Energy Efficient Wideband Supply Interpolating Transmitter for Millimeter-Wave 5G System

Amitava Giri

TU Delft Delft University of Technology

Challenge the future

Energy Efficient Wideband Supply Interpolating Transmitter for Millimeter-Wave 5G System

by

Amitava Giri

in partial fulfillment of the requirements for the degree of

Master of Science
in Microelectronics

at the Delft University of Technology,
to be defended publicly on Thursday October 18, 2018 at 14:00.

Supervisors:	dr. S. M. Alavi,	TU Delft
	prof. dr. L. C. N. de Vreede,	TU Delft
Thesis committee:	prof. dr. L. C. N. de Vreede,	TU Delft
	dr. M. Spirito,	TU Delft
	dr. ir. M. A. P. Pertijs,	TU Delft

This thesis is confidential and cannot be made public until December 31, 2020.

TU Delft Delft University of Technology

NXP

To my parents
To my sister
To my cousin, Dalia

"Arise, awake, and stop not till the goal is reached"
— Swami Vivekananda

Preface

During my B.Tech., I realized that we were not being taught much about circuit design. It was this craving to have more in-depth knowledge about circuit design that made me pursue an MSc. I mostly concentrated on European universities and soon found out about TU Delft. I have always wanted to be independent, hence travelling away from home to a completely new land was really enticing to me. But nevertheless, all of this would have still remained a dream without the help and support of my cousin, Dalia. Everything I have been able to accomplish in the past two years, would not have been possible without her support and kindness. I owe a lot to you, Dalia.

It has been the toughest two years of my academic life. Too much to do with too less time in hand. I still remember the long hours spent working on CMOS assignments, staying up late studying for the exams and then grabbing a couple of hours of sleep before waking up at dawn for the last minute revisions. It is probably because of this hard work that I have learned a lot during my Masters. I would like to thank my daily supervisor, Prof. Morteza Alavi for his time and guidance. You have always been available for a discussion and have helped me out in all the phases of this project. I have learned a lot from you. My sincere apologies to your wife and son for keeping you late in the office. I would also like to thank Prof. Leo de Vreede for his guidance during the initial phases of the project and also for accepting me as his student to work on $gENESIs$. Your motivation and inspiring words have helped me to relax and calm my nerves. I would also like to thank my thesis committee members, Prof. Marco Spirito and Prof. Michiel Pertijs for taking the time to read my thesis. Special thanks to NXP Semiconductors and Marcel Geurts for supporting this project and believing in me.

My classmates, especially Sriram, Rushil, Sarthak helped me a lot during my studies and we have spent hours working on numerous assignments. Thanks a lot for your explanations and your efforts to help me understand some key concepts. In this regard, my colleagues on the 18^{th} floor deserve my gratitude as well. Thank you Mohsen Mortazavi and Masoud Pashaeifar for the discussions we had and in guiding me in the right direction. Mohammad Reza, thanks a lot for helping me with the Momentum simulations. Satoshi Malotaux, thank you for the numerous discussions we had about layout and EM simulations. Atef Akhnoukh, thank you for solving all Cadence related issues and giving me tips on layout. Antoon, your help with Linux and other desktop related issues are much appreciated.

My fellow Masters' students, Tim and Marc, thank you for your company. I will definitely miss our 11 am coffee time. Thank you for answering my silly questions about Dutch language and tolerating my ridiculous Dutch pronunciations. Special thanks to Koen for organizing the monthly Friday drinks. It would have been really boring to work alone on the weekends without Gerard and Frederico. Thanks to both of you for providing me company during the weekends. Special thanks to Gerard for sharing your stories and views about life with me. My best wishes to you on starting a new phase in your life.

My heartfelt gratitude to everyone at DelftProject, for organizing the intercultural evenings and for the great food. It was really great to see how each of you welcomed all the international students with open arms. Thanks for spending the Christmas eves with us, away from your families. Hope you continue to support all the future international students in the same way. My meditation instructor, Jurgita Mikelenaite deserves to be mentioned as well. Thank you for volunteering on Sunday afternoons to teach meditation to newbies. The way you welcome all of us with a hug makes me feel at home. Your kind words and spiritual messages have helped me get through tough times. Those forty-five minutes of meditation sessions once a week have helped me calm down and relax during the most stressful times of my thesis.

I would also like to thank all my teachers from my school and college back home, for instilling in me the correct values and ethics; without your teachings, I would not have been able to come this far.

And finally, I want to apologize to my parents for leaving home and coming to Europe. I definitely shattered your dreams by doing so and all I can say is to forgive me. Forgive me for chasing my dreams selfishly. Both of you have done your best to educate me and my sister, and, for that I will be eternally grateful to both of you. Thank you for being such amazing parents and giving us the freedom we wanted by sacrificing your own dreams. I cannot express my love and respect for you in words. Thanks to my little sister, Rumpa for staying back with my parents and taking care of them. My childhood would not have been the same without you dear. You are a wonderful daughter, sister and a selfless soul. God bless you Rumpa.

Amitava Giri
Delft, October 2018

Contents

List of Figures

List of Tables

1

Introduction

1.1. Evolution

WE, i.e., humans are a very special creation of nature. We are so special that it took billions of years of evolution to get us to our present state. Since this is a scientific document, let's not be vague. Let's throw in some numbers to put things into perspective. The ages of stars can be estimated by measuring the abundance of radioactive elements like thorium-232 and uranium-238. This technique of measuring the age of stars and the universe as a whole is called *nucleocosmochronology*. This scientific study estimates the age of our galaxy, the "Milky Way", to be about 13.51 billion years old[12]. However, our own solar system is only 4.571 billion years old, and our "home", the *Earth*, is about 4.543 billion years old, but we humans have been around for only 250,000 years[13]. To make sense of this timeline, imagine that the "Milky Way" is only one year old. Then our "Earth" is only about four months old, and we humans have been here for a mere ten minutes. We have evolved tremendously through out this time so much so that we are a force to be reckoned with. We have changed our environments to fit and serve our purpose, and we have been able to reach this far because of our knack for collective learning; preserving information and then sharing it with one another. This has continued for generations helping us to create new forms of complexity. So, "*INFORMATION*" is very important and plays a crucial role in what we are today.

One of the most successful and necessary inventions in the modern era has been the development of cellular communication technologies. Starting with the telegraph and evolving into pagers and mobile telephones, the world of communication is not what it used to be a couple of decades ago. We have moved from a 2G Global System for Mobile(GSM) to 4G Long Term Evolution-Advanced(LTE-A) system. The motivation is managing more bandwidth while experiencing lower latency. In other words, this means having the capacity to transfer large amounts of data without negligible delays.

While 2G revolutionized digital mobile voice communication, the later variants 3G and LTE-A systems are aimed at providing faster download speed and for making real-time video calls. The data rate has gone up from 64 kbps in 2G to 2 Mbps in 3G to 50-100 Mbps in 4G[14]. Our inherent need is for more and more amounts of data/information combined with the massive world population calls for mobile networks with huge amounts of data. This is promised by the 5G and Beyond (B5G)

mobile networks that can provide very high throughputs per device($\approx$ multiple Gbps) and higher per area efficiency(bps/km^2). It is estimated that, by 2020, about 50 billion devices will be connected to the IP network [14] and that about 50 petabytes of monthly data traffic will be used in 2021 in smartphones. This is about 12 times as much traffic as in 2016[15].

1.2. Technical Objectives for 5G

The next generation of mobile wireless networks, the so-called 5^{th} Generation (5G), is currently under construction, which intends to offer a factor 1000x increase in wireless data handling capacity.

This 1000x increase in data traffic can be achieved by future 5G networks employing three independent mechanisms [14]: First, adopting more (smaller) base-stations within an existing wireless cell (e.g., 56x as many), effectively transferring from a few macro-cells to many small-cells, demands "high integration" of the transceiver. Second, designating (e.g., 3x) more spectral bandwidth through carrier aggregation demands "new architectures" for the transceiver. Third, adopting spatial-division signaling avoids generating interference and boosts the necessary number of transceivers along with their antennas for a wireless node; e.g., using a 6x multi-input/multi-output (MIMO) configuration [15], [16]. MIMO demands the use of "digitalization and signal processing (DSP)" techniques for the transceivers. Combined, these three mechanisms can facilitate a more than 1000x increase in data capacity in the low GHz range (up to 6GHz) while providing connections that are more robust and shorter response times at a lower energy consumption.

Being Electrical Engineers working on designing transmitters, we will only look into the design of transmitters from now on and motivate the design of high frequency power amplifiers in spite of several challenges associated to it. Also we will dive a bit into SOI technology and outline it's several features.

1.3. Motivation for mm-wave Design

According to the Shannon-Hartley theorem, the channel capacity is directly proportional to the channel bandwidth given that the signal power is bounded and that the noise has a Gaussian power-spectral density. Thus, it makes sense that, in order to meet the ever-increasing data needs of the ever so increasing world population, modern day communication systems are moving into Super High Frequency (SHF) and Extremely High Frequency (EHF) bands. According to IEEE standards, SHF includes S, C, X, K_u and K band while EHF includes K_a, V, W and mm/G bands [17].

In the USA, the Federal Communications Commission(FCC) has allocated 27.5-28.35GHz and 37-40GHz for 5G. The above mentioned spectrum is licensed while a 7GHz unlicensed spectrum is available from 64-71GHz. Similarly, South Korea has made 27.5-28.35GHz and 37.5-40GHz band for 5G communication. Japan intends to use 27.5-29.5GHz while China has reserved 24.25-27.5GHz and 27.5-29.5GHz for 5G systems [15]. According to [18], the frequency spectrums under consideration for future 5G systems in Europe are:

- 24.25-27.5 GHz

- 40-43.5 GHz
- 66-71 GHz

Due to the higher path losses involved at these frequencies, communication will be limited to indoor hotspots and outdoor small cells. This means that we will need a large number of cells, which is beneficial as it enables frequency reuse and limits interference over a limited region. This also relaxes the linearity requirements of the transceiver as the path losses isolate one transceiver from another. Thus, this leaves us with certain room for innovation in the direction of high efficiency transceiver designs. In this work, a novel technique is investigated that can provide high efficiency operation even when operating in the 30GHz frequency band where the Power Gain (G_p) of the transistor is highly compromised. Also, mm-wave frequencies enable the design of on-chip passive components due to the small wavelengths involved in these frequencies. As a result, we can have a fully integrated transceiver that will bring down the cost of implementing 5G systems. Figure 1.1 summarizes the fundamental aspects of 5G wireless communication system.

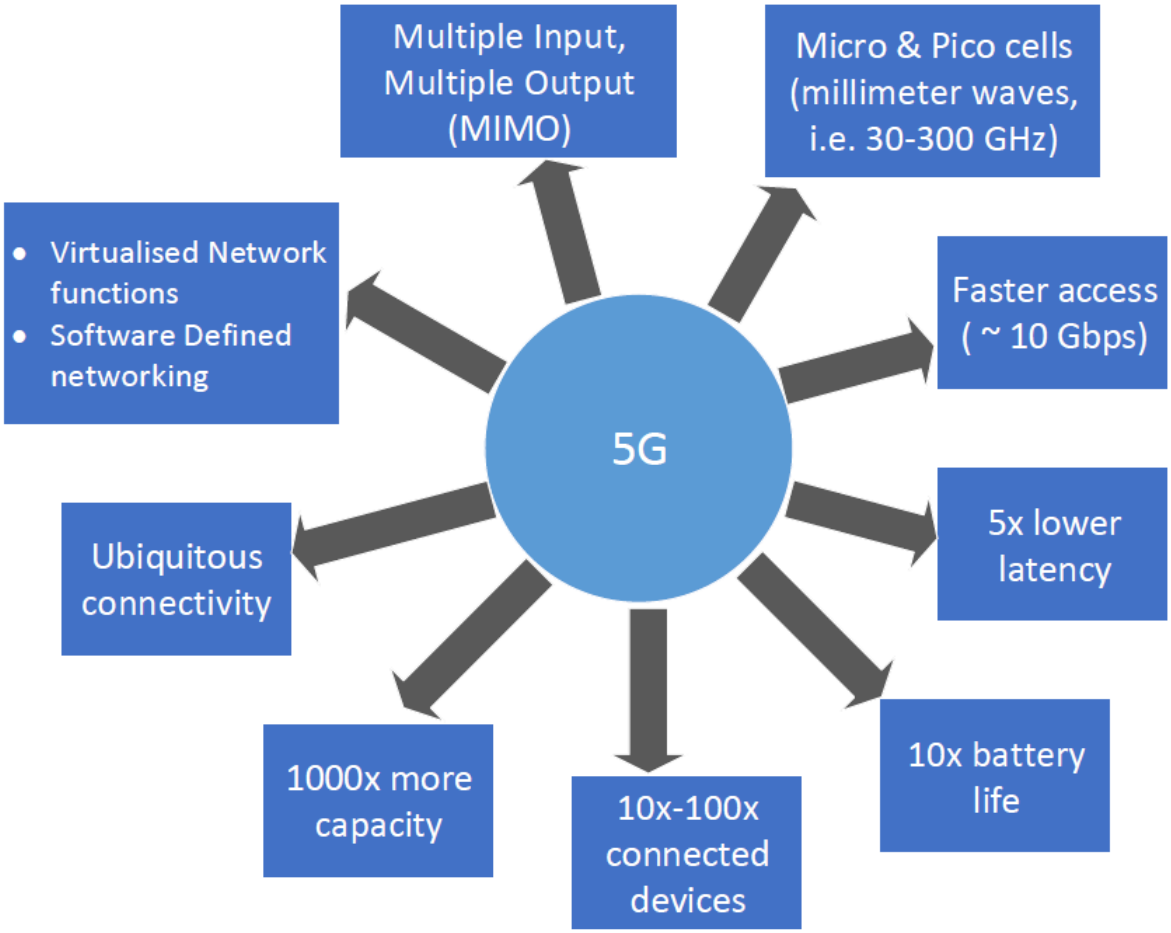


Figure 1.1: **Specifications of 5G**

The craze for high frequency mm-wave design is here to stay for a while because of its use in automotive radar which is gaining popularity as most automobile manufac-

turers turn to the lucrative market of driver assistance systems and self-driving cars. mm-wave radars provide better range resolution and accuracy due to their smaller wavelengths.

1.4. Silicon-On-Insulator (SOI)

As the name suggests, SOI is a variant of the popular bulk CMOS technology. SOI technology consists of an added layer of Oxide that separates the body/bulk from the active area. This way, the parasitic source and drain capacitances can be drastically reduced. This was the major motivation for us to design in SOI technology. 22FDX (22nm Fully Depleted Silicon On Insulator (FDSOI) technology) has been recently launched by GlobalFoundries. Although SOI has been available for a long time, it was partially depleted SOI that suffered from a few problems like the 'Pass-Gate' leakage [19], history effect due to S/D coupling [20], and extra noise margin. However, with the advent of FDSOI, these issues have been resolved. FDSOI was designed to provide FinFET like performance at 28nm die cost [21]. Also, since FDSOI provides different flexibilities in terms of devices with different threshold voltages and the possibility of static/dynamic back-biasing, they can be used for both high performance and low-power designs. The emerging technologies like IoT and 5G RF systems can take advantage of these flexibilities. The forward body biasing(FBB) and reverse body biasing(RBB) mechanisms allow playing with the threshold voltage of the devices in order to have an extended range of operation. For GlobalFoundries, the maximum voltage can be as high as +2/-2 depending on whether FBB or RBB is desired and the type of device NMOS or PMOS. Also, unlike bulk CMOS devices, the range of FBB is not limited by latch-up [21].

All electrical paths between the source and the drain is confined to the narrow region under the gate which ensures better electrostatic control of the channel. This ensures a significant improvement in sub-threshold behaviour and drain-induced barrier lowering(DIBL). At the transistor level, this ensures lower threshold voltage and higher speed. Unlike bulk CMOS, FDSOI does not need any doping or pocket implants in the channel to control the threshold voltage or for better electrostatic control. This ensures superior mismatch as the random dopant fluctuation can be avoided[22].

GlobalFoundries also provides MOS devices in conventional wells and in Flipped wells. A conventional well is when an NMOS is placed in a P-well. These models have regular V_t. The Flipped well models have an NMOS in an N-well. These models are again available in two different models, i.e., a low V_t model and a super-low V_t model.

1.5. Thesis Contribution

Figure 1.2 illustrates a conventional IQ wireless transmitter. This work focuses on the implementation of the in-phase and quadrature phase LO generation, the mixer and the power amplifier. A wideband "Supply Interpolating" transmitter with two amplifier branches has been implemented in this work. The $Main$ branch works in the 8dB power back-off (PBO) region while the $Peak$ is operational in the high power region. A peak output power of 18.4dBm has been generated. The control signal to switch between the two branches has been generated as well. An IM3 of 25dB can be achieved both at peak output power and at 8dB power back-off. A two-stage polyphase filter

has been implemented to generate the IQ LO signal. An image rejection ratio (IRR) > 32dB has been achieved at 30GHz. IRR better than 30dB can be achieved from 26-33GHz. A passive mixer has been used in this work to upconvert the baseband signal to the LO frequency. The amplifier branches, the passive mixer, and the polyphase filter have been laid out. Passives like a center-tapped inductor and the output balun have also been laid out too, EM simulations have been performed to determine their characteristics. Post layout simulation results have been provided in the final chapter. Peak power-added-efficiency (PAE) of 22% has been achieved with 15% PAE at 8.6dB PBO. Supply interpolation enhances the PAE by more than 12% at PBO compared to a conventional PA without any efficiency enhancement scheme.

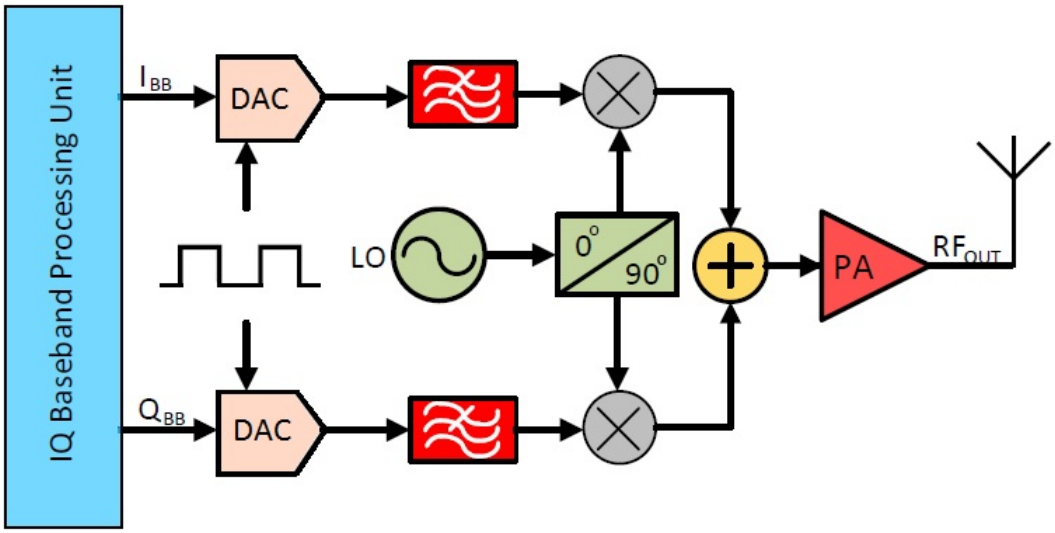


Figure 1.2: **Block diagram of a conventional Cartesian wireless transmitter**

1.6. Outline

In Chapter 2, the prior art in high output power(18-20dBm) will be discussed. Focus will be put on mm-wave high frequency power amplifier designs. This will be followed with a discussion about the various high efficiency techniques and the problems posed by 5G systems that make them unusable.

In Chapter 3, Supply Interpolation will be presented, and we will follow it up with the design choices that we made in order to arrive at our final design.

Chapter 4 will provide a peek into the layout along with a discussion about the important parameters to be taken care of in the layout of mm-wave systems. Chapter 5 presents the post-layout simulation results.

A summary of this work along with recommendations for future research efforts will be provided in Chapter 6.

2

Literature Survey of mm-wave Power Amplifiers

Before we dive into the details of various power amplifiers, it is apt to provide an overview of power amplifiers. A power amplifier (PA) is almost the last stage of any transmitter (TX) that generates desired output power while being energy efficient and reliable. This means that the DC power consumption needs to be reduced. There is no problem in reducing the DC power consumption; in fact, a reduced supply will ensure that the devices will not breakdown and thus leads to a reliable operation. Another criteria that is highly desired from a PA is the ability to accurately replicate the input signal at the output with some constant gain. Simply stated, this means that the spectral purity of the entire system cannot be compromised. However, power amplifiers are also expected to provide high output power, sometimes in the order of watts. For this work, more than 18dBm of output power ($\approx 65mW$) is required. That is where the problem lies. It is not possible to drastically reduce the DC power as that also leads to less output power. One solution would be to design a super efficient amplifier that generates output power almost of the same order as its DC power consumption. There are switching amplifiers like Class-D, E, F and S [23] that can provide very high efficiencies. However, the problem is that the output signal is not linear. Linearity is also very essential as we cannot cause any interference with the adjacent channels and, as a result, transmitting signals in other neighbouring frequency channels is forbidden. The linearity of a system is used to quantify the amount of power being spilled on to the adjacent channels.

A proper design of power amplifier has to take care of all of the above mentioned issues. The design becomes more difficult when migrating to newer technology nodes. The higher technology nodes cannot withstand very high voltages and are more prone to non-idealities like short-channel effects, drain induced barrier lowering (DIBL), and velocity saturation which hampers the performance of MOS devices. Also, another issue when working in the mm-wave domain is the reduced power-gain (G_p) of the devices. Switching amplifiers are not feasible as we cannot generate a proper switching signal at such high frequencies. Simulation shows that an ideal Class-E amplifier, when working with an ideal switching signal can provide efficiencies >90%, but when we replace the perfect switching signal with a sinusoid the efficiency drops to about

70% which means that it is working as a Class-B amplifier (Figure 3.13b). As is evident, the switching behaviour is solely dependent on the driving signal.

A power amplifier design needs a multi-dimensional balance of different design parameters like power-gain, knee voltage, on-chip passives, and matching network efficiency. Even the interconnects become essential and can deteriorate the performance at mm-wave frequencies.

2.1. Basics

As already mentioned, a power amplifier is not a small signal system. However, one needs to take care of both large signal and small signal parameters during its design. The large signal parameters include the power-added efficiency, drain efficiency, saturated output power, -1dB compression point, power gain, and input and output third order intercept points(IIP_3 & OIP_3). On the other hand, the small signal parameters like the S-parameters that are essential for providing proper matching also needs to be considered. Return losses are measured using S_{11} & S_{22}. -3dB bandwidth and stability factors are also important small-signal parameters. Rollett stability factor(K) along with $|\triangle|$ are used to determine if the amplifier is unconditionally stable. Another often used parameter to judge stability is μ. More detailed descriptions about these parameters have been provided in Section 3.3.3.

