FACULTY OF INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING
DEGREE PROGRAMME IN WIRELESS COMMUNICATIONS ENGINEERING

# MASTER'S THESIS

# 2 GHz +14 dBm CMOS power amplifier for Low Power Wide Area Networks

| | |
|---|---|
| Author | David Zapata |
| Supervisor | Aarno Pärssinen |
| Second Examiner | Timo Rahkonen |
| Technical Advisor | Marko Pessa |

June 2020

**Zapata D. (2020) 2 GHz +14 dBm CMOS power amplifier for Low Power Wide Area Networks.** University of Oulu, Faculty of Information Technology and Electrical Engineering, Degree Programme in Wireless Communications Engineering. Master's Thesis, 95 p.

# ABSTRACT

**The design of a radiofrequency power amplifier (RF PA) for narrowband low-power wide area networks is presented in this thesis. Particularly, this RF PA is compliant with the 3GPP TS 36.101 standard for a NB1 device within the Power Class 6. To minimize silicon area consumption, this CMOS RF PA employs a single-ended single-stage topology, avoiding inter-stage matching network inductors and output baluns. This RF PA produces +14 dBm of output power with a PAE of 25% and an EVM better than 4% (-28 dB). Also, its out-of-band and spurious emissions satisfy the standard specifications with a large margin. Furthermore, it provides high ruggedness, tolerating an antenna mismatch with a VSWR of 8:1.**

**Key words: linear power amplifier, integrated circuit, single-ended, LDMOS, CMOS, Narrowband IoT.**

# TABLE OF CONTENTS

# FOREWORD

I came to Oulu with the goal of learning everything about RF electronics. However, after writing this thesis and learning a little bit about RF power amplifiers, I realized that I am far away from this goal. As physicist John Wheeler said, "*we live on an island surrounded by a sea of ignorance. As our island of knowledge grows, so does the shore of our ignorance*".

I have big appreciation for Aarno Pärssinen, who gave me the opportunity to come to Oulu; also, for Tarmo Ruotsalainen and Olli Narhi for giving me the chance of testing my skills as a summer trainee in Nordic Semiconductor. I am also deeply grateful to Mikko Lintonen who supervised my summer internship with lots of patience.

Finally, I have immense admiration for Marko Pessa, my master thesis technical supervisor, who is an electronics wizard. I hope someday to be like him.

Oulu, June, 2020

David Zapata

# LIST OF ABBREVIATIONS AND SYMBOLS

| | |
|---|---|
| ACLR | Adjacent channel leakage ratio |
| AC | Alternating current |
| AM | Amplitude modulation |
| BJT | Bipolar junction transistor |
| MB | Center of the frequency band |
| CG | Common-gate |
| CS | Common-source |
| CMOS | Complementary metal oxide semiconductor technology |
| DC | Direct current |
| EVM | Error vector magnitude |
| FET | Field effect transistor |
| H4 | Fourth harmonic frequency |
| GaN | Gallium Nitride |
| GaAs | Gallium Arsenide |
| HBT | Heterojunction bipolar transistor |
| HEMT | High electron mobility transistor |
| HB | High end of the frequency band |
| I/O | Input/Output |
| IC | Integrated circuit |
| LDMOS | Lateral diffused MOSFET |
| LB | Low end of the frequency band |
| MOS | Metal oxide semiconductor |
| MOSFET | Metal oxide semiconductor field effect transistor |
| MESFET | Metal semiconductor field effect transistor |
| OMN | Output matching network |
| PM | Phase modulation |
| PAE | Power added efficiency |
| PA | Power amplifier |
| PSD | Power spectral density |
| PCB | Printed circuit board |
| QAM | Quadrature amplitude modulation |
| Q | Quality factor |
| RF | Radio frequency |
| RFPA | Radio frequency power amplifier |
| H2 | Second harmonic frequency |
| SRF | Self-resonance frequency |
| SC2H | Short circuit at the second harmonic |
| Si | Silicon |
| OIP3 | Third-order intercept point referred to the output |
| H3 | Third harmonic frequency |
| VSWR | Voltage standing wave ratio |

| | |
|---|---|
| $\omega$ | Angular frequency |
| k | Balun coupling coefficient between windings |
| $L_m$ | Balun magnetizing inductance |

| | |
|---|---|
| $\mathrm{M}$ | Balun mutual inductance |
| $N$ | Balun turns ratio |
| $I_{bias}$ | Bias current |
| $I_{\mathrm{drain,bias}}$ | Bias drain current |
| $R_{bias}$ | Bias resistor for the constant transconductance circuit |
| $V_{bias,CG}$ | Bias voltage for the CG device |
| $V_{BD}$ | Breakdown voltage |
| $CS_{sample}$ | Common-source transistor of the PA sample |
| $P_{DC}$ | DC dissipated power |
| $V_{DG,CG}$ | DC drain-to-gate voltage for the CG device |
| $V_{DS,CS}$ | DC drain-to-source voltage for the CS device |
| $V_{ov}$ | DC overdrive voltage |
| $C_{dec}$ | Decoupling capacitance |
| $I_{ds}$ | Drain-to-source current amplitude |
| $I_{ds,1}$ | Drain-to-source current amplitude at the fundamental frequency |
| $V_{ds}$ | Drain-to-source voltage amplitude |
| $\mu_n$ | Electron mobility |
| $R_{fb}$ | Feedback resistor |
| $\Delta\omega$ | Frequency offset |
| $V_{GS,breakdown}$ | Gate-to-source breakdown voltage |
| $V_{gs}$ | Gate-to-source voltage |
| $Z_{1H}, Z_{2H}, Z_{3H}, Z_4$ | Impedance at the fundamental, second, third and fourth harmonic, respectively |
| $L_1$ | Inductance of the primary winding |
| $L_2$ | Inductance of the secondary winding |
| $C_{in}$ | Input capacitance of the PA |
| $R_{in}$ | Input resistance of the PA |
| $V_{knee}$ | Knee voltage |
| $I_{L,1}$ | Load current amplitude at the fundamental frequency |
| $R_L$ | Load resistance |
| $V_{L,1}$ | Load voltage amplitude at the fundamental frequency |
| $V_{max,DG,CG}$ | Maximum $V_{DG,CG}$ |
| $V_{ds,max}$ | Maximum drain-to-source voltage |
| $V_{ds,min}$ | Minimum drain-to-source voltage |
| $R_{opt,Mod}$ | Optimum load impedance of the modulator |
| $R_{L,opt}$ | Optimum load resistance |
| $C_{out,PA}$ | Output capacitance of the PA |
| $I_{out}$ | Output current |
| $P_{out}$ | Output power |
| $P_{out,Mod,1-dB}$ | Output power delivered by the modulator at the 1-dB compression point |
| $L_{gnd,pkg}$ | Parasitic inductance of the package in the ground node |
| $L_{out,pkg}$ | Parasitic inductance of the package in the RF output path |
| $R_{gnd,pkg}$ | Parasitic resistance of the package in the ground node |
| $R_{out,pkg}$ | Parasitic resistance of the package in the RF output path |
| $I_{peak}$ | Peak RF current |
| $G_{PA}$ | Power amplifier power gain |

| | |
|---|---|
| $P_{in,PA,1-dB}$ | Power at the input of the PA at the 1-dB compression point |
| $P_{L,1}$ | Power delivered to the load at the fundamental frequency |
| $\eta$ | Power efficiency |
| $V_{dd}$ | Power supply voltage |
| $X_{Cout}$ | Reactance of the output capacitance of the PA |
| $\Gamma$ | Reflection coefficient |
| $i_{Mod,1-dB}$ | RF current delivered by the modulator at the 1-dB compression point |
| $i_{drain,1-dB}$ | RF drain current at 1-dB compression point |
| $i_{drain,RMSmod}$ | RF drain current at the RMS power of the modulation |
| $v_{drain,1-dB}$ | RF drain voltage at 1-dB compression point |
| $v_{inc}$ | RF incident voltage amplitude |
| $v_{inc,OMN}$ | RF incident voltage amplitude in the output matching network |
| $P_{RF\ in}$ | RF input power |
| $P_{RF\ out}$ | RF output power |
| $P_{drain,1-dB}$ | RF power at the 1-dB compression point measured at the drain terminal |
| $P_{drain,RMSmod}$ | RF power at the RMS power of the modulation measured at the drain terminal |
| $v_{ref}$ | RF reflected voltage amplitude |
| $v_{ref,OMN}$ | RF reflected voltage amplitude in the output matching network |
| $v_{in,RMS}$ | RF RMS voltage at the PA input |
| $v_{in,1-dB}$ | RF voltage at the input of the PA at the 1-dB compression point |
| $v_{in,peak}$ | RF voltage at the PA input |
| $S_{11}$ | S-parameter from port 1 to port 1 |
| $S_{21}$ | S-parameter from port 1 to port 2 |
| $\Delta IM$ | Third-order harmonic suppression |
| $V_{Th}$ | Threshold voltage |
| $g_m$ | Transconductance |
| $g_{m,CG}$ | Transconductance of the CG transistor of the PA |
| $g_{m,CS}$ | Transconductance of the CS transistor of the PA |
| $L$ | Transistor channel length |
| $W$ | Transistor width |
| $C_T$ | Tuning capacitance of the balun |

# 1 INTRODUCTION

The Internet of Things (IoT) is generally understood as a model in which network connectivity and computing capability is extended to a variety of objects, devices and everyday items allowing them to generate, exchange and consume data with minimal human intervention [1]. Major telecommunication companies, such as Cisco and Ericsson, estimate about ten billion of IoT devices connected in the first few years of the 2020 decade [2]. Applications of IoT reach over a large variety of scenarios, such as [1]:

- Human: IoT devices attached or inside the human body to monitor and maintain human health and productivity.
- Home: Connected devices controlling home appliances and managing security.
- Retail commerce: IoT devices supporting self-checkout, real-time offers and inventory optimization in stores, banks, and restaurants.
- Offices: IoT for improved energy management and security in office buildings.
- Factories: IoT devices enhancing operation efficiency, optimization of manufacturing equipment and inventory.
- Transportation: IoT devices are present in autonomous vehicles, real-time routing, shipment tracking and condition-based vehicle maintenance.
- Cities: Adaptive traffic control, environmental monitoring and resource management can be leveraged by using IoT.

One of the technologies that enables IoT is NB-IoT (narrowband IoT). This technology is designed to support a massive number of ultra-low-cost low-power devices connected to a modern LTE cellular network. It is important to note that long-rage wireless networks composed of low power devices (such as NB-IoT devices) are known as Low Power Wide Area Networks (LPWAN) [2] as conceptualized in Figure 1.

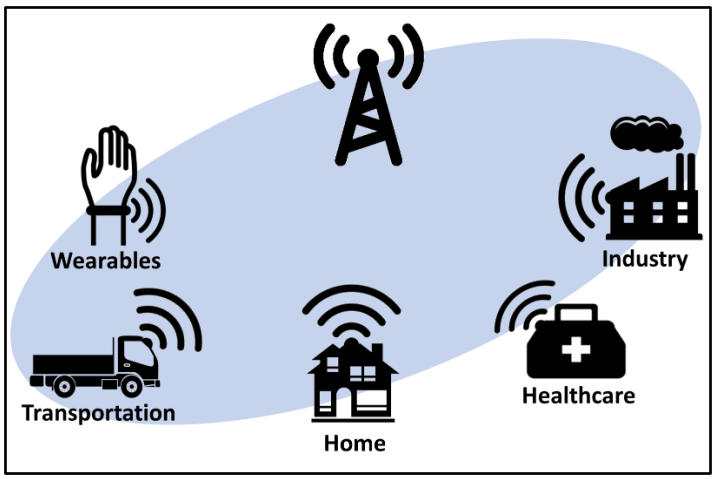


Figure 1: Conceptual diagram of IoT devices in a LWPAN.

An important part of an IoT device (and in fact, of any radio-communication device) is the radio frequency power amplifier (RF PA). This electronic component amplifies the signal “to be transmitted” to a high-power level so it can reach far distances [3].

This thesis presents the design of a NB-IoT RF PA. This RF PA is compliant with the 3GPP standard TS 36.101 in the category of NB1 devices (NB-IoT) in the Power Class 6. This

standard sets stringent requirements for linearity and therefore the design process is focused on achieving them. Moreover, typical PA requirements regarding ruggedness, power efficiency and stability are considered here.

This work also explores the idea of minimizing silicon area by avoiding integrated inductors. This results in a RF PA employing a single-ended single-stage topology. This topology comes with limitations that are examined here.

This document is structured as follows. The main semiconductor technologies for RF PA implementation are outlined in Chapter 2, showing the limitation of CMOS for this application. Chapter 3 describes the key theoretical concepts related to RF PAs, such as reliability, load-line impedance matching and gain compression. Moreover, the classical classes for linear PA are described. The design specifications are detailed in Chapter 4. These specifications are established by the 3GPP standard and by the team of engineers responsible of creating the other blocks of the transmitter. A thorough design workflow for the linear RF PA is presented in Chapter 5. Each block of the PA is examined, and its main parameters are computed based on the design specifications. The test results are given in Chapter 6, using visual aids to convey a better idea about the performance of the PA. Finally, the conclusions are offered in Chapter 7, in which a brief discussion about the design process and possible improvements of the PA performance.

# 2 RELEVANT TECHNOLOGIES FOR INTEGRATED RF POWER AMPLIFIERS FOR MOBILE APPLICATIONS

## 2.1 Semiconductor materials

Semiconductor materials used in the manufacturing of integrated circuits determine the electrical and thermal properties of the transistors. Table 1 contains the most relevant properties of the most common semiconductor materials for the fabrication of RF power electronics [4].

The maximum DC supply voltage for a transistor depends mainly on the bandgap energy and the breakdown electric field of its construction material. A larger supply voltage allows a larger optimum load resistance for a RF power amplifier (as shown in sections 3.5 and 3.7), which improves the performance of the output matching network, reduces electromigration and self-heating issues on the integrated passive elements and allows smaller power transistor size for a given output power. In this respect, Gallium Nitrite (GaN) devices offer largely superior capabilities than Silicon (Si) and Gallium Arsenide (GaAs).

Regarding operation at high temperatures, a larger bandgap energy is most desirable [4]. From Table 1 it can be seen that GaN is also superior in this aspect. Moreover, heat dissipation, which is also crucial for high power applications, is determined by the thermal conductivity of the material. Silicon and GaN have comparable performance in this regard.

Transconductance is also a crucial characteristic for RF devices since it determines in part the maximum operating frequency for a transistor as well as the required MOS transistor size for a specific output current. Since the transconductance of a transistor depends on the electron mobility of its construction material [5], a high mobility is desired. In this regard, GaAs exhibit the largest mobility. However, the higher velocity saturation of GaN compensates in part its low electron mobility for high frequency applications, making its performance comparable to that of GaAs [6].

Table 1: Physical properties of semiconductors used in transistor manufacturing

| | **Si** | **GaAs** | **GaN** |
|---|---|---|---|
| **Bandgap energy** [$eV$] | 1.12 | 1.43 | 3.4 |
| **Electron mobility** [$cm^2/Vs$] | 1350 | 8500 | 1200 |
| **Electron velocity saturation** [$10^7 cm/s$] | 1 | 1 | 2.5 |
| **Breakdown electric field** [$MV/cm$] | 0.25 | 0.3 | 3 |
| **Thermal conductivity** [$WK/cm$] | 1.5 | 0.5 | 1.3 |

From the previous analysis, it can be concluded that GaN is the best choice for RF power applications. However, GaAs offers a slight RF performance advantage over GaN for lower power applications because of its high electron mobility [7]. Moreover, silicon does not offer the physical properties required for RF power applications.

Moreover, Figure 2 shows the trending lines for RF power amplifiers for different semiconductor technologies [8]. The information was obtained by surveying more than 2500 scientific papers published after the year 2000. Each line represents an upper bound for the development of power amplifiers by considering the saturated power and operating frequency as performance indicators. In other words, most of the RF power amplifiers implemented using a given technology lie below the corresponding line. The data show that silicon CMOS development on RF power amplifiers falls behind all other technologies when analysing saturated power for all the frequency spectrum.

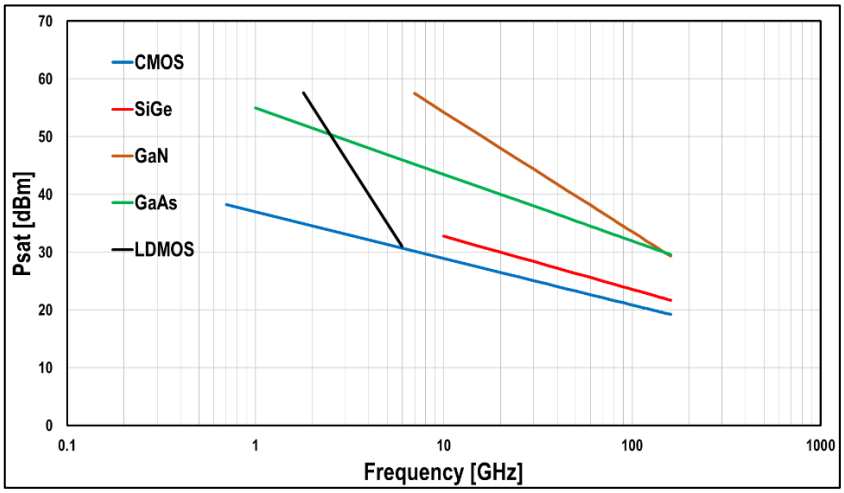



Figure 2: Trending lines for the RF power amplifiers implemented with different semiconductor technologies [8].

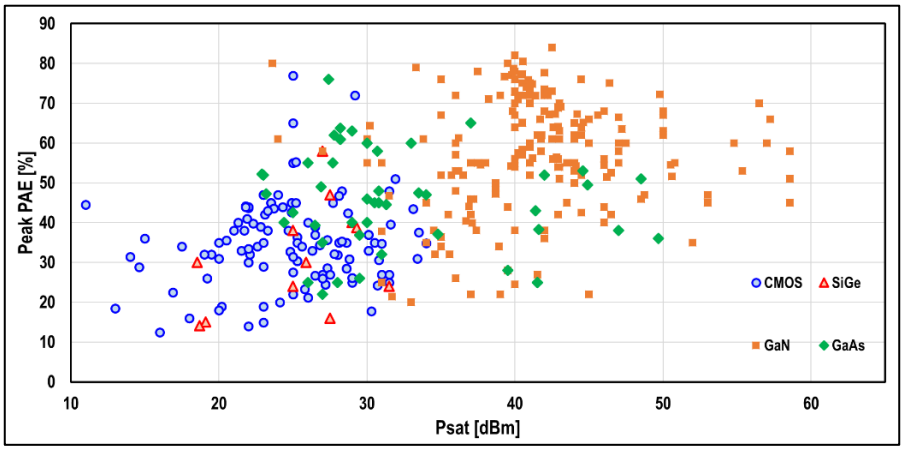



Figure 3: PAE vs saturated power for RF power amplifiers in the 2 - 6 GHz frequency band [8].

Finally, Figure 3 shows the power added efficiency (PAE) for RF power amplifiers operating in the 2 - 6 GHz frequency band for different semiconductor technologies (taken from the same study). In this case, silicon CMOS implementations have the lowest power efficiency for a given output power compared to other RF power amplifier technologies.

## 2.2 Categories of transistors for RF power amplifiers

Most transistors can be categorized as bipolar junction transistors or as field effect transistors. The main characteristics of these devices regarding RF power applications are presented next.

### *2.2.1 Bipolar junction transistor (BJT):*

The output current of a BJT is controlled by its input current. This type of transistor is not appropriate for applications that require high-frequency and high-power at the same time. For instance, large currents are achieved by increasing the size of the emitter terminal which generates as a side effect a large input parasitic capacitance, impairing the maximum operating frequency. Also, high temperature increases the BJT output current which in turn increases

temperature, creating a positive feedback loop known as *thermal runaway*. This issue can be mitigated by some techniques that compromise gain and efficiency [7].

### *2.2.2 Field effect transistor (FET):*

The output current of a FET is controlled by its input voltage. These devices are more appropriate than BJT for high-power and high-frequency applications, since FET do not display many of the same trade-offs as BJT. For instance, FET do not exhibit thermal runaway. Also, increased output current can be achieved by increasing the width of the device, without impairing the maximum operating frequency [7].

## 2.3 IC manufacturing technologies for RF power amplifiers

Many different technologies have been developed for RF power applications. These applications require high speed, large gain, reliable operation at higher voltages, and good thermal performance, all at the same time, making it challenging to produce a transistor that is up to the task.

A brief description of the physical construction of the most relevant RF power IC technologies is presented next, as well as their key advantages and drawbacks.

### *2.3.1 Metal-Oxide-Semiconductor field effect transistor (MOSFET):*

This device allows controlling the charge distribution of a segment of the substrate by means of adjusting the voltage of an oxide-semiconductor interface. This charge distribution can be used to control the current flow between two terminals that are at different potentials [5]. Silicon is typically used for MOSFET manufacturing since this material allows n-channel and p-channel transistors of similar performance for digital circuits [9]. Advanced improvements in MOSFET technology, such as silicon-on-insulator, highly reduce substrate parasitics enabling higher operating frequencies [10].

Silicon-based MOSFET is considered not an appropriate technology for integrated RF high-power applications because of low breakdown voltage. However, this technology is highly available and provides large scale integration at low cost, making it the default choice for mass production of consumer electronics accounting for 80% of the global production [11].

### *2.3.2 Lateral diffused MOSFET (LDMOS):*

This device is basically a MOSFET that allows high voltage operation for frequencies of about 5 GHz [9], compensating the short-comings of MOSFET for RF power applications while maintaining a low production cost [7].

High voltage operation is enabled by reducing the doping concentration below the drain terminal while increasing the distance between the drain and gate terminals (thus extending the size of the depletion region). This larger depletion region increases the maximum drain-to-source voltage before breakdown occurs [12]. Additionally, a grounded “field plate” surrounding the gate terminal blocks part of the electric field between drain and gate, which improves isolation between these terminals. This plate also enables increased drain-to-gate voltage before oxide breakdown occurs and reduces the gate-to-drain capacitance [7].

Besides offering better reliability, one of the main advantages of LDMOS is that these devices can be built in typical Si-MOS wafers without much change to the process itself, allowing PA integration and thus reducing cost compared to having an external PA.

Besides this, the main drawbacks are larger ON resistance (because of increased drain-to-source distance) [12], and lower speeds compared to Si-MOS [9].

#### *2.3.3 Metal-Semiconductor field effect transistor (MESFET):*

This device is basically a MOSFET with a metal-semiconductor rectifying contact (similar to those used in Schottky diodes) instead of the metal-oxide-semiconductor contact for the gate terminal. Additionally, its drain and source terminals are made of metal-semiconductor ohmic contacts instead of the p-n junctions used in MOSFETs.

This device has an almost flat I-V characteristic curve in the saturation region (almost infinite output resistance). Also, MESFETs exhibit a considerably high transconductance and maximum operating frequency three times larger than Si-MOSFET when a high electron mobility material (typically GaAs) is used for its fabrication [5].

One of the main shortcomings of this technology is that available wafers are too small, which reduces manufacturing yield and increases cost [10].

#### *2.3.4 High electron mobility transistor (HEMT):*

Also known as MODFET, TEGFET, SDHT or HFET, this device has been used as an alternative to MESFET for very-high frequency applications. It is basically a FET made of two semiconductors with different bandgap energies, which creates a channel with very low resistance. Typically, GaAs is used together with AlGaAs, reaching a maximum operating frequency 30% higher than GaAs MESFETs. Even higher frequencies (around 600 GHz) can be achieved by using two different alloys of Aluminium, Indium and Arsenic ($Al_xIn_yAs$) [5].

#### *2.3.5 Heterojunction bipolar transistor (HBT):*

This device is basically a BJT in which one or both p-n junctions are made out of two semiconductors with different badgap energies in order to increase its current gain, which improves its high-frequency performance. Most commonly, these devices are made using GaAs together with alloys of aluminium, gallium and arsenic ($Al_xGa_{1-x}As$).

A drawback of these devices is that they require a minimum collector-to-emitter voltage to start generating output current [5]. Additionally, this technology comes with low manufacturing yield and increased cost [10]. Despite these drawbacks, GaAs HBT is the technology of choice for power amplifiers over GaAs MESFET [10].

### 2.4 Integrated passive elements

Integrated RF power amplifiers require passive elements for AC coupling, AC bypassing, DC biasing, impedance matching, resonances, among other roles. Next, a brief description of the features and shortcomings of integrated capacitors and inductors is presented.

### *2.4.1 Capacitors*

Capacitors in CMOS processes can be classified in two categories: capacitors with plates made of polysilicon and capacitors with plates made of metal [13].

The most common type of integrated capacitors in the first category is the MOS capacitor. This type of device is constructed by inserting an N-well layer below a N-channel transistor. The top plate is the gate polysilicon and the bottom plate is the N-well layer. Since both plates are made of high resistivity materials, this capacitor offers low Q factor. This device provides very high capacitance density per unit area. It is usually used when one of the plates of the capacitor is grounded [13]. This type of device is readily available in any CMOS process.

Metal-insulator-metal (MiM) capacitors are constructed by separating two metal layers by a thin oxide layer, and thus belong to the second category. Capacitance density is lower than in the case of MOS capacitors. Manufacturing this type of capacitor requires additional masks and steps, adding complexity to the process.

Another type of integrated capacitor in category two is the metal-oxide-metal (MOM) type. This device is constructed by using the available metal layers of any CMOS process. Usually, the two plates of the capacitor consist of many interleaved fingers, taking advantage of horizontal and vertical capacitances. Although it has low capacitance per unit area, it comes with high Q, making it the device of choice for high frequency applications [13].

### *2.4.2 Inductors*

Integrated inductors are made of spirals of metal wires, since the mutual inductance between adjacent turns give a spiral more inductance than a straight line of wire of the same length [3]. Maximizing the inductance of an inductor is crucial because these devices occupy the largest area compared to other passives.

Integrated inductors come with these significant drawbacks [3]:

- Metal layers in integrated circuits are considerably thin, exhibiting a substantially small cross-section. This increases the inductor DC ohmic losses. Additionally, skin effect further increases AC ohmic losses. Increasing the width of the lines on the inductor mitigates this issue but at the cost of additional silicon area, lower total inductance and higher parasitic capacitance.
- Parasitic capacitances to the substrate and between adjacent turns heavily reduce the self-resonance frequency, affecting the maximum frequency at which the inductor can be used.
- Magnetic and capacitive coupling to the substrate further increase losses because of the resistive nature of the substrate.

These drawbacks are naturally extended to integrated transformers and baluns.

# 3 PRINCIPLES OF RF POWER AMPLIFIERS

## 3.1 Non-linearity in radio systems

Theoretically, a "static" system[1] with output $y(t)$ and input $x(t)$ is linear if [3]:

$$y(t) = \alpha x(t) \tag{1}$$

Where $\alpha$ is a constant. This means that the output of a linear system is a scaled version of its input.

Non-linear static systems are usually modelled by a series expansion of the output signal with respect to the input signal:

$$y(t) = \alpha_0 + \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t) \cdots \tag{2}$$

This representation is used in radio systems for modelling the most relevant non-linear effects, such as harmonic distortion, gain compression and intermodulation.

In radio systems, non-linearities create distortion in the transmitted signal which generates unwanted emissions that cause interference to other systems. Additionally, transmitted signal distortion makes recovery by the receiver end a more challenging task [9].

Non-linearities in the front-end of integrated RF transmitters are mainly generated by the cut-off and clipping behaviour of the transistor. Other effects such as the variation of parasitics and transconductance with respect to the voltages of the transistor terminals also play a crucial role. Furthermore, "memory effects" are non-linearities caused by dynamic deviations from the static behaviour of the transistor due to thermal effects, modulation of the power supply and semiconductor aging [14].

### *3.1.1 Harmonic distortion*

A non-linear system with a pure sinusoidal tone at its input will produce a linear combination of tones at "harmonic" frequencies (integer multiples of the input tone frequency) [3]. This can be shown by substituting $x(t)$ for $A\,cos(\omega t)$ in Equation 2 and using basic trigonometric identities

$$\begin{aligned} y(t) &= \alpha_1 A cos(\omega t) + \alpha_2 A^2 cos^2(\omega t) + \alpha_3 A^3 cos^3(\omega t) + \cdots \\ &= \alpha_1 A cos(\omega t) + \frac{\alpha_2 A^2}{2}\big(1 + cos(2\omega t)\big) + \frac{\alpha_3 A^3}{4}\big(3cos(2\omega t) + cos(3\omega t)\big) + \cdots \\ &= \frac{\alpha_2 A^2}{2} + \left(\alpha_1 A + \frac{3\alpha_3 A^3}{4}\right) cos(\omega t) + \frac{\alpha_2 A^2}{2} cos(2\omega t) + \frac{\alpha_3 A^3}{4} cos(3\omega t) + \cdots. \end{aligned} \tag{3}$$

These additional frequency components produce "harmonic distortion" at the output of the system.

As an example, a pure tone $A\,cos(\omega t)$ that has been half-rectified by a non-linear system can be expressed as [15]:

[1] In contrast to "dynamic" linear systems, whose output follow $y(t) = x(t) * h(t)$

$$y(t) = \frac{A}{\pi} + \frac{A}{2} sin(\omega t) - \sum_{n\ even} \frac{2A}{\pi(n+1)(n-1)} cos(n\omega t). \quad (4)$$

From this expression it can be seen that the half-rectification process (as any other non-linearity) creates new frequency components that were absent at the input.

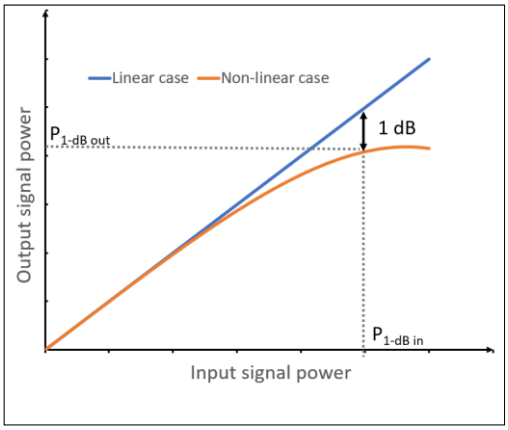
In radio systems, harmonic distortion is relevant because the harmonics of the transmitted signal can produce interference to users of other frequency bands. Therefore, the radio system must filter the harmonic distortion it produces, in accordance with the applicable standards and regulations.

Additionally, there are techniques that rely on altering the amplitude and phase of the harmonics produced by an RF power amplifier with the aim of improving performance. These techniques are generally known as "waveform engineering" [14].

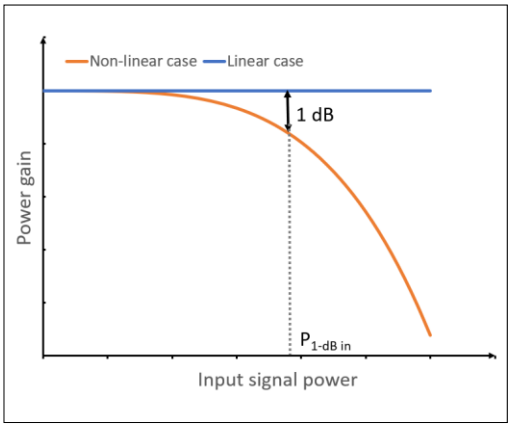
### *3.1.2 Compression gain*

From Equation 3, it can be seen that the amplitude of the output tone at the fundamental frequency is $\alpha_1 A + 3\alpha_3 A^3/4$ (when non-linearity is characterized up to the third-order term). Since the amplitude of the input tone is $A$, then the gain at the fundamental frequency is $\alpha_1 + 3\alpha_3 A^2/4$. It is interesting to note that if the coefficient $\alpha_1$ is positive and $\alpha_3$ is negative (or vice versa), then this gain will decrease as the input level $A$ increases. This effect is known as gain compression [3].

A typical figure of merit for the compression gain is the 1 dB compression point, defined as the input signal power level that causes a 1 dB of gain reduction. The 1 dB compression point can be easily seen in a logarithmic plot of the output power versus the input power or in a logarithmic plot of the power gain versus the input power, as shown in Figure 4.




(a) 1 dB compression point with respect to output power.




(b) 1 dB compression point with respect to power gain.

Figure 4: 1 dB compression point curves.

### *3.1.3 Intermodulation*

Another non-linear effect that can be modelled by Equation 2 is intermodulation. In general, if more than two tones at different frequencies are added at the system input, the output signal

will contain tones at frequencies different that the harmonics of the two input frequencies. This effect is known as intermodulation.

For the particular case of two tones with frequencies $\omega_1$ and $\omega_2$, the input signal is $x(t) = A_1 cos(\omega_1 t) + A_2 cos(\omega_2 t)$ and the output signal is [3]

$$y(t) = \alpha_1[A_1 cos(\omega_1 t) + A_2 cos(\omega_2 t)] + \alpha_2[A_1 cos(\omega_1 t) + A_2 cos(\omega_2 t)]^2 + \quad (5)$$

$$\alpha_3[A_1 cos(\omega_1 t) + A_2 cos(\omega_2 t)]^3 + \cdots.$$

After expanding the right-hand side of this equation and using trigonometric identities, it can be shown that the output signal is composed of tones at frequencies $m\omega_1 \pm n\omega_2$, where $m$ and $n$ are positive integers. This effect is known as *intermodulation*.

If $\omega_1$ and $\omega_2$ are close to one another, the output tones at frequencies $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ (commonly known as third-order intermodulation products) will appear close to the input tones, as shown in Figure 5.

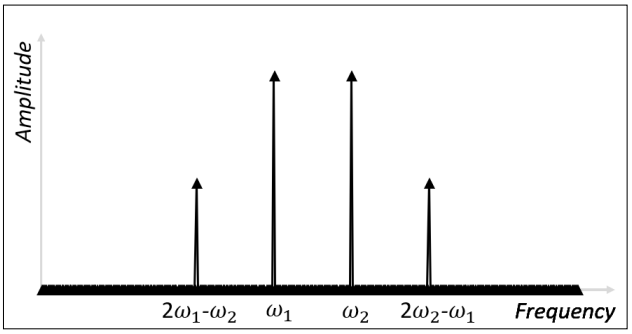



Figure 5: Third order intermodulation products in the output signal of a non-linear system fed with two pure tones at frequencies $\omega_1$ and $\omega_2$.

There are a few important implications of intermodulation for radio systems:

- In modulations using multiple carriers, intermodulation products between carriers can fall inside the signal bandwidth creating inter-carrier interference [16].
- Intermodulation products that fall in adjacent channels produce side bands that can generate interference to other radio channels. This effect is known as *spectral regrowth* since it looks like the signal bandwidth has been spread out [17]. Figure 6 shows an example of this phenomenon.
- An external signal can enter to a transmitter through its antenna and can mix with the transmitted signal generating intermodulation products that can cause interference to other radio channels.
- Third-order intermodulation products are difficult to filter out since they fall in the same frequency band as the transmitted signal [3].

