Technical University of Crete
Electrical and Computer Engineering

ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ
TECHNICAL UNIVERSITY OF CRETE

Design of Stacked High Power RF Amplifiers (RF PA) for 5G New Radio Base Station with GaN and FDSOI Technologies

Diploma Thesis

submitted by

Roussos Nikolaos

Committee:

Professor Bucher Matthias (Supervisor)

Professor Liavas Athanasios

Dr. Vryssas Konstantinos

Chania, December of 2024

Acknowledgements

I would like to express my heartfelt gratitude to Professor Bucher for entrusting me with this fascinating and challenging topic. I am deeply thankful for the knowledge he has imparted to me throughout my studies and for his invaluable guidance during the course of this thesis. Furthermore, I am sincerely grateful for the opportunity he provided me to collaborate with Argo Semiconductors for the completion of my thesis project.

Additionally, I would like to extend my sincere thanks to all the members of Argo Semiconductors for their invaluable assistance. They shared their extensive expertise in the field of microelectronics and introduced me to the Cadence design tool, which was instrumental in my work. In particular, I am deeply grateful to Dr. Vryssas for sharing his profound knowledge in power amplifier design and for guiding me through all stages of the design process.

A special thanks is also due to Dr. Volanis, whose valuable assistance greatly contributed to the success of this work. I deeply appreciate his support throughout the thesis project.

Finally, I am deeply grateful to my family, who supported me throughout my studies.

Abstract

The transition to 5G technology is a significant step forward in mobile communication, as it is expected to achieve higher data rates, ultra-reliable low-latency communications (uRLLC), and massive machine-type communications (mMTC). The 5G New Radio (5G NR) developed by 3rd Generation Partnership Project (3GPP) is a radio access technology (RAT) that supports a wide frequency range, FR1 (410 MHz – 7.125 GHz) and FR2 (24.25 GHz – 71 GHz). 5G NR is important for a broad variety of applications, for example robotics, automotive, factory automation, and healthcare which require enhanced mobile broadband (eMBB) and high-efficiency spectrum use.

There is an increasing use of Gallium Nitride (GaN) High Electron Mobility Transistors (HEMTs) in power amplifiers for high power and broadband base stations in the 5G networks. Unlike CMOS, GaN HEMTs can deliver high power levels at high frequencies. A feature which is essential for mmWave 5G applications.

In this thesis, two differential Stacked FET Power Amplifiers (PAs) are implemented for the FR1 n79 frequency band of 4.4-5 GHz. The first PA is designed using 22nm Fully-Depleted Silicon-On-Insulator (FDSOI) transistor technology (22FDX), and the second PA is designed using GaN HEMT 0.15 µm technology (GH15). A comparison of the performance metrics reveals that while the 22FDX CMOS PA achieves an $OP_{1\text{dB}}$ of 28.1 dBm with a Power Added Efficiency (PAE) of 30% and a gain of 26 dB, the GaN PA significantly outperforms it in terms of output power, delivering an $OP_{1\text{dB}}$ of 44.5 dBm, though with a lower PAE of 16% and a gain of 22.7 dB.

Given its OP1dB of 28.1 dBm, the 22FDX Stacked PA could be effectively utilized as a pre-power amplifier, feeding the GaN Stacked PA (when it comes to FR1 applications) to enhance overall system level of integration. On the other hand, CMOS SOI PA is a perfect candidate for FR2 bands, where massive MiMo allows for lower power levels at the antenna, since a large number of antenna elements – and thus PAs – are used to build the phased array system. This approach highlights the complementary strengths of both technologies, with GaN offering superior power capabilities while CMOS can serve critical roles in the amplification chain.

Contents

Chapter 1

Introduction

1.1 Importance of 5G in modern communication systems

The shift from the fourth generation (4G) to the fifth generation (5G) in mobile communication represents a significant advancement in wireless technology, promising to enhance user experiences by meeting the growing demand for higher data rates and an increasing number of connected devices. The transition to 5G is not merely an incremental improvement but a foundational change in how mobile networks operate, supporting a wide array of applications across various industries, including robotics, automotive, factory automation, healthcare, and education.

5G New Radio (NR), developed by the 3rd Generation Partnership Project (3GPP), is the cornerstone of this revolution. It supports a broad spectrum of frequencies, including the millimeter-wave (mmWave) bands, which offer high spectral efficiency, data speeds up to 20 Gbps, and channel bandwidths up to 2 GHz. These capabilities make 5G NR suitable for various applications, from fixed wireless access to mobile broadband, enabling seamless connectivity in both fixed and mobile environments. The NR framework is designed to support diverse usage scenarios such as ultra-reliable low latency communications (uRLLC), enhanced mobile broadband (eMBB), and massive machine type communications (mMTC), each with distinct system-level performance requirements as depicted in 1.1.

A significant distinction between 4G and 5G, as highlighted in figure 1.2, is the evolution from a network primarily focused on enhancing mobile broadband services for human users to one that supports a wide range of devices and applications, including the Internet of Things (IoT). This transition brings substantial improvements in peak data rate, latency, connection density, and energy efficiency. These advancements are critical not only for enhancing current

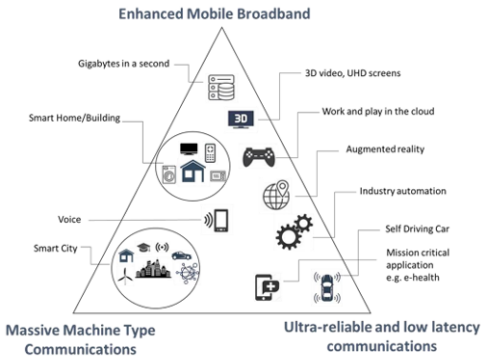

Figure 1.1: Some usage scenarios proposed by International Mobile Telecommunications (IMT) [1]

		4G	5G
Usage Scenarios		• MBB	• eMBB • uRLLC • mMTC
Applications		• High definition videos • Voice • Mobile TV • Mobile internet • Mobile pay	• VR/AR/360° videos • UHD videos • V2X • IoT • Smart city/Factory/Home • Telemedicine • Wearable devices
KPI	Peak data rate	100 Mb/s	20 Gb/s
	Experienced data rate	10 Mb/s	0.1 Gb/s
	Spectrum efficiency	1x	3x that of 4G
	Network Energy Efficiency	1x	10 – 100x that of 4G
	Area traffic capacity	0.1 Mb/s/m²	10 Mb/s/m²
	Connection density	10^5 Devices/km²	10^6 Devices/km²
	Latency	10 ms	1 ms
	Mobility	350 km/h	500 km/h
Technologies		• OFDM • MIMO • Turbo code • Carrier aggregation • HetNet • D2D communications • Unlicensed spectrum	• Mm-Wave communications • Massive MIMO • LDPC and polar codes • Flexible frame structure • Ultra-dense networks • Cloud/fog/edge computing • SDN/NFV/network slicing

Figure 1.2: The network features of 4G and 5G. VR: virtual reality, AR: augmented reality, UHD: ultra-high definition, V2X: vehicle to everything, IoT: internet of things, OFDM: orthogonal frequency-division multiplexing, MIMO: multiple input-multiple output, HetNet: heterogeneous network, D2D: device to device, LPDC: low-density parity check codes, SDN: software-defined networking, NFV: network function virtualization [1]

communication systems, but also for enabling new applications and services that were previously unattainable with earlier generations. However, achieving these capabilities, particularly the high data rates and low latencies, depends heavily on the efficiency and performance of power amplifiers (PAs) used in 5G transmitters.

In the deployment of 5G networks, the FR1 band is primarily utilized for providing widespread coverage and ensuring robust connectivity, especially in densely populated urban areas and indoors. Its lower frequencies allow signals to travel further and penetrate obstacles more effectively than the higher frequencies used in the FR2 band. However, the FR2 band, with its higher frequency range, is critical for delivering the highest data rates and supporting advanced features like 3D beamforming. This makes the design of RF PAs particularly challenging, as they must handle the demanding requirements of broadband modulation, massive MIMO, and beamforming that are essential for achieving the high performance promised by 5G. GaN-based PAs, known for their high power density and efficiency, are especially well-suited to meet these challenges, making them a key component in the successful deployment of 5G networks.

1.2 Motivation for using GaN HEMT and FD-SOI technologies in PA design

The rapid advancement of 5G technology, particularly in the mm-Wave frequency bands, has intensified the demand for efficient and cost-effective RF front-end modules (FEMs). A critical component of these FEMs is the power amplifier (PA), whose performance in terms of

Power Added Efficiency (PAE), output power, linearity, and form factor directly influences the overall efficiency and viability of 5G networks, especially in massive MIMO configurations.

Fully-Depleted Silicon-On-Insulator (FD-SOI) technology has emerged as a strong candidate for low-cost integration of PAs in FEMs, particularly for 5G mm-Wave applications. The key advantage of FD-SOI lies in its ability to integrate complex RF front-end circuitry at a lower cost, which is crucial for the deployment of massive MIMO antennas where numerous PAs are required. This low-cost integration is vital for 5G's success, as it allows for scalable, high-volume production necessary to meet the network's extensive demands.

However, FD-SOI technology is traditionally geared towards low-power applications, which comes with the drawback of lower breakdown voltage. This limitation necessitates the use of advanced design techniques, such as cascode or stacked configurations, to achieve the required high output power without exceeding the device's voltage limits. These configurations effectively distribute the voltage stress across multiple transistors, thereby enhancing the overall robustness and performance of the PA. The ability of FD-SOI to combine low-cost integration with these innovative design approaches makes it a compelling choice for specific 5G applications, despite its inherent limitations in power handling [2], [3], [4], [5].

While FD-SOI offers significant advantages in terms of integration and cost, III-V semiconductor technologies such as GaN HEMTs provide unparalleled performance in power amplification, especially at higher frequencies. GaN HEMTs are well-regarded for their high output impedance, low output capacitance, and high breakdown voltage, which collectively contribute to their superior PAE and higher output power (P_{OUT}). These attributes are particularly beneficial for base-station applications in 5G networks, where efficiency and power density are paramount [6].

However, GaN and other III-V technologies like Gallium Arsenide (GaAs) face challenges related to integration and cost. Their limited ability to integrate multiple RF front-end and control ICs within a single module, coupled with higher production costs, poses potential barriers to their widespread adoption in massive MIMO configurations, where numerous PAs are needed to perform phased-array beamforming.

1.3 Overview of the Power Amplifier Designs and Objectives

The objective of this diploma thesis is to present two differential-ended Power Amplifier designs in the n79 5G NR frequency band (4.4-5 GHz) using CMOS and GaN HEMT technologies. The designs were implemented in 22FDX and GH15 technologies, achieving OP_{1dB} of 28.1 dBm and 44.5 dBm respectively at the frequency of interest. The whole design process will be elaborately discussed, and the verification will be provided via simulations with Cadence Virtuoso and AWR MWO Environments in each case.

Chapter 2

RF Design Fundamentals

2.1 Y, Z, ABCD Parameters

2.1.1 Y Parameters

The electrical behavior of linear RF networks is described using the admittance or Y parameters. These parameters relate the input and output voltages and currents at the two ports of a network. The Y parameters for a two-port network are defined as follows:

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}. \tag{2.1}$$

In general, for a n-port network, the admittance Y_{ij} is defined as:

$$Y_{ij} = \left.\frac{I_i}{V_j}\right|_{V_k=0}, \text{ for } k \neq j. \tag{2.2}$$

This implies that while calculating the admittance all the other ports are connected by short circuit (or arbitrary grounded) while the port j is left open. This is used to express admittance in microwave networks, although it is frequency dependent.

2.1.2 Z Parameters

The impedance parameters or Z parameters relate the input and output voltage and current in a two-port network. They are defined in terms of the open-circuit impedance matrix as follows:

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix}. \tag{2.3}$$

For a n-port network, the impedance Z_{ij} is calculated as:

$$Z_{ij} = \left.\frac{V_i}{I_j}\right|_{I_k=0}, \text{ for } k \neq j. \tag{2.4}$$

This means that all the other ports except port j are assumed to be open circuited, which means that no current flows through the other ports. From these definitions, it is known that:

$$[Z] = [Y]^{-1}. \tag{2.5}$$

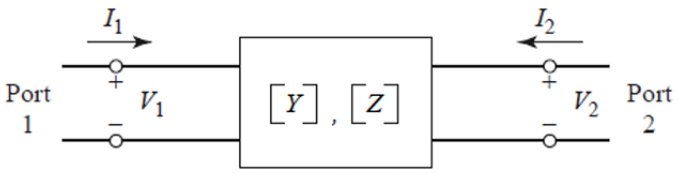

Figure 2.1: Two-port network for representation of Y, Z [7].

2.1.3 ABCD Parameters

Many high-frequency circuits are formed by a series connection of two-port networks. For such networks, relevant and simplified parameters are the ABCD parameters (or chain parameters). The ABCD matrix is defined for a two-port network as shown below:

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} V_2 \\ I_2 \end{bmatrix}. \tag{2.6}$$

$$A = \left.\frac{V_1}{V_2}\right|_{I_2=0}, \quad B = \left.\frac{V_1}{I_2}\right|_{V_2=0}. \tag{2.7}$$

$$C = \left.\frac{I_1}{V_2}\right|_{I_2=0}, \quad D = \left.\frac{I_1}{I_2}\right|_{V_2=0}. \tag{2.8}$$

The n-port network will, in most instances, be symmetric, implying that $A = D$. Moreover, when two two-port networks are connected in series their ABCD matrices can be multiplied to get the ABCD parameters of the overall network. For example, when two networks with parameters $[A_1, B_1, C_1, D_1]$ and $[A_2, B_2, C_2, D_2]$ are cascaded, the total ABCD parameters are:

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix} \begin{bmatrix} A_2 & B_2 \\ C_2 & D_2 \end{bmatrix} \begin{bmatrix} V_3 \\ I_3 \end{bmatrix}. \tag{2.9}$$

Thus, the overall ABCD matrix becomes:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix} \begin{bmatrix} A_2 & B_2 \\ C_2 & D_2 \end{bmatrix}. \tag{2.10}$$

2.2 S-parameters

S-parameters are called scattering parameters and are defined for linear networks specifically for RF and microwave applications at higher frequencies. At such frequencies, the use of Z and Y parameters becomes impractical due to instability. S-parameters are more convenient when describing networks where the signal is made up of traveling waves.

S-parameters relate the incident and reflected voltage waves at the ports of a network, as shown in Figure 2.3. For a two-port network, the incident waves (a_1, a_2) and reflected waves

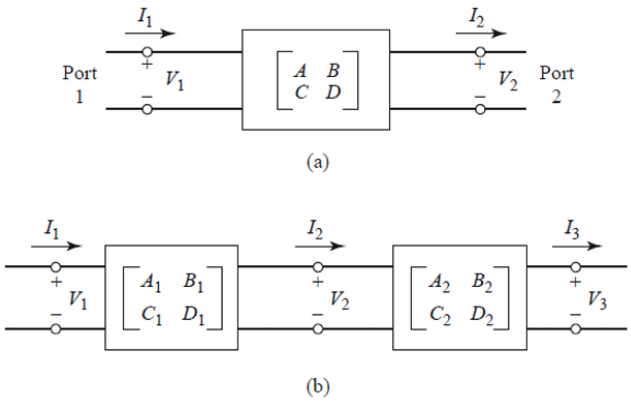

Figure 2.2: (a) Two-port network for ABCD matrix, (b) Cascade of two two-port networks [7]

(b_1, b_2) are described as follows:

$$S_{11} = \left.\frac{b_1}{a_1}\right|_{a_2=0}, \quad S_{12} = \left.\frac{b_1}{a_2}\right|_{a_1=0}, \tag{2.11}$$

$$S_{21} = \left.\frac{b_2}{a_1}\right|_{a_2=0}, \quad S_{22} = \left.\frac{b_2}{a_2}\right|_{a_1=0}. \tag{2.12}$$

The S-parameters are defined as follows:

- S_{11} is the input reflection coefficient.
- S_{21} is the forward transmission coefficient.
- S_{12} is the reverse transmission coefficient.
- S_{22} is the output reflection coefficient.

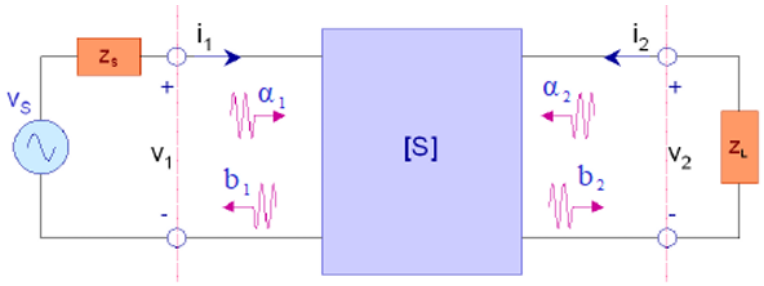

Figure 2.3: Two-port network representing the S-parameter matrix $[S]$. The incident waves (a_1, a_2) and reflected waves (b_1, b_2) are related by the S-parameters [8].

The general form of the two-port S-parameter matrix can be written as:

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}. \tag{2.13}$$

S-parameters are complex numbers, with both a magnitude and a phase. They are often expressed in decibel (dB) scale, where the magnitude is given as:

$$S_{mn,dB} = 20 \log_{10} |S_{mn}|. \tag{2.14}$$

S-parameters are essential for characterizing the behavior of microwave networks and are measured using devices such as vector network analyzers (VNAs). These parameters allow us to evaluate key performance metrics, such as gain, reflection, and isolation.

2.3 Stability Criteria

In contrast to passive components, which never oscillate, an active device may either amplify a signal or sustain continuous oscillations. Since the goal is to design an amplifier, oscillations must be avoided, as they indicate instability in the amplifier's performance. Stability is particularly crucial when the amplifier is expected to operate at a fixed frequency without continuous oscillations. To achieve this, studying the stability conditions of an amplifier is essential.

2.3.1 Reflection Coefficients

To determine the stability criteria, we first need to define the reflection coefficients. The reflection coefficient describes the ratio of a reflected wave to the incident wave at the input or output of the network.

For a general reflection coefficient at port i, we define it as:

$$\Gamma_i = \frac{b_i}{a_i} = \frac{V_i^-}{V_i^+}, \tag{2.15}$$

where:

- Γ_i is the reflection coefficient,
- b_i is the reflected wave amplitude at port i,
- a_i is the incident wave amplitude at port i,
- V_i^- and V_i^+ are the reflected and incident voltages at port i, respectively.

as shown in Figure 2.3.

The reflection coefficient that looks towards the load is:

$$\Gamma_L = \frac{b_L}{a_L} \Rightarrow b_L = \Gamma_L \cdot a_L. \tag{2.16}$$

From Figure 2.3, we observe:

$$V_L^+ = -V_2^+ = V_2^- \Rightarrow a_L = b_2, \tag{2.17}$$

$$V_L^- = -V_2^- = V_2^+ \Rightarrow b_L = a_2. \tag{2.18}$$

Substituting the equations 2.17, 2.18 into the reflection equation 2.16:

$$a_2 = \Gamma_L b_2. \tag{2.19}$$

In a two-port network, the waves at the input and output can be related by the following equations using S-parameters:

$$b_1 = S_{11}a_1 + S_{12}a_2, \tag{2.20}$$

$$b_2 = S_{21}a_1 + S_{22}a_2. \tag{2.21}$$

By combining all the above information, the input reflection coefficient is formed as:

$$\Gamma_{\text{IN}} = \frac{b_1}{a_1} \overset{2.20}{=} \frac{S_{11}a_1 + S_{12}a_2}{a_1} \overset{2.19}{=} S_{11} + \frac{S_{12}\Gamma_L b_2}{a_1} \overset{2.21}{=} S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L}. \tag{2.22}$$

Similarly, using the same procedure, the output reflection coefficient is:

$$\Gamma_{\text{OUT}} = \frac{b_2}{a_2} = S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1 - S_{11}\Gamma_S}. \tag{2.23}$$

2.3.2 Stability Circles

Stability circles provide a visual tool to determine regions in the Smith chart where instability could arise. These circles help in analyzing whether the reflection coefficients Γ_{IN} and Γ_{OUT} are within stable or unstable regions.

For a two-port network, oscillation is likely if either the input or output impedance has a negative real part. This scenario is avoided by ensuring that the reflection coefficients at both the input and output meet the following conditions [9]:

$$|\Gamma_{\text{IN}}| < 1 \quad \text{and} \quad |\Gamma_{\text{OUT}}| < 1. \tag{2.24}$$

When both conditions are satisfied for all passive source and load impedances, the amplifier is said to be unconditionally stable. If these conditions are not met, the amplifier may exhibit conditional stability or instability depending on the impedance values.

Stability circles are characterized by the condition where the magnitude of the reflection coefficients equals one. These conditions are expressed mathematically as:

$$|\Gamma_{\text{IN}}| = 1 \quad \text{and} \quad |\Gamma_{\text{OUT}}| = 1. \tag{2.25}$$

The parameters of the stability circle, specifically its center and radius, can be calculated. For the output stability circle, the center C_L and radius R_L are defined as follows:

$$C_L = \frac{(S_{22} - \Delta S_{11}^*)^*}{|S_{22}|^2 - |\Delta|^2}, \tag{2.26}$$

$$R_L = \frac{|S_{12}S_{21}|}{|S_{22}|^2 - |\Delta|^2}, \tag{2.27}$$

Δ is the determinant of the S-parameter matrix and is given by:

$$\Delta = S_{11}S_{22} - S_{12}S_{21}. \tag{2.28}$$

Similarly, for the input stability circle, the center C_S and radius R_S are expressed as:

$$C_S = \frac{(S_{11} - \Delta S_{22}^*)^*}{|S_{11}|^2 - |\Delta|^2}, \tag{2.29}$$

$$R_S = \frac{|S_{12}S_{21}|}{|S_{11}|^2 - |\Delta|^2}. \tag{2.30}$$

For unconditional stability, the stability circles must lie completely outside the Smith chart. This ensures that no unstable reflection coefficients exist within the practical impedance range. The condition for unconditional stability can be written as:

$$|C_L| - R_L > 1 \quad \text{and} \quad |C_S| - R_S > 1. \tag{2.31}$$

When both of these conditions are met, the reflection coefficients at both the input and output ports are stable for all passive impedance values, guaranteeing that no oscillations will occur in the amplifier.

2.3.3 Rollet's Stability Criterion (k-Factor)

Rollet's stability criterion, also known as the k-factor criterion, provides a simple numerical method to verify unconditional stability. The criterion is based on the S-parameters and provides a value k_f. For unconditional stability, the following conditions must be met [10]:

$$k_f = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}S_{21}|} > 1, \tag{2.32}$$

$$|\Delta| = |S_{11}S_{22} - S_{12}S_{21}| < 1. \tag{2.33}$$

An additional stability parameter, B_{1f}, is used to further assess stability. The B_{1f} parameter is calculated as follows:

$$B_{1f} = 1 + |S_{11}|^2 - |S_{22}|^2 - |\Delta|^2. \tag{2.34}$$

If $B_{1f} > 0$, the amplifier is considered to be stable.

2.3.4 Stability Factors μ and μ'

The μ and μ' stability factors provide a reliable method for assessing the unconditional stability of a two-port network by focusing on the input and output reflection coefficients [11].

The μ' factor is used to assess the stability at the input of a two-port network. It is defined by the following equation:

$$\mu' = \frac{1 - |S_{11}|^2}{|S_{22} - \Delta S_{11}^*| + |S_{12}S_{21}|}, \tag{2.35}$$

$\Delta = S_{11}S_{22} - S_{12}S_{21}$.

Similarly, the μ factor assesses the stability at the output of the amplifier and is defined as:

$$\mu = \frac{1 - |S_{22}|^2}{|S_{11} - \Delta S_{22}^*| + |S_{12}S_{21}|}. \tag{2.36}$$

When both $\mu' > 1$ and $\mu > 1$, the source and load impedances that could cause instability lie outside the Smith chart. This guarantees unconditional stability under all practical operating conditions.

Stability is a crucial factor in amplifier design. Ensuring both input and output stability across the desired frequency range is essential for reliable amplifier performance. Methods

like stability circles, Rollet's criterion, and the B_{1f} stability parameter provide practical ways to analyze and guarantee the stability of a circuit. The μ and μ' factors complement these approaches, offering a straightforward method for determining unconditional stability across a wide range of source and load impedances.

2.4 Power Gain Definitions

In amplifier design, power gain is a crucial metric used to evaluate the performance of the amplifier. Different definitions of power gain are employed, each providing unique insights into the amplifier's behavior depending on the stage of the design process. Below are the most common types of power gain:

2.4.1 Operating Power Gain (G)

Operating power gain is defined as the ratio of the power delivered to the load (P_L) to the power provided at the input (P_{in}). This gain is independent of the source impedance and represents the power transferred from the input to the output of the two-port network. It is expressed as:

$$G = \frac{P_L}{P_{\text{in}}}. \tag{2.37}$$

2.4.2 Available Power Gain (G_A)

Available power gain refers to the ratio of the power available from the two-port network to the power available from the source. This type of gain assumes *conjugate matching* both at the input and output ports, meaning that the source and load impedances are conjugates of the input and output impedances of the two-port network. The expression for available power gain is:

$$G_A = \frac{P_{\text{av}}}{P_{\text{avs}}}, \tag{2.38}$$

P_{av} is the available power from the amplifier, and P_{avs} is the available power from the source.

2.4.3 Transducer Power Gain (G_T)

The transducer power gain defines the ratio of the power delivered to the load to the power available from the source. This gain takes into account the matching conditions at both the input and output, making it a commonly used metric in practical amplifier design. The transducer power gain is calculated as:

$$G_T = \frac{P_L}{P_{\text{avs}}}. \tag{2.39}$$

These three definitions provide a complete framework for understanding the power performance of an amplifier. Each type of gain considers different conditions for impedance matching, making them suitable for different stages of design and optimization.

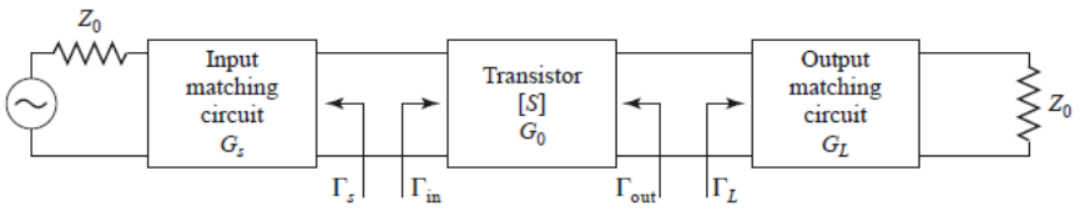

Figure 2.4: Single stage amplifier with it's matching networks [7].

2.5 Impedance Matching

In RF and microwave circuit design, impedance matching is crucial for ensuring maximum power transfer between different stages, such as from a source to an amplifier or from an amplifier to a load. Proper impedance matching helps to minimize signal reflections, optimize power transfer, and maintain the stability of the amplifier across its operating frequency range [7, 12].

In order to avoid signal reflections that cause large power loss, the source impedance and the load impedance shall be matched usually referred to as *impedance matching*. Reflections take place when the impedances between the two stages of a circuit are nonconjugate and part of the signals bounces back towards the source, hence the power that gets through to the load is decreased. This fact is more critical at high frequencies, which leads to substantially reduced performance. According to [7], matching circuits are designed to achieve a condition where the load and source impedances are conjugate matched, which means that:

$$\Gamma_{\text{in}} = \Gamma_S^* \quad \text{and} \quad \Gamma_{\text{out}} = \Gamma_L^*, \tag{2.40}$$

Γ_{in} and Γ_{out} are the reflection coefficients looking into the amplifier's input and output, respectively, and Γ_S and Γ_L are the reflection coefficients of the source and load.

The main fact to understand is the maximum power transfer theorem according to which, the load impedance must be equal to the complex conjugate of the source impedance. This means:

$$Z_S = Z_L^* \quad \Rightarrow \quad \Gamma_S = \Gamma_L^*. \tag{2.41}$$

Solitary if these conditions attain, the maximum power transfer to the load is achieved, and the performance of the circuit is advanced [12]. To establish this condition, with high-frequency applications, matching networks composed of inductors, capacitors or transmission line sections are employed. These types of networks transform the impedance of one stage to be conjugate matched to the next stage in a manner to meet matching conditions at a particular bandwidth.

From Figure 2.4, the gain of a system with impedance matching is influenced by three main factors, namely G_S, G_0, and G_L, which are the gains of the input matching network, the amplifier, and the output matching network, respectively. The overall maximum transducer

power gain is defined as:

$$G_{T,\max} = G_S G_0 G_L = \frac{|S_{21}|^2(1-|\Gamma_S|^2)(1-|\Gamma_L|^2)}{|1-S_{11}\Gamma_S|^2|1-S_{22}\Gamma_L|^2}, \tag{2.42}$$

where

$$G_S = \frac{1-|\Gamma_S|^2}{|1-\Gamma_{\text{in}}\Gamma_S|^2}, \tag{2.43}$$

$$G_0 = |S_{21}|^2, \tag{2.44}$$

$$G_L = \frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2}. \tag{2.45}$$

The equations 2.43, 2.45 highlight that G_S and G_L are functions of the reflection coefficients Γ_S and Γ_L, which depend on how well the input and output are matched [13]. The equations for the reflection coefficients that relate the source and load impedances are:

$$\Gamma_S = \frac{B_1 \pm \sqrt{B_1^2 - 4|C_1|^2}}{2C_1}, \quad \Gamma_L = \frac{B_2 \pm \sqrt{B_2^2 - 4|C_2|^2}}{2C_2}, \tag{2.46}$$

$$B_1 = 1 + |S_{11}|^2 - |S_{22}|^2 - |\Delta|^2, \quad B_2 = 1 + |S_{22}|^2 - |S_{11}|^2 - |\Delta|^2, \tag{2.47}$$

$$C_1 = S_{11} - \Delta S_{22}^*, \quad C_2 = S_{22} - \Delta S_{11}^*. \tag{2.48}$$

These equations define the conditions under which the input and output impedances (Z_S and Z_L) must be adjusted to achieve maximum power transfer [13].

The matching conditions are realized in practical designs using specific matching networks that are tuned to the required operating frequency and bandwidth [12]. Impedance transformation and design of the matching network is often graphically realized by using the Smith Chart [7]. Impedance matching allows amplifiers to operate with minimal reflections and maximum power efficiency, and is a critical factor in achieving high power system performance.

2.6 L-Matching Networks

2.6.1 Series Reactive Element

It is possible to use reactive elements like inductors or capacitors as a technique to achieve impedance matching. These elements are applied to convert the input or output impedance of a circuit to the conjugate impedance of the input or output port, respectively. With this technique, the power gain is enhanced and the signal reflections are reduced.

A series reactive element, X_S, placed in series with a resistance, R, can change the impedance seen at the input, transforming the impedance of a load or source, as shown in figure 2.5. This approach is useful in creating equivalent circuits for impedance matching purposes. The input admittance of this series connection is given by:

$$Y_{\text{in}}(\omega) = \frac{1}{Z_{\text{in}}} = \frac{1}{R + jX_S} = \frac{R}{R^2 + X_S^2} - j\frac{X_S}{R^2 + X_S^2}. \tag{2.49}$$

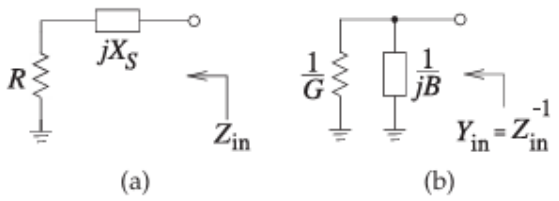

Figure 2.5: Impedance matching using a series reactance X_S: (a) series equivalent circuit, (b) parallel equivalent circuit [14].

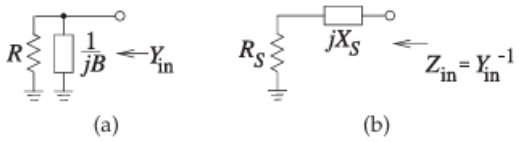

Figure 2.6: A resistor with (a) a shunt parallel reactive element where B is a susceptance, and (b) its equivalent series circuit [14].

From this equation it can be seen that the real part of the input admittance Y_{in} is the conductance G and the imaginary part is the susceptance B. These elements form the equivalent shunt network and are defined as follows:

$$G = \frac{R}{R^2 + X_S^2}, \tag{2.50}$$

$$B = -\frac{X_S}{R^2 + X_S^2}. \tag{2.51}$$

Thus, the series connection of R and X_S is replaced by an equivalent shunt circuit represented by a resistance of $1/G$ and a reactance of $1/B$.

Using this method, the series resistance R is transformed into a new equivalent resistance, R_P, which is greater than R. The transformed resistance is given by:

$$R_P = \frac{R^2 + X_S^2}{R}. \tag{2.52}$$

The resistance R_P can be set to any value that is desired by selecting an appropriate value of X_S as shown above. Nonetheless, for further compensation of the residual reactance, it is required to add a shunt reactive element.

2.6.2 Shunt Reactive Element

In particular, when either a capacitor or an inductor is connected in parallel with a resistive load the equivalent impedance is changed. The input admittance of the shunt circuit, which consists of a resistor R in parallel with a reactive component with susceptance jB, as shown in figure 2.6, is given by:

$$Y_{\text{in}} = \frac{1}{R} + jB. \tag{2.53}$$

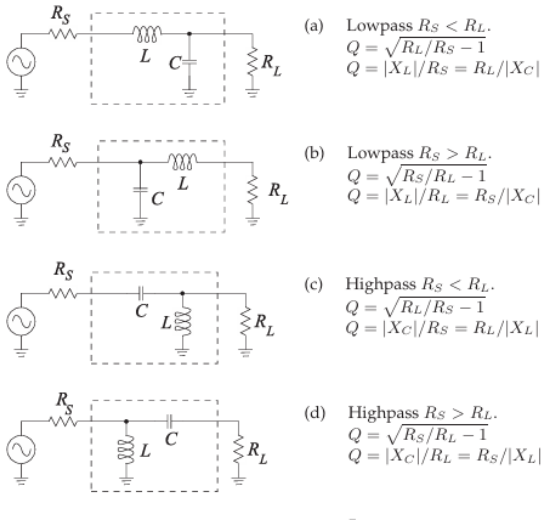

Figure 2.7: L-Matching Networks [14].

To convert the parallel shunt circuit into its series equivalent, we calculate:

$$Z_{\text{in}} = \frac{1}{\frac{1}{R} + jB} = \frac{R}{1 + B^2R^2} - j\frac{BR^2}{1 + B^2R^2}. \tag{2.54}$$

From this, we observe that the equivalent series impedance is composed of a real part R_S and a reactive part X_S, expressed as:

$$R_S = \frac{R}{1 + B^2R^2}, \tag{2.55}$$

$$X_S = -\frac{BR^2}{1 + B^2R^2}. \tag{2.56}$$

Such a change suggests that the real part of the impedance, R_S, is smaller than the initial resistance R. This makes it possible to fine tune the circuit to the right impedance that is required for matching.

2.6.3 Operating Principle

The L-Matching Network is a fundamental technique for matching two different resistance levels, R_S (source) and R_L (load), using a combination of reactive elements (inductor L or capacitor C) to achieve lossless matching and maximum power transfer.

The operating principle of the L-Matching Network involves two main steps:

1. *Step 1:* Use a series (shunt) reactive element to transform a smaller (larger) resistance up (down) to a larger (smaller) value with a real part equal to the desired resistance value.

2. *Step 2:* Use a shunt (series) reactive element to resonate with (or cancel) the imaginary part of the impedance that results from Step 1.

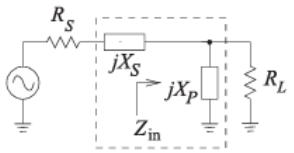

Figure 2.8: The two element matching network topology for $R_S < R_L$. X_S is the series reactance and X_P is the parallel reactance [14].

By using these steps, a resistance can be transformed to any desired value, and four possible L-Matching Networks can be constructed based on whether R_S is greater or less than R_L and whether the network is lowpass or highpass, as shown in figure 2.7.

- *Lowpass* $R_S < R_L$ (Figure 2.7-a): A series inductor and shunt capacitor are used. The quality factor Q is calculated as:

$$Q = \sqrt{\frac{R_L}{R_S} - 1} = \frac{|X_L|}{R_S} = \frac{R_L}{|X_C|} \tag{2.57}$$

- *Lowpass* $R_S > R_L$ (Figure 2.7-b): A shunt capacitor and series inductor are used to achieve matching. The quality factor is given by:

$$Q = \sqrt{\frac{R_S}{R_L} - 1} = \frac{|X_L|}{R_L} = \frac{R_S}{|X_C|} \tag{2.58}$$

- *Highpass* $R_S < R_L$ (Figure 2.7-c): A series capacitor and shunt inductor are used. The quality factor is given by:

$$Q = \sqrt{\frac{R_L}{R_S} - 1} = \frac{|X_C|}{R_S} = \frac{R_L}{|X_L|} \tag{2.59}$$

- *Highpass* $R_S > R_L$ (Figure 2.7-d): A shunt inductor and series capacitor are utilized, and the quality factor is:

$$Q = \sqrt{\frac{R_S}{R_L} - 1} = \frac{|X_C|}{R_L} = \frac{R_S}{|X_L|} \tag{2.60}$$

The matching network topology for $R_S < R_L$ is shown in figure 2.8. The input impedance Z_{in} is given by:

$$Z_{in} = \frac{R_L(jX_P)}{R_L + jX_P} = \frac{R_L X_P^2}{R_L^2 + X_P^2} + j\frac{X_P R_L^2}{R_L^2 + X_P^2}. \tag{2.61}$$

For impedance matching, the objective is the matching network to transform the source impedance Z_S to it's conjugate. From figure 2.8, the source impedance is the series configuration of resistance R_S and reactance X_S, so $Z_S = R_S + jX_S$. Therefore, the matching network needs to achieve the following:

$$Z_{\text{in}} = Z_S^* = R_S - jX_S. \tag{2.62}$$

From the equations 2.61, 2.62:

$$R_S = \frac{R_L X_P^2}{R_L^2 + X_P^2}, \tag{2.63}$$

$$X_S = -\frac{X_P R_L^2}{X_P^2 + R_L^2}. \tag{2.64}$$

From these, we can also write:

$$\frac{R_S}{R_L} = \frac{1}{\left(\frac{R_L}{X_P}\right)^2 + 1}, \tag{2.65}$$

$$\frac{X_S}{R_S} = -\frac{R_L}{X_P}. \tag{2.66}$$

The quantities Q_S and Q_P are defined as:

$$Q_S = \frac{|X_S|}{R_S}, \tag{2.67}$$

$$Q_P = \frac{R_L}{|X_P|}. \tag{2.68}$$

Finally, the design equations for $R_S < R_L$, from equations 2.65, 2.66, are given as:

$$|Q_S| = |Q_P| = \sqrt{\frac{R_L}{R_S} - 1}. \tag{2.69}$$

X_P and X_S will be either capacitive or inductive and they will have the opposite sign. R_S and R_L are determine, as source and load impedance respectively, so the Q factor of the network and thus the bandwidth is defined. Therefore, the designer does not have a choice of circuit's Q factor.

For $R_S > R_L$, the design equations are similarly defined and are

$$|Q_S| = |Q_P| = \sqrt{\frac{R_S}{R_L} - 1}, \tag{2.70}$$

$$Q_S = -Q_P, \tag{2.71}$$

$$Q_S = \frac{X_S}{R_L},\ Q_P = \frac{R_S}{X_P}. \tag{2.72}$$

2.6.4 Impedance Matching with Reactive Source and Load

There are two primary approaches to handling complex sources and loads:

- *Absorption Method*: This method absorbs the reactances from the source and load into the matching network. By transforming the reactive source and load to inductor or capacitor in series or parallel with resistance, the reactive components of the source and load can be included to the matching network.
- *Resonance Method*: Source and load reactances are cancelled through resonance, adding reactive ellements (inductors or capacitors) in parallel or in series to them.

In both approaches, the main objective is to maximize the bandwidth of the matching network by minimizing energy storage in reactive components. Generally, the total energy stored is proportional to the sum of the reactances. Properly incorporating these reactances into the matching network ensures optimal performance. However, when reactances are too large, absorption alone may not be sufficient. In such cases, a combination of resonance and absorption is used to achieve wide bandwidth and efficient matching.

2.7 Monolithic Transformer

2.7.1 Operating Principle

A passive transformer operates based on the principle of mutual inductance between two or more windings, or conductors. It enables alternating current (AC) to be transferred from one winding to the other while suffering minimal power loss. The impedance between the windings is transformed, which provides the opportunity for the correct transmission of the signal (impedance matching). This impedance transformation is described as a ratio between terminal voltage and the change in current across the windings. Additionally, transformers block direct current (DC) and allow different bias potentials at the primary and secondary windings.

To emulate a lumped inductance in microwave circuits, designers often use an electrically short transmission line. In this context, the guided wavelength, λ_g, is significantly smaller than the physical length of the transmission line. The input impedance of such a line, when short-circuited at one end, can be expressed as:

$$Z_{\text{in}} \approx r + j\omega l, \tag{2.73}$$

r represents the series resistance and l denotes the series inductance of the transmission line.

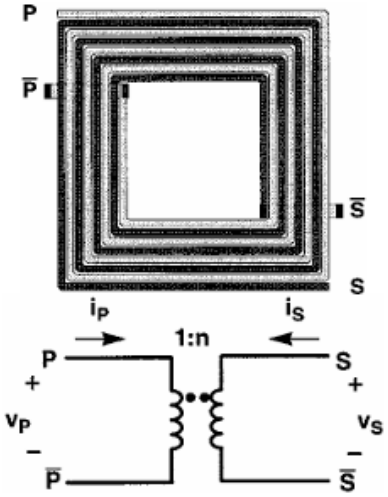

Figure 2.9: Monolithic transformer: (a) Physical layout. (b) Schematic symbol, [15].

On-chip, a microstrip line is commonly used to implement a transmission line inductor, often wound into a spiral to minimize the chip's footprint [16]. Extending this concept, microstrip spiral inductors can be intertwined to magnetically couple separate conductors, effectively forming a monolithic transformer.