Due to device reliability issues, a single-transistor stage is not suitable for generating 18-20dBm output power. Hence, a cascode structure or power combining network at the output is generally used. However, this affects the overall efficiency of the system. The required output power cannot be delivered by having a larger transistor and burning more DC power. The PA needs to "see" the proper load (R_{opt}) to deliver the maximum RF output power by maximizing the output voltage and current swings. This is called maximum power matching. A detailed explanation is provided later in Section 3.3.4. The load (R_l) is usually 50Ω, so R_{opt} needs to be transformed to R_l. This impedance transformation ratio ($m = \frac{R_l}{R_{opt}}$) is also critical because it sets an upper limit to PAE. For a simple L-type matching network, assuming that the Q-factor of the capacitor (Q_C) is much larger than that of the inductor (Q_L), insertion loss (IL) is given by [24] :-

$$IL = 1 + \frac{\sqrt{m-1}}{Q_L} \tag{2.1}$$

[25] states that the impedance transformation ratio $m < 5$ is necessary to limit IL, in order to guarantee proper RF output power along with a decent PAE.

2.2. Prior-art in mm-wave Power Amplifiers

Some of the previous works, focusing on 5G applications are reviewed in this section. We will also take a look at a couple of other amplifier designs in the V (40-75 GHz) band.

2.2.1. Without Efficiency Enhancement

A couple of the most recent power amplifier designs for 5G applications are [1] & [2]. In [1], a 28GHz PA was designed for Phased Arrays in 28nm bulk CMOS. A two stage

design in which the Driver Amplifier (DA) is half the size of the output Power Amplifier (PA) has been implemented. The DA also has twice the bias current density of the PA. This has been done to ensure that the DA can drive the PA into compression. Figure 2.1 shows the two-stage amplifier schematic used in this work. Inductive degeneration has been employed in the output stage so as to broaden the inter-stage-matching bandwidth. This in turn also reduces distortion. Degeneration reduces the gain of the transistor and hence the linear range is increased. In this case, inductive degeneration pushes the P_{1dB} point to higher input power level, but the efficiency at power back-off (PBO) is enhanced [5]. Metal trace lines connecting the sources to the ground plane have been used to implement the degeneration inductor. This helps to reduce the chip area as another passive is not required. However, the drawback is that, without proper design, the losses incurred in the metal traces may be higher than what would take place in a proper inductor. This is because the Q factor of the metal trace based inductor will most definitely be much less than that of an actual inductor, but the fact that routing also becomes a bit difficult when an actual inductor is used might be able to justify the use of a metal trace based degeneration inductor. Sub-threshold biasing has been used to enhance efficiency and linearity. The motivation for using subthreshold region is elaborated further.

In the weak inversion or subthreshold region, highest transconductance (g_m) can be achieved for a given amount of current. The transconductance efficiency ($\frac{g_m}{I_d}$) is largest for subthreshold region of operation whereas the transconductance efficiency is poor in strong inversion. This can be demonstrated with the help of the following equations [26]

$$\begin{aligned} I_{d,weak-inversion} &= I_o \,.\, exp\left(\frac{V_{gs}-V_{th}}{n\,.\,V_T}\right).\left(1-exp\left(\frac{-V_{ds}}{V_T}\right)\right).(1+\lambda\,.\,V_{ds}) \\ I_o &= \mu\,.\,C_{ox}\,.\,\frac{W}{L}\,.\,(n-1)\,.\,V_T^2\ ,\ where\ V_T\ is \\ V_T &= \frac{kT}{q} \end{aligned} \tag{2.2}$$

where V_T is the thermal voltage, k is the Boltzmann constant, T is the absolute temperature and q is the electric charge of an electron. n is the subthreshold swing coefficient, also known as the body effect coefficent [27]. n is defined by [27]

$$n = 1 + \frac{3t_{ox}}{W_{dm}} \tag{2.3}$$

where W_{dm} is the maximum depletion layer width and t_{ox} is the gate oxide thickness. Since V_{ds} is usually much higher than V_T which is usually about 25.85mV at 300K, the third exponential term in Eqn. (2.2) can be ignored. In first order approximation, the channel length modulation can be ignored as well. As a result, the drain current in subthreshold (Eqn. 2.2) can be simply expressed as

$$I_{d,weak-inversion} = I_o \,.\, exp\left(\frac{V_{gs}-V_{th}}{n\,.\,V_T}\right) \tag{2.4}$$

The g_m in the subthreshold region of operation is then purely a function of the drain current as given by

$$g_m = \frac{I_d}{n \, . \, V_{th}} \tag{2.5}$$

Higher transconductance efficiency that is available in the subthreshold region has been utilized to design an efficient and linear transmitter in [1].

The output power (P_{out}) is not too high. P_{sat} of 15.3dBm has been obtained with a 1.1V supply. With 64-QAM OFDM signal and 250MHz RF Bandwidth (RFBW), an EVM of -25dB has been achieved. The EVM was measured at almost 10dB power back-off (PBO). Peak PAE of 36.6% has been achieved but the PAE is only 9% at 10dB PBO.

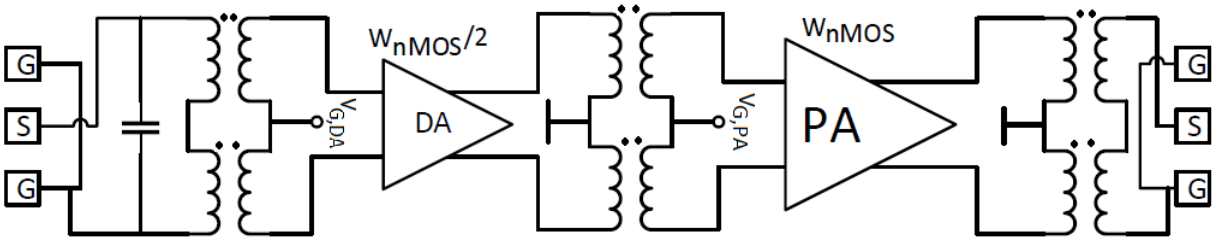


Figure 2.1: **Two stage amplifier adapted from [1]**

In [2], a PA capable of handling a RFBW of 800MHz (which is more in line with the 5G requirements) has been proposed. This work has been implemented in 40-nm CMOS. In this work, a 4-bit gain control (9dB range) for phased-array operation has been presented. Figure 2.2 shows the schematic of the three stage power amplifier. A transformer based matching network with dual-resonance has been used to achieve wideband linearity. Out of the three stages, only the first stage is a Variable Gain amplifier (VGA). The first stage consists of a current steering VGA. 4-bit thermometer code is used to control the gain. The remaining two stages are single-transistor stages with capacitive neutralization. In this paper, a $P_{sat} = 15.1dBm$ with a PAE of 33.7% for a 64QAM OFDM signal has been reported. A RFBW of 800MHz that consists of an 8-component-carrier (8CC), each 90MHz wide with a 10MHz wide guard band has been employed. The P_{out} of 6.7dBm with a PAE of 11% has been achieved with an EVM of -25dB. The drawback is that many passives have been used, however, this is unavoidable when using multiple stages at mm-wave. Inter-stage matching is essential to ensure power operation by minimizing the loss in signal power. The price paid is the increase in chip area. The peak output power is only 15dBm even after using a three stage PA. This is 3dBm less than what we can generate in this work.

Another work in the 5G domain [28] uses Class-J PA to deliver 16.2dBm output power with a PAE of 39%. In this design, 28-nm FDSOI CMOS technology has been used. Post-layout simulation results have been provided in the paper. The only notable thing is that it is a single-ended design and thus lacks the advantages provided by a full/pseudo differential design. No results for the EVM of the system has been mentioned in the paper.

Figure 2.3 shows the schematic of a Class-J PA for 5G applications presented in [3]. A cascode topology has been implemented that can generate more than 18dBm peak output power at 27, 28, and 29GHz. The corresponding peak PAEs are 33.8%, 35.3%

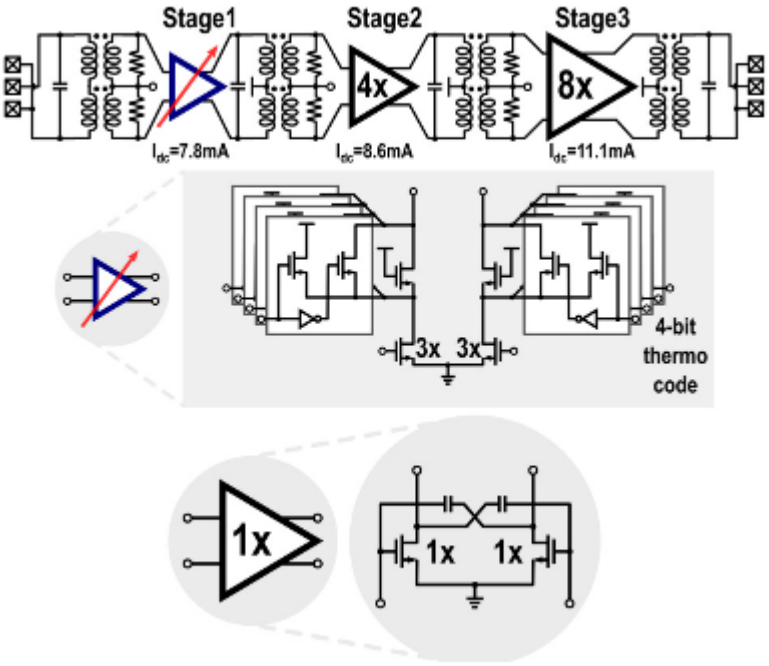


Figure 2.2: **Three stage amplifier along with a cascode VGA used in [2]**

and 34.7%. The PAE at 7dB PBO is only about 9%, except for the 29GHz band in which a PAE of 11.3% has been achieved. The technology used is 120-nm SiGe BiCMOS. No information has been provided about the type of signal used for measurement and also the RFBW and EVM informations are missing. The design is single ended in nature and, hence, is more error prone. It is sensitive to noise and common mode signals. The output matching network has been tuned to provide close to ideal impedance at the fundamental and the 3^{rd} harmonic. The 2^{nd} harmonic impedance was not optimized as a more extensive matching network would have then been necessary. This would have led to a larger chip area and also considerable losses. Due to the above reasons, Class-J PAs are still not a popular choice.

A full Digital Power Amplifier (DPA) has been presented in [4], in which a 6-bit DPA has been designed in 28-nm CMOS. The 60GHz stacked Class-E PA can support a 25Gb/s signal while being linear (EVM < -26dB). Figure 2.4 shows the schematic of the stacked gate-controlled DPA implemented in [4]. A series resistor has been added to the final inverter stage in order to ensure that the resultant RC cut-off frequency is well above the modulation bandwidth but is sufficiently less than the carrier frequency of 60GHz. This DPA can support a wide range of modulation signals from QPSK to 64-QAM, and measurement results have been presented for each of these signals. Measured EVM is about -26dB for a 64QAM signal with a RFBW of 4GHz. This is more than enough to support very high speed 5G systems. However, the main focus of this work is 802.11ay PAN standard [29]. As a result, the peak P_{out} requirement is different (much less). A peak P_{out} of only 6.9dBm has been achieved for a QPSK signal while it is only 0.6dBm for a 64QAM signal. A pattern-dependent DPD has been used in

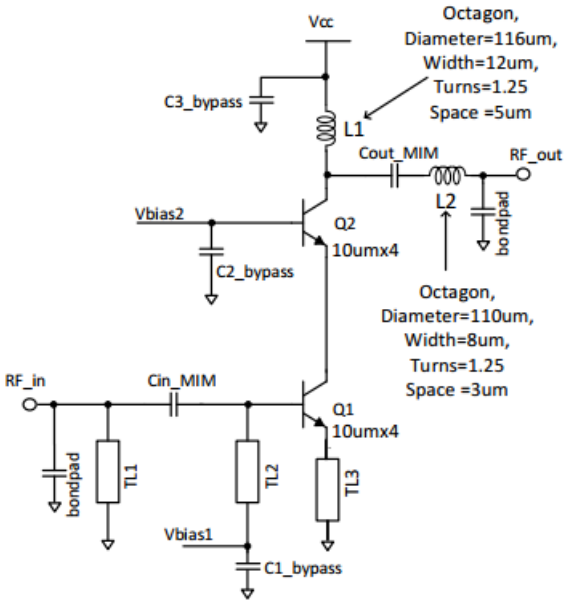


Figure 2.3: **Single stage amplifier operating in Class-J used in [3]**

this design to meet the stringent linearity requirements. The pattern-dependent DPD alleviates the problems of memory effect and bandwidth limitation in the up and down converters. Such a DPD also minimizes the adverse effect of delay mismatch in the amplitude and phase paths. However, the signal bandwidth needs to be sufficiently high if DPD is implemented. This would complicate the design in a 5G system as both high output power and linearity is required. The stacked structure also provides a shielding action and hence there is less leakage to the output.

A high efficiency linear PA has been implemented for 5G systems in [5]. A two stage CS PA with inductive degeneration has been used here. About 18dBm of output power has been reported with peak PAE% of 41.5%. A parallel power combiner has been used at the output. AM-PM distortion has been plotted against different bias voltages to determine the optimum bias voltage. However, [5] states that the optimum bias point found using this method does not provide optimal performance in terms of efficiency and output power. Thus, different bias points were chosen for the DA and the PA in order to simultaneously optimize AM-PM distortion, efficiency, and output power. A 64-QAM signal with RFBW of 1GHz has been measured to provide an EVM of -25dB. An output power of 8.4dBm with a PAE of 8.8% can be obtained at this EVM. Figure 2.5 illustrates the amplifier schematic presented in [5]. Metal traces have been used to implement the degeneration inductors. This helps in saving chip area and also manages to utilize the metal traces to the designer's advantage.

Discussion

- The Class-J based PA proposed in [28] and [3] have not been considered in this work because of their single-ended nature. It is susceptible to common mode errors, ground bounce, and other noise sources. This is the reason why a push-pull amplifier has been used in this thesis. Also, for Class-J amplifiers, proper

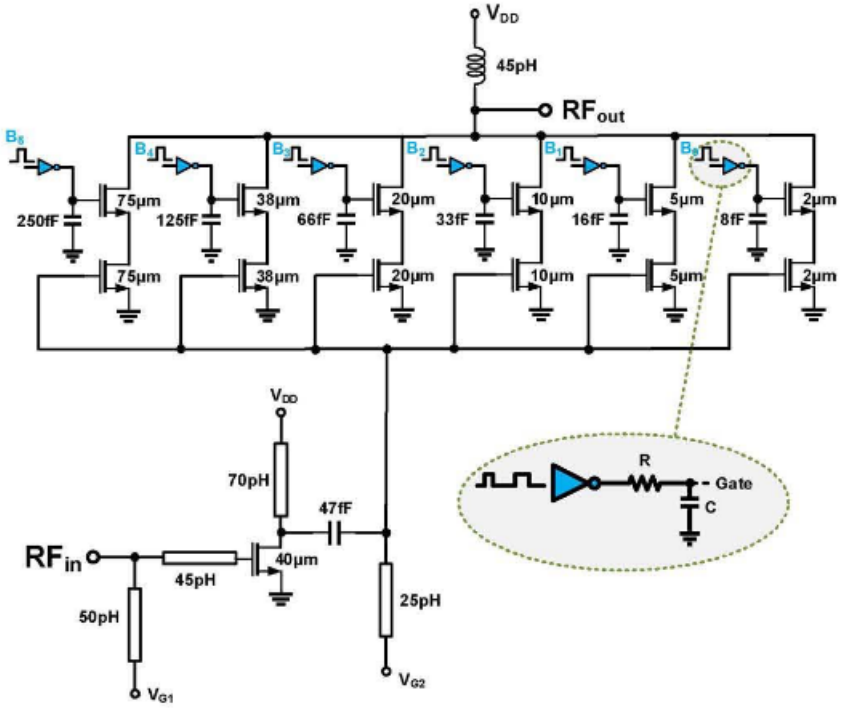


Figure 2.4: **Stacked gate-controlled Class-E DPA used in [4]**

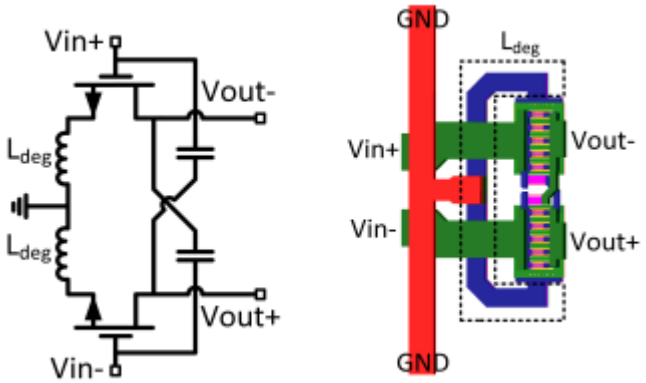


Figure 2.5: **Differential CS pair with inductive degeneration used in [5]**

impedance needs to be provided to the higher harmonics like the 2^{nd} and 3^{rd}. The matching network becomes extremely complicated and hence more losses are incurred in such a scenario. The use of transmission lines also occupies a lot of chip area.

- The work demonstrated in [4] promises to provide error-free operation even at data speeds $> 20Gb/s$. A stacked Class-E DPA has been implemented. The gate of the CG transistor has been used for amplitude modulation by controlling it digitally. Measurement results have been provided for both continuous wave signals and for modulation signals. The system is capable of handling large

RFBW of the order of 4GHz (64-QAM signal has been used so each symbol consists of 6bits. Hence, $24/6\ Gb/s = 4\ GSymbols/s$). However, the system only provides a saturated output power of about 7dBm. Output power in the order of 18dBm is necessary for 5G systems. So, this architecture in its present state does not satisfy all the specifications of a 5G system.

- [5] is based on the designs proposed in [1] & [2]. Both the EVM and RFBW specifications for 5G systems have been accomplished. The use of inductive degeneration leads to a more linear system and helps to meet the EVM specifications. The use of metal traces to implement the degeneration inductor also saves a lot of chip area and simplifies the routing. This design technique has been adopted in this thesis work with slight modifications. The metal traces connecting the source of the PA to the ground line has inductive properties at mm-wave frequencies. Hence, the linearity of the PA will be slightly enhanced in post-layout simulations. This is obtained without much design effort albeit at the cost of output power and less gain. A detailed description about the layout of this thesis work has been provided in a later chapter.

- PA designs proposed in [1] & [2] are more inline with the 5G specifications. A two-stage and three-stage PA has been proposed in these works. Due to its simplicity, this architecture is able to handle large RFBWs and also delivers about 10% PAE at PBO. This is in accordance with some of the other works in the mm-wave regime. The output power is slightly on the low side. So, in this thesis work, attention has been paid to ensure that peak output power of $> 18\ dBm$ can be achieved. The PAE at PBO has been optimized in this thesis with the help of a novel technique.

2.2.2. With Efficiency Enhancement

Most base stations employ some form of advanced architechture like Doherty, outphasing, or Envelope Tracking (ET) in order to tackle the high PAPR values associated with modern day communication standards but, with the demand for larger bandwidth on the rise, all of these techniques lose their charm. However, some of these published works both in the sub-mm-wave and mm-wave regime have been reviewed in this section. Layout schemes proposed in a couple of the works elaborated below have been used in our own design. Lastly, a motivation for the present thesis has been provided.

[30] briefly states some of the techniques used to generate high efficiencies at power back-off. In [31], ET was used to generate more than 43% PAE with an average output power of about 24dBm for WiMAX/3GPP LTE transmitters. A self-biased cascode PA has been used. It has been shown that self-biased cascode SiGe PA leads to lesser on-resistance (R_{on}) than a constant-biased cascode PA. A linear Op-Amp with a Class-AB output stage has been used to design the envelope modulator, but this system is only capable of handling a RFBW of only 5MHz. This work has been demonstrated in 0.35μm SiGe BiCMOS technology as such a system is not capable of handling large RFBW.

A GaN PA has been reported in [32] for 5G applications that can achieve peak $P_{out} = 34dBm$ with a PAE of 30% at P_{1dB} and a peak PAE of 38.7%. Dynamic supply

modulation is used here to achieve higher PAE. This system is capable of handling a maximum RFBW of 20MHz. Using 0.25μm GaN/SiC HEMT technology, Watt-level linear output power has been achieved. This technology is a not a good choice for 5G applications as the maximum reported F_t for this technology is 27GHz [32].

A novel three-bit supply modulator with Digital Predistortion (DPD) that also employs ET has been proposed in [6] to achieve a total system efficiency of 23.9%. Peak P_{out} of 44.5dBm has been reported . The eight voltage levels provide a better resolution than having multilevel voltage sources. This design performs better than [31] as it does not require the use of a less efficient linear Op-Amp in the envelope modulator. Thus the overall system efficiency is improved. Here, the RFBW is limited to 10MHz which is standard for LTE signals. Figure 2.6 shows the multilevel cascaded power-DAC used in [6]. The power-DAC is called so as it converts the digital bitstream directly to analog output power. The DC voltage levels connected to each half bridge converter has been scaled by 2. As a result, $V_{DC}^1 = \frac{V_R}{2}$, $V_{DC}^2 = \frac{V_R}{4}$ and so on.

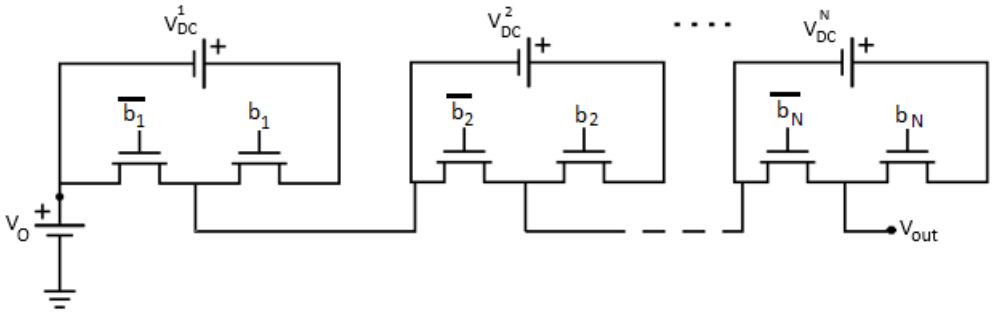


Figure 2.6: **Cascaded multilevel power-DAC used in [6]**

By controlling the value of the bits, the output voltage can be controlled. This can be expressed by [6]

$$V_{out} = \sum_{i=1}^{N} b_i \frac{V_R}{2^i} + V_O \tag{2.6}$$

In the absence of an offset voltage, the dynamic bias needs to provide a voltage swing of 0-48V. In order to prevent the degradation in RF power efficiency, the voltage swing of the dynamic bias has been reduced to about 42V thus an offset voltage of 6V has been used in this design. The voltage swing of the power-DAC is between V_O and V_M, where V_M is given by [6]

$$V_M = V_R\left(1 - \frac{1}{2^N}\right) + V_O \tag{2.7}$$

In this proposed design, the switching frequency of the i^{th} cell is given by [6]

$$f_{sw} = 2^{i-1} f_{in} \tag{2.8}$$

where f_{in} is the input signal frequency and $i = 1$ for the MSB and $i = N$ for the LSB cell. Thus the losses incurred due to switching have been minimized as the faster switching devices are connected to progressively lower DC voltages.