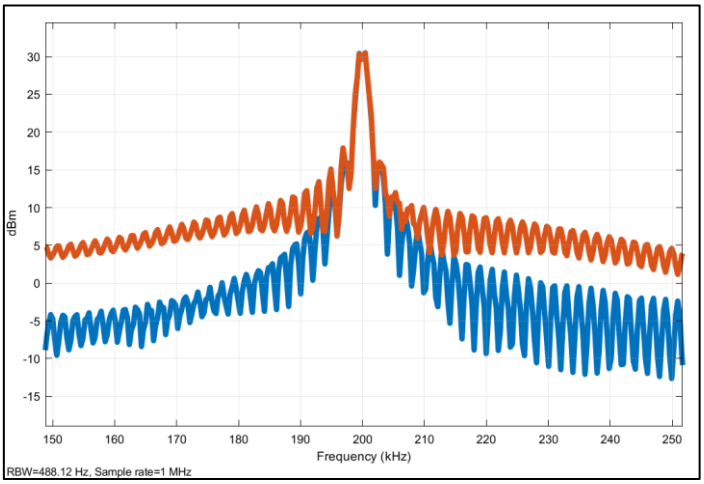



Figure 6: Spectral regrowth. Blue trace represents the power spectral density of an unfiltered 64-QAM modulated signal. Orange trace represents the same signal after being passed through a system with third-order non-linearity and no gain.

Two typical figures of merit for intermodulation in power amplifiers are the third-order intercept point (IP3) and the third-order harmonic suppression ($\Delta IM$), which are related to each other. For obtaining this values, two pure tones are fed into the input of the system, one at frequency $\omega_1$ (in the operating frequency band) and the other at a small frequency offset $\Delta\omega$. Both tones should have the same power and their combined power should not cause compression gain [3]. Then, as shown in Figure 7, the $\Delta IM$ is the ratio between the output power at frequency $\omega_1$ and the power of one of the third-order intermodulation products.

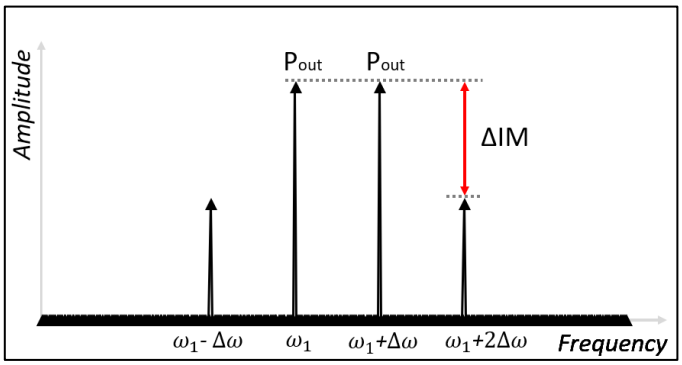



Figure 7: Two tone test for OIP3 and $\Delta IM$.

The IP3 referred to the output side (OIP3) can be obtained by using

$$OIP3 = \frac{\Delta IM}{2} + P_{out}. \quad (6)$$

Another figure of merit for intermodulation, more closely related to measuring spectral regrowth, is the adjacent channel leakage ratio (ACLR). It is obtained by measuring the power spectral density of the transmitted signal and calculating the ratio between the power contained in its assigned channel and the power contained in a nearby channel [17]. This is illustrated in Figure 8.

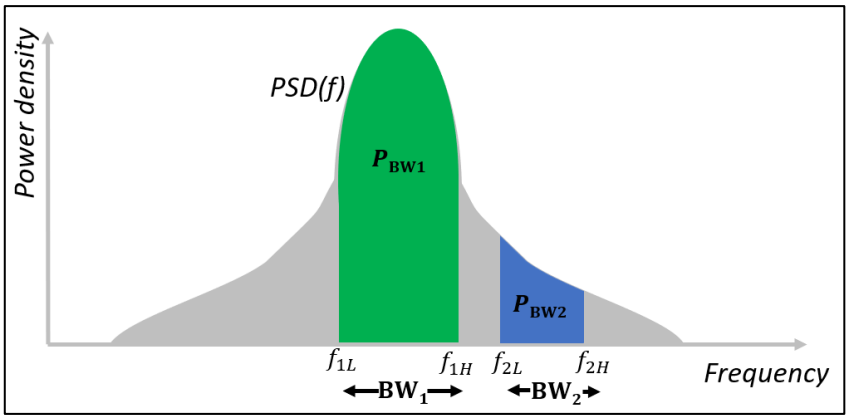



Figure 8: Power spectral density of a transmitted signal. The green area represents the power integrated over the operating bandwidth, while the blue area represents the power integrated over a nearby channel.

Following the notation in Figure 8, the ACLR can be obtained as

$$\text{ACLR} = 10\log\left[\frac{\int_{f_{1L}}^{f_{1H}} PSD(f)df}{\int_{f_{2L}}^{f_{2H}} PSD(f)df}\right] = 10\log\left[\frac{P_{BW1}}{P_{BW2}}\right]. \quad (7)$$

It is important to note that filtering the baseband signal to be transmitted improves ACLR and is necessary to reduce the bandwidth of the modulated signal that has a "sinc" shape in unfiltered format. Also, in Figure 8, the space between the operating channel and the adjacent channel is known as *guard band*.

Finally, spectral regrowth limits known as "spectral masks" are usually set by regulations or by technical standards. A spectral mask is a set of boundaries for the maximum power spectral density that a transmission is allowed to have [17].

### *3.1.4 AM/AM and AM/PM conversion*

A modulated signal can be modelled by

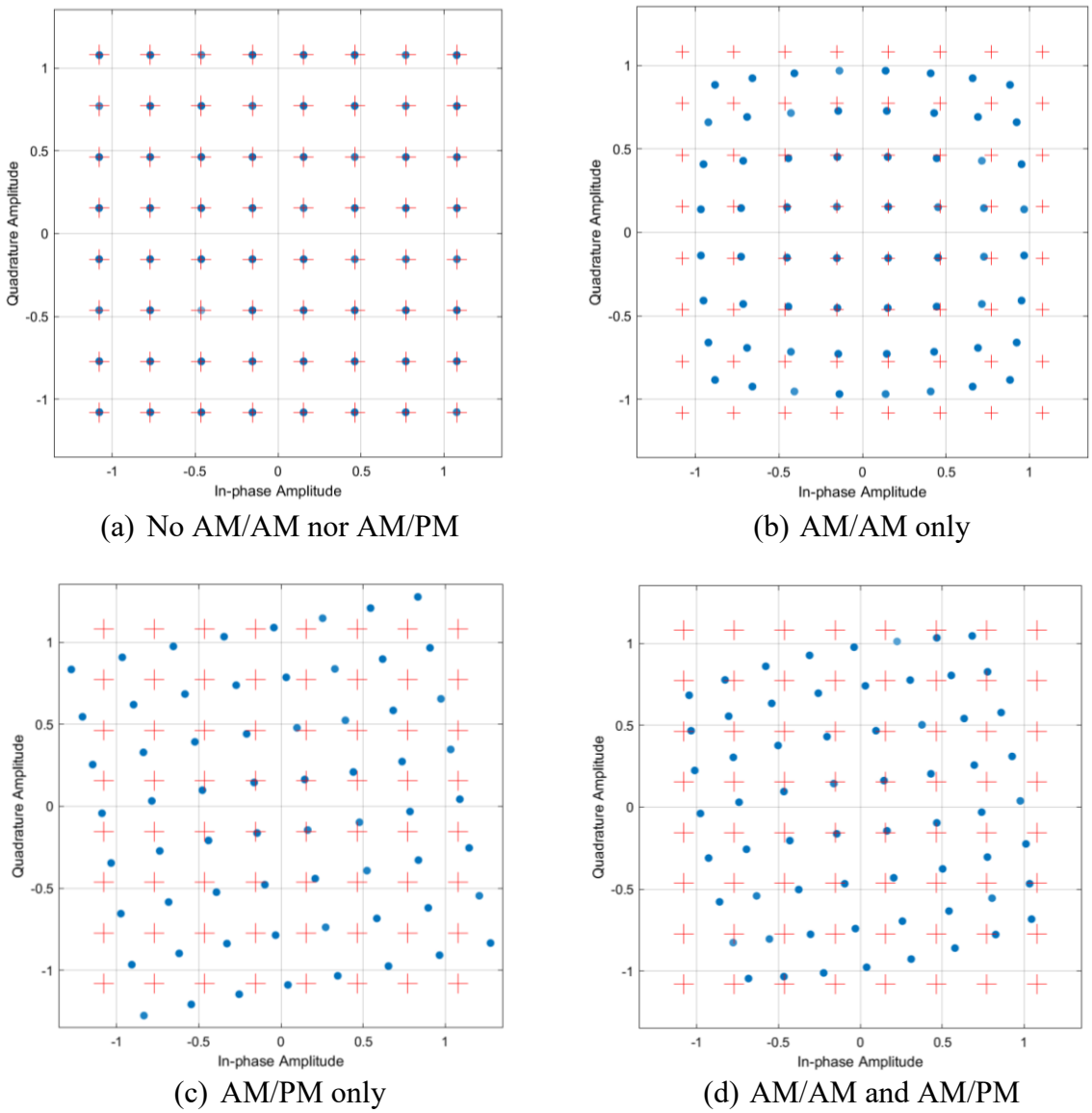
$$x(t) = a(t)cos[\omega t + \varphi(t)], \quad (8)$$

where $a(t)$ represents the amplitude modulation (AM) and $\varphi(t)$ the phase modulation (PM). When the modulated signal is fed into the input of a non-linear system, the output can be approximated by [3]

$$y(t) \approx A[a(t)]\cos\left[\omega t + \varphi(t) + \Phi[a(t)]\right]. \quad (9)$$

That is, the amplitude and phase of the output signal are functions of the amplitude of the input signal. The functions $A[a(t)]$ and $\Phi[a(t)]$ are known as the AM/AM and AM/PM conversion, respectively.

The AM/AM and AM/PM conversion effects on a digital modulation scheme can be observed readily in its constellation diagram. A few examples of degraded constellation diagrams are illustrated in Figure 9.




(a) No AM/AM nor AM/PM (b) AM/AM only

(c) AM/PM only (d) AM/AM and AM/PM

Figure 9: Constellation diagram for a 64-QAM modulation degraded by AM/AM and AM/PM conversion. Reference constellation is represented by red crosses while affected symbols are represented by blue dots.

A typical figure of merit for the degradation of the constellation diagram is the error vector magnitude (EVM). For measuring the EVM, a large sequence of symbols $(I_n, Q_n)$ of a digital modulated signal are fed into the input of the non-linear system and the degraded sequence of symbols at the output of the system $(\hat{I}_n, \hat{Q}_n)$ are recorded. For a specific symbol $(I_i, Q_i)$, an error vector $(\delta I_i, \delta Q_i) = (I_i, Q_i) - (\hat{I}_i, \hat{Q}_i)$ is computed [3]. This computation is illustrated in the Figure 10.

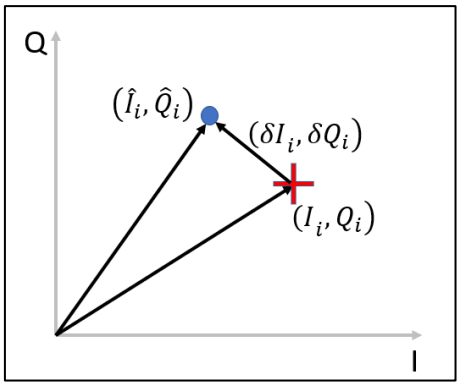



Figure 10: Error vector representation in the I-Q plane of a constellation diagram. The red cross represents an input symbol and the blue dot represents its corresponding output (degraded) symbol.

Then, the EVM is obtaining by applying

$$EVM = \sqrt{\frac{\sum_{i=1}^{N}(\delta I_i{}^2 + \delta Q_i{}^2)}{\sum_{i=1}^{N}(I_i{}^2 + Q_i{}^2)}}, \tag{10}$$

where N is the total number of symbols. The EVM is usually given as a percentage or in decibels (by applying $20log(EVM)$).

### 3.2 Efficiency

The PA is one of the components that consumes the most power in a transceiver [3]. Ideally, all the power consumed by the PA should be converted in the radio waves that carry the transmitted signal. However, in practice, a large fraction of the consumed power is dissipated in the power transistors (in the case of linear PA). Also, a relevant fraction of power is lost in the passive components, in the bias circuit and in digital circuitry that performs intensive computations for the PA (such as those in digital predistortion linearizers).

That is why the efficiency is one of the key indicators for the performance of a PA.

So, efficiency is defined by the ratio between the average power radiated in the assigned channel[2] and the average power that the PA draws from the power supply. In mathematical form [3] [14]

$$\eta = \frac{P_{RF\,out}}{P_{DC}}. \tag{11}$$

However, when the PA has a small gain, its RF input power is a considerable amount of the total dissipated power. That is why an alternative measurement of efficiency is commonly used. This figure of merit is called power added efficiency (PAE), and it is computed as

$$\text{PAE} = \frac{P_{RF\,out} - P_{RF\,in}}{P_{DC}}. \tag{12}$$

[2] The power radiated "outside" the assigned channel is **not** accounted as useful RF power.

That is, the RF input power is subtracted from the RF output power before taking the ratio to the power drawn from the supply. Considering that the RF output power is equivalent to the RF input power times the power gain of the PA ($P_{RF\,out} = G_{PA}P_{RF\,in}$), equation 12 can be expressed in a more revealing form as:

$$\text{PAE} = \frac{P_{RF\,out} - P_{RF\,out}/G_{PA}}{P_{DC}} = \eta\left(1 - \frac{1}{G_{PA}}\right) \tag{13}$$

So, in the limit for very large $G_{PA}$, PAE and η are approximately the same.

### 3.3 Stability

Preventing undesired oscillations is crucial for the correct operation of RF circuits. These oscillations can be produced if the input or the output impedance of the amplifier has a negative real part. Since these impedances depend on the previous and next stages connected to the amplifier as well as the frequency, it is possible that the amplifier is unstable at some frequencies and under some loads [18].

There are a few criteria to test for instability. One of them is plotting the stability circles in the Smith chart and verifying if the input and output reflection coefficients ($\Gamma_{in}$ and $\Gamma_{out}$) of the amplifier are outside these circles. These circles are constructed based on the S-parameters of the amplifier, by determining the areas in that chart in which $|\Gamma_{in}| > 1$ and $|\Gamma_{out}| > 1$ (condition equivalent to input or output impedance with negative real part) [18]. Other tests for instability are the Rollet stability factor and the μ stability factor. These factors are also calculated by using the S-parameters of the amplifier and verifying also when $|\Gamma_{in}| > 1$ and $|\Gamma_{out}| > 1$ [18].

These criteria have mainly two drawbacks. First, they assume a completely linear amplifier, which is usually not the case for RFPAs. This issue can be overcome by checking stability using the S-parameters of the amplifier for different power levels. Secondly, these criteria do not ensure stability for multistage amplifiers [11].

Additionally, bias circuits can generate instability, and it is the most common cause for oscillation in RFPAs [10]. This situation can be mitigated by ensuring good RF isolation between the bias circuit and the PA core.

Finally, differential amplifiers can suffer from common-mode oscillations, since the differential input and output impedances are generally different from the common-mode ones, making it possible to have a negative input resistance in common-mode and not in differential mode [3].

### 3.4 Reliability

Modern integrated circuit technologies aim to achieve the smallest transistors with the highest performance. A common feature of many advanced process nodes is having an exceptionally short channel length together with an extremely thin gate oxide layer. However, this results in significantly high electric fields which compromises transistor expected lifetime for some applications such as power amplifiers [11].

### *3.4.1 Oxide breakdown*

Oxide breakdown happens when the electric field between the gate terminal and the substrate is so large that it gives charge carriers[3] enough energy to create electrical defects in the gate oxide. This is a process that creates a gradual degradation of the transistor capabilities.

At first, these electrical defects produce small conductive paths through the oxide layer, generating a markedly increase of the leakage current. This condition is known as **soft breakdown** and causes significant loss of performance, especially in advanced process nodes.

If the device maintains operation for long time under large electric fields, defects accumulate in the oxide layer, which together with high temperatures, create a large conductive path through the gate oxide. This situation is known as **hard breakdown** and causes the immediate destruction of the transistor [11].

Since high electric field across the gate oxide causes this issue, limiting the gate-to-source and the drain-to-gate voltages is an effective measure for protection. Integrated circuits manufacturers provide maximum figures for safe operation.

### *3.4.2 Hot carrier injection*

Hot carrier injection is also an issue that degrades the transistor performance. It happens when charge carriers in the transistor channel gain high speed due to large drain-to-source voltage. These fast carriers can ionize the substrate producing as a side effect electrical defects in the gate oxide [11].

This type of stress directly degrades the threshold voltage and the transconductance of the device just after a few hours of uninterrupted operation. Although hot carrier injection cannot destroy the affected device by itself, it can accelerate the occurrence of oxide breakdown.

A simple technique to prevent hot carrier injection is avoiding a high drain-to-source voltage and drain current at the same time. However, this might not be possible in all applications, such as in linear power amplifiers.

## 3.5 Principles of linear RF PAs

Traditionally, PAs are classified in two categories: linear amplifiers and switching amplifiers. On one hand, linear amplifiers have sinusoidal-like voltage and current waveforms, good linearity and low efficiency. On the other hand, switching amplifiers have square-like waveforms that minimize the overlap between voltage and current, which highly improves efficiency at the cost of linearity [3] [14].

The linear PA category contains the conventional classes A, B, AB and C. The amount of clipping in the current waveform (controlled by the bias current) is the main difference among these classes. The same linearity/efficiency trade-off is present here: Less clipping results in more linearity but requires larger bias current which increases DC power consumption and thus degrades efficiency.

There are three key concepts to be explained before describing the linear PA classes. These are: knee voltage, load-line matching and harmonic termination.

[3] Electrons for NMOS and holes for PMOS.

### *3.5.1 Knee voltage*

Knee voltage ($V_{knee}$) is the minimum drain-to-source voltage ($V_{ds}$) of the PA output transistor that changes its operation region from saturation to triode [11]. This voltage can be easily observed in the transistor I-V characteristic curve, as illustrated in Figure 11.

For linear operation $V_{ds}$ must be larger than $V_{knee}$, since moving into triode region generates a significant reduction of the drain current ($I_{ds}$).

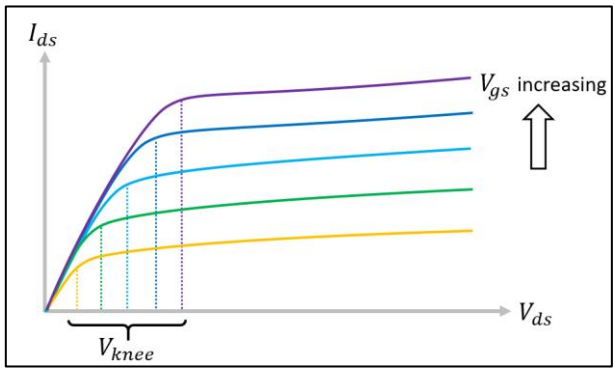



Figure 11: I-V characteristic curve of a FET transistor showing the knee voltage for varying gate-to-source voltages.

### *3.5.2 Load-line matching*

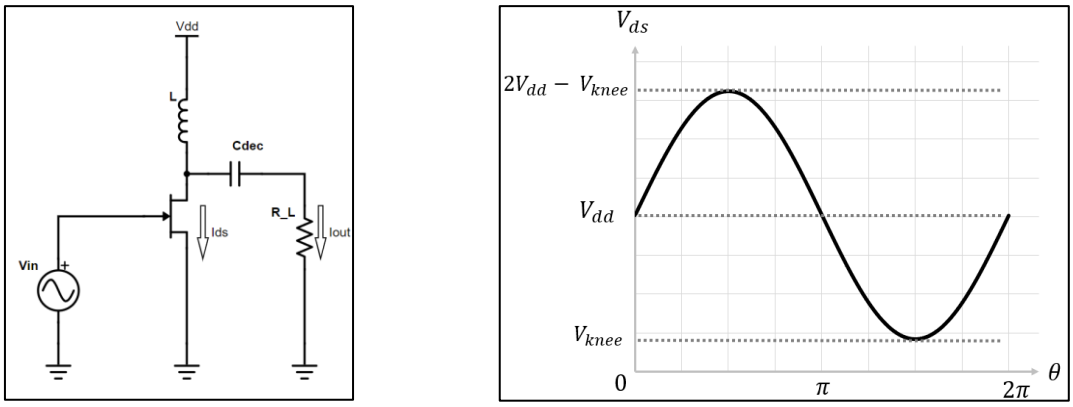
To obtain the most power out of a transistor, its output current $I_{out}$ should be pushed into (or pulled from) a load with a resistance that is as large as possible. The output power is given by [14]

$$P_{out} = \frac{1}{2} {I_{out}}^2 R_L, \tag{14}$$

where $R_L$ is the load resistance. However, a large $R_L$ will create a large output voltage that will shift the transistor operation region from saturation to triode, which is undesirable for linear operation. Therefore, for maximum linear output power the optimum load resistance is given by

$$R_{L\,opt} = \frac{(V_{ds,max} - V_{ds,min})/2}{I_{out}}, \tag{15}$$

where $V_{ds,max}$ and $V_{ds,min}$ are the maximum and minimum drain-to-source voltages in saturation region, respectively. For maximum power, it is desirable to make $V_{ds,max}$ as large as possible, which is achieved by connecting a large inductance between the power supply and the drain terminal, as illustrated in Figure 12. From this figure, $V_{ds,max} = 2V_{dd} - V_{knee}$ and $V_{ds,min} = V_{knee}$.




(a) Simplified schematic. (b) Drain-to-source voltage waveform.

Figure 12: Inductance connected at the drain terminal for maximum output voltage swing.

### *3.5.3 Harmonic termination*

Non-linearities in the transistor operation, such as clipping, saturation or distortion, produce harmonics in the drain current. These current harmonics create harmonics in the drain-to-source voltage when the drain current is pushed into (or pulled from) the load impedance. These voltage harmonics are undesired from a linearity perspective since they can provoke non-linear behaviour by increasing the drain-to-source peak voltage to a point in which it crosses $V_{knee}$, as illustrated in Figure 13.

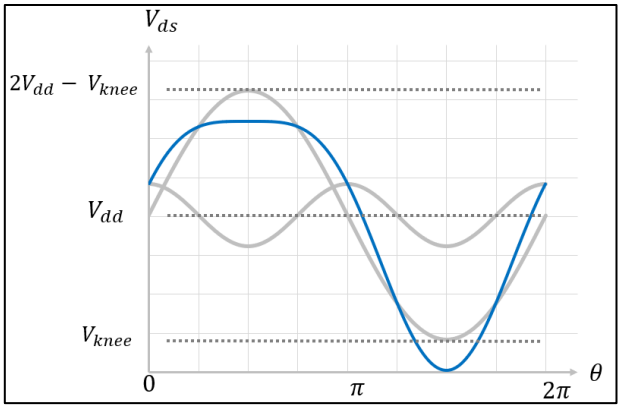
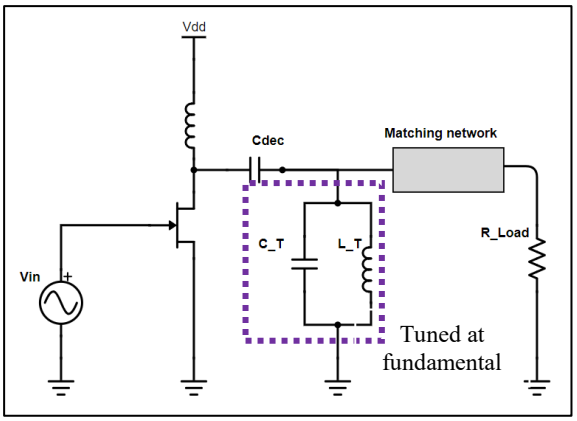


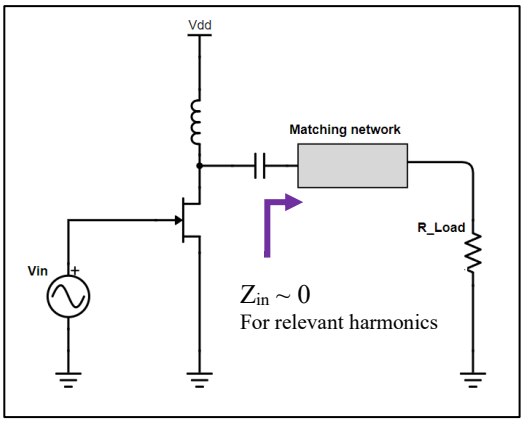

Figure 13: Effect of $V_{ds}$ harmonics on linearity. Fundamental and second harmonic of $V_{ds}$ are represented by grey traces. The blue trace represents the addition of these two waveforms.

One option for solving the issue presented in Figure 13 is reducing the amplitude of the fundamental of $V_{ds}$ until $V_{knee}$ is not crossed anymore. This has the drawback of reducing the output power at the fundamental frequency. Therefore, it makes sense to present the PA with a short circuit for all harmonics to obtain linear operation. This condition is always used when defining the traditional classes of linear PAs [14].

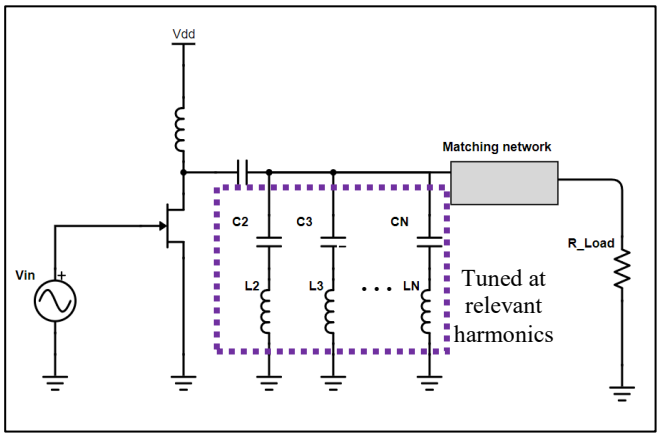
In practice it is enough to present a short circuit at the first few harmonics since they usually carry the most power. Three ways for producing a short circuit at the relevant harmonics are shown in Figure 14.




(a) Parallel resonance at fundamental frequency[4].




(b) Output matching network presents short circuit at relevant harmonics.




(c) Series resonance for each of the relevant harmonics.

Figure 14: Techniques for presenting a short circuit for the relevant harmonics.

In Figure 14(a) a parallel resonance acts as an open circuit at the fundamental frequency and as a short circuit for all harmonics. In Figure 14(b) the output matching network intrinsically presents a very low impedance at harmonic frequencies. Finally, in Figure 14(c) each harmonic is shorted by using a series resonance.

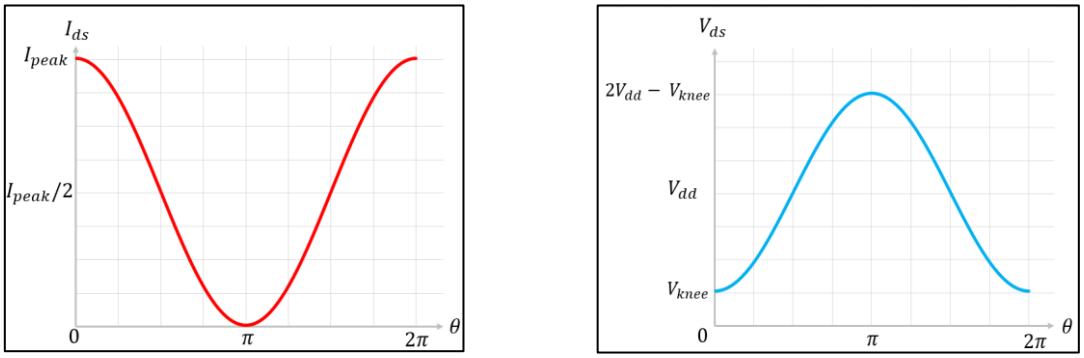
## 3.6 Linear RF PA classes

### *3.6.1 Class-A*

PAs operating in Class-A regime do not exhibit clipping of the drain current waveform. This is achieved by biasing the amplifier with a DC current ($I_{bias}$) equal to half of the peak RF current ($I_{peak}$) [3]. The equations defining the voltage and current waveforms in the transistor terminals in linear regime are

$$I_{ds} = \frac{I_{peak}}{2} cos(\omega t) + \frac{I_{peak}}{2}$$
$$V_{ds} = (V_{dd} - V_{knee}) cos(\omega t + \pi) + V_{dd}. \quad (16)$$

[4] The RF choke can be used as the inductance of the resonator.




(a) Drain current waveform. (b) Drain-to-source voltage waveform.

Figure 15: Class-A drain current and voltage waveforms.

An illustration of those waveforms is presented in Figure 15. With these equations, the output power, optimum load resistance and efficiency can be computed in a straightforward manner.

Now, the output power delivered to the load at the fundamental frequency is determined by

$$P_{L,1} = \frac{1}{2} V_{L,1} I_{L,1}\,, \tag{17}$$

where $V_{L,1}$ and $I_{L,1}$ are the fundamental components of the voltage and current delivered to the load, respectively. For determining $I_{L,1}$, two things should be considered: First, the large inductance connected at the drain terminal prevents the high frequency current from flowing into the power supply causing the load current to be the flipped polarity version of the drain current. Second, the decoupling capacitor between the drain terminal and the load prevents DC current from flowing into the load. Therefore, the fundamental component of the load current is

$$\mathrm{I}_{L,1} = \; -\mathrm{I}_{bias}\cos(\omega t). \tag{18}$$

Similarly, $V_{L,1}$ is given by

$$V_{L,1} = (V_{dd} - V_{knee})\cos(\omega t + \pi). \tag{19}$$

Then, the power delivered to the load at the fundamental frequency is

$$P_{L,1} = \frac{\mathrm{I}_{bias}(V_{dd} - V_{knee})}{2}. \tag{20}$$

Moreover, the optimum load resistance can be calculated according to Equation 15

$$R_{L\,opt} = \frac{V_{dd} - V_{knee}}{I_{bias}}. \tag{21}$$

Finally, the efficiency is determined by

$$\eta = \frac{P_{L,1}}{P_{DC}} = \frac{\mathrm{I}_{bias}(V_{dd} - V_{knee})}{2\,\mathrm{I}_{bias}V_{dd}} = \frac{1}{2} - \frac{V_{knee}}{2\,V_{dd}}, \quad (22)$$

which is 50% if $V_{knee} << V_{dd}$. However, in modern process nodes, $V_{knee}$ is an important fraction of $V_{dd}$ and therefore it should not be ignored [11].

### *3.6.2 Classes AB, B and C*

Classes AB, B and C exhibit clipping in the drain current as opposed to Class-A. These classes are distinguished from one another by the fraction of the time period that the current waveform is clipped, as presented in Table 2.

Table 2: Linear PA classes categorized by drain current clipping

| PA Class | Fraction of the time period in which $I_{ds}$ is clipped | Conduction angle |
|---|---|---|
| Class-A | 0% | 360° |
| Class-AB | Between 0% and 50% | Between 360° and 180° |
| Class-B | 50% | 180° |
| Class-C | < 50% | < 180° |

The drain current equations for classes AB, B and C can be presented concisely as shown below [14]

$$I_{ds} = \begin{cases} (I_{peak} - I_0)\,cos(\omega t) + I_0, & -\alpha/2 < \omega t < \alpha/2 \\ 0, & \text{otherwise} \end{cases}, \quad (23)$$

where $I_0$ is a parameter that controls the amount of clipping[5], $I_{peak}$ is the maximum current provided by the transistor and $\alpha$ is the angle $\omega t$ during which $I_{ds}$ is positive (known as the conduction angle). It can be observed that Class-A is a special case of Equation 23 when $I_0 = I_{\mathrm{peak}}/2$ and $\alpha = 360°$. Examples of the $I_{ds}$ waveform for the linear PA classes are shown in Figure 16.

Since it is assumed that the PA is presented with a load that behaves as a short circuit for all harmonics, the drain voltage waveform is that defined in equation 19 and illustrated in Figure 15 (b).

Now, to achieve maximum output power at the fundamental frequency, the optimum load resistance needs to allow maximum drain voltage swing ($i.e. V_{dd} - V_{knee}$). For that to happen, the optimum load resistance must be

$$R_{L\,opt} = \frac{V_{dd} - V_{knee}}{I_{ds,1}}, \quad (24)$$

where $I_{ds,1}$ is the amplitude of the first harmonic of $I_{ds}$. Moreover, the efficiency is given by

$$\eta = \frac{P_1}{P_{DC}} = \frac{(V_{dd} - V_{knee})I_{ds,1}}{V_{dd}I_{DC}}, \quad (25)$$

[5] Parameter $I_0$ corresponds to the drain bias current.