A planar monolithic transformer constructed from interwound metal conductors is shown in Figure 2.9. Alternating current i_P, which flows to the primary winding, produces an alternating magnetic flux, which induces an alternating current i_S to the secondary winding. The main parameters of interest for a designer in a transformer are the turns ratio n (number of turns of secondary winding devided by the number of turns of primary winding) and the magnetic coupling coefficient k. The turns ratio, which defines the voltage and current relationship between the primary and secondary windings, is:

$$n = \frac{V_S}{V_P} = \frac{I_P}{I_S} = \sqrt{\frac{|Z_S|}{|Z_P|}} = \sqrt{\frac{L_S}{L_P}}. \tag{2.74}$$

The strength of the magnetic coupling between windings is defined by the coupling factor k:

$$k = \frac{M}{\sqrt{L_P L_S}}, \tag{2.75}$$

M is the mutual inductance between the primary and secondary windings, and L_P and L_S are the self-inductances of the primary and secondary windings, respectively.

If the magnetic coupling between the windings is perfect (no leakage of magnetic flux), the coupling factor k is unity, while uncoupled coils have a k-factor of zero. There is poor confinement of the magnetic flux in a monolithic transformer and $M < \sqrt{L_P \cdot L_S}$. Thus, the k-factor is always substantially less than one for a monolithic transformer, however, coupling coefficients as high as 0.9 are realizable on-chip.

The self-inductance of a given winding is the inductance measured at the transformer terminals with all other terminals open circuited. This enables the use of Z-parameters to calculate the electrical parameters of a transformer.

$$L_{P,S} = \frac{\mathrm{Im}\{Z_{11,22}\}}{\omega}, \tag{2.76}$$

$$Q_{P,S} = \frac{\mathrm{Im}\{Z_{11,22}\}}{\mathrm{Re}\{Z_{11,22}\}} = \frac{\omega \cdot L_{P,S}}{r_{P,S}}, \tag{2.77}$$

$$M = \frac{\mathrm{Im}\{Z_{12}\}}{\omega}, \tag{2.78}$$

$$k = \frac{M}{\sqrt{L_P \cdot L_S}} = \frac{\mathrm{Im}\{Z_{12}\}}{\sqrt{\mathrm{Im}\{Z_{11}\} \cdot \mathrm{Im}\{Z_{22}\}}}, \tag{2.79}$$

$Q_{P,S}$ are the quality factors of primary and secondary windings respectively.

2.7.2 Figures of Merit

To be able to compare the different transformer designs and evaluate their performance, several metrics have been used. In its simplest form, a transformer can be considered as two

inductors coupled together. Therefore, fundamental performance parameters of transformers are also similar to those of inductors which are primarily, the Q-factor (Q) and self-resonance frequency (f_{SR}).

Apart from these, there are other parameters which describe the performance of monolithic transformers, including their coupling factor (k), mutual inductance (M), insertion loss and power gain (G), area occupied, and transformer's characteristic resistance (TCR). Each of these factors is inter-related to one another, which creates multiple design trade-offs for monolithic transformers.

Q-Factor of Primary and Secondary Coil

Primary and secondary coils of a transformer each have their own Q-factors, denoted as Q_P and Q_S, respectively, and they collectively determine the overall performance of the on-chip transformer. The quality factor of a passive device is of great importance, as it indicates all the losses and the efficiency of the device. Therefore, one of the main targets when designing monolithic transformers is to maximize the Q-factor for both of the coils. The Q-factor of a transformer, can be given from equation 2.77.

Essentially, the Q-factor is the ratio of energy stored in a coil and energy dissipated in its series resistance, and can be improved by decreasing the series resistance. At high frequencies, additional phenomena like the skin effect and proximity effect need to be considered, which set critical limitations on the maximum Q-factor of a transformer. On the other hand, substrate loss is likely to be one of the major loss mechanisms in inductive devices, imposing a fundamental upper boundary on the Q-factor. Fabricating the transformer on a higher resistivity substrate or by shielding the conductors from the substrate has been shown to increase the Q-factor of on-chip transformers.

Coupling Coefficient, k-factor

The *coupling coefficient* k is a direct measure of the magnitude of magnetic coupling between the two coils of a transformer, making it an essential parameter for assessing transformer performance. Greater overlap area and tighter packing of the transformer's coils contribute to a higher coupling coefficient. The formula for the coupling coefficient is given by equation 2.79. The mutual inductance between the coils is represented by M, and L_P and L_S denote the self-inductances of the primary and secondary coils, respectively. The self-inductances are largely influenced by the coil length, remaining relatively constant if the length is unchanged.

The coupling coefficient k effectively quantifies the mutual inductance between primary and secondary coils. Stacked transformers, due to their compact geometry, exhibit higher coupling coefficients. Conversely, planar interleaved transformers generally exhibit lower k values due to less compact arrangements.

There is an inherent trade-off between the voltage transformation ratio n and k. For applications necessitating substantial voltage or impedance transformation ratios, significant

variations in the turn count of primary versus secondary coils reduce the overlap area, increasing leakage flux and consequently lowering k-factor.

Self-Resonance Frequency (f_{SR})

Like inductors, transformers also resonate at specific frequencies known as the self-resonance frequency of the transformer. One of the main design criteria for RF-CMOS monolithic transformers is to make this resonance frequency sufficiently high to enhance the operating bandwidth of the device. Transformers resonate at these frequencies due to the existence of parasitic capacitances:

- Between the substrate and the coil, which is directly proportional to the area of the coil facing the substrate.
- Between the two coils of the transformer.

These capacitances cause the transformer to resonate at a frequency beyond which capacitive effects become dominant over inductive effects. For high-frequency wireless applications, the self-resonance frequency (f_{SR}) is typically in the order of GHz and can be expressed as:

$$f_{SR} = \frac{1}{2\pi\sqrt{LC_P}}, \tag{2.80}$$

L represents the total inductance of the transformer, which includes both self-inductances and mutual inductance, and C_P is the total parasitic capacitance present in the transformer.

Power Gain (G)

In its ideal form, a transformer is a lossless device that outputs the exact amount of power it receives at its input. However, in practical scenarios, there are inherent losses (i.e., insertion losses) that reduce the output power. The power gain G of a transformer, defined as the ratio of output to input power, is a crucial metric for assessing the performance of monolithic transformers. The analytical formula for power gain is:

$$G = 1 + 2\left(x - \sqrt{x^2 - x}\right), \tag{2.81}$$

$$x = \frac{\mathrm{Re}(Z_{11})\,\mathrm{Re}(Z_{22}) - |\,\mathrm{Re}(Z_{12})|^2}{|\,\mathrm{Im}(Z_{12})|^2 + |\,\mathrm{Re}(Z_{12})|^2}. \tag{2.82}$$

This relation employs the Z-parameters of the device to compute the maximum power gain and is particularly applicable when the monolithic transformer is modeled as a two-port network.

2.7.3 Transformer's Characteristic Resistance (TCR)

Transformer's Characteristic Resistance (TCR) is another parameter of comparison for on-chip transformers. It is used to characterize on-chip transformers when they are used as tuned loads. TCR incorporates both Q and k parameters of a transformer to evaluate its behavior, as shown in the following equation

$$\mathrm{TCR} = \omega Q_{\mathrm{EQ}} L_{\mathrm{EQ}}. \tag{2.83}$$

Q_{EQ} and L_{EQ} are the equivalent quality factor and inductance of a transformer given by:

$$Q_{\mathrm{EQ}} = Q_1 \frac{k^2 Q_1 Q_2}{1 + k^2 Q_1 Q_2}, \quad L_{\mathrm{EQ}} = L_1 \left(1 + \frac{1}{Q_1^2} + k^2 \frac{Q_2}{Q_1}\right). \tag{2.84}$$

Q_1, Q_2 are the quality factors of the primary and secondary coil.

When used as a tuned load in a circuit, the capacitive reactance of a transformer can balance out the inductive reactance, and the circuit can be loaded purely by the resistive component. This resistive component is also called the input parallel resistance of a transformer and can be a measure of the TCR, which is roughly double the value of the parallel input resistance. TCR should be taken into account while designing the geometry and dimensions of a transformer, since maximization of the TCR is also helpful to optimize the available output power and gain.

Even though the TCR parameter is used by some researchers to characterize the device's performance, still it is less commonly used amongst all other parameters.

2.7.4 Physical Layout

To ensure compatibility with standard integrated circuit (IC) fabrication processes, the design of on-chip transformers presents a unique set of challenges. Standard IC processes typically favor *planar, thin-film metallization* techniques, which limit the transformer layout to two-dimensional structures, unlike more versatile three-dimensional topologies. As a result, *planar transformer topologies* are predominantly used in on-chip designs.

However, planar configurations introduce significant limitations:

1. *Increased Chip Area*: The physical constraints of planar layouts require larger substrate areas, increasing the overall chip size.

2. *Low Coupling Efficiency*: The magnetic coupling between windings is often less efficient in planar designs, leading to lower transformer performance and energy transfer.

In order to overcome these challenges, various layout configurations have been provided based upon the need of the application. The two most common topologies are the *interleaved planar layout* and *stacked layout*.

The *interleaved planar layout* is a design method commonly used in on-chip transformers where two planar coils are constructed in an interleaved manner on a silicon chip. Instead of having two coils separately placed, the interleaving method positions one coil in between the other. This arrangement improves the magnetic coupling between the coils, resulting in better transformer performance.

In figure 2.10, we can observe the physical layout of a 2-turn interleaved transformer in (a), and a center-tapped configuration functioning as a balun (balanced to unbalanced), which is shown in (b). Interleaved transformers typically use a *1:1 turn ratio*, where the primary and

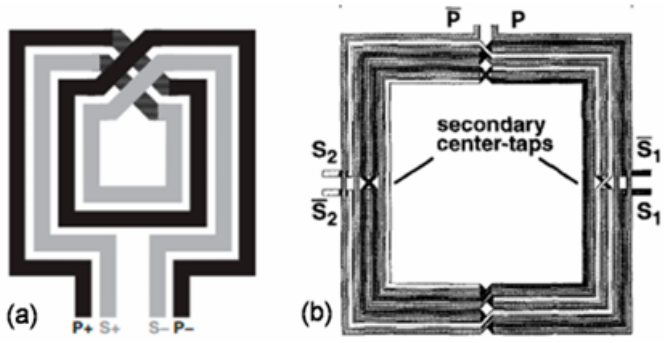

Figure 2.10: Structure of the interleaved transformer: (a) Simple layout of a square spiral interleaved two-turn transformer; (b) Center-tapped interleaved transformer functioning as a balun (balanced to unbalanced, or differential to single ended) [17].

secondary coils have matching lengths. This design enhances coupling and minimizes magnetic flux leakage between the coils.

However, in cases where transformers require different turn ratios (such as *1:N* or *N:1*), the coil lengths between the primary and secondary will differ. This makes magnetic coupling to be poor because of high leakage flux; therefore, the transformers possess lower *coupling factor* than transformers with equal turn ratios.

Despite the benefits of interleaving, a notable drawback is the larger chip area required, which makes the layout less efficient in terms of silicon utilization. Nevertheless, interleaved transformers remain a popular choice for certain RF circuits, such as LNAs (low-noise amplifiers), where they function as center-tapped baluns to convert balanced signals (differential ended) into unbalanced ones (single ended).

Several techniques have been developed to enhance the performance of interleaved transformers. One such method is the use of a *porous silicon layer* between the transformer and the substrate, which acts as a shield to reduce parasitic coupling between the metal layers and the substrate. This shielding method improves the self-resonance frequency and gain of the transformer.

Another effective technique is the *multipath technique*, which involves reducing the width of the inner paths in a step-wise manner. This approach helps to decrease the effects of *skin* and *proximity losses*, allowing for a more constant length-to-width ratio and ensuring even current distribution throughout the transformer loops. These improvements have been shown to increase both the coupling factor and overall gain of the transformer.

In conclusion, the *interleaved planar layout* has the following characteristics:

1. *Simple to Fabricate*: Since the interconnects are in the same metallization plane, the proposed interleaved structure is simpler to implement.

2. *Larger Area*: However, simplicity has its drawback; it takes a much larger area on the chip, and therefore is less area-efficient.

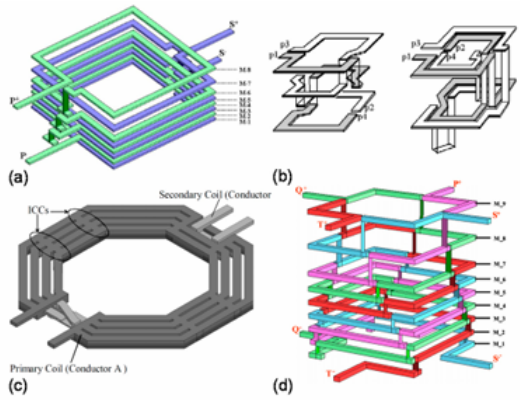

Figure 2.11: Different types of *stacked layout* configuration [17].

3. *Lower Magnetic Coupling*: The interleaved layout in most cases leads to suboptimal coupling between the windings and hence restricts transformer performance.

Interleaved layouts were confronted with some issues such as lower coupling factors large chip area, hence the *stacked layout configuration* was developed to overcome these challenges. In the stacked layer configuration, two or more metal layers are positioned in a way that is one on top of the other (stacked) resulting in a 3D structure that greatly improves magnetic coupling. This configuration keeps the total flux of the magnetic field within the structure, and thereby increases coupling of the coils.

The coupling involved with stacked layout is better, and this takes up a smaller area than the interleaved structures. When the layers of metal are packed closely, the magnetic interaction between the layers is enhanced by providing a high value of coupling coefficient. However, there are also some problematic aspects which are characteristic of the stacked layout. The major issue is the issue of *high parasitic capacitances* because of metal layers which are placed in close vicinity to each other and this negatively impact the self-resonant frequency of the transformer, thus its performance at high frequencies.

Despite these challenges, the stacked layout is becoming more popular in *RF IC applications*, particularly because it offers a better balance between coupling efficiency and area usage. Some examples of stacked layout configuration are shown in figure 2.11. Figure 2.11-(a), shows the typical stacked layout configuration [18], and Figures 2.11-(b), (c), (d) show improved stacked configurations like transformers with half-coil [19], inter-coil connection [20] and quad-coil [21], respectively.

Stacked layout arranges the windings across different metal layers, effectively reducing the substrate area. Stacked layout configuration offers:

- *Improved Magnetic Coupling*: In order to achieve an even better performance the layout is stacked so that the windings are put closer and closer across several metal layers.
- *Smaller Area*: This stacking of layers minimizes space needed for the transformer on the chip.

- *Fabrication Complexity*: Despite the improvements in coupling and area, the stacked layout introduces more *fabrication complexity*, increasing parasitic capacitances between the layers and potentially degrading performance.

General guidelines for the design of the layout of a transformer (or balun) are the following:

- Thicker metal reduces the ohmic losses in the primary and secondary windings.
- Insulating oxide layer reduces parasitic coupling to the semiconducting substrate, thereby improving the operating bandwidth and reducing in-band losses caused by shunt parasitic elements.
- Losses due to the substrate are further reduced through the use of a relatively high resistivity substrate.
- Minimum conductor spacing and metal width improve magnetic coupling (k-factor) between primary and secondary windings at the cost of interwinding capacitance.
- There is a large improvement in k-factor as the number of turns is increased from one to two, because of coupling between adjacent lines. The number of metal lines running in parallel increases with the number of turns, and because the magnetic coupling decreases quickly with the separation between parallel conductors, the k-factor quickly tends to a limiting value for a given metal width.
- k-factor decreases as the turns ratio is increased.
- The total inductances on the primary and secondary windings scale approximately according to equation 2.74.
- The self-inductance and associated parasitics are determined by the length of the winding and increase with increasing winding length.
- For relatively low terminal impedances (on the order of the conductor ohmic losses), the series parasitics (self inductance and resistance of windings) dominate the response and so wider conductors lines with relatively large parasitic capacitances may be used. When terminal impedances on the order of hundreds of ohms are used, shunt parasitics play a more important role in the overall response and therefore narrower conductor widths should be used.

From this analysis, the transformer layout design is seen to play a critical role in enhancing the performance of the on-chip transformers.

2.7.5 Modeling and Characterization

To understand the transformer characteristics deeply, a circuit model is proposed, as shown in figure 2.12. The transformer core consists of two individual inductors (L_P and L_S) and the associated mutual inductance (M). The parasitic line effect is depicted with the series resistance (R_P and R_S). Notably, the parallel capacitor (C_{21}) between the primary and the secondary is used to describe the coupling effect of the transformer at high operating frequencies. Moreover, the substrate effect is depicted by the parallel connection of the resistor and the capacitor between the port terminal and the ground.

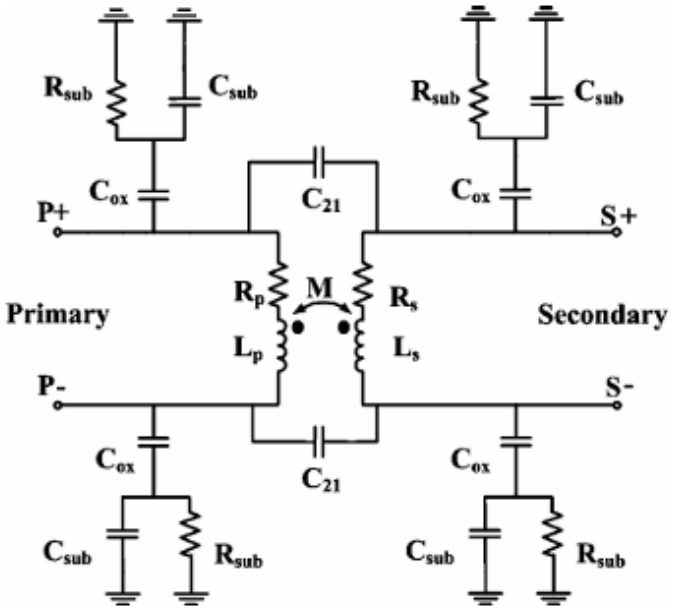

Figure 2.12: Equivalent circuit of on-chip transformer [22].

2.8 Balun

The $1 : n$ transformers described in the previous section 2.7 consist of two independent windings (or conducting filaments) and are classified as bifilar transformers. Multifilament transformers can also be constructed on-chip. These devices are used to implement power dividers/combiners and baluns.

A balun is a device which couples a balanced circuit (differential ended) to an unbalanced one (single ended). There are many structures used to implement baluns at RF and microwave frequencies, although a differential amplifier is the most commonly used method for unbalanced-to-balanced signal conversion on-chip. Microwave balun structures require physical dimensions on the order of the signal wavelength, and so these devices consume too much chip area when operating below approximately 15 GHz [15].

Fig 2.13 shows the schematic of the proposed 5-port extended balun model. In this model, the primary inductor has series inductance L_p, series resistance R_p, and self-capacitance C_p, while the secondary inductor has L_s, R_s, and C_s. The primary and secondary inductors have a mutual inductance M, and the mutual capacitance between the primary and secondary inductor is denoted by C_m.

For convenience, it is assumed that the inductors are equally split into two, and the polarity signs are on the top side for all inductors unless otherwise specified. Parasitic capacitances other than C_p, C_s, and C_m won't be considered.

The proposed model has an added impedance Z_c at the center tap. In balun design, a proper center tap loading is usually desired for optimal performance.

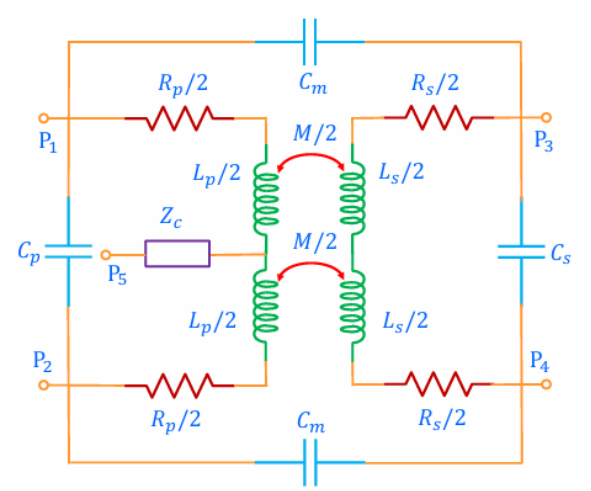

Figure 2.13: Extended Balun Model [23].

In this model, mutual resistance is not included since it is typically negligible. Additionally, the center tap at the primary is only considered, and generalization to a center tap at the secondary is trivial.

2.9 Differential-Ended Amplifier Design

Differential-ended design offers several advantages compared to single-ended design, which make it a preferred choice in most applications [24].

2.9.1 Greater Output Swing

Differential circuits provide a higher output swing, approximately double the amplitude, compared to single-ended signal chains. This is attributed to the symmetric nature of the differential design, where the two terminals exhibit voltage swings with a 180° phase difference. This design enables:

- Achieving higher overall signal swings.
- Delivering the same overall signal swing with a lower supply voltage.
- Reducing power dissipation.

2.9.2 Linearity

Differential circuits inherently reject even-order harmonics because of their odd-signed nature of the input/output transfer characteristic. The differential circuit in essence is two single-ended circuits with 180° phase difference between the inputs as well as the outputs. What this means is that between the input/output characteristics of the two single-ended circuits occurs $y(-x) = -y(x)$. For this to be valid, all the even-order harmonic terms must be zero in equation 3.6. This property compels only odd harmonics to be in place, and thus evidences a profound enhancement in linearity. The output signal $y(t)$ of a differential circuit,

driven by a differential input, can be expressed as:

$$y(t) = \alpha_1 x(t) + \alpha_3 x^3(t) + \alpha_5 x^5(t) + \dots$$

where all the terms of even order $\alpha_{2 \cdot j} = 0$ for every value of j. As a result:

- All even-order harmonics are eliminated.
- The same voltage rails applied to the design exemplify an increase $\approx 3\,\text{dB}$ in $OP_{1\text{dB}}$ and OIP_3.

2.9.3 Noise Immunity

Differential designs offer superior noise rejection compared to single-ended configurations. Some of the key benefits include:

- Signal return paths no longer rely on the ground, reducing ground-loop noise.
- Excellent common-mode rejection ratio (CMRR), which reduces sensitivity to external noise sources.
- Improved power supply rejection ratio (PSRR), which minimizes disturbances from supply variations.
- Enhanced immunity to coupled electromagnetic interference (EMI).

2.9.4 Design Considerations

Although differential circuits may need approximately twice the area of single-end counterparts that may be seen as a disadvantage, all the pro's easily outweigh them: higher linearity, better rejection of noises, and the capability to swing higher levels of signals. The differential-ended design yields important enhancements in the measures of output swing, linearity, and noise. These characteristics make it more desirable for today's demanding RF and mixed signal designs.

Chapter 3

Fundamentals of Power Amplifiers

3.1 Operating Principles

In modern telecommunication systems, the RF PA plays a pivotal role, serving as the final active component in the transmitter chain before the signal is radiated through the antenna. Its primary function is to amplify the power of the RF signal to the required level for transmission, ensuring that the signal reaches the receiver with sufficient strength. However, this amplification process is not without challenges. The RF PA must balance the often conflicting requirements of efficiency and linearity, both of which are critical for effective and reliable communication.

Efficiency refers to the PA's ability to convert the DC power it consumes into RF power, with higher efficiency meaning less energy is wasted as heat. This is particularly important in modern systems where power consumption and thermal management are key concerns.

Linearity, on the other hand, ensures that the amplified signal remains a faithful reproduction of the input, free from distortion that could cause interference and spectral regrowth. The inherent trade-off between these two parameters makes PA design one of the most complex tasks in RF engineering.

RF PAs are classified into various types, such as **Class A, B, AB, C**, and others, based on their operation and efficiency characteristics. Each class represents a different approach to managing the efficiency-linearity trade-off:

- **Class A**: This class offers the best linearity, making it ideal for applications where signal fidelity is paramount. However, Class A amplifiers operate with a bias current that is constant throughout the entire signal cycle, resulting in low efficiency. This makes Class A amplifiers less suitable for applications where power consumption is a critical factor.
- **Class B**: Class B amplifiers improve efficiency compared to Class A by operating each transistor for only half of the signal cycle (180 degrees of the waveform). However, the drawback is that the transition between the active devices can introduce distortion, affecting linearity.
- **Class AB**: Class AB amplifiers aim to strike a balance between the linearity of Class A and the efficiency of Class B. By biasing the transistors slightly above threshold voltage V_{t}, they reduce distortion while maintaining higher efficiency than Class A amplifiers. This makes Class AB amplifiers a popular choice for many RF applications.
- **Class C**: Class C amplifiers offer even higher efficiency by operating for less than 180 degrees of the signal cycle. However, this comes at the cost of significant nonlinearity, making Class C amplifiers suitable only for applications where linearity is not a primary

concern, such as in certain types of RF transmitters where harmonic distortion can be managed.

- **Other Classes (D, E, F, etc.)**: These classes use various techniques such as switching operation and harmonic tuning to achieve even higher efficiencies, often at the cost of increased circuit complexity and reduced linearity. These amplifiers are typically used in specialized applications where efficiency is critical, and signal distortion can be managed or corrected.

3.2 The Role of Power Amplifiers in the Transceiver Chain

The transceiver chain in a communication system consists of both a transmitter and a receiver, each playing a critical role in the transmission and reception of signals. The following is a brief description of each block within the transceiver chain, providing a complete picture of how signals are processed from generation to reception, as illustrated in figure 3.1.

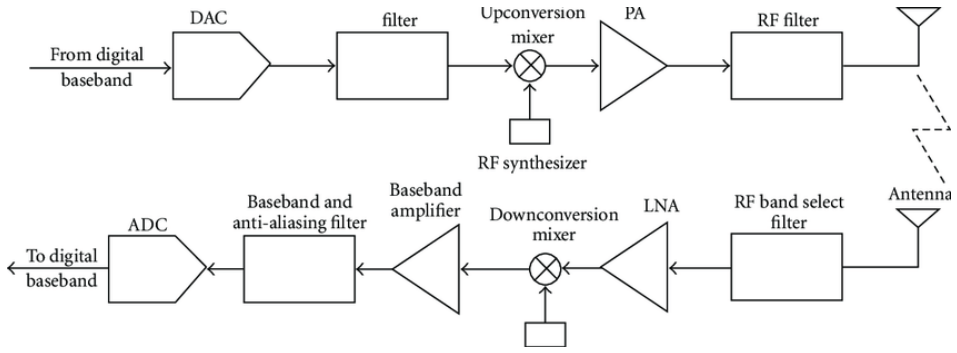

Figure 3.1: Block diagram of the transceiver analog signal chain [25].

Transmitter Chain:

- **Digital-to-Analog Converter (DAC)**:
  The DAC converts the digital baseband signal, which is the processed data in digital form, into an analog signal that can be further processed by the analog components of the transmitter.
- **Filter**:
  After the DAC, the signal passes through a filter to remove any high-frequency components that may have been introduced during the digital-to-analog conversion. This ensures that the signal is within the desired bandwidth and reduces potential interference.
- **Upconversion Mixer**:
  The filtered analog signal is then fed into an upconversion mixer. The mixer combines the baseband signal with a high-frequency signal generated by the RF synthesizer, converting the baseband signal to the desired RF frequency suitable for transmission.
- **Power Amplifier (PA)**:

The RF signal from the mixer is usually too weak to be transmitted effectively over long distances. The Power Amplifier (PA) plays a crucial role here by boosting the power of the RF signal to a level sufficient for transmission through the antenna. The PA is a vital component of the transmitter chain, ensuring that the signal reaches the receiver with adequate strength while maintaining efficiency and linearity.

- **RF Filter**:
  After amplification, the signal passes through an RF filter, which removes any unwanted harmonics or spurious signals generated during amplification. This filter ensures that only the desired RF signal is transmitted, reducing interference with other communication channels.
- **Antenna**:
  Finally, the amplified and filtered RF signal is transmitted through the antenna, which radiates the signal into the air for reception by the intended receiver.

Receiver Chain:

- **RF Band Select Filter**:
  On the receiver side, the incoming RF signal from the antenna first passes through an RF band select filter. This filter selects the desired frequency band, filtering out unwanted signals and noise from other frequency bands.
- **Low-Noise Amplifier (LNA)**:
  The filtered signal is then amplified by a Low-Noise Amplifier (LNA), which boosts the signal strength while minimizing the addition of noise. The LNA is essential for improving the sensitivity of the receiver.
- **Downconversion Mixer**:
  The amplified RF signal is then fed into a downconversion mixer, where it is combined with a signal from an RF synthesizer. This process shifts the RF signal back down to the baseband or IF, making it easier to process.
- **Baseband Amplifier**:
  After downconversion, the signal is further amplified by a baseband amplifier to ensure that it is strong enough for further processing in the digital domain.
- **Baseband and Anti-Aliasing Filter**:
  The signal then passes through a baseband and anti-aliasing filter, which removes any high-frequency components that could cause distortion or aliasing during analog-to-digital conversion.
- **Analog-to-Digital Converter (ADC)**:
  Finally, the Analog-to-Digital Converter (ADC) converts the filtered analog signal back into a digital form, enabling further digital processing and interpretation by the receiver's digital baseband unit.

In the transmitter chain, the Power Amplifier (PA) is a critical component that amplifies

the RF signal to a power level that is sufficient for effective transmission over the air. The PA ensures that the signal reaches the receiver with enough power to be accurately received and decoded, overcoming losses that occur during transmission. However, the PA must also maintain a balance between efficiency and linearity to avoid introducing distortion that could degrade the signal quality or cause interference with other communication channels. Thus, the PA plays a dual role in both strengthening the signal and preserving its integrity, making it indispensable in the overall transceiver chain.

3.3 Key Performance Metrics

3.3.1 Power Gain

Power gain is a critical parameter in RF design and is defined as the ratio of the output power of an amplifier to the input power of the amplifier. This parameter is typically expressed in decibels (dB), which provides a logarithmic measure of the gain. The power gain, G, can be calculated using the following equation:

$$G = \frac{P_{\text{out}}}{P_{\text{in}}}, \tag{3.1}$$

P_{out} is the output power and P_{in} is the input power.

To express the power gain in decibels (dB), the following equation is used:

$$G_{dB} = 10 \log \left(\frac{P_{\text{out}}}{P_{\text{in}}} \right). \tag{3.2}$$

The power gain provides a direct measure of how much an amplifier increases the power of a signal, and it is a fundamental parameter in determining the efficiency and performance of RF amplifiers.

3.3.2 Efficiency

Efficiency is a crucial metric in power amplifier design, representing how effectively the power from the supply is converted into useful output power. Two common measures of efficiency are drain efficiency (η) and Power-Added-Efficiency (PAE).

Drain efficiency, sometimes referred to as DC-to-RF efficiency, is defined as the ratio of the RF output power (P_{out}) to the total DC power drawn from the supply (P_{dc}). Mathematically, it is expressed as:

$$\eta = \frac{P_{\text{out}}}{P_{\text{dc}}}. \tag{3.3}$$

DC power is calculated as

$$P_{\text{dc}} = V_{\text{CC}} \cdot I_{\text{dc}}, \tag{3.4}$$

V_{CC} is the supply voltage, and I_{dc} is the DC harmonic of the current waveform.

Power-Added Efficiency (PAE) is a similar concept to drain efficiency but takes into account the gain of the amplifier. It is defined as:

$$\text{PAE} = \frac{P_{\text{out}} - P_{\text{in}}}{P_{\text{dc}}} = \eta \left(1 - \frac{1}{G}\right). \tag{3.5}$$

For high-gain amplifiers, PAE is nearly the same as drain efficiency, but as the gain decreases (especially in the compression region), PAE also decreases.

Figures 3.2 and 3.3 illustrate the relationship between drain efficiency, PAE, and power gain.

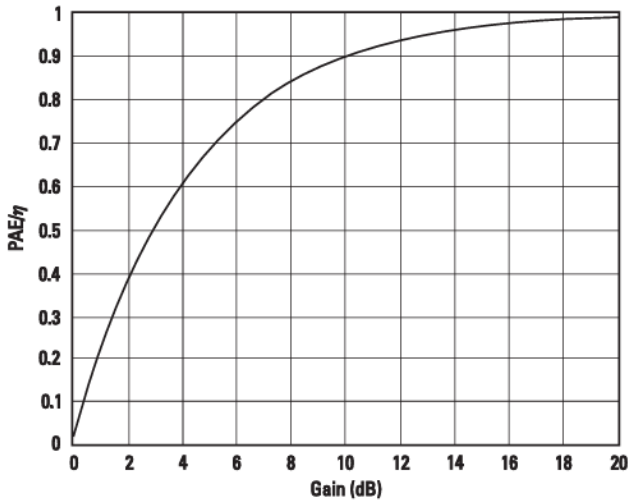

Figure 3.2: Normalized power-added efficiency versus gain. [26]

Figure 3.2 shows the normalized PAE as a function of gain. As gain increases, PAE approaches the drain efficiency (η). This is because, at higher gains, the input power P_{in} becomes negligible compared to the output power P_{out}, making PAE nearly equal to η.

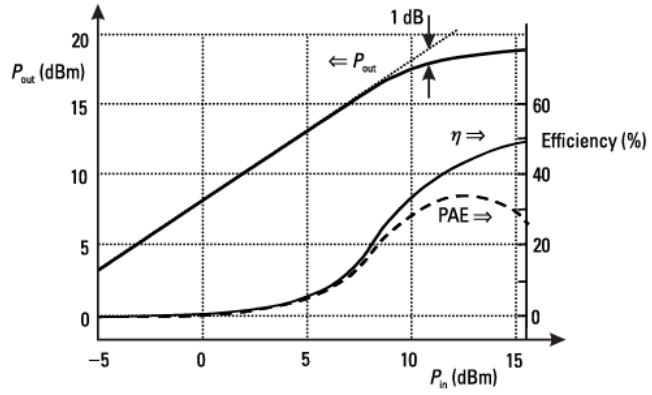

Figure 3.3: Output versus input power. [26]

Figure 3.3 illustrates the output power, efficiency, and PAE versus input power. The graph shows that while efficiency (η) continues to increase with input power, PAE peaks and then

decreases as the amplifier enters the compression region. The peak PAE typically occurs just a few dB beyond the 1-dB compression point, indicating the optimal operating point of the amplifier.

These figures underscore the importance of prioritizing PAE in amplifier design, especially when seeking optimal performance in power-constrained environments. Unlike drain efficiency (η), which only considers the output power delivered to the load, PAE accounts for both the input power and the output power. This makes PAE a more comprehensive measure of efficiency, as it reflects the amplifier's overall effectiveness, including the power provided by the previous stage or source. By incorporating the input power into the efficiency calculation, PAE offers a more accurate assessment of the amplifier's performance in real-world scenarios.

3.3.3 Harmonic Distortion and 1-dB Compression Point

When a sinusoid is applied to a nonlinear system, the output generally exhibits frequency components that are integer multiples (harmonics) of the input frequency. This behavior can be described mathematically by considering a memoryless nonlinear system with an input/output characteristic approximated by [27]:

$$y(t) \approx \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t). \tag{3.6}$$

If the input signal is a simple cosine wave $x(t) = A\cos(\omega t)$, the output can be expressed as:

$$y(t) = \alpha_1 A\cos(\omega t) + \alpha_2 A^2\cos^2(\omega t) + \alpha_3 A^3\cos^3(\omega t). \tag{3.7}$$

Using the following trigonometric identities:

$$\cos^2(\omega t) = \frac{1+\cos(2\omega t)}{2} \tag{3.8}$$

$$\cos^3(\omega t) = \frac{3\cos(\omega t)+\cos(3\omega t)}{4}, \tag{3.9}$$

the output equation expands to:

$$y(t) = \alpha_1 A\cos(\omega t) + \frac{\alpha_2 A^2}{2}\left[1+\cos(2\omega t)\right] + \frac{\alpha_3 A^3}{4}\left[3\cos(\omega t)+\cos(3\omega t)\right]. \tag{3.10}$$

Simplifying further:

$$y(t) = \left[\alpha_1 A + \frac{3}{4}\alpha_3 A^3\right]\cos(\omega t) + \frac{\alpha_2 A^2}{2}\cos(2\omega t) + \frac{\alpha_3 A^3}{4}\cos(3\omega t) + \frac{\alpha_2 A^2}{2}. \tag{3.11}$$

From this expansion, we observe that the output contains components at the fundamental frequency ω, the second harmonic 2ω, and the third harmonic 3ω. The coefficients of these terms indicate the amplitude of each harmonic.

In many RF circuits, harmonic distortion is a significant concern as it can cause interference and degrade system performance. However, in some cases, like narrowband systems, harmonic distortion might be suppressed or irrelevant.

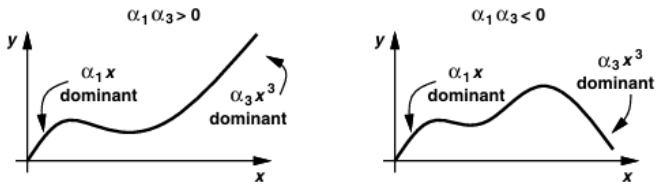

Figure 3.4: Expansive (left) and compressive (right) characteristics [27].

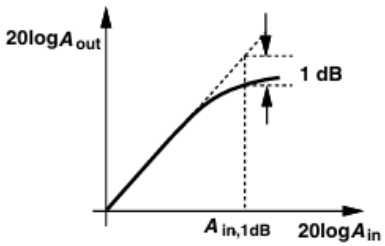

Figure 3.5: Definition of 1-dB Compression Point [27]

The small-signal gain of circuits is usually obtained with the assumption that harmonics are negligible. However, as the input amplitude A increases, the gain experienced by $A\cos(\omega t)$ is given by:

$$G(A) = \alpha_1 + \frac{3}{4}\alpha_3 A^2. \tag{3.12}$$

As A becomes larger, the gain decreases because of the $\frac{3}{4}\alpha_3 A^2$ term, indicating the onset of compression. To have compression, the product $\alpha_1\alpha_3$ must be negative. This condition results in a compressive characteristic, where the gain decreases as the input amplitude increases.

Figure 3.4 illustrates the two possible behaviors of a nonlinear system depending on the sign of the product $\alpha_1\alpha_3$:

- **Expansive Characteristic** ($\alpha_1\alpha_3 > 0$): In this scenario, the gain increases as the input amplitude increases, leading to an expansive behavior. This behavior is less common in practical RF circuits.
- **Compressive Characteristic** ($\alpha_1\alpha_3 < 0$): Here, the gain decreases with increasing input amplitude, which is the typical scenario in RF amplifiers. This characteristic is crucial for understanding gain compression, where the amplifier begins to deviate from its linear response as the input power increases.

As the input signal level A_{in} increases, the output level A_{out} initially follows the ideal linear gain (i.e., it increases proportionally with the input). However, due to the nonlinearity of the amplifier, the output begins to compress at higher input levels, deviating from the expected linear relationship.

Figure 3.5 illustrates this concept by showing the ideal linear response and the actual response of the amplifier. The point where the actual output deviates from the ideal by 1 dB is known as the 1-dB compression point, denoted as $A_{\text{in,1dB}}$. At this point, the gain has dropped by 1 dB from its small-signal value, marking the onset of significant nonlinearity.

To find the input signal level corresponding to the 1-dB compression point, we start by considering the small-signal gain of the amplifier, which is typically expressed as:

$$G = 20 \log |\alpha_1|. \tag{3.13}$$

As the input amplitude A increases, the output signal becomes compressed due to the third-order nonlinearity term $\alpha_3 A^3$. The gain at this level can be approximated by:

$$G_{\text{compressed}} = 20 \log \left(\left| \alpha_1 + \frac{3}{4} \alpha_3 A_{\text{in,1dB}}^2 \right| \right). \tag{3.14}$$

At the 1-dB compression point, the gain is reduced by 1 dB, so:

$$20 \log |\alpha_1| - 1 = 20 \log \left(\left| \alpha_1 + \frac{3}{4} \alpha_3 A_{\text{in,1dB}}^2 \right| \right). \tag{3.15}$$

We can rearrange this equation to isolate $A_{\text{in,1dB}}$:

$$\left| \alpha_1 + \frac{3}{4} \alpha_3 A_{in,1dB}^2 \right| = |\alpha_1| \cdot 10^{-1/20}. \tag{3.16}$$

Given that $10^{-1/20} \approx 0.891$, the equation becomes:

$$\left| \alpha_1 + \frac{3}{4} \alpha_3 A_{\text{in,1dB}}^2 \right| = 0.891 \, |\alpha_1| \, . \tag{3.17}$$

Expanding and solving for $A_{\text{in,1dB}}$, we get:

$$A_{\text{in,1dB}} = \sqrt{0.145 \left| \frac{\alpha_1}{\alpha_3} \right|}. \tag{3.18}$$

Equation 3.18 allows us to calculate the input signal level at the 1-dB compression point, marking the threshold beyond which the amplifier's gain starts to compress significantly, leading to nonlinear behavior.

3.3.4 Intermodulation and Third-Order Intercept Point

In RF design, understanding the effects of nonlinearity is critical, particularly when multiple signals or interferers are present. Previously, we discussed the nonlinearity effects for a single signal, leading to harmonic distortion. However, a more complex and common scenario in RF systems is when two interferers accompany the desired signal, leading to intermodulation (IM) products.

When two signals with frequencies ω_1 and ω_2 are applied to a nonlinear system, the output typically exhibits new frequency components that are not harmonics of either input signal. These components, called intermodulation products, arise from the "mixing" or multiplication of the two input signals as their sum is raised to a power greater than unity.

Assume the input signal is:

$$x(t) = A_1 \cos(\omega_1 t) + A_2 \cos(\omega_2 t) \tag{3.19}$$

The output can be expressed as:

$$y(t) = \alpha_1(A_1 \cos\omega_1 t + A_2 \cos\omega_2 t) + \alpha_2(A_1 \cos\omega_1 t + A_2 \cos\omega_2 t)^2 + \alpha_3(A_1 \cos\omega_1 t + A_2 \cos\omega_2 t)^3 \tag{3.20}$$

Expanding the right-hand side and discarding the dc terms, harmonics, and components at $\omega_1 \pm \omega_2$, we obtain the following intermodulation products:

$$\omega = 2\omega_1 \pm \omega_2 : \quad \frac{3\alpha_3 A_1^2 A_2}{4} \cos(2\omega_1 + \omega_2)t + \frac{3\alpha_3 A_1^2 A_2}{4} \cos(2\omega_1 - \omega_2)t \tag{3.21}$$

$$\omega = 2\omega_2 \pm \omega_1 : \quad \frac{3\alpha_3 A_1 A_2^2}{4} \cos(2\omega_2 + \omega_1)t + \frac{3\alpha_3 A_1 A_2^2}{4} \cos(2\omega_2 - \omega_1)t \tag{3.22}$$

The fundamental components at ω_1 and ω_2 are given by:

$$\omega = \omega_1, \omega_2 : \quad \left(\alpha_1 A_1 + \frac{3}{4}\alpha_3 A_1^3 + \frac{3}{2}\alpha_3 A_1 A_2^2\right) \cos\omega_1 t \tag{3.23}$$

$$\left(\alpha_1 A_2 + \frac{3}{4}\alpha_3 A_2^3 + \frac{3}{2}\alpha_3 A_2 A_1^2\right) \cos\omega_2 t \tag{3.24}$$

Figure 3.6 illustrates the generation of various intermodulation components in a two-tone test. Among these, the third-order IM products at $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ are particularly important because they can appear close to the original signal frequencies if ω_1 and ω_2 are near each other. This proximity can cause significant interference with the desired signal, as shown in Figure 3.7.