An all-digital RF transmitter in 40-nm CMOS has been presented in [7]. Because of the rise in RFBW for high speed operation, a direct quadrature architecture has been used in this transmitter. Thus, a power hungry CORDIC system is not required. Also, a 2D Look-Up table (LUT) has been used to perform digital pre-distortion. The baseband signal is also upsampled to 800MHz. 5x oversampling is used to reduce the far-out noise and to push away the sampling replicas. This is followed by another 2x oversampling after the DPD operation. One thing that makes this design non-portable to other frequency bands is the use of 2x carrier frequency and then using a "divide-by 2" circuit to generate the quadrature signals. Figure 2.7 shows the all-digital quadrature transmitter, and Figure 2.8 illustrates the unit DPA cell proposed in [7]. A current-mode Class-D operation along with a transformer based matching network has been implemented that ensures zero-voltage switching (ZVS) thereby improving the efficiency. The sampling clock for the digital front-end has been derived from the carrier frequency by means of a divider. A duty-cycle-control (DCC) circuit has been used in the LO path to improve the efficiency. The DCC ensures that both sides of the differential pair are not switched on simultaneously. The currents from the I and Q paths have been combined at the output to generate the RF output power. Even when tested with a 160MHz RFBW 64-QAM OFDM signal, the transmitter can provide 13.5dBm output power with -30dB EVM. The RFBW is much larger than what we have seen so far. The AND gate (Figure 2.8) is used to enable or disable a particular leg of the differential pair. This is followed by buffers at the output to compensate for the loss of signal in the interconnects.

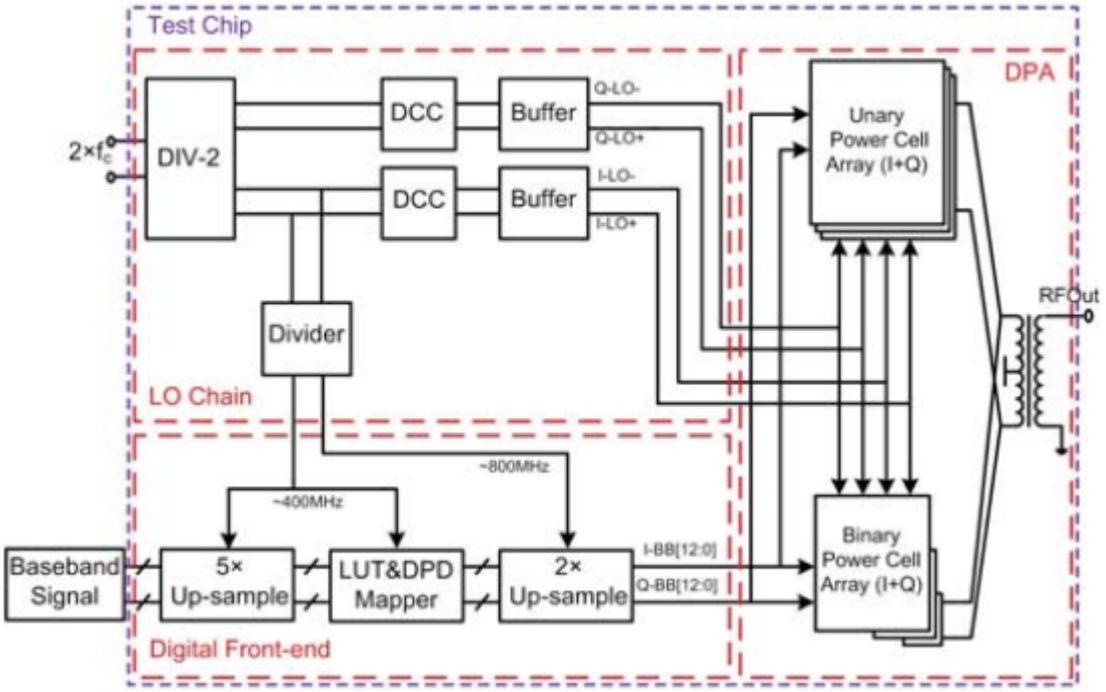


Figure 2.7: **All-digital quadrature transmitter used in [7]**

In [8], a digital intensive DPA/power-DAC topology has been used to generate watt level peak P_{out}. Measurements have been performed in a 33-46GHz frequency band.

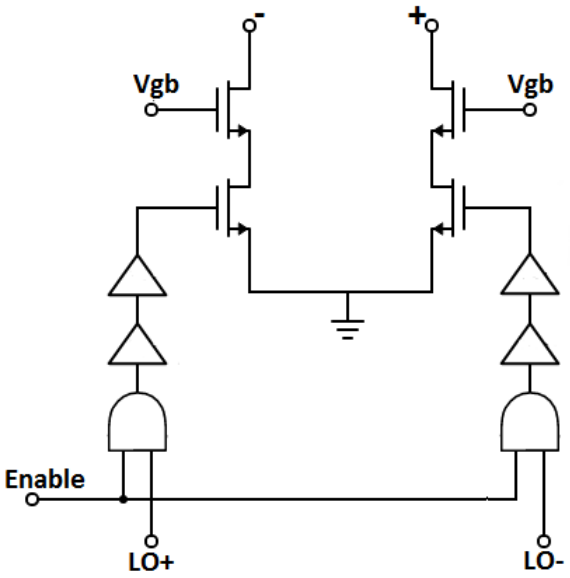


Figure 2.8: **DPA unit cell adapted from [7]**

High output power has been generated by using stacked Class-E like PA unit cells that have been power combined at the output by using a non-isolating power combiner. Supply switching has been employed to minimize the DC power consumption in the OFF unit-cells. The basic architecture has been illustrated in Figure 2.9. The use of non-isolating power combiner enables load modulation which helps to achieve better efficiencies. If an isolating power combiner had been used then the total drain efficiency would have been similar to Class-B [8]. This can be intuitively explained in that isolating power combiners burn the power from the unit-cells in the resistor (for a Wilkinson combiner) to maintain isolation.

Another issue outlined in this work is that it is difficult to create high characteristic impedance (Z_O) transmission lines that are less lossy and, at the same time, meet the electromigration rules. So, a C-L-C π-matching network has been proposed that can meet the high Z_O requirements as well as behave as a quarter-wave transmission line at the desired frequency of operation. The π matching network has been implemented with spiral inductors. The parasitic capacitances of the inductor have been absorbed in the design and form the key component of the C-L-C network. More than 23dBm of output power with a peak PAE of 10.7% for the stacked 8-way combiner configuration has been reported. The PAE at 6dB power back-off is only about 4.5%. The area occupied is also about 2-4 times larger than the other works that have been compared to in the paper. This difference is partly because the output power is about 6-7dBm higher and also because of the use of eight spiral inductors. A DNL of less than 0.5LSB and an INL of 1LSB has been reported with end-point fitting.

Although Doherty PA first came to light almost 80 years ago, it is still a popular choice for most PAs used in the communication industry. This has been proved by [9], in which a multiband linear Doherty PA that operates in 28GHz/37GHz/39GHz bands for 5G applications has been proposed. The conventional output network has been replaced by transformers in both the main and the auxiliary paths. This helps

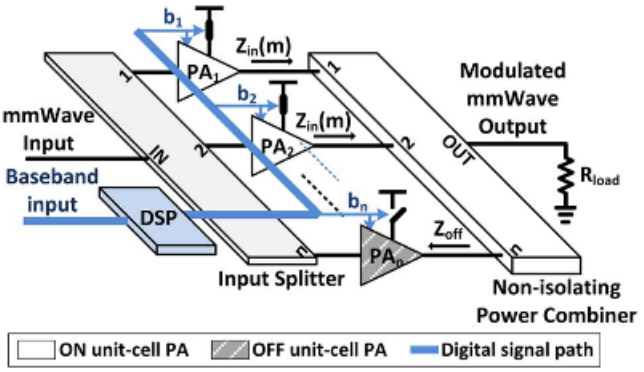


Figure 2.9: **Load modulated power DAC proposed in [8]**

to reduce the impedance transformation ratios (ITR) at PBO. It has been stated that, in the conventional Doherty network, the ITR at 6dB PBO is 4 which can be reduced to 1.65 with the proposed matching network. This drastically improves the efficiency at PBO and also makes it more wideband [33]. Two stages (DA & PA) have been used in this design. Transformers have been employed for inter-stage matching. The input match has been improved by using varacter-tuned transmission lines (TLs) to make it more wideband. The quadrature signal has been generated on chip with a differential hybrid. Figure 2.10 illustrates the proposed design. Peak P_{out} of > 16dBm with a peak PAE $> 20\%$ has been reported in all three bands. At 6dB PBO, PAE's of $> 12.8\%$ has been achieved while it is $> 15\%$ in the 37GHz band. The RFBW used for the measurement is 500MHz, but a RFBW of 1GHz has been reported for 28GHz band with EVM better than -26dB.

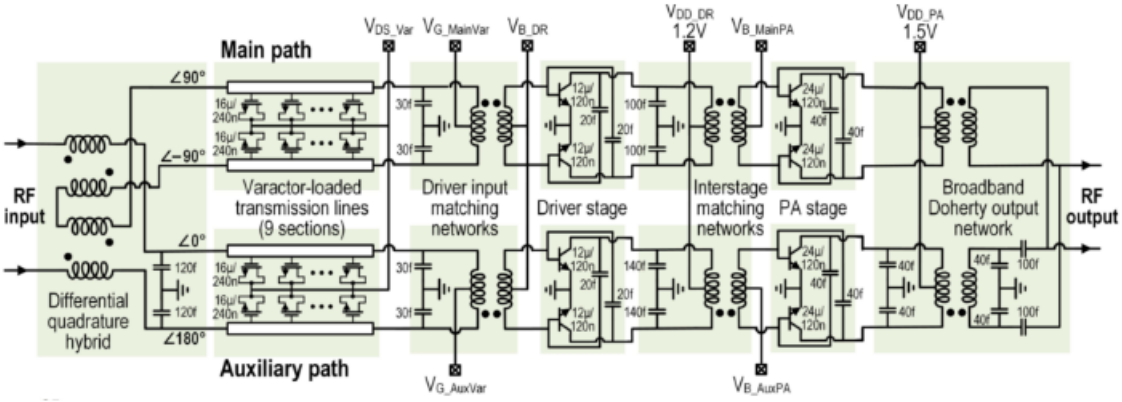


Figure 2.10: **Multiband Doherty PA presented in [9]**

Even though [10] is not exactly for 5G applications, it still addresses the problems faced in mm-wave design. A new transistor layout scheme has been proposed that reduces the device and interconnect parasitics that play a very essential role at mm-wave frequencies. A 60GHz Class-AB PA has been presented for a dual mode of operation, i.e., for high power (HP) and for low power (LP). In the LP mode, one of the transistor branches is shorted at the output to reduce the DC power consumption. This work has been implemented in 40-nm bulk CMOS technology. 17dBm peak P_{out} with a

peak PAE of $> 30.3\%$ has been reported. The PAE at 6dB PBO for the high power mode is only 10% (graphically estimated). However, only continuous wave measurement results have been provided in this paper. Figure 2.11 shows the schematic of the dual mode PA presented in [10].

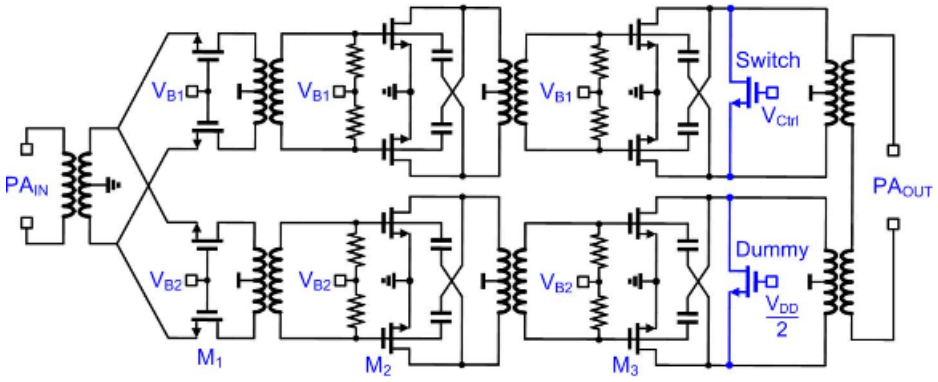


Figure 2.11: **Dual mode PA presented in [10]**

The design proposed in [11] is more relevant to our design. In this work, a cascode structure has been implemented. This is unique in the sense that it consists of a neutralized CS stage and a bootstrapped cascode device (Figure 2.12). A capacitor is placed across the CG device. Due to the Miller effect, this capacitor appears at the drain of the CS device as $(1-A)C_M$, where A is the voltage gain of the CG transistor and C_M is the capacitance placed across CG device. This means that a negative capacitance is created between the drain and the source of the CS stage. This negative capacitor, if tuned correctly, will cancel out the C_{ds} of the CS stage and hence prevent shunting of RF signals to the ground. This leads to higher power gain. This is similar to placing an inductor at the drain of the CS transistor, but placing an inductor complicates the layout. Fortunately, the bootstrapped capacitor can be implemented with metal lines drawn between the source and drain nodes of the CG stage. Hence, the layout complication is reduced. Nevertheless, the use of a bootstrapped capacitor leads to increased instability. Hence, in our work, bootstrapped capacitors were not used.

The measurements have been performed at 73GHz. 18.9dBm of P_{1dB} and a saturated output power of 22.6dBm has been achieved with a maximum PAE of 19.3%. EVM results for 64-QAM signal with a RFBW of 400MHz and 600MHz has been provided. EVM is better than -24dB in both cases.

Discussion

A short comparison of the above mentioned works have been provided. More emphasis has been put on pointing out the shortcomings of each work. A few notable features have also been highlighted.

- Due to the poor efficiency and inability to handle large bandwidths, Op-Amp based envelope modulators proposed in [31] are not a suitable choice for modern communication systems like 5G.
- The GaN based PA [32], although shown to generate very high output power

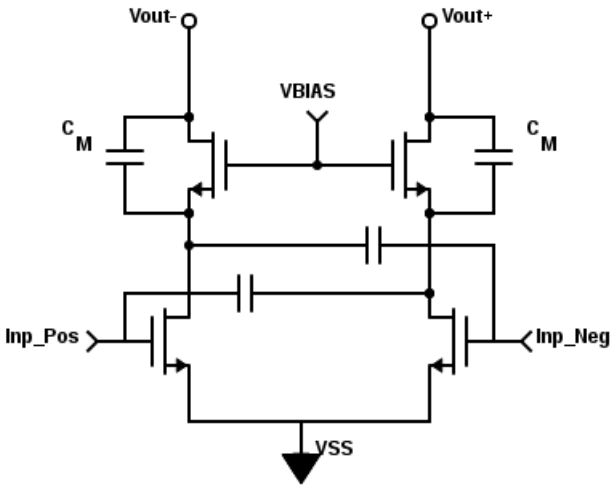


Figure 2.12: **Neutralized bootstrapped cascode PA adapted from [11]**

and PAE, is not suitable for systems with large RFBWs. The 0.25μm technology used in this work is also not suitable for mm-wave design.

- The 3-bit supply modulator proposed in [6] manages to provide better voltage resolution. By employing a binary coding for the DC voltages, the voltage resolution has been minimized to $V_M/(2^N - 1)$. This is lower than other implementations in which the same number of power DAC stages (N) with fixed DC voltages can provide a voltage resolution of only V_M/N. However, the price paid for this higher voltage resolution is the higher switching speed that is necessary in each power DAC stage (Eqn. 2.8). This frequency increases by power of two with each stage. For an input signal of several hundred MHz, generating these high frequency switching signals would be a problem. Hence, this architecture is also not a good choice for 5G systems.

- The all-digital RF transmitter proposed in [7] is not suitable for mm-wave systems as the synthesis of high frequency clocks is difficult. High frequency clocks usually have lower driving capability because of reduced voltage swings. This system also suffers from poor efficiency at PBO.

- The maximum PAE is limited in [8] because of the use of stacked PA topology. Very high output power can be achieved because of stacking. However, a lot of chip area is consumed due to the use of several spiral inductors and long transmission lines that are still used in the DPA unit cells. The use of spiral inductors along with the parasitic capacitances relaxes the design of the power combining network, but this design also relies on load modulation to enhance its efficiency. However, for 5G systems, due to the use of massive MIMO, the load witnessed by the PA and the antenna changes dynamically. Hence, the efficacy of load modulation techniques is severely affected.

- In [9], the system is capable of handling sufficiently large RFBW but, since the design is based on load modulation technique (Doherty), such an architecture may not fare well in massive MIMO systems. However, the PAE at PBO is at least 3-4% better than all of the works (except the GaN PA in [32]) cited above. This can be attributed to the superior nature of Doherty amplifiers and also to the fact that the impedance transformation ratio has been drastically reduced at PBO with the help of the modified transformer based output matching network.
- The cascode PA proposed in [11] illustrates layout techniques for the mm-wave PA that has been adopted in this work. Although the EVM has been reported for slightly smaller RFBWs, the frequency of operation is 73GHz. Thus a superior EVM can be expected from such an architecture at half this frequency.
- For 5G systems, massive MIMO is going to be used for a better link performance. What this means is that the load experienced by the antenna and the PA keeps on changing dynamically. Doherty PA is very sensitive to load modulation. So, in real-time, Doherty PA would not be able to work to the best of its capability.

In this work, load modulation techniques will not be used to achieve higher back-off efficiency since supply modulators are not efficient for signals with large RFBW. Hence a novel technique has been adopted to increase the efficiency at PBO.

Table 2.1 outlines the various performance metrics and their desired values for a modern 5G transmitter.

Metric	Value
System Efficiency	>20%
Output Power	18 dBm ~ 63 mW
Modulation Bandwidth	800 MHz
EVM	~ -37.5 dB (<1.5% with DPD)
Power Back-Off	10 dB

Table 2.1: Performance Requirements

2.3. Conclusion

The different innovative power amplifier design techniques proposed in literature have been reviewed. The shortcomings of various designs, as far as their applicability to 5G systems is considered, have been highlighted. A couple of those presented above are plausible design solutions for 5G systems. Inspiration has been drawn from some of the design and layout choices elaborated in these works. The concept of ”Supply Interpolation” has been introduced in the next chapter along with a basic background of the different circuit components.

3

Supply Interpolation

Before the concept of ”Supply Interpolation” is introduced, background information will be provided about the challenges involved in designing a PA. A high-level description of ”Supply Interpolation” is followed by examining the benefits of SOI technology. A brief description is provided regarding the various pros and cons of SOI and CMOS bulk technology. This is followed by a more circuit level schematic of ”Supply Interpolation”. Admittance parameters are used to provide an insight into the workings of a two-port network. The different configurations of ”Supply Interpolation” are provided along with a motivation that leads to the final schematic. The supporting circuit blocks like the mixer and the polyphase filter is elaborated.

3.1. Recap

As already stated, different techniques have been used to obtain more efficiency from PAs without compromising its linearity. To ensure spectral efficiency, different modulation schemes are used. These modulation schemes affect the PA design because more advanced modulation types have a high peak-to-average-power ratio (PAPR). This means that the input signal reaches its maximum power level for only a small amount of time. In other words, the PA operates in the power back-off regime instead of its peak power. This drastically affects the drain efficiency and the power-added-efficiency (PAE) of the PA.

For example, in BPSK modulation, according to IEEE 802.11a specifications, the maximum expected PAPR is 52 or 17dB. However, this is unlikely as we have a scrambler so not all of the subcarriers in an OFDM symbol are equally modulated. Figure 3.1 shows the plot of PAPR for the ideal case when using BPSK OFDM modulation with 52 subcarriers. It is clear from the plot that most of the power will be concentrated at about 10dB power back-off. This means that, most of the time, the PA is operating from a certain supply voltage but is not generating enough power as the input signal is low. Thus, the efficiency of the system suffers. A simple MATLAB example [34] shows that the difference in PAPR for a 64-QAM signal and OFDM modulation of the 64-QAM signal can be as high as 6dB. Hence, while the PAPR for a 64-QAM signal is only 3.7dB, the corresponding PAPR for 64-QAM OFDM signal is more than 9.4dB.

Now that the importance of efficiency at Power Back-off (PBO) has been successfully established, it is obvious why so many researchers have tried to maximize the

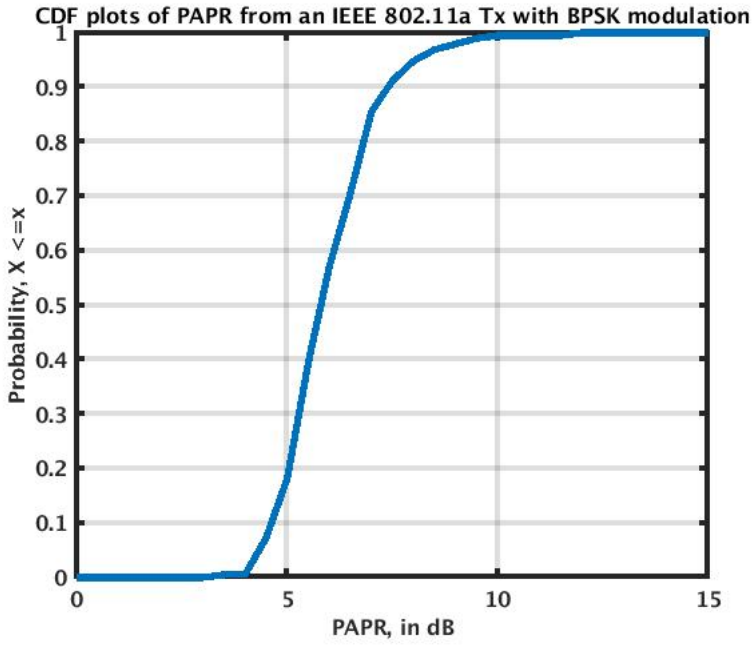


Figure 3.1: **Cumulative Distribution Function of BPSK OFDM signal**

efficiency at PBO. The golden ways of doing this are:

- Supply modulation
- Load modulation

It has been shown that supply modulation is limited to signals with a radio frequency bandwidth (RFBW) in the range of a few tens of MHz. The term video bandwidth is also used to refer to the instantaneous bandwidth of a signal. In this thesis, RFBW has also been used to refer to the instantaneous bandwidth of a signal. The RFBW is usually small as the efficiency of the supply modulator is severely affected when the RFBW increases. However, for 5G systems, supply modulators will not be a good choice as the entire system efficiency will degrade. On the other hand, the performance of load modulation technique like Doherty does not suffer from large RFBW. However, the caveat here is that, due to massive MIMO in 5G systems, the load of the antenna along with that of the PA is dynamically changing. Hence, extra measures are necessary to ensure that the Doherty configuration works properly.

Thus, in this work, a novel technique has been investigated. This technique does not have the problems that are associated with supply modulation. ”Supply Interpolation” as a concept is not new and has been around for almost two decades. A low frequency implementation of the concept can be found in [35]. In this work, the same concept has been utilized to achieve higher efficiencies at PBO but without the limitations of a dynamic supply modulator. Class-G amplifiers are not suitable due to hard-switching between different supply voltages and hence are prone to severe transient disturbances in the output signal. Another used technique is multi-level outphasing. Here, the output power is modulated by switching the supply of a switched-mode PA

to different levels. However, like the normal supply modulation concept, this is also not suitable for 5G applications.