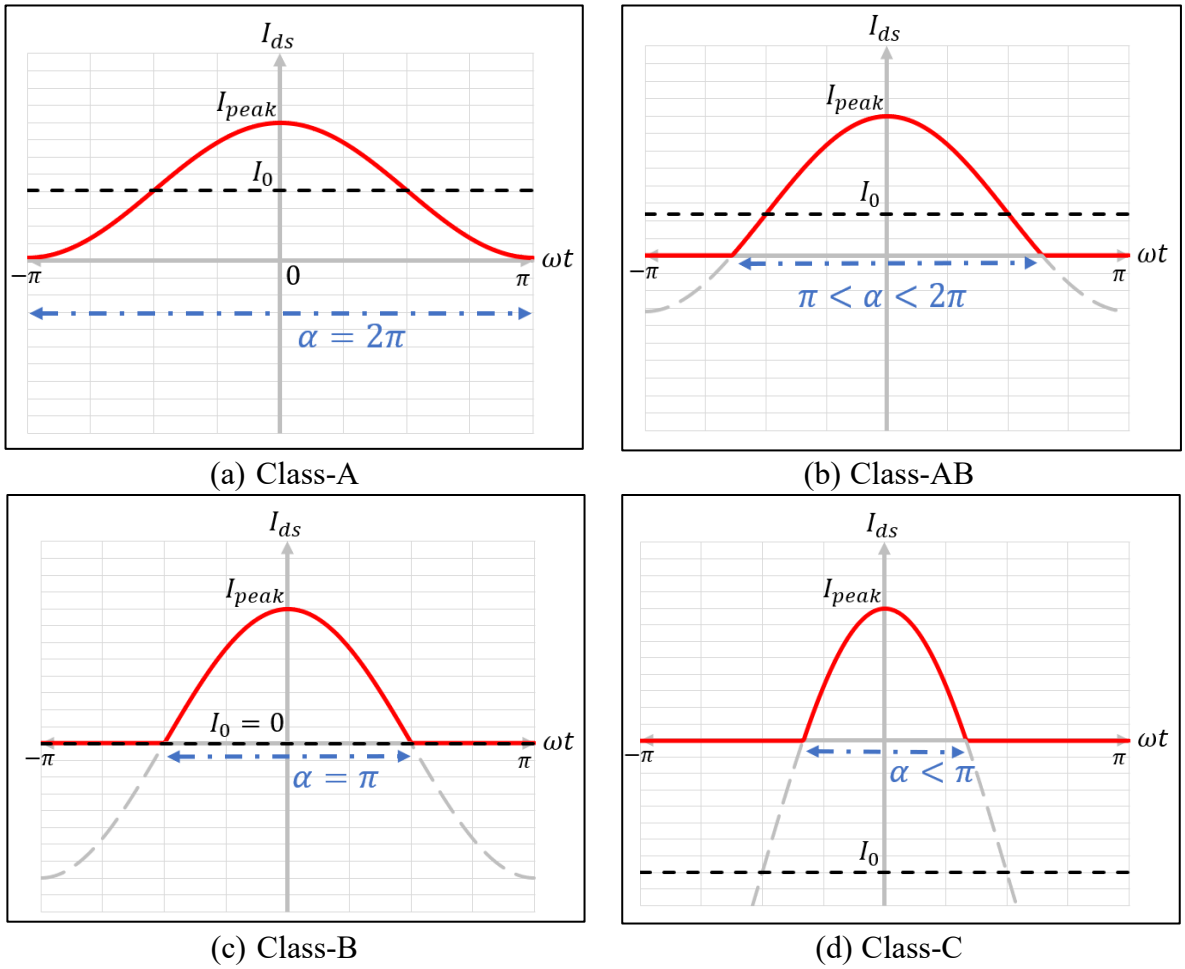



Figure 16: Classes A, AB, B and C drain current waveforms.

where $I_{DC}$ is the DC component of the drain current. The expressions for $I_{DC}$ and $I_{ds,1}$ are [14] [11]

$$I_{DC} = \frac{I_{peak}}{2\pi} \frac{2\,sin(\alpha/2) - \alpha\,cos(\alpha/2)}{1 - cos(\alpha/2)} \tag{26}$$

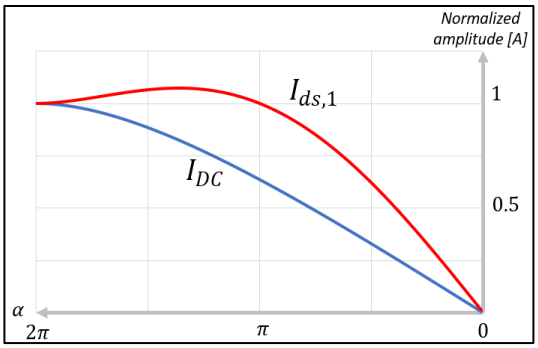
$$I_{ds,1} = \frac{I_{peak}}{2\pi} \frac{\alpha - sin(\alpha)}{1 - cos(\alpha/2)}. \tag{27}$$

With these equations, the output power, optimum load resistance and efficiency can be computed in a straightforward manner. A visual representation for the variation of these parameters with respect to the conduction angle is illustrated in Figure 17.

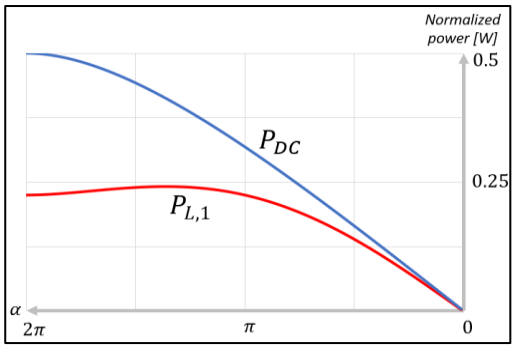
From this figure, a few important facts can be highlighted:

- For class-AB the fundamental component of the drain current is maximized.
- Class-A and class-B amplifiers provide the same fundamental current amplitude. However, the DC current for class-B is lower and therefore it provides higher efficiency.

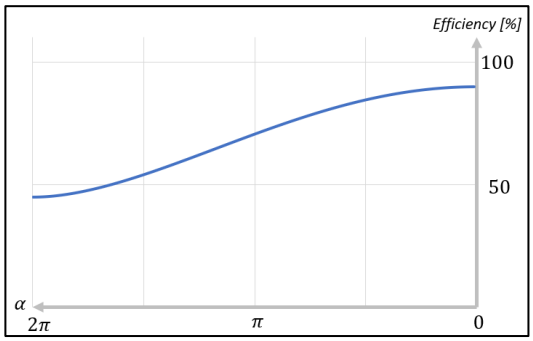
- The knee voltage reduces efficiency. This can be observed in the class-A and class-C cases in which efficiency disregarding $V_{knee}$ would be 50% and 100% respectively. In the presented plots, efficiency accounting for $V_{knee}$ is 45% and 90% respectively.



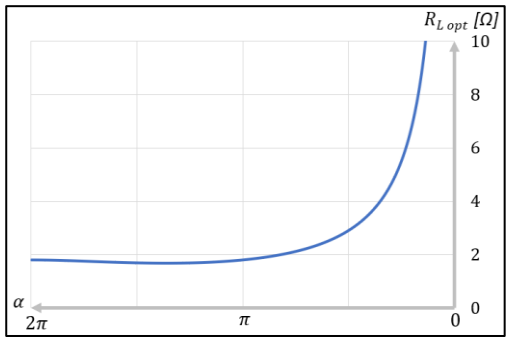

(a) First harmonic and DC component of the drain current




(b) First harmonic output power and DC power




(c) Power efficiency




(d) Optimum load impedance

Figure 17: Dependence of the key parameters for linear PAs with respect to the conduction angle. The power and amplitude were normalized by using $V_{dd} = 1\ V$, $V_{knee} = 0.1\ V$ *and* $I_{peak} = 1\ A$.

Finally, the linearity for the linear PA classes can be assessed by observing the amplitude of the harmonics they produce. Figure 18 shows the variation of the first 5 harmonics with respect to the conduction angle, which was obtained by utilizing Fourier analysis on $I_{ds}$ [14]. From this plot, the following observations can be made:

- The amplitude of all harmonics is zero for class-A, indicating maximum linearity.
- For Class-AB the amplitude of the second harmonic is increased by reducing conduction angle, resulting in lower linearity than class-A.
- Class-B second harmonic amplitude is half of the fundamental amplitude, while higher harmonics are negligible.
- For class-C has the worst linearity since the amplitude of the harmonics is large compared to the amplitude of the fundamental. Therefore, class-C exhibit the worse linearity among the linear PAs.

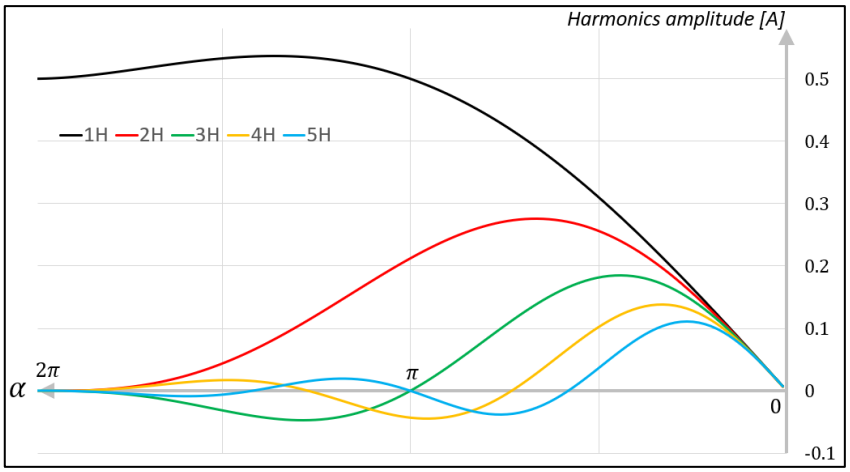



Figure 18: Amplitude of the drain current fundamental component and its harmonics with respect to the conduction angle.

It can be concluded that there is a trade-off between linearity vs efficiency that must be considered when designing a linear PA. It is also important to note that the knee voltage also plays a key role in this trade-off: it reduces efficiency since it limits the maximum output power that can be achieved without compromising linearity (maintaining the output transistors in saturation region).

Another important conclusion is that the results presented in this section apply only if the PA is presented with a short circuit for all harmonics. However, this is a condition that is challenging to achieve in practice. Therefore, the drain voltage waveform, as well as the output power, efficiency, and linearity are drastically different in practice which alters the performance of the PA.

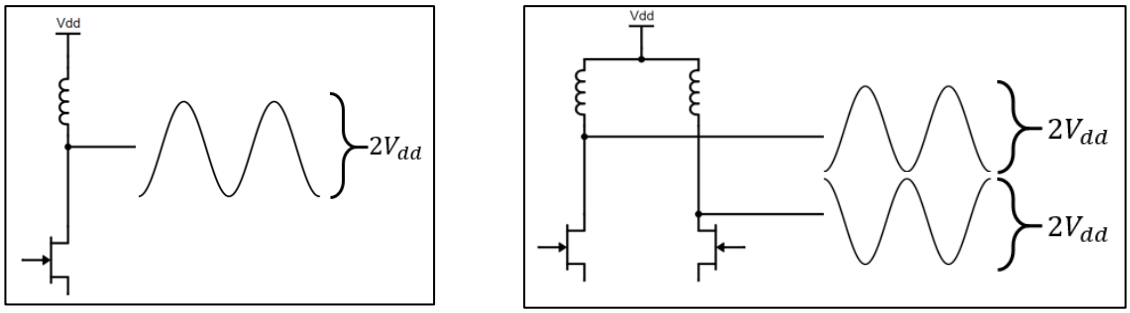
## 3.7 Circuit architectures

### *3.7.1 Single-ended and differential PAs*

Most of the PAs in the literature utilize a differential structure. However, it is important to understand the trade-offs between employing a single-ended PA and a differential one.

First, a differential PA has a larger voltage swing for a given power supply voltage $V_{dd}$. The maximum output voltage that can be achieved by a single-ended design is $2V_{dd}$ by connecting an inductance between the power supply and the drain terminal [3]. Though, a differential design has an effective output voltage twice as large as the single-ended case, as shown in Figure 19.

A larger output voltage results in a larger $R_{L\,opt}$ (according to equation 15) which facilitates the implementation of the output matching network by either improving its bandwidth, reducing the number of lumped elements or reducing its power losses [14]. A larger output voltage also reduces the output current required for delivering a given power to the load. This has various benefits such as minimizing electromigration issues, decreasing self-heating, reducing the size of the power transistors and diminishing losses in the passive elements connected at the PA output [11]. Additionally, smaller output current reduces the effect of the package parasitic inductance and mitigates possible feedback and noise to previous stages through the power supply and ground lines [3].



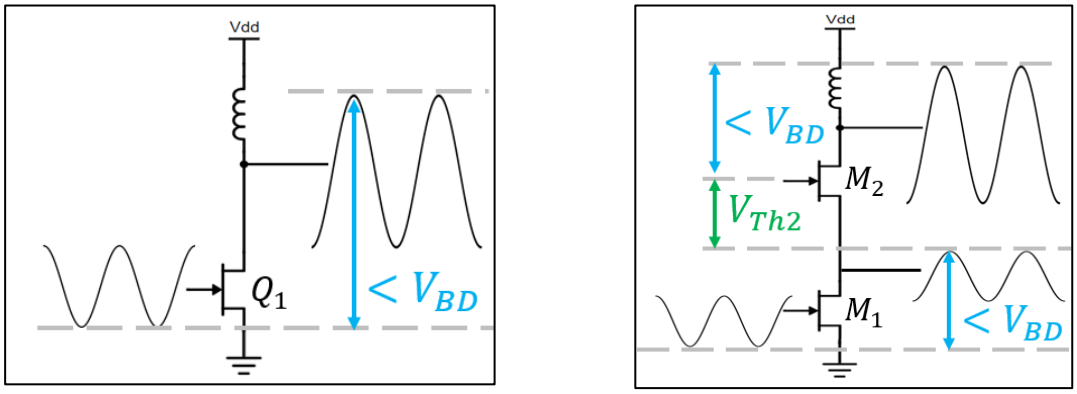

(a) Single-ended (b) Differential

Figure 19: Maximum output voltage for single-ended and differential PA.

A drawback of differential designs is the need of a balun for driving single-ended antennas, which are most common in practice. Integrated baluns (and in general integrated inductors) are significantly large structures occupying a large chip area and introducing considerable losses [3]. Additionally, differential PAs require additional I/O pins which in some cases can result in larger chip area.

### *3.7.2 Single device vs stacked devices*

Most of the PA design in the literature employ a cascode topology. The most important reason for using this topology in integrated RFPAs is the capability of handling larger voltages while mitigating reliability issues [3] [11]. As shown in Figure 20, a cascode structure can safely sustain a larger output voltage than a single device. As explained in sub-section 3.7.1, a larger output voltage has many benefits.




(a) Single device case. (b) Cascode case.

Figure 20: Maximum output voltages for ensured reliability. $V_{BD}$ is the oxide breakdown voltage.

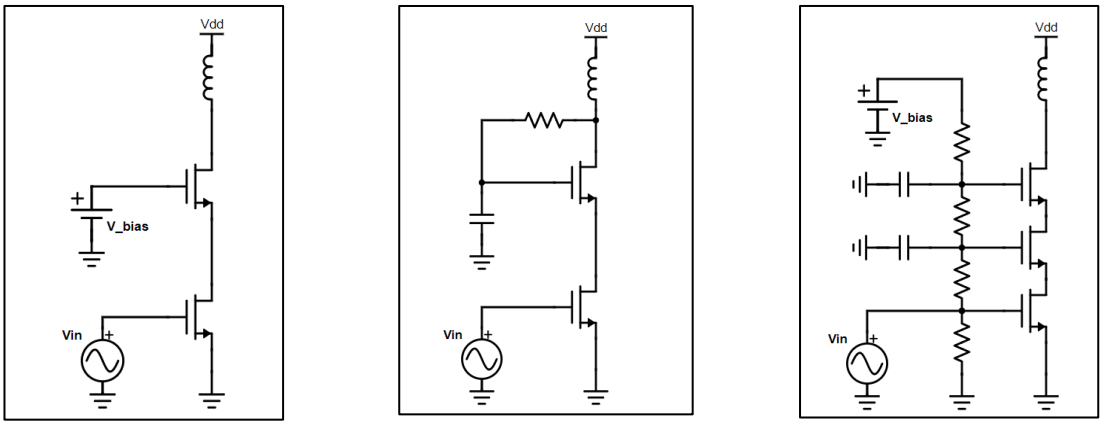
Another advantage of the cascode topology is its high reverse isolation, which mitigates unwanted feedback, improving stability [3].

Nevertheless, this topology comes with a few drawbacks. First, the knee voltage is larger than in the single device case because the knee voltage of the common-gate (CG) transistor is added to the one of the common-source (CS) transistor. A consequence of having increased knee voltage is a reduced voltage headroom at the drain of the common-gate device to keep the devices in linear operating region. Obviously, a larger drain current will be required to achieve

a given output power, which increases the ohmic losses in the passive components present in the output matching network [3].

Another disadvantage of using the cascode topology is that the total silicon area is increased compared to having a single device. In the case of a single device, the transistor has to be wide enough to produce the required transconductance. However, for the cascode configuration, the CS transistor require the same width as in the single device case and additionally the CG device needs to be as wide as possible in order to maximize its drain voltage headroom before entering into triode region.

A few variations of the cascode topology can be found in the literature [11]. The schematics for these variations are presented in Figure 21.




(a) Conventional (b) Self-biased (c) Stacked

Figure 21: Variations of the cascode topology.

In the conventional cascode, the CG device suffer most of the RF stress. In the self-biased cascode, the gate voltage of the CG transistor follows the variations of the drain voltage, which has the effect of balancing the RF stress between the CS and the CG devices. However, this has the effect of reducing its gain when compared to the conventional cascode. Finally, the stacked configuration can tolerate great RF stress at the cost of a more complex bias network and larger knee voltage [11].

## 3.8 Passive matching network topologies

Besides impedance conversion, an output matching network in a PA needs to:

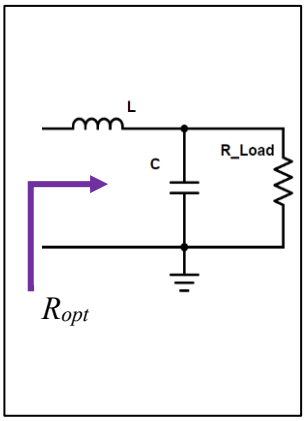
- Provide DC bias to the power transistors.
- Prevent RF power from flowing into the power supply.
- Provide proper harmonic termination for improved linearity or efficiency.
- Filter the signal so no power at high harmonics is consumed by the load.
- Prevent DC power from flowing into the load (DC decoupling).
- Convert differential to single-ended output.

These are challenging requirements that have to be attained while minimizing power losses, maximizing bandwidth and utilizing minimum chip area (or minimum number of components in the case of off-chip networks). Although less complex in nature, inter-stage matching networks have also difficult requirements.

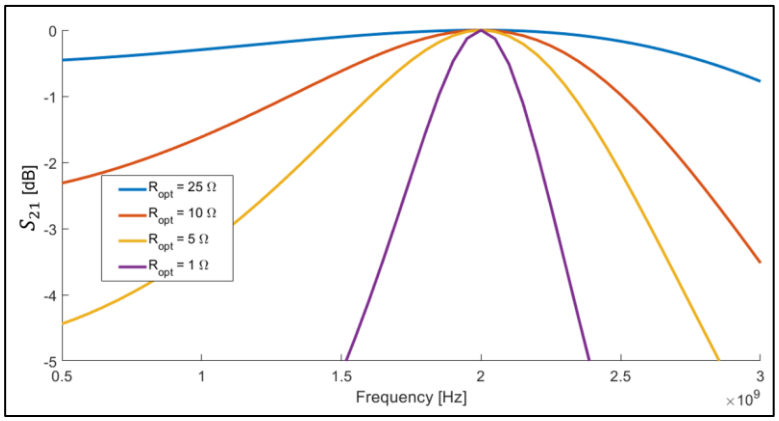
The core role of a matching network (impedance conversion) can be performed at low frequencies by sections of lumped components or by a transformer. These options are explored next.

### *3.8.1 Lumped components matching network*

By using inductors and capacitors, it is possible to perform impedance conversion between the load resistance $R_L$ and the optimum load impedance $R_{L,opt}$ with low losses. Since harmonic filtering is required, a low pass LC matching network is of most interest. This matching network, illustrated in Figure 22(a), can be easily designed by using circuit theory or the Smith chart.




(a) Schematic.

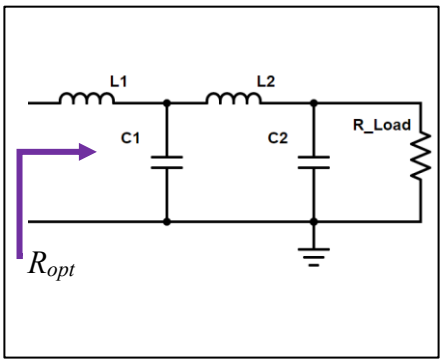



(b) $S_{21}$ for different target resistances. Load is 50 Ω.

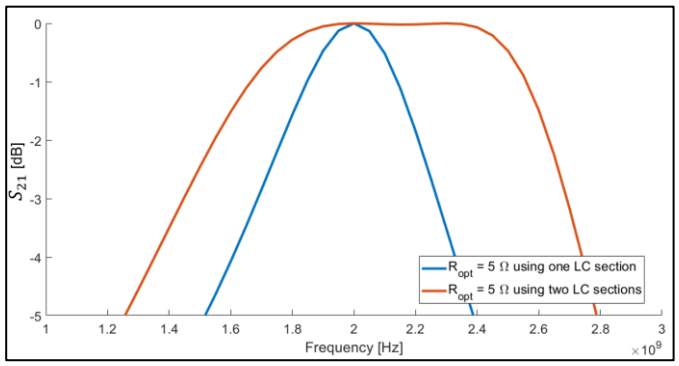
Figure 22: Schematic and loss of a single LC section matching network.

A drawback of this topology is that its bandwidth depends strongly on the impedance transformation ratio $R_L/R_{L,opt}$. Figure 22(b) illustrates this issue for the case in which $R_L = 50\ \Omega$. Unfortunately, $R_L/R_{L,opt}$ is large for integrated RF PAs with low supply voltage and relative high output power, resulting in a poor impedance matching bandwidth [11].

This problem can be solved by using two LC sections in cascade, as shown in Figure 23. The obvious drawback is additional components which increases the cost and the ohmic losses.




(a) Schematic.




(b) $S_{21}$ for one and two LC-sections.

Figure 23: Schematic of a two LC section matching network and loss comparison between one and two LC sections.

The designing of a two LC-sections matching network can be decomposed in the design of one LC-section that matches $R_L$ to $\sqrt{R_L R_{L,opt}}$ followed by another one that matches $\sqrt{R_L R_{L,opt}}$ to $R_{L,opt}$ [14]. However, improved bandwidth can be achieved by using computer-aided optimization, as shown in Figure 24.

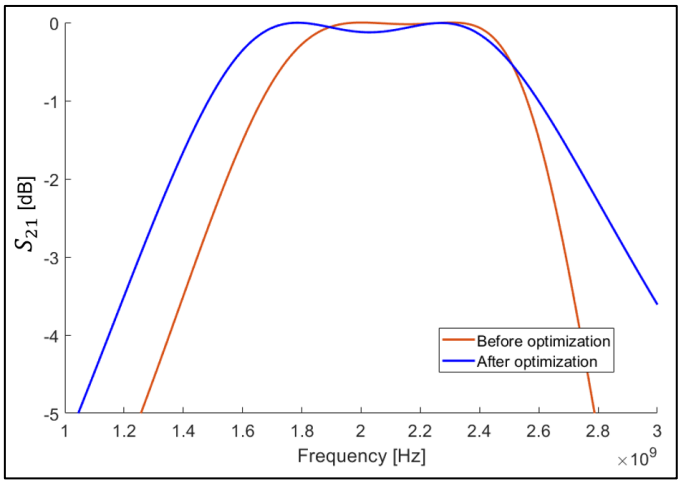



Figure 24: $S_{21}$ of a two LC-section matching network converting 50 Ω to 1 Ω before (orange trace) and after (blue trace) computer-aided optimization.

### *3.8.2 Integrated transformers*

A transformer is a device that couples AC currents between two wire windings by means of their mutual inductance. This device offers efficient power transfer between its two windings while converting their impedance levels. Additionally, a transformer provides DC isolation, allowing biasing both windings at different voltages [19].

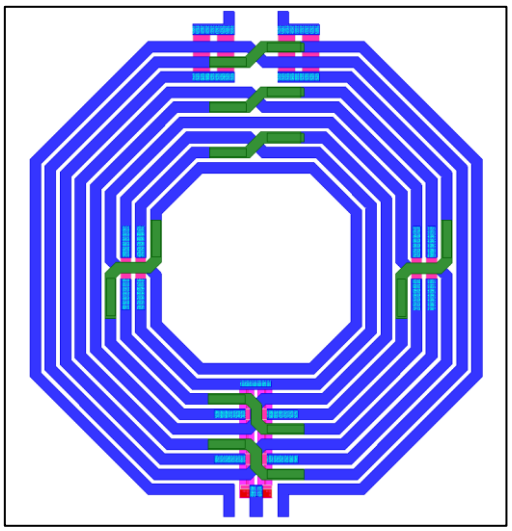


Figure 25: Layout diagram of a 6:3 integrated transformer.

Integrated transformers are built from interwound conducting wires as shown in Figure 25. Their most relevant parameters are the turns ratio $n$ and the coefficient of magnetic coupling $k$. Additionally, the self-inductance of the windings satisfies the relation $n = \sqrt{L_2/L_1}$ while their mutual inductance satisfies $\mathrm{M} = \mathrm{k}\sqrt{L_2 L_1}$.

Integrated transformers suffer from the same loss mechanisms than integrated inductors, namely high metal resistance and magnetic/capacitive coupling to the resistive substrate. These mechanisms can be mitigated by using wide traces of thick metal layers for constructing the transformer windings, as well as using a ground shield placed between the transformer and the substrate [3].

Regarding electrical modelling, lumped-component approximations are widely used for transformers with physical lengths much less than the signal wavelength. These models offer good accuracy for frequencies of a few gigahertz, as well as enough simplicity for hand-calculations and computer-aided design optimization. One of these models is shown in Figure 26 [19].

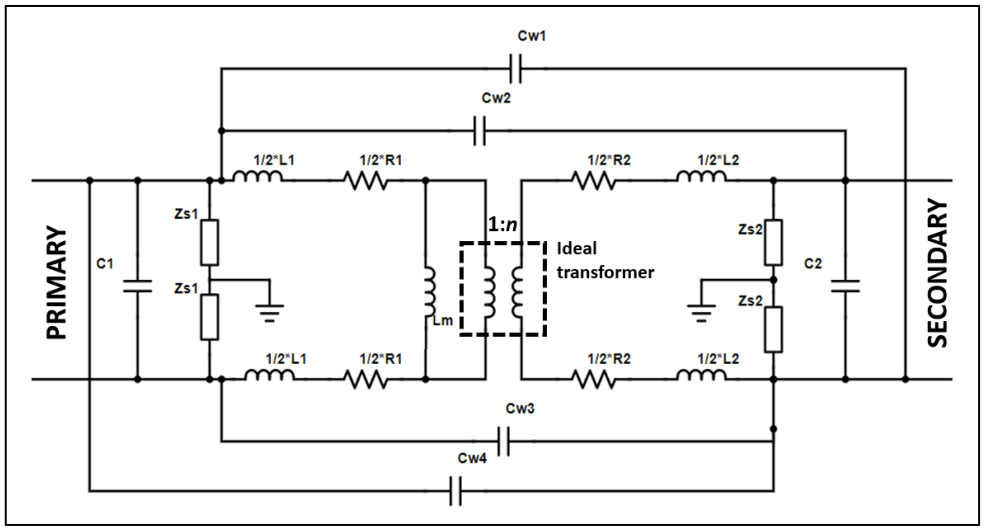



Figure 26: Lumped-component model for an integrated transformer.

This model is centered around an ideal transformer with turns ratio $n$ and magnetizing inductance $L_m$. The self-inductance of the windings is represented by $L_1$ and $L_2$. Also, $R_1$ and $R_2$ model the resistance of the conductors. Elements $C_w$ model the parasitic capacitance between primary and secondary, while $C_1$ and $C_2$ account for the total capacitance seen by each individual winding. Finally, impedances $Z_s$ represent the capacitive coupling between the windings and substrate as well as its associated ohmic losses [19].

An appropriate transformer design achieves the required impedance conversion with minimum power loss and large bandwidth. However, in practice the bandwidth is traded-off for the power loss by means of tuning the transformer using additional capacitances connected at the primary and secondary sides [19] [20]. These capacitances resonate out the equivalent inductances seen at each side, minimizing mismatch losses.

# 4 DESIGN SPECIFICATIONS

This RF PA must comply with the transmitter specifications established in the standard 3GPP TS 36.101 [21] for devices in the category NB1 and NB2 (Narrowband Internet of Things) within the Power Class 6. Also, there are additional specifications for this power amplifier established by the team that designed the other transmitter blocks.

## 4.1 Overall design description

The RF PA designed in this work is meant to be part of a transmitter that has already been designed. With the addition of this PA the transmitter would be able to operate as a device in the Power Class 6 of the 3GPP TS 36.101 standard. A simplified block diagram of the transmitter chain is shown in Figure 27.

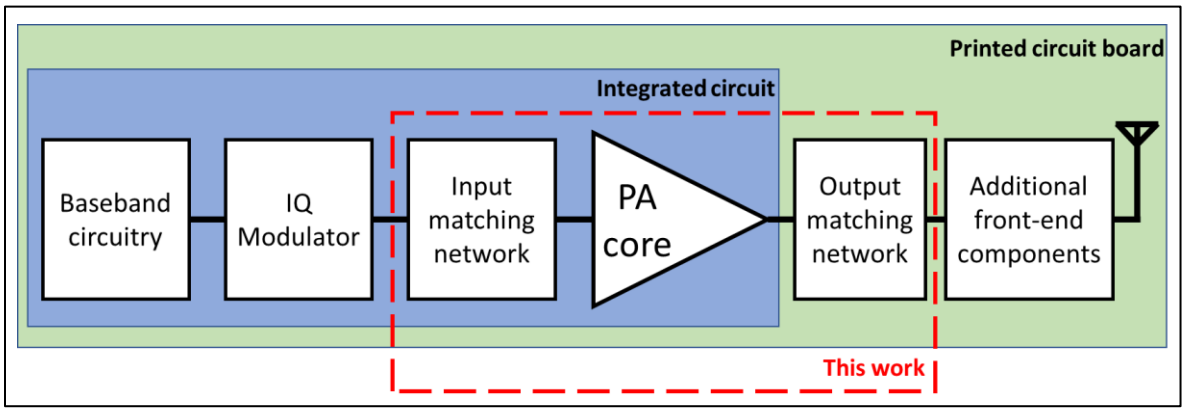



Figure 27: Block diagram of the transmitter.

Moreover, the main goal of this work is to determine if it is possible to implement this RF PA by avoiding the use of integrated inductors and external baluns in order to minimize the consumed silicon and PCB area. This means that the design must be single-ended (to avoid the output balun), and it has to be composed of only one stage (preventing the use of the integrated inductors needed in the matching networks between PA stages).

## 4.2 Information regarding the IQ modulator

The output characterization of the IQ modulator is fundamental the design of the PA. The output stage of this modulator is a transconductance amplifier and therefore it generates current. At the average power of the modulation, the modulator output current amplitude is -59 dBA. At the peak power of the modulation, this amplitude increases up to -53 dBA. However, the modulator has programmable gain and hence, these values correspond to the maximum gain setting. The gain dynamic range of the modulator is 48 dB.

Additionally, if the load impedance of the modulator is larger than 315 Ω, its output stage will saturate, generating undesired signal distortion.

Furthermore, the output signal of the modulator contains residual power at the harmonic frequencies. This residual power depends on several characteristics of the output signal, such as bandwidth, type of modulation, power, among others. For this PA design, the harmonic content of the signal produced by the modulator is assumed to be given in Table 3. The power levels in this table are given with respect to the power at the fundamental.

Table 3: Harmonic content of the modulator output signal

| Harmonic | Power [dBc] |
|---|---|
| Second | -40 |
| Third | -11 |
| Fourth | -28 |
| Fifth | -18 |

### 4.3 Modulation and use of the frequency spectrum

According to the 3GPP standard, the modulated signal can be formed by 1, 3, 6 or 12 tones. For a single tone, the modulation can be QPSK or BPSK and the tone bandwidth can be 3.75 kHz or 15 kHz. For the multi-tone case, each tone has QPSK modulation and 15 kHz of bandwidth.

Moreover, the operating frequency bands for the PA are 1695 MHz – 1758 MHz, 1850 MHz – 1882 MHz and 1920 MHz – 1980 MHz.

### 4.4 Output power

The 3GPP standard establishes a maximum output power of +14 dBm. This value corresponds to the average power measured at the antenna port. Additional 2 dB must be added to compensate for the losses in the circuit board and in the antenna switch. Therefore, the PA must be able to generate +16 dBm of average power.

To determine the peak power that the power amplifier must handle, two parameters should be considered: the maximum power reduction (MPR) and the peak-to-average power ratio (PAPR). The first one is a reduction in the maximum transmitted power allowed by the 3GPP standard in order to meet the adjacent channel leakage requirements [22]. The second one determines the power of the peaks of the modulated signal with respect to its average power [11].

These two parameters vary with the number of tones in the modulation as seen in Table 4.

Table 4: PAPR and MPR for different modulation schemes

| Number of tones | Bandwidth of tone [kHz] | Estimated PAPR [dB] | MPR [dB] |
|---|---|---|---|
| 1 | 3.75 | 1.5 | Not applicable |
| 1 | 15 | 2.1 | Not applicable |
| 3 | 15 | 4 | 0.5 |
| 6 | 15 | 4.9 | 1 |
| 12 | 15 | 6.2 | 1.5 |

Then, the maximum peak power for each modulation scheme is given by adding the maximum average power and PAPR and then subtracting the MPR. In the worst case (corresponding to the 12-tone modulation profile), the PA would need to amplify peaks up to +20.7 dBm.

The 3GPP standard also defines a minimum output power of -40 dBm. This means that the transmitter should be able to transmit at any power between the maximum (+16 dBm) and this minimum, corresponding to a dynamic range of 56 dB. As mentioned previously, the modulator has a dynamic range of 48 dB, leaving a 12 dB range for the PA. Therefore, the PA gain must be controllable over a range of 12 dB.

## 4.5 Linearity

### *4.5.1 EVM*

The EVM of the whole transmitter chain shall not exceed 17.5% according to the 3GPP standard. However, the requirements by the transmitter design team are stricter, corresponding to an EVM of 3.4%.

### *4.5.2 Emission limits*

Additionally, the PA must comply with the limits of emissions out of the operating channel according to the 3GPP standard. There are three types of limits: Spectrum emission masks, adjacent channel leakage ratios (ACLR) and spurious emissions limits.

Regarding the spectrum masks, there are two of them established in the standard. These masks are defined using maximum power spectral densities measured at different frequency offsets. These masks are plotted in Figure 28 and Figure 29. It is important to note that the mask 1 is stricter than the mask 2.

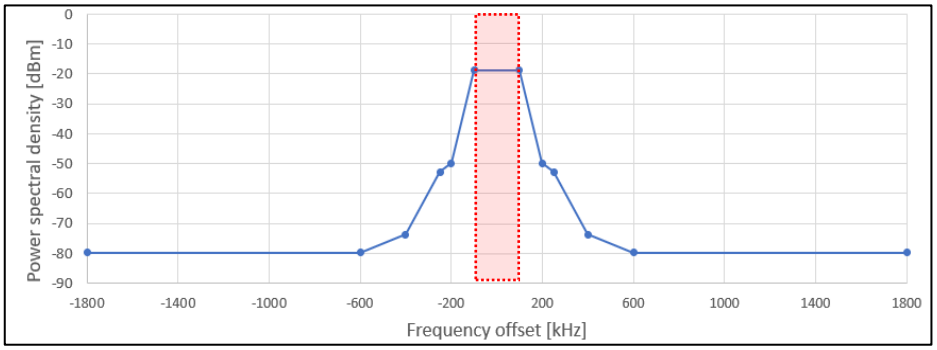



Figure 28: Spectrum mask 1 (blue) and channel bandwidth (red).