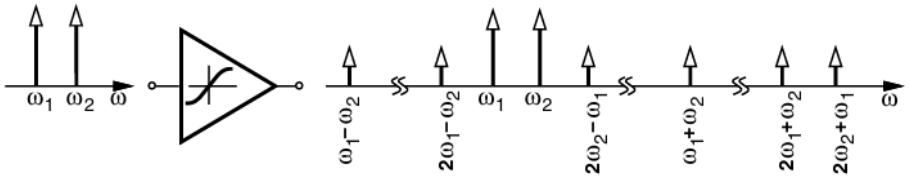

Figure 3.6: Generation of various intermodulation components in a two-tone test [27].

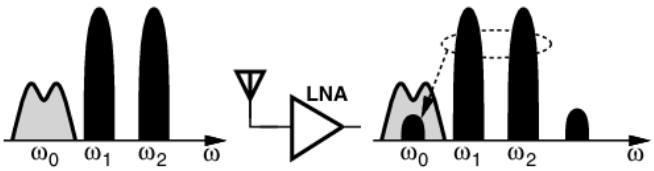

Figure 3.7: Corruption due to third-order intermodulation [27].

The two-tone test is versatile and powerful because it can be applied to systems with arbitrarily narrow bandwidths. A sufficiently small difference between the two tone frequencies ensures that the IM products also fall within the band, providing a meaningful view of the nonlinear behavior of the system. Figure 3.8 contrasts the two-tone and harmonic distortion tests, demonstrating the advantage of the two-tone test in narrowband systems.

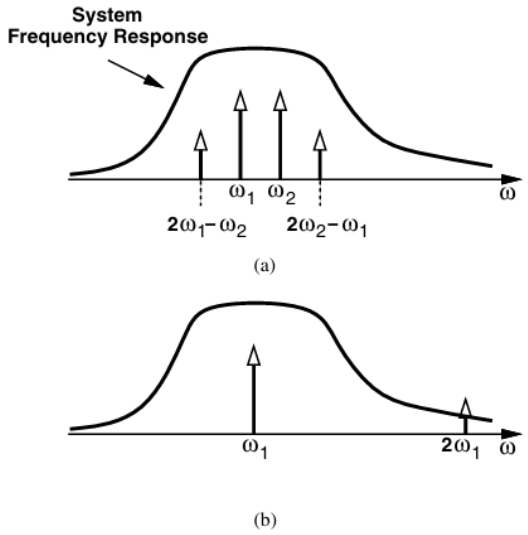

Figure 3.8: Two-tone and harmonic tests in a narrowband system [27].

Third-Order Intercept Point (IP3)

The concept of the Third-Order Intercept Point (IP3) is a key metric for assessing the linearity of RF systems. As the amplitude of each tone rises, the output IM products increase more sharply ($\propto A^3$), as show from equations 3.21, 3.22. Eventually, the amplitude of the IM products becomes equal to that of the fundamental tones at the output. This intersection point, when plotted on a log-log scale, is called the "Third-Order Intercept Point" (IP3), as depicted in Figure 3.9. The input level at which this occurs is known as the Input Third-Order Intercept Point (IIP3), while the corresponding output is represented by OIP3.

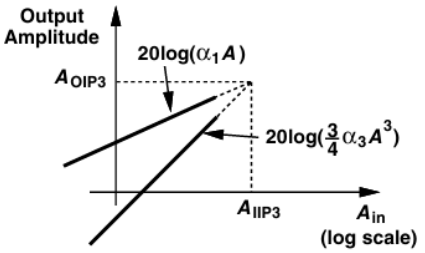

Figure 3.9: Definition of IP3 [27].

To determine the IIP3, we equate the fundamental and IM amplitudes:

$$|\alpha_1 A_{IIP3}| = \left|\frac{3}{4}\alpha_3 A_{IIP3}^3\right| \tag{3.25}$$

Solving for A_{IIP3}, we obtain:

$$A_{IIP3} = \sqrt{\frac{4}{3}\left|\frac{\alpha_1}{\alpha_3}\right|} \tag{3.26}$$

$$= \sqrt{\frac{4}{0.435}}$$

$$\simeq 9.6\,\text{dB} \tag{3.27}$$

Equation 3.27 proves helpful as a sanity check in simulations. Note that this relationship holds for a third-order system and not necessarily if higher-order terms manifest themselves. However, it is essential to note that IP3 is not a directly measurable quantity. Instead, it is an extrapolated value based on large-signal measurements, as the true intercept may exceed the supply voltage. The IP3 concept is invaluable in understanding the linearity limits of RF systems and predicting how a circuit will perform under varying input power levels.

3.3.5 Saturated Output Power

Saturated output power, denoted as P_{sat}, refers to the maximum output power that a power amplifier can deliver once it enters the saturation region. This occurs when the amplifier's P_{out} versus P_{in} curve begins to flatten, indicating that any further increases in input power P_{in} do not result in significant increases in output power P_{out}. At this point, the gain of the amplifier effectively drops to zero. Figure 3.10 illustrates this phenomenon, showing how the output power plateaus beyond a certain input level, marking the saturation threshold.

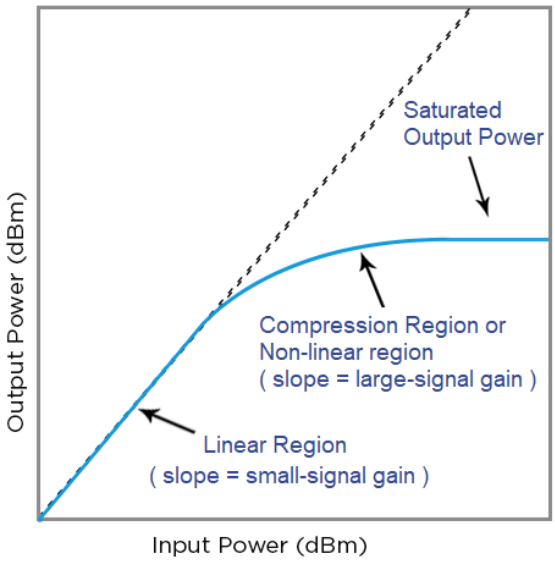

Figure 3.10: Illustration of saturated output power P_{sat} [27].

Saturated output power is a critical metric in power amplifier design, particularly in ensuring compliance with the power levels mandated by various communication standards and

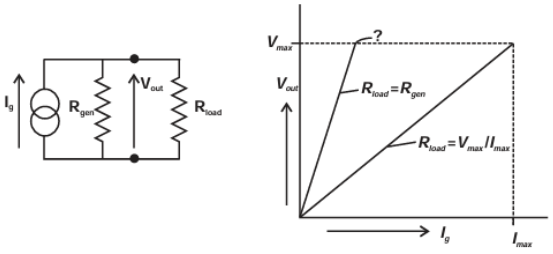

Figure 3.11: Conjugate match and loadline match [28]

protocols. These protocols specify the required power levels for different applications, ensuring effective communication and signal integrity.

Typically, power amplifiers are operated below their saturation point to maintain linearity, reduce distortion, and minimize potential issues related to heat dissipation and energy consumption. By staying within the linear operating region, the amplifier ensures consistent performance across a range of input signals, avoiding the nonlinearities that occur in the saturation region.

3.4 Power Match

In RF power amplifier (PA) design, achieving an optimal power match between the amplifier and the load is crucial for maximizing output power and ensuring efficient operation. Power matching involves tuning the load impedance to align with the optimal load that allows the amplifier to deliver maximum power (load pull analysis). This concept is particularly important in power amplifiers where linearity and efficiency are of paramount concern.

Why Power Match is Needed

Power matching is essential because it enables the amplifier to operate at its maximum output power without the constraints imposed by conjugate matching. As discussed by [28], conjugate matching is often idealized for maximizing power transfer in simple circuits but may not be practical in real-world RF amplifier designs, particularly when dealing with transistors (non-linear circuits).

In practical PA designs, it's important to consider the limitations imposed by the transistor's voltage and current handling capabilities. For example, as illustrated in Figure 3.11, applying the conjugate match theorem to a transistor could result in excessive voltage across the generator terminals, potentially exceeding the transistor's voltage rating. This situation is mitigated by using a loadline or power match, which adjusts the load impedance to ensure the transistor operates within its safe voltage and current limits and especially in its maximum voltage and current swing, in order to deliver the maximum output power to the load.

In the context of a power amplifier, power matching becomes especially important when operating in the nonlinear region, where gain compression and other nonlinear effects are significant. As shown in Figure 3.12, a power match can yield significantly higher 1 dB compression power compared to conjugate matching, making it a preferred choice for amplifiers

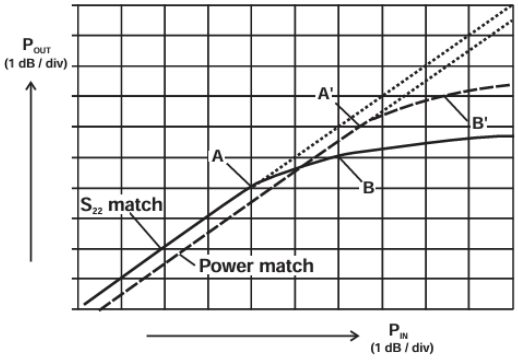

Figure 3.12: Compression characteristics for conjugate (S_{22}) match (solid curve) and power match (dashed curve). The 1 dB compression points (B, B') and maximum linear power points (A, A') show improvements under power-matched conditions [28]

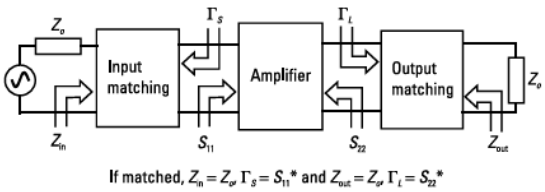

Figure 3.13: Block diagram of amplifier and matching circuits [26]

operating under high drive levels. The power match condition allows the amplifier to maintain higher output power, evidenced by the improvements in the 1 dB compression points (B, B') and maximum linear power points (A, A').

Furthermore, power matching accounts for practical realities such as gain compression and harmonic generation. While conjugate matching may be effective for small-signal operation, it does not account for the nonlinearities that arise in high-power amplifiers. Power matching, on the other hand, optimizes the load impedance, ensuring the amplifier delivers the highest possible power while minimizing distortion and other unwanted effects.

To obtain maximum output power, particularly in large-signal conditions, the amplifier is typically not conjugately matched. Instead, the load is designed to provide the correct voltage and current necessary to deliver the required power, as depicted in the block diagram of amplifier and matching circuits in Figure 3.13. This approach, known as load-pull, involves adjusting the load impedance until the optimal load Γ_{opt} is found, ensuring the amplifier delivers maximum power to the load.

3.5 Load-Pull Measurements

In RF power amplifier (PA) design, one of the crucial steps in the characterization process is understanding the relationship between output power and output matching conditions. This understanding is often facilitated through a method known as load-pull measurements. These measurements are used to optimize the performance of an RF amplifier by systematically

varying the load impedance presented to the amplifier's output and observing how this affects the output power.

Figure 3.14 illustrates a typical set of load-pull data plotted on a Smith chart. The contours in this figure represent constant output power levels, often measured in decibels (dB) relative to a reference value (the maximum achievable output power). The central point, labeled P_{opt}, represents the optimal load impedance where the output power is maximized. Surrounding P_{opt}, we find contours labeled as -1 dB and -2 dB, indicating load impedances that produce 1 dB and 2 dB less output power, respectively, compared to the optimum.

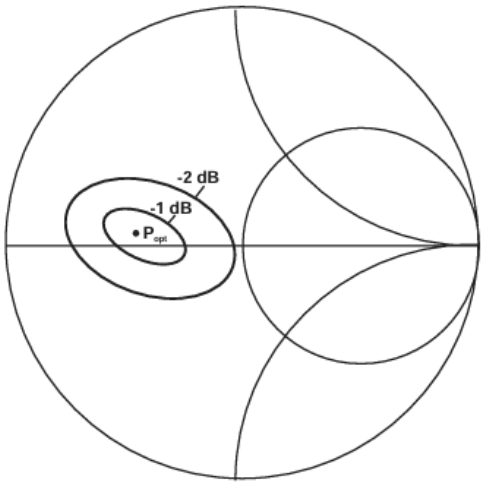

Figure 3.14: Typical load-pull data plotted on a Smith chart [28]. The point P_{opt} indicates the optimal load impedance.

Load-pull measurements provide a powerful tool for RF designers, as they allow for the identification of the optimal load impedance that maximizes output power. By understanding the shape and position of these contours, designers can better tailor the matching networks used in RF circuits to ensure that the PA operates at its highest possible performance.

These measurements are particularly important for high-frequency designs where small changes in impedance can lead to significant variations in performance. The ability to visualize and quantify the effects of impedance mismatches on a Smith chart helps engineers make informed decisions when designing matching networks and adjusting the load presented to the PA.

3.6 Classes of Power Amplifiers

Figure 3.15 illustrates the bias circuit of PA. A "big fat" inductance, BFL, or RF choke, feeds DC power to the drain, and is assumed large enough so that the current through it is substantially constant. The drain is connected to a tank circuit through "big fat" capacitor BFC to prevent any DC dissipation in the load. The tank circuit is used for resonance between the drain and R_L.

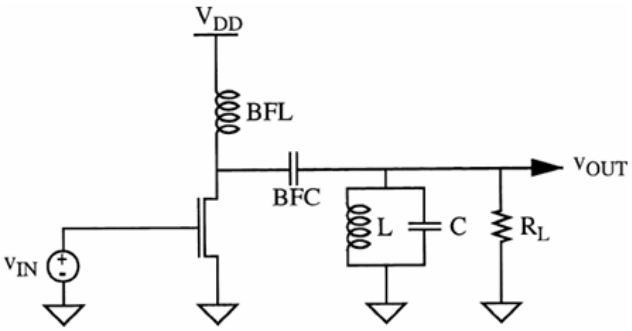

Figure 3.15: Bias circuit of PA [29].

The higher class categories (D, E, F, F1, S, and T) utilize the amplifier as a switch, allowing for very high efficiency to be achieved. However, in communications, the majority of transmitters operating predominantly at high frequencies rely on Class A, AB, or B amplifiers due to the need for low distortion products. Therefore, this section briefly examines Class A, AB, B, and C amplifiers, which differ from each other based on the shapes of their voltage and current waveforms and their conduction angles [9].

3.6.1 Class A

Class A power amplifiers are known for their high linearity and simplicity in design. These amplifiers operate by biasing the transistor so that it is always conducting. This is achieved by setting the gate voltage (V_g) higher than the threshold voltage (V_{th}), ensuring the transistor remains in the active region ($V_g > V_{th}$) throughout the entire input signal cycle, as shown in Figure 3.16. This continuous operation makes Class A amplifiers the most linear, as they faithfully reproduce the input signal with minimal distortion. However, this comes at the cost of efficiency.

The efficiency of a Class A amplifier is inherently low due to the continuous current flow through the transistor for the entire input signal cycle. Let us assume that the quiescent drain current, I_{DC}, is made just large enough to guarantee that the transistor does not ever cut off. That is,

$$I_{DC} = i_{rf}, \tag{3.28}$$

so that the DC power consumption is

$$P_{DC} = I_{DC} \cdot V_{DD} = i_{rf} \cdot V_{DD}. \tag{3.29}$$

The efficiency (η) is given by:

$$\eta = \frac{P_{rf}}{P_{DC}} = \frac{i_{rf}^2 R/2}{i_{rf} V_{DD}} = \frac{i_{rf} R}{2V_{DD}}, \tag{3.30}$$

i_{rf} is the amplitude of the drain current, R is the load resistance, and V_{DD} is the supply voltage.

Now, the absolute maximum that the product $i_{rf} \cdot R$ can have is V_{DD}. Therefore, the maximum theoretical drain efficiency is just 50%.

In Class A configuration, the transistor's drain current can be approximated by:

$$i_D = I_{DC} + i_{rf} \cdot \sin(\omega t), \tag{3.31}$$

I_{DC} is the bias current and i_{rf} is the fundamental harmonic component of drain current.

This drain current produces a corresponding drain voltage, which, due to the action of the inductor BFL and the capacitor BFC, consists of a DC component and an AC component as shown in figure 3.16.

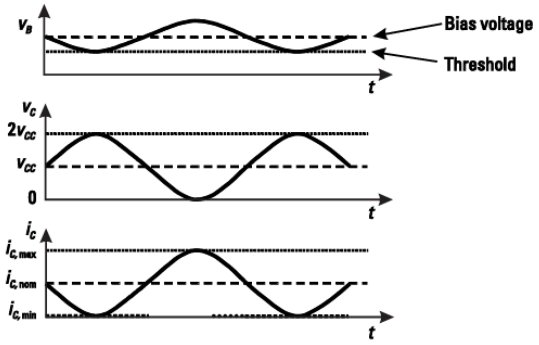

Figure 3.16: Gate voltage, drain voltage and drain current waveforms for an ideal Class A amplifier. The waveforms show that the gate voltage is always higher than the threshold voltage and the drain current and voltage are 180 degrees out of phase with each other [26].

In figure 3.16, it is clear that the drain voltage and current are out of phase by 180°. Despite the excellent linearity, the continuous current flow for the entire signal cycle, leads to significant power dissipation in the form of heat, thus lowering the overall efficiency.

In summary, while Class A amplifiers are the most linear and simplest to design, their low efficiency makes them less desirable for applications where power consumption is a critical factor. Nonetheless, their use is widespread in scenarios where signal fidelity is of outmost importance.

3.6.2 Class B

Class B PAs are designed to improve efficiency over Class A amplifiers by reducing the conduction angle of the transistor. In a Class B amplifier, the gate voltage (V_B) of the transistor is biased at the threshold voltage (V_{th}), such that it conducts only during one half of the input signal cycle. This setup results in a conduction duty cycle of 50%, meaning that the transistor is turned off during the other half of the cycle. As a consequence, Class B amplifiers operate more efficiently but introduce distortion due to the nonlinear operation.

Figure 3.17 illustrates the waveforms of the gate voltage (V_B), drain voltage (V_C), and drain current (I_C) in a Class B amplifier. The gate voltage V_B is set at the threshold voltage

V_{th}, which means the transistor conducts only during the positive half of the input waveform. As shown, the drain current I_C is zero during the negative half-cycle, indicating that the transistor is in the cut-off region during this period. The output tank in Figure 3.15 filters out the harmonics of this current, leaving a sinusoidal drain voltage as in the Class A amplifier.

To compute the output voltage, we first find the fundamental component of the drain current and then multiply this current by the load resistance. For this amplifier, we assume that the drain current is sinusoidal for one half-cycle and zero for the other half-cycle.

$$i_D = \begin{cases} i_{rf} \cdot \sin(\omega_0 t), & 0 < t < \frac{T}{2} \\ 0, & \frac{T}{2} \leqslant t \leqslant T \end{cases}, \tag{3.32}$$

$$i_{\text{fund}} = \frac{2}{T} \int_0^{T/2} i_D \cdot \sin(\omega_0 t)\, dt = \frac{i_{\text{rf}}}{2}, \tag{3.33}$$

$$V_{out} = \frac{i_{rf}}{2} \cdot R. \tag{3.34}$$

Since the maximum possible value of V_{out} is V_{DD}, it is clear that the maximum value of i_{rf} is:

$$i_{\text{rf, max}} = \frac{2V_{DD}}{R}. \tag{3.35}$$

The peak drain current and maximum output voltage for a Class B amplifier are therefore the same as those for a Class A amplifier. However, in a Class B amplifier, only the positive half of the input waveform is amplified, which leads to the higher efficiency that is characteristic of this amplifier class.

The output power is given by:

$$P_o = \frac{v_o^2}{2R}, \tag{3.36}$$

where v_o is the amplitude of the signal across the load resistor. The maximum value of v_o remains V_{DD}, so the maximum output power is:

$$P_{o,\max} = \frac{V_{DD}^2}{2R}. \tag{3.37}$$

Computing the DC input power requires the average drain current:

$$\overline{i_D} = \frac{1}{T} \int_0^{T/2} \frac{2V_{DD}}{R} \sin(\omega_0 t)\, dt = \frac{2V_{DD}}{\pi R}, \tag{3.38}$$

$$P_{DC} = \frac{2V_{DD}^2}{\pi R}. \tag{3.39}$$

Finally, the maximum drain efficiency for a Class B amplifier is:

$$\eta = \frac{P_{o,\max}}{P_{DC}} = \frac{\pi}{4} \approx 0.785. \tag{3.40}$$

Figure 3.17 clearly demonstrates that while Class B amplifiers offer a significant improvement in efficiency compared to Class A amplifiers, they do so at the cost of linearity. The transistor only conducts during the positive half-cycle of the input, leading to harmonic distortion in the

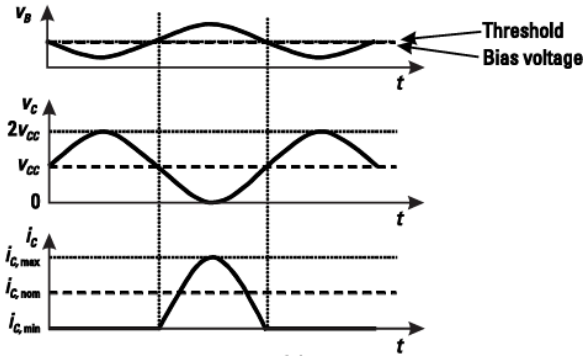

Figure 3.17: Waveforms for a Class B amplifier showing the gate voltage (V_B), drain voltage (V_C), and drain current (I_C). The transistor conducts only during the positive half-cycle of the input signal when $V_B > V_{th}$ [26].

output. Nevertheless, Class B amplifiers are widely used in applications where efficiency is more critical than signal linearity, particularly in RF power amplification.

The Class B amplifier, while more efficient than Class A, still leaves room for improvement efficiency. This pursuit of better performance leads us naturally to the consideration of Class C amplifiers, which further reduce the conduction angle in search of even higher efficiency, though at the cost of even greater distortion.

3.6.3 Class C

A Class C amplifier is characterized by a gate bias that causes the transistor to conduct for less than half of the input signal cycle. This is achieved by setting the gate voltage below the threshold voltage V_{th}. As a result, the drain current consists of a periodic train of narrow pulses, while the drain voltage remains sinusoidal due to the action of the high-Q tank circuit in the output.

Figure 3.18 illustrates the typical waveforms for a Class C amplifier, showing the relationship between the gate voltage, drain voltage, and drain current. The gate voltage waveform indicates that the transistor is only active when $V_g > V_{th}$, which occurs for less than half the cycle. Consequently, the drain current I_D is a series of pulses, while the drain voltage V_{DS} remains sinusoidal due to the filtering effect of the output tank circuit.

In this configuration, the efficiency of the amplifier can be significantly higher than in Class A or Class B amplifiers because the transistor is on for a shorter period, reducing the power dissipation during the non-conducting portions of the cycle. However, this also leads to increased distortion, making the design of the output tank circuit critical for maintaining linearity.

To analyze the efficiency of a Class C amplifier, we start by expressing the drain current as:

$$i_D = I_{DC} + i_{rf} \cdot \cos(\omega_0 t), \quad i_D > 0, \tag{3.41}$$

I_{DC} is the DC component of the drain current, and i_{rf} is the amplitude of drain current.

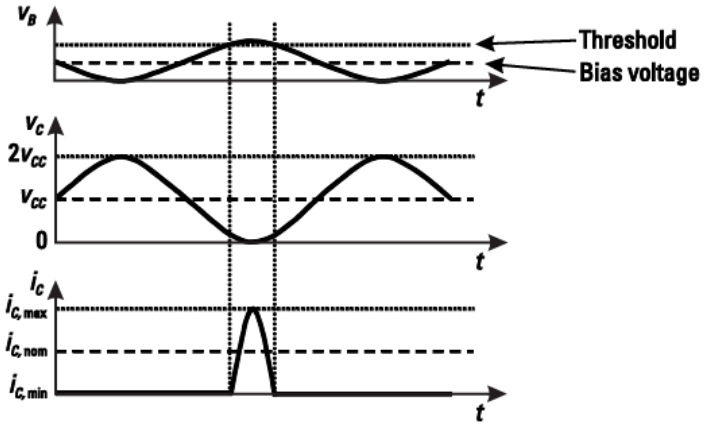

Figure 3.18: Gate Voltage, Drain voltage and Current for an ideal Class C amplifier. The gate voltage is biased below the threshold voltage V_{th}, causing the transistor to conduct only for a small portion of the cycle (dashed verticals), resulting in the characteristic pulsed drain current [26].

The transistor conducts only for a portion of the input cycle, determined by the conduction angle 2Φ, where 2Φ is the angle during which the current i_D is positive. Setting the current, from equation 3.41, equal to zero and solving for the total conduction angle 2Φ yields

$$2\Phi = 2 \cdot \cos^{-1}\left(-\frac{I_{DC}}{i_{rf}}\right). \tag{3.42}$$

From equation 3.42, the bias current I_{DC} can be expressed as:

$$I_{DC} = -i_{rf}\cos\Phi. \tag{3.43}$$

The average drain current $\overline{i_D}$ over one cycle is calculated by integrating the drain current over the conduction angle:

$$\overline{i_D} = \frac{1}{2\pi}\int_{-\Phi}^{\Phi}(I_{DC} + i_{rf}\cos\phi)\,d\phi. \tag{3.44}$$

Substituting the equation 3.43, we get:

$$I_{DC} = \frac{i_{rf}}{\pi}\left[\sin\Phi - \Phi\cos\Phi\right]. \tag{3.45}$$

In a high-Q tank circuit, the load is tuned to the fundamental frequency, and the harmonic components are filtered out. Thus, the fundamental component of the drain current, i_{fund}, is given by the fundamental term in the Fourier series:

$$\begin{aligned} i_{\text{fund}} &= \frac{2}{T}\int_0^T i_D\cos(\omega_0 t)\,dt \\ &= \frac{1}{2\pi}\left(4I_{DC}\sin\Phi + 2i_{rf}\Phi + i_{rf}\sin(2\Phi)\right). \end{aligned} \tag{3.46}$$

Substituting for I_{DC} the equation 3.45, we obtain

$$i_{\text{fund}} = \frac{i_{rf}}{2\pi}\left[2\Phi - \sin(2\Phi)\right]. \tag{3.47}$$

The output voltage swing V_{out} across the load resistance R is related to the fundamental component of the current as follows:

$$V_{out} = i_{\text{fund}} \cdot R. \tag{3.48}$$

Substituting the equation 3.47, we get:

$$V_{out} = \frac{i_{rf} R}{2\pi} \left[2\Phi - \sin(2\Phi)\right]. \tag{3.49}$$

The output power P_{out} delivered to the load is given by:

$$P_{out} = \frac{V_{out}^2}{2R}. \tag{3.50}$$

Substituting the equation 3.49, we get:

$$P_{out} = \frac{i_{rf}^2 R}{8\pi^2} \left[2\Phi - \sin(2\Phi)\right]^2. \tag{3.51}$$

The DC input power P_{DC} is given by:

$$P_{DC} = I_{DC} V_{DD}. \tag{3.52}$$

Substituting I_{DC} from Equation 3.45 and assuming that $V_{DD} = V_{out}$, we have:

$$P_{DC} = i_{rf}^2 \cdot \frac{R}{2\pi^2} \cdot \left[\sin\Phi - \Phi\cos\Phi\right] \cdot \left[2\Phi - \sin(2\Phi)\right]. \tag{3.53}$$

The efficiency η is defined as the ratio of the output power to the DC input power:

$$\eta = \frac{P_{out}}{P_{DC}}. \tag{3.54}$$

Substituting the equations 3.51, 3.53, we get:

$$\eta = \frac{2\Phi - \sin(2\Phi)}{4\left[\sin\Phi - \Phi \cdot \cos\Phi\right]}. \tag{3.55}$$

As the conduction angle shrinks toward zero, the efficiency approaches 100%. While this sounds promising, the gain and output power unfortunately also tend toward zero at the same time, since the fundamental component in the ever-narrowing slivers of drain current shrinks as well. All of these tradeoffs force the attainment of less than 100% efficiency in practice, since we generally want a reasonable amount of output power as well as high efficiency. Mainly, the efficiency can be large, but at the cost of reduced power-handling capability, gain, and linearity [29].

3.6.4 Class AB

Class AB amplifiers are designed to combine the benefits of Class A and Class B amplifiers by conducting more than 50% but less than 100% of the time, depending on the chosen bias levels. This operating condition allows the Class AB amplifier to achieve a compromise

between the efficiency of Class B amplifiers and the linearity of Class A amplifiers. The result is an amplifier that is more efficient than a Class A amplifier but more linear than a Class B amplifier, making it the most popular choice.

The key characteristic of a Class AB amplifier is its conduction angle, which is greater than 180° but less than 360°, meaning that it conducts for more than half but not the full duration of each signal cycle. This allows the amplifier to provide better efficiency compared to Class A while maintaining a reasonable level of linearity.

The equations used to analyze Class AB amplifiers are the same as those used for Class C amplifiers. So, the efficiency for Class A, B, AB and C amplifiers is given by 3.55.

Overall, the Class AB amplifier is a practical solution that balances the trade-offs between efficiency and linearity, making it the most commonly used topology in power amplification.

3.6.5 Summary

The conduction angle and efficiency are summarized for the above classes in table 3.1.

Class	Conduction Angle (°)	Efficiency (max theoretical) (%)
A	360	50
AB	360–180	50–78.5
B	180	78.5
C	180–0	78.5–100

Table 3.1: Summary of Power Amplifier Classes

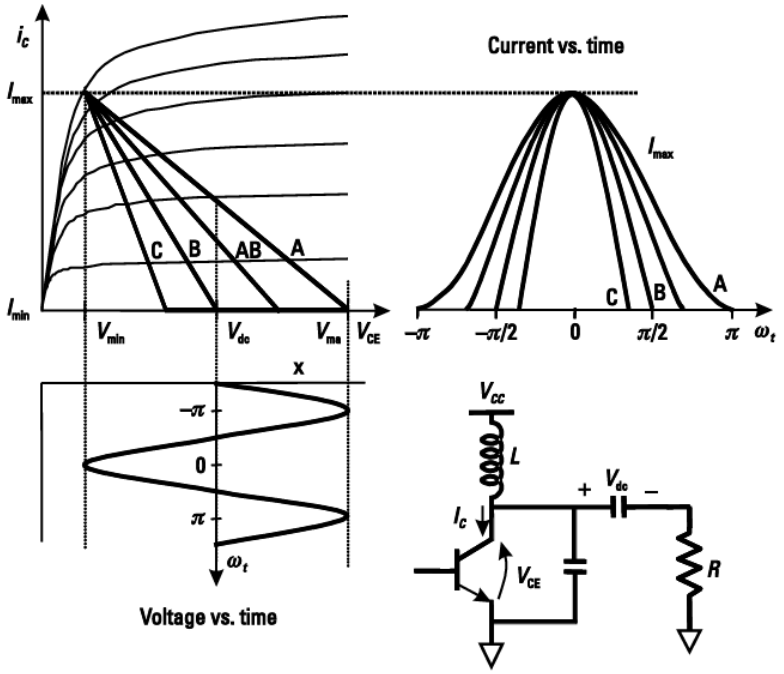

Figure 3.19: Loadline curves for different classes of amplifiers (A, B, AB, C) [26].

The loadline curves depicted in figure 3.19 show the operating characteristics of a transistor for different amplifier classes (Class A, B, AB, and C). Each curve represents the relationship between the drain current (I_D) and the drain-source voltage (V_{DS}) over one cycle of the input signal.

- Class A: The loadline curve for Class A shows that the transistor conducts for the entire input cycle (360° conduction angle). The current remains continuous, leading to high linearity but low efficiency.
- Class B: The Class B curve indicates that the transistor conducts for half of the input cycle (180° conduction angle), improving efficiency but introducing some distortion.
- Class AB: The Class AB curve demonstrates conduction for more than half but less than the full cycle (between 180° and 360° conduction angle). This class balances the linearity and efficiency characteristics of Classes A and B.
- Class C: The Class C curve shows that the transistor conducts for less than half of the input cycle (less than 180° conduction angle). This class offers high efficiency but at the expense of linearity.

The figure also includes the corresponding voltage and current waveforms, highlighting how the conduction angle affects the transistor's performance in each class. Additionally, the circuit diagram shows the basic configuration of the transistor with its associated components in a typical amplifier circuit.

Chapter 4

Description of design Technologies

4.1 CMOS FDSOI

FDSOI stands for Fully-Depleted-Silicon on Insulator. It is a planar process technology developed as an alternative solution to address some of the limitations of bulk CMOS technology, particularly at smaller silicon nodes and reduced geometries. FDSOI enables improved performance, reduced power consumption, and greater control over the transistor characteristics.

The FDSOI process incorporates two key features:

- Buried Oxide Layer: An ultra-thin buried oxide (BOX) layer is placed on top of the silicon substrate. This buried-oxide layer electrically isolates the transistor from the bulk silicon, reducing parasitic capacitances and leakage current to substrate.
- Thin Silicon Channel: A very thin silicon layer is deposited above the BOX layer to form the transistor channel. This thin channel eliminates the need for doping, resulting in a fully depleted channel under the gate.

The FDSOI process is illustrated in figure 4.1. On the left, a wafer is shown, consisting of the ultra-thin buried oxide layer and the ultra-thin top silicon layer. On the right, a cross-sectional view of the FDSOI transistor is presented. The figure highlights:

- The ultra-thin buried oxide (BOX) layer.
- The thin silicon layer forming the fully depleted channel.
- The raised source and drain regions.
- The gate, which controls the channel current.

FDSOI technology offers several distinct benefits over bulk CMOS:

- *Improved Gate Control:* The thin top-silicon layer and buried oxide layer ensure better electrostatic control, reducing short-channel effects.

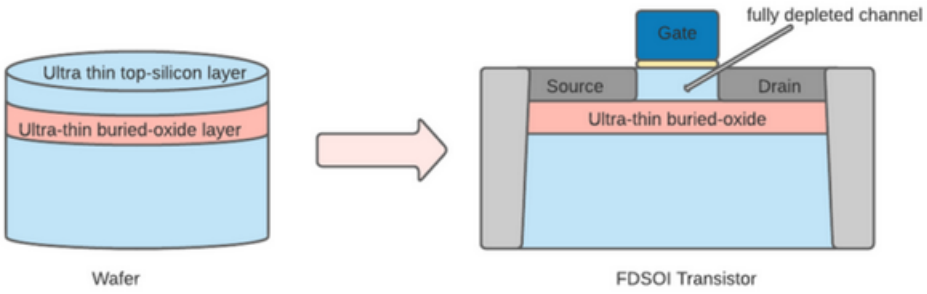

Figure 4.1: FDSOI transistor [30].

- *Reduced Leakage Current:* The BOX layer blocks leakage currents from the channel to the substrate, which is a significant issue in bulk CMOS.
- *Enhanced Performance:* The FDSOI channel allows faster switching with a steeper subthreshold slope, reducing threshold voltage (V_{th}) and enabling higher performance at lower supply voltages.
- *Body Biasing:* The FDSOI architecture allows forward and reverse body biasing. Forward body biasing reduces V_{th} to improve performance, while reverse biasing increases V_{th} to reduce leakage current.
- *Lower Parasitic Capacitances:* The buried oxide layer isolates the channel from the substrate, reducing parasitic capacitances and enabling faster operation.
- *Simplified Manufacturing:* The FDSOI planar process is simpler than 3D technologies like FinFET, as it reuses steps from mature bulk CMOS manufacturing processes, leading to reduced costs and design complexities.

In conventional bulk CMOS, short-channel effects such as drain-induced barrier lowering (DIBL) and threshold voltage variability are exacerbated as the gate length shrinks. These effects degrade transistor performance and increase leakage power. In FDSOI, the thin BOX layer and fully depleted channel effectively suppress short-channel effects, resulting in superior threshold voltage control and reduced leakage current.

The buried gate structure of FDSOI transistors allows body biasing, which enhances transistor flexibility. Forward body biasing boosts performance by reducing the threshold voltage, while reverse body biasing minimizes leakage currents. Different voltage levels can be independently applied to the top and bottom gates of the transistor, providing a wide range of performance and power consumption options.

FDSOI technology supports design reuse due to its planar structure. Designers can adapt existing bulk CMOS intellectual property (IP) blocks for FDSOI with minimal modifications. This scalability, combined with lower manufacturing complexity, makes FDSOI an attractive alternative to FinFET technology for achieving high performance and low power consumption in advanced technology nodes. Also, the FDSOI technology eliminates dopant fluctuations due to the fully depleted channel, so the FDSOI technology is independent of process variations.

For analog designs, meeting the noise, power, leakage and variability requirements is becoming very challenging due to degradation in transistor characteristics as technology shrinks. As FDSOI transistors have improved device matching, gain and parasitics, it can greatly simplify the analog design. The back bias capability in FDSOI also has a lot of potential in designing analog circuits. FDSOI has very high f_t, $f_{\max}$ and high mmWave self gain which is important for RF blocks. These key characteristics help in superior RF and analog designs.

4.2 GaN HEMT

Gallium Nitride (GaN) has emerged as a transformative technology in RF and power applications due to its superior material properties compared to traditional silicon. Table 4.1

summarizes the key advantages of GaN, and its benefits are detailed below.

The bandgap of a material defines the energy required for electrons to move freely from the valence band to the conduction band. GaN has a bandgap that is three times wider than silicon, enabling it to operate effectively at much higher temperatures, up to 800°C.

GaN exhibits high electron mobility, approximately $2000\,\mathrm{cm}^2/\mathrm{V}\cdot\mathrm{s}$, which allows electrons to move quickly under a low electric field. This property reduces on-resistance and switching losses, enabling operation at switching frequencies up to ten times higher than silicon. As a result, GaN is ideal for high-speed power applications.

The high breakdown electric field of GaN, approximately $3.3\,\mathrm{MV/cm}$, is over ten times higher than that of silicon. This characteristic enables GaN devices to sustain higher voltages without failure, making them ideal for high-power applications such as power amplification and switching power supplies.

While GaN itself has moderate thermal conductivity, combining it with Silicon Carbide (SiC) substrates improves its thermal performance significantly. GaN-on-SiC devices can dissipate heat efficiently, allowing operation at higher power levels without requiring excessive cooling solutions. This flexibility makes GaN suitable for various applications, including 5G millimeter-wave technology, automotive power systems, and IoT devices.

GaN's material properties enable compact transistor designs without compromising performance. GaN transistors can achieve high power densities while maintaining efficiency, making them suitable for miniaturized, high-performance electronic systems.

Property	**Si**	**SiC**	**GaN**
Bandgap (E_g) [eV]	1.12	3.26	3.425
Breakdown Field (E_c) [MV/cm]	0.3	3	3.3
Electron Mobility (μ_n) [$\mathrm{cm}^2/\mathrm{V}\cdot\mathrm{s}$]	1500	1000	2000
Thermal Conductivity [$\mathrm{W}\cdot\mathrm{K/cm}$]	1.5	4.9	1.3
Max Operating Temp ($T_{\max}$) [°C]	150	760	800

Table 4.1: Material Properties Comparison: Si, SiC, and GaN

Gallium Nitride High Electron Mobility Transistors (GaN HEMTs) are designed similarly to conventional transistors, consisting of three key components: the source, drain, and gate. However, GaN HEMTs utilize advanced material growth techniques like Metal-Organic Chemical Vapor Deposition (MOCVD) to form multiple layers on either silicon or silicon carbide (SiC) substrates.

As illustrated in figure 4.2, the active region of the GaN HEMT, where current flows, is located at the interface between the AlGaN and GaN layers. This layered structure includes:

- Gate and Field Plate: Controls the current flow.
- Passivation Layer: Provides surface stability and reliability.
- AlGaN Layer: Introduces a piezoelectric effect due to lattice strain.

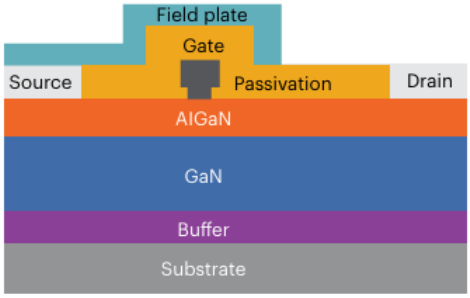

Figure 4.2: Typical structure of RF AlGaN/GaN HEMT.

- GaN Layer: Acts as the primary channel for current flow.
- Buffer and Substrate: Supports the device and isolates it electrically.

GaN devices benefit from their unique crystal structure. At the interface between the AlGaN and GaN layers, bonding asymmetries in the wurtzite crystal structure generate built-in electric fields. These fields cause an electric field difference (Φ_B) at the AlGaN/GaN junction, leading to the formation of a triangular quantum well.

In this quantum well, electrons form a highly dense layer called the two-dimensional electron gas (2DEG). This layer:

- Offers extremely high electron mobility ($> 1500\,\mathrm{cm}^2/\mathrm{V}\cdot\mathrm{s}$).
- Enables efficient current flow without additional doping processes.
- Contributes to high switching speeds and low power losses.

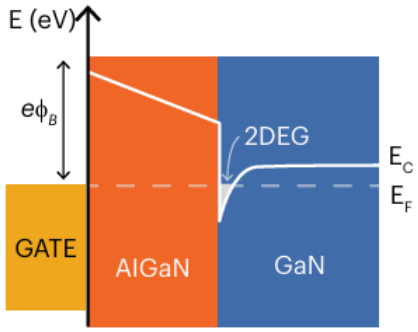

Figure 4.3: Formation of 2DEG (two-dimensional electron gas) in GaN HEMTs.