3.2. Supply Interpolation

This section explains the concept of "Supply Interpolaton" (SI) and also presents a high level implementation of the concept.

Instead of switching the supply to various discrete levels, we have different amplifier branches connected to different supply voltages. Output power can be controlled by turning on or off the amplifier branches connected to the most appropriate supply voltages. This way, we can reap the benefits of supply modulation. This method is not dependent on the bandwidth of the signal, and now the control can move to the baseband side. This makes processing much easier. The gradual transition between the amplifier branches can be controlled through its bias or by controlling the gain of the amplifiers. This can also be done digitally by using digital control word to control a RFDAC that then drives the main/peak amplifier branch. In this thesis, however, we have concentrated on the analog implementation of this concept. A conceptual schematic has been illustrated in Figure 3.2. It consists of a digital signal processing (DSP) block that generates the digital bitstreams. The analog/digital upconverter converts the digital signal to an analog signal and upconverts it to the LO frequency. The controller block is responsible for generating signals that determine the switching on/off of the different amplifier branches, i.e., whether to turn on the "Peak" branch or to continue using the "Main" branch. The "Peak" branch is connected to a higher voltage source and hence is used to generate higher output power. In the low power regime, the "Peak" branch is switched off to enhance efficiency. At PBO, the "Main" amplifier branch is used to reduce DC power consumption. This provides a peaking of efficiency at PBO. The outputs of the different amplifier branches can be connected as depicted in Figure 3.2, or they can be connected to power combining networks. This will be elaborated upon later.

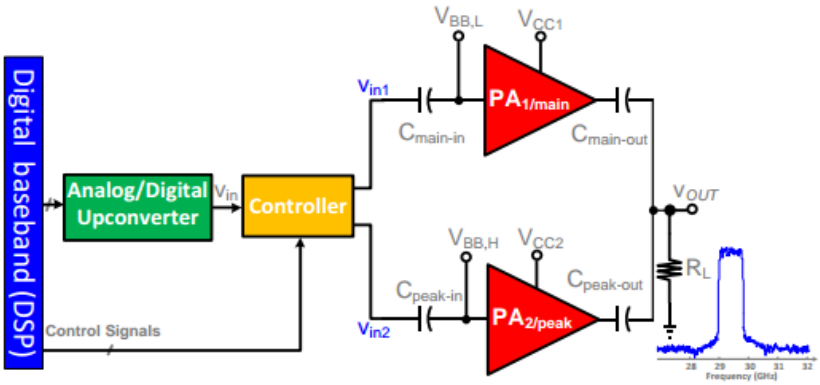


Figure 3.2: **Conceptual schematic of Supply Interpolating Transmitter**

Figure 3.3 illustrates the basic working principle. When the input amplitude is small, the amplifier connected to the lower V_{DD} is used. From Figure 3.3, we can see that PA_1 is on when $v_{in} < v_{in,max}$. During this phase, PA_2 is off. This saves the DC

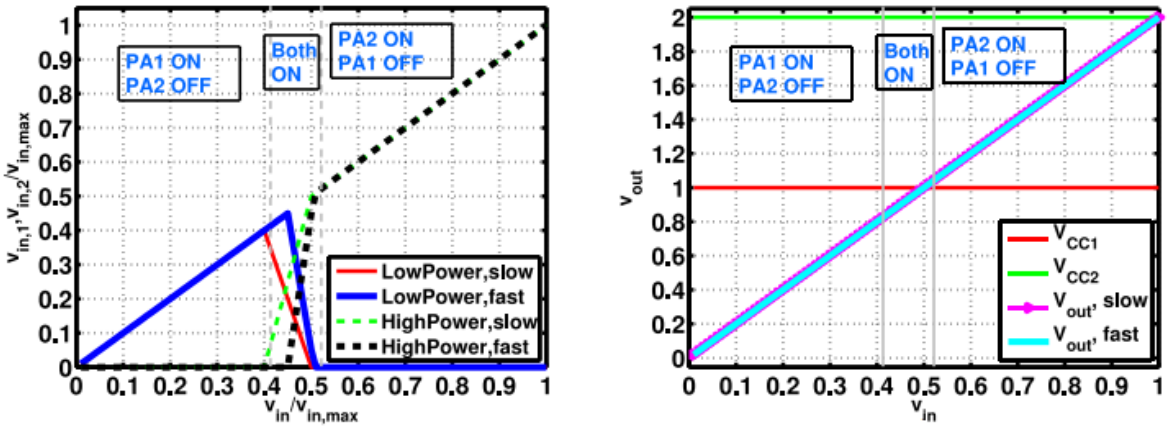


Figure 3.3: **Conceptual working principle of Supply Interpolating Transmitter**

power consumption in the "Peak" branch. Close to the threshold, PA_1 starts to turn off, and PA_2 starts to deliver more output power. This transition can be performed in two ways. PA_1 and PA_2 can be switched on and off rapidly, or this can be performed steadily. During the transition phase, both of the PAs are on, however, the power delivered to the output from each branch is different. This transition needs to be performed with care; otherwise, the linearity will be affected. Figure 3.3 portrays the operation of such a transmitter with respect to the input and the output amplitudes. For an ideal Class-B operation (all even harmonics shorted) along with smooth transition between the low and the high power region, a perfectly linear transfer can be achieved. The transients that are present at the output need to be reduced in order to ensure linearity and to be able to meet the EVM requirement for 5G systems. The rapid switching between the branches leads to common-mode oscillations at the output. Due to the non-ideality of a real circuit, these common-mode oscillations can get converted to differential signals. As a result, the linearity of the system is compromised. A steady switching between the branches is essential as it helps to alleviate the problem of common mode oscillation. Such an operation should provide efficiency peaks at PBO similar to that of the Doherty technique. However, the operating principle is drastically different in Supply Interpolation. This novel operating technique does not depend on load modulation. Thus, an SI based PA will not be affected by the dynamically changing load conditions in a massive MIMO system. Figure 3.4 delineates the drain efficiency versus output power for an SI transmitter. Compared to the efficiency of a conventional Class-B PA, this technique provides much higher efficiencies at PBO. By using two or more different amplifier branches connected to different voltage sources, efficiency peaks at deep PBO can be achieved.

3.2.1. Technology : FDSOI

SOI stands for Silicon-On-Insulator. SOI MOSFETs consist of a film of crystalline silicon that is separated from the substrate by a layer of SiO_2. SOI devices are available in two models: PDSOI and FDSOI. PDSOI stands for Partially Depleted SOI and FDSOI stands for Fully Depleted SOI.

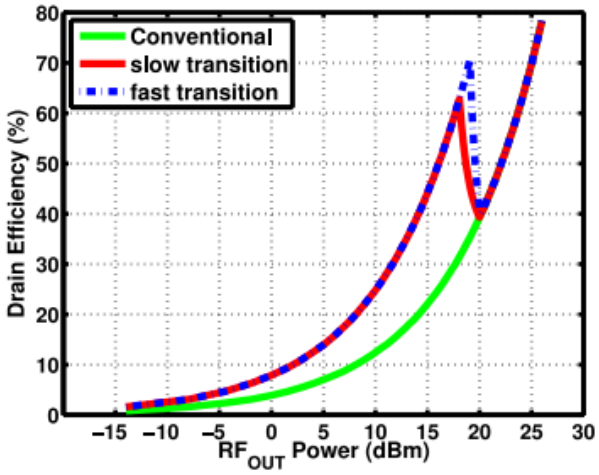


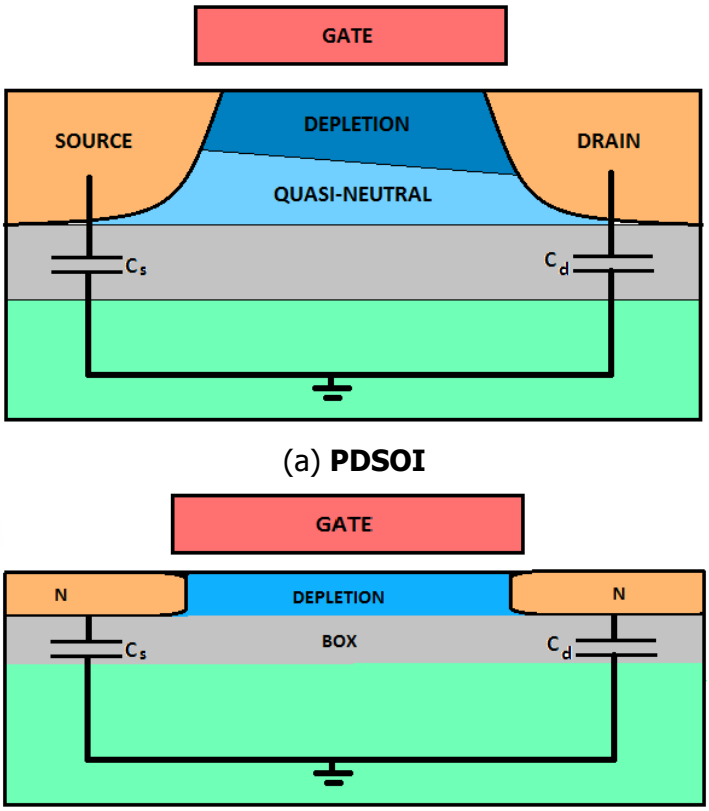
Figure 3.4: **Efficiency at PBO in Supply Interpolating Transmitter**

PDSOI has a thicker top silicon layer (channel) in the active region. This thickness is usually 50-90nm. Due to the presence of the thick top silicon, the channel/body can only be partially depleted of mobile carriers. The depletion region extends into the body but not all the way through to the BOX layer. The depletion region is wider near the drain than it is near the source. This can be explained by the fact that the drain is always at a much higher potential than the source. The channel electrons near the drain acquire very high energy due to this high electric field. Hence, electron-hole pairs are generated by impact ionization. The electrons get absorbed by the drain while the holes start to accumulate in the floating body. As a result, the potential of the body increases leading to a reduction in the threshold voltage. The reduction in the threshold voltage acts like a positive feedback which leads to a further increase in the drain current. Consequently, there is again an increase in the number of holes accumulating in the floating body. This process continues until the source-body junction is forward biased at which point the holes are no longer trapped and are released. Due to the reduction in the threshold voltage, the drain current abruptly increases leading to a discontinuity in the I_{ds}/V_{ds} plot [36] [37] [38]. This is called the ”kink effect”.

In the case of FDSOI, the top silicon layer is very thin, approximately 5-20nm [39]. It is usually $1/4$ the gate length. As the silicon is so thin, the channel is fully depleted of mobile charges. This leads to the superior behavior of FDSOI. Due to the presence of a fully depleted channel, FDSOI does not suffer from the ”floating body effect”. As a result of the fully depleted nature of the body, the potential barrier to holes at the source end is small. So, there is very little accumulation of holes and, as such, ”kink” cannot appear [40].

In PDSOI, the BOX layer is very thick ($\approx 100 - 200nm$) but, the BOX is only about $5 - 50nm$ in thickness in FDSOI [39]. Figure 3.5 illustrates the two different types of SOI.

The distinguishing feature between an NMOS in bulk CMOS and in a regular-well FDSOI process is the presence of the Buried Oxide (BOX) layer in FDSOI (Figure 3.6b).



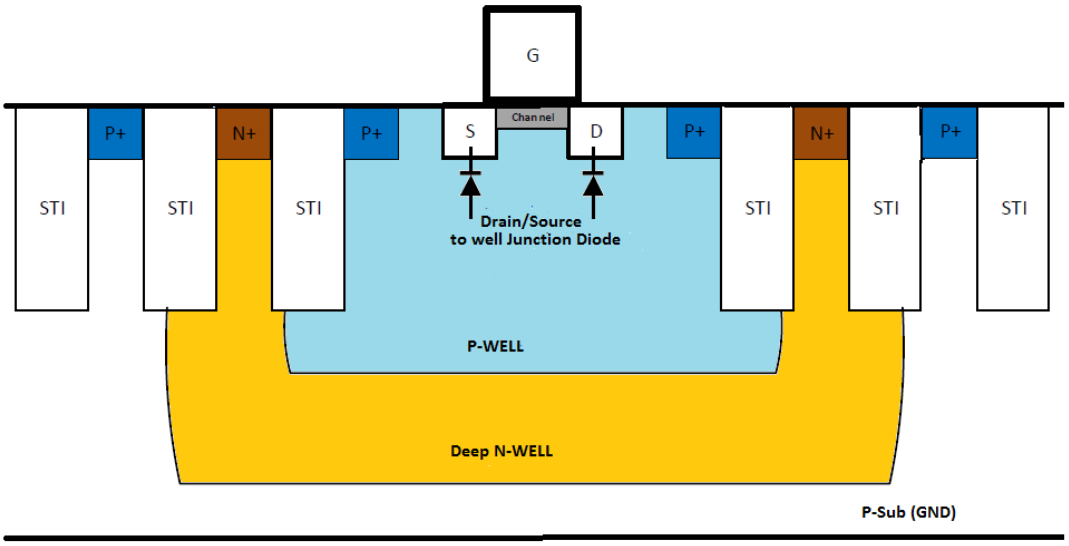
(a) **PDSOI**

(b) **FDSOI**

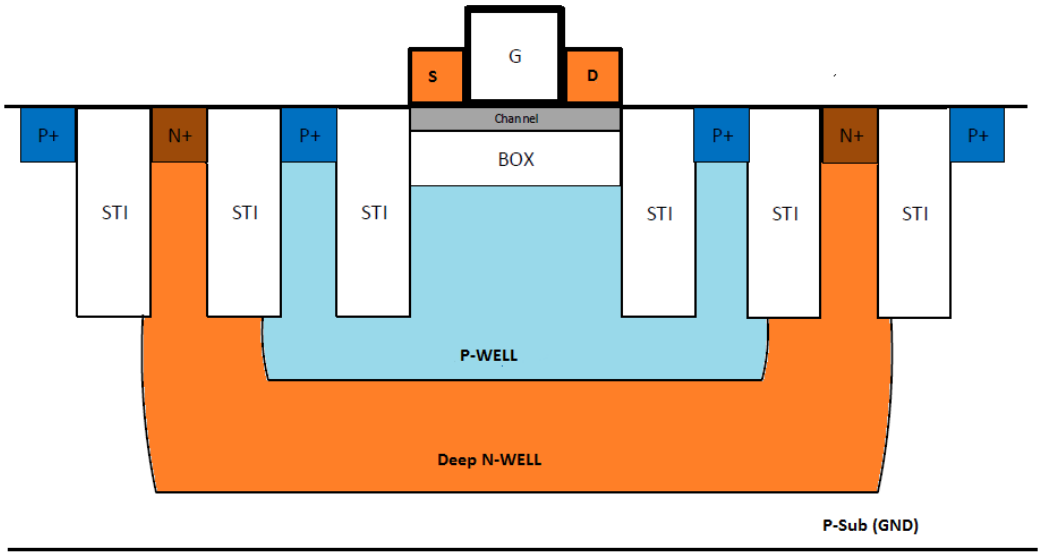
Figure 3.5: **Different types of SOI MOS transistor**

There are two different "Thin-Oxide" devices available in the PDK library. One model is called "Regular V_t" (RVT) and the other is "Low V_t" (LVT). We also have a "Super Low V_t" (SLVT) model. RVT is the conventional MOS, so it has a p-type substrate for NMOS. However, both the LVT and SLVT models have their wells flipped. This means that an NMOS is placed in an N-well and a PMOS in a P-well. The V_t is increased by using a lightly p-doped channel implant for an NMOS to create a "Low V_t" (LVT) and a lightly n-doped channel implant is used for a PMOS to create a "Low V_t" (LVT) variant. SLVT models do not have such channel implants [21]. Of course, such channel implants can also be used in the conventional MOS devices, to generate RVT and HVT (High V_t) variants. Figure 3.7 shows the MOS devices with a flipped well. "STI" and "DTI" stands for Shallow Trench Isolation and Deep Trench Isolation, respectively. Isolation is needed to prevent latchup. Trench isolation also enables higher packing density, i.e., the density is not limited by latchup. Without isolation, the n+ and p+ regions would need to be separated by longer distances.

The $I_d - V_{gs}$ plot of the three different variants of "Thin-Oxide" NMOS is illustrated in Figure 3.8. The effect of Body Biasing (BB) can be endorsed by these plots. The blue plot is when no BB is used. As soon as we forward bias the body, the drain current increases. The reverse effect is seen with reverse biasing the body. The threshold voltage can be seen to vary by at least 100mV under BB. Figure 3.9 portrays the $I_d - V_{ds}$ curves of the three different NMOS devices for different gate-to-source voltages (V_{gs}).



(a) **NMOS in conventional Bulk Process**



(b) **NMOS in regular-well FDSOI Process**

Figure 3.6: **NMOS in Bulk and FDSOI Process**

However, for the sake of brevity, not all of the different BB configurations are depicted in Figure 3.9. The BB voltage can be as high as $\pm 2V$. In this case, $\pm 1.5V$ has been used. This is much higher than what can be used in bulk CMOS. In bulk CMOS, the BB is limited by the parasitic diodes present between the n+ source and the p-substrate in NMOS and between the p+ source and the n-well in PMOS (Figure 3.6a). FBB voltage is limited to 300mV in bulk CMOS because of these parasitic diodes [21]. In FDSOI, such junctions do not exist and, therefore, there are no parasitic diodes. This enables the use of a much higher voltage in BB. Moreover, since the channel is fully depleted, the effect of BB is enhanced. The threshold voltage can vary by 70mV/V in FDSOI compared to a mere 25mV/V in bulk CMOS [21]. The performance of LVT

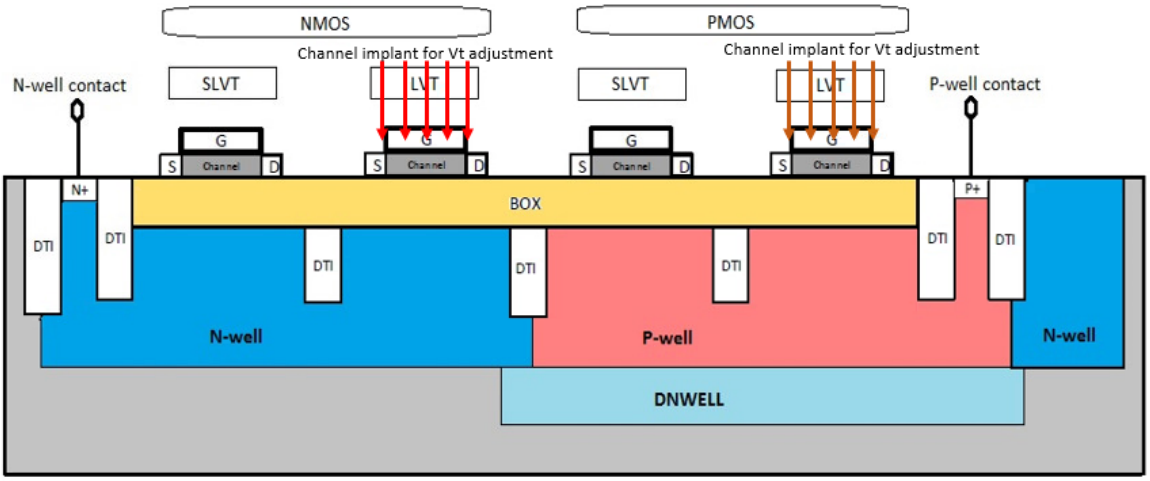


Figure 3.7: **NMOS and PMOS with Flipped Well**

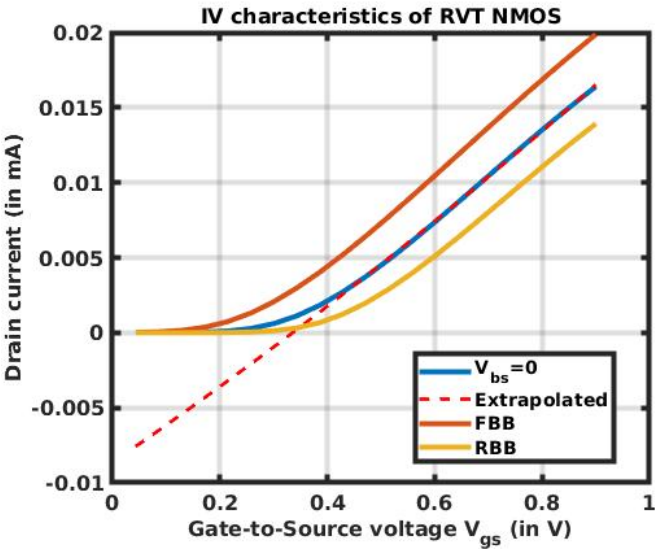
is slightly inferior to that of RVT. This can be explained by the presence of channel dopants in LVT as explained previously. SLVT in the absence of such dopants perform extremely well under FBB and gives superior performance than RVT.

FBB enables high performance at the cost of power. Due to the low V_t of these devices, there is a lot of power consumption in the standby mode. Thus, if one desires to minimize the standby power, then FBB is not an option. Leakage power (standby power) can be reduced by using RBB. Therefore, a trade-off exists between power and performance, which is universally known.