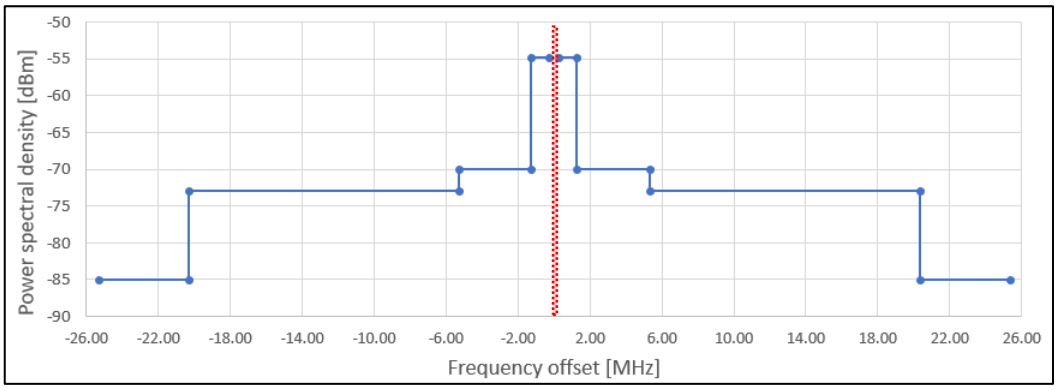



Figure 29: Spectrum mask 2 (blue) and channel bandwidth (red).

With respect to ACLR, the standard defines two limits which are summarized in Table 5.

Table 5: 3GPP standard ACLR limits

| | **ACLR1** | **ACLR2** |
|---|---|---|
| **ACLR** | 20 dB | 37 dB |
| **Adjacent channel center frequency offset** | ±200 kHz | ±2.5 MHz |
| **Adjacent channel measurement bandwidth** | 180 kHz | 3.84 MHz |

However, the transmitter design team agreed on an additional margin of 10 dB for the ACLR.

With respect to the spurious emission, the limits established in the 3GPP standard are summarized in Table 6. These limits only apply for frequencies that are more than 1.7 MHz away from the edge of the operating channel (the spectrum masks apply in that range).

Table 6: Spurious emission limits

| **Frequency Range** | **Maximum Level [dBm]** | **Measurement bandwidth** |
|---|---|---|
| 9 kHz ≤ f < 150 kHz | -36 | 1 kHz |
| 150 kHz ≤ f < 30 MHz | -36 | 10 kHz |
| 30 MHz ≤ f < 1000 MHz | -36 | 100 kHz |
| 1 GHz ≤ f < 12.75 GHz | -30 | 1 MHz |
| 12.75 GHz < f < 26 GHz | -30 | 1 MHz |

The transmitter design team uses an additional margin of 10 dB for the spurious emission limits. Combining the spurious emission limits with this 10 dB margin and with the harmonic content of the modulator output, the attenuation (filtering) that the PA must provide at the harmonic frequencies is presented in Table 7.

Table 7: PA required attenuation for compliance with spurious emission limits

| **Harmonic** | **Attenuation [dB]** |
|---|---|
| Second | 14 |
| Third | 43 |
| Fourth | 26 |
| Fifth | 36 |

### *4.5.3 Transmit intermodulation*

The standard defines transmit intermodulation as the intermodulation generated between the signal transmitted by the device under test and a foreign signal that enters this device through its transmitter antenna. The standard also defines limits for the transmit intermodulation products for two scenarios.

In the first scenario, the signal transmitted by the device under test is formed by only one tone of 15 kHz of bandwidth. The frequency and power difference between this signal and the foreign signal are 180 kHz and -40 dB, respectively. Then the intermodulation product must be lower than -20 dBc when measured at the antenna port of the device under test.

In the second scenario, the frequency difference is 360 kHz and the intermodulation product limit is -39 dBc.

## 4.6 Ruggedness

The PA must tolerate changes of antenna impedance that produce a VSWR up to 8:1. This means that the internal components of the PA have to endure the large voltages and currents generated in antenna mismatch conditions.

Moreover, the stability of the PA must be guaranteed for antennas impedances corresponding to a VSWR of 5:1.

# 5 RF POWER AMPLIFIER DESIGN

## 5.1 Power amplifier core

### *5.1.1 Initial remarks about the available process node*

The first step in any electronics design is understanding the available process node. For this design, a 28 nm node is used. This process has "core" transistors (used for digital applications), I/O transistors (used for analog applications) and LDMOS transistors. A few parameters of these transistor types are compiled in Table 8.

Table 8: Characteristics of active components in the available process node

| Device type | Oxide thickness type | Drain-to-source breakdown voltage [V] | Gate-to-source breakdown voltage [V] | Drain-to-gate breakdown voltage [V] | Minimum length [nm] |
|---|---|---|---|---|---|
| Core | Thin | 2.7 | 0.9 | 0.9 | 28 |
| I/O | Thick | 3.7 | 1.9 | 1.9 | 150 |
| LDMOS | Thick | 10 | 1.9 | 3.6 | 270 |

On one hand, breakdown voltages are critical for ensuring reliable operation in all scenarios (including high VSWR created by variations of the antenna impedance). These maximum voltages are determined by the gate oxide thickness and overall device construction. Table 8 shows that "core" devices offer the least ruggedness, while LDMOS give the best reliability in the available process node. This table also shows that LDMOS can endure larger drain-to-gate voltage than gate-to-source voltage. This asymmetry is caused by its "field plate" whose main purpose is to improve isolation between drain and gate (see section 2.3.2).

On the other hand, the minimum length together with the oxide thickness provide a rough idea of the transconductance of the devices because of two reasons. First, a thinner gate oxide layer increases the oxide capacitance, which directly affects the transconductance according to the well-known approximation $g_m \approx \mu_n C_{ox} \frac{W}{L}\left(V_{gs} - V_{Th}\right)$. The second reason is that a smaller channel length enables a higher transconductance for a given device width, which can also be seen in the previous equation. Consequently, from Table 8 it can be concluded that the "core" transistors have the largest transconductance per width unit, followed by the I/O devices and LDMOS come in last place.

It is important to note that the breakdown voltages in Table 8 refer to DC. According to [11], these values can be used to assess reliability when a large RF signal is present, by comparing its RMS voltage against the DC hard breakdown limits.

### *5.1.2 Circuit topology selection*

In order to obtain a large output voltage swing from a transistor, using an inductor to connect the power supply to the PA makes the most sense since this passive component by itself allows a maximum voltage swing of $2V_{dd}$.

Regarding the circuit topology, a cascode configuration is selected because of its advantages over a single transistor (as outlined in section 3.7.2). In addition, since the goal is implementing a single-stage PA, a large gain is required. Therefore, the conventional cascode topology (shown in section 3.7.2) is selected over the self-bias and the stacked configurations because it provides the largest gain.

In a cascode PA, the common-source (CS) transistor generates the PA output current by means of its transconductance, while the common-gate (CG) transistor protects the CS device against large output voltages. Therefore, a thin oxide device as the CS transistor will provide the highest transconductance for a given silicon area, while a thick oxide device as the CG transistor will offer increased ruggedness.

However, selecting between I/O and LDMOS for the CG transistor should be made carefully, as shown next.

### *5.1.3 Device selection for the CG transistor*

There is a trade-off when selecting between an I/O and an LDMOS transistor for the CG device in a cascode PA. On one hand, an I/O device offers a smaller knee voltage compared to an LDMOS device of the same width, as the minimum length of the I/O device is smaller. A lower knee voltage enables larger output voltage amplitude in linear operation. On the other hand, LDMOS offers significantly better ruggedness at the cost of additional silicon area and reduced voltage headroom. Therefore, an I/O transistor is preferred only if it can tolerate the maximum voltage present at the PA output.

In order to determine this maximum voltage, the worst case is considered. According to the design specifications, the PA must tolerate a VSWR of 8:1 (equivalent to an antenna reflection coefficient of $|\Gamma| = 0.78$). This means that the maximum peak voltage at the antenna port is

$$\begin{aligned} |v_{max}| &= |v_{inc}| + |v_{ref}| = |v_{inc}| + |v_{inc}\Gamma| = |v_{inc}|(1 + |\Gamma|) \\ &= 1.78|v_{inc}|, \end{aligned} \tag{28}$$

where $v_{inc}$ and $v_{ref}$ are the incident and reflected peak voltage at the antenna port, respectively, as illustrated in Figure 30. Also, if the output matching network (OMN) has a impedance transformation ratio of $N$ from port 1 to port 2, the relation between its port voltages is $|v_{port1}| = |v_{port2}|/\sqrt{N}$. Then, the voltage reflected from the antenna will come back to the OMN through its port 2 and will come out from its port 1 with an amplitude given by

$$\begin{aligned} |v_{ref,OMN}| &= \frac{1}{\sqrt{N}}|v_{ref}| = \frac{1}{\sqrt{N}}|v_{inc}\Gamma| = \frac{1}{\sqrt{N}}|\sqrt{N}\, v_{inc,OMN}\Gamma| \\ &= |v_{inc,OMN}\Gamma|. \end{aligned} \tag{29}$$

In the worst case this reflected voltage will add up with the voltage produced by the PA core, generating a maximum voltage at port 1 of the OMN equal to

$$\begin{aligned} |v_{max,OMN}| &= |v_{inc,OMN}| + |v_{ref,OMN}| = |v_{inc,OMN}| + |v_{inc,OMN}\Gamma| \\ &= 1.78|v_{inc,OMN}|. \end{aligned} \tag{30}$$

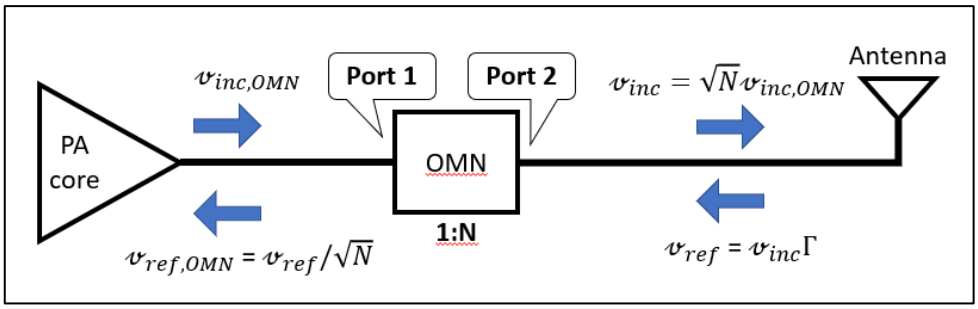



Figure 30: VSWR at antenna port and PA core output.

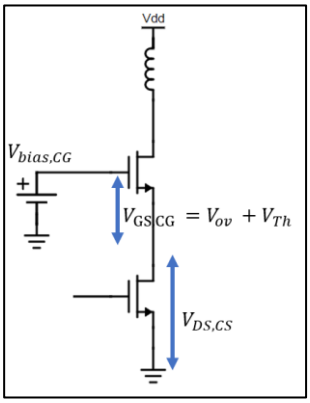
Since at the highest output power, the core of the PA will produce a voltage of amplitude $|v_{inc,OMN}| = V_{dd} - V_{knee}$, then this amplitude in an antenna mismatch condition will increase up to $|v_{max,OMN}| = 1.78\,(V_{dd} - V_{knee})$. Assuming $V_{knee} = 0.4\,V$ will result in $|v_{max,OMN}| = 3.2V$.

This maximum voltage corresponds to the maximum drain voltage of the CG transistor. Now, it should be determined whether this maximum voltage exceeds the breakdown limits presented in Table 8. To do this, the maximum drain-to-gate, drain-to-source and gate-to-source voltage for the CG device need to be computed. The drain-to-gate voltage will be calculated first since it is the most critical one. To do this, first the CG bias voltage needs to be obtained.

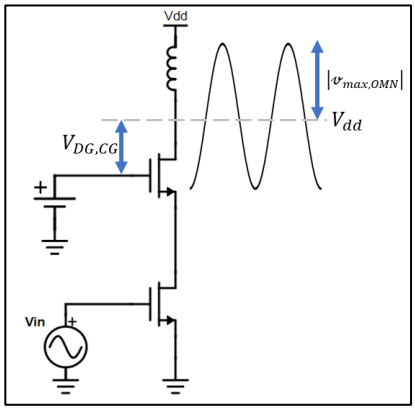
For this, the Figure 31 (a) will be used as a guide. First, it is assumed that the CG device has a DC overdrive voltage (that is, $V_{ov} = V_{GS} - V_{Th}$) of 400 $mV$ (this assumption will be validated later). Also, the threshold voltages for the I/O and the LDMOS transistors are needed; those are 625 $mV$ and 645 $mV$, respectively[6]. Then, to ensure a $V_{DS,CS}$ (DC drain-to-source voltage for the CS device) of 800 $mV$ (a protection margin of 100 $mV$ before breakdown), the bias voltage for the CG device is

$$V_{bias,CG} = V_{ov} + V_{Th} + V_{DS,CS}. \tag{31}$$

Hence, $V_{bias,CG}$ for I/O and LDMOS transistors are 1825 $mV$ and 1845 $mV$, respectively.




(a) $V_{bias,CG}$ calculation diagram




(b) Maximum $V_{DG,CG}$ calculation diagram

Figure 31: Diagram for the CG transistor calculations.

[6] The body effect was accounted for since the source terminal of the CG transistor is at a potential higher than ground in the cascode topology.

Also, from Figure 31(b), it can be seen that the maximum $V_{DG,CG}$ is

$$V_{max,DG,CG} = \sqrt{\frac{1}{2}\left|v_{max,OMN}\right|^2 + (V_{dd} - V_{bias,CG})^2}. \quad (32)$$

Considering that the breakdown of the oxide layer is triggered by the RMS voltage and not by the peak voltage [11], the calculation of $V_{max,DG,CG}$ is done by computing the RMS value of a sine wave (of amplitude $\left|v_{max,OMN}\right|$) with a DC offset (of $V_{dd} - V_{bias,CG}$). Then, the previous equation gives a $V_{max,DG,CG}$ for the I/O transistor of $2.294\ V$ and for the LDMOS of $2.29\ V$. Observing the Table 8, the drain-to-gate breakdown voltages for the I/O transistor is $1.98\ V$ and for the LDMOS is $3.6\ V$. Therefore, the CG device must be an LDMOS to avoid breakdown. Compliance with the other breakdown limits will be checked later in the design process.

Coming back to the DC overdrive voltage assumption, it can be argued that assuming a value larger than $400\ mV$ would allow using an I/O transistor. For this to happen, the overdrive voltage would have to be increased up to $1060\ mV$. Unfortunately, this would require an excessively large bias current or a very small device, both cases seriously affecting the performance of the PA.

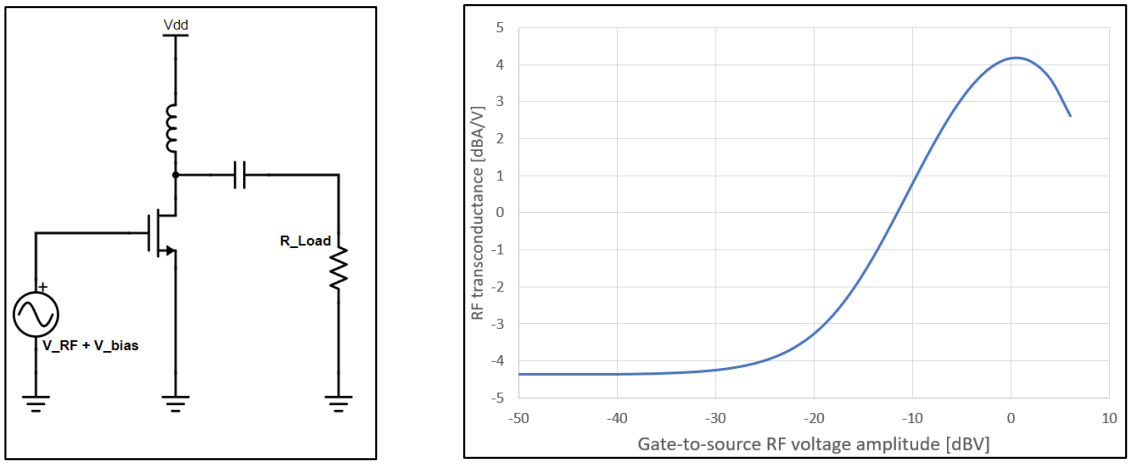
Furthermore, it is important to note that in the calculation of the maximum drain voltage the losses of the OMN were not included. Additionally, there is an antenna switch between the OMN and the antenna, which contribute with more losses. These losses have the effect of reducing the amplitude of the reflected voltage seen in the drain terminal of the CG transistor. However, even when considering these losses, the drain-to-gate voltage is still above the breakdown limit for the I/O device.

### *5.1.4 Linearity of the available transistors*

Understanding the sources of non-linearity in the transistors of the available process node is fundamental for the design of a linear PA. Besides clipping, the main cause of non-linearity in a transistor is its transconductance [14]. Although other sources are also important, such as its non-linear parasitic capacitances, the transconductance will be the focus of this sub-section.

The RF transconductance[7] of a "core" transistor of large width and minimum channel length is presented in Figure 32. Here, the device was biased at $50\ mV$ above the threshold voltage. Also, the load resistance was set to a very low value ($0.1\ \Omega$) to limit the amplitude of the drain voltage, delaying entry into triode region. It can be seen that the RF transconductance increases after $-26\ dBV$ ($50\ mV_p$). At this point, the drain current starts clipping (since the transistor is transitioning from class-A to class AB). Since a clipped sine wave has a DC component (see equation 4 at the beginning of this document), the effective bias current will be higher, boosting the transconductance. This behavior is known as "self-biasing".

[7] Defined here as the ratio between the amplitude of the first harmonic of the drain current and the amplitude of the first harmonic of the gate-to-source voltage.




(a) Schematic (b) RF transconductance

Figure 32: RF behavior of "core" transistor.

Also, at 0 $dBV$ (1 $V_p$) the transconductance starts saturating. At this point, the drain voltage amplitude is so large that it crosses the knee voltage, sending the device into triode region. However, at 1 $V_p$ the gate-to-source RMS voltage exceeds the breakdown limit, meaning that in this case the transistor would destroy itself before entering triode region.

### *5.1.5 PA basic calculations*[8]

#### *5.1.5.1 Required drain current and optimum load impedance*

The 1 dB compression point output power after the OMN should be 21 $dBm$ according to the design specifications. Assuming 1.5 $dB$ loss in the OMN, the power at the drain of the CG transistor should be 22.5 $dBm$ (or 178 $mW$) at the compression point. Then, assuming a knee voltage of 500 $mV$,

$$P_{drain,1-dB} = \frac{1}{2}\left|\mathcal{v}_{drain,1-dB}\right|\left|\mathcal{i}_{drain,1-dB}\right| = \frac{1}{2}(V_{dd} - V_{knee})\left|\mathcal{i}_{drain,1-dB}\right|$$
$$\left|\mathcal{i}_{drain,1-dB}\right| = \frac{2\,P_{drain,1-dB}}{(V_{dd} - V_{knee})} = \frac{2 \cdot 178\ mW}{2.2\ V - 0.5\ V} = 210\ mA. \quad (33)$$

Then, the required load resistance that creates a voltage swing of $V_{dd} - V_{knee}$ for a drain current with amplitude of 210 mA is

$$R_{opt} = \frac{\left|\mathcal{v}_{drain,1-dB}\right|}{\left|\mathcal{i}_{drain,1-dB}\right|} = \frac{1.7\ V}{210\ mA} = 8.1\ \Omega. \quad (34)$$

So, the OMN needs to convert 50 Ω down to 8.1 Ω. The OMN design is presented later.

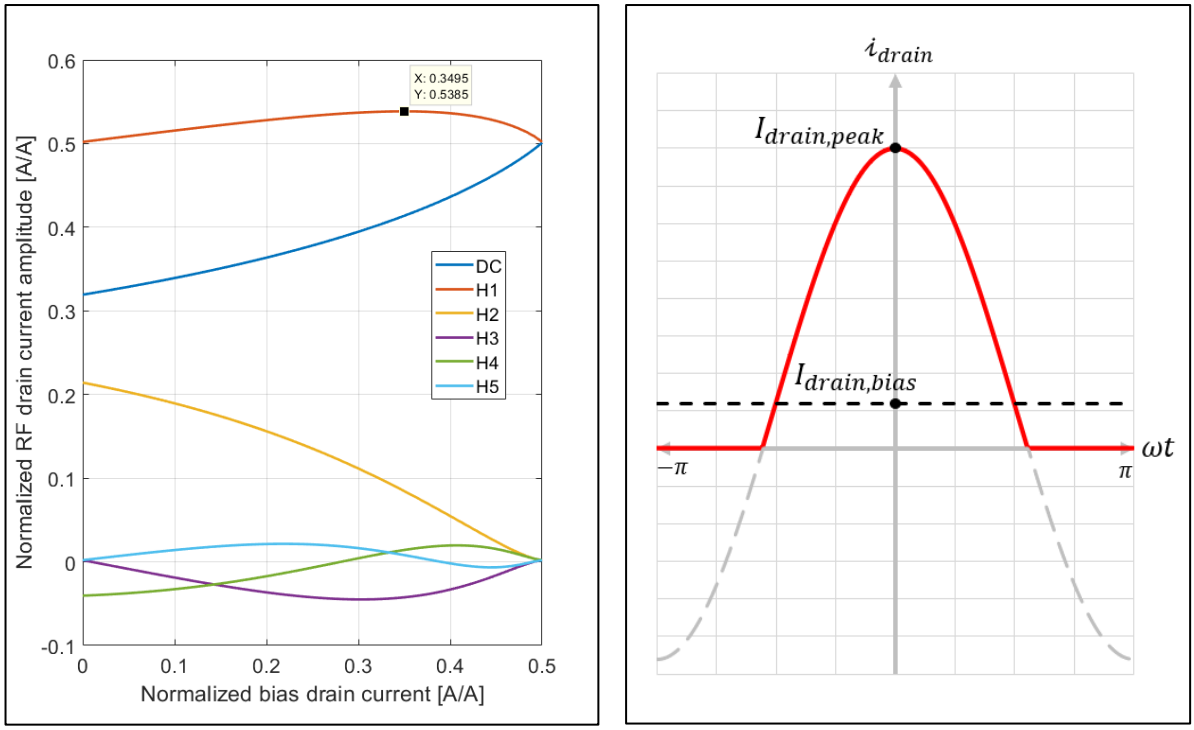
[8] In these calculations, upper case letters (such as $V$ and $I$) represent DC quantities, while script letters (such $\mathcal{v}$ and as $\mathcal{i}$) represent RF quantities.

### *5.1.5.2 Bias current*

Now, the bias drain current $I_{\text{drain,bias}}$ needs to be found. This parameter is crucial since it determines in part the transconductance (and hence the power gain) of the PA, the conduction angle (and hence the PA class) and its efficiency.

To determine $I_{\text{drain,bias}}$ the PA conduction angle (and thus the PA class) is chosen at the RMS power of the modulation, instead of at the peak power of the modulation. To do this, the harmonic components of a theoretical rectified sine wave drain current are plotted with respect to the bias current, as presented in Figure 33(a). This is the same plot shown in Figure 18, with the difference that the $x$-axis is the bias current instead of the conduction angle.

In this plot, class-A corresponds to a bias current of 0.5 times the peak current, class-B corresponds to a zero bias current and class-AB corresponds to bias current between these two values. For clarity, the peak current and the bias current are indicated in the drain current waveform in Figure 33(b).




(a) Amplitude of harmonics (b) Reference diagram

Figure 33: Amplitude of the harmonics of the drain current with respect to the bias current. All currents were normalized with respect to the peak drain current $I_{drain,peak}$.

Since this is a PA for a mobile application with considerable linearity requirements, selecting class-AB makes sense, presenting a good balance between linearity and efficiency. Furthermore, the bias point that produces the maximum fundamental current is selected, corresponding to a bias equal to 35% of the peak current (marked in Figure 33(a)).

Considering that the RMS power of the modulation is

$$P_{drain,RMSmod} = P_{drain,1-dB} - PAPR = 22.5\,dBm - 5\,dB = 17.5\,dBm. \quad (35)$$
$$\equiv 56\,mW$$

Then, the drain current amplitude at that power level is $118mA$, which is found by using

$$P_{drain,RMSmod} = \frac{1}{2}\left|i_{drain,RMSmod}\right|^2 R_{L,opt}. \tag{36}$$

And 35% of that current is 41 $mA$, corresponding to the bias current $I_{drain,bias}$.

### *5.1.5.3 Required PA input impedance*

The PA input impedance is a key parameter because it helps to determine the required transconductance of the CS device and the size and loss of the input balun. The required input impedance can be achieved by using feedback.

As noted in section 5.1.4, the PA input voltage amplitude can be as large as needed without affecting the linearity of the RF transconductance. The limiting factor for the input voltage amplitude is the gate-to-source breakdown voltage. Therefore, considering $150\ mV$ of safety margin before breakdown and assuming a gate-to-source bias voltage of $50\ mV$ above the threshold (this assumption will be revisited later), the maximum input voltage amplitude is given by

$$\begin{aligned} v_{in,RMS} + V_{bias,CS} &< V_{GS,breakdown} - V_{margin} \\ v_{in,peak} &< \sqrt{2}\left[V_{GS,breakdown} - V_{margin} - (V_{Th} + 50\ mV)\right] \\ v_{in,peak} &< 1.4142 \cdot [900\ mV - 150\ mV - (380\ mV + 50\ mV)] \\ v_{in,peak} &< 452\ mV. \end{aligned} \tag{37}$$

This peak voltage will be present at the PA input at the 1 dB compression point (that is, $v_{in,1-dB} = 452\ mV$), in which the modulator generates an output current of $i_{Mod,1-dB} = -54\ dBA_p$ $(2\ mA_p)$. Since the optimum load impedance[9] of the modulator is $R_{opt,Mod} = 315\ \Omega$, then its output power is

$$P_{out,Mod,1-dB} = \frac{1}{2}\left|i_{Mod,1-dB}\right|^2 R_{opt,Mod} = \frac{1}{2}(2\ mA)^2 \cdot 315\ \Omega = 630\ \mu W. \tag{38}$$

Furthermore, assuming $4\ dB$ for the balun loss, the power delivered to the PA input is $P_{in,PA,1-dB} = 251\ \mu W$. So, at this power level the PA input resistance needs to produce a voltage of $v_{in,1-dB} = 452\ mV$. For this to happen

$$\begin{aligned} P_{in,PA,1-dB} &= \frac{1}{2}\frac{\left|v_{in,1-dB}\right|^2}{R_{in,PA}} \\ R_{in,PA,1-dB} &= \frac{1}{2}\frac{\left|v_{in,1-dB}\right|^2}{P_{in,PA,1-dB}} = \frac{1}{2}\frac{(452\ mV)^2}{251\ \mu W} = 407\ \Omega. \end{aligned} \tag{39}$$

[9] The optimum load impedance of the modulator was given in the design specifications. More details about this parameter can be found in section 5.4.

### *5.1.5.4 Required transconductances*

With the parameters found in the previous steps, the transconductances of the CS and the CG devices can be determined.

For the CS device, at the 1 dB compression point it needs to produce a drain current of $i_{drain,1-dB} = 210\ mA$ for a input voltage amplitude of $v_{in,1-dB} = 452\ mV$. Assuming that the main source of the PA non-linearity comes from the transconductance, at the 1 dB compression point the transconductance should have decreased down to 90% of its small-signal value[10]. Therefore, the transconductance of the CS device needs to be

$$g_{m,CS} = \frac{1}{0.9} g_{m,CS,1-dB} = \frac{i_{drain,1-dB}}{0.9\ v_{in,1-dB}} = \frac{210\ mA}{0.9 \cdot 452\ mV} = 515\ mA/V. \qquad (40)$$

For the CG device, its transconductance has to be maximized to obtain maximum linearity. To understand why, first it needs to be noted that the input impedance of a CG amplifier can be approximated by a resistance of $1/g_{m,CG}$ in parallel with its gate-to-souce capacitance. In a large transistor, the real part of this impedance is dominant. Therefore, as seen in Figure 34, in a cascode stage the voltage swing at the drain of the CS transistor is approximated by $i_{drain}/g_{m,CG}$.

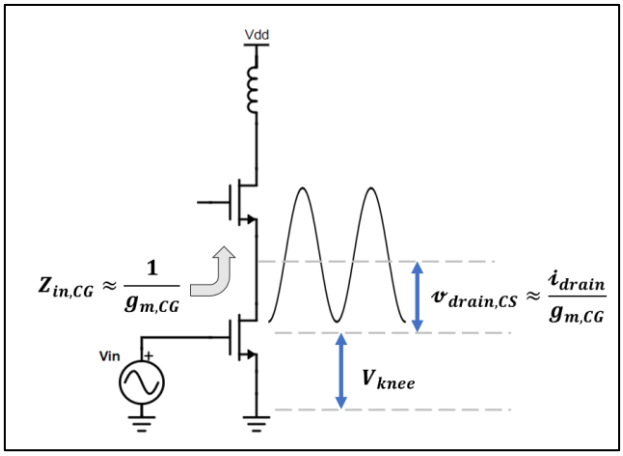



Figure 34: Effect of the transconductance of the CG device in the linearity of PA.

If the drain voltage amplitude at the CS drain terminal is large, it can exceed the knee voltage affecting linearity. So, to minimize this voltage amplitude the transconductance of the CG transistor should be maximum.

### *5.1.5.5 Transistor sizes*

Having the transconductances and bias current defined allows to determine the size of the power transistors.

For the CS device, a unitary cell is defined with 6 fingers, each one 1 $\mu m$ wide. Also, a channel length close to the minimum is used. The number of cells is swept until the required RF transconductance is obtained. This is done by using the schematic shown in Figure 32 with the bias current previously found, a 1 $mV_p$ input voltage and a load impedance of 1 Ω (to prevent

[10] Because $-1dB$ corresponds to an amplitude reduction factor of $10^{-1/20} = 0.9$

exceeding the knee voltage). The result can be found in Figure 35. A transconductance of $515\ mA/V$ is obtained by using 230 cells.

For this transistor size, the gate-to-source bias voltage needed to obtain the PA bias current of $41\ mA$ is $60\ mV$ over the threshold. This value is not far from the $50\ mV$ assumed in section 5.1.5.3 for the computation of the maximum input voltage amplitude.

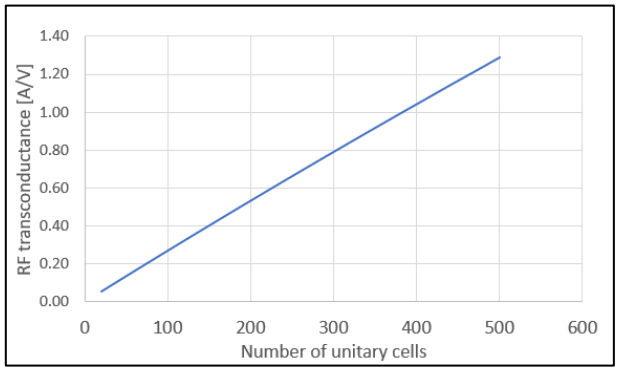



Figure 35: RF transconductance vs CS device size.

However, the size of the CG device is determined in a different way. At first glance, its size should be maximized because the required transconductance needs to be as large as possible. However, the limit for its size is given by its output capacitance since a physically large transistor will exhibit a high output capacitance which is detrimental for the PA performance. A large capacitance will sink a significant fraction of the output current, as shown in Figure 36, reducing the current flowing into the load. The fraction of the output current that is lost in the output capacitance is dependent on the ratio between the load impedance and the output capacitance reactance (a current divider is formed).

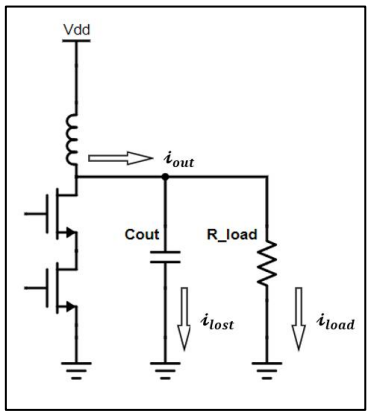



Figure 36: Effect of the output capacitance on the current delivered to the load.

So, to achieve a good balance between a large transconductance of the CG device with low output capacitance, it is proposed an output capacitance reactance two times larger than the optimum load impedance, that is $X_{Cout} = 16.2\ \Omega$. At the center frequency, the output capacitance would be $5.3\ pF$.

An LDMOS unitary cell of minimum length and two 1 $\mu m$-wide fingers are used. A total of 850 cells will exhibit the calculated output capacitance with a resulting RF transconductance of $450\ mA/V$.

### *5.1.5.6 Feedback for input impedance adjustment*

The input impedance of the PA using the size of the transistors and the bias current found in the previous section is $2.38 - j63.2\ \Omega$, which corresponds to a $1680\ \Omega$ resistance in parallel with a $1.36\ pF$ capacitance. Since the required PA input resistance is $407\ \Omega$, resistive feedback can be used to reduce the $1680\ \Omega$ to the required level. A simplified schematic for this scenario is shown in Figure 37.

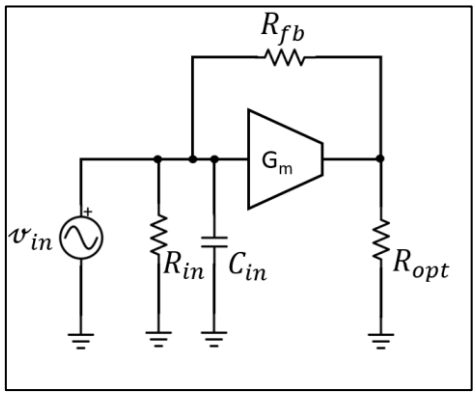



Figure 37: Resistive feedback simplified schematic.

With the resistive feedback, it can be shown that the input admittance seen by the voltage source is

$$Y_{in,fb} = \frac{1}{R_{fb}} + \frac{1}{R_{in}} + \frac{1}{j\omega C_{in}} + \frac{G_m - 1/R_{fb}}{1 + R_{fb}/R_{opt}}. \tag{41}$$

For this particular case, the feedback resistor can have a value of a few $k\Omega$, $G_m$ is around $0.5\ A/V$ and $R_{opt}$ is $8.1\ \Omega$. Then, the last term in the input admittance expression can be approximated to $G_m R_{opt}/R_{fb}$, and thus

$$Y_{in,fb} \approx \frac{1}{R_{in}} + \frac{1}{j\omega C_{in}} + \frac{1 + G_m R_{opt}}{R_{fb}}. \tag{42}$$

This means that the feedback adds a resistance of value $R_{fb}/(1 + G_m R_{opt})$ in parallel with the input of the PA, effectively lowering its input resistance without modifying its input capacitance.

So, to obtain a $407\ \Omega$ total resistance from a parallel between $R_{fb}/(1 + G_m R_{opt})$ and $1680\ \Omega$, it would require a $R_{fb} = 2780\ \Omega$. In Figure 38, the actual PA input resistance and input capacitance is plotted with respect to $R_{fb}$, validating the previous calculation.