Figure 4.3 demonstrates the energy band alignment where electrons are confined below the Fermi level (E_F) in a quantum well at the AlGaN/GaN interface. This results in superior power performance for GaN HEMTs [31].

Chapter 5

Stacked FET Circuit Analysis

Figure 5.1 illustrates the schematic of a stacked-FET power amplifier (PA). The circuit structure is based on the series connection of a common-source (CS) transistor followed by multiple common-gate-like transistors. Unlike cascode amplifiers, where the gate of the common-gate transistor is grounded at the operating frequency, in the stacked-FET configuration, the gate of the common-gate-like transistor is connected to a finite impedance and experiences a voltage swing. This setup ensures that the drain voltages of the transistors add in phase, while the drain current remains constant through all the transistors in the stack.

The gate voltage swing of the transistors in the stack is regulated by introducing appropriate capacitances C_k at the gates of the stacked transistors. The combination of C_k and the gate-source capacitance $C_{\mathrm{gs},k}$ forms a voltage divider network, determining the gate voltage levels at each stage. Compared to cascode designs, this method minimizes the drain-gate and drain-source voltage swings under large-signal conditions, enabling the transistors to operate reliably even at high aggregate voltage levels.

Despite its advantages, the stacked-FET PA exhibits a lower gain compared to the cascode amplifier. However, the tradeoff is acceptable given the higher output power and improved drain efficiency of the stacked-FET design. This makes the stacked-FET configuration particularly suitable for applications requiring high output power levels, even when a large number of transistors are connected in series. The slight reduction in gain is offset by the significant improvements in power efficiency and overall performance.

From figure 5.1, the impedance $Z_{d,k-1}$ seen at the drain of transistor $(k-1)$ assuming linear

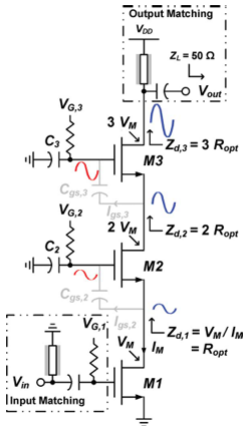

Figure 5.1: Schematic Design of 3-Stacked PA [5].

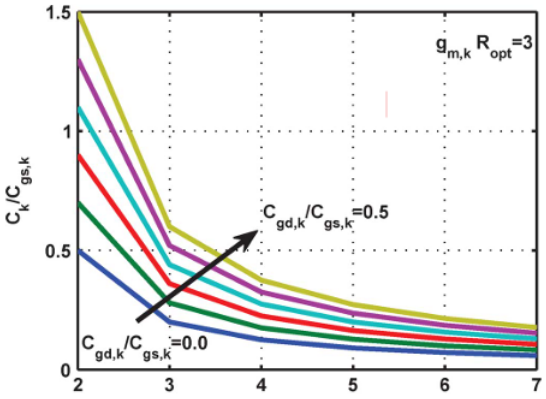

Figure 5.2: $C_k/C_{gs,k}$ for various $C_{gd,k}/C_{gs,k}$ for $g_{m,k} \cdot R_{\text{opt}} = 3$ [5].

operation and neglecting the FET small-signal output resistance and drain-source capacitance is:

$$Z_{d,k-1} = \frac{C_{\text{gs},k} + C_k + C_{\text{gd},k}(1 + g_{\text{m},k} Z_{d,k})}{(g_{\text{m},k} + sC_{\text{gs},k})(C_{\text{gd},k} + C_k)}. \tag{5.1}$$

where $C_{\text{gd},k}$ is the gate-drain device capacitance and $g_{\text{m},k}$ is the transconductance of the k-th transistor in the stacked-FET PA.

From equation 5.1, $Z_{d,k}$ is real when the operating frequency f_0 is much smaller than the cutoff frequency f_t of the transistor. To provide the optimum load line impedance to each of the transistors and ensure that the drain-source voltages are equally distributed among the stacked devices, the impedance $Z_{d,k}$ should be adjusted to $k \cdot R_{\text{opt}}$, where R_{opt} is the loadline impedance of a single device.

In III-V technologies, the gate-drain capacitances are typically much smaller compared to the gate-source capacitances and can often be neglected. However, in scaled CMOS FETs, the gate-drain capacitance cannot be ignored, especially at millimeter-wave frequencies. Since $sC_{\text{gs},k}$ is significantly smaller than $g_{\text{m},k}$, even at these high frequencies, the expression for $Z_{d,k-1}$ can be simplified as:

$$\begin{aligned} Z_{d,k-1} \approx \frac{C_{\text{gs},k} + C_k + C_{\text{gd},k}(1 + g_{\text{m},k} Z_{d,k})}{g_{\text{m},k}^2 (C_{\text{gd},k} + C_k)} g_{\text{m},k} \\ - \frac{C_{\text{gs},k} + C_k + C_{\text{gd},k}(1 + g_{\text{m},k} Z_{d,k})}{g_{\text{m},k}^2 (C_{\text{gd},k} + C_k)} sC_{\text{gs},k}, \quad k = 2, 3, \ldots, K. \end{aligned}$$

Assuming that $Z_{d,k}$ is primarily real and is set equal to $k \cdot R_{\text{opt}}$, the gate capacitance C_k is derived by matching the real part of $Z_{d,k-1}$ to $(k-1) \cdot R_{\text{opt}}$. The resulting gate capacitance C_k is given by:

$$C_k = \frac{C_{\text{gs},k} + C_{\text{gd},k}(1 + g_{\text{m},k} R_{\text{opt}})}{(k-1) g_{\text{m},k} R_{\text{opt}} - 1}, \quad k = 2, 3, \ldots, K. \tag{5.2}$$

Figure 5.2 illustrates the variation of C_k, normalized to $C_{\mathrm{gs},k}$, for different ratios of $C_{\mathrm{gd},k}/C_{\mathrm{gs},k}$ with a fixed gain of 3. The graph emphasizes the significance of including $C_{\mathrm{gd},k}$ in the calculation of C_k.

Furthermore, it is noted that the gate capacitance C_k becomes very small for higher values of k, making its precise determination highly sensitive to modeling inaccuracies. This introduces one of the critical challenges in stacking multiple transistors within the design process.

One of the critical aspects of stacked FET power amplifier (PA) design is the proper adjustment of the DC gate voltages. This ensures efficient and reliable operation while maintaining the DC and RF voltages V_{gs}, V_{gd}, and V_{ds} below their respective breakdown limits.

In Class AB operation, as the RF input power P_{in} increases, the DC bias current rises accordingly. The DC gate voltages are typically fixed independent of P_{in}. To avoid issues, the gate voltages should be carefully adjusted to meet the current levels required for maximum P_{in}. Improper adjustment may result in insufficient voltage swing for the top stacked FETs, leading to early breakdown and compression of the bottom Common-Source (CS) FET.

The DC gate voltage for the k-th stacked transistor in a K-stacked PA can be expressed as:

$$V_{\mathrm{G},k} = \left(\frac{k-1}{K}V_{\mathrm{DD}} + V_{\mathrm{GS},k-\mathrm{sat}}\right), \quad k = 2, 3, \ldots, K \tag{5.3}$$

where $V_{\mathrm{GS},k-\mathrm{sat}}$ represents the gate-source voltage of the k-th transistor at the saturation power level.

It is essential to ensure that the gate-drain voltage swing remains below the breakdown threshold. The optimal choices of capacitance C_k, gate voltage $V_{\mathrm{G},k}$, and load resistance R_{opt} help avoid breakdown. Under ideal conditions, the gate-drain voltage $V_{\mathrm{gd},k}$ for the k-th stacked device is given by:

$$V_{\mathrm{gd},k} = \frac{1 + g_{\mathrm{m},k}R_{\mathrm{opt}}}{g_{\mathrm{m},k}R_{\mathrm{opt}}}V_{\mathrm{opt}}, \quad k = 1, 2, \ldots, K \tag{5.4}$$

where $g_{\mathrm{m},k}$ is the transconductance of the k-th transistor, and V_{opt} is the optimal load voltage swing.

From equation 5.4, it can be inferred that the peak gate-drain voltage magnitude is equal to:

$$V_{\mathrm{opt}} + V_{\mathrm{gs}},$$

where V_{gs} is the gate-source voltage needed to drive the RF current. However, if the transistor size is not properly designed, the sum of these voltages may exceed the gate-drain breakdown voltage, posing reliability concerns.

To address this, additional constraints must be considered on the transistor size, ensuring the device operates within safe voltage limits.

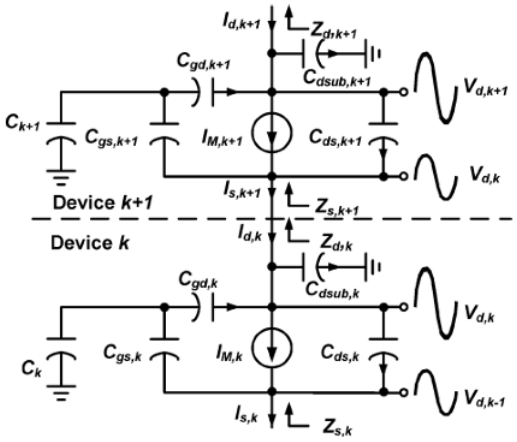

Figure 5.3: Simplified small signal model of stacked transistors [5].

For PAs utilizing CS amplifiers in scaled CMOS processes, there are three primary factors that constrain the achievable saturated output power. These factors include the transistor breakdown voltage, the maximum gate width that ensures sufficient gain, and the availability of a realizable matching network that aligns with the desired loadline impedance.

Furthermore, for a CS amplifier implemented in a scaled CMOS process, the matching network often becomes the dominant limiting factor. By stacking K FETs while maintaining a constant drain current, the output power can theoretically increase proportionally by a factor of K. Additionally, the required load impedance increases linearly with K, presenting benefits for stacked configurations compared to other configurations.

At low frequencies, $Z_{d,k}$ can be approximated as a resistive load. However, at millimeter-wave frequencies, the intermediate node impedances exhibit significant reactance due to the transistor capacitances. This behavior reduces efficiency for two primary reasons:

1. As illustrated in figure 5.1, a portion of the transistor RF current flows through the gate-source capacitance $C_{gs,2}$ and other parasitic capacitances at the drain of M_1, preventing it from reaching the load.

2. The voltage waveforms at the intermediate and top drains are not perfectly phase-aligned, which limits the maximum voltage swing achievable at the topmost drain.

The simplified small-signal model of stacked transistors, as illustrated in figure 5.3, is utilized to determine the optimal impedances at the drain terminals of the FETs. The optimal admittance $Y_{\mathrm{opt},k}$ for the k-th stacked transistor can be approximated as:

$$\begin{aligned} Y_{\mathrm{opt},k} &\approx \frac{1}{kR_{\mathrm{opt}}} - \frac{s}{k}\left(C_{ds,k} + kC_{\mathrm{sub},k} + C_{gd,k}\right) \\ &= \frac{1}{kR_{\mathrm{opt}}} - \frac{s}{k}\left(C_{\mathrm{eqv},k}\right), \quad k = 1, 2, \ldots, K. \end{aligned} \tag{5.5}$$

Here, $C_{\text{eqv},k}$ represents the equivalent capacitance seen at the drain node of the k-th transistor. This optimum load admittance ensures that all drain-source voltages and drain currents are aligned, leading to highest output power and best efficiency.

For low frequencies, the gate capacitance C_k is carefully selected to ensure that each stacked transistor presents an optimal load resistance. However, the susceptance seen by the k-th transistor should ideally be inductive to counteract the capacitances at the drain. In reality, the $(k+1)$-th transistor presents a capacitive load to the k-th transistor, which complicates the load matching process.

The admittance looking into the source of the $(k+1)$-th transistor is given as:

$$Y_{s,k+1} = \frac{1}{kR_{\text{opt}}} - \frac{sC_{\text{ds},k+1}}{k} + \frac{sC_{\text{gs},k+1}}{kg_{m,k+1}R_{\text{opt}}}. \tag{5.6}$$

Using the above expression, the phase angle of the impedance presented by the $(k+1)$-th transistor to the k-th transistor can be determined as:

$$\Phi_{s,k+1} = \arctan\left(\omega\left(\frac{C_{\text{gs},k+1}}{g_{m,k+1}} - C_{\text{ds},k+1}R_{\text{opt}}\right)\right). \tag{5.7}$$

To ensure proper matching, the optimal load at the drain of the k-th transistor must exhibit a phase angle given by:

$$\Phi_{\text{opt},k} = \arctan\left(-\omega\left(C_{\text{eqv},k}R_{\text{opt}}\right)\right), \tag{5.8}$$

where $C_{\text{eqv},k}$ is defined in equation 5.5.

To account for the capacitive effects and ensure phase alignment, additional matching components are required to phase-rotate $Y_{s,k+1}$ by:

$$\Phi_k = -\Phi_{\text{opt},k} + \Phi_{s,k+1}. \tag{5.9}$$

This phase tuning is critical to achieve the highest voltage swing at the topmost drain node and maximize the performance of the stacked FET power amplifier.

The relationship between phase errors Φ_k and the power-combining efficiency of stacking is derived. The total output power of a K-stacked PA can be approximated by:

$$\eta_{\text{stacking}} \approx \left(\prod_{k=1}^{K}\cos(\Phi_k)\right)^2 \tag{5.10}$$

$$P_{\text{out, K-stack}} \approx \eta_{\text{stacking}}P_{\text{out, ideal K-stack}} \tag{5.11}$$

where K represents the total number of stacked transistors, and $P_{\text{out, ideal K-stack}}$ is the ideal output power of the K-stacked PA when all currents and voltages are perfectly aligned.

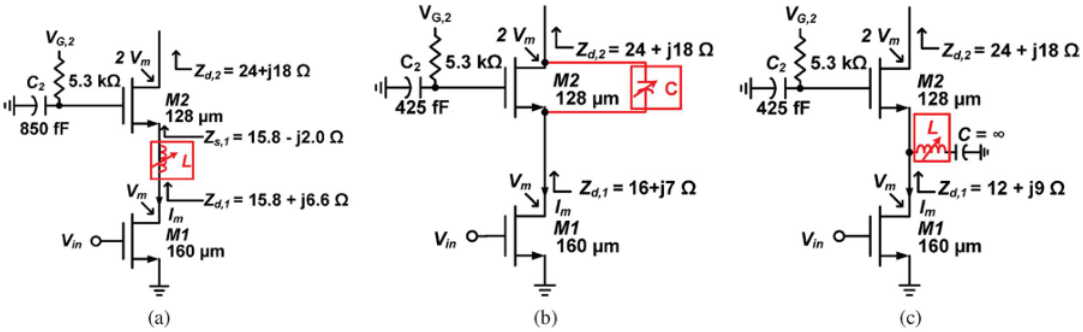

Figure 5.4: 2-Stacked PA schematic with different intermediate node tuning technigues: a) Shunt inductive tuning, b) Shunt-feedback C_{ds} tuning, c) Series inductive tuning [5].

However, phase misalignment between the stacked transistors can cause a significant degradation in efficiency. Even small degrees of phase errors lead to observable efficiency loss. This phase alignment effect poses a key challenge to the successful implementation of many stacked FETs. This becomes significant in mm-Wave frequencies, while this phase misalignment does not exist or it is negligible at sub-10GHz region.

At millimeter-wave frequencies, significant efficiency and power reduction can occur without any reactive tuning. To address this, recent literature proposes three different circuit approaches to implement the proper complex impedance at the intermediate node.

- Figure 5.4 a) presents a shunt L tuning technique.
- Figure 5.4 b) demonstrates a shunt-feedback drain–source capacitor tuning method.
- Figure 5.4 c) highlights the use of series inductance between the transistors.

The inductances in these methods serve to tune out the parasitic capacitances at the intermediate nodes, acting as either a series LC or parallel LC circuit. Notably, the shunt-feedback C_{ds} approach achieves a similar effect since the capacitance across the transistor effectively behaves like a negative capacitance, as shown in equation 5.6.

Chapter 6

Design of Stacked FET PA in CMOS FDSOI and GaN HEMT technology

6.1 CMOS FDSOI

Current Density is defined as $J = \frac{I_{\text{bias}}}{W}$, where I_{bias} is the bias current of the transistor and W is its total gate width. Constant current density design procedure reduces the impact of statistical process variation, temperature and bias current variation on circuit performance. It is proven that characteristic current densities also remain invariant for the most common circuit topologies across different technology nodes and gate lengths [32].

To fully characterize the slvtnfet_rf device, which is a NMOS device with super low threshold voltage V_{th}, from Globaldoundries 22nm Fully-Depleted Silicon-on-Insulator (FD-SOI) process technology, the following metrics need to be evaluated in function of current density J:

1. G_{max} - maximum transducer power gain (conjugate matching at input and output)
2. f_{max} - frequency at which maximum power gain becomes unity ($G_{\text{max}} = 0\,\text{dB}$)
3. NF_{min} - minimum noise figure (noise match at the input port)
4. K_{f}, μ, μ', $B_{1\text{f}}$ - stability factors

From sections 2.3.3, 2.3.4, unconditional stability is secured when

$$K_{\text{f}},\ \mu,\ \mu' > 1 \text{ and } B_{1\text{f}} > 0. \tag{6.1}$$

The schematic design used to characterize the slvtnfet_rf device is shown in figure 6.1. An ideal current source (instance I_1) provides the bias current I_{bias} which flows directly to the ideal current controlled current source (CCCS, instance F_0). The gate of slvtnfet_rf device and capacitor C_1 block the DC current. CCCS F_0 copies the bias current to the reference ideal voltage source, V_2, which is connected at the source terminal of slvtnfet_rf device. Therefore, the circuit configuration shown in Figure 6.1, biases the slvtnfet_rf device with a specified current I_{bias}. Instances F_0, V_2 form an ideal current mirror. The bulk and source terminals of slvtnfet_rf device are biased to ground. The drain terminal is biased to the overdrive voltage of the device which is 0.9V.

The capacitors C_0, C_1 are used for AC coupling and inductors L_0, L_1 (RF chokes) block the AC (or RF) signals and bias the transistor terminals with the appropriate voltages. The

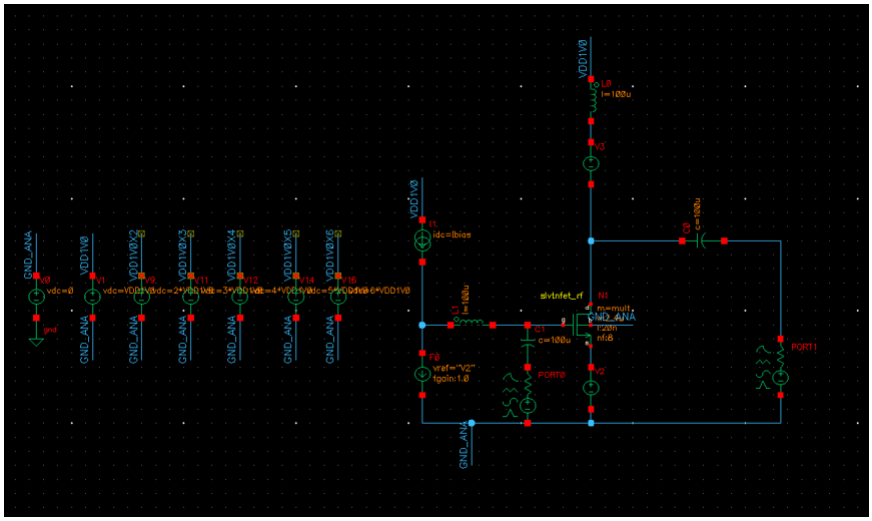

Figure 6.1: Schematic Design for Characterization of slvtnfet_rf device

slvtnfet_rf device has a number of fingers $n_f = 8$, a finger width of $w_f = 300$um and multipliers mul = 64.

The stability factors, maximum power gain $G_{\max}$, minimum noise figure $NF_{\min}$, and transition frequency $f_{\max}$ are plotted as functions of current density J in figures 6.2-6.8. Specifically, the width of the slvtnfet_rf device remains constant while I_{bias} is swept. For each bias current, the stability factors, $G_{\max}$ and $NF_{\min}$ are measured using SP analysis at 5 GHz and then plotted against the corresponding current density $J = \dfrac{I_{\text{bias}}}{W}$, where $W = \text{mul} \cdot n_f \cdot w_f$. To calculate the transition frequency $f_{\max}$, the $G_{\max}$ is plotted as a function of frequency for each bias current, and the frequency at which $G_{\max} = 0\,\text{dB}$ corresponds to $f_{\max}$ for each bias current. Then, $f_{\max}$ can be plotted as a function of current density J.

From figures 6.5-6.8 and equation 6.1, stability is guaranteed when $J > 60\,\frac{\text{mA}}{\text{mm}}$.

From figures 6.2-6.4, $f_{\max}$ reaches its maximum at $325\,\frac{\text{mA}}{\text{mm}}$, $G_{\max}$ is inversely proportional to the current density J, and $NF_{\min}$ reaches its minimum at $142\,\frac{\text{mA}}{\text{mm}}$. These results are consistent with relevant scientific references [32].

For the design of a power amplifier, noise figure is not a significant metric. Also, even at low current densities, $f_{\max} > 200\,\text{GHz} \gg 5\,\text{GHz}$, so the slvtnfet_rf device can operate as an amplifier at the operating frequency of the power amplifier. Therefore, the main objective is to bias the slvtnfet_rf device to achieve maximum power gain and ensure stability. Based on the previous analysis, the optimal design choice is $70\,\frac{\text{mA}}{\text{mm}} < J < 100\,\frac{\text{mA}}{\text{mm}}$. In this current density range, maximum power gain $G_{\max}$ is achieved, and stability is ensured.

The next goal is to specify the multiplier and bias current of the common source (CS) device, slvtnfet_rf. Each multiplier will be tested separately, and for $J = 70\,\frac{\text{mA}}{\text{mm}}$ and $J = 100\,\frac{\text{mA}}{\text{mm}}$, the bias current will be calculated using the following equation:

$$I_{\text{bias}}\,(\text{mA}) = J \cdot \text{mul} \cdot n_f \cdot w_f \tag{6.2}$$

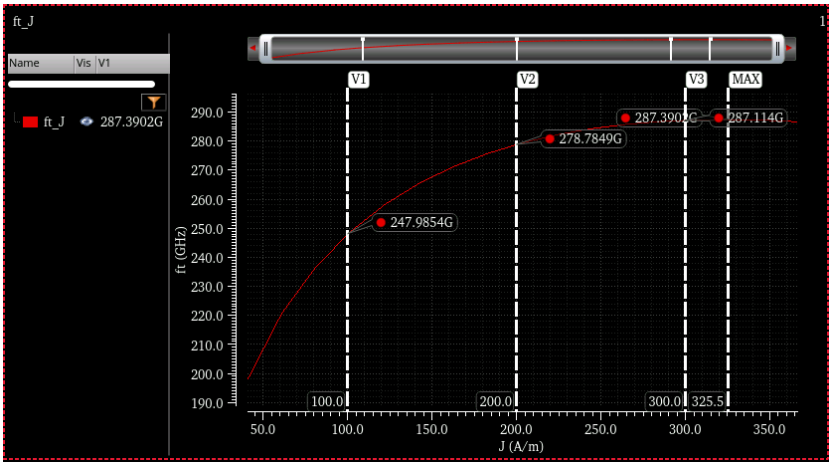

Figure 6.2: $f_{\max}$ (GHz) vs $J\,(A/m)$

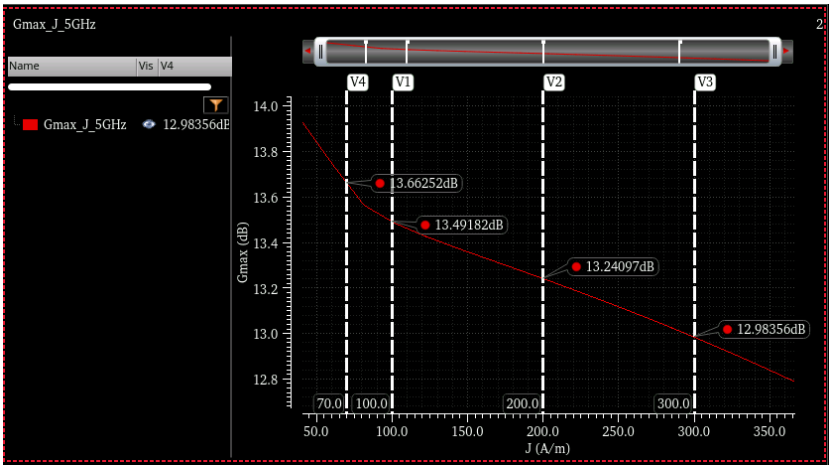

Figure 6.3: $G_{\max}$ (dB) vs $J\,(A/m)$

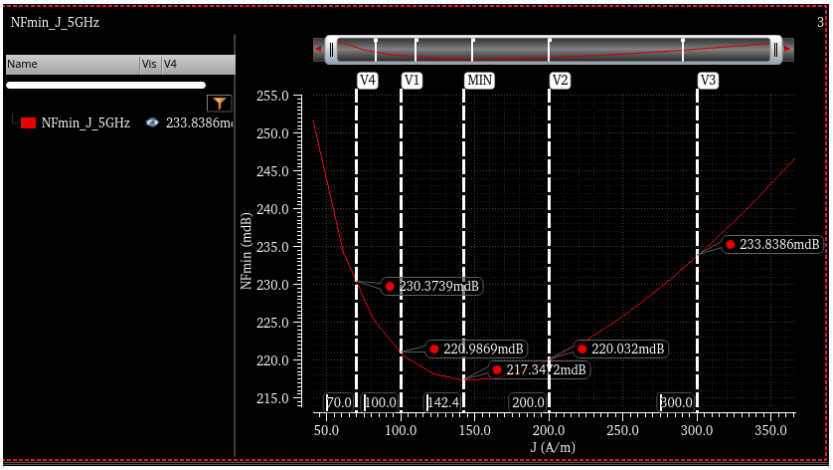

Figure 6.4: $NF_{\min}$ (dB) vs $J\,(A/m)$

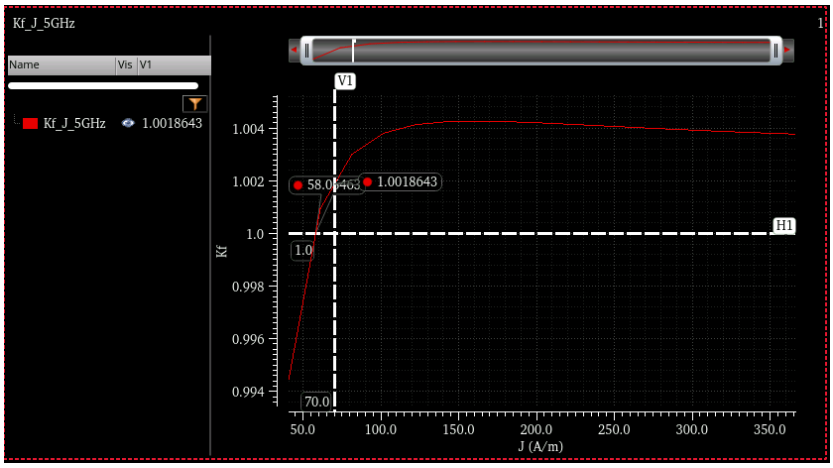

Figure 6.5: $K_f\,(-)$ vs $J\,(A/m)$

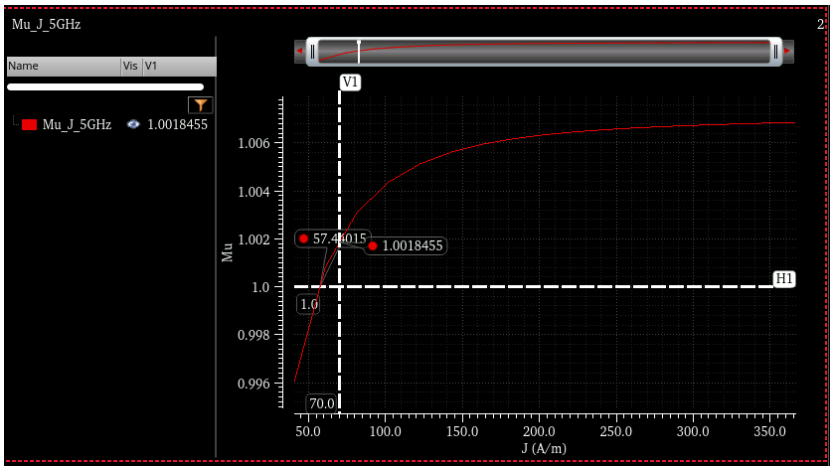

Figure 6.6: $\mu\,(-)$ vs $J\,(A/m)$

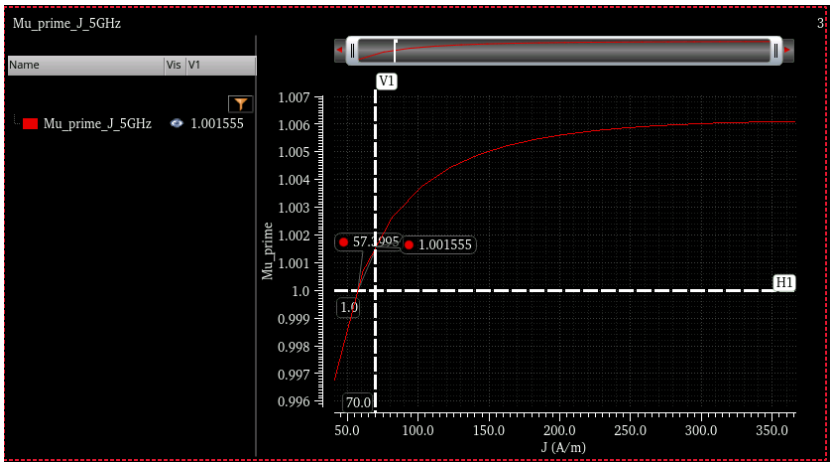

Figure 6.7: $\mu\prime\,(-)$ vs $J\,(A/m)$

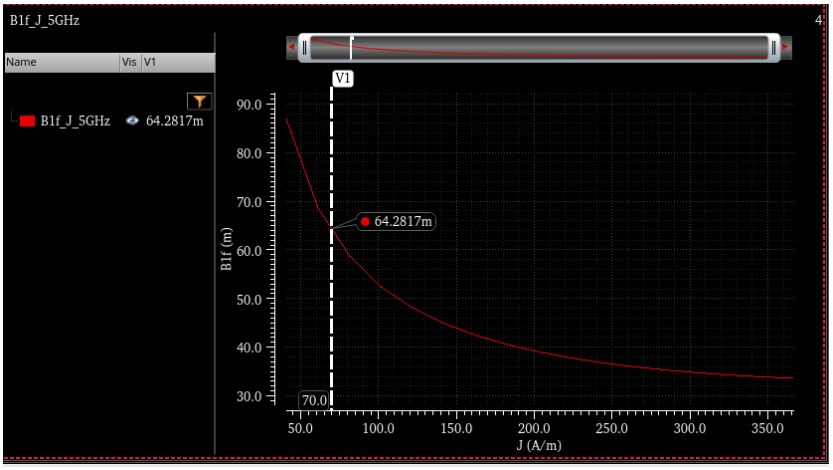

Figure 6.8: $B_{1f}\,(-)$ vs $J\,(A/m)$

Multiplier	J (mA/um)	I_{bias} (mA)	Z_{opt} (Ω)	$OP_{1\text{dB}}$ (dBm)	Gain (dB)	Z_{in} (Ω)
2048	0.0712	350	2.88+0.22j	19.98	8.00	0.348-0.052j
	0.102	500	2.88 + 0.22j	20.12	8.55	0.308-0.044j
1024	0.069	170	5.80 + 0.22j	17.23	6.92	0.673-0.810j
	0.102	250	2.88 + 0.22j	18.92	6.50	0.512-0.035j
512	0.069	85	9.07 + 0.21j	14.33	7.19	1.320-0.170j
	0.098	120	5.80 + 0.22j	15.77	6.42	1.040-0.081j
256	0.07	43	16.89+0.19j	11.38	6.98	2.580-0.319j
	0.099	61	12.733+0.204j	12.77	6.74	2.100-0.188j
128	0.0716	22	35.35+8.65j	8.39	7.39	5.350-0.389j
	0.098	30	27.10+0.15j	9.60	6.89	4.280-0.419j
64	0.0716	11	66.85+16.16j	5.42	7.22	10.540-0.724j
	0.098	15	55.8+6.835j	6.53	7.10	8.740-0.632j
32	0.0716	5.5	148	2.37	7.29	20.810-2.850j
	0.1	7.7	115.55	3.52	7.12	17.200-1.800j

Table 6.1

For each multiplier and bias current, a load pull simulation will be conducted to calculate the optimum load impedance (Z_{opt}) and $OP_{1\text{dB}}$. To perform a load pull simulation, harmonic balance analysis is used. The load pull simulation identifies the optimum load impedance, displayed on a Smith Chart, that allows the slvtnfet_rf device to achieve maximum current and voltage swing, thereby producing maximum power. Additionally, the achieved power gain will be recorded from the harmonic balance simulation. The input impedance Z_{in} of the device is calculated from SP analysis at 5 GHz by plotting the $Z_{m,1}$ metric. The input impedance will be used later to match the CS device with the input port of 50 Ohms at 5 GHz. The results are summarized in table 6.1.

From Table 6.1, it is observed that as the number of devices (multiplier) and bias current (I_{bias}) are doubled, Re$\{Z_{\text{in}}\}$ and Re$\{Z_{\text{opt}}\}$ decrease by a factor of 2. When the multiplier is doubled, the number of devices connected in parallel also doubles, resulting in the input resistance, Re$\{Z_{\text{in}}\}$, and output resistance, Re$\{Z_{\text{opt}}\}$, of the devices decreasing by a factor

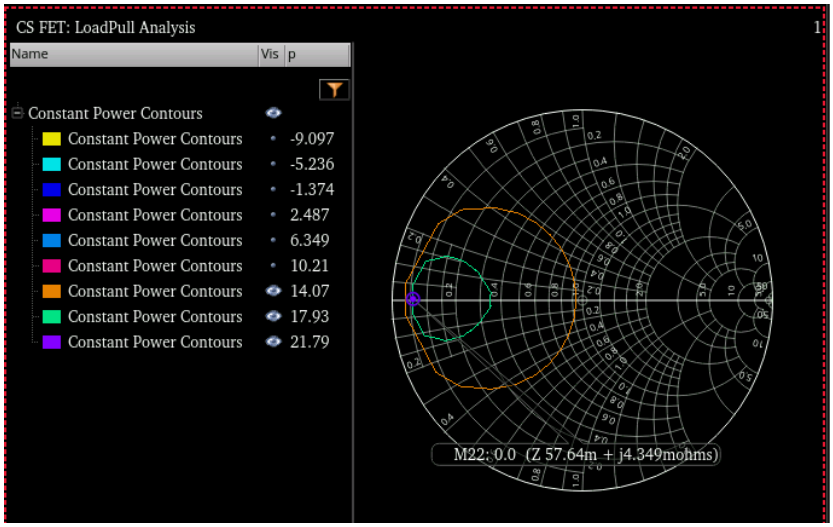

Figure 6.9: Load Pull Simulation of CS FET - Bias Point selected from table 6.1

of 2. Additionally, if Z_{in} is converted from series reactance to parallel, we observe that as the number of devices increases, the gate capacitance of the common-source (CS) FET also increases. This outcome is logical because the parasitic capacitances of the CS FET increase with the number of devices.

Similarly, the parasitic capacitance at the drain terminal increases with the multiplier of the device. For the CS FET to produce maximum power to the load, the voltage and current swings at the drain terminal must be in phase, meaning the impedance seen from the drain of the CS FET needs to be resistive. To cancel out the parasitic capacitances at the drain, an inductive load is required, and the parallel inductance increases as the multiplier increases [33]. This observation is validated by the data in Table 6.1.

As the number of devices (multiplier) and bias current (I_{bias}) are doubled, $OP_{1\text{dB}}$ increases by 3 dBm, except when the multiplier reaches 2048. This 3 dBm increase in $OP_{1\text{dB}}$ signifies that the power delivered to the load has doubled, which is expected due to the doubling of both the devices and bias current. The design goal is to maximize the power delivered to the load while minimizing power consumption. This objective leads to the design choice highlighted in Table 6.1, as for multipliers greater than 2048, the power delivered to the load no longer increases and plateaus.

On the other hand, a very large device indicates very low R_{opt}, which will be needing a very large cascode device to present this R_{opt} to the main device. The large cascode periphery will entail large parasitic capacitances, limiting the performance at higher frequencies. Additionally, a low R_{opt} will scale up the transistor totem to a value $N \cdot R_{\text{opt}}$ (where N is the number of cascodes), which if it is low, then the transformation to $50\,\Omega$ using the balun will not be an easy task due to transformation losses: these losses are due to low coupling coefficient values achieved, since primary and secondary windings will be very different in terms of diameter, resulting thus in low coupling coefficient values and thus lower power transferred to $50\,\Omega$ load.

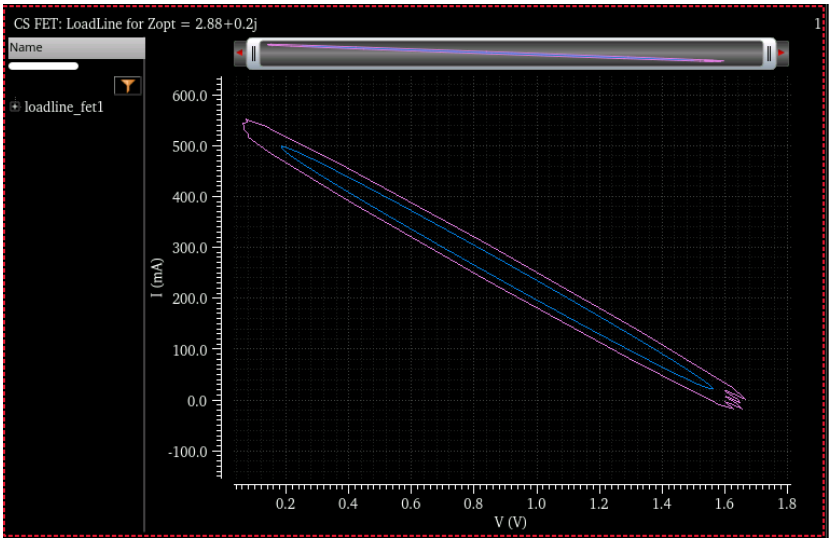

Figure 6.10: Loadline curves of CS FET for load $Z_{\mathrm{L},0} = 2.88 + 0.22j\ (\Omega)$

The load pull simulation of the CS FET at the bias point highlighted in table 6.1 is shown in figure 6.9. The optimum impedance of the CS FET, as shown in figure 6.9, is normalized to $Z_0 = 50\,\Omega$ and is calculated using the following equation:

$$Z_{\mathrm{opt},0} = (57.64 + 4.349j) \cdot 10^{-3} \cdot 50\ (\Omega) = 2.88 + 0.22j\ (\Omega) \tag{6.3}$$

The CS FET supports a maximum current swing that is double its bias current, $I_{\mathrm{swing}} \approx 2 \cdot I_{\mathrm{bias}} = 500\,\mathrm{mA}$. In terms of voltage, the maximum swing that the CS FET can accommodate is dictated by its drain-source maximum operating voltage, $V_{\mathrm{swing}} = 2 \cdot V_{max} = 2 \cdot 0.9\,\mathrm{V} = 1.8\,\mathrm{V}$. For the CS FET to efficiently transfer maximum power to the load, it's crucial that the load is capable of handling both the highest possible current swing and the highest possible voltage swing [28]. The load-line curves should be examined. If the load is purely resistive, the load-line will be a straight line. If not, and the load exhibits reactive components, the load-line will take on an elliptical shape.

The loadline curves for the load $Z_{\mathrm{L},0} = 2.88 + 0.22j\ (\Omega)$, as determined by the load-pull analysis, are examined in figure 6.10. From figure 6.10, the maximum voltage swing of the load is approximately 1.7 V, and the maximum current swing exceeds 500 mA. The appropriate load has a relatively higher resistance, as higher resistance decreases the current swing while increasing the voltage swing. The loadline curves are plotted using harmonic balance analysis, where each curve corresponds to a specific input power level $\mathrm{p_{in}}$ of the signal. The loadline curve that shows the maximum voltage and current swings represents input power levels that drive the power amplifier to the saturation of its output power.

The optimal load for maximizing the current and voltage swing is $Z_{\mathrm{L},0} = 3.5 + 0.2j\ (\Omega)$, as shown in figure 6.11. As indicated in Figure 6.11, this load supports a maximum voltage swing of 1.8V and a maximum current swing of 500 mA. The compression curve is presented in figure 6.12. The output-referred compression point, OP_{1dB}, reaches 19.5 dBm, which is higher than the corresponding highlighted value shown in table 6.1.