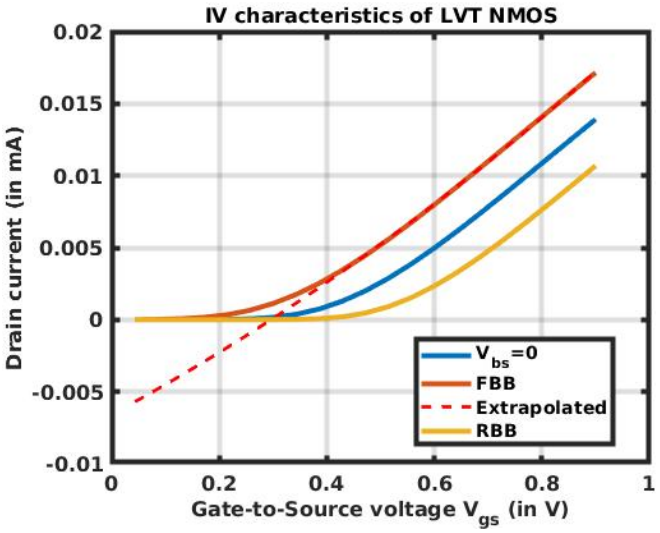
In this design, we wanted to minimize the power consumption at all costs, hence, the conventional well (RVT) models with 0 BB voltage have been used. Table 3.1 shows the different device operating points. As can be seen, the g_m of both SLVT and RVT devices are high under FBB. The threshold voltage V_{th} has been estimated from the I_{drain} versus V_{GS} plots shown in Figure 3.8. The extrapolated line has only been shown for the normal usage scenario of each device (i.e., FBB for SLVT and LVT and 0 BB for RVT). The V_{GS} has been kept at 0.5V for simulating the g_m and the g_{DS} of the devices. Similar simulations can be done with the other devices present in the

Parameter	Device Name	No BB	FBB	RBB
g_m	SLVT	34.16m	38.46m	25.17m
	LVT	23.16m	29.96m	11.56m
	RVT	30.72m	35.94m	24.58m
V_{th}	SLVT	303.25m	216.46m	379.2m
	LVT	368.35m	292.4m	422.6m
	RVT	335.8m	270.7m	379.2m
g_{DS}	SLVT	3.161m	3.825m	2.206m
	LVT	2.024m	2.742m	971.7u
	RVT	2.948m	3.549m	2.308m

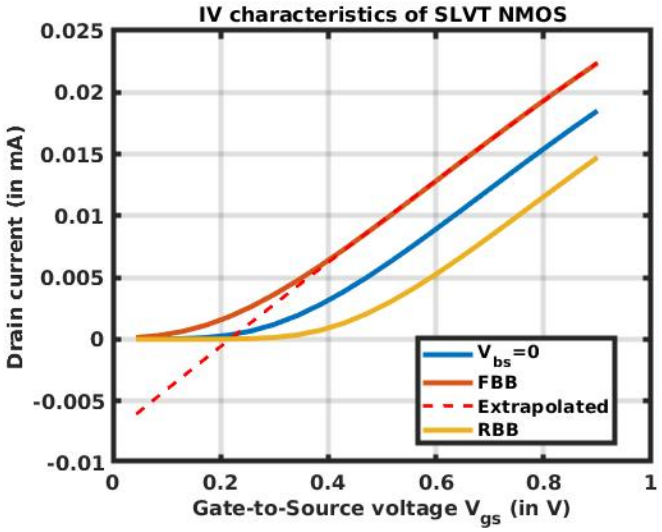
Table 3.1: Parameters for different NMOS devices $V_{DS} = 0.8V$



(a) **IV plot of Regular V_t NMOS (RVT)**



(b) **IV plot of Low V_t NMOS (LVT)**

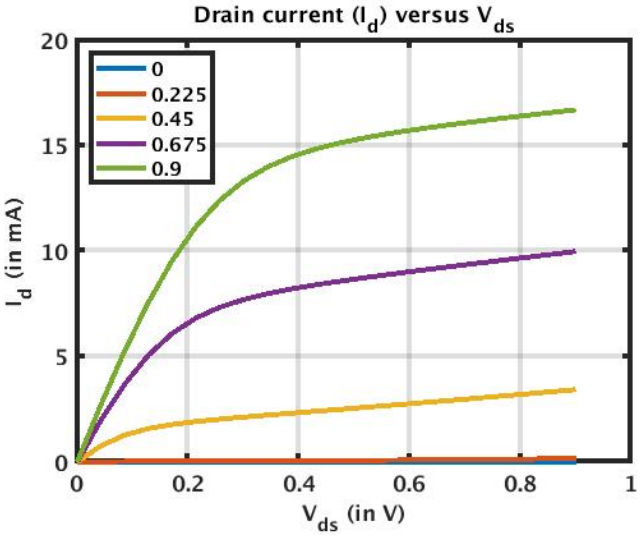


(c) **IV plot of Super Low V_t NMOS (SLVT)**

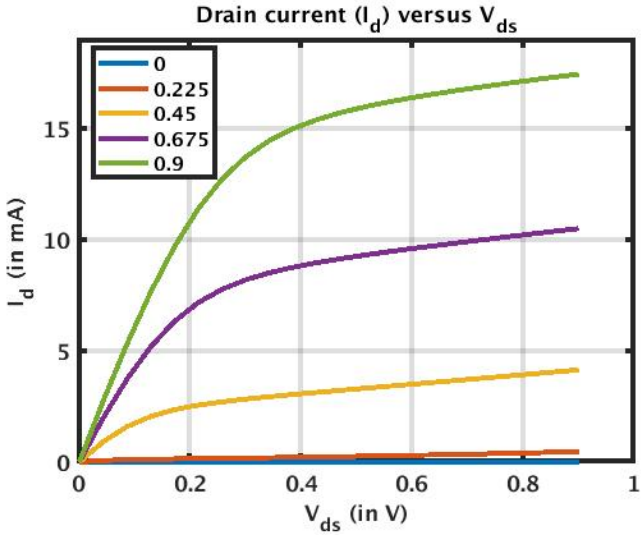
Figure 3.8: I_d-V_{gs} **plots of the different Thin-Oxide NMOS devices**

library, however, for the sake of brevity, they are not shown here. Table 3.2 provides an overview of the advantages and disadvantages of CMOS, FDSOI, and PDSOI.

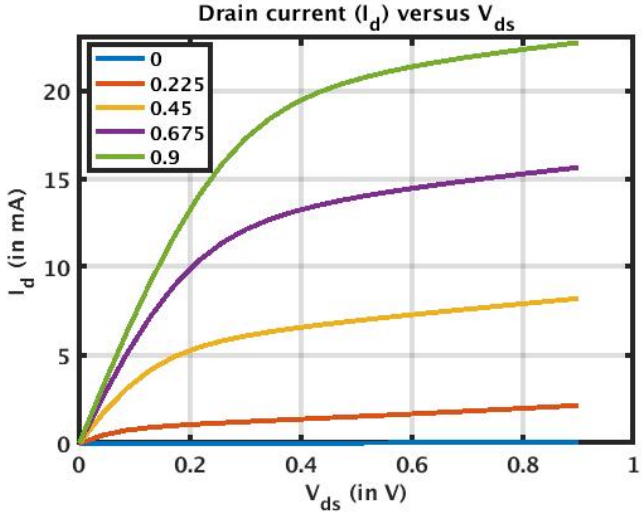
This section, though a bit of a stray from our discussion in the last section, is essential in order to understand the characteristics and the nature of the devices available in the PDK. Based on this knowledge, we can make an informed decision about which devices to be used in our design. Due to the high g_m and the low leakage characteristics of RVT, they are a perfect choice for the CS input stage of our PA. For the cascode device, we are going to use a ”Medium-Oxide” NMOS (EGV Low-V_t) model. The EGV model has a channel length of 100nm and can withstand much higher voltages than the RVT model. This is essential for generating high power and, at the same time, maintaining reliability.



(a) **Regular V_t NMOS (RVT) with no BB**



(b) **Low V_t NMOS (LVT) with FBB**



(c) **Super Low V_t NMOS (SLVT) with FBB**

Figure 3.9: **I_d-V_{ds} plots of the different Thin-Oxide NMOS devices**

We return back to the implementation details of the SI transmitter. Figure 3.10 shows the analog intensive supply-interpolating transmitter. The Digital Baseband Processing block generates the complex I-Q signal from the digital bitstream. Since, the signal bandwidth for 5G systems is in the order of 1GHz, a sampling frequency (f_s) of at least 2GHz is required. This clock can also be generated by the DSP block. A Digital-to-Analog Converter (DAC) is required to convert this digital information into analog signals. This is followed by a Low-Pass filter, or a Reconstruction filter as it is usually called. The task of this filter is to attenuate the high frequency aliases that occur at the harmonics of f_s. This filter smoothens out the stair-case like output of the DAC to generate a smooth continuous signal. Such a filter may also include a correction for the "Hold" operation that affects the amplitude of the higher frequencies. This is followed by a mixer that translates the signal to 30GHz. Since IQ modulation has

Technology	Pros	Cons
CMOS	• Fabrication is relatively easy. • Heat can easily dissipate into the substrate due to the absence of an insulating material.	• The latch-up effect is detrimental to circuit performance. • The presence of source-body and drain-body junctions lead to higher junction capacitance. These junctions also lead to the presence of parasitic diodes. These parasitic diodes limit the voltage range for forward and reverse body bias. • Limited variation in V_t is possible with forward body bias.
FDSOI	• "Floating body" effect is absent. • No "kink effect". • Random fluctuations of V_t are absent. • Heat dissipation is much better than PDSOI as the BOX layer is much thinner. • No latch-up effect. • The threshold voltage can be varied by about 100mV for every 1V of body bias [21]. • SOI devices are more suitable for high temperature operation as the leakage current is drastically lower than bulk CMOS [37].	• Challenging to manufacture very thin body. • Higher fabrication costs. • Relatively difficult to control the thickness of the ultra-thin body. • Threshold voltage sensitive to the thickness of the silicon film [38].
PDSOI	• Easier to manufacture than FDSOI. • "Floating body" effect can be used to enhance the performance • Heat cannot easily dissipate due to the presence of a thicker BOX. This drastically elevates the device temperature which leads to a reduction in the mobility of the electrons. • Threshold voltage less sensitive to the uniformity of the silicon film thickness [38]. • No latch-up effect.	• Presence of "kink effect". • Shows history effect which is responsible for different switching rates on subsequent switching edges. • The addition of body tie to reduce "kink effect" increases the cost of the die. • Body tie contacts also introduce charges into the body. This affects the speed of operation as some time is required to discharge these charges.

Table 3.2: Comparison between the different technologies

been used, a quadrature LO signal is necessary. How such a signal can be generated is illustrated later. After upconversion, the signal is fed to a PA. The control signal

determines the appropriate PA branch to be operated.

As an example, Figure 3.11a depicts an SI based on digitally controlled PA (DPA). The DPA is turned on/off by the digital bitstream. Due to the switched mode PA operation, the DPA is highly energy efficient. For example, if a 10-bit ($2^{10} = 1024$) resolution is desired, then two separate bitstreams of 9-bit ($2^9 = 512$) each can be generated. When the MSB is high, then the DPA connected to V_{DD2} is turned on or else only the DPA connected to V_{DD1} is in operation. The output matching network ensures that we are operating in Class-E mode. Figure 3.11b portrays how the two branches are used. In the low-power regime, the amplifier connected to V_{DD1} is on. As the Amplitude Command Word (ACW) increases, the amplifier connected to V_{DD2} starts to turn on, and the other amplifier slowly turns off. This has been elucidated by plotting the DC power consumption of each branch (Figure 3.11b). Efficiency and P_{out} versus ACW is illustrated in Figures 3.12a & Figure 3.12b. It is clear that this technique can deliver very high efficiencies at PBO. All of this can be achieved even when handling signals with large RFBW. By refining the cross-over from one branch to another, a more linear transfer can be obtained.

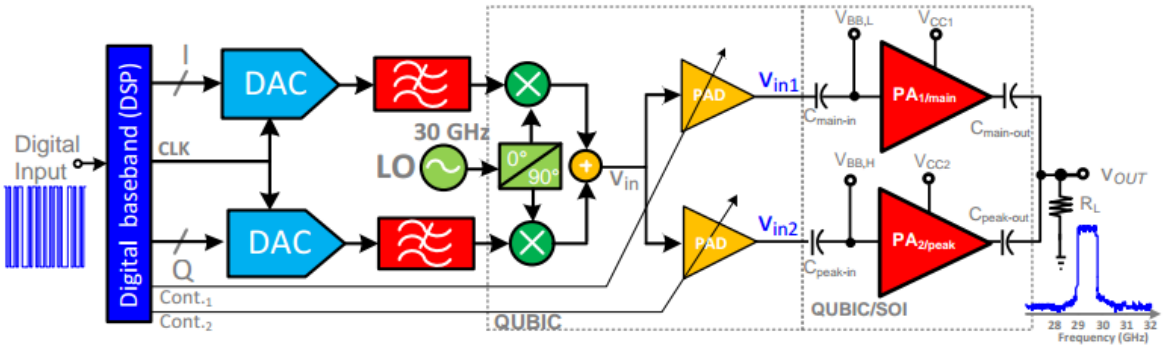


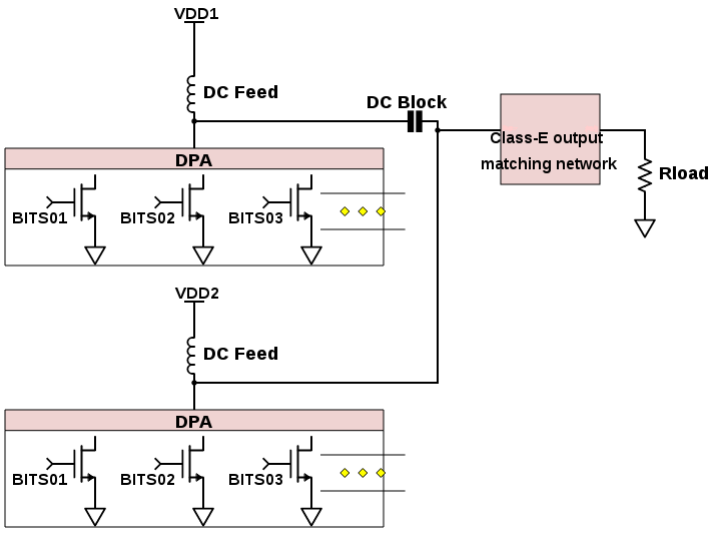
Figure 3.10: **Analog Intensive Supply Interpolating Transmitter**

However, unfortunately, the efficiency depends on the on-resistance (R_{on}) of the devices. Hence, one needs wider devices to ensure low R_{on}. A simple simulation can be used to demonstrate this. Here, the MOS devices have been replaced with ideal switches and the on-resistance R_{on} of the switch has been swept. An ideal pulse of 30GHz is used for this simulation with 0.1psec rise and fall time. This is done to show the effect of R_{on} only. A plot of efficiency versus R_{on} is plotted in Figure 3.13a, but this efficiency will be even lower in the real scenario as we cannot possibly generate a pulse at mm-wave frequencies with ideal rising and falling edges. Consequently, if the pulse excitation is replaced with a sinusoid in the same schematic, then the efficiency drops drastically. This can be observed in Figure 3.13b.

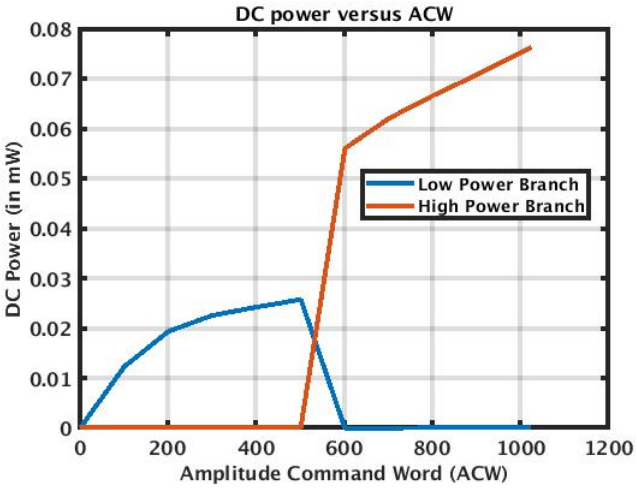
However, we cannot meet the linearity (EVM) criterion with a Class-E based PA as switched mode PAs are very non-linear. Thus, we are going to implement a more linear PA using a Class-AB mode of operation.

3.3. Basic Architecture

Polar power amplifiers are very popular and widely used. Let us take a look at them and see how they fare under the strict requirements of 5G. In a Polar architecture,

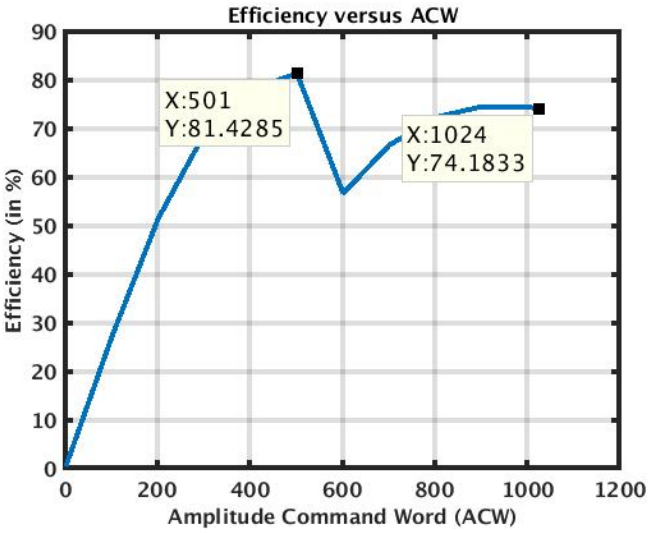


(a) **Simple Example of Supply Interpolation**

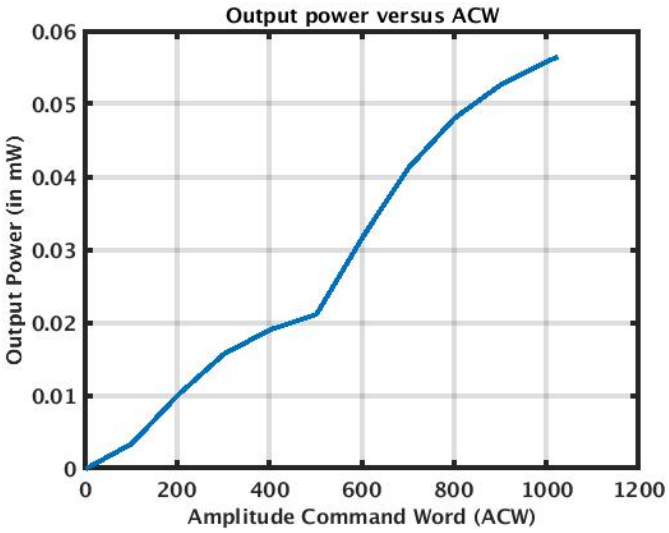


(b) **DC Power consumption versus ACW**

Figure 3.11: **DPA based SI Transmitter along with its control profile**



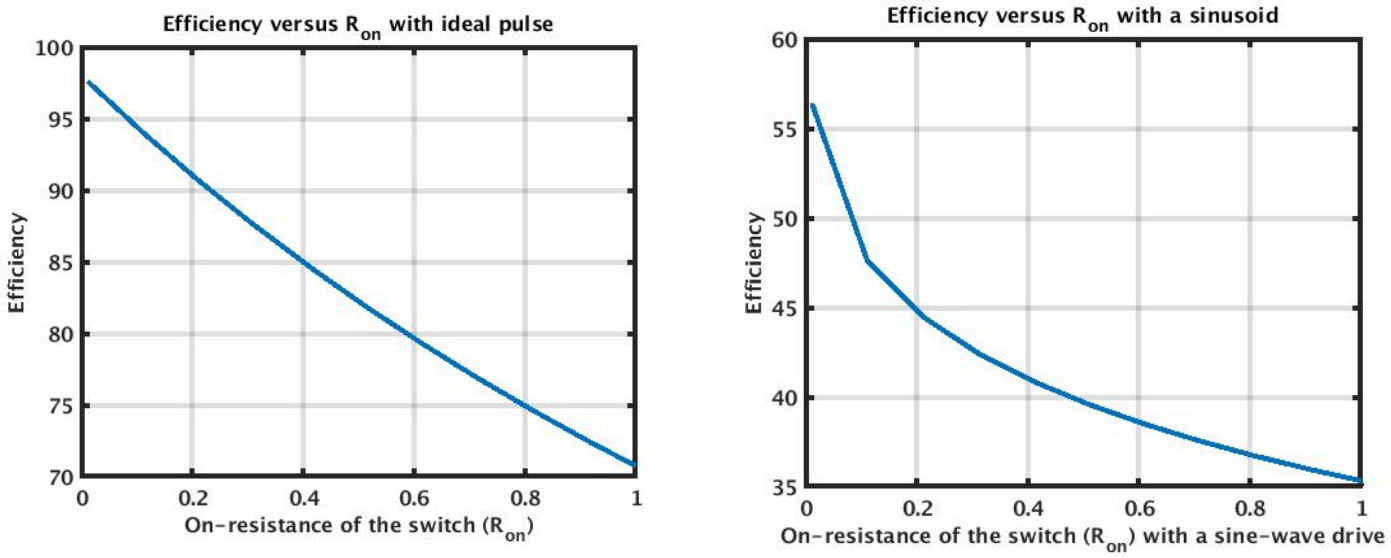
(a) **Efficiency versus ACW**



(b) **Output Power versus ACW**

Figure 3.12: **DPA based SI Transmitter along with its Efficiency and P_{out}**

the amplitude information and the phase information of the baseband is applied separately. The amplitude information can be used to modulate the supply voltage by Envelope Tracking (ET) or with the help of a switched-mode power supply (SMPS). Both the AM and PM signals can be merged in a saturated PA driver followed by a linear (less efficient) PA [41]. Figure 3.14a shows such a design. Here, the amplitude information from the baseband is provided by the DAC directly into the source of the PA. The phase information is encoded into the carrier.



(a) **Efficiency versus** R_{on}

(b) **Efficiency versus** R_{on} **with a sinusoid**

Figure 3.13: **Effect of** R_{on} **and the type of excitation on the Efficiency of a switched-mode amplifier**

An IQ signal can be expressed in polar form by r & θ, where r is the envelope and $theta$ is the phase. In terms of the rectangular coordinate plane, r is the distance of the point from the origin, and θ is the angle in radian, from the positive x-axis to the line connecting origin to the point. Equation 3.1 shows this relationship.

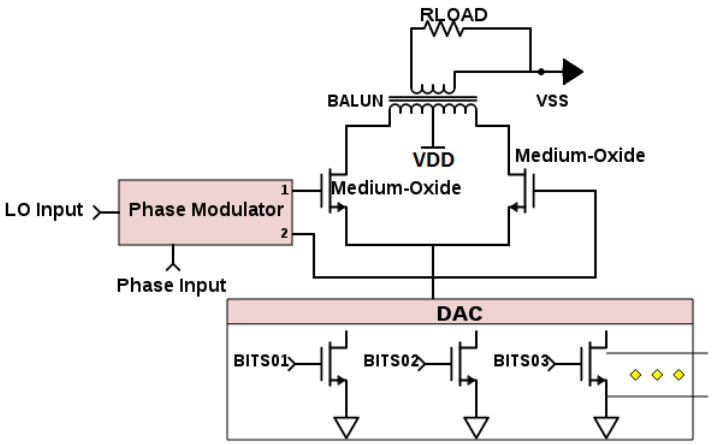
$$IQSignal = I + j * Q = r\angle\theta \tag{3.1}$$

The instantaneous frequency of modulation can be expressed as [41] :-

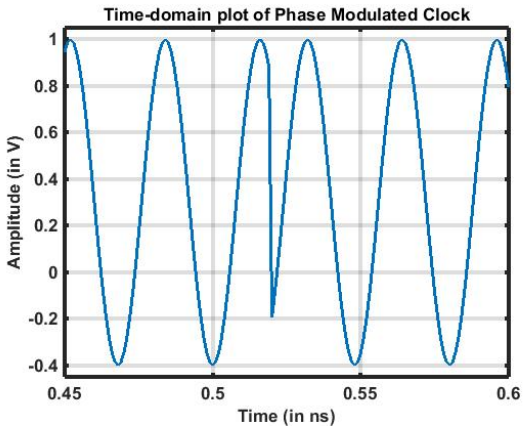
$$\frac{d\theta}{dt} = \frac{d}{dt}(\arctan\frac{Q(t)}{I(t)}) = \frac{I.dQ/dt + Q.dI/dt}{I^2 + Q^2} \tag{3.2}$$

From Equation 3.2, it is clear that there is a significant frequency drift from the carrier frequency when I and Q signals pass near the origin. For advanced modulation schemes like 64-QAM OFDM, the PAPR is almost 9-10dB, and hence there will be a large number of trajectories near the origin [41]. This leads to large bandwidth expansion which degrades the in-band EVM performance and, at the same time, produces out-of-band spectral regrowth when applied to a nonlinear PA [41]. Figure 3.14b and Figure 3.15a illustrate the time-domain and the frequency-domain plot of the phase modulated carrier signal, respectively. A 2-tone signal with a RFBW of about 980MHz was provided and, in Figure 3.15b, the third order distortion products can be seen. An IM3 of only 9dB is obtained. Also, since the AM and PM signals pass through different circuit blocks, they experience different delays. This differential delay leads to Intermodulation Distortion (IMD). The analysis in [42] shows that, for a differential delay of τ, an IMD voltage proportional to τ^2 is generated.