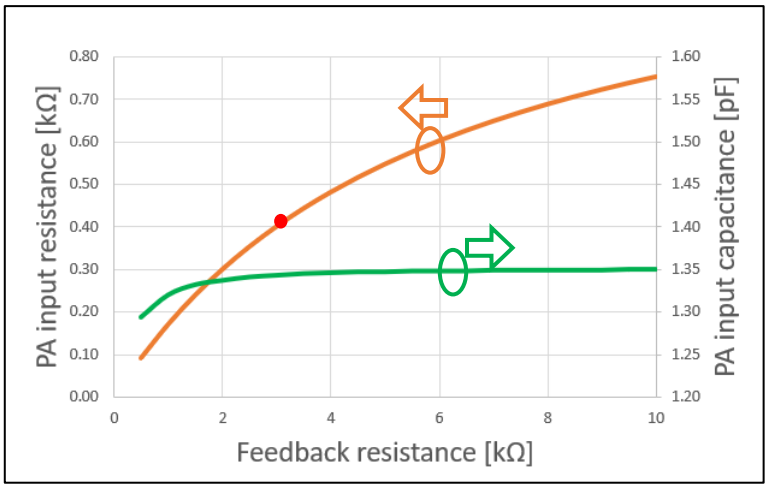



Figure 38: PA input impedance modified by the resistive feedback.

## 5.2 Bias circuit

The main objectives of the bias circuit are:

- To provide a steady bias operating point to the PA.
- To minimize variation on the PA performance across power supply, process and temperature variations.
- To present a low input impedance at the modulation envelope frequencies in order to minimize signal distortion.

Then, in this section a strategy is proposed to achieve these objectives.

### *5.2.1 Constant-transconductance bias circuit*

Reducing the effect of process, temperature and supply variations can be done by utilizing a bias constant-transconductance circuit [23]. This circuit, shown in Figure 39, forces the transconductance of a transistor to be dependent on a resistor value.

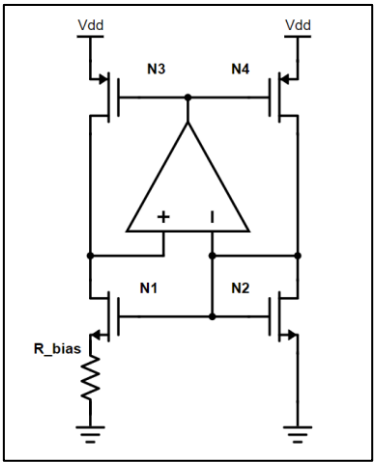



Figure 39: Constant-transconductance bias circuit

In this circuit, the operational amplifier forces the drain voltage of the PMOS devices to be equal resulting in equal drain currents. Then, by using the first order approximation for the drain current in transistors $N_1$ and $N_2$, it can be shown that [23]

$$g_{m,2} = \frac{2}{R_{bias}}\left(1 - \sqrt{\frac{W_2/L_2}{W_1/L_1}}\right), \tag{43}$$

and if transistor $W_1/L_1 = 4W_2/L_2$ then $g_{m,2} = 1/R_{bias}$. So, any transistor (of the same type as $N_2$) biased by this circuit will have also a transconductance depending only on $R_{bias}$.

However, this circuit has a few drawbacks. First, if $R_{bias}$ is an integrated component, its resistance will have a ~20% tolerance, which directly translates into a variation of the bias transconductance. Also, its resistance will vary with temperature. Second, this circuit has two feedback loops. A negative feedback loop created by the operational amplifier and transistor $N_3$ and a positive feedback loop created by the operational amplifier and transistors $N_4$ and $N_2$. The negative feedback loop can be compensated by connecting a large capacitor at the operational amplifier output, while the positive feedback loop has very low closed-loop gain [23]. Third, this circuit needs a start-up circuit because it has a stable operation point in which all currents are zero, increasing the complexity of the solution.

To validate the performance of the constant-$g_m$ circuit, the variation of the transconductance for a single test transistor across process corners and temperature is compared against the variation of the transconductance of transistor $N_2$ in the constant-$g_m$ circuit. To make a fair comparison, the test transistor has same size and similar bias current than $N_2$.

First, the change of the transconductance of the test transistor is shown in Table 9. To obtain that data, the test transistor was set in the diode-connected configuration and it was biased with a fixed current. To see the variations more clearly, percentage figures were added, using as reference the transconductance for the typical process corner at 25 °C (that is why the variation percentage is 0% for that corner).

It can be seen that the maximum change of transconductance for a single transistor is about $\pm 20\%$ across process corners and temperatures when compared to the typical case.

Table 9: Transconductance variation for a single transistor

| | **Transconductance [$mA/V$]** | | | **Transconductance variation with respect to typical corner at 25°$C$** | | |
|---|---|---|---|---|---|---|
| **Temperature [°$C$]** | **Typical corner** | **Fast corner** | **Slow corner** | **Typical corner** | **Fast corner** | **Slow corner** |
| **-40** | 1.413 | 1.432 | 1.391 | 20.7% | 22.3% | 18.8% |
| **25** | 1.171 | 1.182 | 1.156 | 0.0% | 0.9% | -1.3% |
| **+110** | 0.949 | 0.956 | 0.941 | -18.9% | -18.4% | -19.7% |

Second, the variation of the transconductance of transistor $N_2$ in the constant-$g_m$ circuit is shown in Table 10. To obtain this data, an ideal resistor was used as $R_{bias}$ in order to isolate the performance of the constant-$g_m$ circuit from the effect of temperature dependence of a real on-chip resistor. By using this compensation technique, the maximum change of transconductance is less than $\pm 2\%$ across process corners and temperatures when compared to the typical case.

Table 10: Transconductance variation for transistor $N_2$ in the constant-$g_m$ circuit

| | **Transconductance [$mA/V$]** | | | **Transconductance variation with respect to typical corner at 25 °$C$** | | |
|---|---|---|---|---|---|---|
| **Temperature [°$C$]** | **Typical corner** | **Fast corner** | **Slow corner** | **Typical corner** | **Fast corner** | **Slow corner** |
| **-40** | 1.161 | 1.155 | 1.169 | -0.8% | -1.3% | -0.1% |
| **25** | 1.170 | 1.165 | 1.176 | 0.0% | -0.4% | 0.5% |
| **+110** | 1.176 | 1.171 | 1.181 | 0.5% | 0.1% | 0.9% |

### *5.2.2 Voltage mismatch mitigation*

Now, the stable bias voltage provided by the constant-$g_m$ circuit can be used by the PA core. However, since the common-source (CS) power transistor in the PA core has a length close to the minimum one, its transconductance is a strong function of its drain-to-source voltage. So, connecting the constant-$g_m$ bias voltage directly would increase the transconductance variation for the PA core caused by the drain-to-source voltage mismatch between the bias circuit ($V_{DS,bias} \approx 400\ mV$) and the PA core ($V_{DS,CS} \approx 600\ mV$). This situation is illustrated in Figure 40.

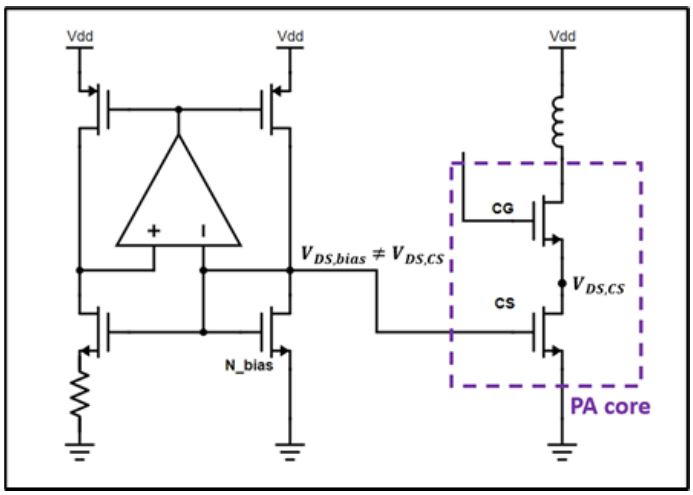



Figure 40: Drain-to-source voltage mismatch between bias circuit and PA core in a direct connection scenario.

Moreover, the effect of this mismatch also changes with process and temperature variations. This mismatch can increase the transconductance variation up to 13%, as seen in Table 10. To obtain the data in this table, the PA core was directly connected to the constant$-g_m$ circuit as shown in Figure 40 and the transconductance of transistor $CS$ was observed.

Table 11: Transconductance variation of PA core due to drain-to-source voltage mismatch

| | **Transconductance [$mA/V$]** | | | **Transconductance variation with respect to typical corner at 25 °$C$** | | |
|---|---|---|---|---|---|---|
| **Temperature [°$C$]** | **Typical corner** | **Fast corner** | **Slow corner** | **Typical corner** | **Fast corner** | **Slow corner** |
| **-40** | 635.0 | 692.7 | 591.4 | 3.4% | 12.8% | -3.7% |
| **25** | 614.0 | 660.2 | 578.8 | 0.0% | 7.5% | -5.7% |
| **+110** | 595.8 | 633.6 | 567.0 | -3.0% | 3.2% | -7.7% |

A way to mitigate this issue is presented in Figure 41. In this circuit, a small sample of the PA core is used to generate a bias voltage $V_{bias,PA}$ by means of the constant-$g_m$ circuit. The CS device in this PA sample has its drain-to-source voltage equal to the one of the actual PA core, mitigating the voltage mismatch.

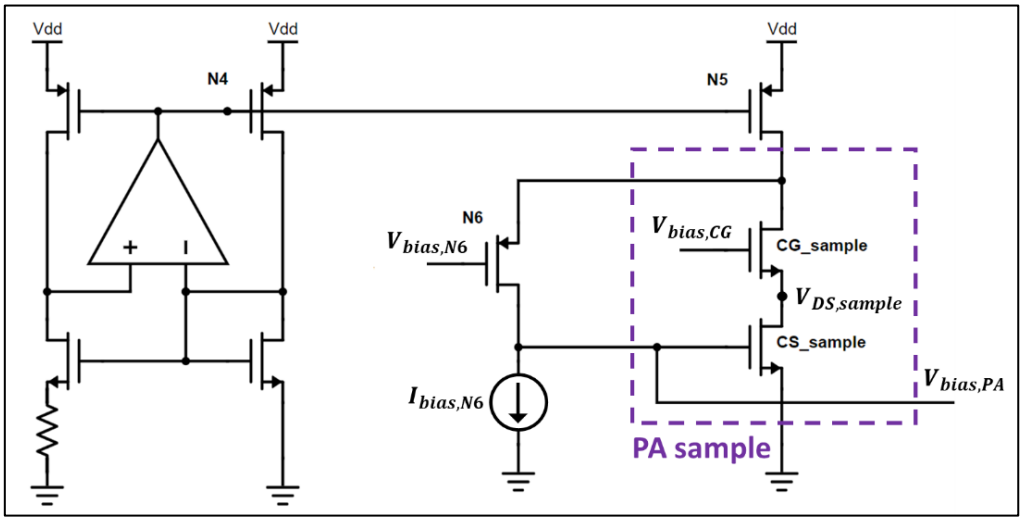



Figure 41: Bias circuit with improved drain-to-source voltage matching. $V_{bias,CG}$ is used to bias the CG device of the PA core.

In this bias circuit, transistors $N_4$ and $N_5$ form a current mirror, taking a copy of the constant-$g_m$ current, which is used to bias the PA sample. Additionally, transistor $N_6$ completes a closed loop, setting the PA sample into a *diode-connected configuration*. In this configuration, the PA sample generates a gate-to-source voltage linked to its drain current just as in the case of a single diode-connected transistor. Therefore, voltage $V_{bias,PA}$ is linked to the constant-$g_m$ current, thus providing a stable transconductance to the PA sample as well as to the PA core.

Moreover, $N_6$ provides a low impedance to the drain of the CG device in the PA sample, highly reducing the gain of the closed loop, thus improving its stability (since it is a negative feedback loop). To further reduce the closed-loop gain at high frequencies, a capacitor can be connected from the gate terminal of transistor $CS_{sample}$ to ground.

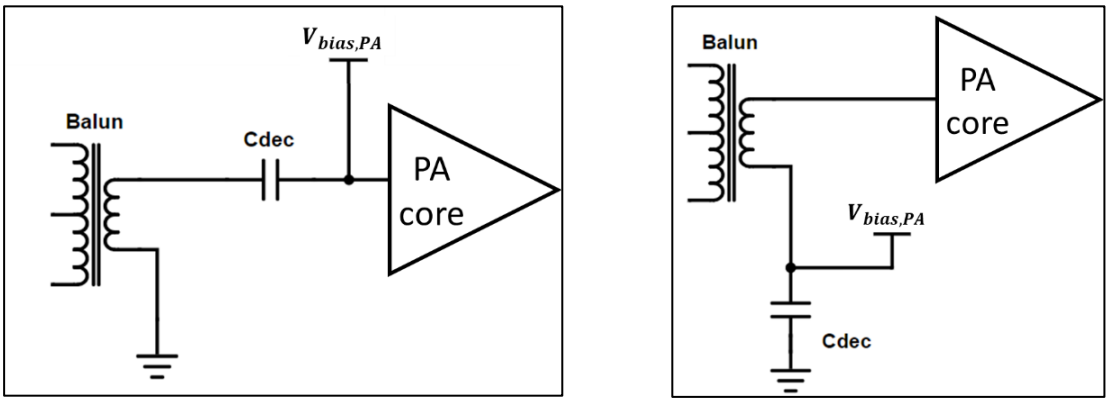
The transconductance variation across process corners and temperature is presented in Table 12. To obtain this data, the PA sample was connected to the constant-$g_m$ circuit as illustrated in Figure 41. Moreover, the node $V_{bias,CG}$ was used to bias the PA core. With this scheme the variation across corners was reduced down to 3%.

Table 12: Transconductance variation of PA core after voltage mismatch mitigation

| | **Transconductance [$mA/V$]** | | | **Transconductance variation with respect to typical corner at 25 °$C$** | | |
|---|---|---|---|---|---|---|
| **Temperature [°$C$]** | **Typical corner** | **Fast corner** | **Slow corner** | **Typical corner** | **Fast corner** | **Slow corner** |
| **-40** | 616.00 | 614.40 | 616.00 | -1.3% | -1.6% | -1.3% |
| **25** | 624.32 | 623.04 | 624.32 | 0.0% | -0.2% | 0.0% |
| **+110** | 635.52 | 639.04 | 632.96 | 1.8% | 2.4% | 1.4% |

### *5.2.3 Connecting bias circuit and PA core*

Typically, the bias voltage is fed into the PA core as shown in Figure 42(a) [11]. However, this approach has one significant drawback. The decoupling capacitance $C_{dec}$ must be implemented as a metal-oxide-metal capacitor since neither of its two terminals is connected to ground. This capacitor implementation has parasitic capacitance towards the substrate, creating a path for the RF signal to leak into ground, potentially reducing the gain of the whole PA. Moreover, $C_{dec}$ is large (around $20\ pF$) because it needs to provide a low impedance path at RF frequencies. This results in a large parasitic capacitance, aggravating the RF leakage to ground. This is the reason why the approach illustrated in Figure 42(b) is employed.




(a) Typical bias feed. (b) Bias feed through the balun

Figure 42: Bias feeding approaches for a single-ended PA.

### *5.2.4 Output impedance at baseband frequencies*

The decoupling capacitor $C_{dec}$ is connected in parallel to the bias circuit as shown previously in Figure 42 (b). This large capacitor connects the balun to the ground by offering a low impedance path (4 Ω) at the RF frequency. However, at baseband frequencies the impedance of $C_{dec}$ is 4 kΩ, which is far from a low impedance.

This is a serious issue because any low frequency component of the input signal will add to the bias voltage of the PA affecting its performance. Although the input signal by itself has no energy at low frequencies, the non-linear input impedance of the PA introduces distortion to the input signal generating a DC level (similar to the DC component that appears when rectifying a sinewave). This DC level accumulates in the capacitor $C_{dec}$ since there is no discharging path towards ground, drastically affecting the bias operational point of the PA.

For illustration purposes, Figure 43 shows the time-domain waveform of the input voltage of the PA core when an hypothetical bias circuit with a large output impedance at envelope frequencies is used. A large simulation time was selected in order to show the variation of the input signal envelope. It can be observed that the low frequency component of the input voltage of the PA (marked in red) is slowly increasing, drastically affecting the bias voltage of the PA.

A way to reduce this effect is designing the bias circuit so it has a low output impedance at low frequencies. The diode-connected PA sample cell previously shown in Figure 41 has an output impedance equal to $1/g_{m,CSsample}$. To see why, consider Figure 44. If a small voltage perturbation $v_x$ is present at the PA cell output (the gate of transistor $CS_{sample}$) then current $i_1$ would be $v_x\ g_{m,CSsample}$. Since transistor $N_5$ is a DC current source, it behaves as an open

circuit for a small-signal perturbation and thus $i_2 = -i_1$. Also note that $i_x = -i_1$. Then, the output impedance of the PA cell is $v_x/i_x = 1/g_{m,CSsample}$.

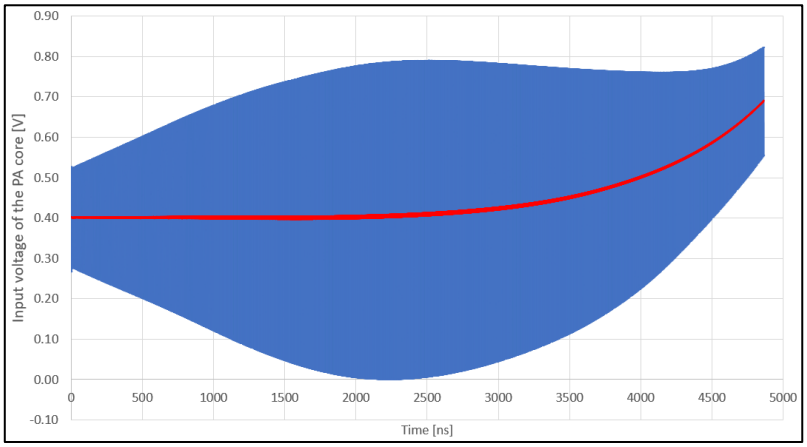



Figure 43: Input voltage of the PA (in blue) for a hypothetical bias circuit with large output impedance at low frequency. Its low frequency component is shown in red.

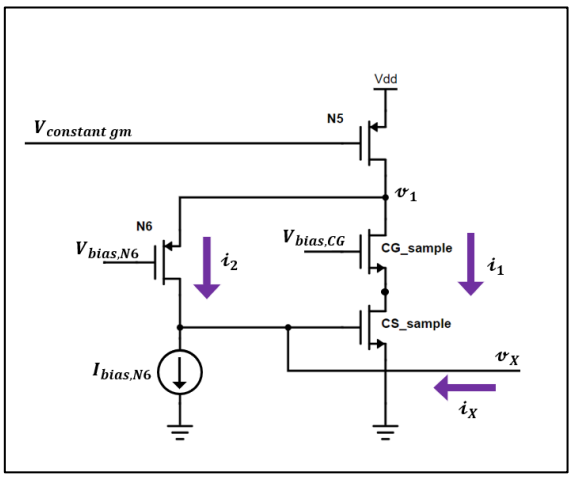



Figure 44: Schematic for the calculation of the output impedance of the PA cell.

If the size ratio between the PA cell and the PA core is 1:200, then $g_{m,CSsample} = g_{m,CS,PA}/200 = (500\,mA/V)/200 = 2.5\,mA/V$ and therefore, the output impedance of the PA cell is 400 Ω. Even though this is not a significantly low impedance, it is low enough to enable a discharge path for $C_{dec}$, (with a time constant of $8\,ns$), sending the low frequency component to ground.

Figure 45 shows the input voltage of the PA core with this diode-connected PA sample cell. Now the bias voltage of the PA is not changing with the envelope of the input signal.

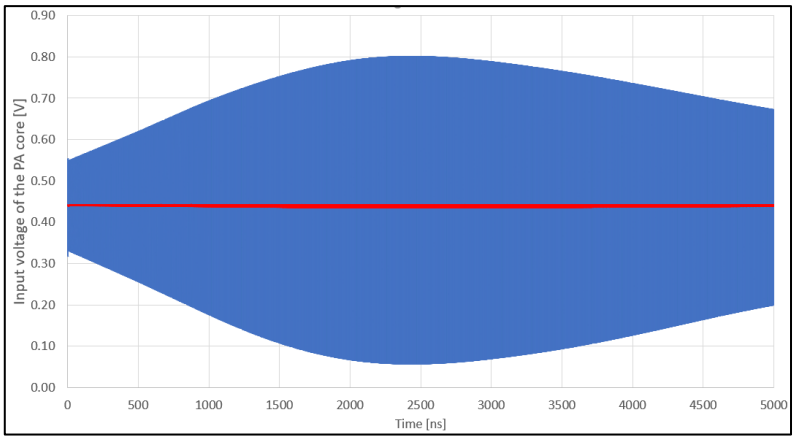



Figure 45: Input voltage of the PA (in blue) when using the proposed bias circuit. Its low frequency component is shown in red.

## 5.3 Output matching network

The main objective of the output matching network (OMN) for a PA is to provide the impedance transformation between the load (usually 50 Ω) and the optimum load impedance (usually lower than 50 Ω). Additional goals that must be achieved are:

- Deliver bias to the power transistors.
- Prevent RF power from flowing into the power supply.
- Provide proper harmonic termination for improved linearity or efficiency.
- Filter the signal so no power at high harmonics is consumed by the load.
- Prevent DC power from flowing into the load (DC decoupling).

Typically, the OMN is implemented off-chip because the substantial power level at the PA output makes unfeasible using low-Q integrated passive components. However, off-chip components also come with disadvantages. They are expensive, occupy valuable area in the printed circuit board (PCB), have low self-resonance frequency (SRF) and only a few capacitance and inductance values are available.

### *5.3.1 PCB components model*

For this PA design there are two available PCB component packages: 01005 and 0201. Generally, the 01005 package offers a smaller physical size and higher SRF at the cost of decreased power handling capability and lower Q factor when compared to the 0201 package, as shown in Table 13 and Table 14.

Table 13: PCB capacitor performance for different packages

| | **Self-resonance frequency [MHz]** | | **Series resistance at 1850 MHz [Ω]** | |
|---|---|---|---|---|
| **Capacitance [pF]** | **01005 pkg** | **0201 pkg** | **01005 pkg** | **0201 pkg** |
| 1.8 | 7811 | 7534 | 0.4 | 0.2 |
| 3.3 | 6799 | 5371 | 0.3 | 0.1 |
| 12 | 3607 | 2893 | 0.2 | 0.1 |
| 33 | 2011 | 1780 | 0.2 | 0.1 |

Table 14: PCB inductor performance for different packages

| | **Self-resonance frequency [MHz]** | | **DC series resistance [Ω]** | | **Q at 1850 MHz** | |
|---|---|---|---|---|---|---|
| **Inductance [nH]** | **01005 pkg** | **0201 pkg** | **01005 pkg** | **0201 pkg** | **01005 pkg** | **0201 pkg** |
| 1.2 | >3000 | >3000 | 0.2 | 0.05 | 23 | 45 |
| 3.3 | >3000 | >3000 | 0.4 | 0.1 | 21 | 38 |
| 7.5 | >3000 | 2080 | 1.1 | 0.3 | 20 | 31 |
| 12 | 1910 | 1490 | 1.8 | 0.4 | 16 | 21 |

For the capacitors in the OMN, the 01005 package is better suited for this design because it provides higher SRF which results in a smaller capacitance variation inside the PA frequency band (since the effective capacitance grows rapidly near the SRF). For the inductors in the OMN, the 0201 package is better suited because it offers the smaller DC series resistance and higher Q which is desirable for a component that is typically located in the DC power supply path and the RF power path.

The initial design of the OMN can be done by utilizing ideal reactances as a model for the PCB components. However, ideal reactances cannot model the SRF and the losses of real components. Therefore, a more accurate model needs to be used.

As a first order approximation, a PCB capacitor can be modelled as an RLC series circuit. Using the data compiled in the previous tables, it is possible to model all 01005 package capacitors by using a $190\ pH$ parasitic series inductance with a $0.3\ \Omega$ series resistance.

Moreover, a PCB inductor can be modelled by an RLC parallel circuit. From the previous tables, a $800\ fF$ capacitance approximately replicates the SRF of those components. However, the Q-factor and the equivalent parallel resistance varies too much to be modelled by a single value. Therefore, the worst case will be used, namely a parallel resistance of $700\ \Omega$.

With these models in place, a computer-aided optimization process can be utilized to obtain the OMN component values by maximizing the bandwidth while maintaining reasonable resistive losses.

### *5.3.2 Matching network topology*

To select the most suitable topology, the following key performance indicators will be used:

- Bandwidth: The OMN should provide a load impedance to the PA close to $R_{L,opt} = 8.1\Omega$ in the operating frequency band (1700 MHz – 2000 MHz). A 10% variation for the load impedance is allowed, which corresponds to a $|S_{11}| = -25\ dB$.
- Maximum acceptable in-band losses of $-1.5\ dB$.

Considering that non-linearities in the PA generate harmonics that must be filtered to comply with the out-of-band emission limits, a low-pass matching network is selected. At first, an LC-section topology is examined given that it is the simplest case. The schematic for a single LC-section and two LC-sections matching network was previously shown in Figure 22 and Figure 23 in section 3.8.

The optimum performance of a single LC-section and two cascaded LC-sections is shown in Figure 46 and Figure 47, respectively. In these simulations, the parasitics of each component were included.

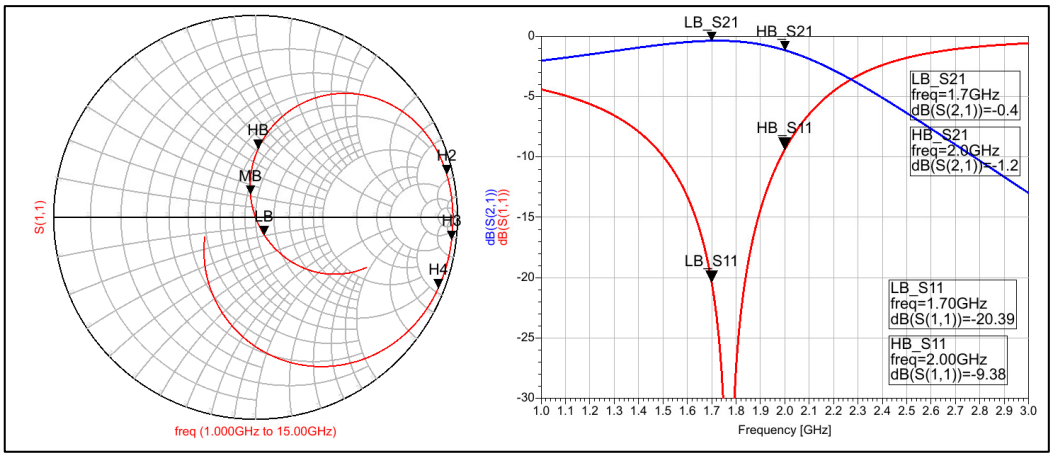



Figure 46: S-parameters for a single LC-section matching network.

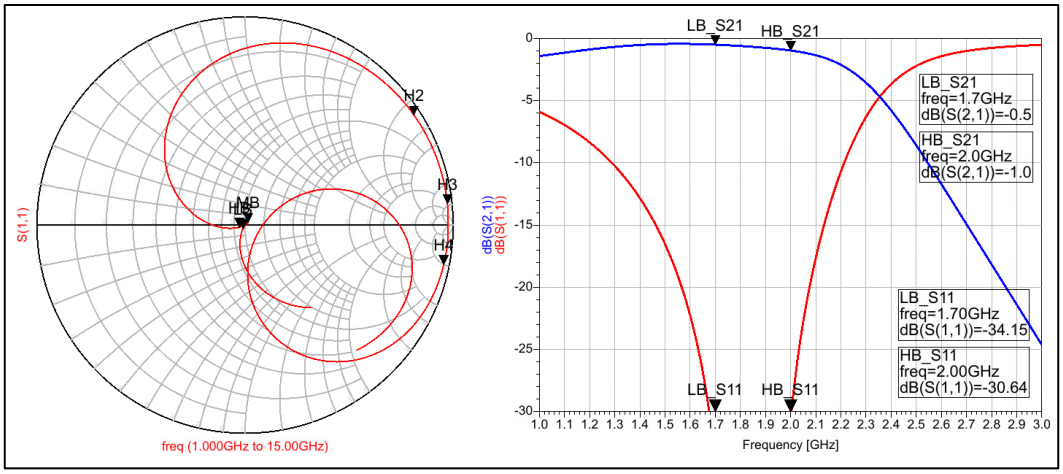



Figure 47: S-parameters for a two LC-sections matching network.

In these figures, the S-parameters were normalized by using a characteristic impedance equal to $R_{L,opt}$. Also, markers were placed at the center frequency (MB), low edge of the band (LB) and high edge of the band (HB) for clarity. Furthermore, markers were placed for the second, third and fourth harmonic frequencies (named H2, H3 and H4).

From these results, it can be concluded that two LC-sections produce the required bandwidth performance.

Also, it can be seen that the impedance presented to the PA output by the OMN at the harmonic frequencies is predominantly reactive and significantly large when compared with $R_{L,opt}$. Particularly, for the two LC-section case: $Z_{2H} = 1.6 + j27.1\ \Omega$, $Z_{3H} = 31.5 + j146.3\ \Omega$ and $Z_{4H} = 11.4 + j81\ \Omega$. This is problematic scenario because even if the PA drain current harmonic components have low amplitude, the drain voltage at these harmonics would be exceedingly large, seriously affecting its gain and linearity.

However, there is a key (and certainly false) assumption made while obtaining these results: the antenna impedance is 50 Ω for all the frequencies. This assumption will be used in this study because there is no information available regarding the antenna characteristics at higher frequencies.

### *5.3.3 PA output parasitics*

The OMN simulation results presented above did not included the effect of the PA output capacitance ($C_{out,PA}$) nor the parasitic inductance present in the interface between the PA output and the OMN input ($L_{out,pkg}$). This inductance is inherent to the metal traces from the PA drain terminal to the package bounding ball, the bounding ball itself and the PCB trace from the bounding ball to the OMN input. A simplified schematic of this situation is illustrated in Figure 48. Additionally, the parasitic inductance that exists in the interface between the PA ground and the PCB ground ($L_{gnd,pkg}$) will affect the impedance matching and must be considered as part of the OMN. These inductances have also resistive losses, namely $R_{out,pkg}$ and $R_{gnd,pkg}$.

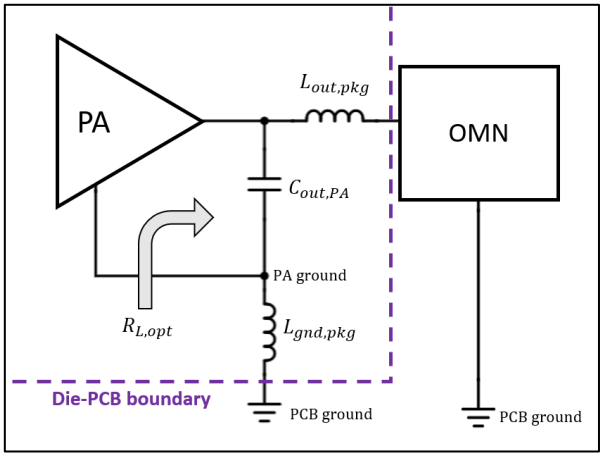



Figure 48: PA output parasitics.

The estimated values for these parasitics are listed in Table 15.

Table 15: Estimated values for PA output parasitics.

| **Parameter** | $C_{out,PA}$ | $L_{out,pkg}$ | $L_{gnd,pkg}$ | $R_{out,pkg}$ | $R_{gnd,pkg}$ |
|---|---|---|---|---|---|
| **Value** | 6.6 pF | 200 pH | 100 pH | 200 mΩ | 100 mΩ |

The result after optimizing the OMN including all these parasitics is presented in Figure 49.

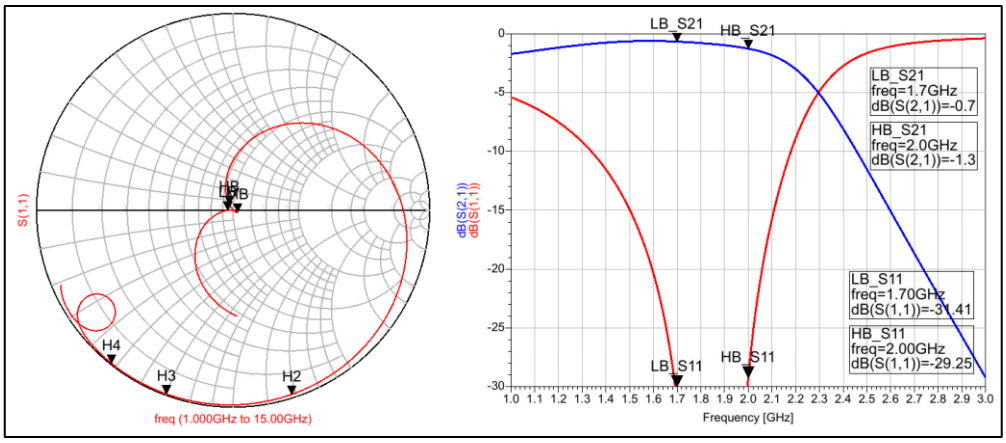



Figure 49: S-parameters for a two LC-sections matching network including PA output parasitics.

Fortunately, the large PA output capacitance produces a low impedance for the third and fourth harmonics. Specifically, $Z_{3H} = -j5.7\ \Omega$ and $Z_{4H} = -j3.9\ \Omega$. Though, for the second harmonic the impedance presented to the PA is $Z_{2H} = 0.2 - j11.1\ \Omega$, which is still large to some extent. This can affect PA performance as explained next.

### *5.3.4 Second harmonic short circuit*

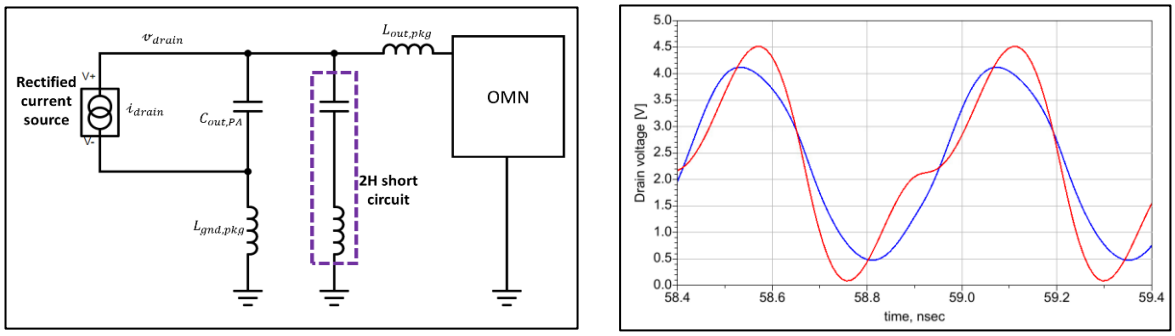
In a class-AB PA the drain current has a second harmonic component that increases with output power, since at very low power it behaves as a class-A (where the second harmonic amplitude is non-existent) while at large output power it behaves almost as a class-B (where the second harmonic amplitude is significant). For this particular design, the amplitude of the second harmonic component of the drain current varies from 0% to 40% of the amplitude of its fundamental.

