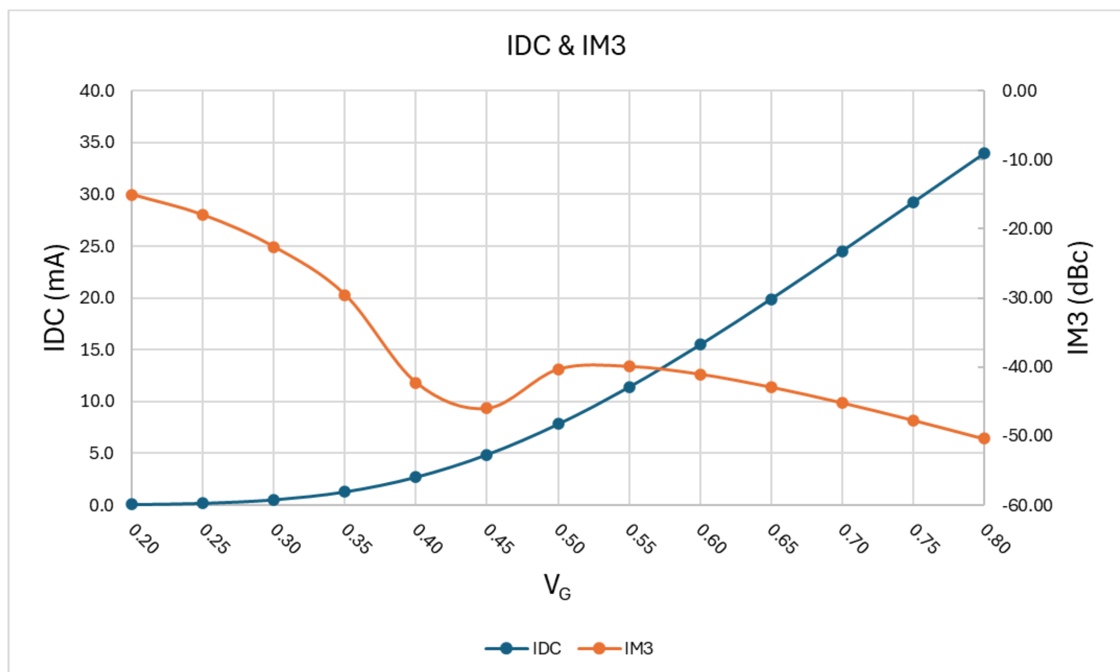

Figure 28. The DC bias current and IM3 linearity of the output current as a function of gate bias voltage V_G.

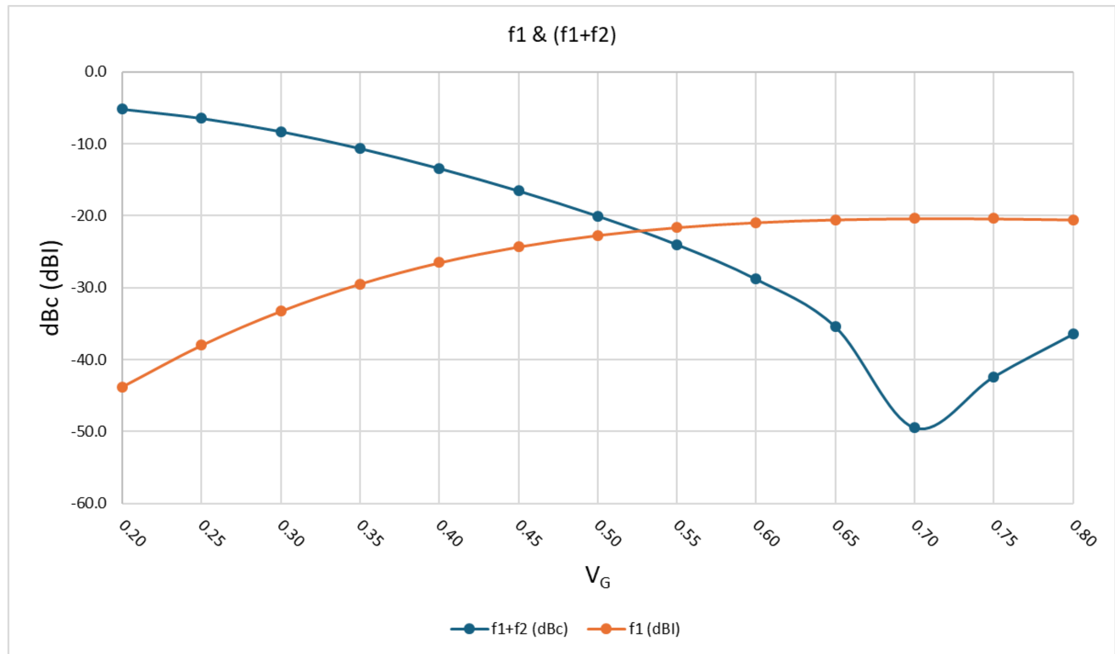

Figure 29. The magnitude of the current at the first wanted frequency tone *f1*, and the relative magnitude of the current at 2nd order intermodulation tone *f1+f2*.

Figure 30 shows the drain current waveform zoom at the peak of the envelope with $V_{in,rms} \approx 70\ \mathrm{mV}$.

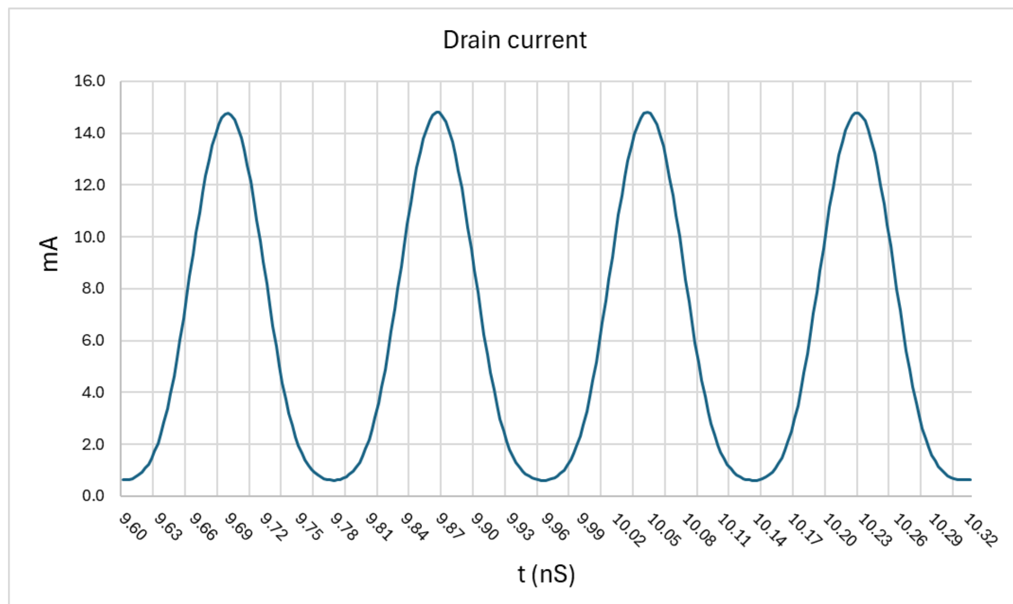

Figure 30. Drain current waveform zoom at max envelope point.

When the bias point and the input signal level resulting in $IM3 \leq -44\ \mathrm{dBc}$ have been defined, the next step is to scale the design to achieve the 16 dBm antenna power. Again, assuming 1 dB loss in the matching network the power generated by the core is 17 dBm. Therefore,

$$I_{rms_17dBm} = \sqrt{\frac{P_{CORE_17dBm}}{R_{LD}}} = \sqrt{\frac{1 \cdot 10^{-3} \cdot 10^{\left(\frac{17}{10}\right)}\ \mathrm{W}}{12\ \Omega}} \approx 65\ \mathrm{mA}. \qquad (5.8)$$

The RMS signal current, I_{rms} in the previous analysis is 3.6 mA meaning that the multiplier, m the number of parallel unit transistors must be

$$m = \frac{I_{rms_{17dBm}}}{I_{rms}} = \frac{65\ \mathrm{mA}}{3.6\ \mathrm{mA}} \approx 18. \qquad (5.9)$$

DC current consumption is scaled accordingly. DC current consumption, I_{DC_unit} of a unit transistor at the linearity optimum according to Figure 30 is 5 mA so total current consumption per “leg” is

$$I_{DC} = mI_{DC_unit} = 18 \cdot 5\ \mathrm{mA} = 90\ \mathrm{mA}.$$

At this stage, it is evident that the RMS signal amplitude is less than the bias current. However, the peak amplitude of a two-tone signal, twice the RMS value, exceeds the bias current value. Additionally, the gate bias voltage only marginally exceeds the threshold voltage. Consequently, the operational condition with this DC bias current is categorized as class-AB. The size of the cascode transistor was defined so that the source impedance $1/g_{m2}$ seen by the input stage is roughly $R_{LD}/10$.