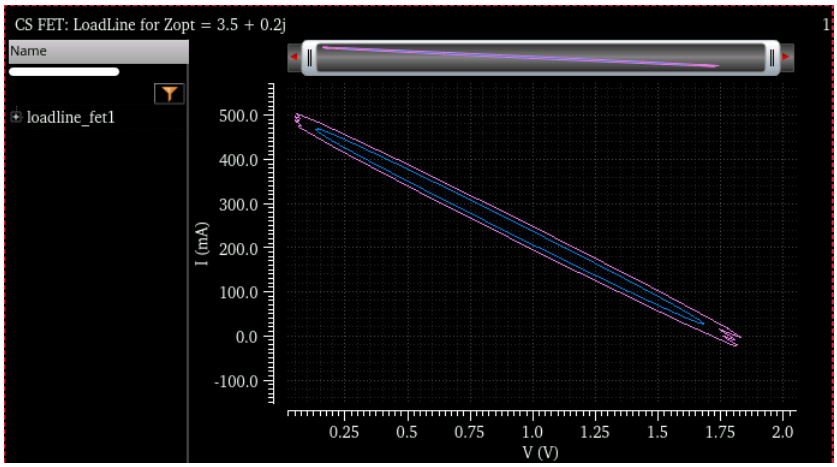

Figure 6.11: Loadline curves of CS FET for load $Z_{\mathrm{L},0} = 3.5 + 0.2j\ (\Omega)$

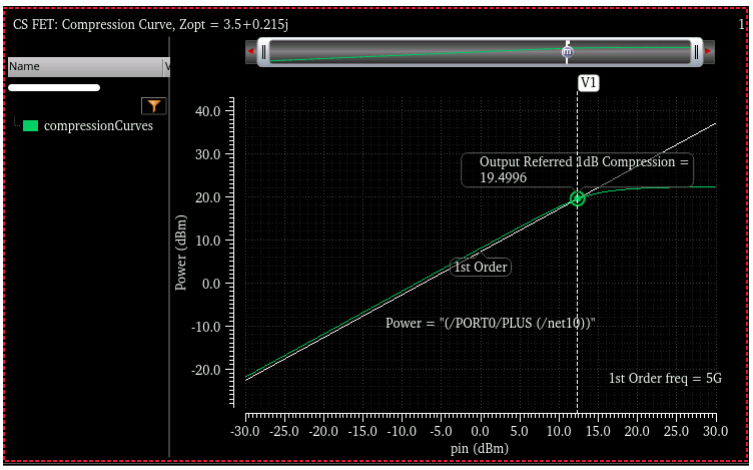

Figure 6.12: $OP_{1\mathrm{dB}} = 19.5\,\mathrm{dBm}$ of CS FET for $Z_{\mathrm{L},0} = 3.5 + 0.2j\ (\Omega)$

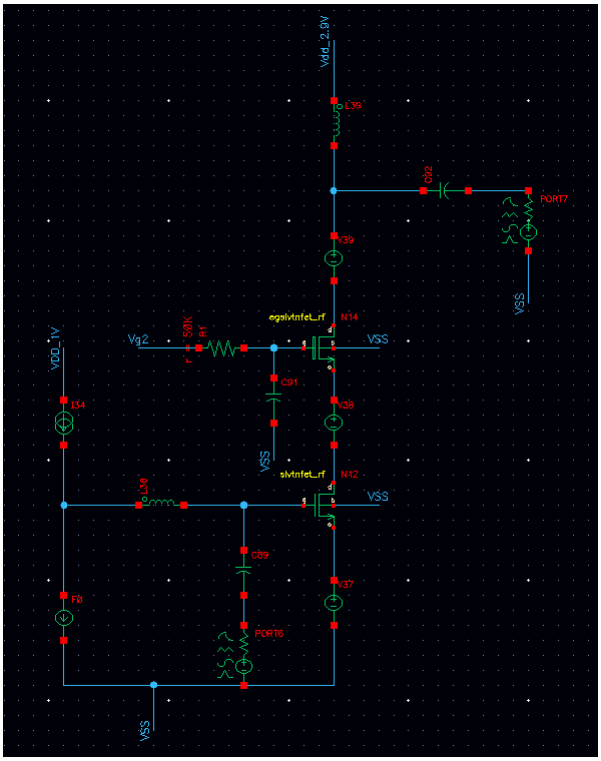

Figure 6.13: Schematic Design of 1-Stacked FET PA

To increase the output power delivered to the load, another device will be stacked onto the CS FET. Transistor stacking refers to a series interconnection of transistors where common-gate-like amplifiers are stacked on top of a common-source amplifier as shown in figure 6.13 [34]. The stacked configuration is similar to the cascode configuration, with the addition of a resistor and capacitor at the gate of the stacked device.

The capacitor at the gate is used to create a voltage swing that is in phase with the voltage swing at the drain. It also "shapes" the value of the input impedance "seen" when looking into source terminal of this common gate device, as described in equation 5.6. This improves the reliability of the stacked device by ensuring that $V_{dg} < 2 \cdot V_{\max}$, where $V_{\max}$ is the maximum power supply voltage of the device.

The stack topology enhances the voltage handling capability of the circuit and hence allows higher power supply voltage V_{DD} to be used for the PA. For an N-stack topology comprised of N identical transistors, V_{DD} can be scaled to $N \cdot V_{\max}$. However, the scaling of V_{DD} is on the condition that the gate bias voltages $V_{G,n}$, $(n = 1, 2, ..., N)$ are set correctly such that V_{DD} is distributed equally among the individual transistors in the stack (i.e., $V_{DS,1} = V_{DS,2} = ... = V_{DS,N}$).

With the equal DC drain source voltage $V_{DS,n}$ across each transistor, the RF drain-source voltages of the individual transistors $V_{ds,n}$, must also be tuned to have equal magnitude and phase by loading the gates of the upper-stack transistors with the proper values of capacitance. The gate capacitance of the n-th stacked device shapes its input impedance to match the

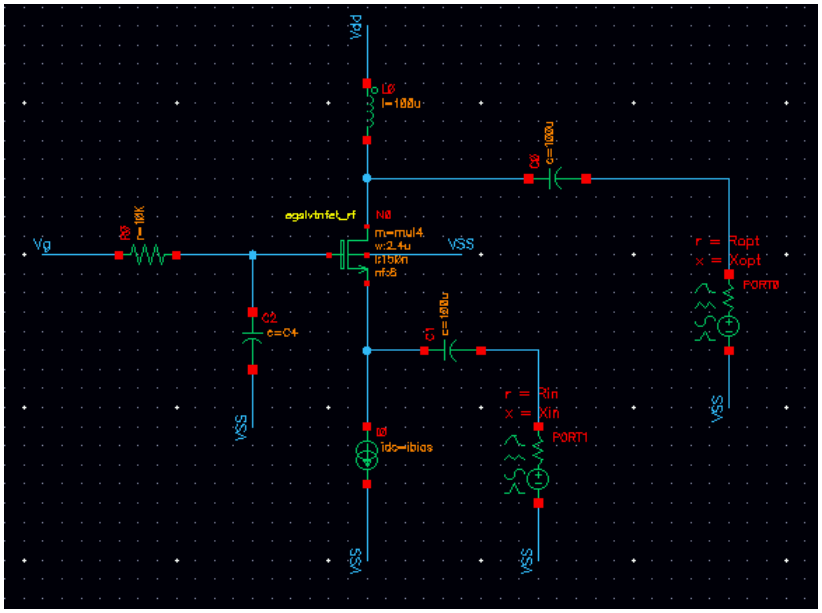

Figure 6.14: Schematic design for the calculation of the input impedance of a stacked FET

optimal output impedance of the $(n-1)$-th stacked device, $Z_{\mathrm{opt,n-1}}$. Ultimately, the equal distribution of both DC and RF voltages allows the individual transistors to operate within the drain-source breakdown voltage limit of a single transistor under a large-signal. The gate resistor in stacked FETs, isolates the DC bias voltage from the gate's voltage swing.

With the increase in the number of stacked FETs, the voltage swing at the drain of the topmost stacked FET increases. That's why larger voltage swing on the gate is needed, leading to a reduction in gate capacitance. This decrease in gate capacitance ultimately limits the practical number of stacked FETs that can be used [5].

In principle, by stacking N transistors, the RF output power can be increased by a factor of N from that of a single transistor. However, in reality, the output power is lower than the theoretical maximum value due to the losses in the transistors and non-ideal or out-of-phase drain-source voltage combining.

To ensure the CS FET delivers maximum power to the load, it must "see" the optimum load impedance Z_{opt} at its drain. This impedance is determined from load pull analysis and loadline curves. In our case, from prior analysis, $Z_{\mathrm{d,0}} = Z_{\mathrm{L,0}} = 3.5 + 0.2j\ (\Omega)$. To achieve this, the input impedance of the first Stacked FET, $Z_{\mathrm{s,1}}$, should match $Z_{\mathrm{d,0}}$.

The stacked device is egslvtnfet_rf (medium oxide super low threshold voltage nfet) and has an overdrive voltage $V_{\mathrm{max,egslv}} = 2V$. The egslvtnfet_rf device has $n_f = 8$, $w_f = 300\,\mathrm{nm}$ and $L = 150\,\mathrm{nm}$. From the above, to produce the maximum power to the load, each stacked FET must be biased to its overdrive voltage. The first stacked FET will have a bias voltage at the drain

$$V_{\mathrm{D,1}} = V_{DS,1} + V_{DS,0} = V_{\mathrm{max,egslvt}} + V_{\mathrm{max}} = 2.9V \tag{6.4}$$

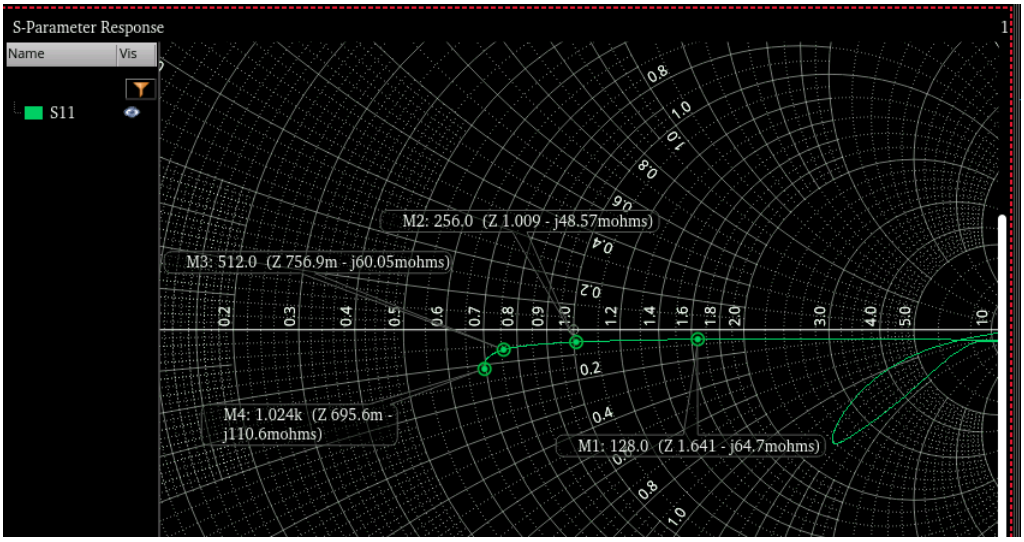

Figure 6.15: Input Impedance of first stacked FET, $Z_{\text{in},1}$, versus multiplier on a Smith Chart.

The input impedance of the first stacked FET is measured using the schematic shown in Figure 6.14. In this configuration, the input impedance is measured at the source of the first stacked FET using SP analysis. The optimum output impedance is connected at the drain of the first stacked FET, as it influences the input impedance of this device. Additionally, the gate capacitance of the first stacked FET serves as AC coupling. It is initially set to a large value, 3 pF, to minimize its impact on the input impedance. The DC bias voltages and currents in the configuration shown in figure 6.14 are consistent with those used in the 1-Stacked PA.

As previously mentioned, the input impedance of the first stacked FET must match the optimum load impedance of the CS FET, as determined through power matching. To accomplish that, the multiplier of the first stacked FET is varied to ensure that $\text{Re}\{Z_{\text{in},1}\} \simeq \text{Re}\{Z_{\text{opt},0}\}$. The Smith Chart, as illustrated in figure 6.15, is normalized to $\text{Re}\{Z_{opt,0}\} = 3.5\,\Omega$, and it depicts the input impedance of the first Stacked FET as a function of its multiplier. The input impedance is measured using SP analysis at the operation frequency of the stacked PA, 5 GHz. As shown in figure 6.15, the input impedance of the stacked FET is capacitive and the optimum load impedance of the CS FET is inductive. This indicates an interstage mismatch between the CS FET and the first stacked FET, which introduces power loss [35]. This is why only the real parts of the above impedances can be matched.

It is evident from the figure that the multiplier must be 256 or higher to approximately match the input impedance of the first stacked FET with the optimum load impedance of the CS FET. To determine the multiplier of the first stacked FET (256, 512, or 1024), the OP_{1dB}, PAE, and power gain will be compared, with the primary criterion being OP_{1dB}. For a multiplier of 256, after performing load pull analysis and evaluating the loadline curves, the optimum load for the 1-Stacked PA is found to be $Z_{\text{opt},1} = 11 + 0.2j$. The loadline curves are presented in Figure 6.16, demonstrating that the two FETs achieve maximum voltage and current swing. The compression curve, power-added efficiency (PAE), and power gain are obtained using harmonic balance analysis at 5 GHz. The compression curve and PAE are

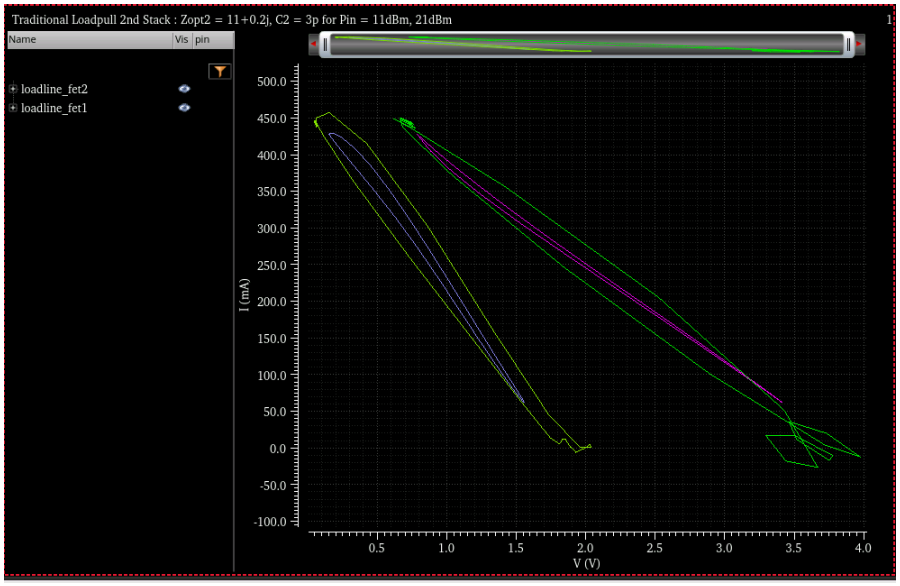

Figure 6.16: Loadline curves of CS FET (left) and first stacked FET (right) with multiplier of 256 for load $Z_{\text{opt},1} = 11 + 0.2j$.

Multiplier	256	512	1024
$OP_{1\text{dB}}$ (dBm)	23.2	24.2	24.5
PAE (%)	43.9	52.6	54.55
Power Gain (dB)	12.4	13.3	13.3

Table 6.2: Metrics of 1-Stacked PA in function of the multiplier of the first stacked FET

shown in Figures 6.17 and 6.18, respectively.

The same procedure is followed for a multiplier of 512. The optimum load, determined through load pull analysis and loadline curves, remains unchanged at $Z_{\text{opt},1} = 11 + 0.2j$. The loadline curves are shown in figure 6.19. Additionally, the compression curve and power-added efficiency (PAE) for the operating frequency of 5 GHz are presented in figures 6.20 and 6.21, respectively. The output-referred 1 dB compression point ($OP_{1\text{dB}}$) increases by 1 dB, and the PAE improves by 8%, as demonstrated in the referenced figures.

The optimum load remains unchanged for the multiplier of 1024, as demonstrated by the loadline curves in Figure 6.22. The compression curve and power-added efficiency (PAE) are shown in Figures 6.23 and 6.24, respectively. Doubling the stacked FET devices does not result in an improvement in output compression or PAE. The power gain achieved for the 1-Stacked PA with each multiplier is illustrated in Figure 6.25. Both the 512 and 1024 devices of the stacked FET achieve the same power gain, which is 1 dB higher than that of the 256 devices.

The metrics achieved by the 1-Stacked PA based on the multiplier of the stacked FET are summarized in table 6.2. From Table 6.2, the configuration with a multiplier of 512 is selected due to its improved PAE, $OP_{1\text{dB}}$, and power gain compared to the 256-device configuration. Additionally, doubling the number of devices to 1024 does not further enhance these metrics,

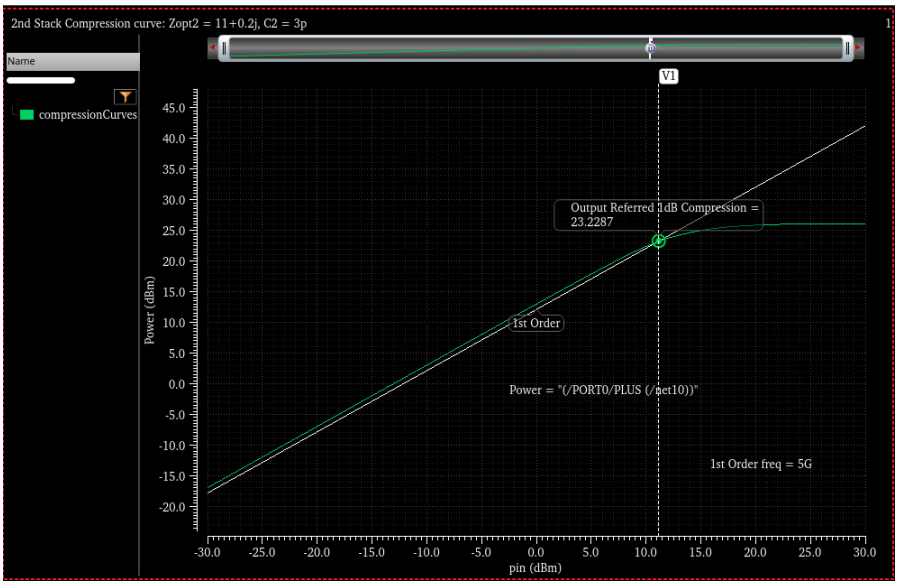

Figure 6.17: Compression curve of 1-Stacked PA for multiplier of 256, $OP_{1\mathrm{dB}} = 23.23\,\mathrm{dBm}$

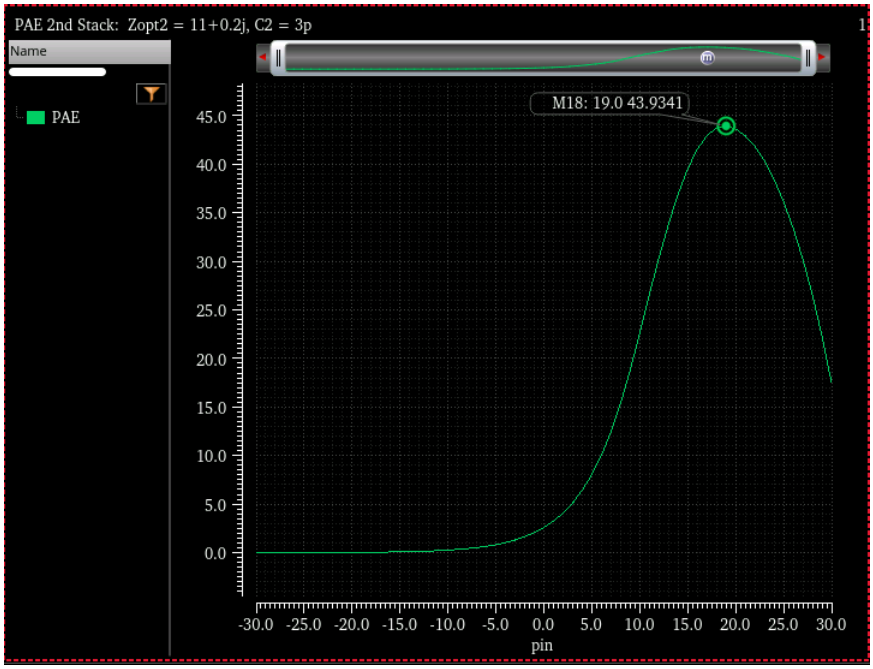

Figure 6.18: PAE of 1-Stacked PA for multiplier of 256, PAE = 43.9%

Figure 6.19: Loadline curves of CS FET (left) and first stacked FET (right) with multiplier of 512 for load $Z_{\mathrm{opt},1} = 11 + 0.2j$.

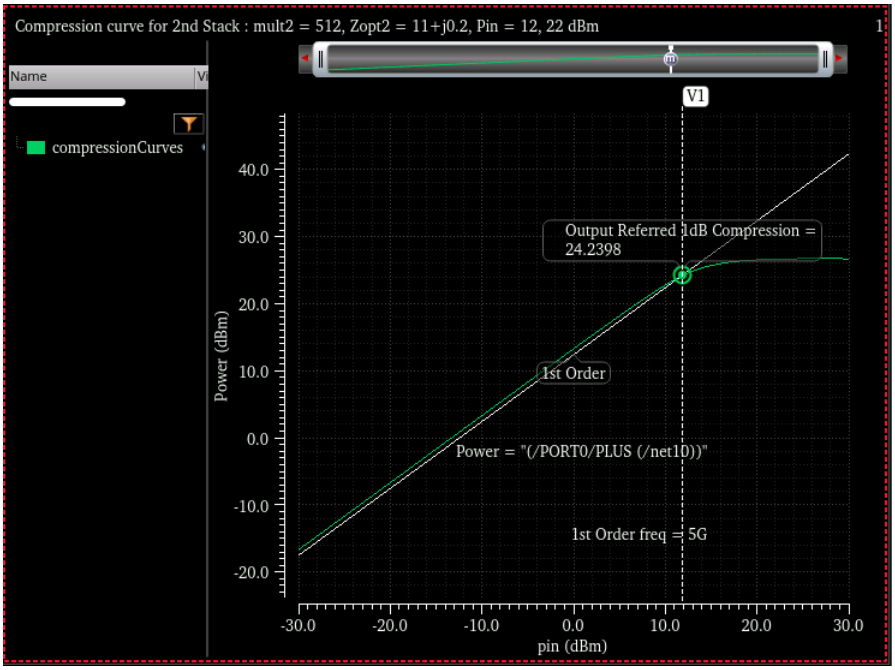

Figure 6.20: Compression curve of 1-Stacked PA for multiplier of 512, $OP_{\mathrm{1dB}} = 24.24\,\mathrm{dBm}$

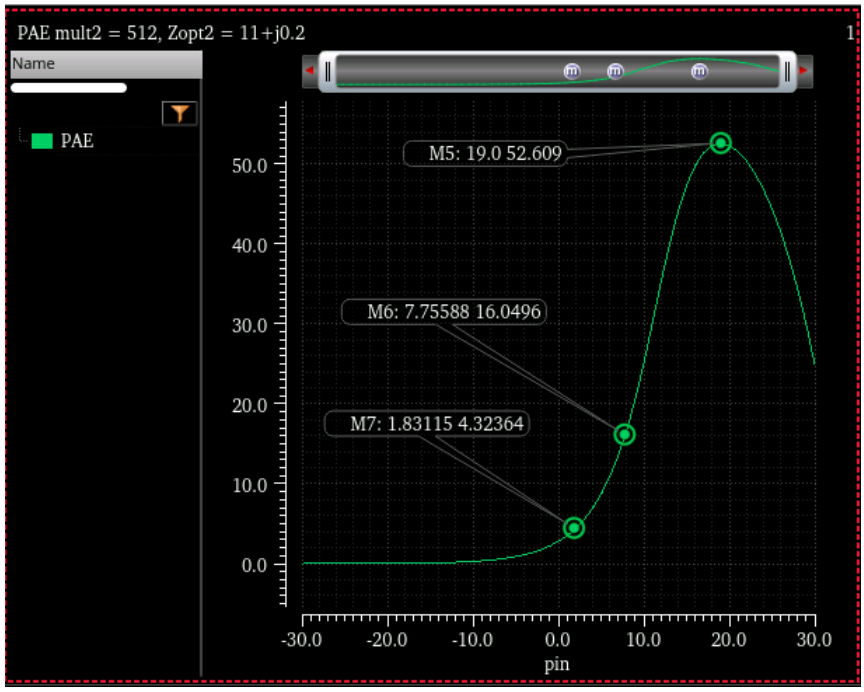

Figure 6.21: PAE of 1-Stacked PA for multiplier of 512, PAE = 52.6%

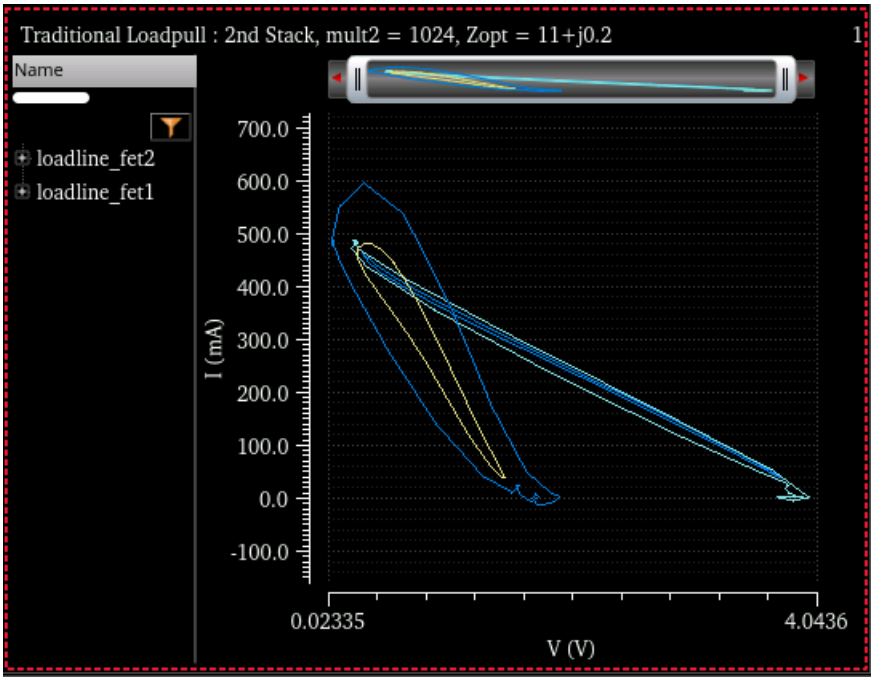

Figure 6.22: Loadline curves of CS FET (left) and first stacked FET (right) with multiplier of 1024 for load $Z_{\mathrm{opt},1} = 11 + 0.2j$.

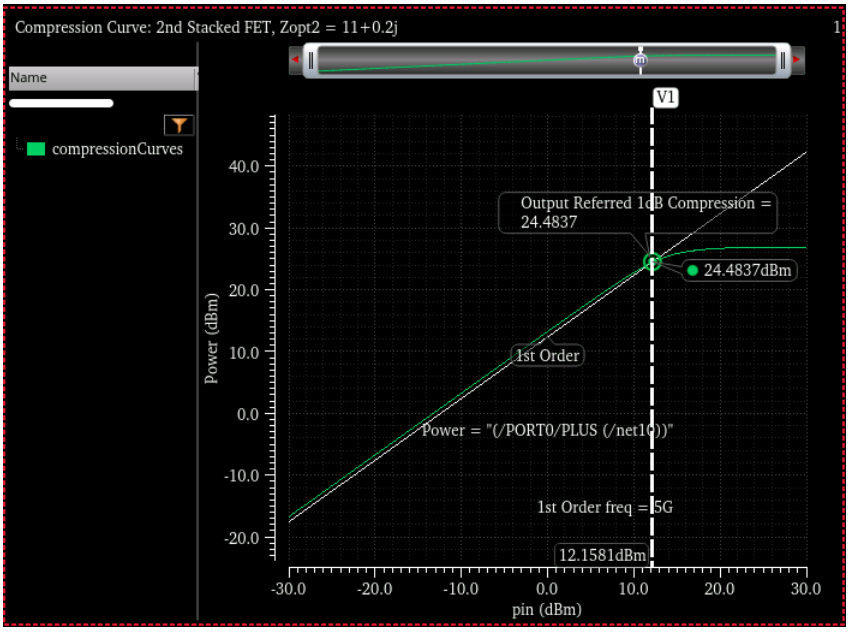

Figure 6.23: Compression curve of 1-Stacked PA for multiplier of 1024, $OP_{1\mathrm{dB}} = 24.5\,\mathrm{dBm}$

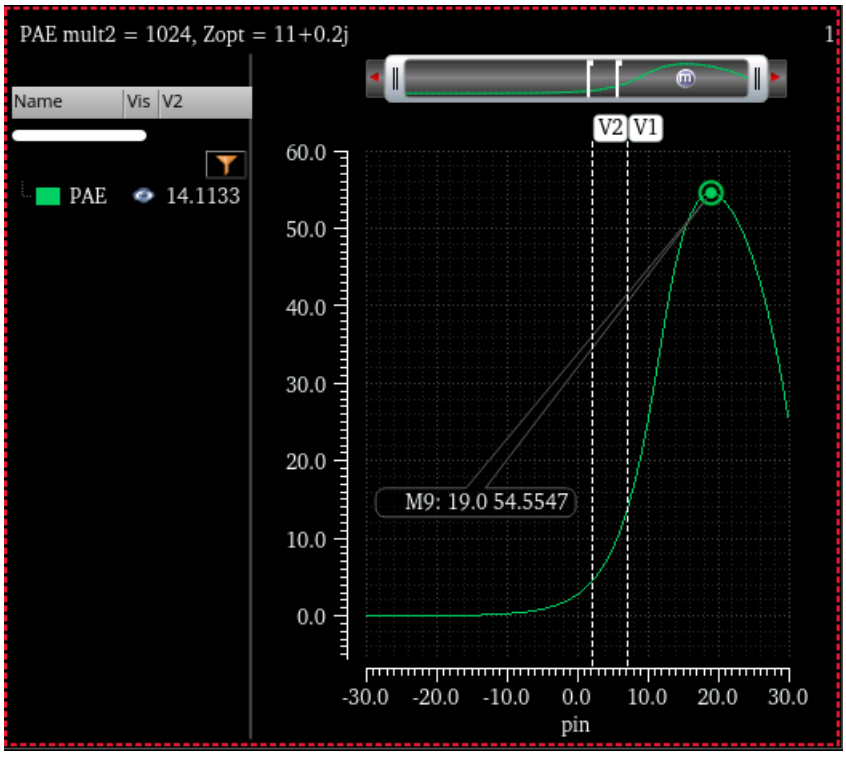

Figure 6.24: PAE of 1-Stacked PA for multiplier of 1024, PAE = 54.5%

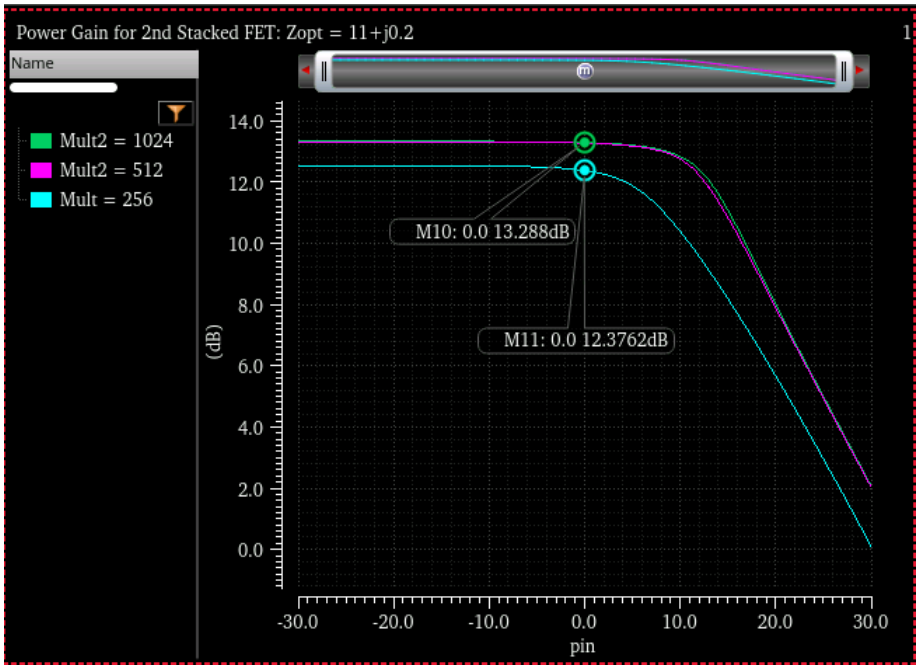

Figure 6.25: Power Gain of 1-Stacked FET PA for multipliers 256, 512 and 1024.

as they remain unchanged.

The shunt capacitor at the gate of the second stacked FET is set to 1.5 pF. This capacitor is chosen to ensure that, at the 1-dB compression point, the V_{gd} and V_{ds} do not exceed their breakdown limits. For the CS FET (thin oxide slvtnfet_rf), $V_{\text{max,slvt}} = 0.9\,\text{V}$, implying that $V_{\text{gd,ds}} \leqslant 1.8\,\text{V}$. For the first stacked FET (medium oxide egslvtnfet_rf), $V_{\text{max,egslvt}} = 2\,\text{V}$, implying that $V_{\text{gd,ds}} \leqslant 4\,\text{V}$.

Theoretically, with the addition of the first stacked FET device, the supply voltage increases to 2.9 V from 0.9 V, implying that the output power delivered to the load must increase by a factor of $\frac{2.9}{0.9} = 3.2$, or $10 \cdot \log(3.2) = 5.05\,\text{dB}$. The increase in $OP_{1\text{dB}}$ with the addition of the stacked FET is 4.74 dB, comparing figures 6.12 and 6.20. This result is very close to the theoretical value.

A second `egslvtnfet_rf` device will be stacked, increasing the supply voltage of the 2-Stacked PA by $V_{\text{max,egslvt}} = 2\,\text{V}$, resulting in $V_{DD} = 4.9\,\text{V}$. The schematic design of the 2-Stacked PA is illustrated in figure 6.26.

The multiplier of the second stacked FET is swept to match the input impedance of the second stacked FET to the optimum load of the 1-Stacked PA, ensuring that $\text{Re}\{Z_{\text{in,2}}\} \approx \text{Re}\{Z_{\text{opt,1}}\} \approx 11\,\Omega$. As shown in figure 6.27, the appropriate multipliers are 128, 256, 512, and 1024, which allow for impedance matching. This approach ensures the input impedance of the second stacked FET aligns with the optimum load impedance of the preceding stage, maintaining performance and minimizing power loss.

For a multiplier of 128, $V_{\text{G,2}} > 4.9\,\text{V}$ is required to achieve $V_{\text{d,1}} = 2.9\,\text{V}$. Consequently, $V_{\text{gs,2}} > 2\,\text{V}$, causing the gate-source junction to exceed its breakdown voltage. Therefore, a

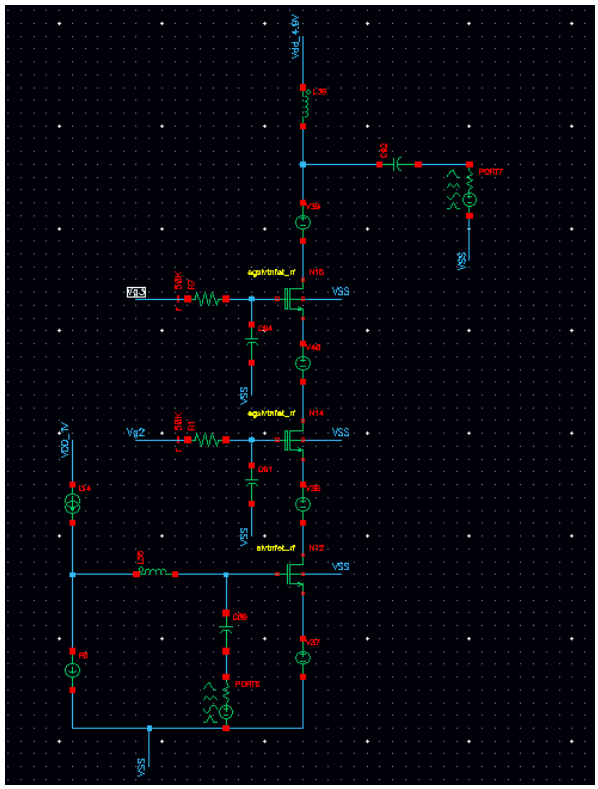

Figure 6.26: Schematic Design of 2-Stacked FET PA

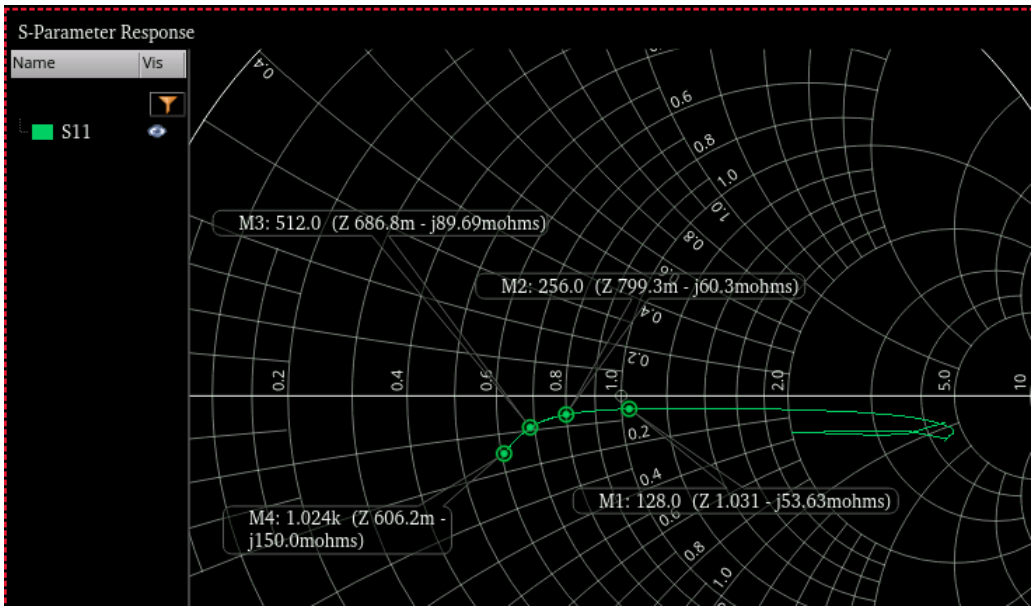

Figure 6.27: Input impedance of the 2nd Stacked FET versus the multiplier. The characteristic impedance of the Smith Chart is $Z_0 = \mathrm{Re}\{Z_{\mathrm{opt},1}\} = 11\,\Omega$.

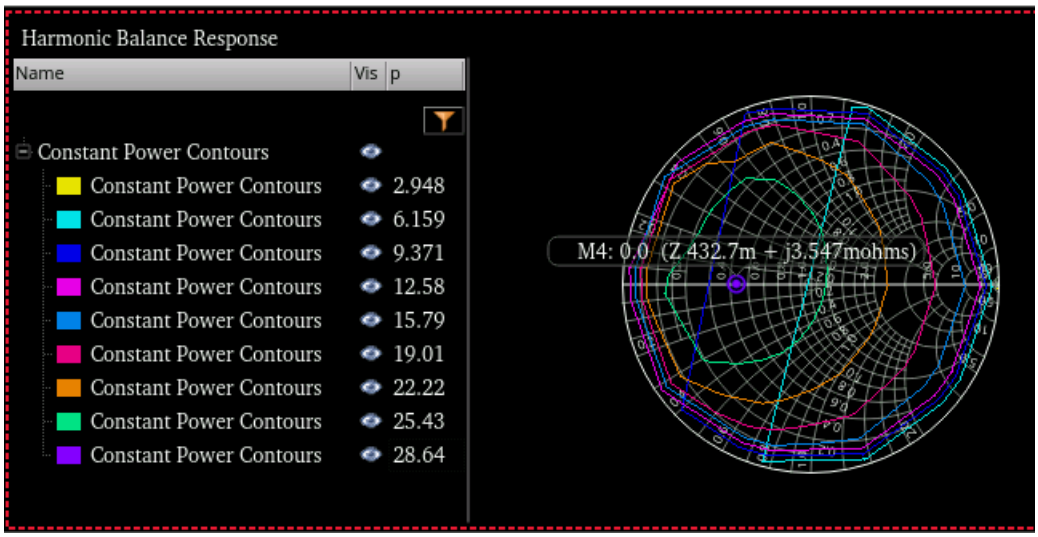

Figure 6.28: From Load-Pull Analysis: $Z_{\text{opt},2} = (432.7 + 3.547 \cdot j) \cdot 10^{-3} \cdot 50\,(\Omega) = 21.635 + 0.177j\,(\Omega)$

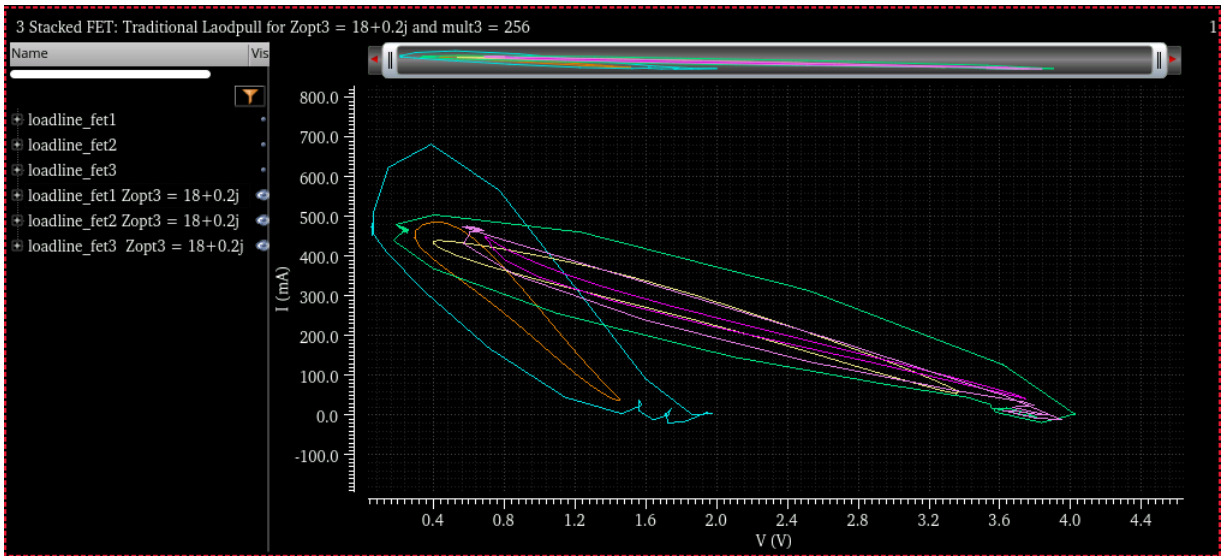

Figure 6.29: Load-line curve for $Z_{opt,2} = 18 + 0.2j\,(\Omega)$

multiplier of 128 is not selected.

For a multiplier of 256, load pull analysis is conducted, and the Smith Chart is presented in figure 6.28. The optimum load for the 2-Stacked PA, with the topmost FET having a multiplier of 256, is $Z_{\text{opt},2} = 21.635 + 0.177j\,(\Omega)$. However, the load line curve indicates that the current swing of the 2nd Stacked FET is low, suggesting the need for a smaller impedance.

Figure 6.29 demonstrates that the load $Z_{opt,2} = 18 + 0.2j\,(\Omega)$ achieves the maximum current and voltage swing across all stacked FETs, making it the optimal load for power matching. This load is also optimal for multipliers of 512 and 1024.