This large second harmonic of the drain current is of main concern, because it generates a large second harmonic in the drain voltage, forcing the power transistors to enter into triode region rapidly. This problem exacerbates if the OMN provides a large input impedance at the second harmonic.

This issue can be understood by analyzing a simplified model. In Figure 50(a), the PA is modelled by an ideal rectified current source with amplitude and DC level equal to those of the calculated drain current at the 1 dB compression point (see section 5.1.5). With this model two tests were made: In the first one, a two LC-section OMN was used; in the second one, a short circuit at the second harmonic (SC2H) was added to the output of the PA and the two LC-section OMN was designed again taking into account the impedance of this SC2H at the fundamental[11]. In Figure 50(b), the drain voltage waveforms for these two tests are presented. It can be seen that the presence of the second harmonic increases the voltage swing by $400\ mV$, crossing the knee voltage sooner (drain voltage fundamental component is the same in both cases).

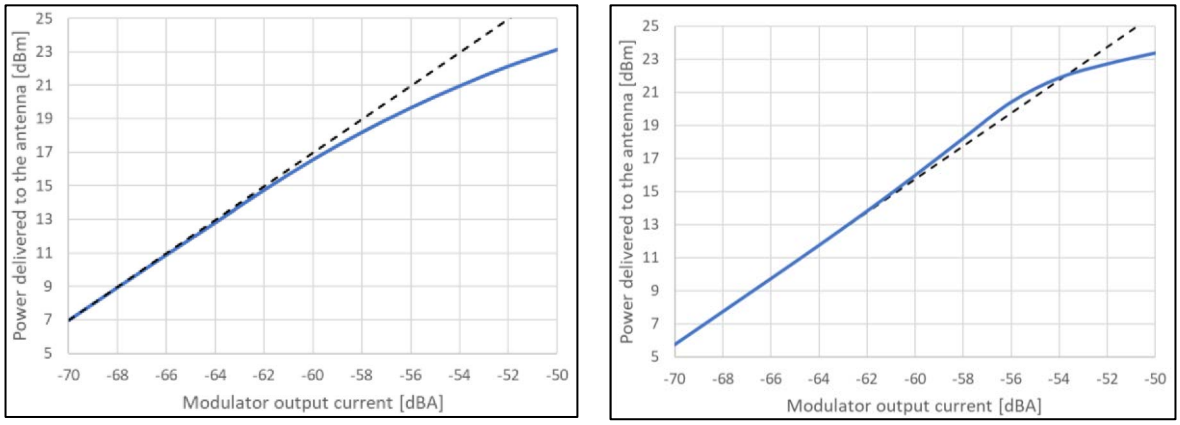
A more realistic scenario was also analyzed by simulating the power transistors and including the input balun model and the bias circuit. The 1 dB compression point curves are shown in Figure 51.

[11] The SC2H is capacitive at the fundamental, a short circuit at the second harmonic and behaves as an inductor for higher frequencies.




a) Diagram for simplified test b) Drain voltage waveforms

Figure 50: Simplified SC2H test. Drain voltage waveform for OMN with (blue) and without (red) the SC2H.




(a) No second harmonic short circuit (b) With second harmonic short circuit

Figure 51: PA 1 dB compression point curves demonstrating the linearity improvement given by the SC2H.

Finally, the proposed implementation of the second harmonic short circuit is a 3.67 $pF$ on-chip metal-insulator-metal capacitor in series with an inductance of 400 $pH$ formed by a bounding ball in series with a PCB metal trace.

### *5.3.5 RF choke and DC decoupling*

To deliver the power supply voltage to the PA and enable twice the voltage amplitude at the PA output, a large inductor is typically used. From the available PCB components, the only large inductor with low DC resistance is a 12 $nH$ inductance contained in the 0201 package. Its DC resistance is 0.3 $\Omega$.

Moreover, to prevent DC into the load while introducing minimum loss in the RF path, a large capacitance is traditionally utilized. Unfortunately, a large capacitor will self-resonate with its own package parasitic inductance at very low frequencies, acting as an inductor at the RF frequency. For instance, a 1 $nF$ capacitor inside a 01005 package has a SRF of 600 MHz. On top of that, a small capacitance presents a large impedance which is undesirable. For example, a 12 $pF$ capacitor in a 01005 package has an impedance of 5 $\Omega$.

An interesting fact about PCB capacitors is that they act as series RLC circuits, providing a very low impedance at the SRF. This fact can be exploited by selecting a capacitor with SRF inside the PA frequency band. For instance, a 33 $pF$ capacitor in a 01005 package self-resonates at 2032 $MHz$ exhibiting a 0.2$\Omega$ impedance at that frequency. Furthermore, in the band 1650 − 2500 $MHz$ it has an impedance lower than 1 $\Omega$, so any SRF variation caused by manufacturing tolerances will still give a low impedance at the PA frequency.

### *5.3.6 OMN using S-parameter files provided by the manufacturer*

The OMN schematic is shown in Figure 40. The calculated component values obtained through computer-aided optimization are listed in Table 16. It also contains the available PCB components that have the closest value to the optimum. For instance, for $C_2 = 3.48\,pF$, the closest available values are 3.3 $pF$ and 3.9 $pF$ in the 01005 package. In addition, Table 16 contains the PCB available component values that gives the best performance (closest to the optimum performance).

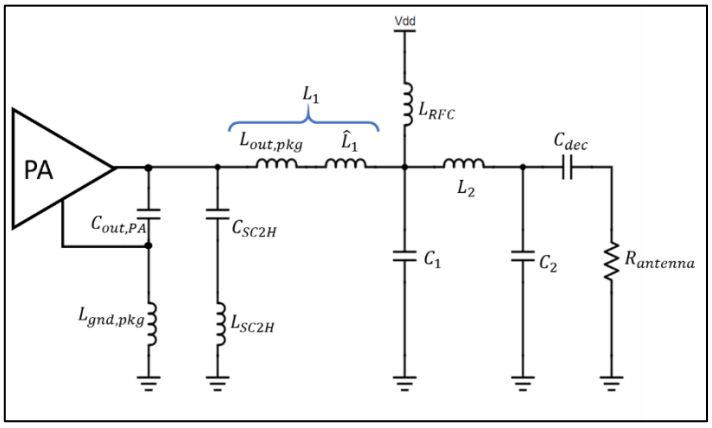



Figure 52: OMN schematic.

Table 16: Component values for the OMN. Component package in parenthesis

| **Component** | $L_1$ | $C_1$ | $L_2$ | $C_2$ |
|---|---|---|---|---|
| **Optimum value** | 770 pH | 9.96 pF | 2.05 nH | 3.48 pF |
| **Closest available value** | $L_{out,pkg}$ + 600 pH (0201) | 10 pF (0201) | 2 nH (0201) | 3.3 pF (01005) |
| **Best performance value** | $L_{out,pkg}$ + 600 pH (0201) | 9.7 pF + 2.7 pF (01005) | 2.4 nH (0201) | 1.8 pF + 1.8 pF (01005) |

Using the PCB components with the inductance/capacitance numerically closest to the optimum ones gives unacceptable performance for two main reasons: First, the inductance/capacitance values provided by the manufacturer are measured at low frequency, which is inaccurate for RF since the inductance/capacitance changes significantly for higher frequencies. Second, there are no available values for some capacitance/inductance ranges; for example, there are not capacitances with values between 9 $pF$ and 12 $pF$ or between 12 $pF$ and 22 $pF$ for the 01005 package. So, a large error is introduced when selecting the numerically closest value (instead of selecting the component that gives the closest performance).

In order to obtain a performance similar to the optimum one, the S-parameter files published in the manufacturer website were utilized[12]. For the capacitors, the best result is obtained by using two components in parallel to create the desired capacitance since the values offered by the manufacturer are quite limited. Although this approach is more expensive, it produces an OMN that behaves similar to the optimum one, which is difficult to achieve when $R_{L,opt}$ is smaller than $10\,\Omega$. Also, having two capacitors instead of one makes the OMN physically larger, requiring larger transmission lines, adding losses and parasitic reactances.

Finally, the best performance of the OMN obtained by using the S-parameter files of the PCB components is shown in Figure 53. For comparison, the optimum OMN and the OMN produced by using the numerically closest component value are also presented in that figure.

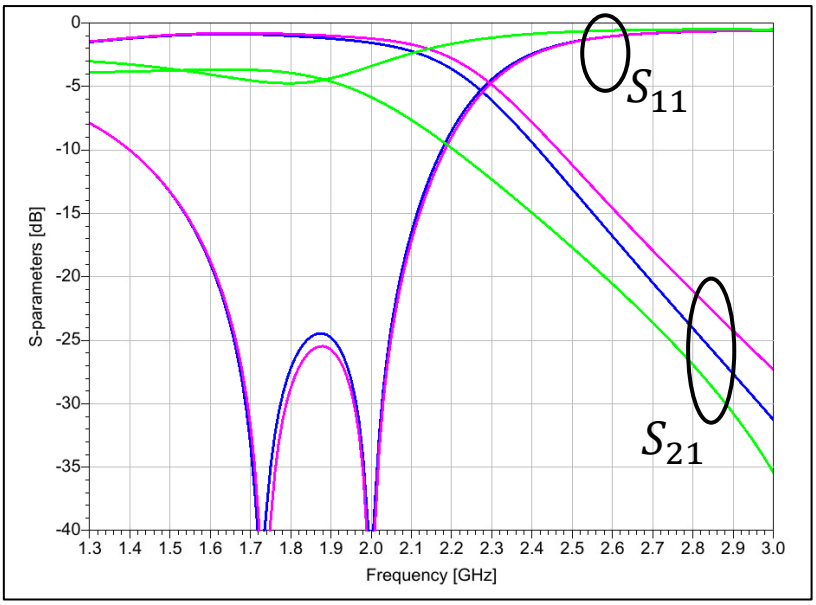



Figure 53: OMN S-parameters for different component values: Optimum values (blue), closest values (green), best values (purple).

## 5.4 Input matching network

### *5.4.1 Balun design calculations*

According to the design specifications, the stage before the PA is a modulator with differential output. So, a balun is required to couple this differential output with the PA single-ended input. This balun needs to provide a load impedance $R_{opt,Mod}$ of 315 Ω to the modulator in order to extract maximum power without compromising the signal integrity. On one hand, if the modulator load impedance is smaller, the power delivered to the balun (and therefore to the PA) will be smaller because the modulator output stage is sinking/sourcing current ($P_{out} = 0.5\,{I_{out}}^2 R$). On the other hand, if this load impedance is larger, the voltage swing at the modulator output stage will be too large, causing its transistors to enter into triode region resulting in signal distortion.

[12] https://ds.murata.co.jp/simsurfing/index.html

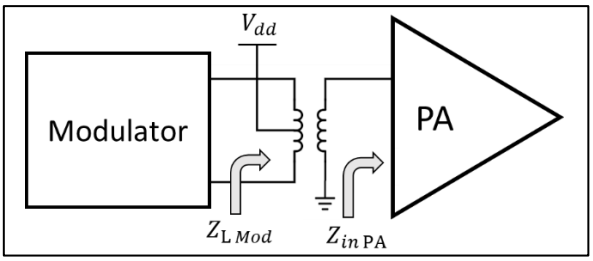



Figure 54: Impedance matching between modulator and PA.

Additionally, as described in before, the PA input impedance $Z_{in,\mathrm{PA}}$ is equivalent to a 407 Ω resistance in parallel with a 1.35 $pF$ capacitance. So, the balun needs to transform $Z_{in,\mathrm{PA}}$ into $Z_{\mathrm{L},Mod}$.

Now, a balun design has three degrees of freedom: the inductance of the primary winding $L_1$, the inductance of the secondary winding $L_2$ and the coupling factor $k$ between these two. For minimum losses, the coupling factor should be as large as possible [19] [24], which leaves only two degrees of freedom.

In order to obtain the inductance for both windings that achieve the required impedance transformation, a simplified version of the balun model previously presented in Figure 26 is used. As shown in Figure 55, in this simplified model the winding capacitances were neglected, because those mainly affect the behaviour at higher frequencies. Also, the magnetic flux leakage (due to imperfect coupling) was moved to the primary side. Moreover, the substrate losses were omitted for simplicity.

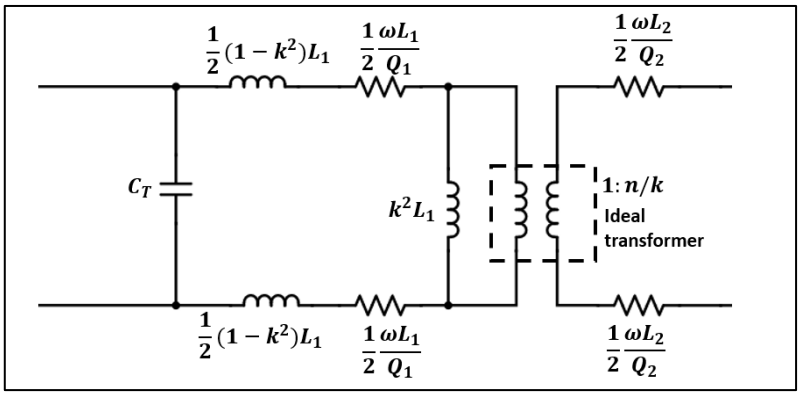



Figure 55: Simplified balun model [19].

Notice also that capacitor $C_T$ is added to the primary side with the goal of tuning the operating frequency, reducing the power loss [19], [24]. This capacitor is implemented as a switchable capacitor.

For this design, it is assumed that $k = 0.83$ and $Q_1 = Q_2 = 8$, since those were the typical values obtained after a few trials of 3D electromagnetic simulation for different balun layouts by using the metal layers available in the process node.

Although the model presented in Figure 55 is relatively simple, the mathematical expressions derived from it are significantly complex. For instance, the resistive losses of the balun ($L_{balun}$) are given by:

$$L_{balun} = \frac{R_{in,PA}\left\|\frac{1}{1+j\omega C_{in,PA}}\right\|^2}{R_{in,PA}\left\|\frac{1}{1+j\omega C_{inPA}}\right\|^2 + \frac{\omega L_2}{k^2 Q_1}\left\|1+\frac{\frac{\omega L_2}{Q_2}+\frac{R_{in,PA}}{j\omega C_{in,PA}R_{in,PA}+1}}{j\omega L_2}\right\|^2 + \frac{\omega L_2}{Q_2}} \quad (44)$$

Therefore, a computer-aided design is used instead of the mathematical model. To do this, the balun model of Figure 55 is recreated in a simulation software and the PA input impedance is connected to the balun secondary. At the primary, an RF current source is connected, modelling the modulator.

Then, the inductance of the secondary winding is determined by selecting the value that minimizes power loss at the center of the frequency band (1850 $MHz$). From Figure 56, the optimum value is 3.87 $nH$. This value is independent of $L_1$, $n$ or $C_T$ as seen in equation 44, so in the simulation any value for these two parameters can be used without affecting the optimum $L_2$ value.

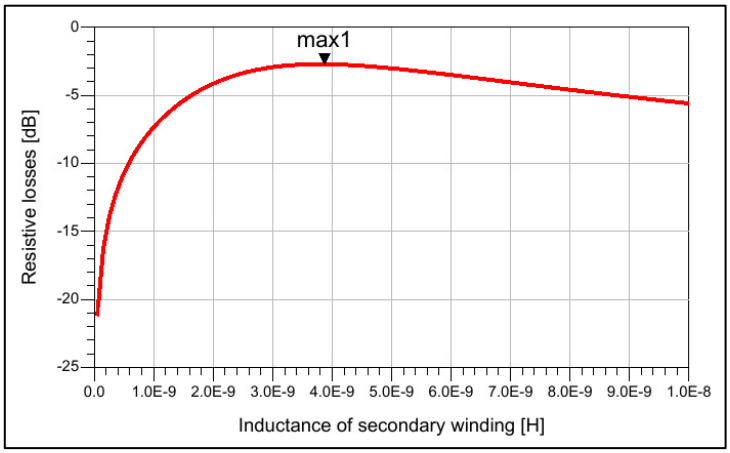



Figure 56: Resistive losses vs $L_2$.

After that, $L_1$ is found by selecting the value that generates the desired impedance transformation ratio, i.e. $L_1$ should generate an input resistance at the primary winding of 315 Ω. From Figure 57, the required inductance is 5.17 $nH$.

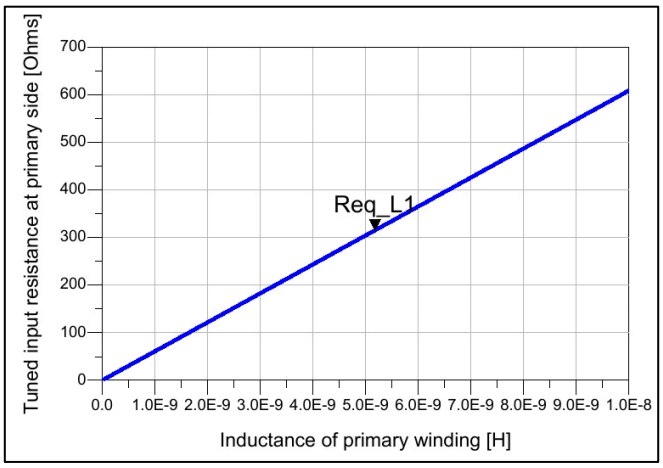
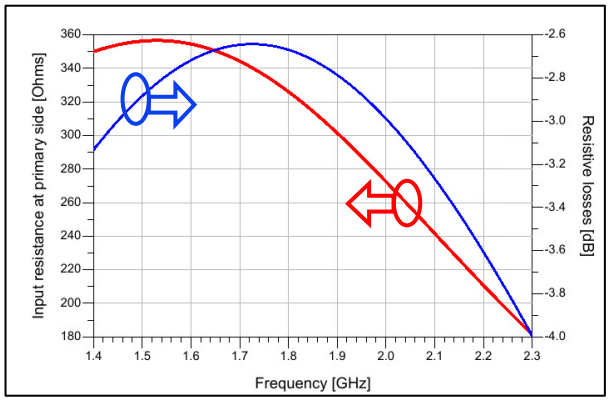
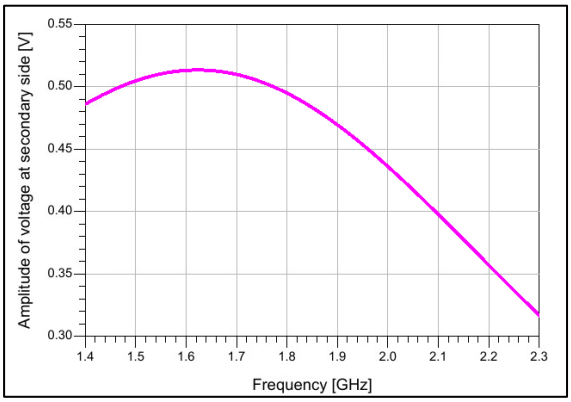



Figure 57: Input resistance at primary side vs $L_1$.

Now, the frequency response in the operating band ($1700\ MHz$ – $2000\ MHz$) is verified. Figure 58 shows the resistive losses and the input resistance seen at the primary winding. It is evident that these parameters are not centered in the operating band.



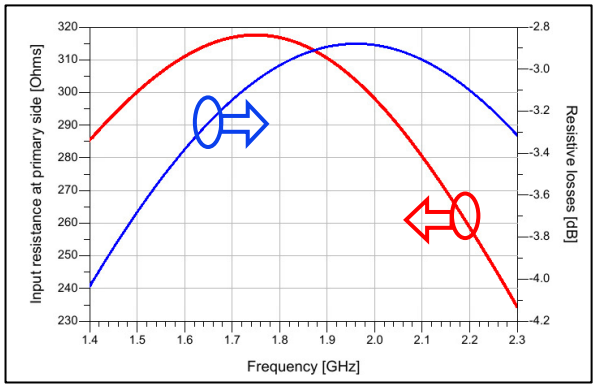

a) Input impedance at the primary and resistive losses



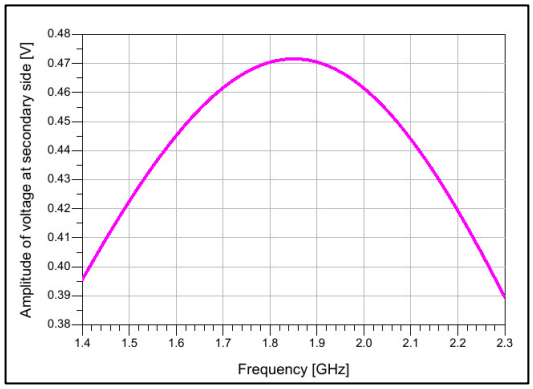

b) Amplitude of the voltage in the secondary side

Figure 58: Balun frequency response before fine tuning.

This issue can be mitigated by fine tuning the balun inductance values. Although it is not feasible to center both the resistive losses and the impedance level at the primary side, the voltage amplitude at the secondary can be (which is the relevant parameter in terms of PA performance). Figure 59 shows these parameters by using the tuned inductance values ($L_1 = 3.5\ nH$ and $L_2 = 2\ nH$). It can be observed that the voltage amplitude at the secondary is close to the one calculated in the PA basic calculations section.




a) Input impedance at the primary and resistive losses




b) Amplitude of the voltage in the secondary side

Figure 59: Balun frequency response after fine tuning.

It is important to mention that the turns ratio of this balun is $n = \sqrt{L_2/L_1} \approx 3/4$. Therefore, the layout topology to be chosen must allow a 3:4 turns ratio.

## 5.5 Gain control

According to the design specifications the power gain of the PA needs to have a control range of 12 dB with steps of less than 1 dB for high levels and up to 2 dB for low levels.

There are several approaches for varying the gain of a linear PA. The most common approach is reducing the transconductance of the PA. This is done by either lowering its bias current [25] [26] or by decreasing its size [27]. The latter option is used in this work.

The PA core is then divided into unitary cells that can be turned on and off to change the PA core effective size. Dividing the PA core into 16 equally-sized cells allows for a gain step of less than 1 dB for higher power levels as shown in Table 17.

Table 17: Power gain reduction step for 16 unitary cells

| Number of active cells | Power gain reduction [dB] | Power gain step [dB] |
|---|---|---|
| 16 | 0.00 | 0.00 |
| 15 | 0.56 | 0.56 |
| 14 | 1.16 | 0.60 |
| 13 | 1.80 | 0.64 |
| 12 | 2.50 | 0.70 |
| 11 | 3.25 | 0.76 |
| 10 | 4.08 | 0.83 |
| 9 | 5.00 | 0.92 |
| 8 | 6.02 | 1.02 |
| 7 | 7.18 | 1.16 |
| 6 | 8.52 | 1.34 |
| 5 | 10.10 | 1.58 |
| 4 | 12.04 | 1.94 |
| 3 | 14.54 | 2.50 |
| 2 | 18.06 | 3.52 |
| 1 | 24.08 | 6.02 |

To turn cells on and off, a transistor acting as a switch is placed in every cell as shown in Figure 60. This switch transistor has minimal length and its width is large enough to minimize the impact on the performance of the PA cell when the switch is on. If the switch transistor is not large enough, its on-resistance creates a feedback loop that lowers the PA gain (known as source degeneration). Also, the voltage drop across the switch adds up to the knee voltage, degrading linearity.

One important drawback of this approach is its low power efficiency. To understand why, note that the DC drain current $I_{DS}$ (and so the transconductance of the PA $g_{m,PA}$) increases proportionally with the number of active cells $N_{on}$. Then efficiency is proportional to $N_{on}$ since

$$\eta = \frac{P_{RF}}{P_{DC}} = \frac{|i_{drain}|^2 R_{L,opt}}{V_{dd} I_{DS}} = \frac{\left|v_{in,PA}\, g_{m,PA}\right|^2 R_{L,opt}}{V_{dd} I_{DS}}$$
$$= \frac{\left|v_{in,PA} N_{on} g_{m,PA}\right|^2 R_{L,opt}}{V_{dd} N_{\mathrm{on}} I_{DS,cell}}. \quad (45)$$

Therefore, reducing the number of active cells decreases the efficiency.

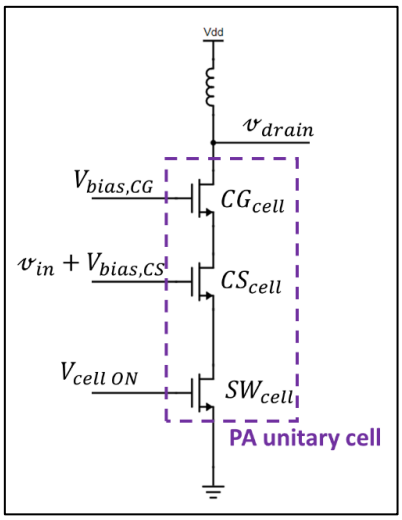


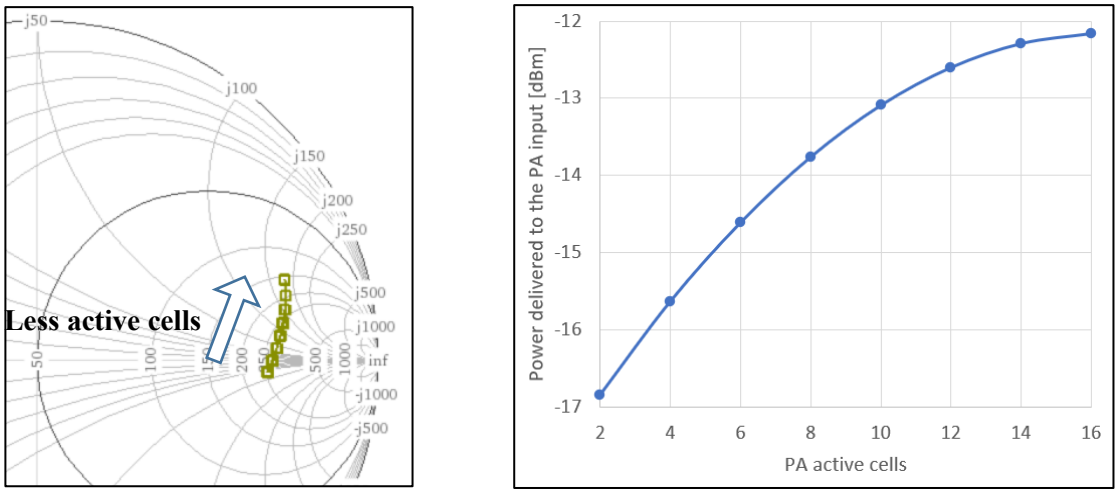

Figure 60: Schematic of a unitary cell of the PA

One important note about this gain control scheme needs to be made at this point. A deactivated cell does not have drain current and thus, the drain terminal of its CS transistor will go up to $V_{dd}$ (2.2 V for this PA). This is an undesired situation because the drain-to-gate breakdown limit of that transistor is 0.9 V. Then, to avoid this condition from happening, the drain terminal of the CS transistor of all cells are connected together. Therefore, the deactivated cells will keep the same drain voltage for the CS transistor than in the activated cells, protecting the devices from oxide breakdown.

### *5.5.1 Capacitance compensation scheme*

An issue when deactivating cells is the reduction of the input capacitance of the PA. This has a few effects on the performance of the PA. First, it degrades the input impedance matching, since the balun was designed for a specific PA input impedance. Second, it shifts the tuning frequency of the balun, which adds resistive losses [17]. Third, it changes the power delivered by the modulator since the modulator generates power by injecting its output current into a load impedance.

The variation of the input impedance of the PA is translated into a variation of the input impedance of the balun (which is the same as the load impedance of the modulator) as shown in Figure 61 (a). This variation causes a drop in the power delivered by the modulator as well as at the input of the PA. As presented in Figure 61 (b), the power at the input of the PA can drop up to 5 dB.

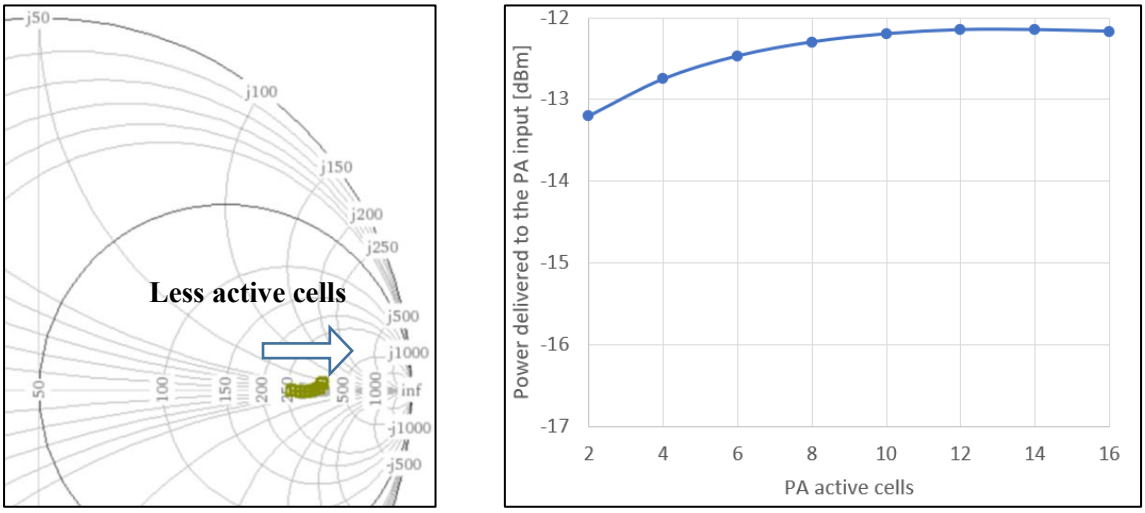



(a) Input impedance of the balun (b) Power delivered to the input of the PA

Figure 61: Effect of the number of active PA cells on the PA performance.

To compensate the lack of capacitance when deactivating cells, a switchable capacitor can be placed in parallel to the PA input. Each time one cell is deactivated, this capacitor would increase its capacitance by an amount equal to the input capacitance of one cell. However, it was found that the compensation capacitance can be added to the switchable capacitor that tunes the frequency of the balun (named $C_T$ as shown previously in Figure 55) reducing design complexity. Increasing $C_T$ by 21.5 $fF$ for each deactivated cell was found to nearly compensate the lack of capacitance at the input of the PA.

Figure 62 (a) shows the impact of this compensation scheme. It can be seen that the variation of the input impedance of the balun was drastically reduced. Moreover, the power at the PA input changes only by 1 dB instead of 5 dB for the whole range of active cells, as illustrated in Figure 62 (b).




(a) Input impedance of the balun (b) Power delivered to the input of the PA

Figure 62: Performance of the PA after capacitance compensation.

### *5.5.2 AM/AM expansion compensation scheme*

Another issue created when deactivating cells is the expansion in the AM/AM response, as shown in Figure 63. The gain expansion is higher than 1 dB when four or more cells are deactivated. It is worth mentioning that 1 dB of gain expansion is as harmful for linearity as 1 dB of gain compression.

This expansion is caused by the increase of the DC level of the drain current with higher input power (since it is a rectified sine wave). This additional DC current raises the bias operating point of the PA giving it more gain (as previously shown in Figure 32).

It is worth mentioning that this effect is less noticeable when there are more active cells because gain compression occurs earlier than expansion in this case.

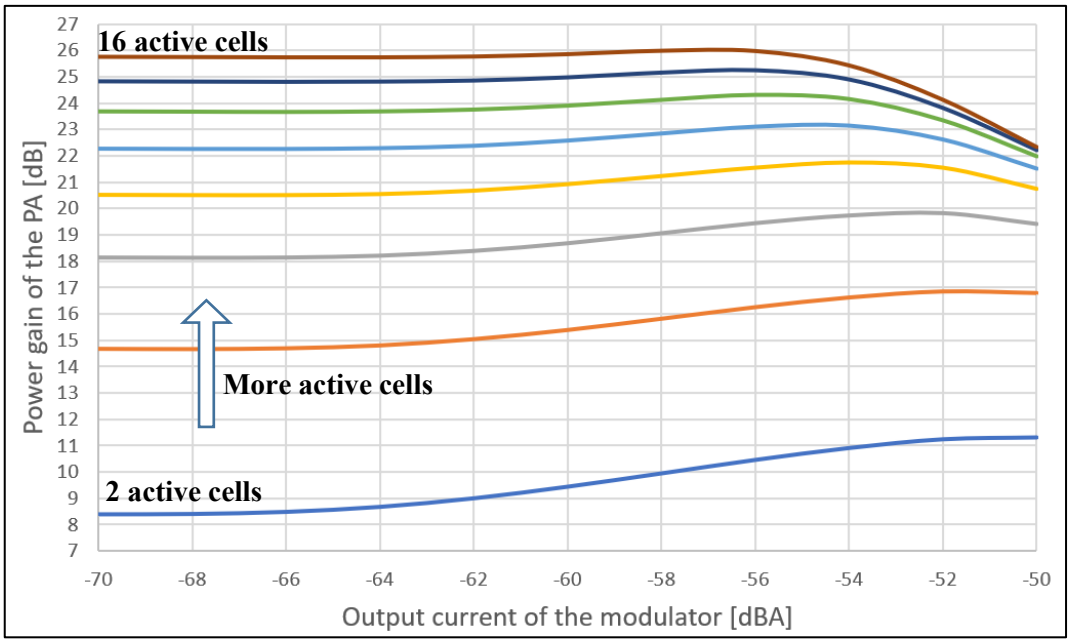



Figure 63: AM/AM curves with respect to the number of active cells.

To mitigate this behavior, a possible solution is illustrated in Figure 64. An additional bias current $I_{AM/AM}$ is injected into the "PA sample" located in the bias circuit. The additional bias current is increased with the number of deactivated cells, moving the PA closer into class-A, thus improving its linearity.

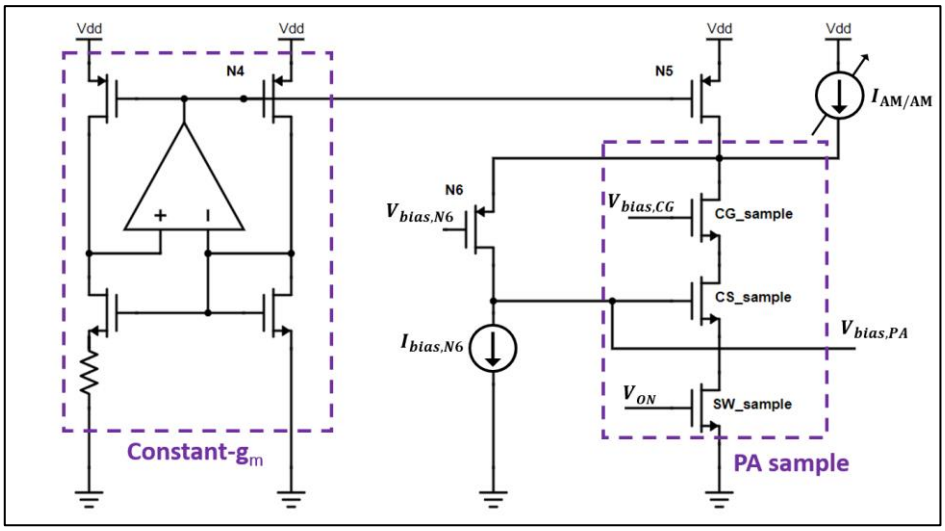



Figure 64: Compensation scheme for the expansive AM/AM behavior.