5.4 Bias circuit

The oscillatory tendencies of RF power transistors have long been recognized as an unfortunate aspect of RF power amplifier design. In the low frequency area ranging from MHz to several hundreds of MHz, the common-mode terminating impedances of RF PA is often defined by the bias networks. An RF PA that operates in Class AB will exhibit AC currents through its biasing circuits

with the variations in signal envelope and the impedance that is connected in the bias supply path will generate voltage swing accordingly that will affect the gain and phase of the PA. Therefore, the design of bias networks, and considering the common-mode waveform generation, plays a significant role when ensuring a linear and stable operation. [18]

Bias voltage is generated by supplying tunable current into a diode-connected MOSFET that is physically embedded in the transistor core. This approach was selected to achieve good layout matching and to ensure that the diode and the actual PA common-source transistors experience the same junction temperature to avoid possible thermal runaway. According to simulations the temperature difference between the core and perimeter area can be tens of Celsius degrees.

Bias voltage is finally fed to the amplifier core through a differential inductor as shown in Figure 31. Diode-connected M1 is physically located inside transistor core.

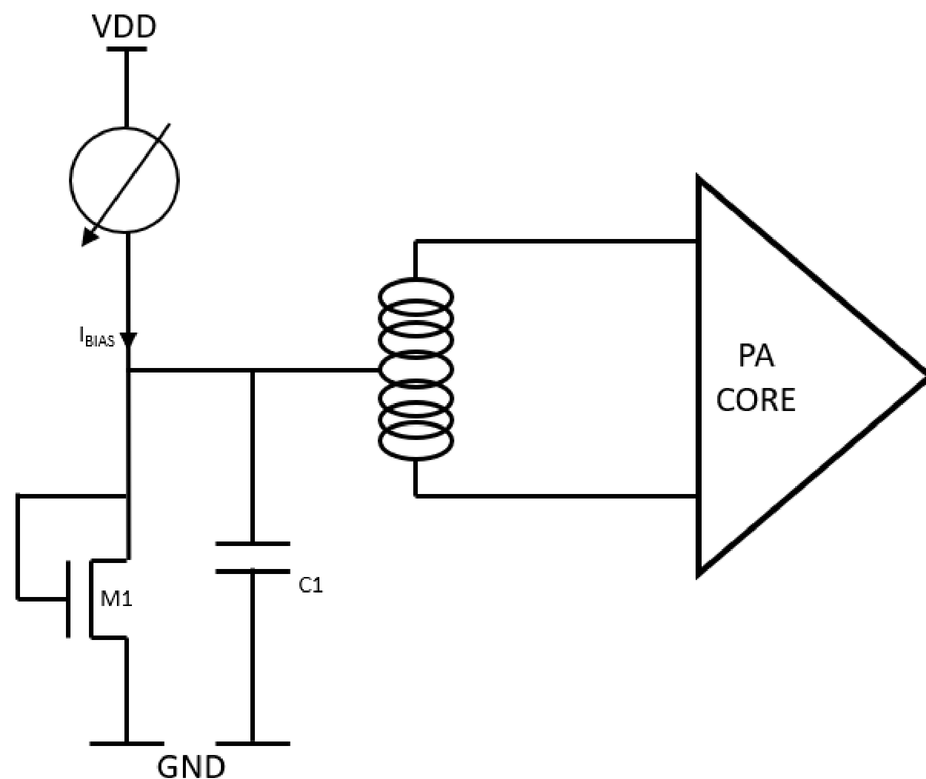

Figure 31. Bias voltage feed for the amplifier core.

The advantage of using an inductor is that the common-mode impedance seen by the common-source gates is defined by the parallel connection of M1 and C1. Non-linear input impedance of an amplifier introduces distortion that can be seen as a variating bias level as the gate-to-source capacitance changes with the gate voltage. This low-frequency variation occurs at $2f_{bb}$ frequency and depends on bias circuit impedance. The bias circuit must be able to sink/supply gate charge to maintain DC point as the input capacitance of the core varies with envelope amplitude. [9]

When analyzing the low frequency common-mode operation, the core input transistors are effectively parallel to the bias circuit. Therefore, it becomes intuitive that the lower the bias circuit impedance, the less significant is the core input impedance variation.

5.5 Output matching network

As discussed in 5.2 a balanced-to-unbalanced (balun) transformer is needed with a differential circuit structure when connecting to a single-end antenna. This section briefly discusses the key parameters and considerations with the transformer and output matching network used in this work.

There are two main challenges for the output circuit. The PA output and LNA input shares the same connection point at the antenna node meaning that the matching network must be compatible for both PA and LNA operation mode. Another challenge is the high voltage swing at the antenna node that restricts the use of MOSFET switches. Another and simpler solution would be to use an external RF switch that offers 30…40 dB isolation between the PA/LNA [36]. This approach would give a great degree of freedom to optimize the PA matching and the LNA matching separately. However, this would result in additional cost and an additional output pin would be needed to route PA and LNA separately to the RF switch on the printed circuit board (PCB).

Figure 32 shows the circuit that is used to share the antenna node between the PA and LNA. A transformer with an impedance transformation ratio of

$$N^2 = \left(\frac{N_p}{N_s}\right)^2 = \left(\frac{1}{2}\right)^2 = 0.25 \qquad (5.10)$$

is selected as the antenna impedance is 50 Ω and the target load value R_{LD} according to Equation 5.6 for the PA core is 12 Ω. N_p is the number of primary windings and N_s is the number of secondary windings. Capacitors C1 and C2 are used for tuning and the MOSFETs M1 and M2 operate as switches. Tuning is needed to minimize the insertion loss (IL) [37]. In PA operating mode the M1 connects the secondary to ground and M2 enables the transformer secondary tuning capacitance C2.

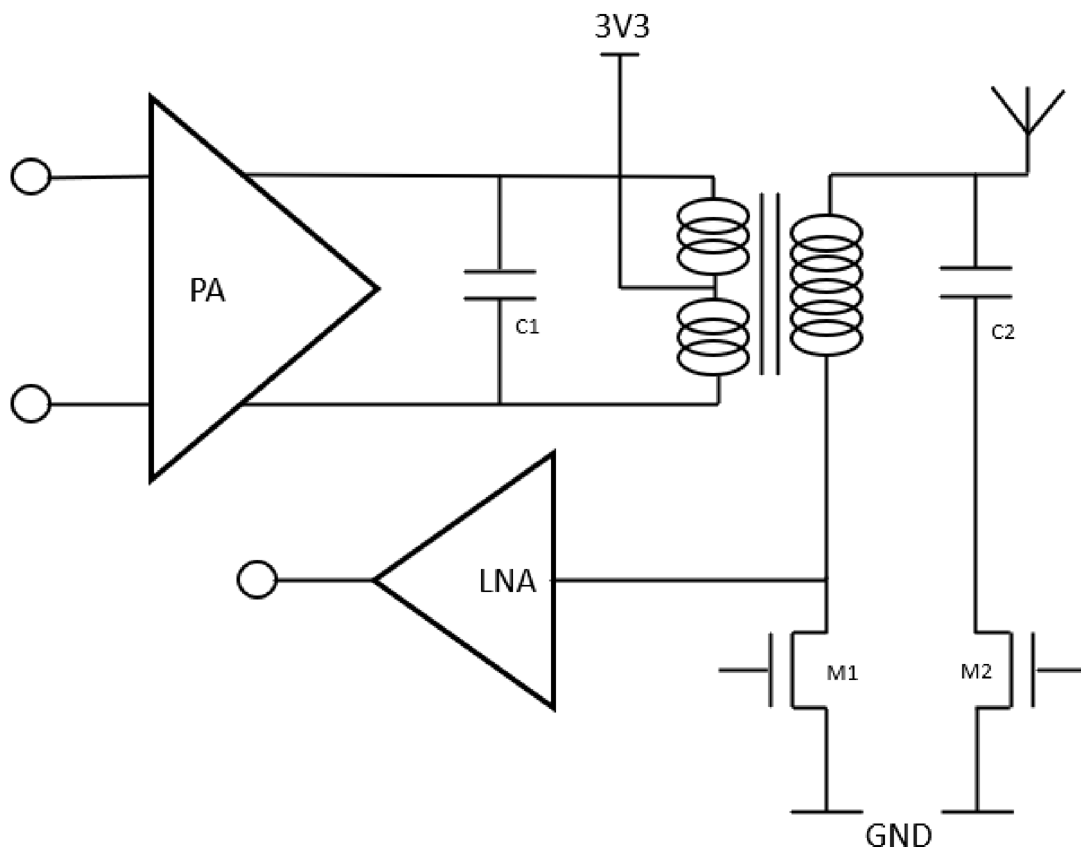

Figure 32. Output matching network with PA and LNA sharing the antenna node.