The analysis above establishes that an IQ transmitter is the only option for 5G systems. In this work, the IQ transmitter converts the baseband signal to 30GHz

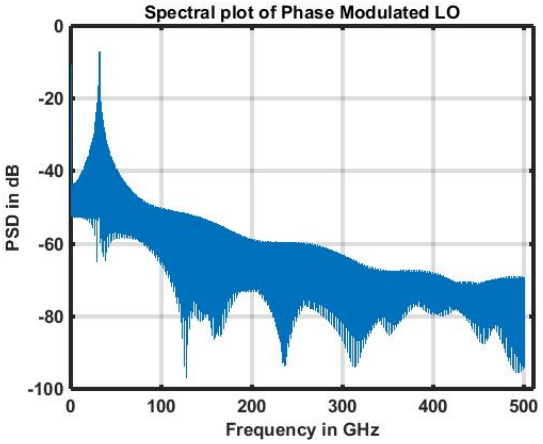


(a) **DAC based Polar Power Amplifier**

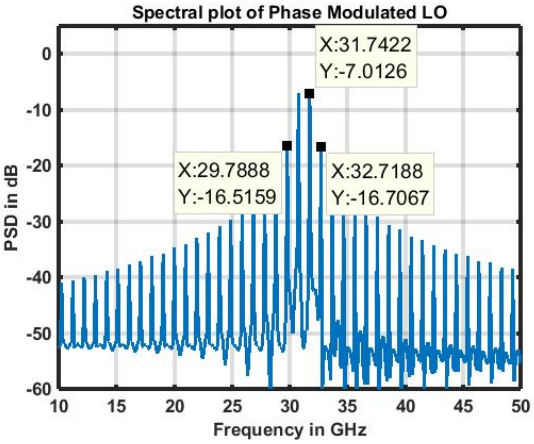


(b) **Time-domain plot of a Phase Modulated Clock**

Figure 3.14: **Polar Amplifier along with the Phase Modulated Clock**



(a) **Spectral plot of a Phase Modulated Clock**



(b) **Zoomed in Spectral plot of a Phase Modulated Clock**

Figure 3.15: **Phase Modulated Carrier Signal**

before feeding it into the PA. The differential baseband signal and a single ended carrier signal comes from off-chip. The single ended carrier signal is converted to a differential carrier with the help of a Balun. Then, this differential carrier signal is passed through a Poly-Phase Filter (PPF) to generate the quadrature signals (I and Q). The differential baseband IQ signal comes directly from a signal generator.

3.3.1. Passive Mixer

A mixer is used for signal upconversion. Apart from upconverting the baseband signal to the carrier frequency, a mixer must also provide a sufficient signal swing for the subsequent stages (e.g., PA). The conversion gain as well as the power consumption of a mixer are two important design parameters. The mixer should also be able to drive the input capacitance of the following stage, which is typically a PA in a

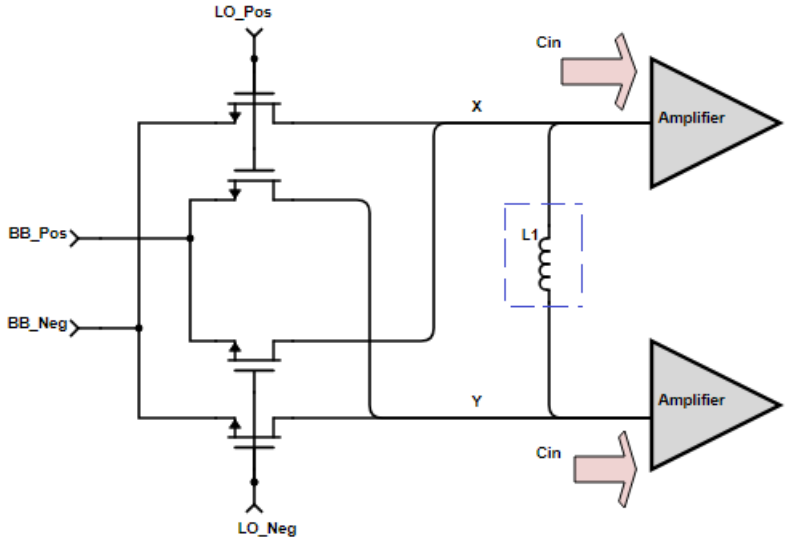


Figure 3.16: **Double Balanced Passive Mixer**

transmitter. Lastly, since a transmitter needs to be linear, the mixer output should also be linear. The main function of a PA is to deliver high power to a load. However, such an operation should not affect the power budget of the transmitter. Hence, a PA needs to be energy efficient. Apart from providing high output power, a PA is also responsible for transferring the modulation from the input to the output. This implies that a PA needs to be linear in its operation. Regrettably, a highly efficient PA is non linear thus the mixing operation needs to be linear so as not to put the extra burden of linearity on the PA.

In order to relax the design of such a mixer, buffers can be used in between the mixer and the PA, but these extra buffer stages lead to non linearity [43]. Thus, in this work no additional stage has been used in between the mixer and the PA.

A passive mixer has been used for the upconversion in this thesis. This has been done to decrease the DC power consumption in the mixing operation. Also, passive mixers are highly linear and are thus suitable for upconversion [43]. The conversion gain of the passive double balanced mixer is $\frac{2}{\pi}$ [43]. However, this is only valid when using an "ideal" LO waveform, i.e., a square wave. When operating at very high frequencies, this is not possible. As a result, during the transition period (when the LO signals change polarities), all of the transistors are on, and the input signal is treated as a common mode input. This leads to a decrease in the conversion gain [43]. This issue can be minimized by using an LO signal with a very large amplitude. However, this is also difficult to achieve at such high frequencies.

A few things need to be considered while designing a passive mixer. The on-resistance (R_{on}) of the switch is important. This is because the R_{on} along with the output capacitance of the switches and the input capacitance (C_{in}) of the following stage (or amplifier) forms a low pass filter that can limit the bandwidth at nodes X and Y [Figure 3.16]. 5G systems will be dealing with huge RFBWs so proper care

needs to be taken during the design phase to ensure that the parasitic capacitance and resistance at nodes X and Y are within permissible limits. It may not be possible to tune/reduce C_{in} beyond a certain limit as the amplifier needs to deliver a certain amount of power. In such a situation, an inductor can be used to resonate out the input capacitance of the PA and the capacitance present at the nodes X and Y at the frequency of operation. Using Equation 3.3, we can find the suitable value of the inductor. The factor 2 is because we are using the capacitances considering a single ended operation as $C_{X,Y}$ is the capacitance at node X **or** node Y. The input capacitance of the amplifier can be estimated with an S-parameter simulation. However, due to the finite quality factor (Q) of the passives, the mixer will see the parallel equivalent resistance (R_p) of the inductor at resonance. The value of this resistance is given by Equation 3.4. This needs to be kept in mind as a similar situation will be encountered in this work during the design of the Polyphase filter in Section 3.5.

$$L1 = \frac{2}{(2\pi f_{LO})^2 (C_{X,Y} + C_{in})} \tag{3.3}$$

$$R_p = Q * L1 * 2\pi f_{LO} \tag{3.4}$$

Another option is to reduce the value of R_{on} by using wider switches. This will provide favorable results until the point where the input capacitance of the switches start loading the LO ports. As has been elaborated later on, the input capacitance of the mixers also need to be within a certain limit. This is because the stages before the mixer (in this case, a Polyphase filter) has to deliver enough power to drive the input capacitance of the mixer. As a result, it is convenient to have a mixer with smaller input capacitance. Also, wider switches add to the capacitance at nodes X and Y [Figure 3.16].

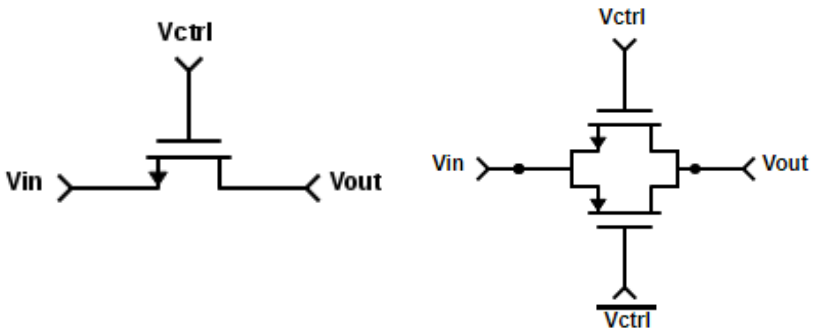
Apart from bandwidth limitations, issues like carrier feedthrough and offset in baseband signals need to be taken care of. The carrier feedthrough is equal to $\frac{2V_{offset}}{V_{amp}}$ [43], where V_{offset} and V_{amp} is the baseband offset voltage and amplitude of the baseband signal, respectively.

3.3.2. Pass Gate

Although it is presumed that the reader is well aware of the advantages of using a pass gate rather than a single MOS switch, a short description has been provided nonetheless for the sake of completeness. Figure 3.17 shows two voltage controlled switches. In Figure 3.17a, a control signal is used to turn on/off the NMOS in order to control the flow of the signal from the input port to the output port. However, this simple structure has a serious drawback. The control signal by itself is not self-sufficient. This means that the state of the switch cannot be determined by the control signal itself. The switch can be in the on or off state for the same value of the control signal depending on whether the source voltage is high or low. The gate-source voltage thus determines the state of the switch. Therefore, even if the control signal is high, the switch will be in the off state if the source voltage is high as then the V_{gs} of the device is below the threshold voltage. This simple structure can only pass signals that are one threshold voltage below V_{DD}. Also, the threshold voltage fluctuates due to the body effect. So, the source and the body terminals need to be tied together

in bulk CMOS. In a FDSOI technology, FBB can be employed to reduce the threshold voltage of the device. As a result, FBB has been used in this work to demonstrate its importance in enhancing the performance of a system.

On the other hand, a pass gate is more versatile. The NMOS can pass signals when the PMOS is not working and the PMOS is on when the NMOS is in an off state. Thus, in this way, both high and low signals can pass through. When the source signal is low, then the high control signal enables the NMOS, and the signal passes through. The PMOS is in the off state during this operation as it is driven by the complementary clock, therefore, its V_{sg} is low. Now, when the source signal is high, then NMOS is off as its V_{gs} is below the threshold but the PMOS is on as its V_{sg} is high. Consequently, the addition of a PMOS takes care of all of the deficiencies of the single MOS switch.



(a) **Single MOS switch** (b) **Pass Gate**

Figure 3.17: **Voltage controlled switch**

In this design a pass-gate based passive mixer has been used for the reasons stated above. Figure 3.18 illustrates the basic pass gate based mixer schematic and the mixers in both of the I and Q paths. It is similar to the passive mixer depicted in Figure 3.16; the only difference is that, here, we have a pass-gate based mixer. The I and the Q paths are combined at the output of the mixer. This is also beneficial as, otherwise, we would need a separate PA for the I and the Q paths and then combine them at the "Drain" of the PA [44] (although it is a digital approach, but it is based on the same idea). This method ensures that we would not need extra PAs, which would have made the output matching more complicated because of the large output capacitances associated with these PAs. Also, from the matching point of view, this technique is superior as the mismatch between the I and the Q paths is reduced.

Figure 3.19a and Figure 3.19b demonstrate the advantage of using IQ modulation for signals with large bandwidth, i.e., for 5G systems. The spectral plot is free of the distortion products that can be seen in Figure 3.15b.

3.3.3. Two Port Network

In order to design a microwave amplifier, a little bit of insight is necessary. A system needs to be completely characterized by parameters that can be measured at its terminals (ports). This is extremely necessary and only then can we gain insight into the true workings of a network. Different parameters like Impedance parameters (Z-parameters), Admittance parameters (Y-parameters), and scattering parameters

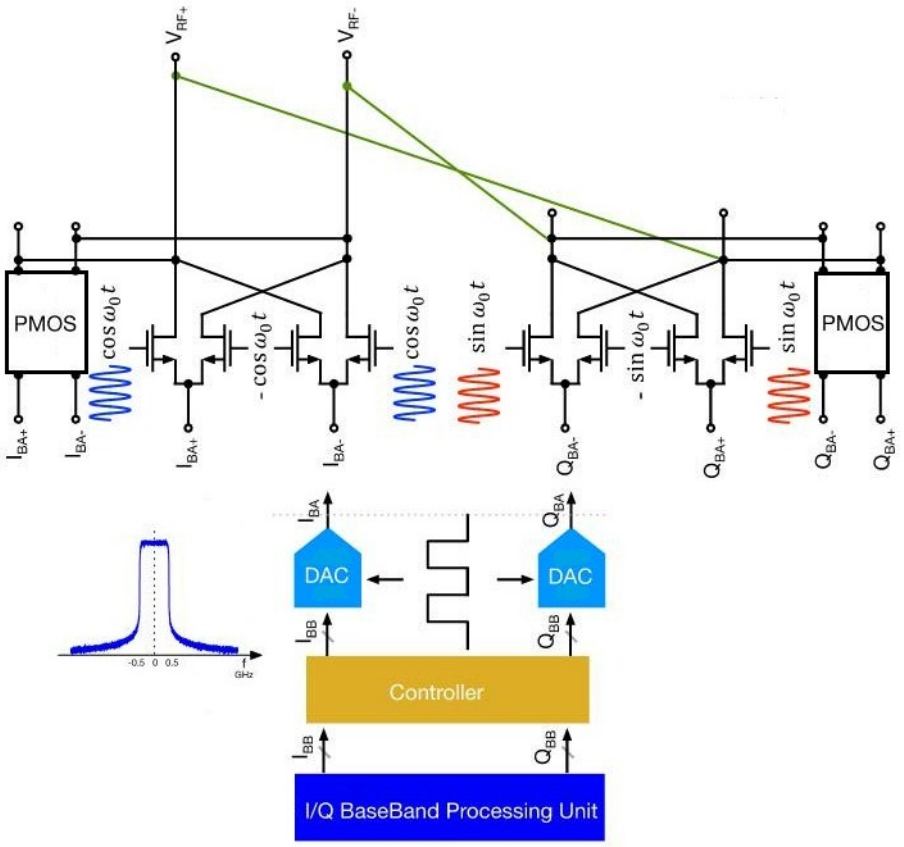


Figure 3.18: **Passive Mixer**

(S-parameters) are available and widely used. Figure 3.20 shows a two port network. The Norton equivalent circuit of a two port network is also illustrated.

Using Y-parameters, we can derive the input and output admittance of the 2-port network. The input admittance Y_{in} and the output admittance Y_{out} can be represented by [24]

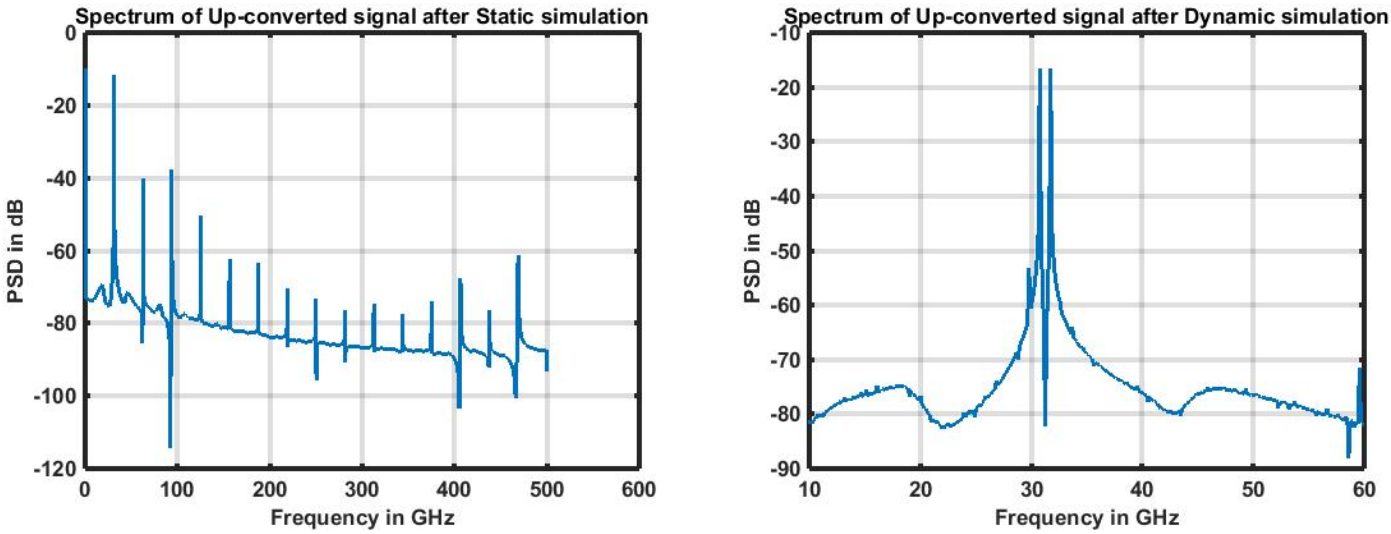
$$\begin{aligned} Y_{in} &= Y_{11} - \frac{Y_{12}Y_{21}}{Y_L + Y_{22}} \\ Y_{out} &= Y_{22} - \frac{Y_{12}Y_{21}}{Y_S + Y_{11}} \end{aligned} \tag{3.5}$$

For a unilateral network, i.e., for $Y_{12} = 0$, Eqn. 3.5 reduces to the form $Y_{in} = Y_{11}$ and $Y_{out} = Y_{22}$. Using Figure 3.20 different powers and gains associated with a two port network can be calculated. The different powers can be calculated easily using Y-parameters [24]. The transducer power gain (G_T) can be expressed as :-

$$G_T = \frac{Power\ delivered\ to\ the\ load}{Power\ available\ from\ the\ source} = \frac{P_L}{P_{AVS}} = \frac{\frac{|V_2|^2}{2}Re(Y_L)}{\frac{|I_S|^2}{8Re(Y_S)}} \tag{3.6}$$

Another gain term often used is operating power gain (G_P). This is defined by

$$G_P = \frac{Power\ delivered\ to\ the\ load}{Power\ input\ to\ the\ network} = \frac{P_L}{P_{IN}} = \frac{|V_2|^2 Re(Y_L)}{|V_1|^2 Re(Y_{in})} \tag{3.7}$$



(a) **Spectral plot of up converted IQ signal** (b) **Zoomed in Spectral plot**

Figure 3.19: **Up converted IQ signal**

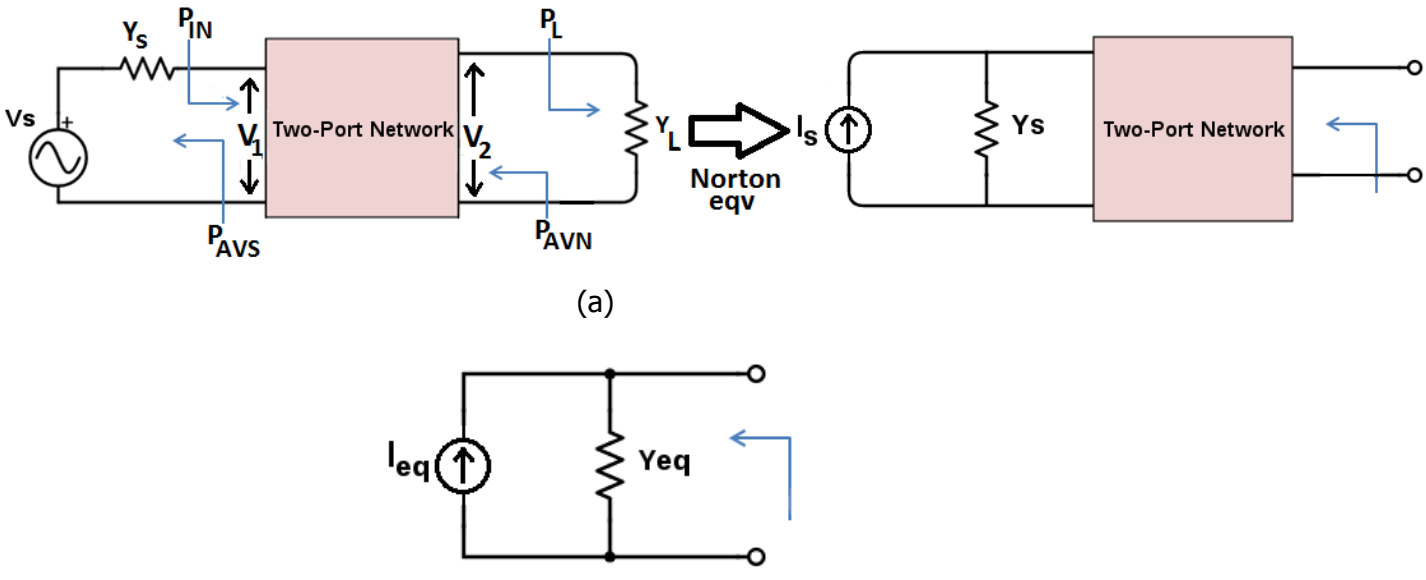


Figure 3.20: **Two-port network**

The available power gain (G_A) is defined as

$$G_A = \frac{Power\ available\ from\ the\ network}{Power\ available\ from\ the\ source} = \frac{P_{AVN}}{P_{AVS}} = \frac{|I_{eq}|^2 Re(Y_S)}{|I_S|^2 Re(Y_{eq})} \tag{3.8}$$

where $I_S = V_1(Y_{11} + Y_S)$ is the current delivered by the source (Figure 3.20a), and I_{eq} can be derived by using a Norton equivalent of the 2-port network (Figure 3.20b).

I_{eq} can be expressed as [24]:

$$\begin{aligned} I_{eq} &= I_2 = Y_{21}V_1 \\ I_{eq} &= \frac{Y_{21}}{Y_{11} + Y_S} I_S \end{aligned} \tag{3.9}$$

The operating power gain G_P, available power gain G_A can be expressed as [24]

$$\begin{aligned} G_P &= \frac{|Y_{21}|^2}{|Y_L + Y_{22}|^2} \frac{Re(Y_L)}{Re(Y_{in})} \\ G_A &= \frac{|Y_{21}|^2}{|Y_{11} + Y_S|^2} \frac{Re(Y_S)}{Re(Y_{eq})} \end{aligned} \tag{3.10}$$

where Y_{eq} is the Norton equivalent admittance.