The metrics achieved by the 2-Stacked PA based on the multiplier of the second stacked FET are summarized in table 6.3. From table 6.3, the configuration with a multiplier of 256 is selected because increasing the number of devices to 512 or 1024 does not further improve the

Multiplier	256	512	1024
$OP_{1\text{dB}}$ (dBm)	26.1	26.3	26.2
PAE (%)	51.2	53.0	52.6
Power Gain (dB)	15.4	15.3	15.0

Table 6.3: Metrics of 2-Stacked PA in function of the multiplier of the second stacked FET

above metrics. The shunt capacitance at the gate of 2nd stacked FET, designed to prevent breakdown for $P_{\text{in}} \leqslant IP_{1\text{dB}}$ is $C_2 = 150\,\text{fF}$. This is validated by transient and harmonic balance analysis, plotting the gate and drain voltages of the stacked FETs and subtracting them for each stacked FET. $OP_{1\text{dB}}$ is increased by 1.9 dB with the addition of the 2nd Stacked FET, compared to the results in table 6.2. Theoretically, the addition of the second stacked FET should enhance the $OP_{1\text{dB}}$ by $10 \cdot \log \frac{V_{DD,\text{2-stacked}}}{V_{DD,\text{1-stacked}}} = 2.3\,\text{dB}$, which aligns closely with the experimental result.

The stacked PA will be power matched to an output port of $50\,\Omega$. This requires an output balun with a transformation ratio given by:

$$n = \sqrt{\frac{|Z_{\text{port}}|}{|Z_{\text{opt,2}}|}} = \sqrt{\frac{50}{\sqrt{18^2 + 0.2^2}}} = 1.67 \tag{6.5}$$

This implies that a 1:2 output balun will be used. However, such a balun would have a low coupling coefficient, resulting in significant power loss [17]. To address this issue, a third stacked FET will be added to increase Z_{opt}, enabling a 1:1 transformation ratio [36].

The schematic design of the 3-Stacked PA is shown in figure 6.30. The supply voltage is increased by $V_{\text{max,egslvtnfet}} = 2\,\text{V}$, resulting in $V_{\text{DD}} = 6.9\,\text{V}$. The multiplier of the third stacked device, $Z_{\text{in,3}}$, is swept to match its input impedance to the optimum output impedance of the 2-Stacked PA, $Z_{\text{opt,2}}$. The input impedance as a function of the multiplier is illustrated in figure 6.31 using a Smith Chart normalized to $\text{Re}\{Z_{\text{opt,2}}\} = 18\,\Omega$. It is evident that appropriate multipliers are 128, 256, 512, and 1024, ensuring that $\text{Re}\{Z_{\text{in,3}}\} \simeq \text{Re}\{Z_{\text{opt,2}}\}$.

For a multiplier of 256, $V_{\text{g,3}} > 6.9\,\text{V}$ is required to achieve $V_{\text{s,4}} = 4.9\,\text{V}$, causing the DC gate-source voltage to exceed its breakdown limit. Therefore, a multiplier of 256 is not selected, leading to the choice of a larger multiplier to reduce the drain-source resistance. Ultimately, a multiplier of 512 is chosen, as increasing the number of devices in the third stacked FET does not improve the performance of the 3-Stacked PA, as shown in table 6.4. After load pull analysis and loadline curves, the optimum load for power matching is determined to be $Z_{\text{opt,3}} = 27 + 11j\,\Omega$. The transformation ratio is calculated as

$$n = \sqrt{\frac{|Z_{\text{port}}|}{|Z_{\text{opt,3}}|}} = \sqrt{\frac{50}{\sqrt{27^2 + 11^2}}} = 1.3 \tag{6.6}$$

This results in 1:1 impedance transformation ratio for the output balun, minimizing transformation loss.

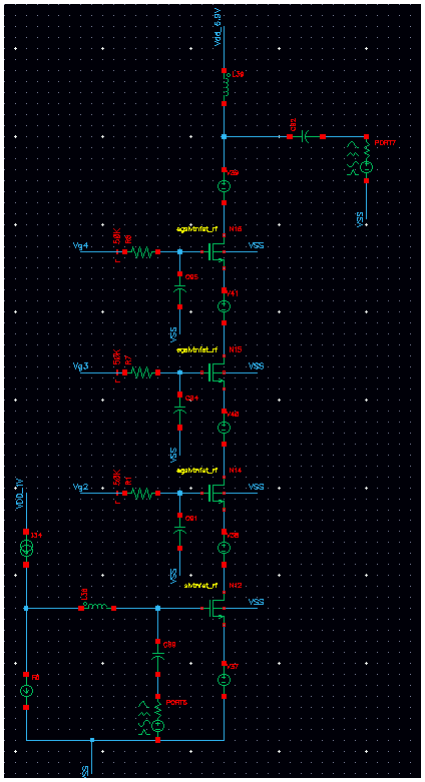

Figure 6.30: 23) Schematic Design of 3-Stacked PA

Multiplier	512	1024
$OP_{1\text{dB}}$ (dBm)	27.4	27.6
PAE (%)	52	54.1
Power Gain (dB)	17	17.5

Table 6.4: Metrics of 3-Stacked PA in function of the multiplier of the third stacked FET

The loadline curves for the CS FET and stacked FETs are shown in figures 6.32–6.35. Additionally, the compression curve, PAE, and power gain achieved by the 3-Stacked PA are shown in figures 6.36–6.38, respectively. The $OP_{1\text{dB}}$ of the 3-Stacked PA is increased by 1.3 dB compared to the 2-Stacked PA, which is close to the theoretical value of $10 \cdot \log \frac{V_{\text{DD,3-stacked}}}{V_{\text{DD,2-stacked}}} = 1.48\,\text{dB}$.

The single-ended amplifier design becomes differential, to enhance the linearity of the 3-Stacked PA and increase the $OP_{1\text{dB}}$ by 3 dB, as explained in section 2.9. The schematic design of the differential 3-Stacked PA is shown in figure 6.39. The single-ended 3-Stacked PA has been duplicated, and ideal baluns are inserted in input and output to convert between single ended and differential signals. The bias at the gate of the CS FETs is given from the center tap of the ideal balun in the input, using a diode-connected transistor and forming a current mirror. The bias at the drains of the third stacked FETs is given from the center tap of the output balun. The input and output ports have an impedance of 50 Ω. The differential 3-Stacked PA is going to be power matched to the output port and conjugated matched to

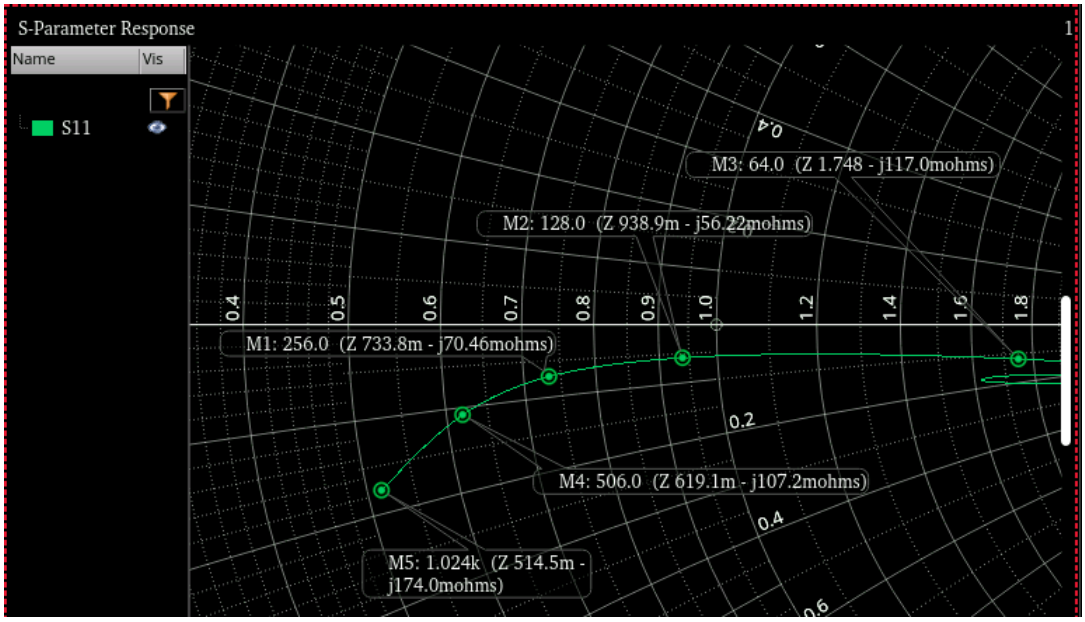

Figure 6.31: Input impedance of the 3rd Stacked FET versus the multiplier. The characteristic impedance of the Smith Chart is $Z_0 = \mathrm{Re}\{Z_{\mathrm{opt},2}\} = 18\,\Omega$.

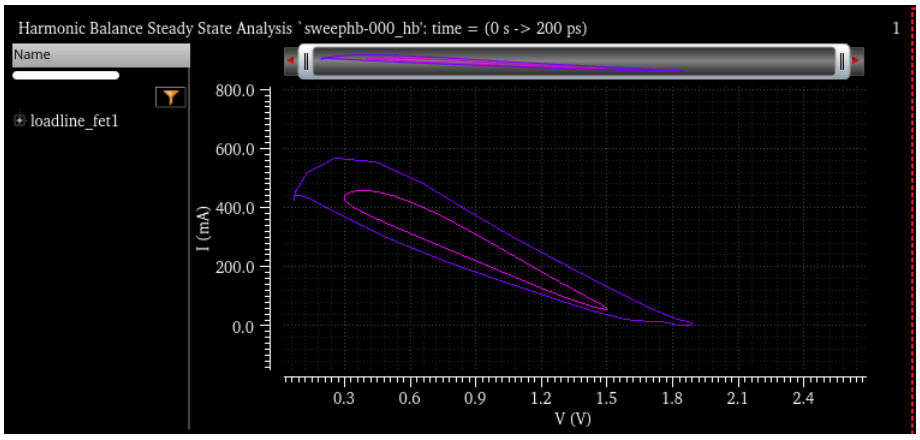

Figure 6.32: Loadline curve of CS FET

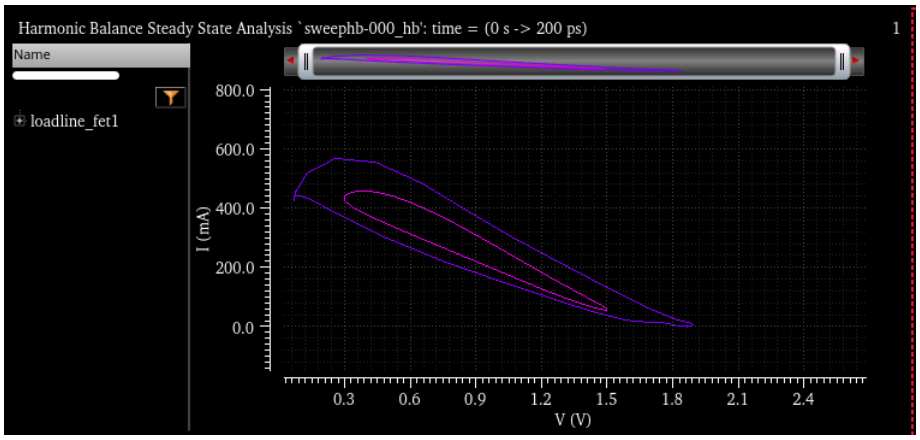

Figure 6.33: Loadline curve of first Stacked FET

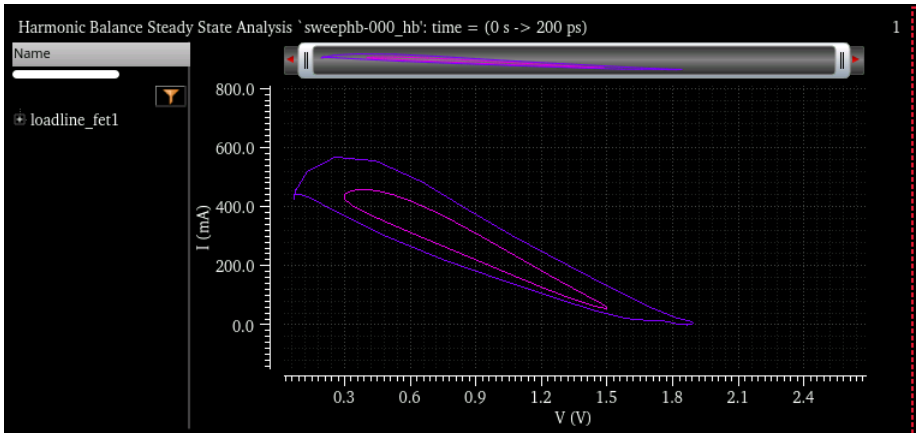

Figure 6.34: Loadline curve of second Stacked FET

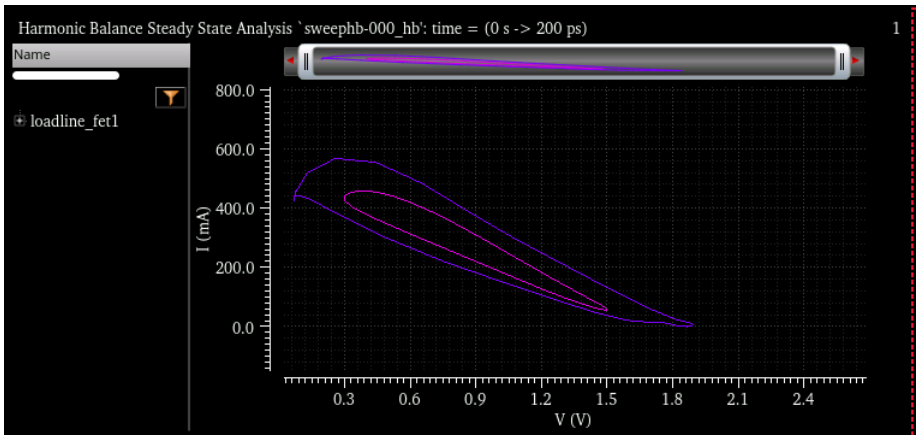

Figure 6.35: Loadline curve of third Stacked FET

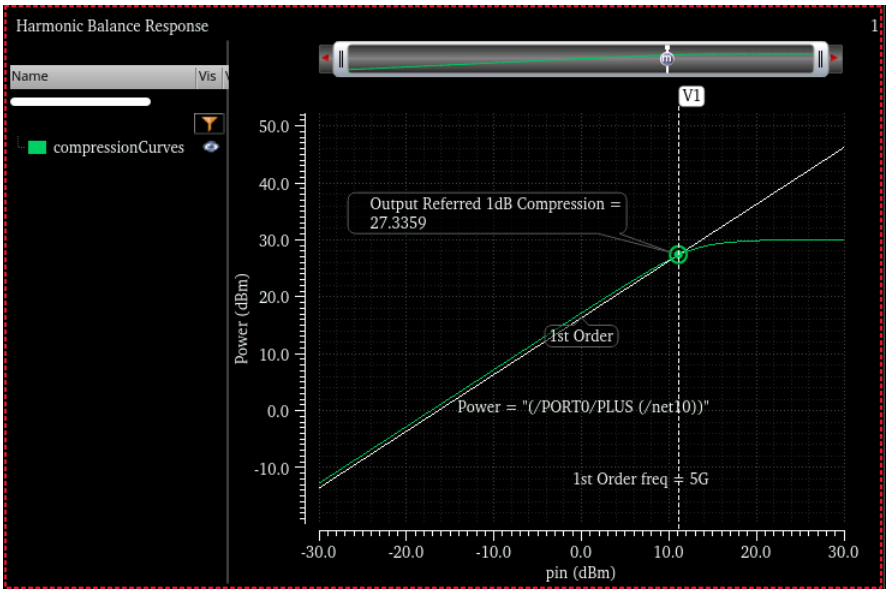

Figure 6.36: Compression curve of 3-Stacked PA, $OP_{1\mathrm{dB}} = 27.3\,\mathrm{dBm}$

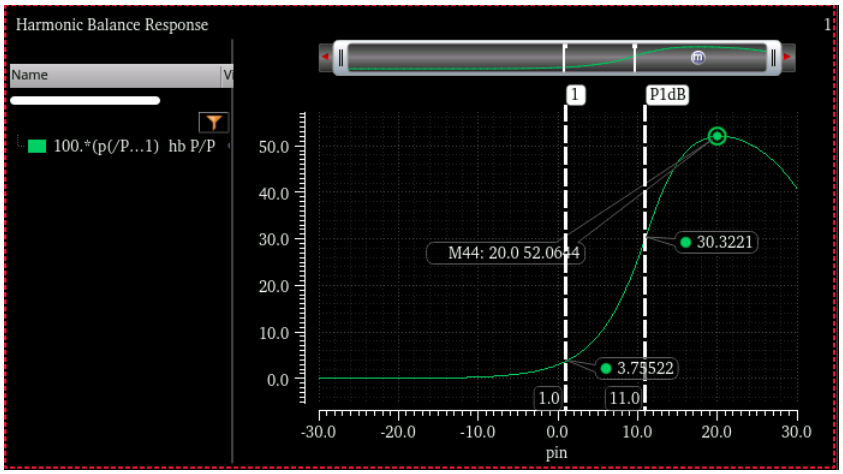

Figure 6.37: PAE of 3-Stacked PA, PAE = 52.1%

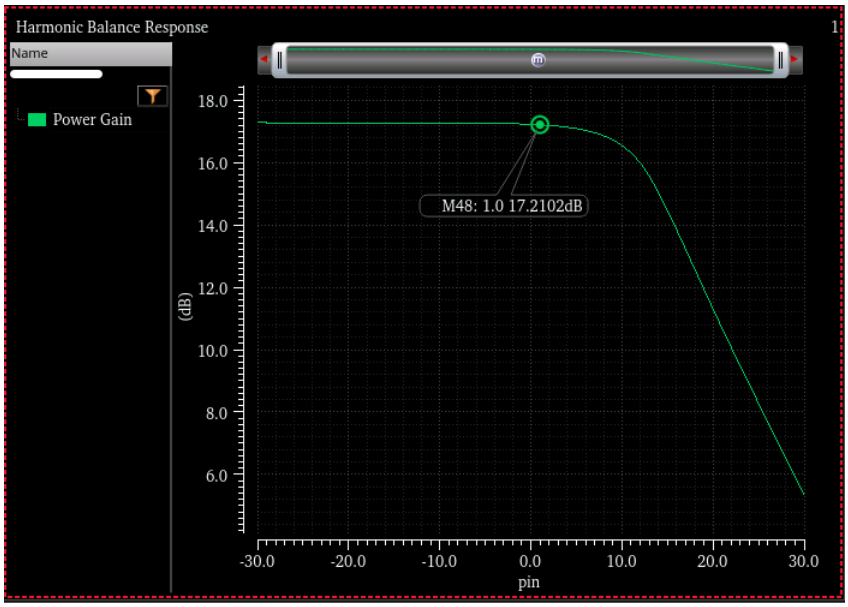

Figure 6.38: Power Gain of 3-Stacked PA, $P_{\text{gain}} = 17.2\,\text{dB}$

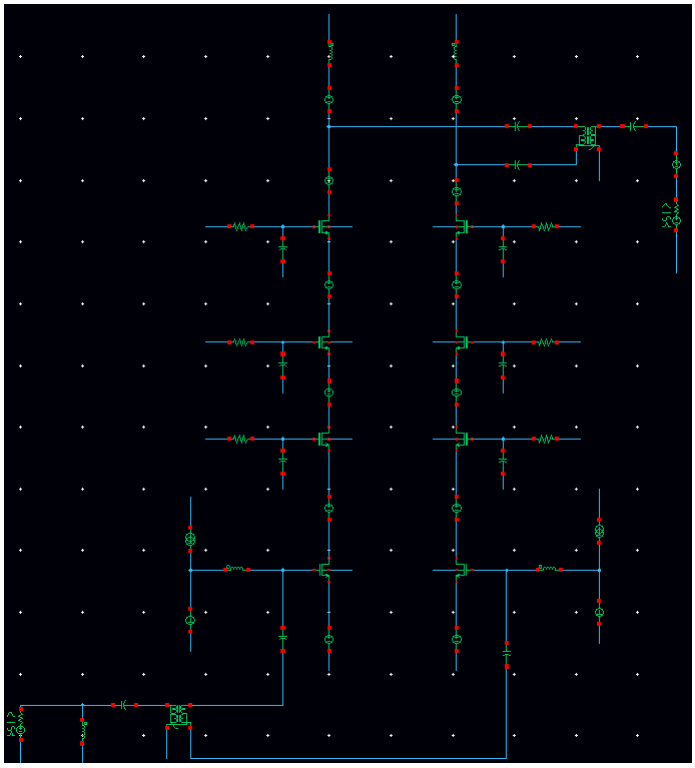

Figure 6.39: Schematic Design of the Differential 3-Stacked FET PA

input part using a L-matching network. The differential 3-Stacked PA is going to operate at n79 5G NR frequency band, which is at 4.4-5 GHz [37].

The single-ended amplifier design is converted to a differential design to enhance the linearity of the 3-Stacked PA and increase the $OP_{1\text{dB}}$ by 3 dB, as explained in section 2.9. The schematic design of the differential 3-Stacked PA is shown in figure 6.39. The single-ended 3-Stacked PA has been duplicated, and ideal baluns are inserted at the input and output to convert between single-ended (unbalanced) and differential (balanced) signals. The gate voltage bias of the CS FETs is provided from the center tap of the ideal balun at the input, using a diode-connected transistor to form a current mirror. The drain voltage bias of the third stacked FETs is supplied from the center tap of the output balun. The input and output ports each have an impedance of $50\,\Omega$.

The differential 3-Stacked PA will be power matched to the output port and conjugated matched at the input using an L-matching network. It is designed to operate in the n79 5G NR frequency band, which spans 4.4–5 GHz [37].

As indicated by the load pull analysis and loadline curves, the optimum load for the differential 3-Stacked PA is $Z_{\text{opt}} = 50 + 8.8j\,\Omega$. Consequently, a balun with a 1:1 transformation ratio is required at the output to power match the PA to the output port impedance of $50\,\Omega$. With the optimum load impedance applied at the output port, harmonic balance analysis demonstrates a 3 dB increase in the output 1 dB compression point, achieving $OP_{1\text{dB}} = 30.86$ dBm at 5 GHz.

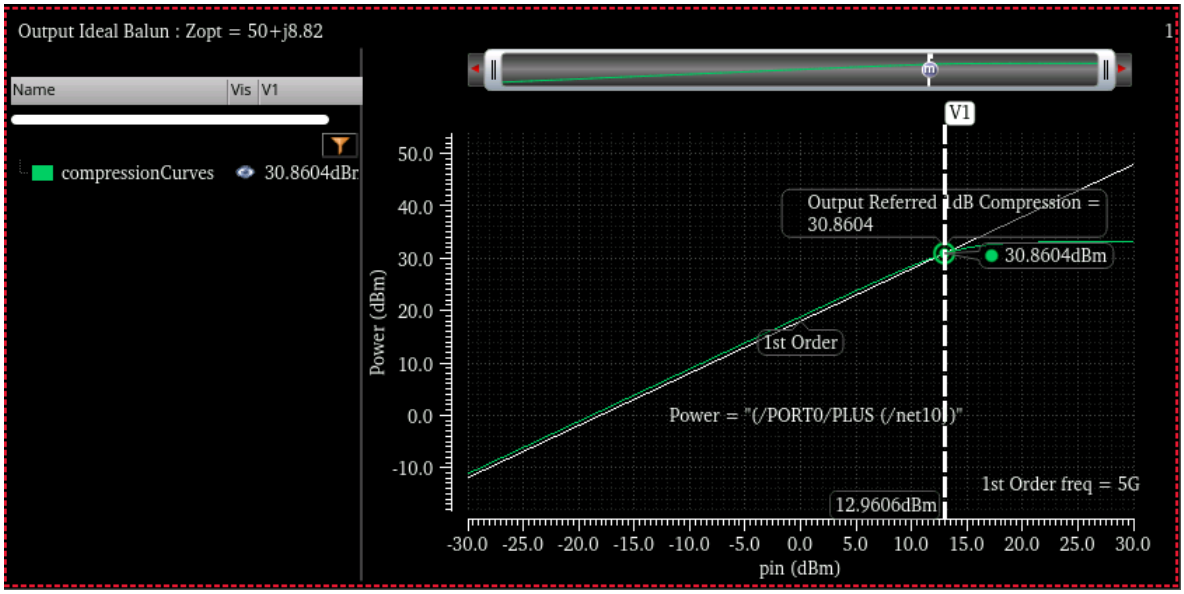

Figure 6.40: Compression curve of differential 3-Stacked PA with ideal balun at output, $OP_{1\mathrm{dB}} = 30.86$ dBm.

To validate that 1:1 transformation ratio is needed to the output port, an ideal xfmr (transformer) is used. The primary and secondary turns are set $n_1 = n_2 = 2$, respectively. The schematic design using the xfmr at the output is shown in figure 6.41. It is known that:

$$Z_{\mathrm{p}} = \left(\frac{n_1}{n_2}\right)^2 \cdot Z_{\mathrm{L}} \tag{6.7}$$

where Z_p is the impedance observed from primary coil and Z_L is the load impedance, $Z_{\mathrm{L}} = 50\,\Omega$.

The 1:1 impedance transformation ratio results in full current and voltage swing for the CS FET and 3 Stacked FETs, as shown in figures 6.41-6.44. According to the figures, all the FETs exhibit elliptic loadline curves with full current and voltage swing.

The ideal xfmr will be replaced with a non-ideal balun model. The circuit schematic of the non-ideal balun and it's symbol view are shown in figures 6.46 and 6.47, respectively. The primary inductor, L_{p}, is magnetically coupled to the secondary inductor, L_{s}, with a coupling coefficient, $0 \leqslant k \leqslant 1$. At $k = 1$, magnetic flux transfers from the primary inductor to the secondary inductor without loss. For this design, the coupling coefficient is set to $k = 0.6$. The series resistance of all inductors is determined by their quality factors. For this design, $Q_{\mathrm{p}} = 12$, $Q_{\mathrm{s}} = 4$. The resistances of the inductors are calculated by the following equation

$$R_{\mathrm{p,s}} = \frac{\omega \cdot L_{\mathrm{p,s}}}{Q_{\mathrm{p,s}}} \tag{6.8}$$

The input impedance of the balun, with its primary and secondary inductances swept (where $L_p = L_s$), is measured in the schematic shown in figure 6.48 at 5GHz. The resulting input impedance is depicted in figure 6.49. It can be observed that there is a significant

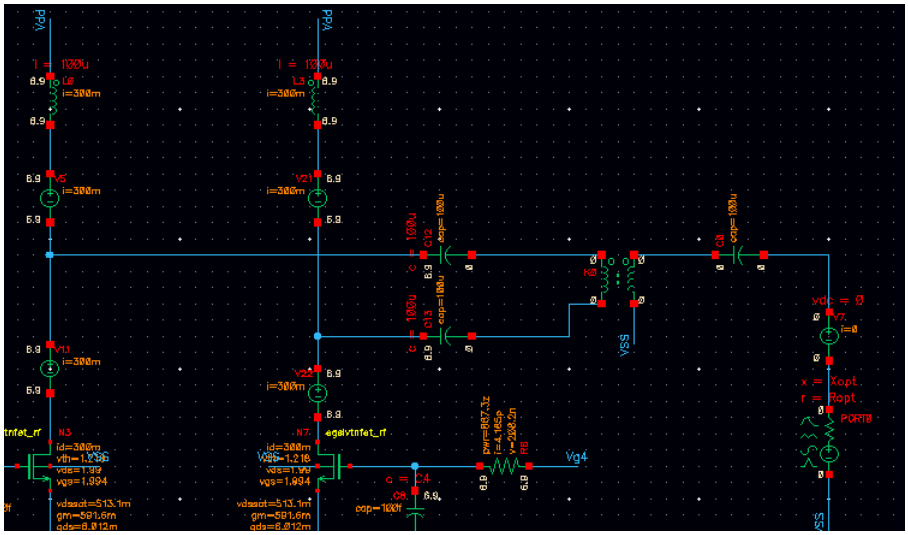

Figure 6.41: Schematic Design using xfmr (ideal transformer) at output

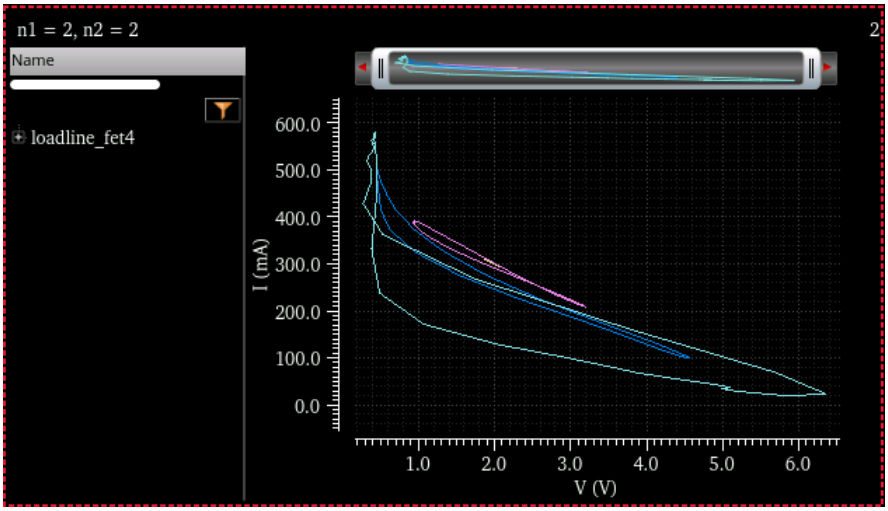

Figure 6.42: Loadline curve of 3rd Stacked FET

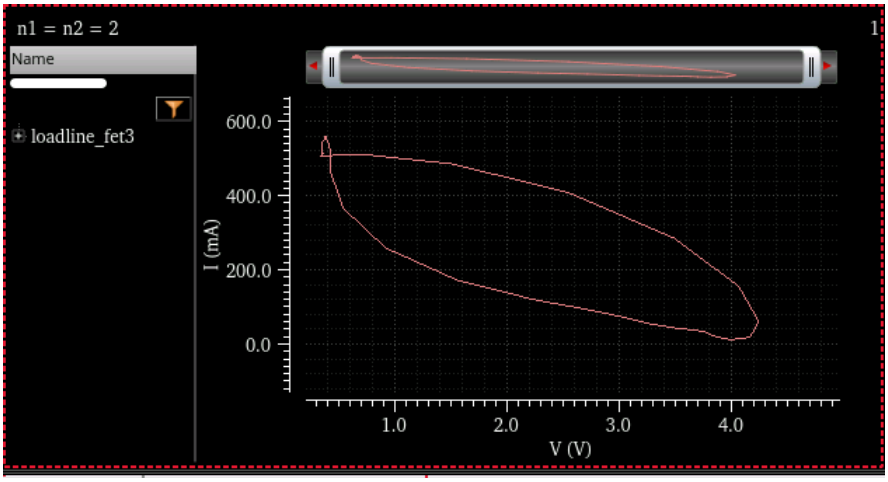

Figure 6.43: Loadline curve of 2nd Stacked FET

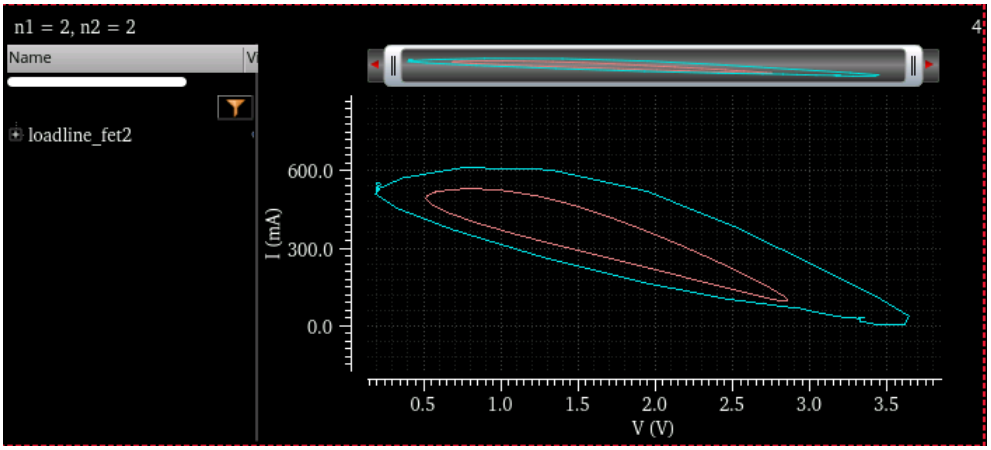

Figure 6.44: Loadline curve of 1st Stacked FET

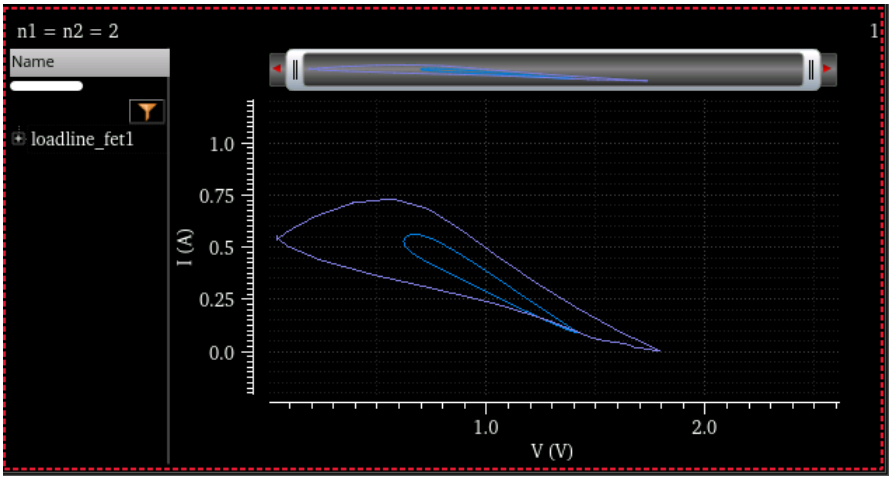

Figure 6.45: Loadline curve of CS FET

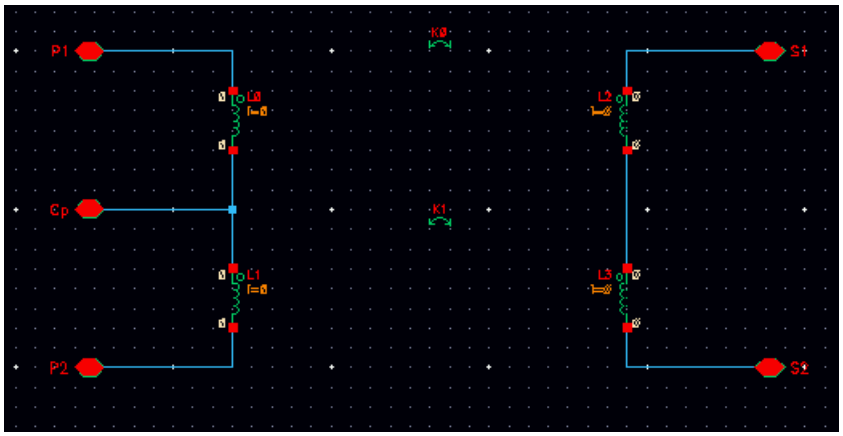

Figure 6.46: Circuit schematic of non-ideal balun model.

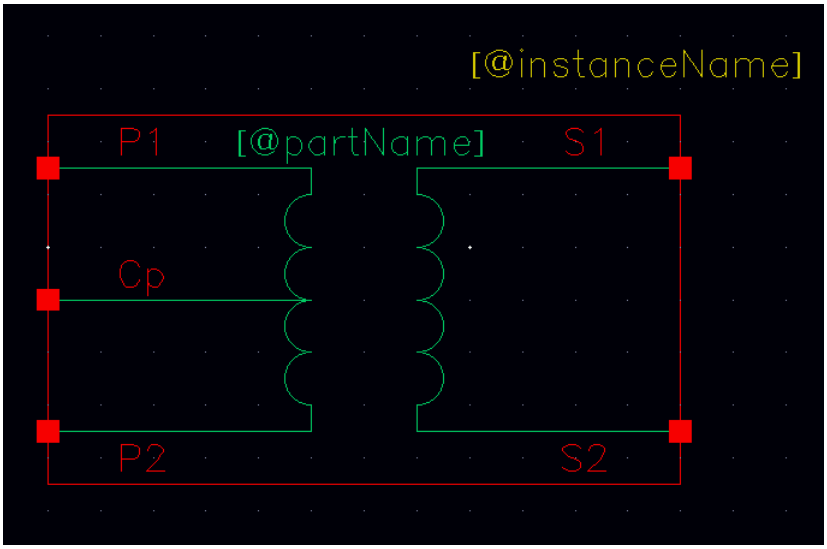

Figure 6.47: Symbol view of non-ideal balun model.

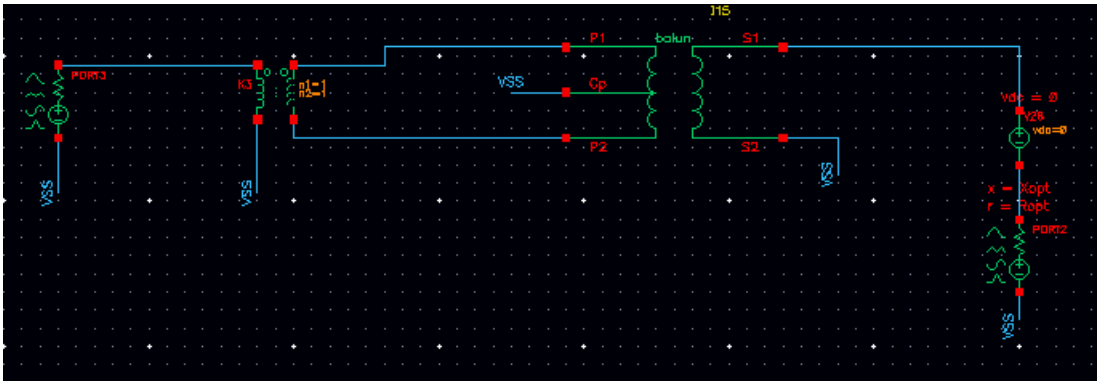

Figure 6.48: Schematic Design to calculate input impedance of balun

inductive component. To compensate for this, a capacitor is added in parallel with the balun's output. The primary and secondary inductances selected are $L_p = L_s = 0.5\,\text{nH}$, and the parallel capacitor used has a value of $C = 2\,\text{pF}$. After resonance, the input impedance of the output balun is shown in figure 6.50. It can be observed that the inductive component is canceled out, and the real part of the input impedance of the balun closely matches the real part of the optimum impedance, $\text{Re}\{Z_{\text{opt}}\}$, of the differential 3-stacked PA.

The bias current of the differential 3-Stacked PA is reduced to enable operation in Class AB. In Class AB, gain expansion is observed. With gain expansion, the power gain is reduced due to the lower bias current, but $OP_{1\text{dB}}$ is increased. To achieve Class AB operation and observe gain expansion, the CS FETs must be biased near their threshold voltage (V_{th}). A gain expansion of less than 0.5 dB is preferred, as it provides the best compromise between $OP_{1\text{dB}}$ and the linearity of the PA, which directly impacts the EVM (Error Vector Magnitude) [38].