The design of output transformer is beyond the scope of this thesis. However, few key points effecting the design are briefly explained. The main issue is that

as the output node is shared between the PA and LNA the transformer cannot be fully optimized for the PA operation. As can be seen, in the LNA operating mode, the secondary winding forms a series inductance between the antenna and LNA. The capacitance C1 is connecting parallel to this inductance considering the magnetic coupling of the transformer, and further decreases the self-resonance frequency of the series inductor. The reactance X_{C1} of capacitance C1 seen by the secondary circuit is scaled with the transformation ratio N and becomes X_{C1}/N^2. When the secondary inductor forms a parallel LC circuit with the tuning capacitors, no power can flow into the LNA at resonance, and this can be seen as a steep increase in the LNA noise figure AC response when the secondary is approaching the resonance. Secondary tuning capacitor C2 has been designed switchable and can be disabled to push the resonance frequency further away from the frequency of operation in LNA mode. The shape and size of the transformer has been designed to achieve sufficiently high self-resonance for the secondary inductor to mitigate the above-mentioned problem. [37]

5.6 Parasitic and electromagnetic modeling

At gigahertz frequencies the interconnection routings and parasitics play significant role in the performance. Therefore, proper modeling of the parasitics as well as magnetic behavior of the layout structures must be considered. Figure 33 shows the layout area that has been modeled using electromagnetic (EM) simulation tool. The modeled area contains all the transformers and inductors to include the magnetic coupling between structures in the simulated model. All the active device block layouts such as PA core and LNA core have been modeled using parasitic RC extractor tool that accounts for parasitic resistive losses and parasitic capacitive couplings only.

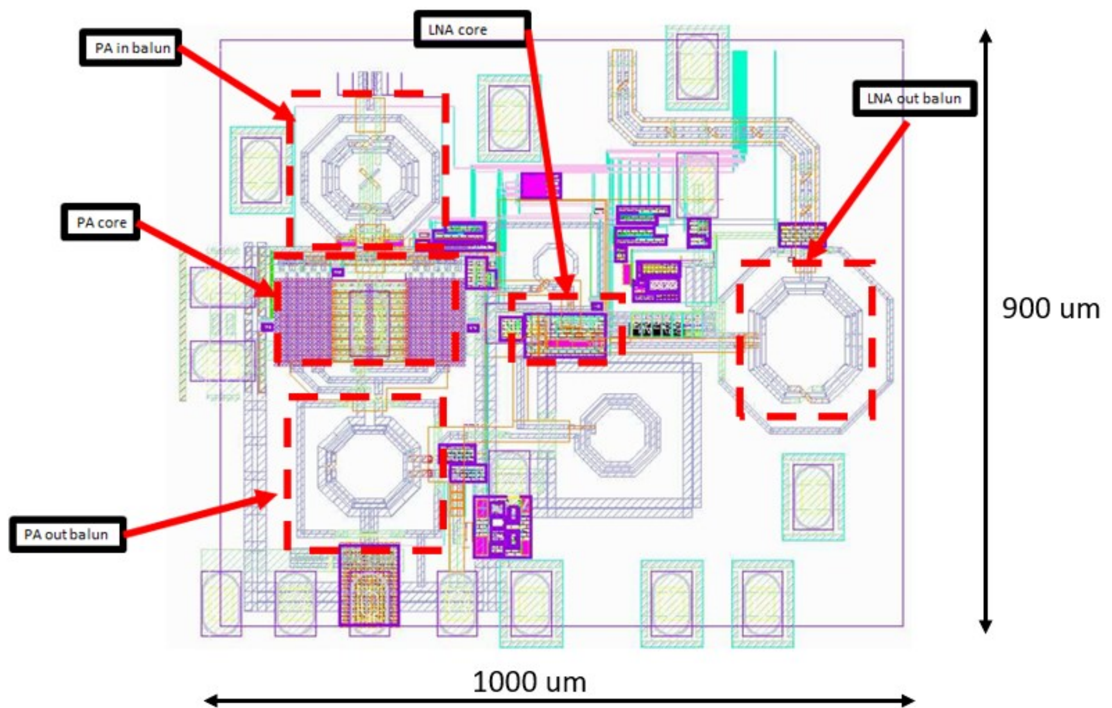

Figure 33. EM modeled layout area.

5.7 Thermal considerations

Due to the high-power consumption a thermal analysis was performed for the device area. Figure 34 shows a temperature analysis for the PA core area. Layout design efforts has been made by maximizing the substrate conductors near the PA core to conduct the heat away from the PA core area to the ground pads of the die. The cascode transistor M2 dissipated the highest power as it has the highest DC voltage drop. However, the input transistor M1 size is smaller that results in higher power density compared to that of the cascode transistor. According to the analysis the junction temperatures can reach as high as 90…95 °C temperatures for the input transistor M1 and slightly less for the cascode M2. The PA core has three on-chip temperature sensor circuits that are indicated by the orange triangles.

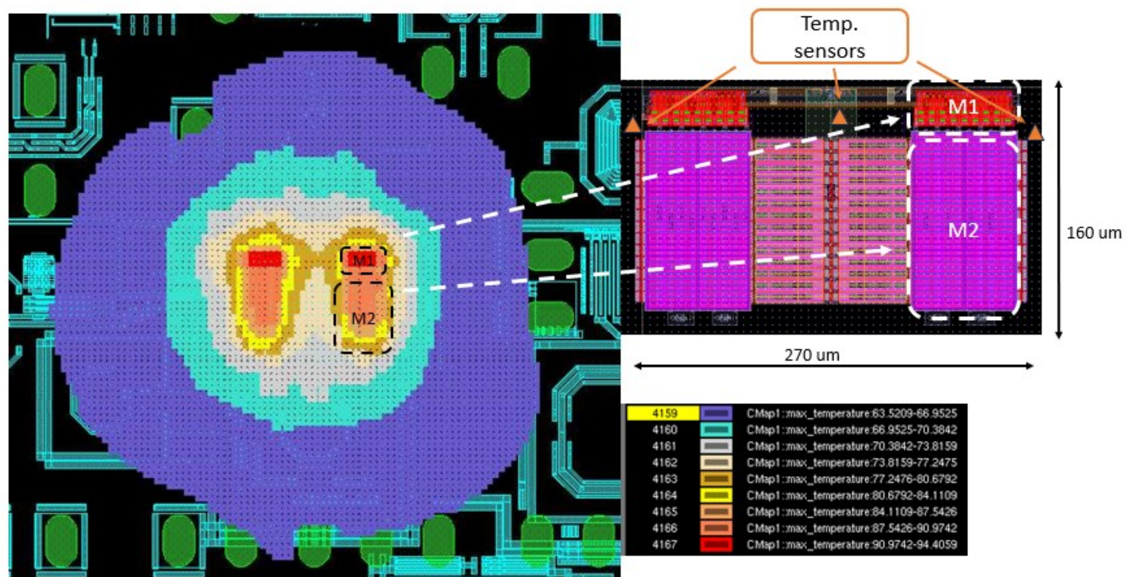

Figure 34. Simulated temperature analysis for the PA core.

6 RESULTS

This section presents the simulation results for the designed PA. Simulation cases have been selected to evaluate how the design compares with the target specifications in Table 6. Simulation results are presented for the 5G band PA. The simulation frequency is mid-band frequency 5.5 GHz and temperature is 80 °C unless otherwise mentioned. The core structure and the topologies of the input/output resonator networks are identical between the bands.

6.1 Voltage gain

The gain partition of the system has been defined as voltage gains for different blocks instead of power gains. Figure 35 shows the PA power gain for different process corners. The FF process corner refers to fast corner sample (i.e. low RC, low V_t, and high g_m) and SS refers to “slow” corner sample (i.e. high RC, high V_t, and low g_m). The process corner TT refers to typical sample. As assumed the FF corner results in the highest gain while the SS corner provides the lowest gain. Also, the gain compression is clearly visible in the SS corner. According to the voltage gain specifications of 24 dB, it seems that the gain falls 2…3 dB short of the target in SS and TT corner. However, it is possible to compensate this shortcoming in the gain with PPA output resonator design. The voltage gain specification for the PPA is 9 dB and the gain of the PPA has been adjusted using a resistor at the output. Few additional decibels can be achieved by readjusting the PPA gain.

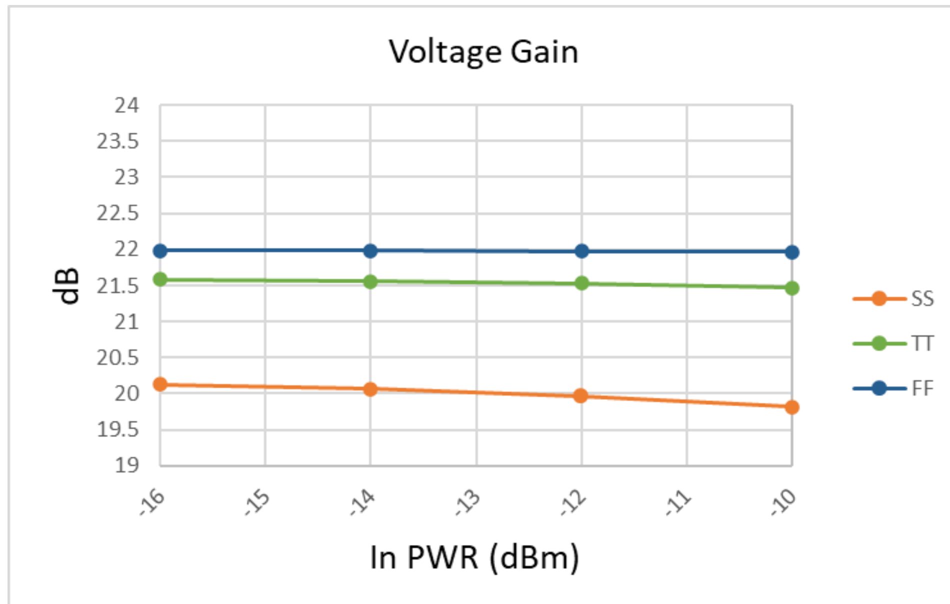

Figure 35. Voltage gain of the PA for different process corners.