The Norton equivalent admittance is equal to the output admittance of a 2-port network. So $Y_{eq} = Y_{22}$. We can also express G_T as [24]

$$\begin{aligned} G_T &= \frac{0.5Re(Y_L)|V_2|^2}{\frac{|I_S|^2}{8Re(Y_S)}} \\ G_T &= 4Re(Y_L)Re(Y_S)\left|\frac{V_2}{I_S}\right|^2 \end{aligned} \tag{3.11}$$

The ratio of V_2/I_S can be derived using the Y-parameters of the 2-port network. Finally, G_T can be expressed in terms of Y-parameters by [24]

$$G_T = \frac{4Re(Y_L)Re(Y_S)|Y_{21}|^2}{|(Y_S + Y_{11})(Y_L Y_{22}) - Y_{12}Y_{21}|^2} \tag{3.12}$$

Figure 3.21 shows the different gain versus frequency plot. For a unilateral system, $Y_{12} = 0$ and the gain from such a system is referred to as the unilateral power gain. Figure 3.21 shows the maximum unilateral power gain ($GUMX$). Since a real system is never unilateral, this is only of theoretical interest. H_{21} is a hybrid parameter that is used to determine the F_t of a transistor. As is shown in the plot, F_t is the point where H_{21} crosses the X-axis.

With the help of Figure 3.20, it can be stated that $P_L \leq P_{AVN}$ where the equality is only achieved for a matched load ($Z_L = Z_o$). Thus using Eqns. 3.6 and 3.8, it can be stated that $G_T \leq G_A$. The maximum G_T can only be achieved for a conjugate match, i.e., when the load is conjugately matched to the two-port output impedance. A conjugate match cancels the reactive component of the impedances, enabling maximum power transfer.

Using similar reasoning as above, it can be stated that $P_{IN} \leq P_{AVS}$ [Figure 3.20]. The above condition leads to equality only when the two-port is conjugately matched to the source. As a result, we can state that $G_T \leq G_P$. In order to maximize all of these three different gain terms, a bi-conjugate match is necessary. This means that both the input and the output ports are simultaneously conjugately matched.

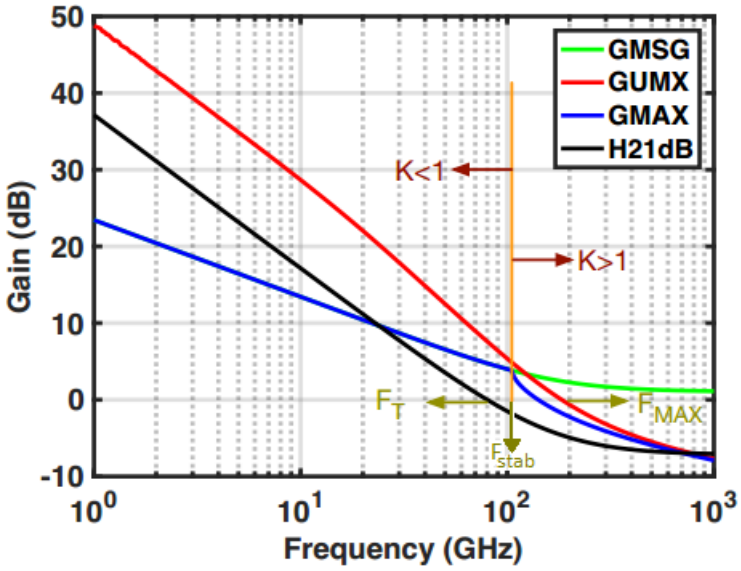


Figure 3.21: **Different Gains versus frequency**

Hence, optimum source and load impedance/admittance necessary to maximize all of the different gain terms is [24]

$$\begin{aligned} Y_{in} &= Y_{11} + \frac{Y_{12}Y_{21}}{1 - Y_{22}} = Y_S^* \\ Y_{out} &= Y_{22} + \frac{Y_{12}Y_{21}}{1 - Y_{11}} = Y_L^* \end{aligned} \tag{3.13}$$

Because the amplifier should never oscillate, one needs to ensure that the input and output port never present a negative resistance. The conditions $K > 1$ and $|\Delta| < 1$ are used to check for stability. These parameters are usually used in S-parameter form, hence Eqn.3.14 presents both the Y and S parameter representations of K and Δ [24] [45].

$$\begin{aligned} K &= \frac{2Re(Y_{11})Re(Y_{22}) - Re(Y_{12}Y_{21})}{|Y_{12}Y_{21}|} \\ &= \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}||S_{21}|} \\ |\Delta| &= |(1 + Z_O Y_{11})(1 + Z_O Y_{22}) - Z_O^2 Y_{12}Y_{21}| \\ &= |S_{11}S_{22} - S_{12}S_{21}| \end{aligned} \tag{3.14}$$

where Z_O is the characteristic impedance at each port.

Another way to test for stability is to use the μ Stability test. The advantage of this method is that only a single parameter needs to be evaluated. Another advantage is that the magnitude of μ directly provides a measure of the stability, e.g., a two-port with a larger μ is more stable.

Under a simultaneous conjugate match, G_T can be expressed as [24]:-

$$G_{T,max} = G_{MAX} = \frac{|Y_{21}|}{|Y_{12}|}(K - \sqrt{K^2 - 1}) \tag{3.15}$$

When $K = 1$, the value of $G_{T,max}$ is called the maximum stable gain (MSG) and is defined as :-

$$G_{MSG} = \frac{|Y_{21}|}{|Y_{12}|} \tag{3.16}$$

Referring back to Figure 3.21, we can state that for $K < 1$, $G_{T,max}$ (Eqn. 3.15) is not defined. In such a scenario, the maximum stable gain is used (G_{MSG}) but, for $K > 1$, the maximum available gain ($G_{T,max}$) is defined (Eqn. 3.15). The F_{MAX} is defined as the frequency where the gain plots intersect the X-axis or where the gain is unity. Since unilateral systems do not exist at mm-wave frequencies, in this work, F_{MAX} is taken as the frequency where $G_{MAX} = 0\,dB$. The interesting thing to be noted here is the intersection of the curves G_{MSG} and G_{MAX}. The point or the frequency at which these curves start to diverge is called the stability break point (F_{stab}). This is because, for frequencies less than F_{stab}, the system is conditionally stable. This means that oscillations can occur if the source and load impedance are not proper. However, for frequencies greater than F_{stab} the system is unconditionally stable and will never oscillate. This region has been marked in the plot. We will revisit this in the discussion about capacitive neutralization in Section 3.4.3.

These various gain terms has been plotted for our technology of interest (22FDX SOI) in Figure 4.7b.

3.3.4. Gain & Power Match

The analysis in the previous section should be sufficient for emphasizing the importance of matching in order to extract the best possible gain from the amplifier. Two types of matching are possible, i.e., Conjugate (Gain) match or Load-line (Power) match [46]. Figure 3.22 shows that, in the case of an ideal current source, the generator can deliver maximum power to the load when the load resistance is conjugately matched. This means that the real part of the load resistance is equal to the real part of the generator impedance. Any reactive component needs to be resonated out. The term ”Gain-match” will be used when referring to such a matching condition. For a

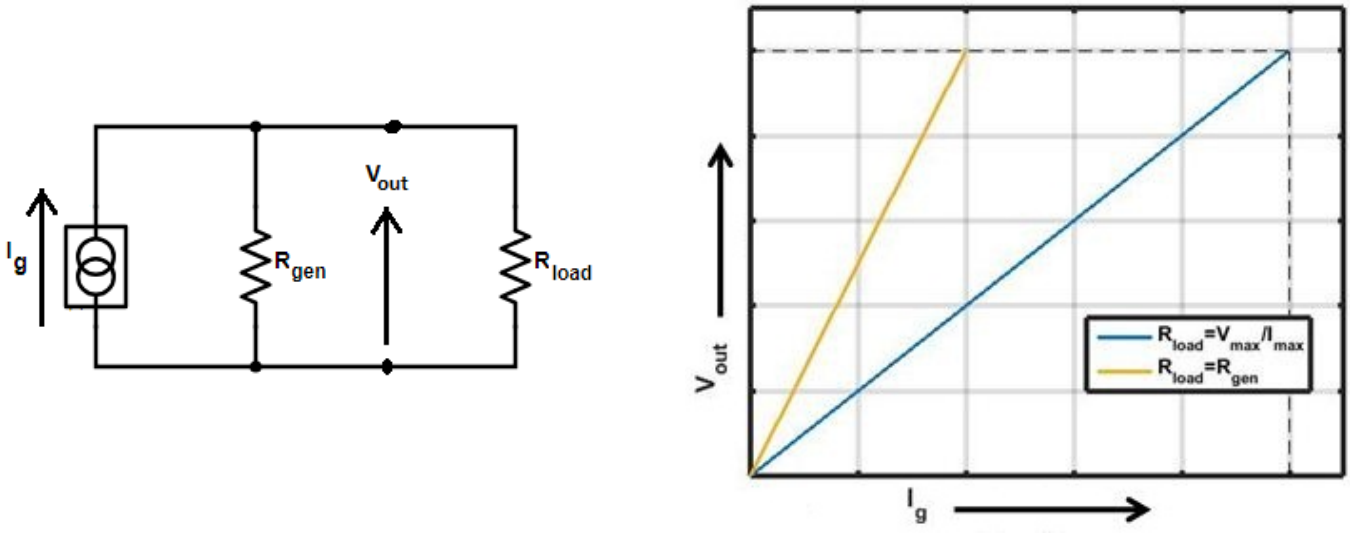


Figure 3.22: **Loadline matching**

real device like the PA, we are limited by the maximum voltage and current ratings of the transistor. These limitations need to be satisfied for a reliable operation. So, for real devices, an optimum resistance is chosen that meets their maximum voltage and current ratings. This kind of a matching will be referred to as a "Power-match" from here on.

Power match has certain advantages over Gain match. Figure 3.23 shows the output power versus input power for an ideal amplifier when "Gain-matched" versus when it is "Power-matched". From this plot, it is clear that even though power match leads to lower output powers at low input power levels, it provides sufficiently higher output power when the input power is above a certain level. This is just the opposite of what one can achieve with a "Gain-match". For a gain-matched scenario, the amplifier provides higher gain for lower input power levels, but the output power is significantly lower for higher input power levels. In Figure 3.23, the 1dB compression point has been marked with an asterisk. As can be observed, the 1dB compression point is lower than what can be achieved for a power-matched case. This means a power-matched amplifier can provide linear output powers for a much larger input power range than would be possible in a gain-matched amplifier. The difference in the 1dB compression points for these two cases can be in the order of 2dB, but it can vary in the range of 0.5 to 3/4dB [46].

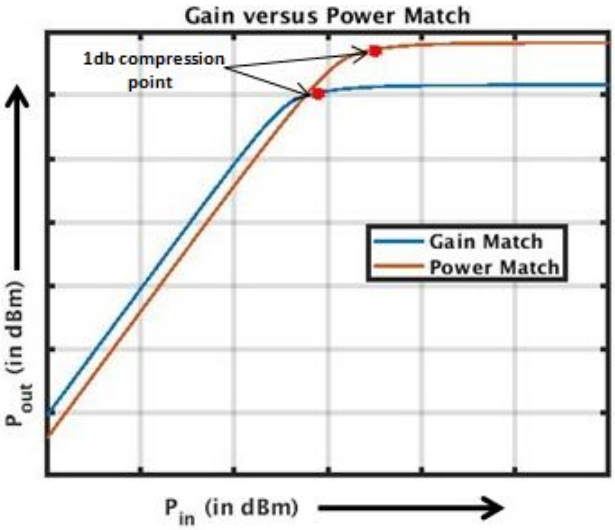


Figure 3.23: **Gain versus Power match**

3.3.5. Loadline/Power Match

In the case of a Class-A power amplifier, in Figure 3.24, the optimum load resistance R_{opt} is given by

$$R_{opt} = \frac{V_{DD}}{I_{max}/2} = \frac{V_{DD}}{I_{DC}} \tag{3.17}$$

This is because, if the device is excited with a sinusoidal waveform, then a voltage swing of V_{DD} is seen at its drain node. The peak voltage at its drain goes up to $2 * V_{DD}$. This is because the drain voltage is also sinusoidal in nature with an amplitude of V_{DD}

across the DC bias voltage of V_{DD}. Similar reasoning leads us to the conclusion that the drain current goes up to $2 * I_{DC}$. Thus the load seen by the amplifier must be given by Equation 3.17. This has been illustrated in Figure 3.25

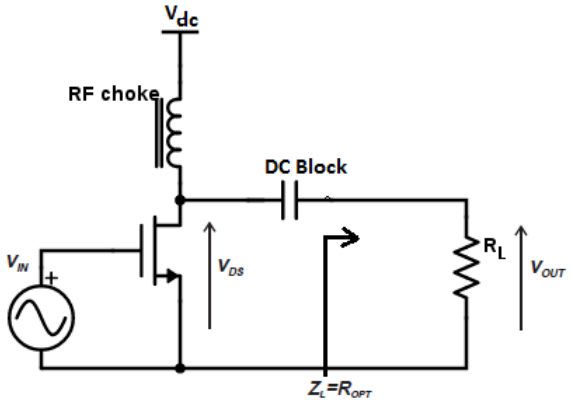


Figure 3.24: **Class-A Power Amplifier**

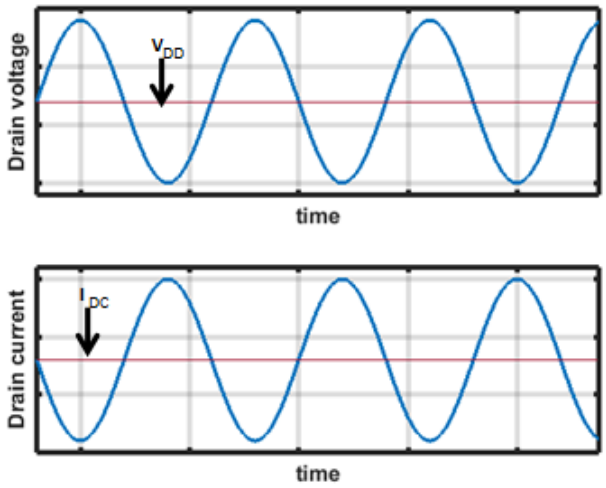


Figure 3.25: **Drain voltage and current waveforms for a Class-A PA**

Now let us outline the requirements of the power amplifier to be used in our system, and then we can utilize the information provided in the last couple of sections to derive some numerical values for some of the circuit elements.

3.3.6. Power Amplifier

The power amplifier in Figure 3.26 receives the combined I and Q signals from the mixers and has to amplify them but, obviously, there are certain problems associated with it. Let us review them a bit :-

- We need to generate a certain amount of power, i.e., 18dBm in this case. Now, for a balun with 1:1 transformation ratio, the voltage swing at the drain of the amplifier is too large for a reliable operation. That is why, in Figure 3.26,

"Medium-Oxide" NMOS devices have been used. "Medium-Oxide" devices have a thicker gate-oxide and can withstand higher DC voltages. For 22FDX SOI Process Design Kit (PDK), the "Medium-Oxide" device has a channel length of 100-nm and can withstand a DC voltage of 1.5V. Of course, the RF voltage amplitude can go up to $\sqrt{2}$ times this value, but even that is not sufficient for a reliable operation.

  - **Solution**: Using power combiners at the output to generate the desired power. Then the output from each PA unit can be lower and hence a single-transistor (Common-Source (CS)) configuration with "Thin-Oxide" devices can be used. This would provide more power gain and hence better Power Added Efficiency (PAE). However, the passive power combiners are lossy and hence it may not benefit the over all system efficiency.
  - **Solution**: Using a cascode structure. Then we will not need a lossy power combining network. Also, the elimination of such a large passive network means a smaller chip area is required. A cascode device is less efficient than a CS configuration, but using a cascode configuration also leads to greater isolation between the output and the input.

- Another issue with using a "Medium-Oxide" device is that the output swing from the mixer needs to be higher. The threshold voltage for a "Medium-Oxide" device is higher than the "Thin-Oxide" devices which is around 0.25-0.3V. As stated in Section 1.4 and elaborated in Section 3.2.1, different models of devices are available. For "Medium-Oxide", the low V_t model has a threshold voltage of around 0.7V which, for "Thin-Oxide" devices, is around 0.25-0.3V. We have not used the flipped well "Super low V_t" model as it is leaky and leads to more DC power consumption. So, using a "Medium-Oxide" device would mean that our mixer also needs to be designed with "Medium-Oxide" devices. As we all know, the switching action is most important in mixers. Therefore, by using a "Medium-Oxide" device we compromise the performance of the mixer. Also, the linearity of the mixer is affected if a large voltage swing needs to be generated at its output. In order to generate a linear output, the size of the devices in the mixer should also be scaled depending on the desired output swing. Wider NMOS and PMOS devices in the mixer is not desired as it will be difficult to drive it because of the large gate-source capacitance of the device. The polyphase filter driving the mixer will have a hard time driving it if the input capacitance of the mixer is huge.

  - **Solution**: To use a cascode structure with a "Thin-oxide" device as the CS input. We only need a $0.6V_{peak-peak}$ signal to drive our PA. Thus the mixer can generate an output with low distortion levels. Figure 3.27 shows the modified schematic.

If the total output power is 100mW and we are limited to a maximum V_{DD} of 1.1V due to the device reliability issues then, from the discussion in Section 3.3.5, we can derive the optimum load that should be presented to the power amplifier to ensure a "Power-match". Output power of $100/2 = 50$mW is required from a single ended amplifier. Putting in the values of 50mW as P_{out} and $V_{DD} = 1.1$ in Equation 3.18, we

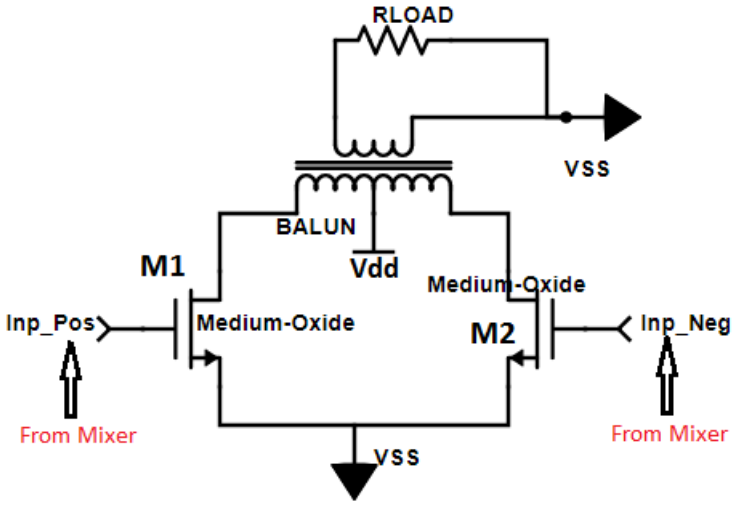


Figure 3.26: **Basic Schematic of the Power Amplifier**

can determine that the optimum load resistance (single ended) required to ensure a loadline or power match is $R_{opt} = 12.1\Omega$. Therefore, the differential R_{opt} is 24.2Ω.

$$P_{out} = \frac{V_{DD}^2}{2R_{opt}}$$
$$R_{opt} = \frac{V_{DD}^2}{2P_{out}} \tag{3.18}$$

If the balun is designed to have a coupling coefficient of $k_m = 0.707$, then, the effective turns ratio of the balun becomes $n = \frac{1}{k_m} = 1.414$. Therefore, a 50Ω load at the output will be transformed to a differential impedance of $50/n^2 = 25$. Simulation results show that $k_m = 0.65$, and, therefore a differential $R_{opt} = 21.125\Omega$ has been achieved finally. This is close to the theoretical value of R_{opt} derived above.

3.4. Different Supply Interpolating Amplifiers

Continuing from our discussion in Section 3.3, the cascode PA shown in Figure 3.27 is the basic building block of our Supply Interpolating Transmitter. The amplifier consists of *neutralizing capacitors* to minimize the effect of C_{GD} capacitance of the CS device. The C_{GD} forms a negative feedback in the transistor that degrades the stability of the MOS transistor. The neutralizing capacitors (C_{neut}) are connected between the gate of a MOS and the drain of the other input device in a differential pair. This takes advantage of the 180° out of phase nature of the drain and gate voltages of a MOS. In Figure 3.27, a part of the output signal (at the source of the cascode device) flows to the input through the C_{GD} capacitances of M1 and M2. This signal is neutralized by an equal and opposite signal that is fed by the C_{neut} capacitors. C_{neut} connected between the gate of M1 and the drain of M2 feeds in a signal that is 180° out of phase with the output signal fed back in by C_{GD} of M1 [47]. A similar explanation applies for the other CS device M2.

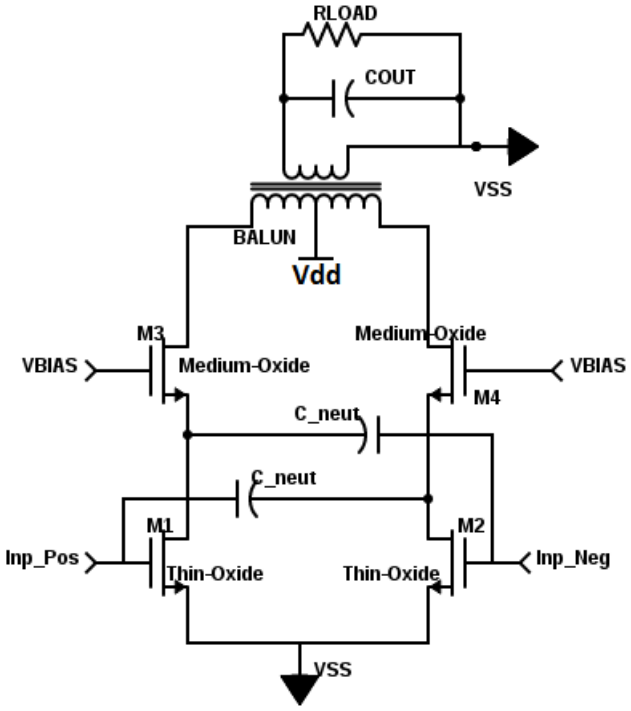


Figure 3.27: **Modified Schematic of the Power Amplifier**

3.4.1. Differential mode stability

As already stated in Section 3.3.3, the Rollet stability factor (K) and Δ (Eqn. 3.14) are sufficient criterions to check for stability of a system. A simple S-parameter simulation can be performed to determine the value of the neutralizing capacitor which satisfies the condition $K > 1$ and $|\Delta| < 1$. It can be observed that the system is unconditionally stable over a range of values for the neutralizing capacitor. To gain further insight into the stability of the system, the poles of the system need to be analyzed as a function of C_{neut} [48].

Figure 3.28 demonstrates a differential amplifier with differential excitation [48]. An equivalent circuit for the same can be derived to analyze the poles and zeros of the system. The transfer function of the equivalent circuit can be derived in terms of the input admittance of the system [48].

By analyzing the transfer function, it can be shown that with the increase in the value of C_{neut}, the poles shift away from the right-half-plane. Eventually, for $C_{neut} \approx 4 * C_{gd}$, the poles move into the right-half-plane. Thus the system becomes unstable at this point. We can also explain this phenomenon intuitively. As already mentioned, C_{neut} feeds back a part of the output signal into the gate of the transistor to neutralize/cancel out the signal that gets fed in through the C_{gd} of the transistor. When C_{neut} is close to the value of C_{gd}, the signal fed back through C_{neut} is able to cancel out the feedback through C_{gd}. When C_{neut} is much larger than C_{gd}, then the feedback signal from C_{neut} is much higher than the signal fed back through the transistor's C_{gd}. Thus at this point, the feedback signal from C_{neut} leads to instability.