The next step involves replacing the analogLib components (capacitors, inductors, resistances) with their corresponding elements from the 22FDX PDK. Additionally, the simplified model of the output balun is replaced by a stacked balun from 22FDX technology. The final

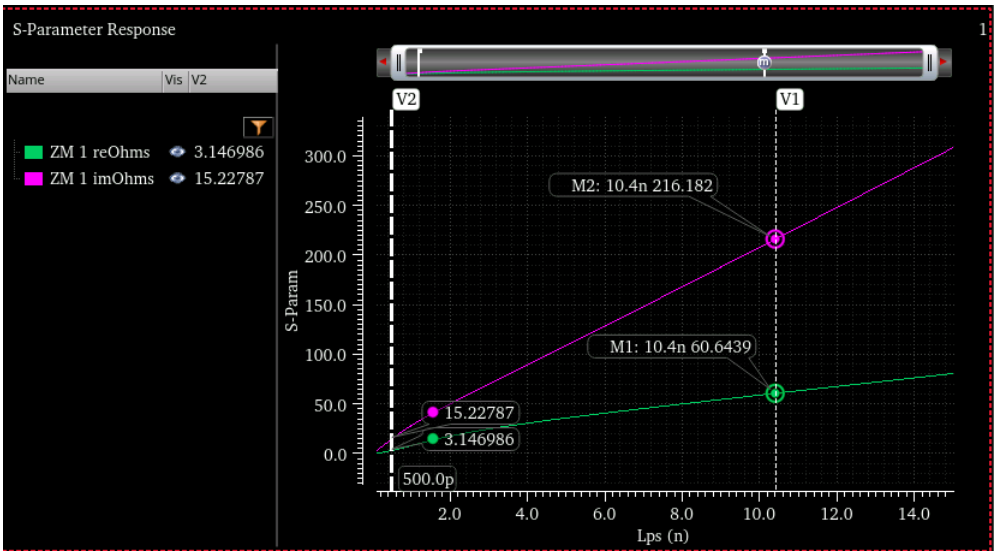

Figure 6.49: Input Impedance of output balun

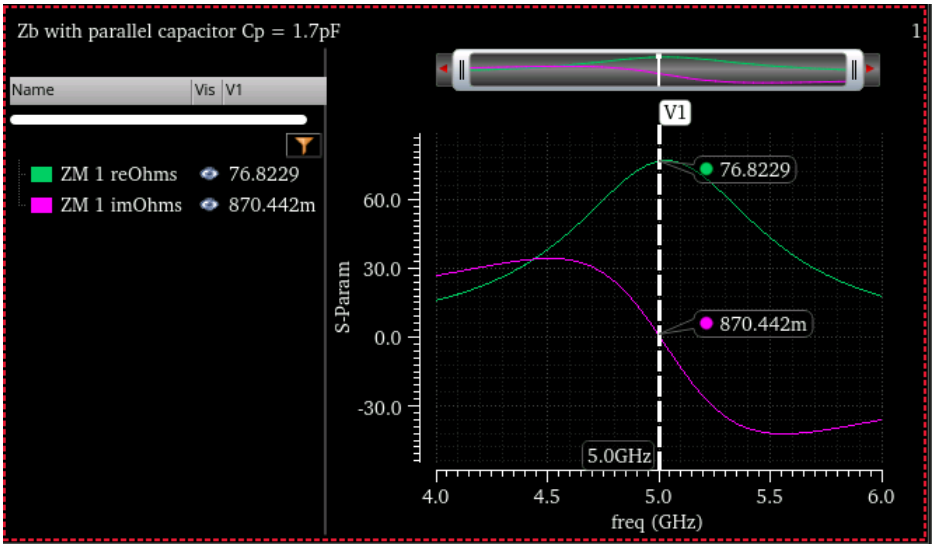

Figure 6.50: Input Impedance of output balun with shunt capacitor

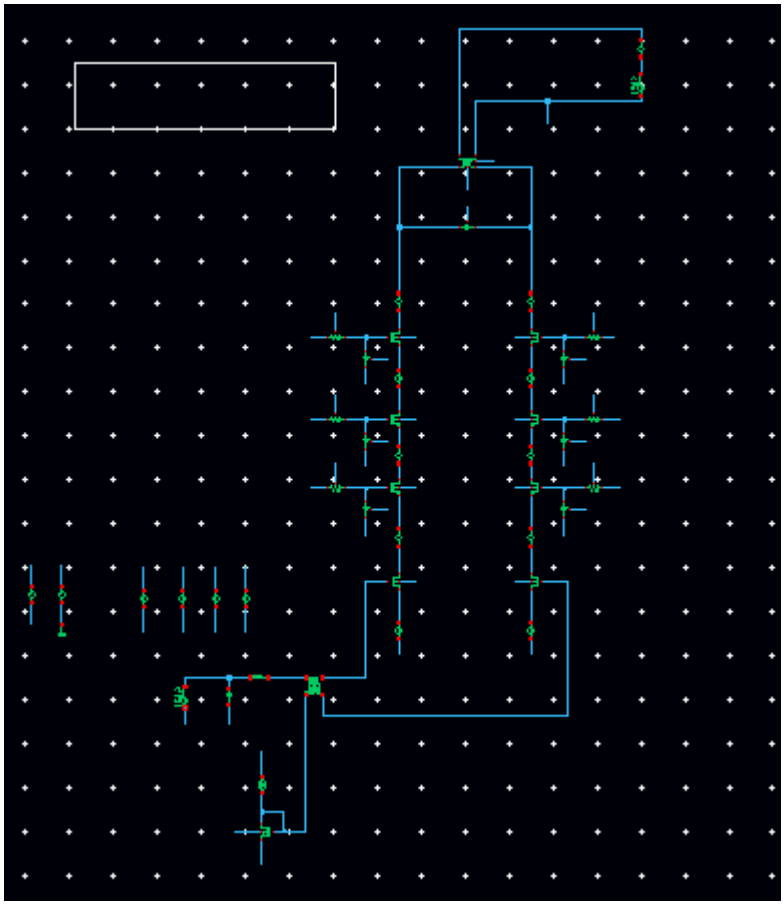

Figure 6.51: Final schematic design of differential 3-Stacked PA.

schematic design is shown in figure 6.51. At the center frequency of the C-band, 4.7 GHz, the differential 3-Stacked PA achieves $OP_{1\text{dB}} = 28.14\,\text{dBm}$, $P_{\text{gain}} = 26\,\text{dB}$, and PAE = 30%, as shown in figures 6.52–6.54.

6.2 GaN HEMT

The design procedure follows the same design methodology as in CMOS FDSOI. First, the GaN HEMT device, with the model name GH15NHF (GaN HEMT 150 μm Non-Linear Hot FET model), needs to be characterized. The current density characteristics are shown in figures 6.55–6.58. The CS FET is biased with 480 mA and has a gate width of $w_f = 100\,\mu\text{m}$, number of fingers $n_f = 12$, and multiplier mul = 8. The drain voltage is biased at 20 V, and the gate voltage is set to $-2.8\,\text{V}$, above the threshold voltage of the device, $V_{th} = -3.2\,\text{V}$.

As shown in figures 6.55 and 6.56, for $J > 100\,\text{mA/mm}$, G_{max} and f_{max} remain nearly constant. For $J < 50\,\text{mA/mm}$, G_{max} and f_{max} decrease significantly. From Figure 6.57, NF_{min} remains almost constant across the current density range. From Figure 6.58, the device is not unconditionally stable across the current density range. By using cascode FETs and an output balun, $|S_{12}|$ will be reduced (improving reverse isolation), and the stability factor K_{f}

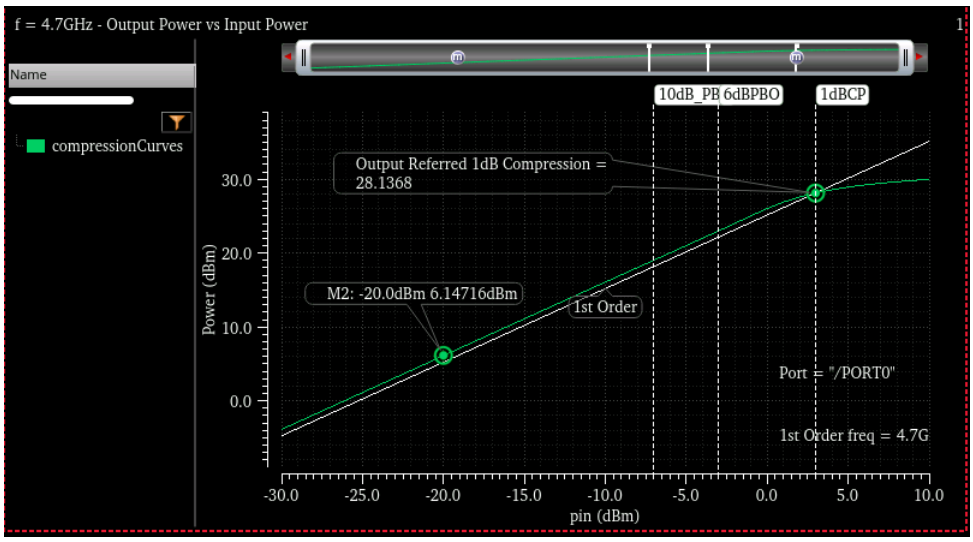

Figure 6.52: Compression curve of differential 3-Stacked PA, $OP_{1\mathrm{dB}} = 28.14\,\mathrm{dBm}$

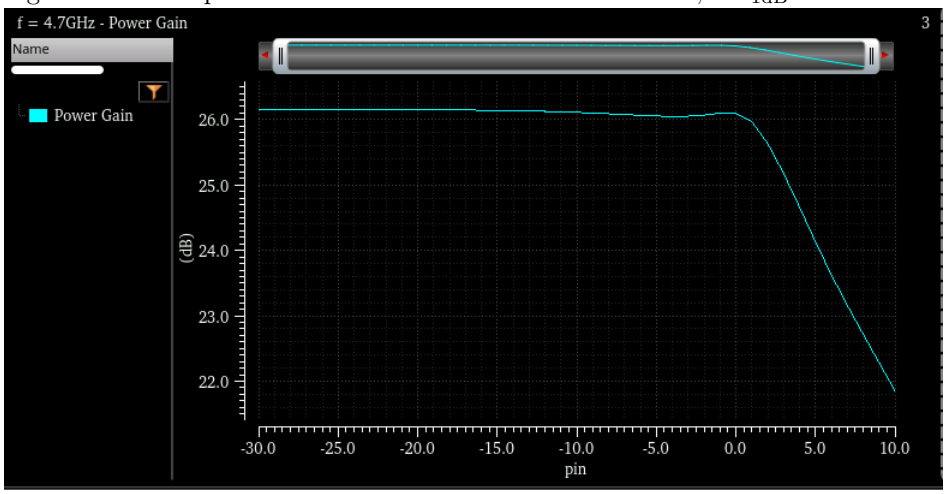

Figure 6.53: Power gain of differential 3-Stacked PA, $P_{\mathrm{gain}} = 26\,\mathrm{dB}$

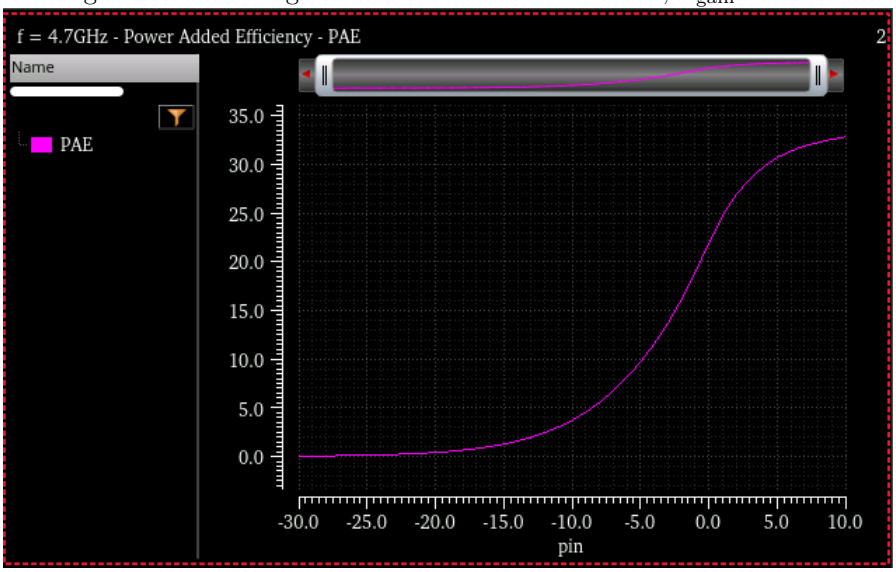

Figure 6.54: PAE of differential 3-Stacked PA, PAE = 30%

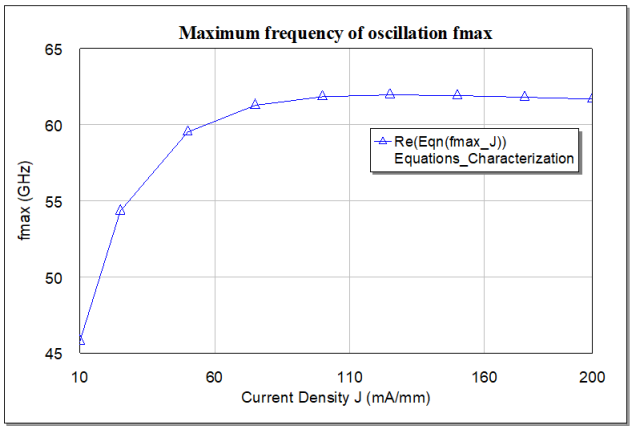

Figure 6.55: $f_{\max}$ (GHz) in function of current density J (mA/mm)

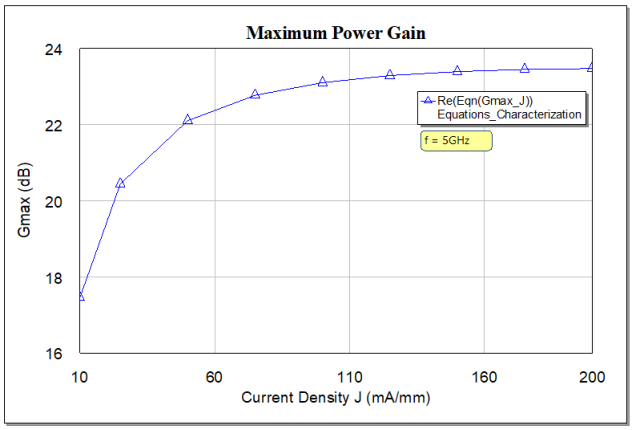

Figure 6.56: $G_{\max}$ (dB) in function of current density J (mA/mm)

will increase, given by:

$$K_{\mathrm{f}} = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2\,|S_{12} \cdot S_{21}|} \tag{6.9}$$

Based on these observations, the device will be studied at $J = 50\,\mathrm{mA/mm}$ and $J = 100\,\mathrm{mA/mm}$. The same table as table 6.1 will be calculated for the GaN HEMT technology. The results are presented in table 6.5. For the CS FET, a multiplier of 8 is chosen with a bias current of $I_{\mathrm{bias}} = 480\,\mathrm{mA}$. The key metrics are highlighted in table 6.5. The current density $J = 50\,\mathrm{mA/mm}$ is selected due to its lower power consumption. A multiplier of mul = 16 is not selected because it exhibits lower power gain compared to mul = 8, although $OP_{1\mathrm{dB}}$ is increased by 5.5 dB in the former case.

The design methodology for the GaN HEMT stacked PA is the following:

1. Select the multiplier (e.g. 2,4,8) for the stacked device, while keeping the gate width (e.g. 100 μm) and the number of fingers (e.g. 12) fixed.

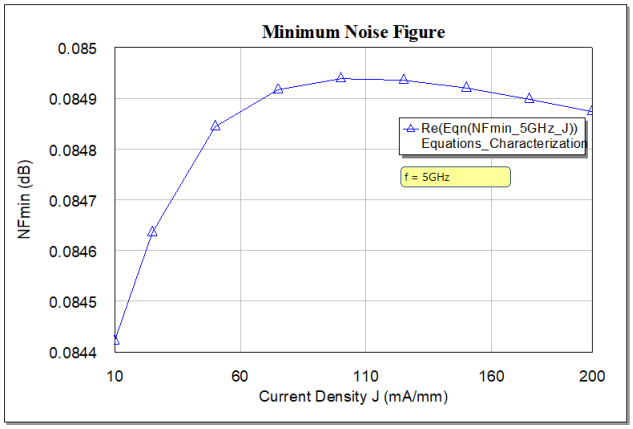

Figure 6.57: $NF_{\min}$ (dB) in function of current density J (mA/mm)

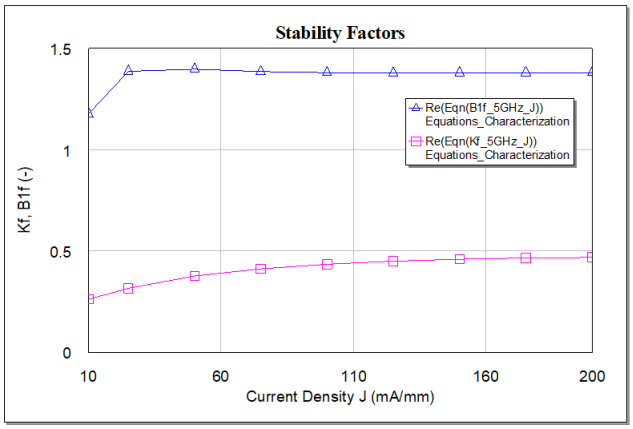

Figure 6.58: K_f, B_{1f} $(-)$ in function of current density J (mA/mm)

Multiplier	Width (mm)	J (mA/mm)	I_{bias} (mA)	Z_{opt} (Ω)	Z_{in} (Ω)	$OP_{1\text{dB}}$ (dBm)	Gain (dB)	PAE @1dBCP (%)
1	1.2	50	60	$20+32j$	$0.24-12j$	17.5	32	4.5
		100	120	$19+35j$	$0.29-11j$	25.2	32	13.8
2	2.4	50	120	$12+20j$	$0.32-5.7j$	23.1	27.8	8.3
		100	240	$10+15.7j$	$0.26-5.8j$	28.6	29.7	15
4	4.8	50	240	$5+7j$	$0.12-3.4j$	25.8	29.6	7.8
		100	480	$6.4+9.3j$	$0.25-2.89j$	32.4	26.7	17.9
8	9.6	50	480	$3.3+5.5j$	$0.11-1.63j$	32.8	26.8	19.2
		100	960	$4+3.5j$	$0.23-1.7j$	40.4	24.4	56.3
16	19.2	50	960	$3+1.3j$	$0.2-1j$	38.3	21.6	34.3
		100	1920	$3.4+2j$	$0.2-1j$	41.7	22.3	37.7

Table 6.5: Metrics for CS FET with $W_{\text{f}} = 100\,\text{um}$, $N_{\text{f}} = 12$, $V_D = 20\,V$.

Multiplier	C_2 (pF)	OP_{1dB} (dBm)	P_{Gain} (dB)	PAE (%)	Z_{in} (Ω)	Z_{opt} (Ω)
8	1.5	37.2	27.4	26.5	$0.18 - 1.72j$	$4.24 + 9.2j$
4	1.7	38.1	27.4	33.5	$0.24 - 1.86j$	$6.7 + 9.7j$
2	2.6	38.1	31.7	33.6	$0.18 - 1.80j$	$16.5 + 18.5j$

Table 6.6: 1-Stacked PA, $I_{bias} = 480$ mA, class A

2. Initially, choose a large capacitor (e.g. 10 pF) at the gate of the stacked device to serve as an AC ground.

3. Perform load pull analysis to determine the optimum load impedance for the stacked PA that maximizes the saturation output power, P_{sat}.

4. Sweep the shunt capacitor at the gate of the stacked device. Adjust its value to match the input impedance of the stacked FET to the optimum load of the stacked PA. Ensure the drain-gate voltage, V_{dg}, does not exceed the recommended operating ratings (ROR) of the device ($V_{dg} < 65V$).

5. Once the multiplier, load impedance and shunt capacitor are determined, calculate OP_{1dB}, PAE and P_{Gain}.

6. Choose the multiplier, load impedance and shunt capacitor at the gate with the best metrics. The main purpose is to maximize OP_{1dB} and P_{Gain} since the primary goal of a PA is to produce output power for low input power levels in a transmitter chain.

The metrics of the 1-Stacked PA are shown in table 6.6, along with the chosen capacitor at the gate of the first stacked FET. The bias current, I_{bias}, will be reduced to allow the Stacked FET to operate in class AB. In class AB, the PA exhibits Gain Expansion (G.E), which increases OP_{1dB} and PAE. Generally, G.E should be limited to under 0.5 dB to achieve an optimal trade-off between OP_{1dB} enhancement, PAE improvement, and maintaining low nonlinearity. Based on the results in table 6.7, a multiplier value of 8 is selected for the first stacked device. The multiplier of 8 offers the highest OP_{1dB}, P_{Gain}, and the lowest output impedance, minimizing the mismatch with the input impedance of the 2nd Stacked FET. The supply voltage of the 1-Stacked PA is $V_{DD} = 2 \cdot V_{max,GH15NHF} = 40V$. The bias current is 270 mA and the gate voltage of the CS FET is biased close to the threshold voltage at -2.96V.

The compression curve and power gain of the 1-Stacked PA are shown in figures 6.59 and 6.60, respectively. The loadline curves of the CS FET and the first Stacked FET are presented in figures 6.61 and 6.62, respectively, demonstrating full voltage and current swings. The capacitor at the gate of the first Stacked FET, in addition to enabling power matching between the CS FET and the first Stacked FET, ensures the reliability of the devices. At the 1dB compression point, the drain-gate voltages of the devices remain well below 65 V, which is the maximum recommended drain-gate voltage difference. The drain-gate voltages at the 1dB compression point are shown in figure 6.63.

Multiplier	I_{bias} (mA)	C_2 (pF)	$OP_{1\text{dB}}$ (dBm)	P_{Gain} (dB)	G.E (dB)	Z_{in} (Ω)	Z_{opt} (Ω)
8	270	1.5	45.3	25.4	0.6	$0.18-1.7j$	$4.24+9.2j$
4	240	1.7	44	21.7	0.5	$0.24-1.9j$	$6.7+9.7j$
2	220	2.6	42.1	23.7	0.5	$0.18-1.8j$	$16.5+18.5j$

Table 6.7: 1-Stacked PA, class AB

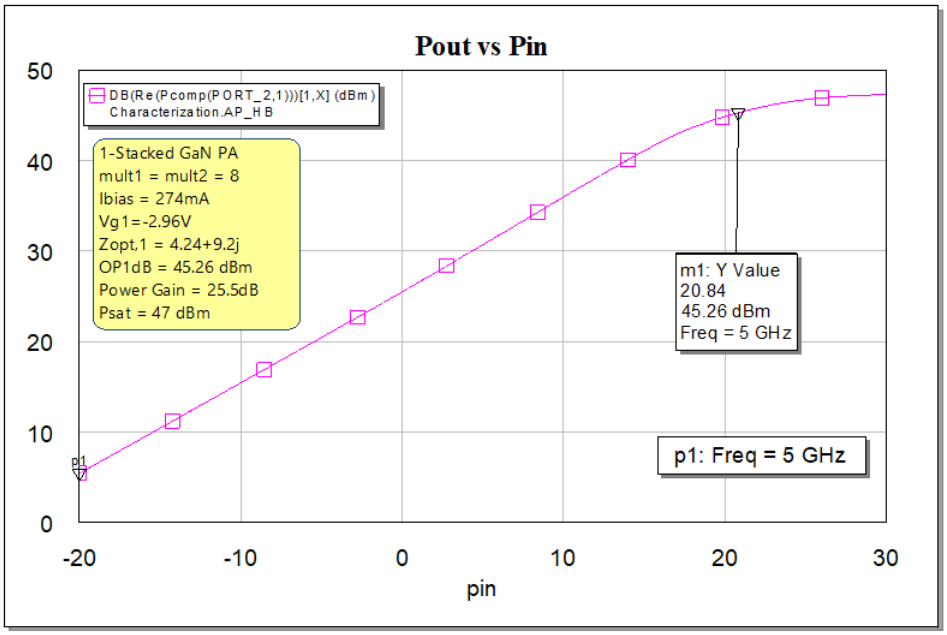

Figure 6.59: Compression curve of 1-Stacked PA, $OP_{1\text{dB}} = 45.3\,\text{dBm}$

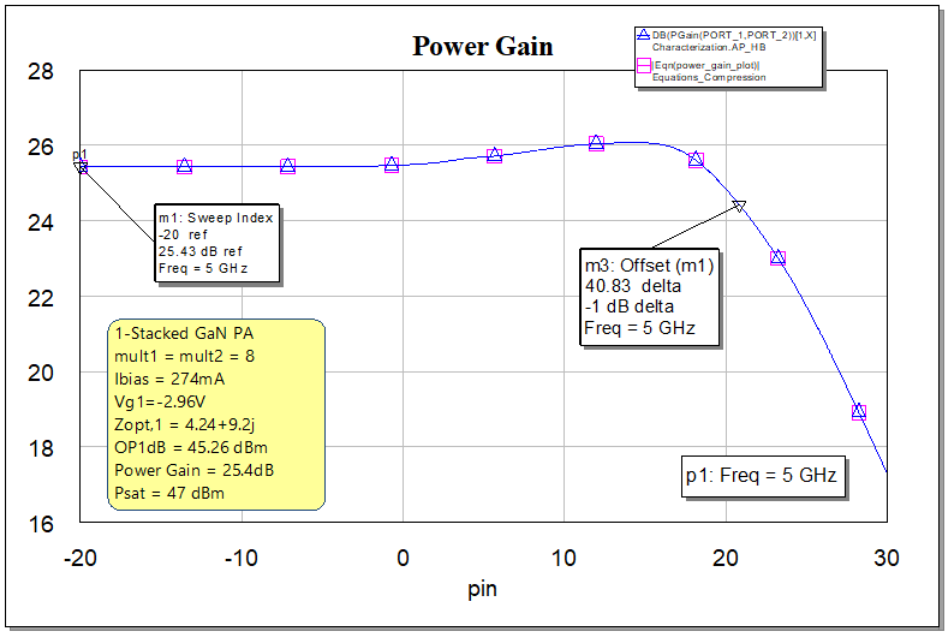

Figure 6.60: Power gain of 1-Stacked PA, $P_{\text{gain}} = 25.4\,\text{dB}$

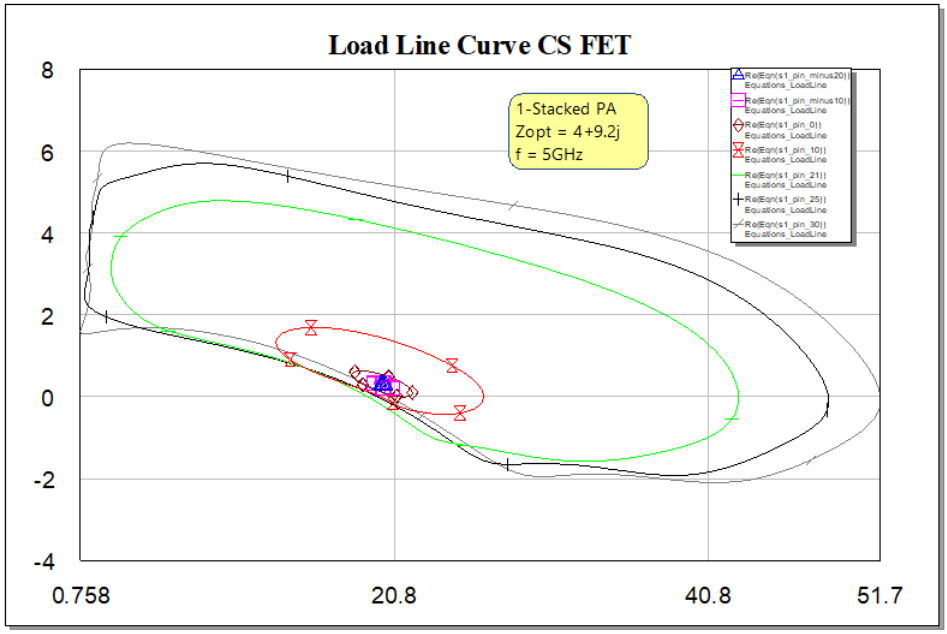

Figure 6.61: Loadline curve of CS FET in 1-Stacked PA.

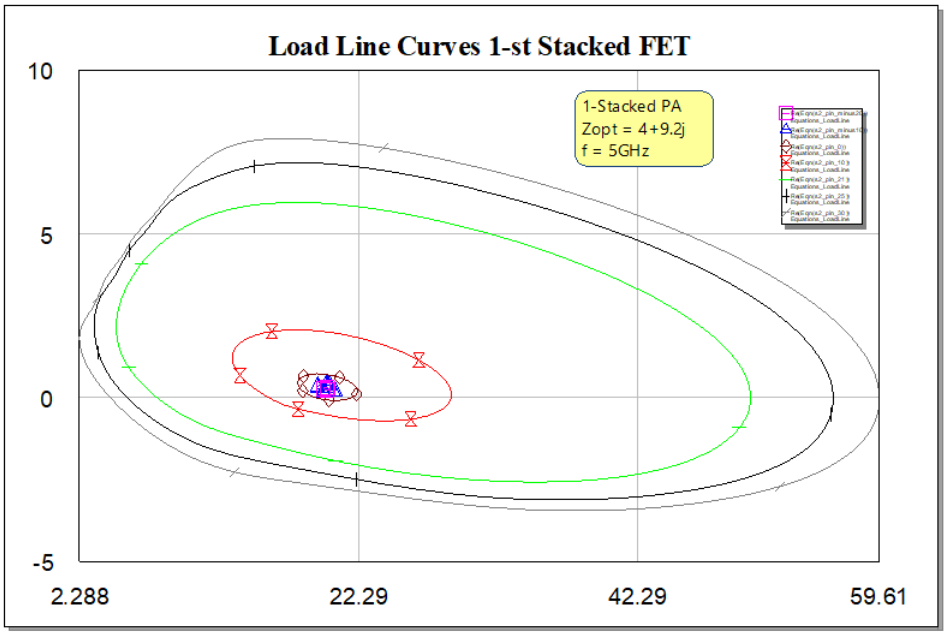

Figure 6.62: Loadline curve of first stacked FET in 1-Stacked PA.

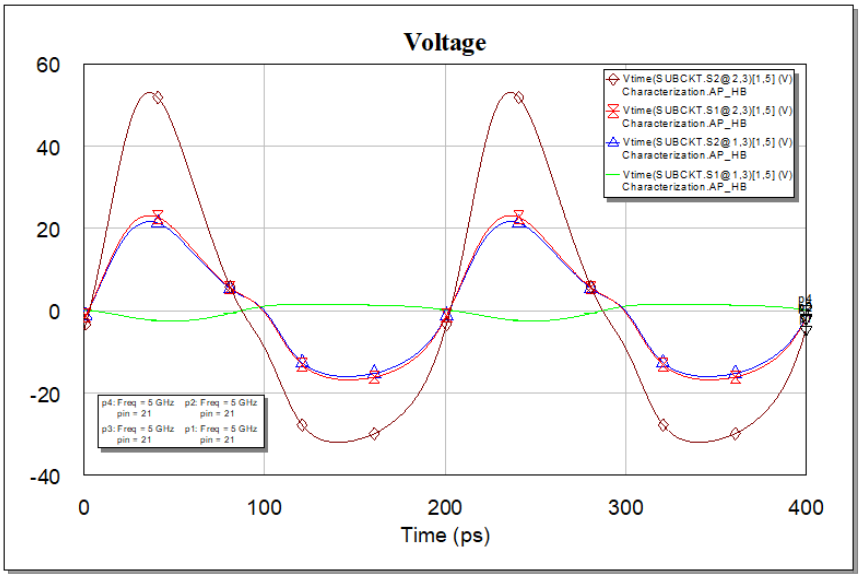

Figure 6.63: Drain-Gate Voltages of CS FET (S1) and first Stacked FET (S2) at 1dB compression point.

Multiplier	I_{bias} (mA)	OP_{1dB} (dBm)	P_{Gain} (dB)	G.E (dB)	PAE (%)	Z_{in} (Ω)	Z_{opt} (Ω)
8	230	47.85	25.9	0.5	45.3	$0.45-2.4j$	$8+7j$
4	186	46.6	25.7	0.5	40.2	$0.39-2.3j$	$13+6j$
2	159	41.4	27.7	0.6	30	$0.35-2.4j$	$20+14j$

Table 6.8: 2-Stacked PA, class AB

The design methodology for the second stacked FET follows the same process as that of the first stacked FET. The design choice for the second stacked FET is presented in table 6.8. A multiplier of 8 is selected for the second stacked FET with a bias current of $I_{\text{bias}} = 230\,\text{mA}$, as it achieves the highest OP_{1dB}. The bias current of 230 mA is chosen to enable Class AB operation, facilitating gain expansion in the 2-Stacked PA. Compression, power gain, PAE, and loadline curves are generated and shown in figures 6.64-6.69, respectively.

The drain-gate voltages of all FETs at the 1 dB compression point are illustrated in figure 6.70, confirming their reliability as $V_{dg} < 65\,\text{V}$. The reliability of the devices is also ensured at the gate-source junction, as shown in figure 6.71, where $V_{gs} > -15\,\text{V}$. The drain and gate voltages of all FETs are depicted in figure 6.72.

For each stacked device added, the drain voltage swing is doubled, contributing to the increase in output power of the stacked PA. The drain and gate voltages are in phase, and the gate voltage swing of the k-th stacked FET is slightly less than the drain voltage swing of the $(k-1)$-th stacked FET. This occurs due to the voltage divider effect caused by the gate-source capacitance of each FET. This behavior ensures the reliability of the devices at the drain-gate junction.

The single-ended 2-Stacked PA will be converted to a differential configuration. Ideal baluns are incorporated at the input and output. At the input, two series-connected L-matching networks are utilized to achieve a broader bandwidth [14]. The schematic design is illustrated in figure 6.73.

The multipliers selected from the single-ended PA are (8, 8, 8). To maintain the same multiplier ratio across all FETs and preserve interstage matching, the multipliers will be proportionally reduced to save chip area and enhance the output impedance $Z_{\text{opt,2}}$. This adjustment minimizes magnetic losses in the output balun. Load-pull analysis will be conducted for the multiplier configurations (8, 8, 8), (4, 4, 4), and (2, 2, 2). For each configuration, the optimum load impedance determined by the load-pull analysis will be used to generate compression, PAE, and power gain curves.

The results are summarized in table 6.9. Based on table 6.9, the multipliers (4, 4, 4) are selected as they reduce I_{bias} and improve PAE, with only a 1dB reduction in OP_{1dB}.

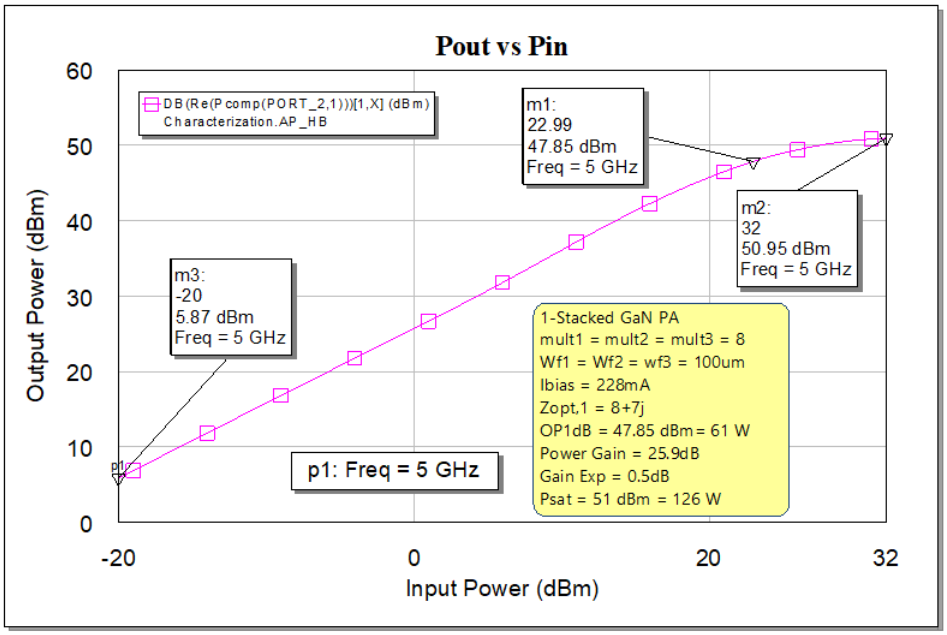

Figure 6.64: Compression curve of 2-Stacked PA.

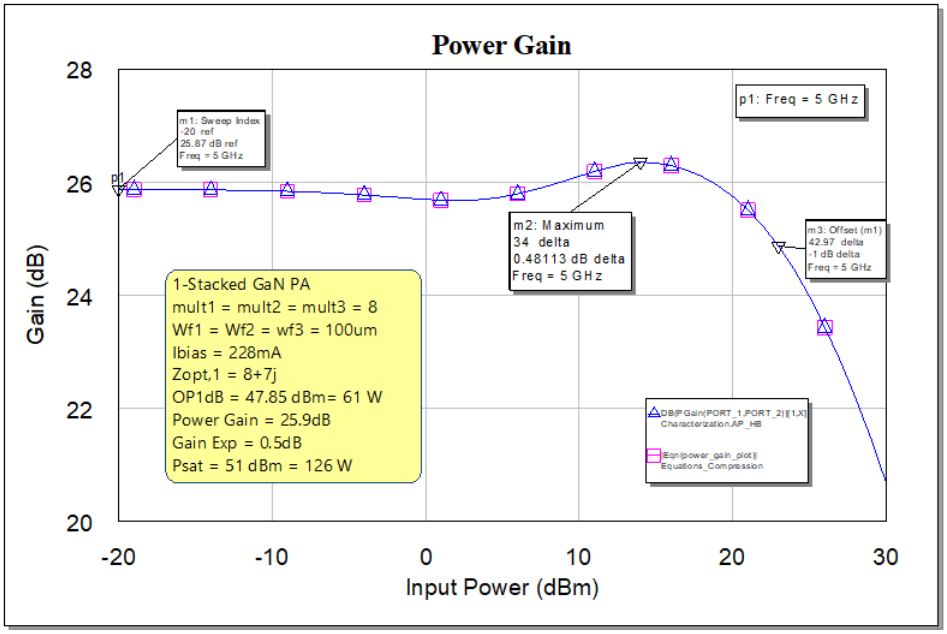

Figure 6.65: Power Gain of 2-Stacked PA.

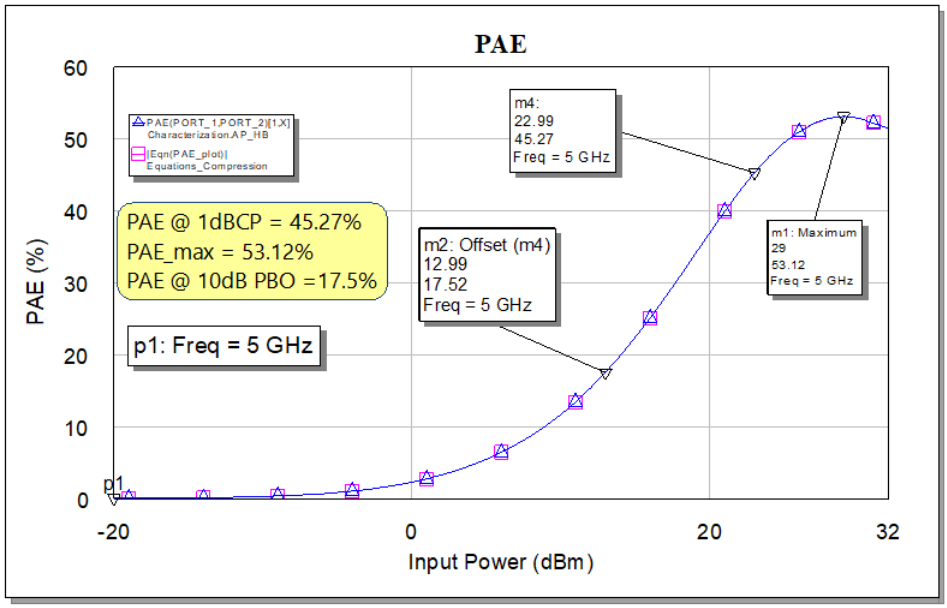

Figure 6.66: PAE of 2-Stacked PA.

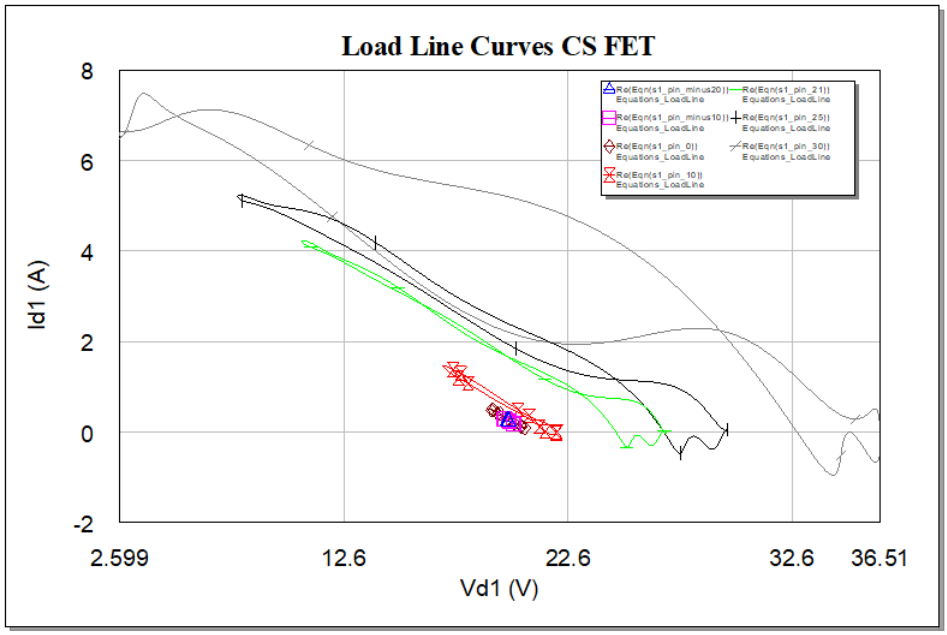

Figure 6.67: Loadline of CS FET of 2-Stacked PA.

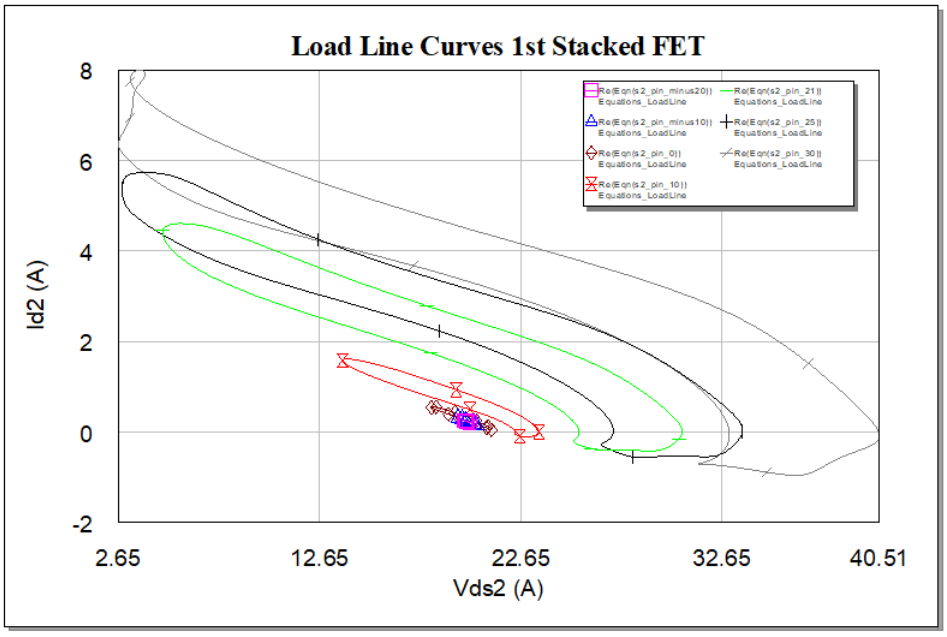

Figure 6.68: Loadline of first stacked FET of 2-Stacked PA.