Figure 36 shows the output power sweep when only carrier wave (CW) tone is fed to the PA. As can be seen the saturating is clearly visible and rough extrapolation for the saturated power P_{sat} is between 27…28 dBm. Dashed markers indicate the 1 dB compression point that occurs at 24.1 dBm.

Figure 37 shows the PAE as a function of signal power. As can be seen, the efficiency is only few percents at the 16 dBm output power.

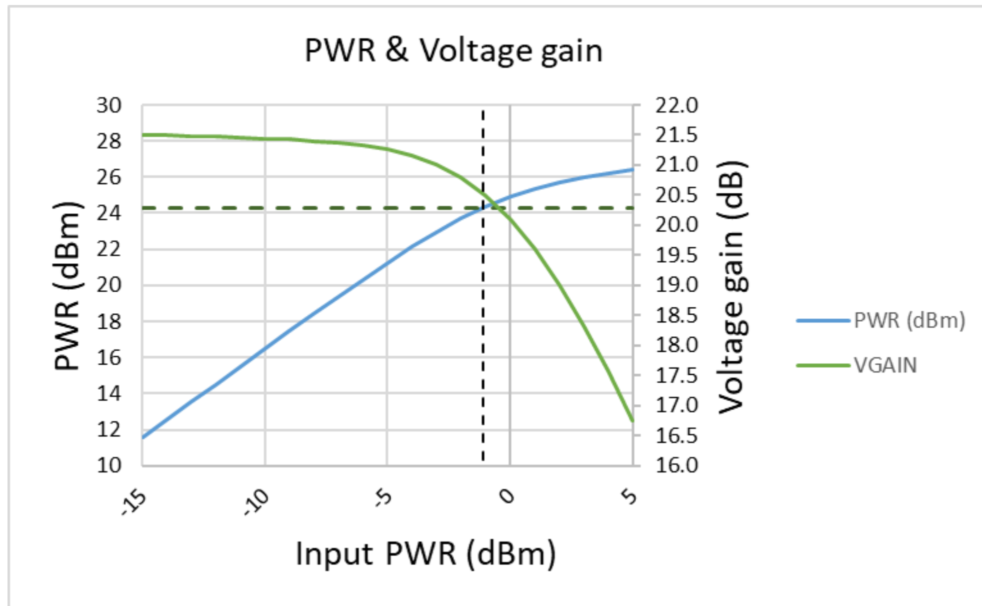

Figure 36. CW power and voltage gain

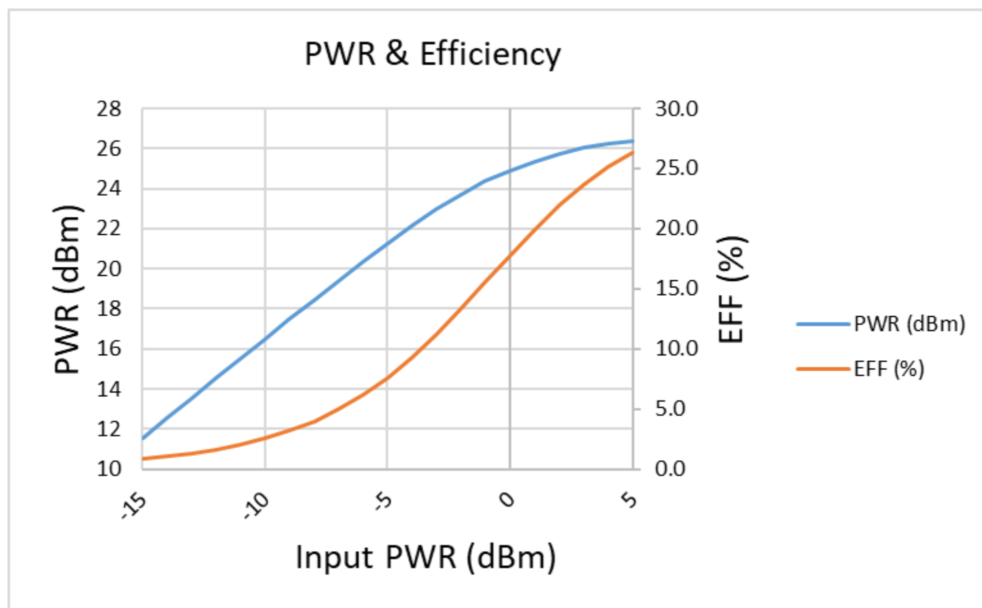

Figure 37. Power added efficiency as a function of signal power.

Power gain

Figure 38 shows the power gain for different process corners. As discussed in 6.1 the system gain partition has been defined as voltage gain. The power gain specification of 30 dB uses the assumption that the input impedance is four

times greater than the 50 Ω antenna impedance resulting in the additional 6 dB compared to the voltage gain. The difference of simulated voltage gain and power gain is between approximately 4.5…5.5 dB depending on the process corner. This can be explained with the input impedance that according to simulation in 6.3 is slightly less than the original estimation.

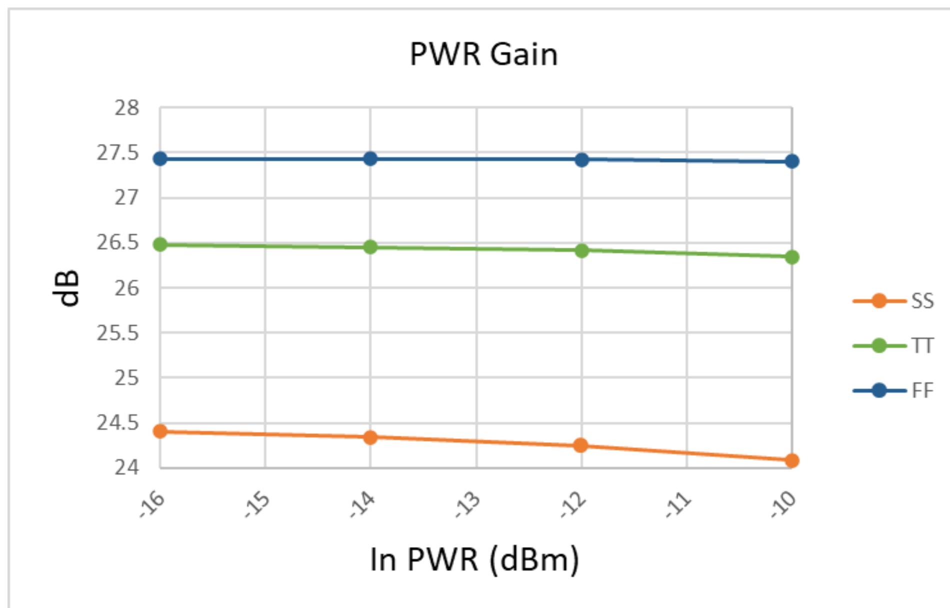

Figure 38. Power gain of the PA for different process corners.

6.2 Linearity and current consumption

Figure 39 shows the linearity and the current consumption as a function of gate bias voltage. It can be observed that the highest OIP3 with gate voltage of 600 mV that results in the current consumption of approximately 520 mA. It is worth noting that the variation of OIP3 is only 1 dB within the current consumption variation range of 350 mA…600 mA. Therefore, it is possible to scale down the current consumption when the absolute maximum linearity performance is not needed with lower MCS indexes. This also mitigates the possible heating issues as the power consumption is directly proportional to the current consumption.

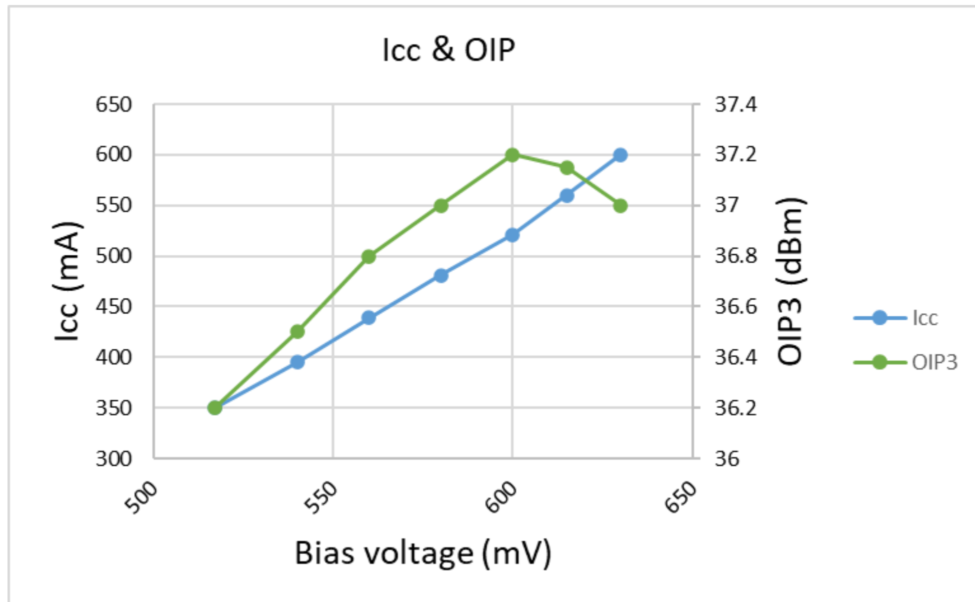

Figure 39. Linearity and Current consumption as a function of bias point

Figure 40 Shows IM3 linearity as a function of power. Dashed markers indicate the 16 dBm point at which the IM3 linearity was specified to be ≤ –44 dBc. As can be seen the linearity performance is approximately 2 dB less than the target specification.