The AM/AM response after applying the proposed compensation scheme is shown in Figure 65. It can be seen that after compensation, the expansive behavior is mitigated. Furthermore, the EVM before and after this compensation scheme is shown in Figure 66, demonstrating the improvement in the linearity of the PA.

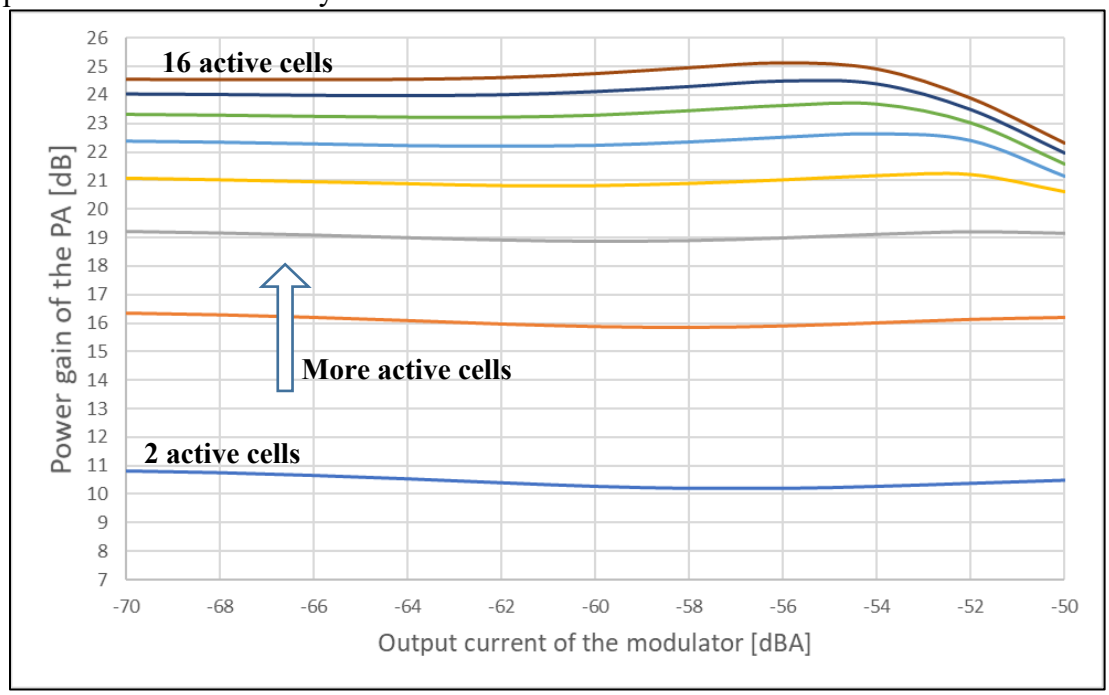



Figure 65: Compensated AM/AM response.

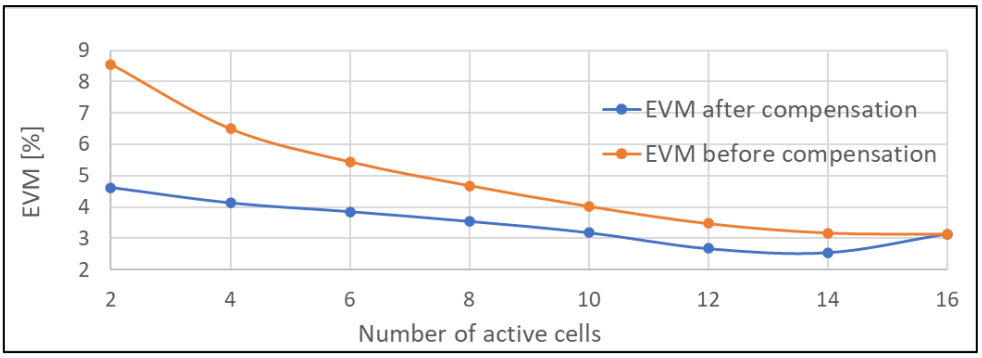



Figure 66: EVM for different number of active cells using the linearity compensation scheme.

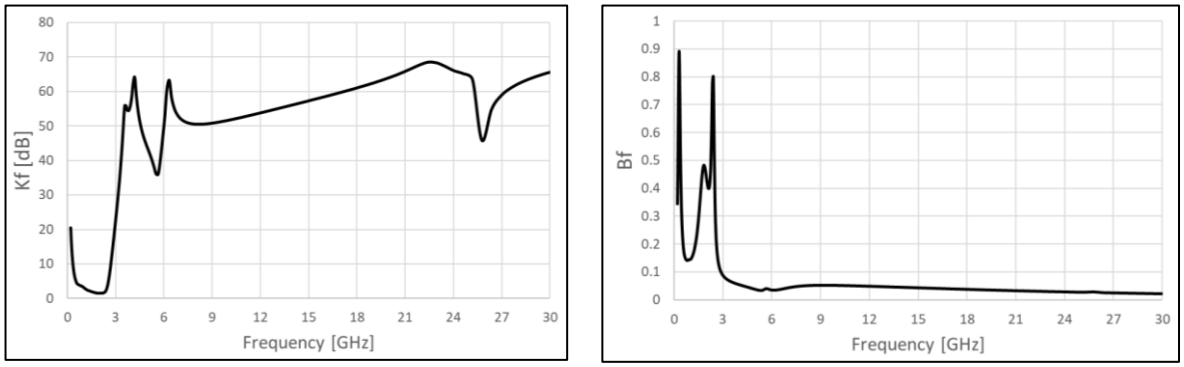
## 5.6 Stability

The PA must be stable for all load impedances with a VSWR of 5:1 according to the design specifications. There are a few methods for verifying the stability of an RF circuit. One of them, which is readily available in the circuit simulator tool used in this work, is based on the S-parameters of the PA. This method was created by Gonzalez in [28], who mathematically proved that unconditional stability is ensured if the following two conditions are both satisfied:

$$K_f = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |S_{11}S_{22} - S_{12}S_{21}|^2}{2|S_{12}S_{21}|} > 1$$
$$B_{1f} = 1 + |S_{11}|^2 - |S_{22}|^2 - |S_{11}S_{22} - S_{12}S_{21}|^2 > 0. \quad (46)$$

That is, if these two conditions are satisfied for all frequencies, then the PA is stable for all load and source impedances.

These two conditions were evaluated for the designed PA by sweeping the frequency up to 30 GHz. The $K_f$ and $B_{1f}$ curves are shown in Figure 67. For clarity, the $K_f$ was plotted in logarithmic scale since its values vary dramatically (from 1 to $10^7$ approximately); this means that the condition $K_f$>1 becomes 10log($K_f$)>0.




(a) $K_f$ in logarithmic scale (b) $B_{1f}$

Figure 67: Frequency sweep for $K_f$ and $B_{1f}$ factors.

As seen in Figure 67, 10log($K_f$)>0 and $B_{1f}$>0, meaning that the PA is unconditionally stable.

## 5.7 Final adjustments

A few key adjustments were made to the PA for achieving the linearity requirements. These adjustments are described next.

When running the first few simulations with a modulated input signal, it was found that the obtained EVM (14.4%) far exceeded the target EVM (3.4%). The constellation diagram showing the unmodulated symbols at the antenna port is shown in Figure 68. As an interesting note, these symbols have a minor phase rotation. This additional phase is caused by the time delay introduced by the PA. If the four symbols $(\pm 1/\sqrt{2}\,;\ \pm 1/\sqrt{2})$ were used as reference for the EVM calculation, the additional phase would add a floor EVM. Then, to have a meaningful result, the EVM in this work is calculated by using as reference the rotated version of the symbols $(\pm 1/\sqrt{2}\,;\ \pm 1/\sqrt{2})$ that results in the minimum EVM. These symbols are marked in red color in the constellation diagrams presented in this work.

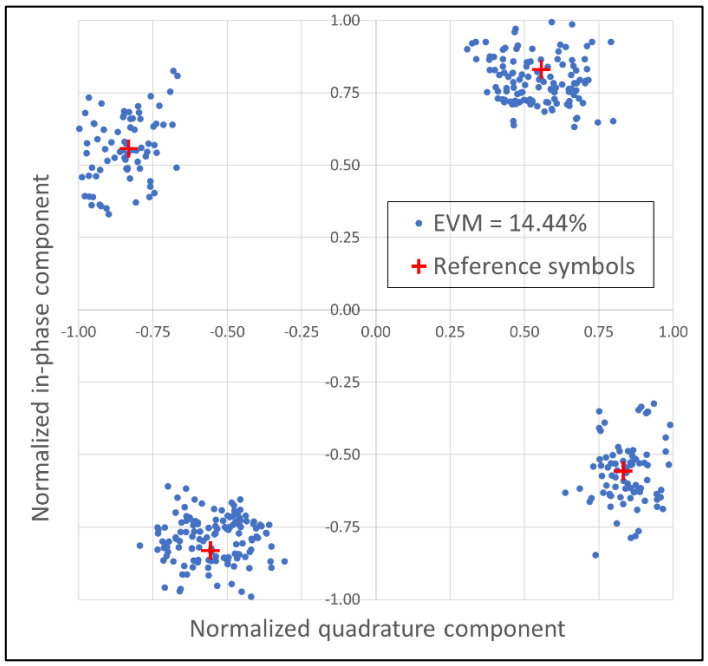
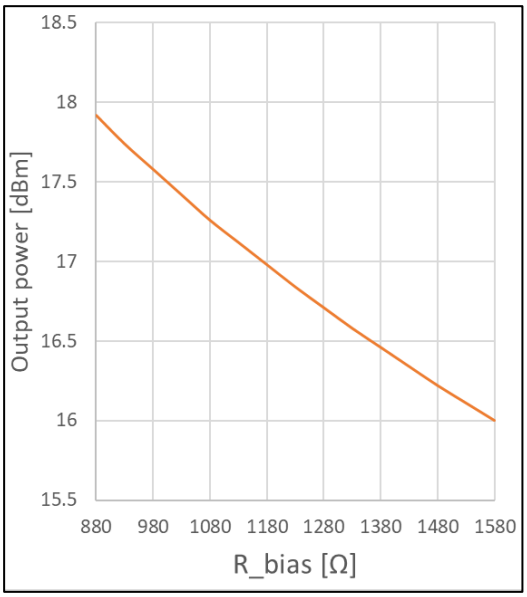



Figure 68: Constellation diagram at the antenna port.

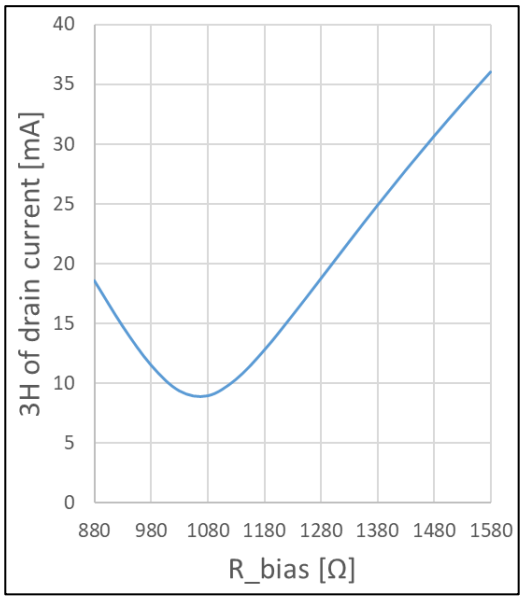
One of the main causes for the high EVM presented in Figure 68 is that the gain of the PA was too large. It was observed that the RMS output power at the antenna port was $+18$ dBm, which is 2 dB over the target of $+16$ dBm. This additional power reduces the back-off needed to amplify without much compression the peaks of the envelope of the modulated signal. Therefore, a gain reduction is needed.

Lowering the gain is possible by reducing the bias current of the PA. This has the additional benefit of slightly improving the efficiency since less DC power is consumed. One of the parameters that directly controls the bias current of the PA is the resistance $R_{bias}$ in the constant-$g_m$ cell of the bias circuit. The variation of the output power of the PA versus $R_{bias}$ is presented in Figure 69(a). So, $R_{bias}$ needs to be increased to 1580 Ω from its initial value of 880 Ω to obtain the target output power. However, this has the side effect of increasing by 75% the amplitude of the third harmonic of the drain current, as seen in Figure 69(b). A larger third harmonic in the drain current directly translates into a larger third harmonic in the drain voltage,

which degrades linearity[13] (similar to the effect caused by its second harmonic as explained in the OMN section).




(a) Output power at the RMS level of the modulation




(b) Third harmonic of the drain current

Figure 69: Effect of sweeping $R_{bias}$ on the PA performance.

Fortunately, there is a minimum in the third harmonic of the drain current for $R_{bias} = 1070\ \Omega$. Although, this value of point $R_{bias}$ only reduces the output power by about 1 dB, it is selected because of this additional benefit. As an interesting note, the same minimum point on the third harmonic of the drain current can be seen in Figure 33, as previously shown.

Another way for reducing the power gain of the PA is lowering the feedback resistance $R_{fb}$, since it controls the voltage amplitude at the input of the PA core. The variation of the output power versus the feedback resistance is illustrated in Figure 70. So, $R_{fb}$ needs to be reduced to 1900 Ω from its initial value of 2800 Ω to obtain the target output power. This adjustment has two additional benefits. First, it can be easily reverted if additional gain in required after designing the layout. Second, it reduces the balun losses by 0.5 dB.

[13] Actually, the effect of the third harmonic of the drain voltage on the PA linearity depends on its phase. For some phases, it can improve linearity.

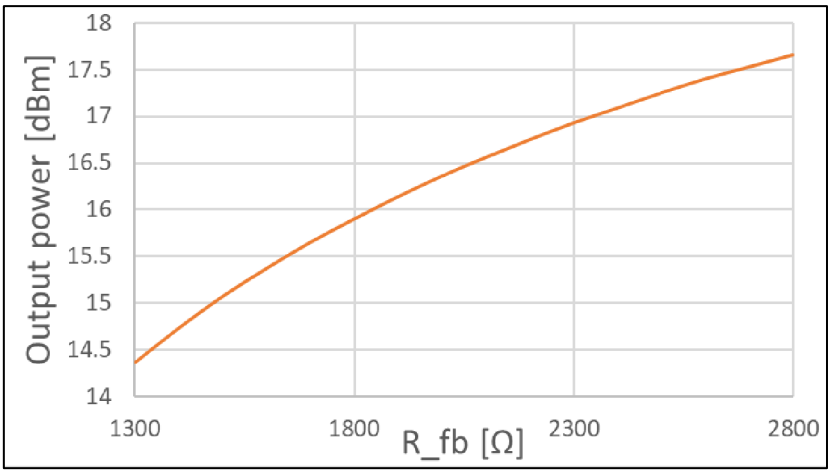



Figure 70: Effect of the feedback resistance on the output power.

With the target output power achieved, the EVM is 8.3%. Although this was a considerable improvement, is still about two times higher than the target EVM.

To further improve the linearity of the PA, the DC drain-to-source voltage of the common-source transistor $V_{DS,CS}$ is tuned. This voltage is particularly important for linearity because if it is too low, the CS transistor will enter into triode region too early (see Figure 71) which reduces its transconductance producing gain compression. If it is too high, the CG transistor will enter into triode region too early reducing its transconductance, which increases the load impedance seen by the CS transistor, hence increasing $v_{drain,CS}$ and pushing CS into triode region. Modifying $V_{DS,CS}$ is possible by changing the bias voltage of the CG transistor ($V_{bias,CG}$).

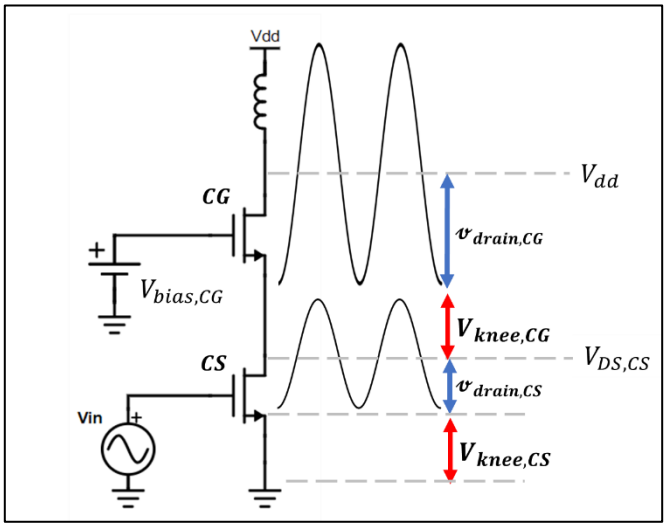



Figure 71: Reference diagram for the effect of $V_{DS,CS}$ on the PA performance.

Sweeping $V_{DS,CS}$ has a noticeable effect on the EVM, as shown in Figure 72. The minimum EVM is 2.94% and it is achieved by using a $V_{DS,CS} = 650\ mV$. The constellation diagram for this case is shown in Figure 73.

In conclusion, the above mentioned adjustments improved the EVM approximately from 14% to 3%.

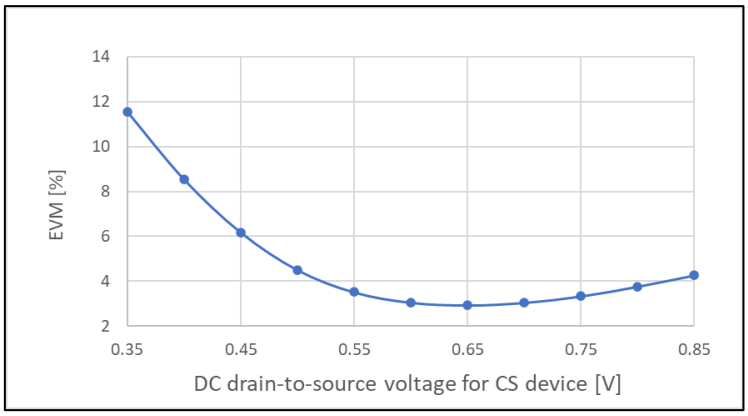



Figure 72: Effect of $V_{DS,CS}$ in the PA linearity.

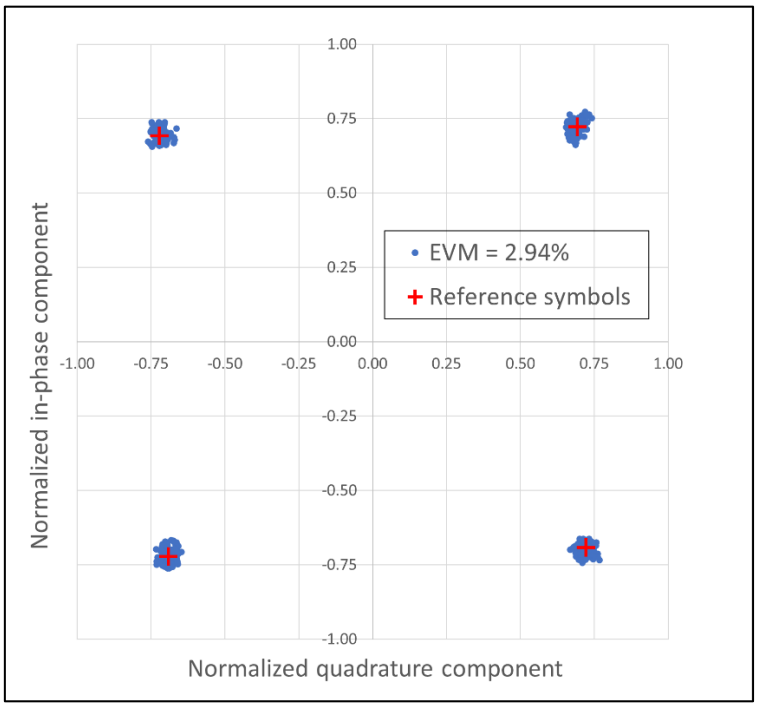



Figure 73: Constellation diagram for the adjusted PA.

# 6 RESULTS

## 6.1 Frequency response and output power

The output power in the operating frequency band for process and temperature variation is shown in Figure 74. The difference in output power across corners is about 1 dB.

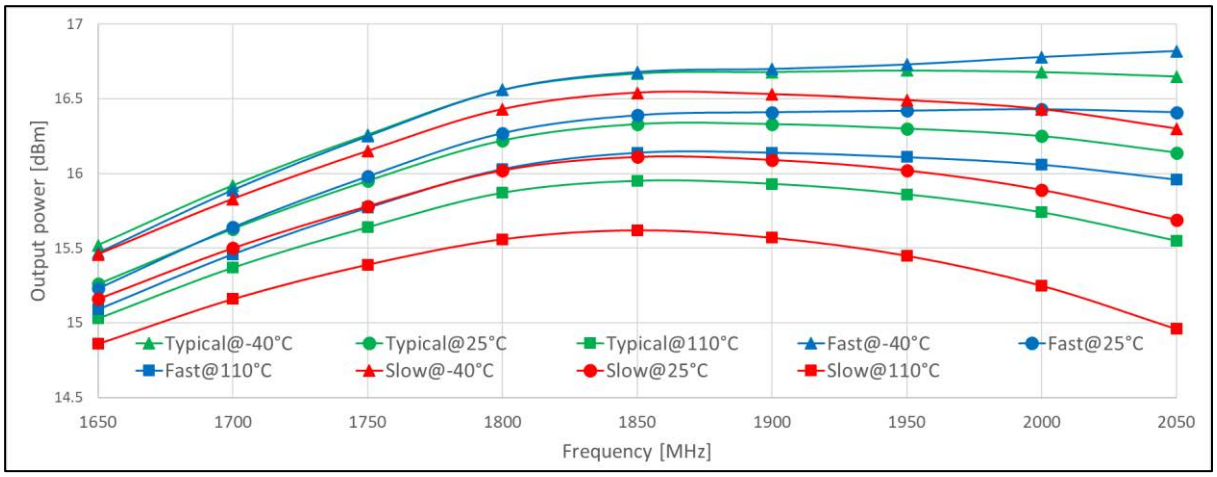



Figure 74: Output power vs frequency for process and temperature variation.

## 6.2 Linearity

To assess the linearity of the PA, the EVM was obtained from simulations for 50 frames of the modulated signal, from different modulation profiles and across process and temperature corners. In order to do a meaningful EVM benchmark, the output power of the PA was set to maximum (+16 dBm), since using lower power improves linearity. The EVM results are registered in Table 18 with a color-coded marking for clarity. Values complying with the design specifications are marked in green, values exceeding the specification by less than 10% are marked in yellow and values exceeding the specification by more than 10% are marked in red.

The specification is not satisfied for the 12-tone and 6-tone modulation profiles in the slow process corner only. The reason for this is that the higher threshold voltage of the CG transistor in this corner increases the total knee voltage of the PA, since the bias compensation scheme was created for the CS transistor only. This effect is present in modulation profiles with higher PAPR since those profiles drive the PA into compression more than the low PAPR ones.

Moreover, for the 1-tone modulation profile the EVM was exactly zero for all temperatures and process corners because this modulation has constant envelope (and therefore its PAPR is $0\ dB$) completely avoiding the compressive behavior of the PA.

Regarding the out of band emissions of the PA, the output power spectral density (PSD) for all the modulation profiles, as well as the 3GPP standard spectrum masks, is presented in Figure 75. To test the worst case, two actions were taken. First, the output power was set to maximum. Second, in the cases where the signal bandwidth is lower than 180 kHz, the spectrum of the signal was positioned near the edge of the channel (instead of in the center of the channel) moving the out of band emission closer to the spectrum masks.

Table 18: EVM across modulation profiles, temperature and process corners

| | | **EVM [%]** (Spec = 3.4%) | | |
|---|---|---|---|---|
| | | **Temperature** | | |
| **Modulation profile** | **Process corner** | **-40°C** | **25°C** | **110°C** |
| **12 tones** | **Slow** | 3.26 | 3.64 | 5.25 |
| | **Typical** | 2.91 | 2.5 | 3.48 |
| | **Fast** | 3.49 | 2.38 | 2.13 |
| **6 tones** | **Slow** | 2.98 | 3.8 | 5.3 |
| | **Typical** | 1.86 | 2.24 | 3.5 |
| | **Fast** | 2.03 | 1.41 | 1.95 |
| **3 tones** | **Slow** | 1.7 | 1.53 | 1.76 |
| | **Typical** | 1.47 | 1.05 | 0.97 |
| | **Fast** | 1.7 | 1.05 | 0.34 |
| **1 tone** (*both bandwidths*) | **Slow** | 0 | 0 | 0 |
| | **Typical** | 0 | 0 | 0 |
| | **Fast** | 0 | 0 | 0 |

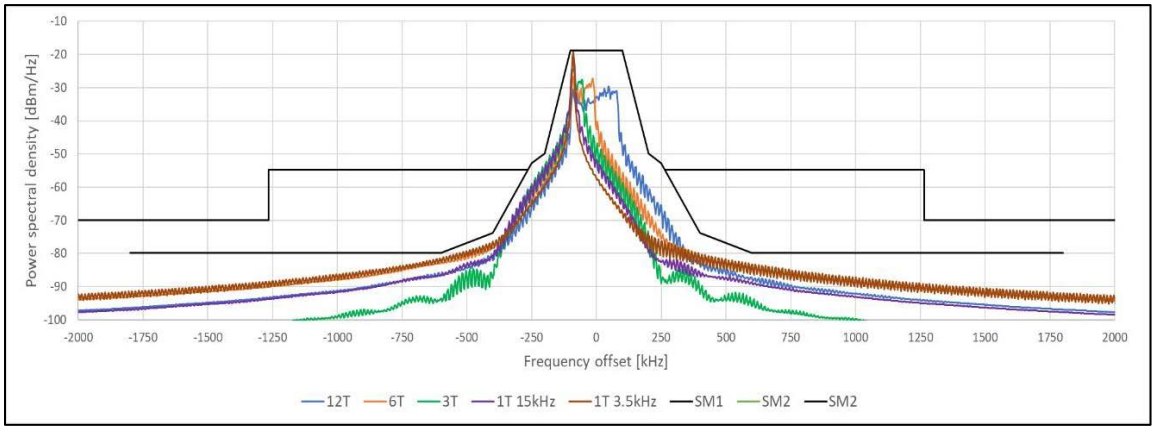



Figure 75: Spectrum masks and PSD for all modulation profiles.

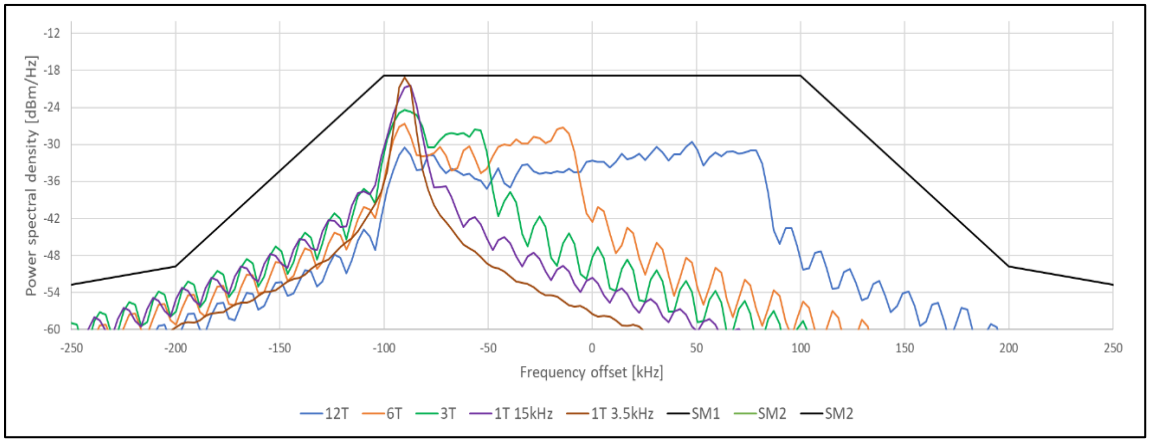



Figure 76: Close up of PSD for all modulation profiles.

It can be seen that the spectrum masks are satisfied in all cases. It is important to mention that the peak PSD is higher for modulation profiles with lower bandwidth since the transmitted signal power was set to +16 dBm in all cases, resulting in higher power per unit frequency for the low bandwidth profiles. For instance, the peak PSD for the 1-tone modulation profile with 3.5 kHz bandwidth is $+16\ \text{dBm} - 10\log(3.5\ \text{kHz}) = -19.44\ \text{dBm/Hz}$ which is 1 dB less than the spectrum mask limit, as seen in Figure 76.

Regarding the ACLR, the results are presented in Table 19. The values are color-coded for clarity. A green highlight indicates compliance with the design specification, yellow highlight is used when the specification is exceeded by less than 1 dB and red is used otherwise.

It can be observed that only in 4 cases the ACLR was exceeded by less than 1 dB. However, there are two crucial factors in this test that need to be addressed. First, the ACLR in the design specifications has a 10 dB margin over the 3GPP standard specifications, resulting in a more challenging case. Second, the modulated signal source available in the simulation software did not satisfy the design specifications, as seen in Table 20[14]. This table also contains in parentheses the difference between the output ACLR and the input ACLR, which was at most 1 dB[15]. These delta values demonstrate the low out of band emissions of the designed PA.

Table 19: ACLR of the output signal for all modulation profiles

| **Modulation profile** | **ACLR1 Upper Channel [dBc]** (Spec=-30 dBc) | **ACLR1 Lower Channel [dBc]** (Spec=-30 dBc) | **ACLR2 Upper Channel [dBc]** (Spec=-47 dBc) | **ACLR2 Lower Channel [dBc]** (Spec=-47 dBc) |
|---|---|---|---|---|
| 12 tones | -37.31 | -35.41 | -50.77 | -49.93 |
| 6 tones | -43.09 | -31.94 | -47.26 | -46.47 |
| 3 tones | -52.42 | -30.29 | -58.76 | -57.5 |
| 1 tone 15kHz | -50.41 | -30.58 | -51.7 | -50.32 |
| 1 tone 3.5kHz | -46.39 | -34.8 | -47.24 | -46.11 |

Table 20: ACLR for the input signal

| **Modulation profile** | **ACLR1 Upper Channel [dBc] (Delta)** | **ACLR1 Lower Channel [dBc] (Delta)** | **ACLR2 Upper Channel [dBc] (Delta)** | **ACLR2 Lower Channel [dBc] (Delta)** |
|---|---|---|---|---|
| 12 tones | -38.11 (0.8) | -35.98 (0.57) | -51.02 (0.25) | -50.16 (0.23) |
| 6 tones | -44.04 (0.95) | -32.19 (0.25) | -47.55 (0.29) | -46.79 (0.32) |
| 3 tones | -52.44 (0.02) | -30.54 (0.25) | -58.53 (-0.23) | -57.23 (-0.27) |
| 1 tone 15kHz | -50.27 (-0.14) | -30.67 (0.09) | -51.61 (-0.09) | -50.21 (-0.11) |
| 1 tone 3.5kHz | -46.43 (0.04) | -34.97 (0.17) | -47.27 (0.03) | -46.13 (0.02) |

For illustration purposes, the PSD of the input signal (1-tone 3.5 kHz-bandwidth modulated signal compliant with the 3GPP standard according to the simulator software manual) and the signal produced by the PA, are presented in Figure 77. Since there is a large difference in power between these signals, the input PSD is presented using a shifted vertical axis with the same decibels per division. Both signals have essentially the same PSD shape.

[14] These values are already improved by using a 6th-order Butterworth baseband filter.

[15] Negative numbers indicate an output ACLR better than the input ACLR, which can be explained by numerical noise in the data.

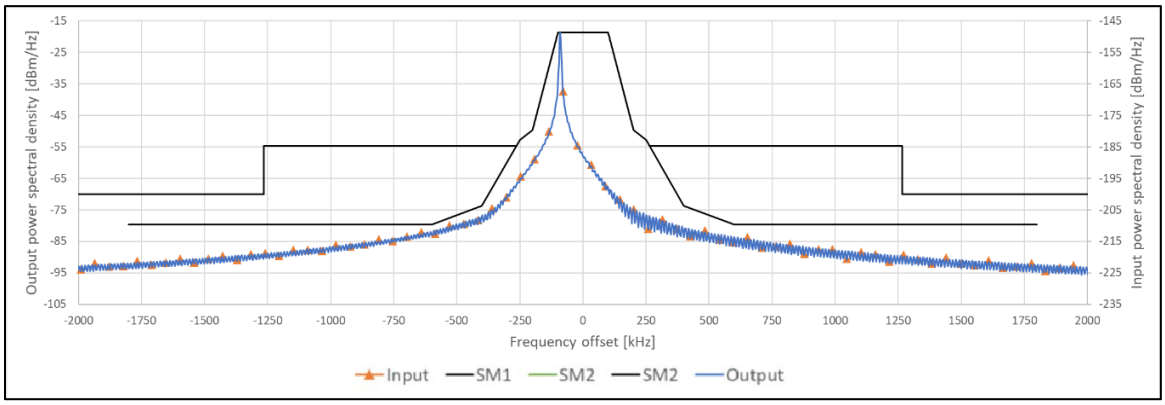



Figure 77: Power spectral density of the input and the output signal.

### 6.3 Gain programmability

The average output power for different number of active PA cells is presented in Table 21. Also, the gain reduction step is presented, which was calculated as the ratio of the output power using $N$ cells and the output power using $N-1$ cells. It can be seen that the gain reduction step is less than 1 dB for higher power levels and less than 2 dB for lower power levels, complying with the design specifications.

Table 21: Average output power for each gain reduction setting

| **Number of active cells** | **Output power [dBm]** | **Gain reduction step [dB]** |
|---|---|---|
| 16 | 16.37 | 0 |
| 15 | 16.08 | 0.29 |
| 14 | 15.75 | 0.33 |
| 13 | 15.39 | 0.36 |
| 12 | 14.97 | 0.42 |
| 11 | 14.5 | 0.47 |
| 10 | 13.96 | 0.54 |
| 9 | 13.32 | 0.64 |
| 8 | 12.55 | 0.77 |
| 7 | 11.64 | 0.91 |
| 6 | 10.51 | 1.13 |
| 5 | 9.11 | 1.40 |
| 4 | 7.30 | 1.80 |
| 3 | 4.85 | 2.44 |
| 2 | 1.25 | 3.60 |

Regarding linearity, the power spectral density for different number of active cells is shown in Figure 78, as well as the spectrum masks of the 3GPP standard. In this plot, the PA was driven with the maximum output power from the modulator. It can be observed that both spectrum masks are satisfied.