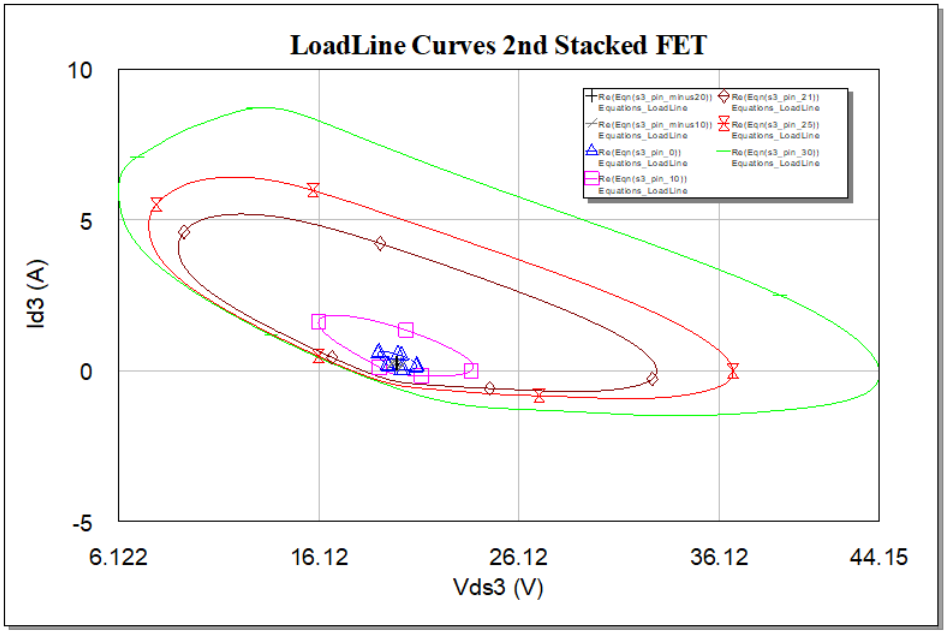

Figure 6.69: Loadline of second stacked FET of 2-Stacked PA.

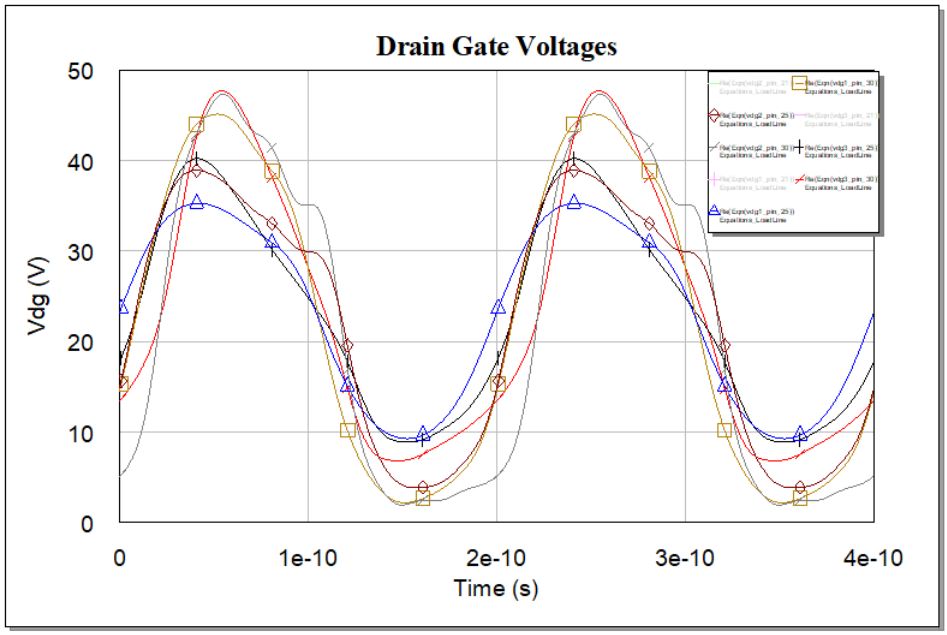

Figure 6.70: Drain-gate voltages of FETs of 2-Stacked PA.

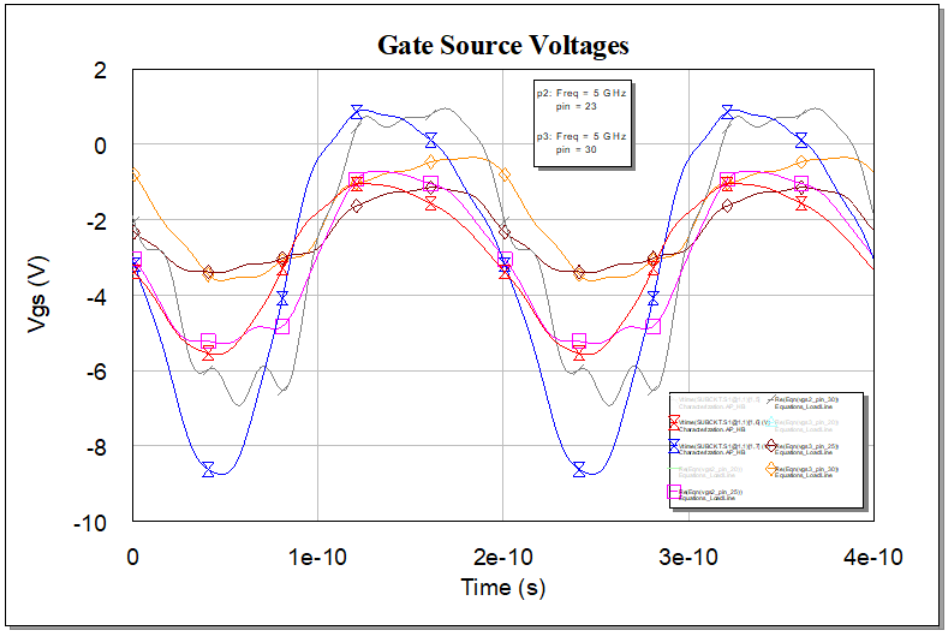

Figure 6.71: Gate-source voltages of FETs of 2-Stacked PA.

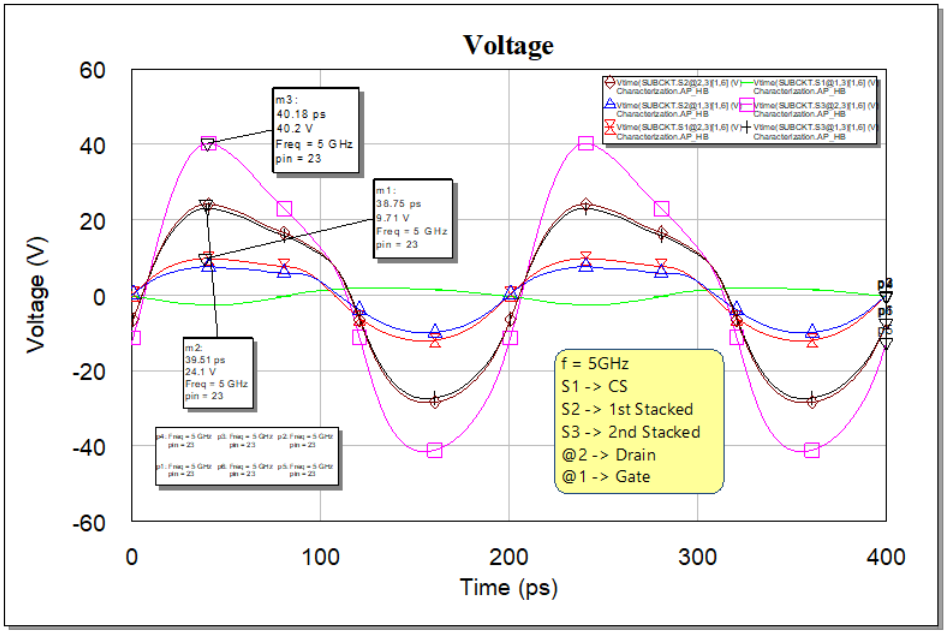

Figure 6.72: Drain and Gate voltages of FETs of 2-Stacked PA.

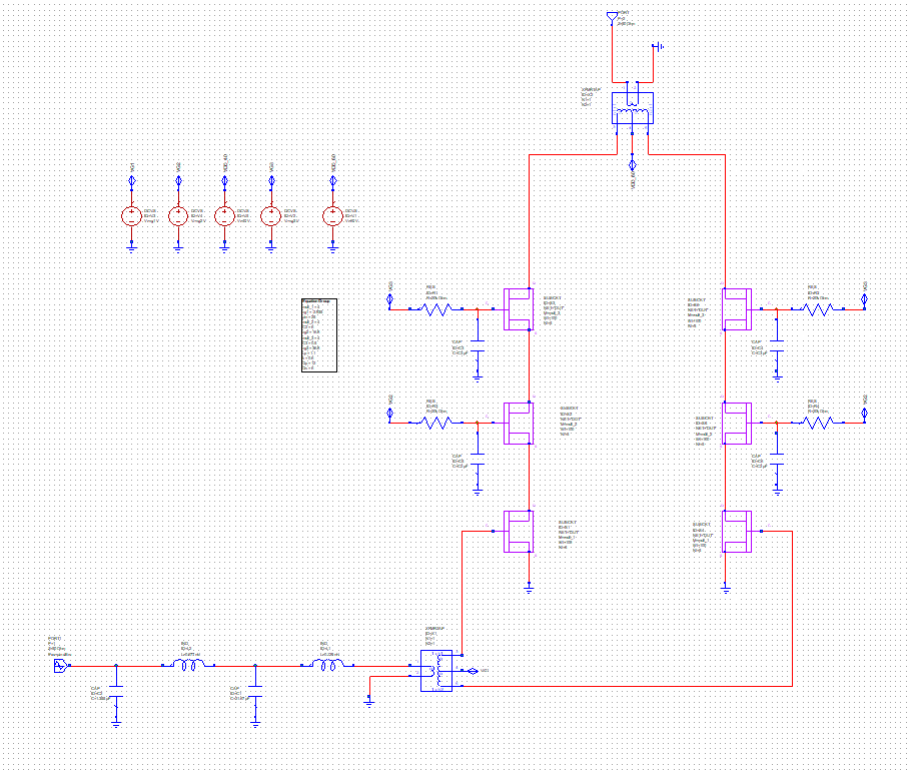

Figure 6.73: Schematic Design of differential 2-Stacked PA

Multipliers	I_{bias} (mA)	$OP_{1\text{dB}}$ (dBm)	P_{sat} (dBm)	P_{Gain} (dB)	G.E (dB)	PAE (%)	Z_{opt} (Ω)
(8,8,8)	408	47.8	48.9	22.0	0.50	27.8	$10+8j$
(4,4,4)	136	46.7	48.2	24.1	0.35	37.0	$16+4j$
(2,2,2)	117	45.2	46.5	25.7	0.50	35.4	$19+5j$

Table 6.9: Metrics of 2-Stacked Differential PA

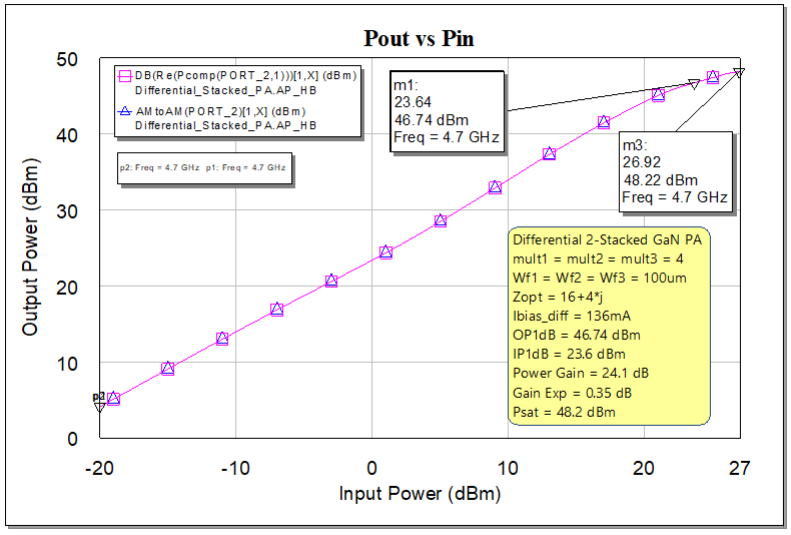

Figure 6.74: Compression curve of differential 2-Stacked PA for the highlighted operating point in table 6.9.

Additionally, the increase in $\text{Re}\{Z_{\text{opt}}\}$ enables the use of a 1:1.5 turns ratio instead of 1:2, which further reduces magnetic losses in the output balun. The multipliers (2, 2, 2) are not selected because PAE remains unchanged while $OP_{1\text{dB}}$ decreases compared to the multipliers (4, 4, 4).

The performance of the differential 2-Stacked PA with multipliers (4, 4, 4) is shown in the following figures. The compression curve, power gain, and PAE are illustrated in figures 6.74-6.76. The loadline curves of the FETs are presented in figure 6.77. The drain-gate voltages of the FETs at output power saturation are depicted in figure 6.78, validating their reliability. The same reliability holds for the gate-source voltages, as shown in figure 6.79. Finally, the drain and gate voltages of the FETs at output power saturation are shown in figure 6.80.

The output ideal balun is replaced with the simplified balun model described in the 22FDX Stacked FET design. This model is used to simulate the magnetic losses of the balun. The parameters of the simplified balun model are the following:

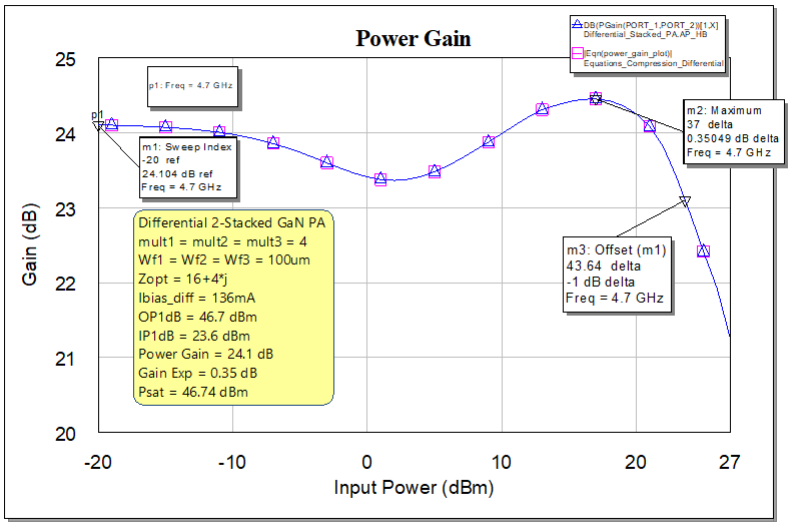

Figure 6.75: Power Gain of differential 2-Stacked PA for the highlighted operating point in table 6.9.

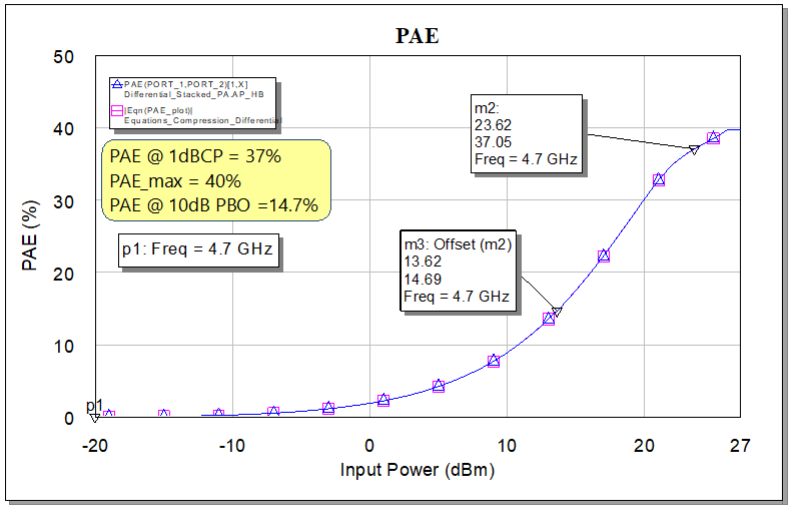

Figure 6.76: PAE of differential 2-Stacked PA for the highlighted operating point in table 6.9.

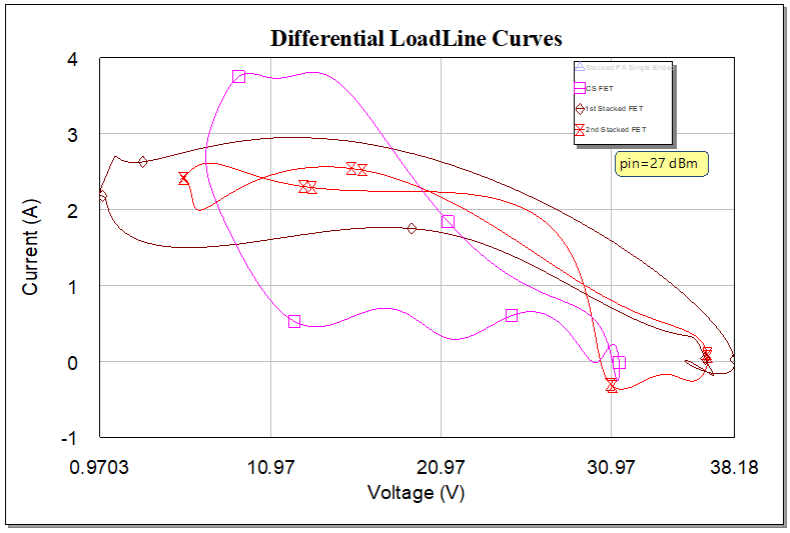

Figure 6.77: Loadline curves of differential 2-Stacked PA.

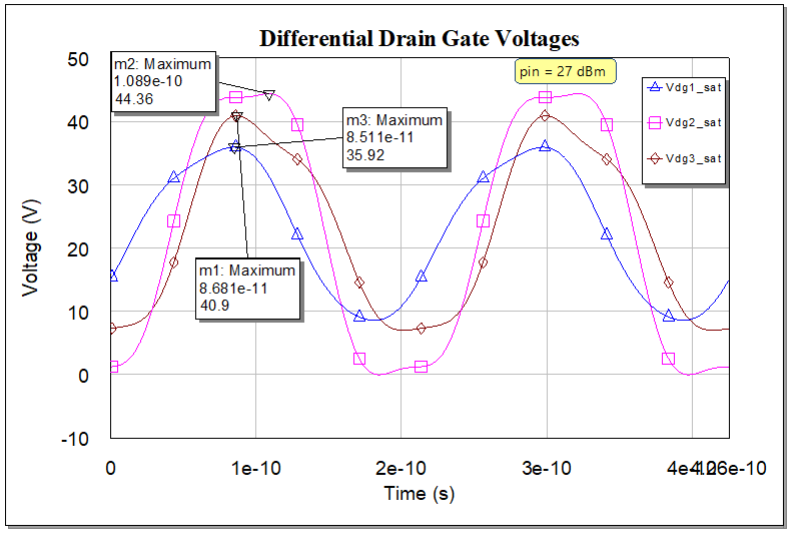

Figure 6.78: Drain-gate voltages of differential 2-Stacked PA in saturation of output power.

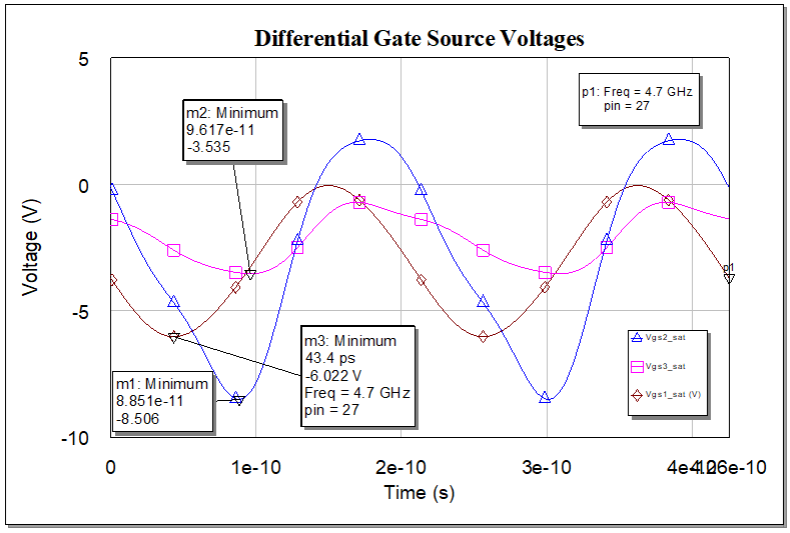

Figure 6.79: Gate-source voltages of differential 2-Stacked PA in saturation of output power.

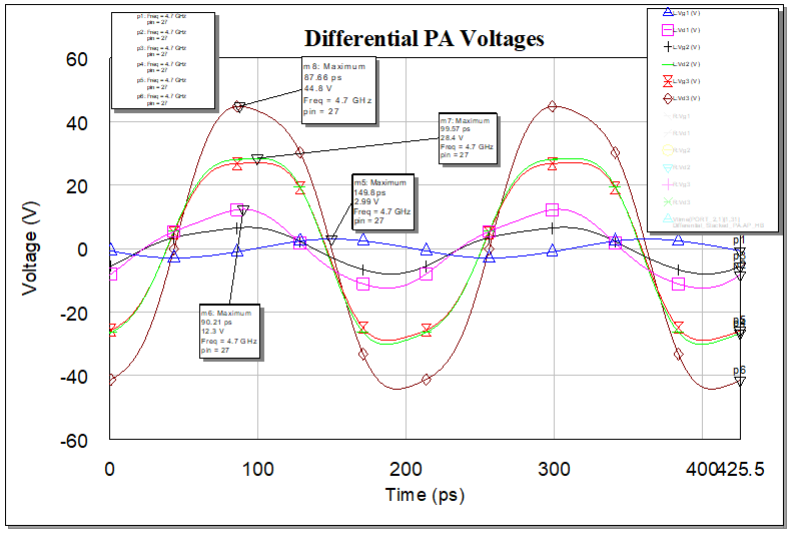

Figure 6.80: Drain and gate voltages of differential 2-Stacked PA in saturation of output power.

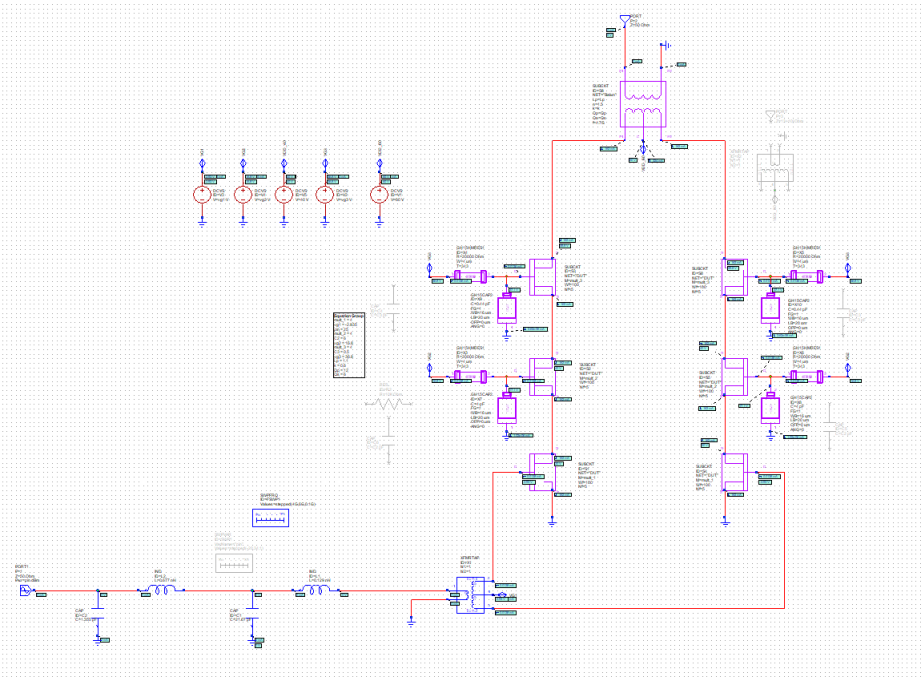

Figure 6.81: Schematic design of differential 2-Stacked PA.

- Coupling Coefficient: $k = 0.6$
- Turn Ratio: $n = 1.5$
- Quality Factors : $Q_\mathrm{p} = 12$, $Q_\mathrm{s} = 6$
- Inductances of Windings: $L_\mathrm{p} = 1.1\,\mathrm{nH}$, $L_\mathrm{s} = n^2 \cdot L_\mathrm{p} = 2.5\,\mathrm{nH}$

The ideal capacitors and resistors are replaced with the corresponding elements provided by the UMS-GH15 PDK:

- Capacitors: Low Density ($175\,\mathrm{pF/mm^2}$) MIM (Metal-Insulator-Metal)
- Resistors: TiWSi ($1000\,\Omega/\square^2$)

The bias current is swept to determine the optimum quiescent operating point for Class AB operation, ensuring reasonable gain expansion. The final schematic design is shown in figure 6.81. The response of the differential 2-Stacked PA is presented in the following figures at the center frequency of 4.7 GHz. The compression curve, depicted in figure 6.82, indicates an OP_1dB of 44.5 dBm. The power gain, shown in figure 6.83, is $P_\mathrm{gain} = 22.7\,\mathrm{dB}$. The PAE, illustrated in figure 6.84, achieves a maximum value near the 1dB compression point, reaching 18.3%.

The S-parameters of the differential 2-Stacked PA are shown in figures 6.85-6.87. From S_{11}, it is evident that the PA is perfectly matched at the input at a frequency of 4.7 GHz. The S_{12} is approximately $-50\,\mathrm{dB}$, validating the excellent isolation of the stacked FET design. The S_{21} is maximized at 4.7 GHz, achieving a gain of 22.7 dB, and the $-3\,\mathrm{dB}$ bandwidth of the design spans 4.4-5 GHz (n79 frequency band, C band). The S_{22} is below $-0.5\,\mathrm{dB}$, indicating a reasonable output match for a PA design.

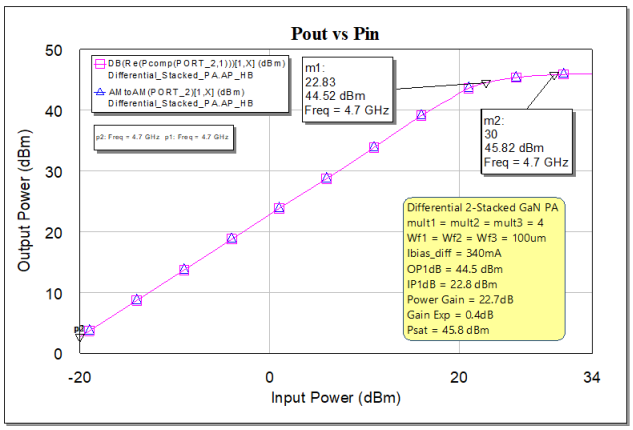

Figure 6.82: Compression curve of differential 2-Stacked PA.

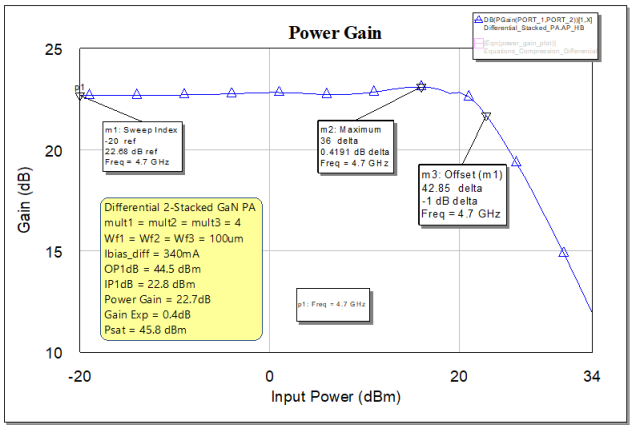

Figure 6.83: Power gain of differential 2-Stacked PA.

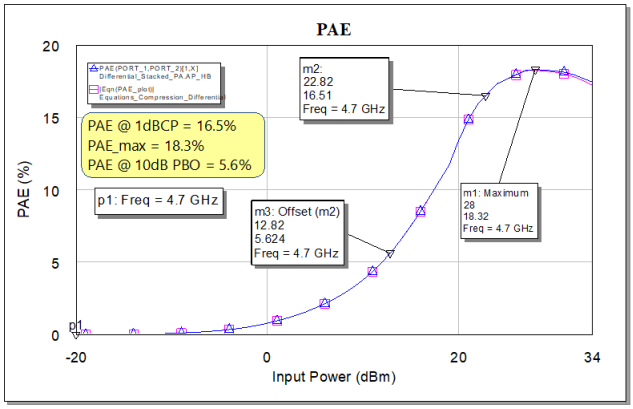

Figure 6.84: PAE of differential 2-Stacked PA.

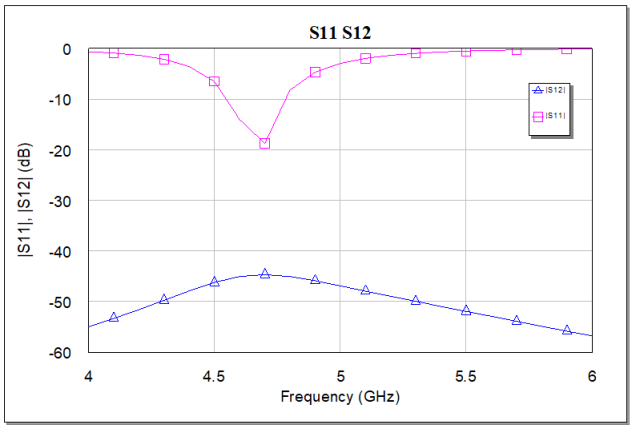

Figure 6.85: S_{11}, S_{12} (dB) of differential 2-Stacked PA.

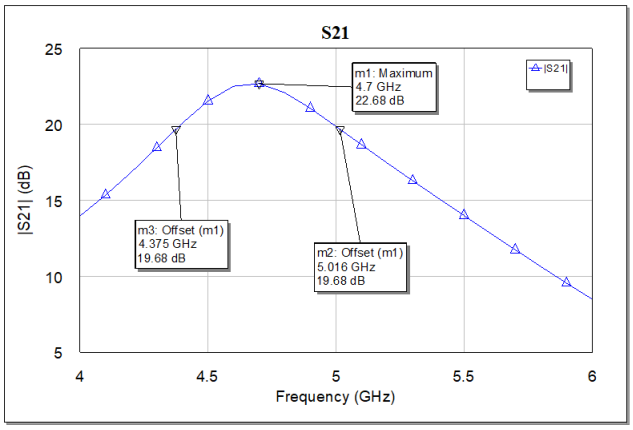

Figure 6.86: S_{21} (dB) of differential 2-Stacked PA.

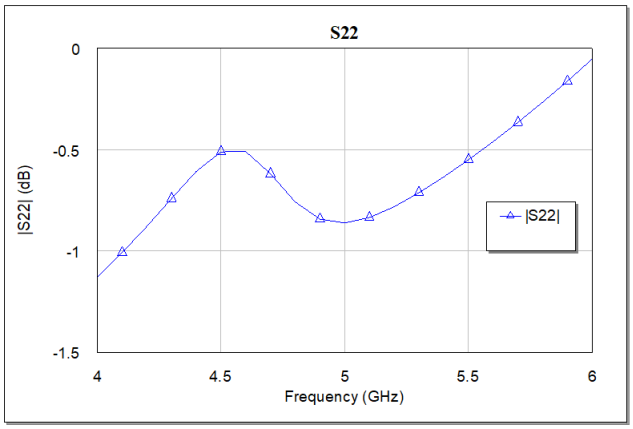

Figure 6.87: S_{22} (dB) of differential 2-Stacked PA.

6.3 Comparison with other works

Ref	**Proc**	**Freq** (GHz)	$\mathbf{OP}_{\mathrm{1dB}}\,(dBm)$	$\mathbf{P}_{\mathrm{sat}}\,(dBm)$	$\mathbf{P}_{\mathrm{gain}}\,(dB)$	**PAE** (%)
This work	22nm FDSOI	4.4-5	28.1	30	26	30
[39]	180nm CMOS	4.2-5.2	25.1	27.8	11.6	32
[2]	28nm FDSOI	4.5-5.5	NA	23	16.5	NA
This work	150nm GaN	4.4-5	44.5	46	22.7	16
[40]	250nm GaN	5-5.8	NA	46	21	46
[41]	250nm GaN	4.4-5.1	NA	42	17	62

Table 6.10: Comparison of Power Amplifiers Designed in Different Processes

In this section, the high-power PAs designed in this thesis are compared with state-of-the-art high-power PAs. From table 6.10, it is evident that the 22nm FDSOI Stacked PA design presented in this thesis achieves higher output power levels and significantly higher power gain while maintaining approximately the same power-added efficiency (PAE), as compared with the other CMOS PAs.

On the other hand, the 150nm GaN Stacked PA presented in this thesis delivers output power levels that are slightly higher or comparable to those of other GaN designs listed in table 6.10, with similar power gain but lower PAE. This suggests that the voltage supply or bias current should be reduced at the cost of some output power.

As highlighted in the table, GaN-based PAs achieve much higher output power (up to 40 times larger) compared to their CMOS counterparts, while also exhibiting higher power gains. This is attributed to the significantly higher breakdown voltages of GaN HEMT devices, making them ideal for PA design.

In conclusion, the PA designs presented in this thesis demonstrate comparable performance to state-of-the-art designs. However, it is important to note that the current results are based on schematic-level design. Layout implementation, which introduces parasitic effects, will slightly degrade performance, as would be demonstrated in post-layout simulations.

6.4 Conclusion and Future Work

This thesis demonstrates two high-power Stacked PA designs for 5G NR base station applications. The designs utilize GaN and CMOS FDSOI technologies, operating in the n79 frequency band (4.4–5 GHz). The GaN PA design achieves high output power levels of 46 dBm, which translates to 40 W. In contrast, the CMOS PA design also achieves high output power levels, comparable to other CMOS PA designs, at 30 dBm, corresponding to 1 W. The substantial difference in output power levels between GaN and CMOS FDSOI PAs arises from the significantly higher breakdown voltage of GaN technology. The designs have been thoroughly analyzed, with the corresponding design choices explained in detail. The Stacked PA architecture has been described step by step for both technologies.

For future work, the layout implementation of both designs should be completed, and post-layout verification should be conducted. To further enhance the performance, particularly the power gain of the GaN PA, the CMOS stacked PA will be used as its driver stage. Due to the cascaded configuration of the above two Stacked PAs, the final power gain is expected to reach levels of 50 dB within the n79 frequency band.

The CMOS FDSOI stacked PA can effectively drive the GaN PA whose input-referred 1 dB compression point is $IP_{1\mathrm{dB}} = 23\,\mathrm{dBm}$, which is significantly lower than the output-referred 1 dB compression point of the CMOS stacked PA, $OP_{1\mathrm{dB}} = 28.2\,\mathrm{dBm}$. The driver stage (CMOS stacked PA) and the PA (GaN stacked PA), connected in a cascaded configuration, will be taped out and their performance will be evaluated through electrical measurements in the laboratory.

Bibliography

[1] K. Harrouche and F. Medjdoub, "Gan-based hemts for mm-wave applications," *Nitride Semiconductor Technology: Power Electronics and Optoelectronic Devices*, 2020.

[2] B. Martineau, E. Mercier, and P. Vincent, "Opportunity of cmos fd-soi for rf power amplifier," in *2017 IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S)*, pp. 1–2, IEEE, 2017.

[3] F. Hameau, J. Zaini, T. Taris, D. Morche, B. Martineau, and P. Audebert, "New design opportunities exploiting fdsoi technology for rf power amplifier and lna design," in *2019 17th IEEE International New Circuits and Systems Conference (NEWCAS)*, pp. 1–4, IEEE, 2019.

[4] J. C. Mayeda, J. Lopez, and D. Y. Lie, "Highly-efficient broadband medium power amplifier design in 22nm cmos fd-soi for mm-wave 5g," in *2020 IEEE Texas Symposium on Wireless and Microwave Circuits and Systems (WMCS)*, pp. 1–4, IEEE, 2020.

[5] H.-T. Dabag, B. Hanafi, F. Golcuk, A. Agah, J. F. Buckwalter, and P. M. Asbeck, "Analysis and design of stacked-fet millimeter-wave power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 4, pp. 1543–1556, 2013.

[6] D. Y. Lie, J. C. Mayeda, Y. Li, and J. Lopez, "A review of 5g power amplifier design at cm-wave and mm-wave frequencies," *Wireless Communications and Mobile Computing*, vol. 2018, no. 1, p. 6793814, 2018.

[7] D. M. Pozar, *Microwave engineering: theory and techniques*. John wiley & sons, 2021.

[8] M. Medina, "Rf power amplifiers for wireless communications," *Departement Elektrotechniek*, 2008.

[9] S. Voinigescu, *High-frequency integrated circuits*. Cambridge University Press, 2013.

[10] S. Kassim and F. Malek, "Microwave fet amplifier stability analysis using geometrically-derived stability factors," in *2010 International Conference on Intelligent and Advanced Systems*, pp. 1–5, 2010.

[11] G. Lombardi and B. Neri, "Criteria for the evaluation of unconditional stability of microwave linear two-ports: a critical review and new proof," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, no. 6, pp. 746–751, 1999.

[12] G. Gonzalez, *Microwave Transistor Amplifiers: Analysis and Design*. Prentice Hall, 1997.

[13] M. Edwards and J. Sinsky, "A new criterion for linear 2-port stability using a single geometrically derived parameter," *IEEE Transactions on Microwave Theory and Techniques*, vol. 40, no. 12, pp. 2303–2311, 1992.

[14] M. Steer, *Fundamentals of Microwave and RF Design*. NC State University, distributed by the University of North Carolina Press, 2019.

[15] J. R. Long, "Monolithic transformers for silicon rf ic design," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 9, pp. 1368–1382, 2000.

[16] J. R. Long and M. A. Copeland, "Modeling of monolithic inductors and transformers for silicon rfic design," in *Proceedings of 1995 IEEE MTT-S International Topical Symposium on Technologies for Wireless Applications (Conjunction with INTER COMM'95)*, pp. 129–134, IEEE, 1995.

[17] R. Bajwa and M. K. Yapici, "Integrated on-chip transformers: recent progress in the design, layout, modeling and fabrication," *Sensors*, vol. 19, no. 16, p. 3535, 2019.

[18] H.-M. Hsu, J.-H. Huang, T.-H. Peng, and N.-C. Liu, "Design of coil length of on-chip transformer with high turn ratio and high coupling performance," *IEEE transactions on electron devices*, vol. 59, no. 11, pp. 3061–3068, 2012.

[19] W.-Z. Chen and K.-C. Hsu, "Miniaturized 3-dimensional transformer design," in *Proceedings of the IEEE 2005 Custom Integrated Circuits Conference, 2005.*, pp. 285–288, IEEE, 2005.

[20] C.-C. Lim, K.-S. Yeo, K.-W. Chew, J.-M. Gu, C. Alper, S.-F. Lim, C. C. Boon, P. Qiu, M. A. Do, and L. Chan, "High self-resonant and area efficient monolithic transformer using novel intercoil-crossing structure for silicon rfic," *IEEE electron device letters*, vol. 29, no. 12, pp. 1376–1379, 2008.

[21] H.-M. Hsu and C.-T. Chien, "Multiple turn ratios of on-chip transformer with four intertwining coils," *IEEE Transactions on Electron Devices*, vol. 61, no. 1, pp. 44–47, 2013.

[22] H.-M. Hsu, C.-W. Tseng, and K.-Y. Chan, "Characterization of on-chip transformer using microwave technique," *IEEE transactions on electron devices*, vol. 55, no. 3, pp. 833–837, 2008.

[23] K. Chen, Z. Liu, X. Hong, R. Chang, and W. Sun, "Balun modeling for differential amplifiers," *Proc. World Congr. Eng. Comput. Sci.(WCECS)*, 2019.

[24] B. Razavi, "Design of analog cmos integrated circuits," 2000.

[25] D. Qiao, Y. Li, and Y. Zhang, "Energy efficient video transmission over fast fading channels," *EURASIP Journal on Wireless Communications and Networking*, vol. 2010, pp. 1–12, 2010.

[26] J. W. Rogers and C. Plett, *Radio frequency integrated circuit design*. Artech House, 2010.

[27] B. Razavi and R. Behzad, *RF microelectronics*, vol. 2. Prentice hall New York, 2012.

[28] S. C. Cripps *et al.*, *RF power amplifiers for wireless communications*, vol. 250. Artech house Norwood, MA, 2006.

[29] T. H. Lee, *The design of CMOS radio-frequency integrated circuits*. Cambridge university press, 2003.

[30] AnySilicon, "FDSOI Technology Overview," 2023.

[31] K. Technologies, *The Essential Guide to GaN Semiconductor Device Modeling*. Keysight Technologies, 2023.

[32] T. O. Dickson, K. H. Yau, T. Chalvatzis, A. M. Mangan, E. Laskin, R. Beerkens, P. Westergaard, M. Tazlauanu, M.-T. Yang, and S. P. Voinigescu, "The invariance of characteristic current densities in nanoscale mosfets and its impact on algorithmic design methodologies and design porting of si (ge)(bi) cmos high-speed building blocks," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 8, pp. 1830–1845, 2006.

[33] Y. Kim and Y. Kwon, "Analysis and design of millimeter-wave power amplifier using stacked-fet structure," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 2, pp. 691–702, 2015.

[34] M. Jung, *Ka-band Single-ended 3-Stack Power Amplifier in 45-nm RF-SOI*. PhD thesis, Carleton University, 2020.

[35] J. Du Preez and S. Sinha, *Millimeter-wave power amplifiers*. Springer, 2017.

[36] F. Costanzo, V. Camarchia, N. Carvalho, P. Colantonio, A. Piacibello, R. Quaglia, V. Valenta, and R. Giofre, "A gan mmic stacked doherty power amplifier for space applications," in *2022 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR)*, pp. 29–31, IEEE, 2022.

[37] Wikipedia, "5g nr frequency bands." `https://en.wikipedia.org/wiki/5G_NR_frequency_bands`.

[38] K. Vryssas, *Power Amplifiers for Wireless Metropolitan Area Networks Based on the 802.16 Standard*. PhD thesis, National Technical University of Athens, Athens, Greece, 2008.

[39] F. Chen, Y. Wang, Y.-H. Hsiao, J.-L. Lin, Y.-C. Chen, and H. Wang, "A 4.6-ghz class-f-1 high power cmos power amplifier," in *2017 IEEE International Symposium on Radio-Frequency Integration Technology (RFIT)*, pp. 183–185, IEEE, 2017.

[40] C. Florian, R. Cignani, A. Santarelli, and F. Filicori, "Design of 40-w algan/gan mmic high power amplifiers for *c*-band sar applications," *IEEE transactions on microwave theory and techniques*, vol. 61, no. 12, pp. 4492–4504, 2013.

[41] G. Lv, W. Chen, L. Chen, and Z. Feng, "A fully integrated c-band gan mmic doherty power amplifier with high gain and high efficiency for 5g application," in *2019 IEEE MTT-S International Microwave Symposium (IMS)*, pp. 560–563, 2019.