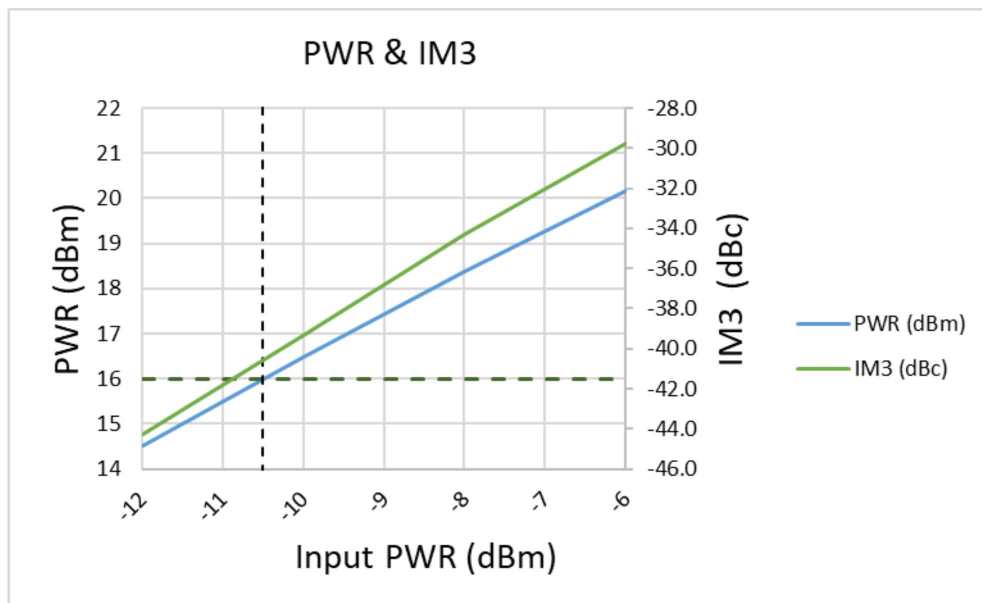

Figure 40. Output power and IM3 as a function of input power.

Figure 41 shows the variation of linearity and output power as a function of junction temperature between 80…120 °C. The power drops less than 1 dB and OIP drops less than 0.5 dB.

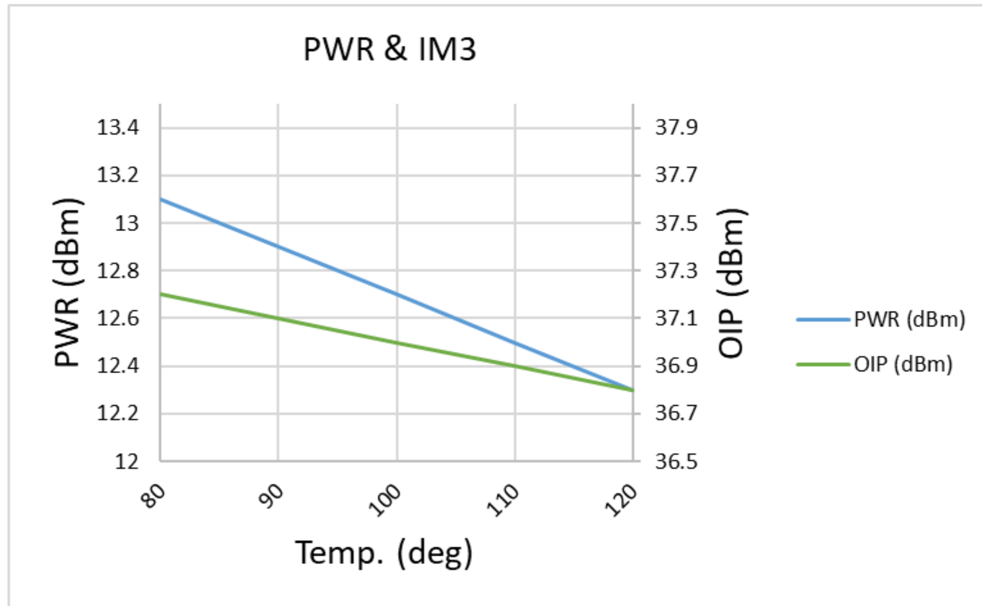

Figure 41. Power and Linearity variation as a function of temperature.

Figure 42 shows an EVM analysis using Matlab. The simulation uses a memoryless baseband equivalent model for the PA that has been created by extracting the AM-AM and AM-PM characteristics using CW tone at 5.5 GHz.

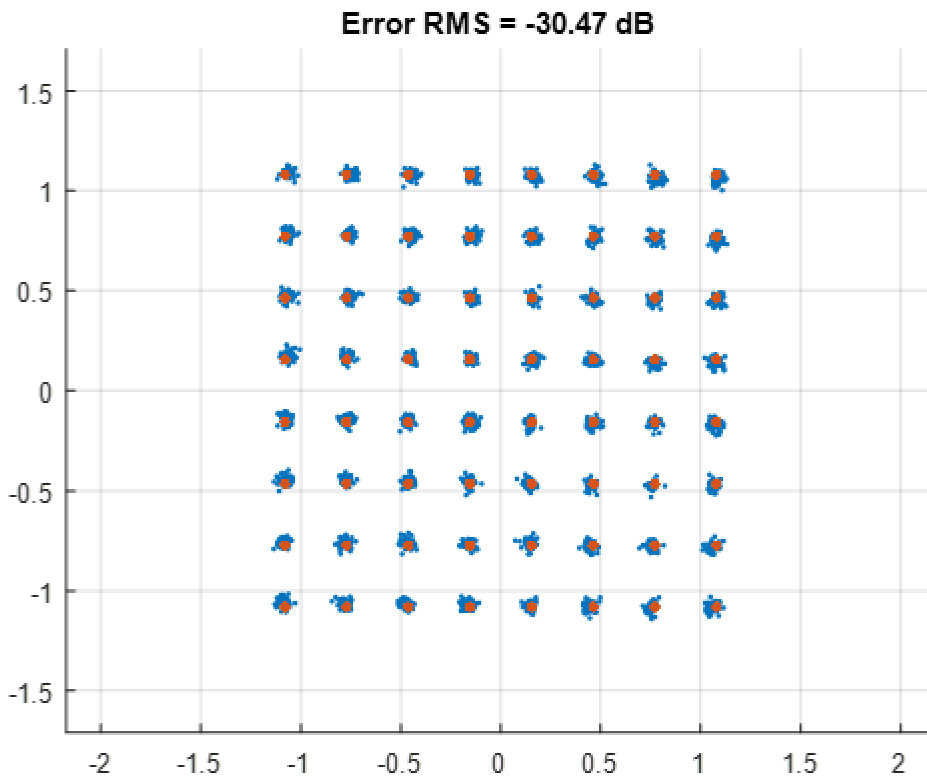

Figure 42. The EVM of 64-QAM modulated signal at 16 dBm average power using memoryless baseband equivalent model.

6.3 Input/output impedance

Figure 43 illustrates a smith chart representation for the input and output impedance of the PA. Input impedance is about 160 Ω at resonance. Tunable capacitor is connected between the PPA PA interface to resonate the PPA DC feed inductor and PA input bias inductor. The capacitor bank has been designed to cover the required frequency range (5150…5895 MHz) for the 5G band PA according to Wi-Fi 6E specification.

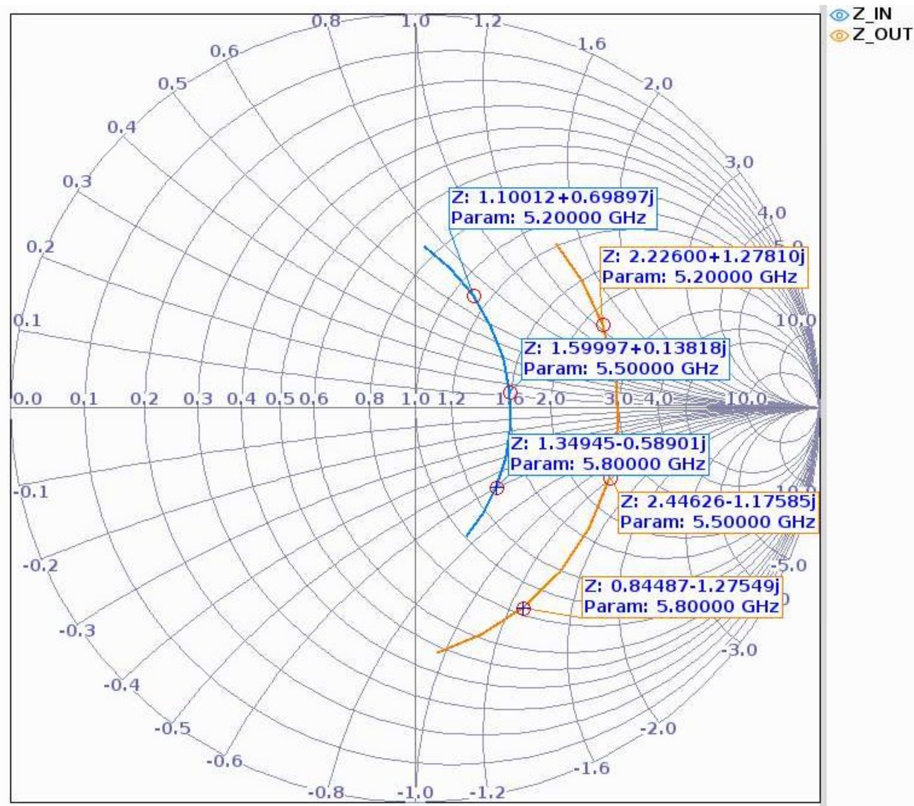

Figure 43. Input (Z_IN) and output impedance (Z_OUT). Reference impedance for the input impedance is 100 Ω and 50 Ω for the output impedance.