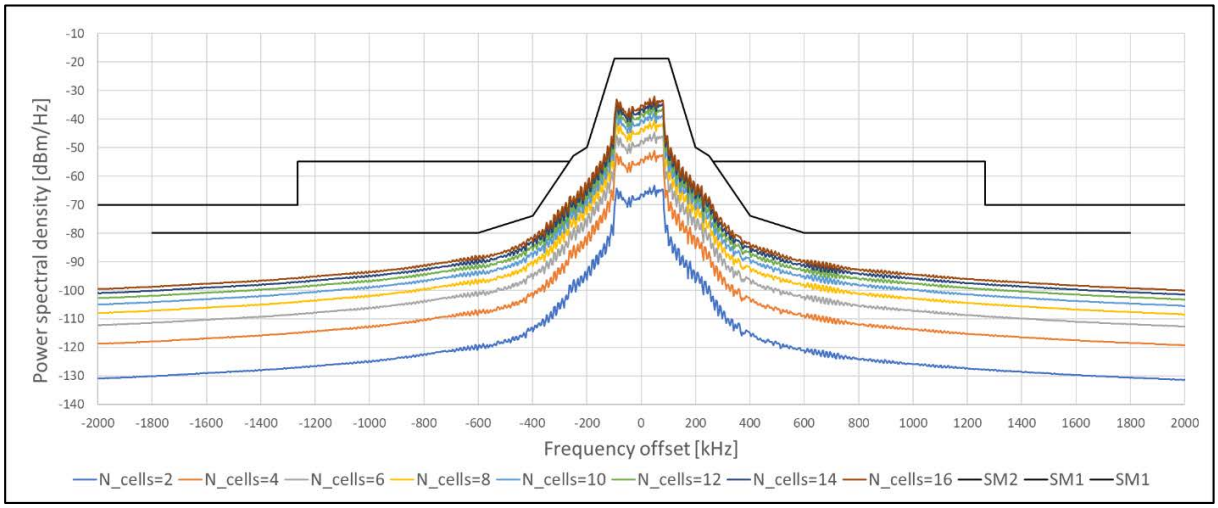



Figure 78: Power spectral density at the antenna port for different number of active cells.

Furthermore, the ACLR for different number of active cells is given in Table 22. The ACLR1 was satisfied with a margin of 6 dB while the ACLR2 the margin was 3 dB.

Table 22: ACLR vs number of active cells

| **Number of active cells** | **ACLR1 Upper Channel [dBc]** (Spec > -30 dBc) | **ACLR1 Lower Channel [dBc]** (Spec > -30 dBc) | **ACLR2 Upper Channel [dBc]** (Spec > -47 dBc) | **ACLR2 Lower Channel [dBc]** (Spec > -47 dBc) |
|---|---|---|---|---|
| 2 | -36.44 | -35.07 | -50.7 | -49.85 |
| 4 | -36.53 | -35.19 | -50.7 | -49.86 |
| 6 | -36.52 | -35.28 | -50.68 | -49.84 |
| 8 | -36.49 | -35.35 | -50.67 | -49.84 |
| 10 | -36.44 | -35.4 | -50.66 | -49.84 |
| 12 | -36.39 | -35.44 | -50.66 | -49.84 |
| 14 | -36.34 | -35.47 | -50.65 | -49.83 |
| 16 | -36.3 | -35.46 | -50.65 | -49.83 |

The EVM variation with the number of active cells is illustrated in Figure 79. The 3.4% target EVM is not satisfied when less than 9 cells are activated. This can be explained by the fact that the real part of the input impedance of the PA grows with the number of deactivated cells, increasing the amplitude of the input voltage of the PA (which could be previously seen in Figure 62). This higher drive strength can trigger non-linearity in the PA, reducing the EVM.

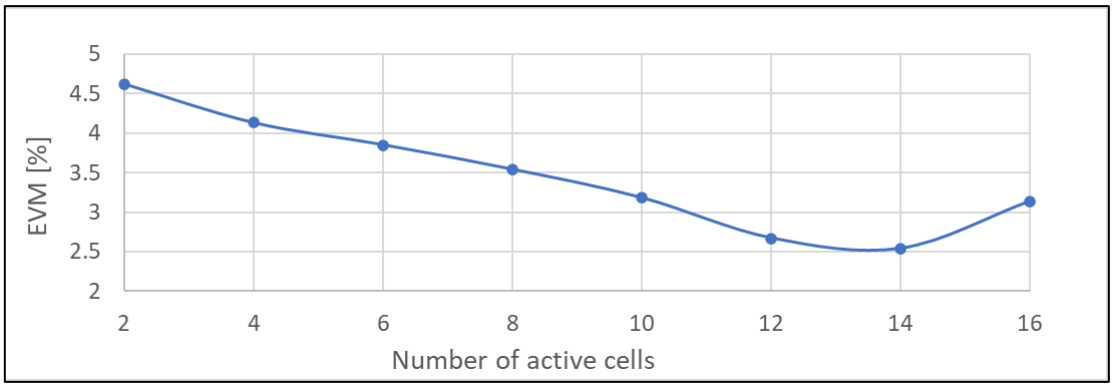



Figure 79: EVM for different number of active cells.

It is important to note that for these linearity tests the 12-tone modulation was used.

## 6.4 Efficiency

The power added efficiency for different number of active cells is shown in Figure 80. As proved previously, the efficiency decreases linearly with the number of deactivated cells. However, it can be seen from Figure 80 that the efficiency also decreases with the input drive. Then, an interesting question would be: what should be reduced for achieving a given output power in the most efficient way, the number of active cells or the input drive strength? By observing Figure 81, it can be seen that reducing active cells is more efficient than reducing the input drive for a given output power.

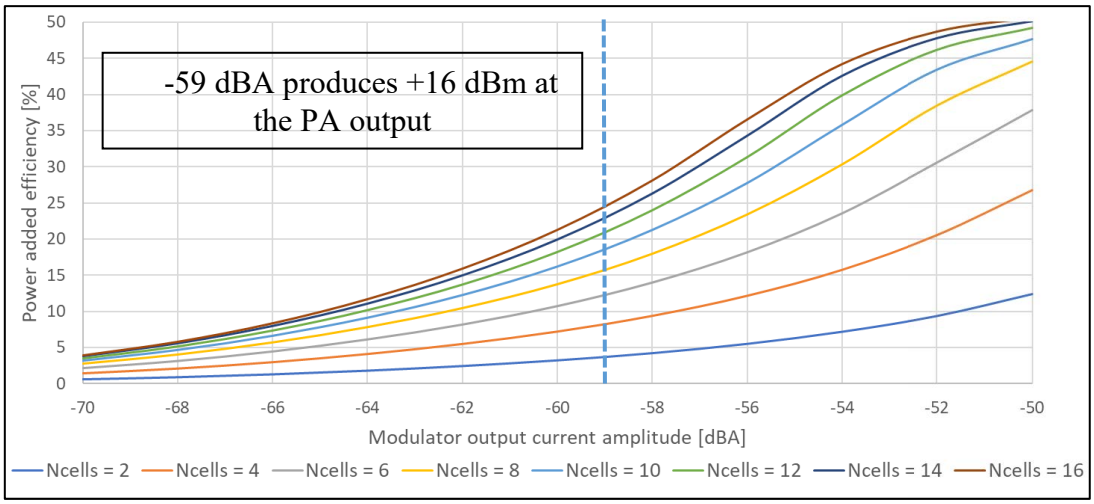



Figure 80: Power added efficiency for different number of cells.

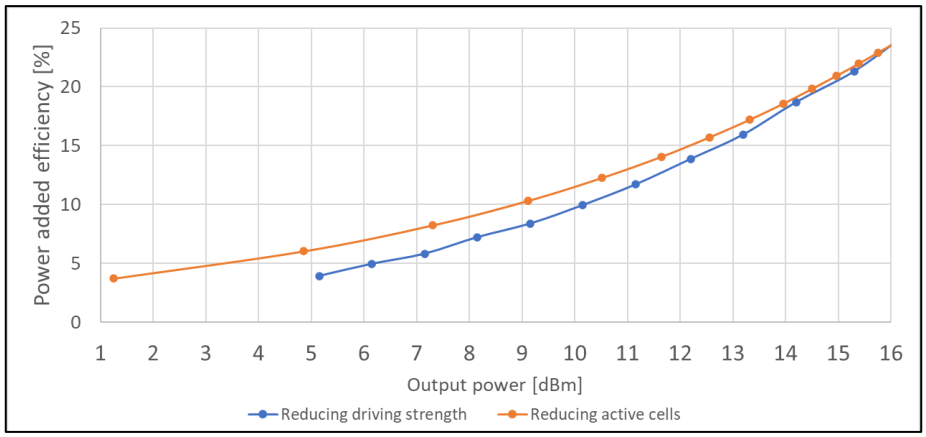



Figure 81: Power added efficiency for two ways of reducing output power.

## 6.5 Transmit intermodulation

The transmit intermodulation test is performed by injecting an external signal into the antenna port with a power 40 dB lower than the transmitted signal, and then measuring the intermodulation products created by the mixing effect of the non-linear components of the PA. This test is performed with a frequency offset between the external signal and the transmitted signal of 180 kHz and 360 kHz. A diagram of the power spectrum at the antenna port showing the transmitted signal, the external signal and the intermodulation products is shown in Figure 82. This diagram shows the power spectrum for continuous wave transmission at the maximum average power (+16 dBm) in orange trace, and at peak power (+22 dBm) in purple trace. The intermodulation products are larger when the transmitted power is +22 dBm than +16 dBm because of the compressive behavior of the PA at peak power.

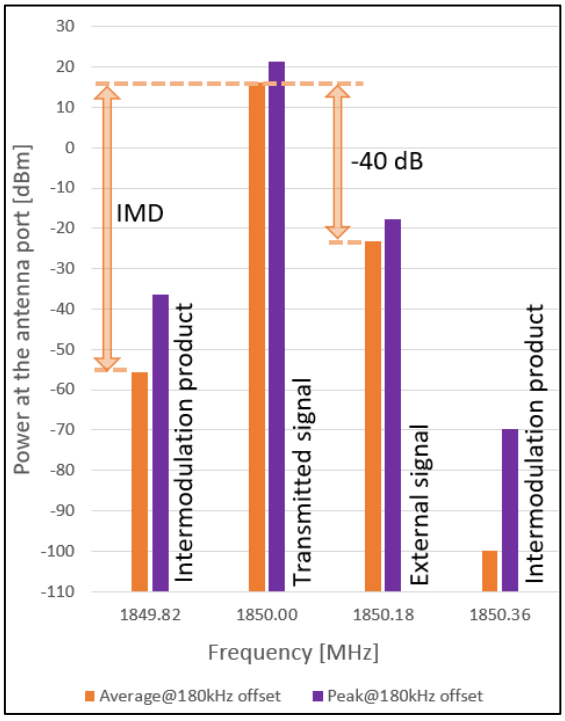



Figure 82: Intermodulation products at the antenna port for average and peak output power.

The figure of merit for this test is the ratio of the power of the transmitted signal to the power of the intermodulation product. To make this test meaningful, the intermodulation product with largest power is used (the one at the lower frequency in Figure 82). Also, the peak power of the transmitted signal used. The results for this test for the process and temperature corners is presented in Table 23. The design specification is satisfied for all test points.

Table 23: Transmit intermodulation results across corners

| | **Transmit intermodulation [dBc]** | | | | | |
|---|---|---|---|---|---|---|
| | **Frequency offset = 180 kHz** (Spec = -20 dBc) | | | **Frequency offset = 360 kHz** (Spec = -39 dBc) | | |
| **Temperature [°C]** | **Fast corner** | **Slow corner** | **Typical** | **Fast corner** | **Slow corner** | **Typical** |
| **-40** | -55.17 | -54.15 | -54.46 | -55.17 | -54.14 | -54.46 |
| **25** | -56.55 | -56.51 | -56.23 | -56.53 | -56.5 | -56.22 |
| **110** | -59.85 | -63.63 | -61.07 | -59.83 | -63.61 | -61.05 |

## 6.6 Ruggedness

Testing the ruggedness of the PA is done here by verifying whether the RMS voltage between each pair of terminals for each transistor exceeds the breakdown voltages when sweeping the VSWR of the antenna from 1:1 to 8:1 while using the peak power of the PA.

Each VSWR value is mapped to a circle in the Smith chart with center corresponding to the center of that chart and radius $|\rho|$ given by $\text{VSWR} = (1 + |\rho|)/(1 - |\rho|)$ (here $\rho$ is the reflection coefficient of the antenna). This means that testing a single VSWR value involves testing all the points in a circle in the Smith chart.

To simplify this test, $|\rho|$ is swept from 0.178 to 0.778 (corresponding to a VSWR between 1.4:1 and 8:1) in steps of 0.1 and each circle is sampled every 30°, for a total of 84 testing points.

Thus, the RMS voltage between each pair of terminals for both transistors and the corresponding breakdown voltage in presented in Figure 83, Figure 84 and Figure 85.

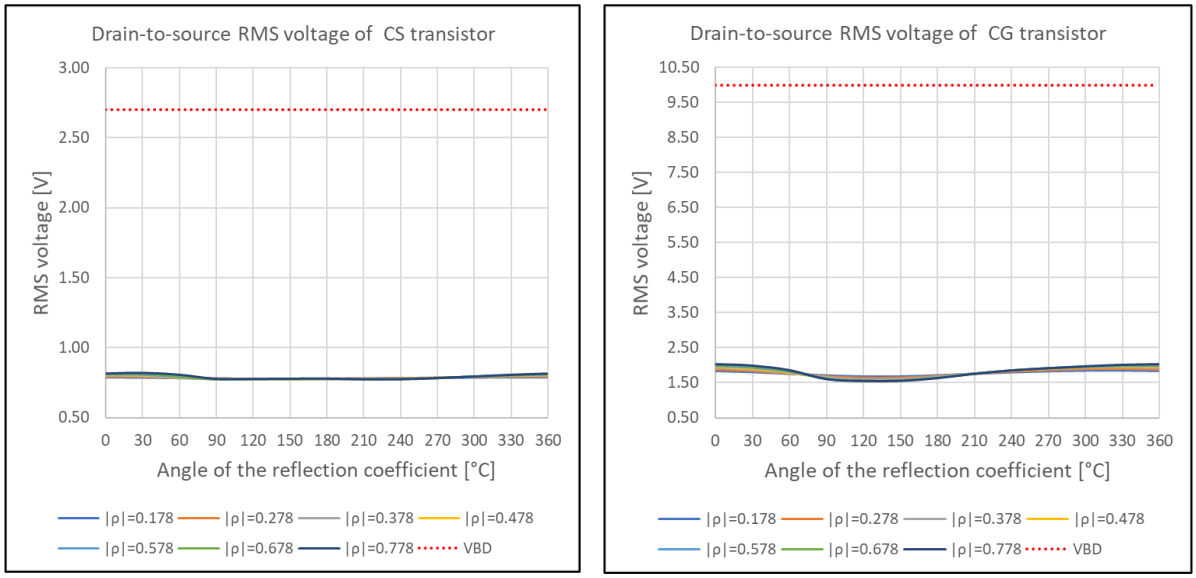



Figure 83: Drain-to-source RMS voltage for VSWR lower than 8:1.

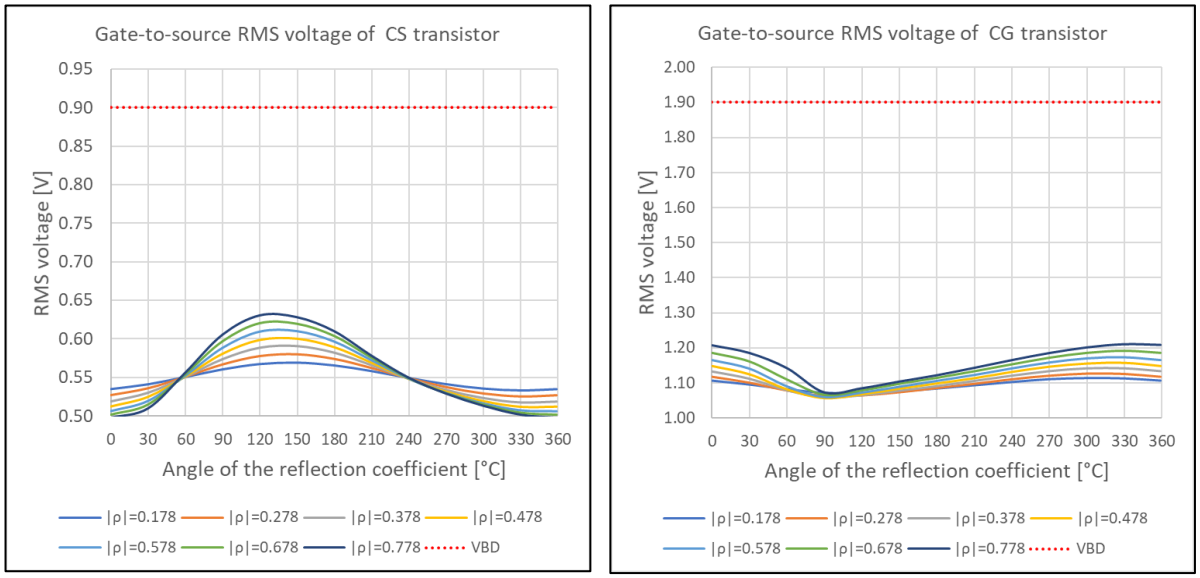



Figure 84: Gate-to-source RMS voltage for VSWR lower than 8:1.

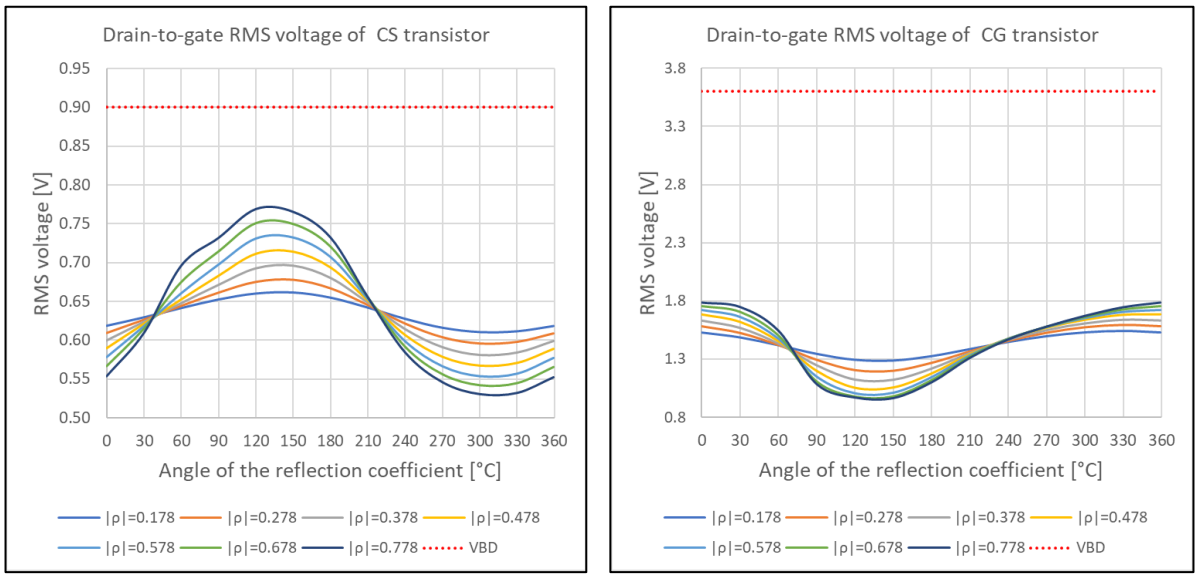



Figure 85: Drain-to-gate RMS voltage for VSWR lower than 8:1.

It can be observed that no breakdown voltage is exceeded. Moreover, coming back to the decision of using a LDMOS or a I/O transistor as the CG device, it can be seen from Figure 85 that the maximum drain-to-gate RMS voltage of this device is 1.8 V, which is 0.1 V lower than the breakdown limit for the I/O transistor. A safety margin of 0.1 V might not be adequate for this application considering the risk of the destruction of the device, and thus the decision of using an LDMOS transistor is justified.

## 6.7 Spurious emissions

According to the design specifications, the PA must provide attenuation at the harmonic frequencies in order to comply the spurious emission limits established in the 3GPP standard.

For this test, first the output power is computed by driving the PA with a current source of amplitude -59 dBA (which produces +16 dBm of output power) and sweeping the frequency of this current source from 100 Hz to 26 GHz. Then, +16 dBm are subtracted from the output power frequency response, thus obtaining the attenuation of the PA.

The result of this test and the design specification limits are plotted in Figure 86. A few observations can be made about it. First, the attenuation curve exhibits a low-pass frequency response because of the output matching network filtering properties. Second, the attenuation at the second harmonic (3700 MHz) is large because of the "second harmonic short circuit" present in the output matching network. Third, the attenuation has a band-pass behavior around the fundamental frequency (1850 MHz) as the input matching network is tuned using parallel resonance.

Finally, the attenuation of the PA satisfies the design requirements.

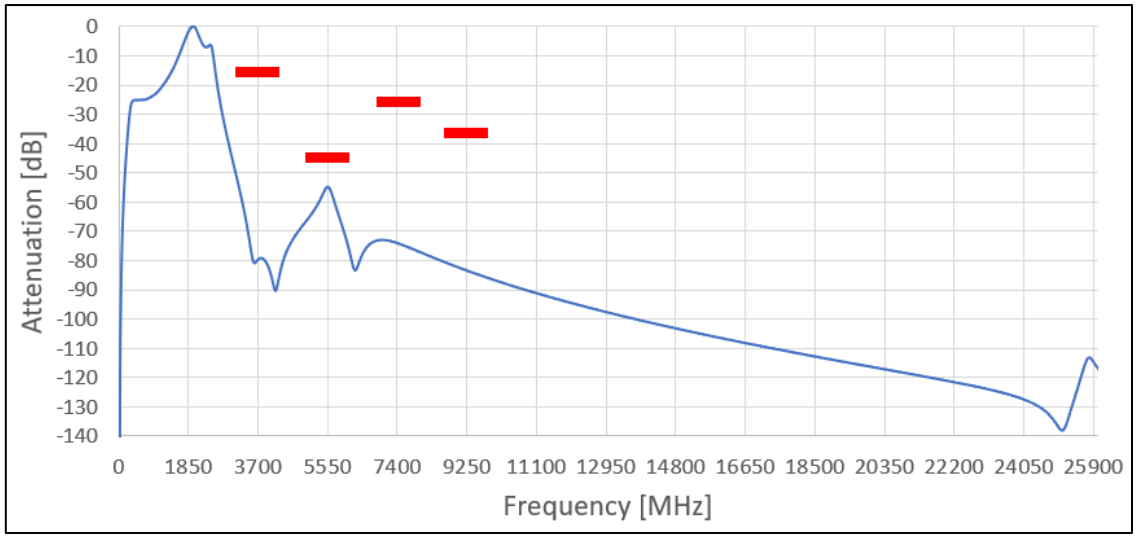



Figure 86: PA attenuation (blue) and the design specification limits (red).

## 6.8 Comparison with other works

The comparison of this work with other PA designs is presented in Table 24. The Georgia Tech PA performance survey [8] was used to find these other works. This survey studied more than 3200 papers on RF PA design published after the year 2000. It contains the most relevant parameters for each studied PA in a database format, facilitating the search for designs with specific properties.

Since all other semiconductor technologies are superior to CMOS regarding RF power applications, only PAs implemented using CMOS technology were considered. Also, in order to make a fair comparison, designs with similar saturated power ($P_{sat}$) and average power were considered. Moreover, only PAs operating between 900 MHz and 3000 MHz were included.

The PAE and the EVM were used as metrics for this comparison. Regarding the PAE, most papers provide the value for the peak power (PAE max), which does not represent the real power efficiency for modulations in which power peaks are sporadic. A more meaningful metric

is the PAE measured at the average power (average PAE), which describes the efficiency of the PA in continuous operation. Nevertheless, both PAE values were added in the comparison table.

A very important remark needs to be made. The PA design presented in this work has not been fabricated and measured as all other works in Table 24 have. Therefore, this comparison has to be taken just as an indication of the potential that the first iteration of the design of this RF PA has.

Table 24: Performance comparison with other works

| **Ref.** | **Frequency [GHz]** | **$P_{sat}$ [dBm]** | **PAE max [%]** | **EVM [dB]** | **Average $P_{out}$ [dBm]** | **Average PAE [%]** | **Process node** |
|---|---|---|---|---|---|---|---|
| [29] | 2.8 | 21.7 | 38 | -28 | 15.1 | 24 | 65 nm |
| [30] | 2.4 | 21.5 | 33 | -25.4 | 15 | 9 | 0.18 um |
| [31] | 2.4 | 23 | 47 | -28 | 15.6 | 23 | 45 nm |
| [32] | 2.3 | 22.4 | 38 (*) | -32 | 15.3 | 24(*) | 45 nm |
| [33] | 2 | 20.5 | 20 | -28.9 | 14.5 | 12 | 65 nm |
| [34] | 1.5 | 21.4 | 31 | -32.5 | 15.2 | 25 | 40 nm |
| **(**)** | **1.8** | **23** | **50** | **-28.5** | **16** | **24** | **28 nm** |

(*) Drain efficiency.

(**) This work.

# 7 CONCLUSIONS

## 7.1 Summary of this work

First, a brief description of the relevant technologies for integrated PA for mobile applications was presented. There, it was shown that PAs implemented in CMOS technology falls behind other technologies when comparing saturated power and efficiency for a given operating frequency.

Then, the principles of linear RF PAs were described. The most important manifestations of non-linearity in an RF circuit were presented, and concepts unique to RF PA design such as transistor reliability, load-line matching, and harmonic terminations were described. Also, the significance of the trade-off between linearity and power efficiency was shown while explaining the main classes of linear RF PAs.

After that, the design specifications were introduced. Primarily, the PA needed to satisfy the requirements of the standard 3GPP TS 36.101 for devices in the category NB1 within the Power Class 6 (+14 dBm). These requirements were made stricter by adding large implementation margins, which were established by the design team that implemented the other blocks of the transmitter chain. Most importantly, the main goal for the design of this PA was achieving the specifications without using integrated inductors or baluns, in order to minimize the silicon area consumption.

Then, a design methodology for the PA was proposed. Here, a detailed mathematical formulation, supported on computer-aided optimization, was used to justify the design decisions taken for each block of the PA. Also, a few techniques were proposed to improve the PA base performance.

Finally, the result of the tests of the PA performance were presented. In here, it was shown that the PA satisfied the design specifications with a few cases that require further improvement.

## 7.2 Discussion

The results show that it is feasible to implement a RF PA using a single-ended single-stage topology, satisfying the design specifications presented in Chapter 4. The PA provides the required output power, EVM and gain programmability, while complying with the emission limits.

In early stages of the design two approaches were studied but later abandoned. In the first one, the class-J PA [14] (also known as second harmonic enhancement PA [3]) was considered since it provides better power efficiency than a class-AB PA with comparable linearity. However, this PA has two issues that were discovered late. First, it requires two times more drain current than a class-AB PA for the same output power. Second, its efficiency falls dramatically when the gain is reduced by deactivating cells. Therefore, the class-AB topology was ended up being used.

The second initial approach was using an I/O transistor instead of a LDMOS as the common-gate device. Only late in the design process it was found that the I/O device could not withstand the large voltages creating under antenna mismatch situations.

These two failed attempts suggest that the design of a PA is an iterative process, in which the specifications are progressively satisfied until a road-block is found, forcing to return to the first design steps.

Finally, the limitations of CMOS technology for the implementation of RF PAs were evident while working in this design. The issue of low power supply voltage (needed to avoid breakdown) makes single-ended PA design an even more challenging task.

## 7.3 Opportunities for improvement

There are many items to be improved in the design presented in this document. The most important ones are described next.

1. Bias circuit: The constant$-g_m$ circuit needs a reference resistor to set a stable transconductance across process corners and temperatures. In case of implementing this resistor as an integrated component, a trimming and calibration routine needs to be utilized to minimize the effect of its resistance variations with the process and temperature.
2. Linearity: The EVM specification was slightly exceeded for a few corners, since the PA was tuned for the typical case. The PA can be re-tuned for improved EVM at those lacking corners. Some performance can be traded-off in the typical case, as there is some margin available.
3. Gain control: Linearity is reduced when using only a few cells of the PA. A compensation technique was proposed for this situation, improving the EVM up to a factor of 2. However, it is not enough, and a better compensation scheme needs to be designed.

## 7.4 Further development

This design requires further development by using three simulation methods: Post-layout simulation, 3D electromagnetic simulation and Monte Carlo simulation.

First and most important, the post-layout simulation will show the effect of individual component parasitics as well as the parasitics created by component layout. These effects will degrade the PA performance and therefore, further adjustment to the design will be required. Though, as explained in section 5.7, there are 2 dB of additional gain that can be obtained by adjusting the feedback resistance, if needed after post-layout simulation.

Second, the 3D EM simulation can show the effect of the electromagnetic coupling among components. This is critical for determining if any instability is introduced by the magnetic coupling between metal traces carrying the output signal and the balun metal traces carrying the input signal, considering the proximity of these two components. Moreover, the balun design presented in this work needs to be tested by using 3D EM simulation.

Third, Monte Carlo simulation accounts for the statistical variations of the process parameters among components, showing the effect of random mismatch on the performance of the PA. This simulation will exhibit issues where large matching ratios are present, such as between the bias circuit and the PA core (1:200 ratio).

# 8 REFERENCES


[1] Internet Society, "The Internet of Things: An overview," 2015.

[2] O. Liberg, M. Sundberg, E. Wang, J. Bergman and J. Sachs, Cellular Internet of Things: Technologies, standards and performance, Academic Press, 2018.

[3] B. Razavi, RF Microelectronics, second edition, Pearson Education, 2012.

[4] A. Salemi, " (PhD Thesis) Silicon Carbide technology for high and ultra high voltage bipolar junction transistors and PiN diodes," Purdue University, 2017.

[5] S. M. Sze and M. K. Lee, Semiconductor devices: Physics and technology, John Wiley & Sons, 2012.

[6] M. F. A. Katz, "GaN comes of age," *IEEE Microwave Magazine,* vol. 11, no. 7, 2010.

[7] M. Kamper, "(PhD thesis) Differential switched mode RF power amplifiers," FAU University Press, Erlangen, 2018.

[8] 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 a. M. E. D. Smith, "Power Amplifiers Performance Survey 2000-Present," Georgia Tech Electronics and Micro-System Lab (GEMS), [Online]. Available: https://gems.ece.gatech.edu/PA_survey.html.

[9] G. Liu, "(PhD thesis) Fully integrated CMOS power amplifier," University of California, Berkeley, 2006.

[10] C. Fallensen, "Design techniques for sub-micron RF power amplifiers," Technical University of Denmark, 2001.

[11] H. Solar and R. Berenguer, Linear CMOS RF power amplifiers: A complete design workflow, Springer, 2014.

[12] S. Lotfi, "Design and characterization of RF-LDMOS transistors and Si-on-SiC hybrid substrates," Uppsala University, 2014.

[13] P. Allen and D. Holberg, CMOS analog circuit design - Third edition, Oxford University press, 2012.

[14] S. Cripps, "RF power amplifiers for wireless communications - second edition," Artech House, 2006.

[15] E. Kreyszig, "Advanced engineering mathematics - tenth edition," John Wiley & Sons, 2011.

[16] J. Pedro and N. Carvalho, "Intermodulation distortion in microwave and wireless circuits," Artech House, 2003.

[17] A. Luzzatto and M. Haridim, "Wireless transceiver design: Mastering the design of modern wireless equipment and systems," Wiley, 2017.

[18] D. Pozar, Microwave engineering, John Wiley & Sons, 2012.

[19] J. Long, "Monolithic transformers for silicon RF IC design," *IEEE Journal of solid-state circuits,* vol. 35, no. 9, 2000.

[20] I. Aoki, S. Kee, D. Ritledge and A. Hajimiri, "Distributed active transformer - A new power-combining and impedance-transformation technique," *IEEE Transactions on microwave theory and techniques,* vol. 50, no. 1, 2002.

[21] 3rd Generation Partnership Project, "3GPP TS 36.101 v16.5.0," 2020.

[22] National Instruments, "Introduction to LTE device testing: From theory to transmitter and receiver measurements," [Online]. Available: http://download.ni.com/evaluation/rf/Introduction_to_LTE_Device_Testing.pdf.

[23] J. M. Carusone, Analog integrated circuit design, 2012.

[24] Aoki, Kee, Rutledge and Hajimiri, "Distributed active transformer - A new power-ccombining and impedance-transformation technique," *IEEE Transactions on microwave theory and techniques,* vol. 50, no. No 1, 2002.

[25] M. Kaixue, T. Kumar and K. S. Yeo, "A reconfigurable K-/Ka-band power amplifer with high PAE in 0.18-um SiGe BiCMOS for multi-band applications," *IEEE transactions on microwave theory and techniques,* vol. 63, no. 12, 2015.

[26] R. Bagger and H. Sjoland, "An 11 GHz-Bandwidth variable gain Ka-band power amplifier for 5G applications," *IEEE 45th European Solid State Circuits Conference,* 2019.

[27] J.-C. Wen and L.-L. Sun, "A variable gain and output power CMOS PA with combination switch controls," *IEEE International Conference on Solid-State and Integrated Circuit Technology,* 2010.

[28] G. Gonzalez, Microwave transistor amplifiers: Analysis and design., Prentice Hall, 1997.

[29] Z. Bai, A. Azam and J. S. Walling, "A Frequency Tuneable Switched-Capacitor PA in 65nm CMOS," *2019 IEEE Radio Frequency Integrated Circuits Symposium (RFIC),* 2019.

[30] P.-C. Huang, Z.-M. Tsai, K.-Y. Lin and H. Wang, "A High-Efficiency, Broadband CMOS Power Amplifier for Cognitive Radio Applications," *IEEE Transactions on Microwave Theory and Techniques,* 2010.

[31] R. Hezar, L. Ding, A. Banerjee, J. Hur and B. Haroun, "A PWM Based Fully Integrated Digital Transmitter/PA for WLAN and LTE Applications," *IEEE Journal of Solid-State Circuits,* 2015.

[32] D. Jung, S. Li, J.-S. Park, T.-Y. Huang, H. Zhao and H. Wang, "A CMOS 1.2-V Hybrid Current- and Voltage-Mode Three-Way Digital Doherty PA With Built-In Phase Nonlinearity Compensation," *IEEE Journal of Solid-State Circuits,* 2020.

[33] W. Yuan, V. Aparin, J. Dunworth, L. Seward and J. S. Walling, "A Quadrature Switched Capacitor Power Amplifier," *IEEE Journal of Solid-State Circuits,* 2016.

[34] L. Xiong, T. Li, Y. Yin, H. Min, N. Yan and H. Xu, "4.2 A Broadband Switched-Transformer Digital Power Amplifier for Deep Back-Off Efficiency Enhancement," *2019 IEEE International Solid- State Circuits Conference,* 2019.

[35] A. Vasylyev, "(PhD thesis) Integrated RF power amplifier design in silicon-based technologies," TU Berlin, 2006.