6.4 Stability factor

Figure 44 represent an ac simulation for the stability μ-factors. Analysis frequency range is from 1 MHz up to 100 GHz. Stability has been analyzed for both input and output and for differential and common mode stability. It can be determined that all the μ-factors remains ≥1 indicating that the nearest impedance resulting in potential instability lies outside the smith chart.

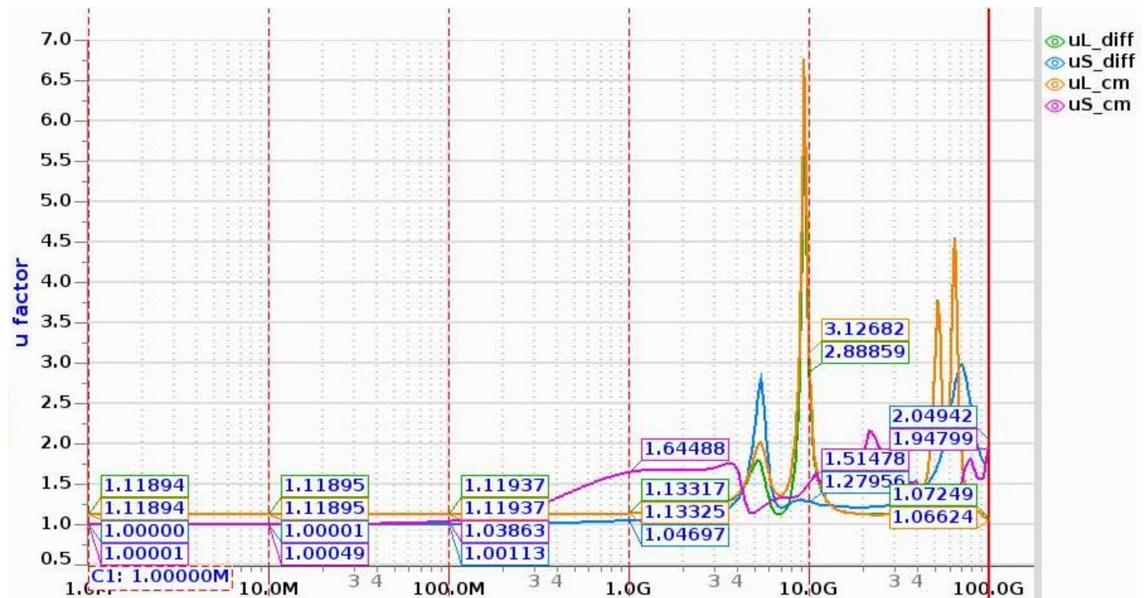

Figure 44. Differential and common-mode stability factors (µ) for the input/output plane.

6.5 Measurement results

During the writing of this thesis, the measurement for the transmitter was started and the work for finding the correct setups was not complete. This section presents the measurement results at the time of writing this thesis. The simulation results in Chapters 6.1…6.4 refer to the performance of PA only and a direct measurement of the exact same parameters is not possible as the measurement refers to the performance of the entire transmitter consisting of multiple building blocks. However, based on the system partitioning, the assumption is that the linearity performance of the PA dominates the overall linearity performance of the transmitter. Additionally, the DC characteristics of the PA core can be obtained by measuring the supply current for the PA, and the DC voltage at the gate of the PA core input transistor that is routed to an external test pin through a multiplexer (MUX). The DC characteristics and the PA core temperature is shown in Figure 45. The simulation uses the typical process corner and temperature is fixed 80 °C. As seen the difference between the simulation and the measurement is visible. It must be noted that the temperature of the measurement varies over the V_{gs} unlike the simulation that

has a fixed temperature of 80 °C. However, the measured DC current at 80 °C should be identical with the measurement if the core matched perfectly with the simulation model. Possible explanation for the difference is a process variation.

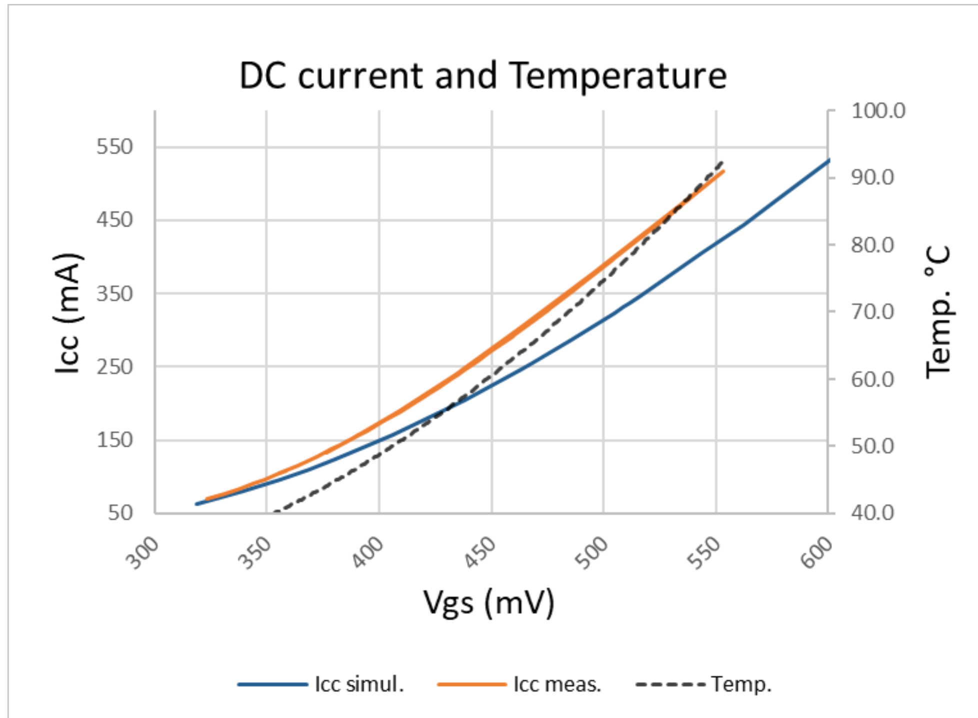

Figure 45. A simulated and measured DC characteristics and a measured temperature as a function of gate-to-source voltage *Vgs*.

Figure 46 shows *OIP3* and output power measurement as a function of frequency. The *OIP3* peaks at about 35 dBm and the variation over the band is less than 3 dB. Figure 47 and Figure 48 shows the output power and *EVM* for MCS0 modulation respectively with different drive levels. *EVM* remains well below the specification –18 dB.

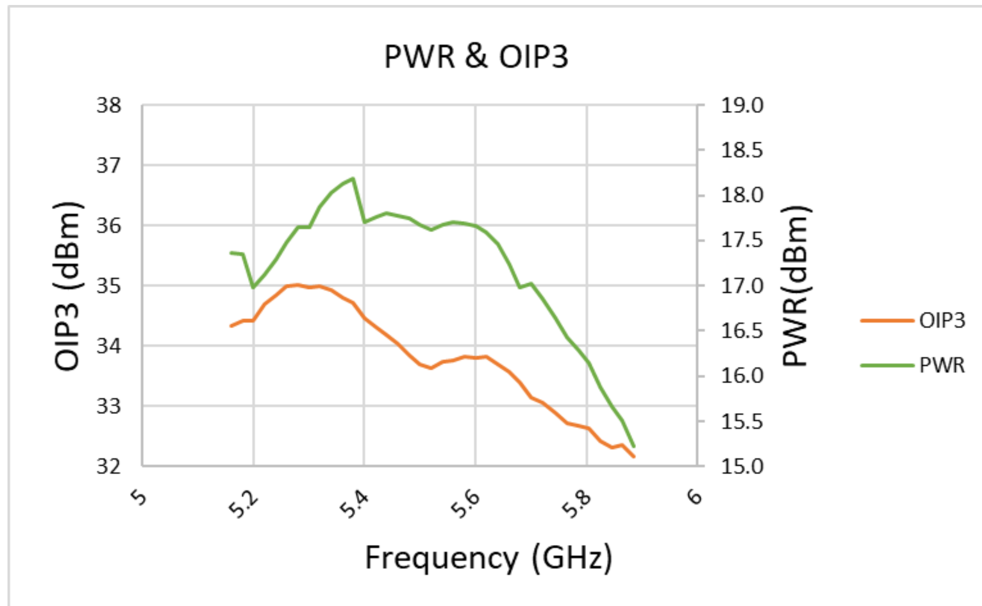

Figure 46. Output power and OIP_3 as a function of frequency.

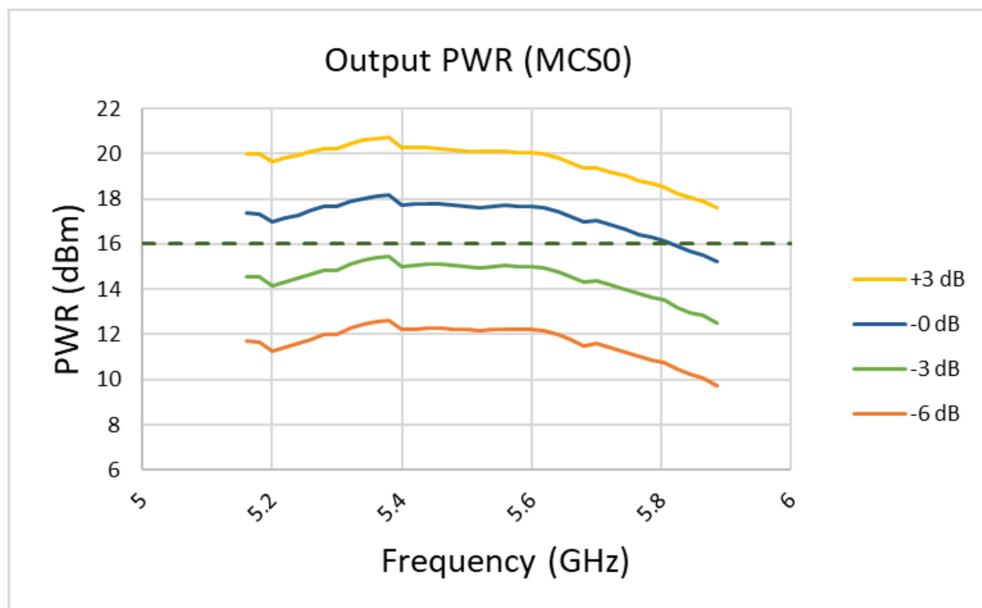

Figure 47. Output power of 20 MHz BPSK modulated signal with different drive levels.

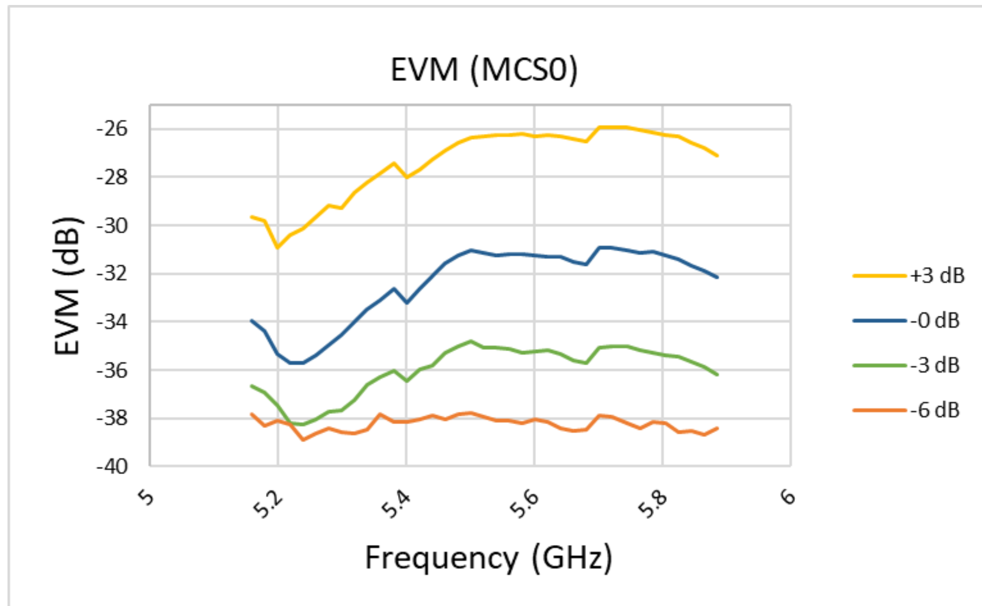

Figure 48. EVM of 20 MHz BPSK modulated signal with different drive levels.

Figure 49 and Figure 50 shows the output power and *EVM* for MCS7 modulation respectively with different drive levels. The *EVM* specification for MCS7 is –27 dB and indicated by the dashed line in Figure 50. As can be seen the maximum output power for MCS7 while meeting the *EVM* specification is around 18 dBm.

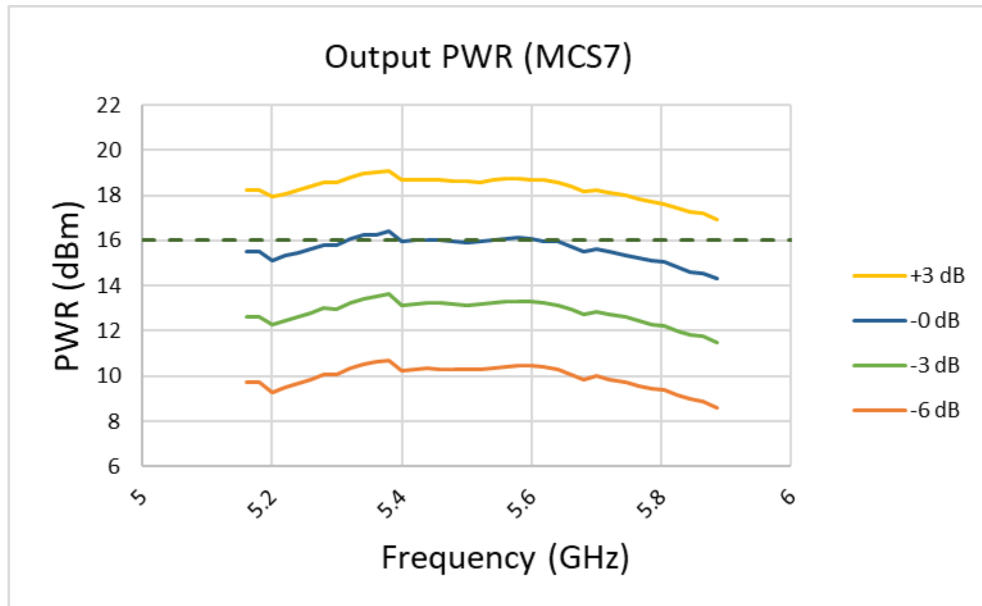

Figure 49. Output power of 20 MHz 64-QAM modulated signal with different drive levels.

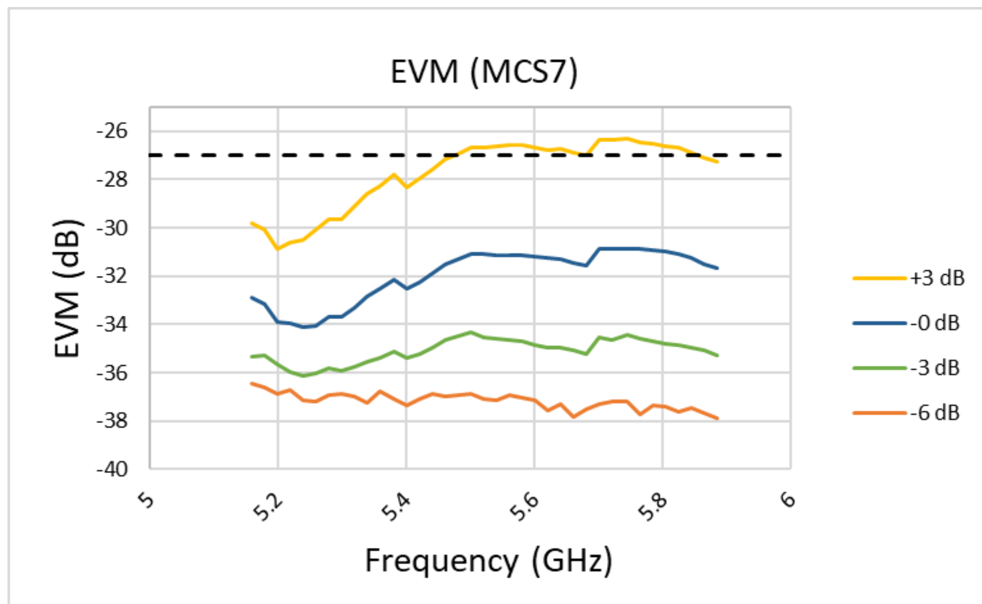

Figure 50. EVM of 20 MHz 64-QAM modulated signal with different drive levels.

7 CONCLUSIONS

In this study, the focus was on the feasibility of an integrated Wi-Fi 6 RF PA designed into a transceiver test chip, utilizing CMOS22 semiconductor fabrication technology. The primary objective was to investigate the feasibility of integrating the power amplifier into a CMOS22 SoC to potentially eliminate the need for external and costly PA modules. First a literature review was conducted, the design was presented, and simulations were conducted to evaluate its performance against target specifications.

The simulation results have provided valuable insights into the capabilities and limitations of the integrated power amplifier. While some aspects of the performance met or slightly deviated from the target specifications, there were notable challenges encountered. Voltage and power gains, while slightly below the targets, demonstrated promise, with proposed adjustments offering potential improvement.

However, the most significant obstacle identified was the linearity performance, specifically the IM3 linearity, which fell few dBs short of the target specification. Linearity is a crucial factor in RF power amplifiers, particularly for applications such as Wi-Fi 6, where signal integrity is crucial. Addressing this shortfall in linearity will important when taking the next steps.

In conclusion, while the integrated Wi-Fi 6 power amplifier shows promise for integration into the SoC, further research and development efforts are necessary to address the identified shortcomings, particularly in linearity performance. The next important step is to study any possible performance differences between the simulated and measured results and find the root causes. Possible reasons for the differences are insufficient modeling of the layout parasitics/electromagnetics or there might be shortcomings on the accuracy of the transistor models provided by the technology vendor especially at RF frequencies.

This study contributes to the growing body of knowledge in the integrated RF power amplifier design and underscores the importance of evaluation and optimization in achieving desired performance metrics for advanced wireless communication technologies such as Wi-Fi 6E.

REFERENCES

[1] The Beacon. ”Wi-Fi® by the numbers: Technology momentum in 2023.” fi-fi.org. https://www.wi-fi.org/beacon/the-beacon/wi-fi-by-the-numbers-technology-momentum-in-2023

[2] A. Salemi, “Silicon Carbide technology for high and ultra high voltage bipolar junction transistors and PiN diodes,” Ph.D. dissertation, Dept. Inform. and Comm. Tech., KTH Royal Institute of Tech. Stockholm, Sweden, 2017. [Online]. Available: urn:nbn:se:kth:diva-197913

[3] Adel S. Sedra and Kenneth C. Smith, *Microelectronics Circuit, 6*th ed. Oxford, United Kingdom, Oxf. Univ. Press, 2011

[4] P. G. Neudeck, R. S. Okojie and Liang-Yu Chen, "High-temperature electronics - a role for wide bandgap semiconductors?," in Proceedings of the IEEE, vol. 90, no. 6, pp. 1065-1076, June 2002, doi: 10.1109/JPROC.2002.1021571

[5] R. Caverly, *CMOS RFIC Design Principles*, 1st ed. Artech House, 2007.

[6] ”The Difference Between GaAs and GaN RF Power Amplifiers.” eliterfllc.com. https://eliterfllc.com/the-difference-between-gan-and-gaas

[7] A. Katz and M. Franco, "GaN Comes of Age," in IEEE Microwave Magazine, vol. 11, no. 7, pp. S24-S34, Dec. 2010, doi: 10.1109/MMM.2010.938583

[8] S. M. Sze and M. K. Lee, *Semiconductor devices: Physics and technology*, 3rd ed. John Wiley & Sons, 2012.

[9] D. Zapata, ”2 GHz +14 dBm CMOS power amplifier for Low Powe Wide Area Networks,” M.S. thesis, Fac. Inform. Tech., Univ. of Oulu, Finland 2020. [Online]. Available: https://urn.fi/URN:NBN:fi:oulu-202007042728

[10] H. Wang, F. Wang, S. Li, T.-Y. Huang, A. S. Ahmed, N. S. Mannem, J. Lee, E. Garay, D. Munzer, C. Snyder, S. Lee, H. T. Nguyen and M. E. D. Smith, ”Power Amplifiers Performance Survey 2000-Present,” Available: https://gems.ece.gatech.edu/PA_survey.html (accessed Feb. 1, 2024).

[11] H. Wang and K. Sengupta, *RF and Mm-Wave Power Generation in Silicon*, 1st ed. Elsevier Science & Technology, 2015.

[12] Intel Corporation. ”Intel Research Fuels Moore’s Law and Paves the Way to a Trillion Transistors by 2030.” intel.com. https://www.intel.com/content/www/us/en/newsroom/news/moores-law-paves-way-trillion-transistors-2030.html (accessed Feb. 1, 2024).

[13] M. K. Kazimierczuk, *RF Power Amplifiers*, 2nd ed. John Wiley & Sons, Incorporated, 2014.

[14] E. H. Armstrong, "A Method of Reducing Disturbances in Radio Signaling by a System of Frequency Modulation," in Proceedings of the Institute of Radio Engineers, vol. 24, no. 5, pp. 689-740, May 1936, doi: 10.1109/JRPROC.1936.227383

[15] A. B. Carlson, P. B. Crilly and J.C Rutledge, *Communication systems : An introduction to signals and noise in electrical communication*, 4th McGraw-Hill, 2002.

[16] P. B. Kenington, *High-linearity RF amplifier design*, Artech House, 2000.

[17] T. Ilari, ”Digital Predistortion Linearization Methods for RF Power Amplifiers, Doctoral Dissertation,” Ph.D. dissertation, Dept. Micro and Nanosciences, Univ. Tech. Espoo, Finland 2008. [Online]. Available: https://urn.fi/URN:ISBN:978-951-22-9546-3

[18] S. Cripps, *RF Power Amplfiers*, 2nd ed. Artech House, 2008.

[19] J. Vuolevi and T. Rahkonen, *Distortion in RF Power Amplifiers*, 1st ed. Artech House, 2003.

[20] H. Ku, M. D. McKinley and J. S. Kenney, "Quantifying memory effects in RF power amplifiers," in IEEE Transactions on Microwave Theory and Techniques, vol. 50, no. 12, pp. 2843-2849, Dec. 2002, doi: 10.1109/TMTT.2002.805196.

[21] B. Razavi, *RF Microelectronics*, Pearson Education, 1998.

[22] M. Trott, *The Mathematica Guidebook for Programming*, Springer New York, 2004.

[23] O. Väänänen, “Digital modulators with crest factor reduction techniques,” Ph.D. dissertation, Dept. Elect. Comm. Eng., Univ. Tech. Espoo, Finland 2006. [Online]. Available: https://urn.fi/urn:nbn:fi:tkk-006673

[24] B. Witte, ”OFDMA improves spectrum use in Wi-Fi 6,” 5gtechnologyworld.com. https://www.5gtechnologyworld.com/ofdma-improves-spectrum-use-in-wi-fi-6 (accessed Feb. 1, 2024).

[25] G. Poitau and A. Kouki, "MILC: Modified Implementation of the LINC Concept," 2006 IEEE MTT-S International Microwave Symposium Digest, San Francisco, CA, USA, 2006, pp. 1883-1886, doi: 10.1109/MWSYM.2006.249782

[26] F. Sechi and M. Bujatti, *Solid-State Microwave High-Power Amplifiers*, 1st ed. Artech House, 2009.

[27] Keysight Technologies, ”Modulation Accuracy-EVM,” keysight.com. https://www.keysight.com/fi/en/lib/resources/user-manuals/modulation-accuracyevm-332561.html (accessed Feb. 1, 2024).

[28] Cadence Design Systems Inc., ”The RF Stability Factor Explained,” cadence.com. https://resources.pcb.cadence.com/blog/2024-the-rf-stability-factor-explained (accessed Feb. 1, 2024).

[29] M. L. Edwards and J. H. Sinsky, "A new criterion for linear 2-port stability using a single geometrically derived parameter," in IEEE Transactions on Microwave Theory and Techniques, vol. 40, no. 12, pp. 2303-2311, Dec. 1992, doi: 10.1109/22.179894

[30] Cisco Systems Inc., ”Wi-Fi 6: The Next Generation of Wireless,” White Paper, 2019. [Online]. Available: https://meraki.cisco.com/lib/pdf/meraki_whitepaper_wifi6.pdf

[31] X. Zhou, ”IP Network eBook Series WIFI6,” Huawei Technologies, 2021. [Online]. Available: https://support.huawei.com/enterprise/en/doc/EDOC1100201276

[32] *Transmitting equipment for the Amplitude Modulated (AM) sound broadcasting service; Harmonised Standard covering the essential requirements of article 3.2 of Directive 2014/53/EU,* ETSI EN 302 017 2014/53/EU, ETSI, December 2016

[33] A. Buritica, ”802.11ax – High Efficiency Wireless,” microwavejournal.com. https://www.microwavejournal.com/articles/27221-11ax-high-efficiency-wireless (accessed Feb. 1, 2024).

[34] "IEEE Standard for Information Technology--Telecommunications and Information Exchange between Systems Local and Metropolitan Area Networks--Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 1: Enhancements for High-Efficiency WLAN," in IEEE Std 802.11ax-2021 (Amendment to IEEE Std 802.11-2020) , vol., no., pp.1-767, 19 May 2021, doi: 10.1109/IEEESTD.2021.9442429

[35] M. M. Hella and M Ismail, *RF CMOS Power Amplifiers: Theory, Design and Implementation*, 1st Springer, 2001.

[36] Qorvo Inc., ”QM11024 Broadband High Linearity DP4T Routing Switch,” qorvo.com. https://www.qorvo.com/products/p/QM11024 (accessed Feb. 1, 2024).

[37] J. R. Long, "Monolithic transformers for silicon RF IC design," in IEEE Journal of Solid-State Circuits, vol. 35, no. 9, pp. 1368-1382, Sept. 2000, doi: 10.1109/4.868049