The enhancement in stability is only applicable for a differential mode excitation. This enhancement strategy takes a toll on the common mode stability of the system.

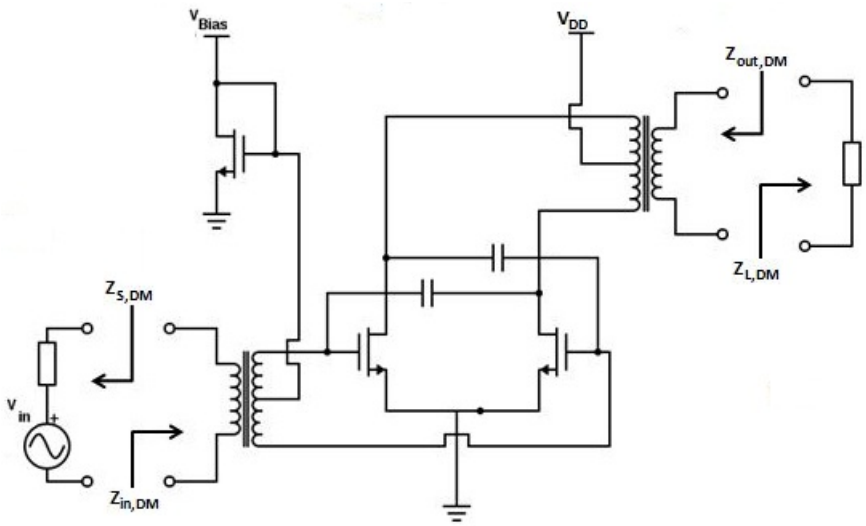


Figure 3.28: **Differential pair in differential mode excitation**

3.4.2. Common mode stability

It can be observed from Figure 3.29 that, in the common mode, the differential pair is connected in parallel. This means that the neutralizing capacitors are also connected in parallel. The issue in common mode stability can be understood by analyzing the equivalent small signal schematic of a neutralized NMOS differential pair under common mode excitation. From the small signal circuit, we can derive the K-factor of a differential pair in common mode excitation. It can be shown that K increases with the increase in the value of C_{neut} but it is still way smaller than 1 [48].

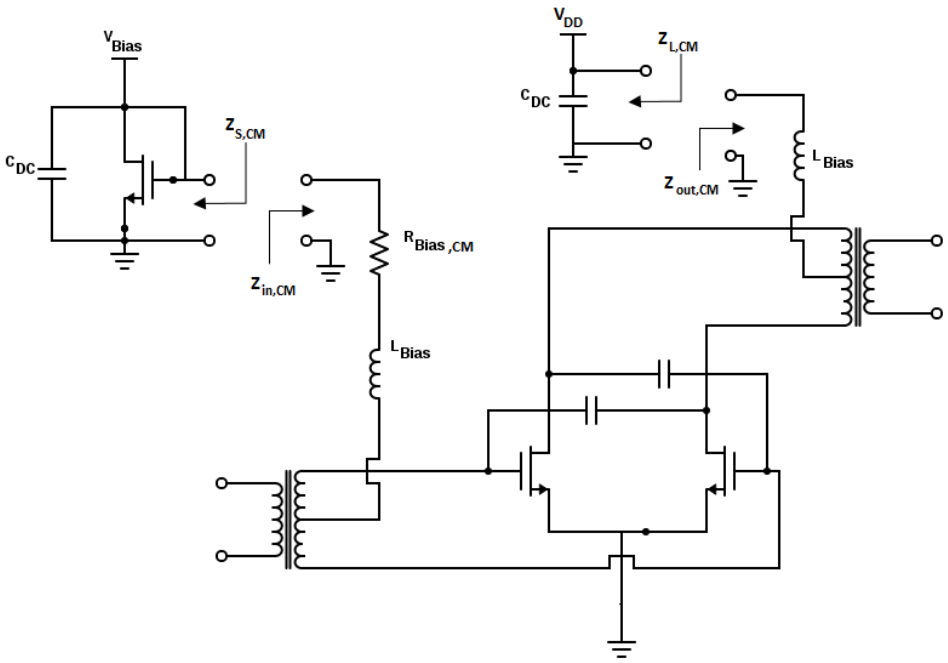


Figure 3.29: **Differential pair in common mode excitation**

For better insight into the common mode stability, we have to analyze the schematic

of the differential pair in common mode excitation (Figure 3.29). $Z_{L,CM}$ has very low impedance even at sufficiently high frequencies because of the presence of decoupling capacitor C_{DC}. Similar reasoning leads us to the conclusion that $Z_{S,CM}$ is also really low. It has been demonstrated in [48] that the poles of such a system lie in the right-half-plane for all values of C_{neut}.

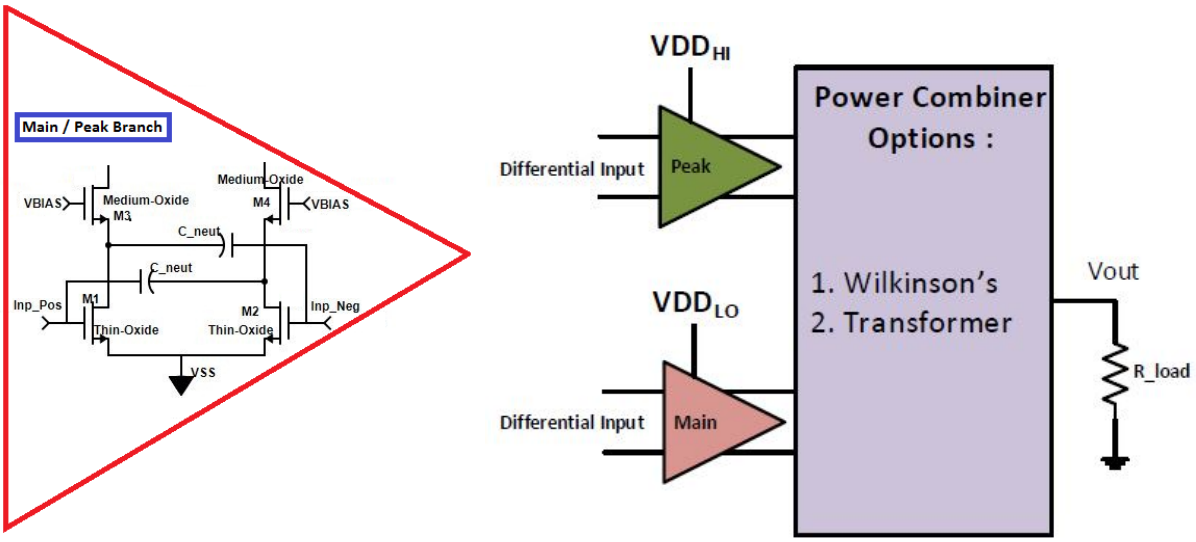
Capacitive neutralization thus helps to restore *differential mode* stability without affecting the power gain and the F_{max}. The common-mode stability is worsened by the use of neutralization capacitors. The poles move towards the right in the right-half-plane for an increase in the value of C_{neut}. The common mode stability can be improved if one can move the poles towards the left into the left-half-plane. Common mode stability can be restored by increasing the series bias resistance in the common-mode bias line [49].

3.4.3. Design of Supply Interpolating Transmitter

After the detailed discussion about the stability issues associated with neutralization, it is apt to turn our attention to the implementation details of the SI transmitter. We start this discussion with the neutralization capacitor. As already stated, "Thin-Oxide" devices are used to implement the neutralization capacitors. The drain node of "Thin-Oxide" RVT has been shorted with that of the source terminal in order to make it behave like a capacitor. This was done to ensure that the neutralization would not vary with PVT conditions. Matching is important at such higher technology nodes, hence using identical devices as our input transistor for implementing C_{neut} ensures a much better neutralization as the capacitors (i.e., both C_{gd} and C_{neut}) change in the same direction. Nevertheless, the Q-factor of MOS capacitors is at least one order of magnitude smaller than MOM capacitors [50]. Therefore, more losses are incurred.

Figure 3.30a illustrates the amplifier structure that has been used in both the $Main$ and the $Peak$ branch. An SI technique requires the use of at least two amplifier branches, therefore, the two branches need to be connected to the load. This implies that the two amplifier branches are also connected to each other, either directly or indirectly. Hence, the amplifier branches will interact with each other. One possible way to minimize such an interaction is to use a Wilkinson's power combiner. However, the isolation resistor is lossy and hence degrades the performance. Apart from that, transmission lines are bulky and occupy a lot of chip area. As a result, a power combiner with isolation capabilities has not been implemented in this work. Another motivation for not using a power combiner with isolation is the different operating principle of SI. As elucidated previously, the two amplifier branches operate at different instants (Figure 3.3). There is only a small region where both the amplifier branches are operating simultaneously. Hence the interaction between the branches is limited.

Transformer based power combiners were the next choice. The "Main" and the "Peak" branch can be connected in series or in parallel (Figure 3.30b). However, for a series combiner, the impedance presented to the PA decreases with each additional stage. For example, if the power from two amplifier branches need to be combined with a transformer of 1:1 turns ratio, the differential impedance presented to each PA is $50/2 = 25\Omega$. When the load impedance is reduced, the transistor size needs to be increased (to deliver more current to the load) in order to produce a certain output power. A transformer based parallel combiner can be used to surpass these issues.



(a) **Basic Cell of Supply Interpolating Amplifier**

(b) **Power Combining options**

Figure 3.30: **Supply Interpolating Amplifier with a power combiner**

The impedance presented to the PA for a 1:1 turns ratio parallel combiner is $50 * 2 = 100\Omega$ (for two amplifier branches). However, the output power that can be delivered by a parallel combiner is usually less than that of a series combiner (for the same parasitics) [51]. Parallel combiners also suffer from low self-resonance frequency and are not suitable for mm-wave operation [51]. Besides, a transformer based combiner with multiple primary coils occupies a large chip area. Due to the above mentioned concerns, neither parallel nor series based transformer power combiners were used.

In order to minimize the use of passives and reduce the chip area, a single primary balun with a 1:1 turns ratio was designed. This means that each amplifier branch is presented with a differential load of 50Ω in the ideal case. The ideal situation is when the output impedance of the branch in the off state is very high such that it is not loading the other branch. However, due to the large transistor size and the frequency of operation, the amplifier in the off state also loads the output. This has been elaborated later on.

Now, the different configurations of the SI transmitter need to be investigated. One way to connect the outputs of the $Peak$ and $Main$ amplifier would be to hardwire them at the output, i.e., current summation. The supply of each branch is provided by connecting a negative voltage source to its VSS terminal while the center tap of the balun is connected to "Ground". This has been illustrated in Figure 3.31a. Hence, the "Main" branch has a supply of $0 - (-VDD_{Low}) = VDD_{Low}$ and, similarly, the "Peak" branch is connected to VDD_{High}. The benefit of this configuration is that a single balun can be used to provide the supply of several amplifier branches. For example, if three amplifier branches need to be connected, then no extra modifications are required. The new branch can be connected to the existing schematic with its own supply voltage (negative supply) connected to its VSS terminal. A slight disadvantage of this schematic is that negative supply voltages are necessary.

Another possibility is to connect the $Main$ branch capacitively to the $Peak$ branch. In this case, the supply voltage for the $Main$ can be connected to the center tap of the inductor as depicted in Figure 3.31b. In this configuration, negative supply voltages are not necessary. However, this comes at the price of extra passives. Simulation shows that the coupling capacitors need to be $> 0.5pF$ in order not to attenuate the output from the $Main$ branch. On the other hand, a higher value of C_{coup} poses a serious issue for the $Peak$ amplifier branch. A higher C_{coup} provides very small resistance and hence the $Peak$ amplifier branch modulates the output impedance of the $Main$ branch on account of its higher drain voltage swing. As a result, the output power from the $Peak$ branch gets attenuated but, of course, this problem is also present in the schematic of Figure 3.31a. The remedy to this issue is simple and does not pose any extra burden on the design. This will be described shortly.

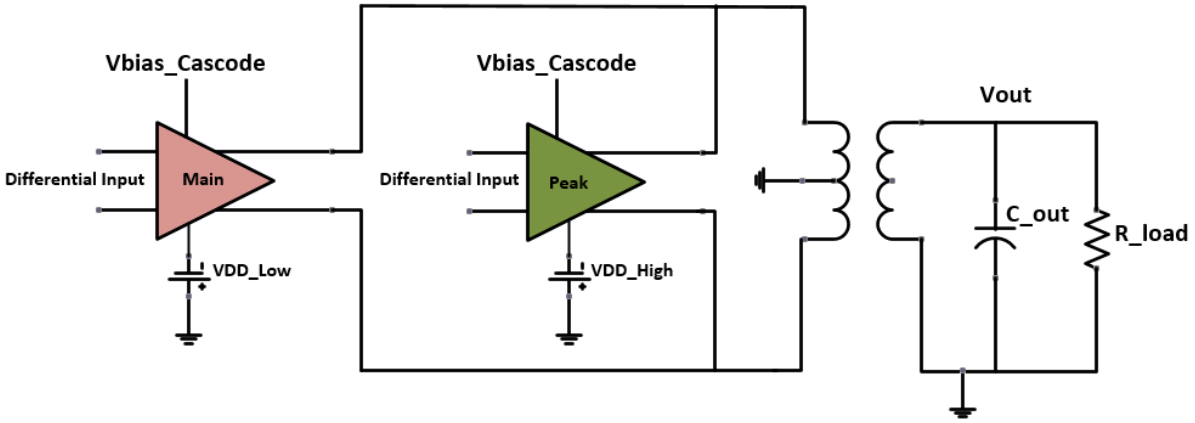
Figure 3.31c portrays yet another configuration of an SI transmitter. The V_{DD} is connected to the center tap of the balun. The supply of the $Main$ stage is offset by connecting its VSS terminal to a DC voltage rather than connecting it to ground. Thus the effective supply voltage for the $Main$ stage is $VDD_{High} - VDD_{Low}$ and that of the $Peak$ branch is $VDD_{High} - 0 = VDD_{High}$. As a result, only a single balun is necessary to provide the supply voltages to all of the branches. If another amplifier branch is necessary, then it can simply be added to the existing configuration with an offset voltage connected to its VSS terminal (similar to the $Main$ stage depicted in Figure 3.31c). This schematic is slightly superior than the previous ones in the sense that we do not need a negative supply voltage and also no passives are required. The Supply Interpolating Transmitter used in this thesis work is based on this configuration (Figure 3.31c).

3.4.4. Practical Considerations

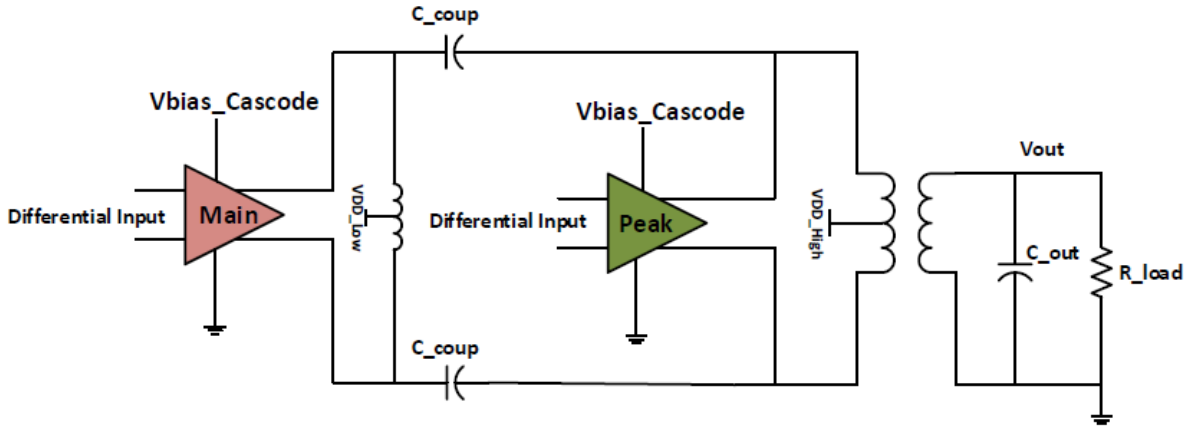
Apart from what has been illustrated in Figure 3.31, a passive mixer and a PPF are required to ensure the proper operation of the SI transmitter. In the configuration depicted in Figure 3.31b, since both the amplifier branches are connected to ground, they can both be biased at the same gate voltage. Both of the branches have a "Thin-Oxide" CS input transistor and hence the DC value of the gate voltage can be the same for both the branches. The LO signal and the baseband signals coming from off-chip can already be biased at the necessary DC voltage. This implies that it is not necessary to bias the PAs separately at the input. However, the same cannot be said about the other configurations. Let us inspect them one by one.

For the schematic in Figure 3.31a, both the $Peak$ and the $Main$ are connected to different DC voltages at their sources. This necessitates the use of a bias network at the input of the PA. For example, if the $Main$ is connected to -0.8V, then the gate voltage should swing around a DC of -0.5V. In essence, the gate of the $Main$ is then biased at $-0.5 - (-0.8) = 0.3V$, i.e., just at the threshold voltage of a "Thin-Oxide" device. Similar reasoning for the $Peak$ branch leads to the conclusion that a bias network needs to be implemented before the amplifier branches. The bias network can be implemented with a R-C network or with a L-C network. Since an inductor occupies more chip area and the Q-factor of a capacitor is usually superior to that of an inductor, the natural choice would be to go for a R-C based biasing circuit.

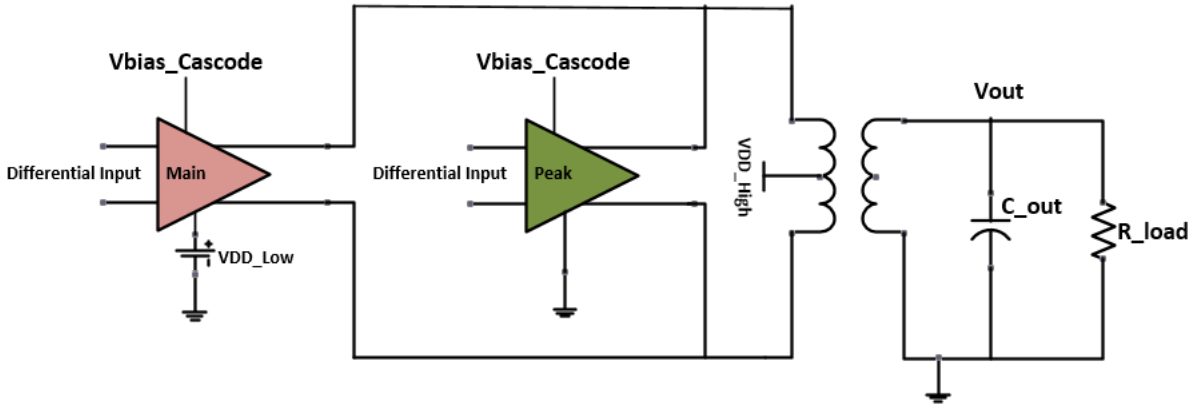
A quick inspection of the configuration depicted in Figure 3.31c suggests that only



(a) **Supply Interpolating Amplifier with Negative Supply Voltages**



(b) **Supply Interpolating Amplifier with Capacitive coupling**



(c) **Supply Interpolating Amplifier with voltage offset at its source**

Figure 3.31: **Different Supply Interpolating Amplifiers**

a single bias network is necessary in this case. However, even though the $Peak$ branch is connected to ground, a dummy bias network is necessary here. Before this is elaborated further, we need to discuss the low level implementation details, i.e., the

layout of the circuit.

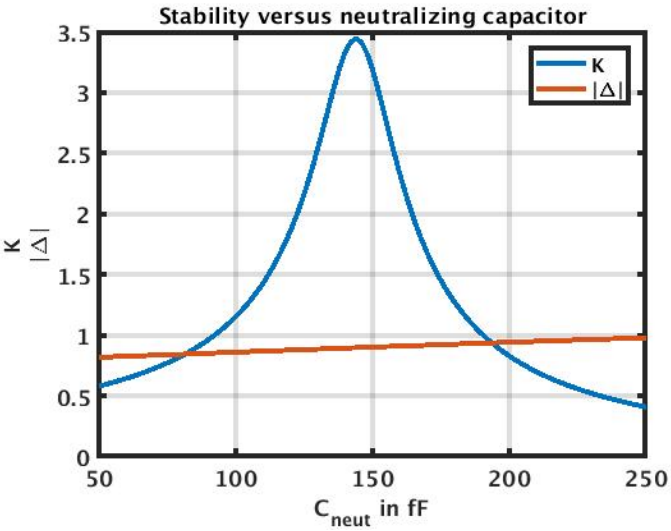
Essentially both the configurations depicted in Figure 3.31c and 3.31a are the same. The bias of the LO and the baseband signals only need to be altered. And since we have not designed the biasing network on chip hence this can be easily done by changing the bias off-chip. The only difference in the configuration depicted in Figure 3.31c is that, if we assume VDD_Low to be a battery, then, the current flowing out from VDD_High charges the supply VDD_Low.

The differential input to the PA comes from the passive mixer. Since, this is an IQ transmitter, four different phases are necessary. The differential IQ LO signal is generated by the PPF (elaborated in Section 3.5). The output of the PPF is routed to the mixer input. In order to ensure decent linearity, the N and PMOS transistors in the mixer cannot be very narrow. Simulations show that the input capacitance of the mixer in the I path is 167fF. The PPF cannot drive such a high capacitive load (Section 3.5). As a result, an inductor had to be used to resonate out the input capacitance of the mixer. Hence, it is wise to use this inductor to provide the bias voltage. Thus the configuration in Figure 3.31b is not suitable at all since it requires passives for the $Main$ stage and also an inductor for each branch to tune out the input capacitance of the mixer. This has already been touched upon in Section 3.3.1, although there the inductor was used to resonate out the input capacitance of the stage following the mixer, the fundamental issue is the same.

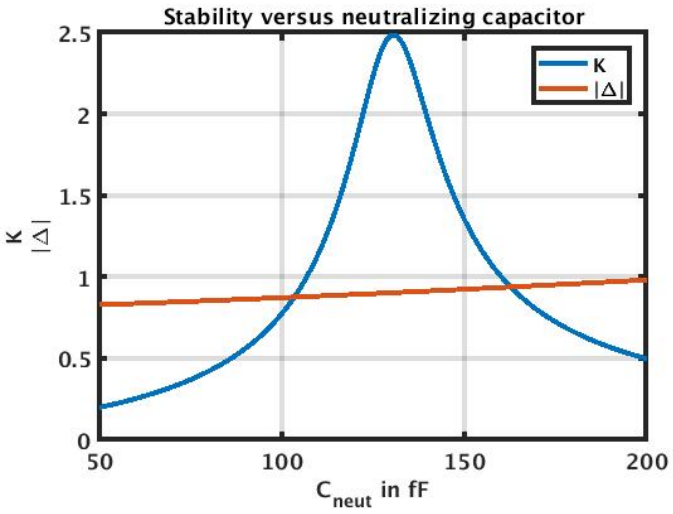
At this juncture, it is easy to imagine why the configuration in Figure 3.31c requires two bias networks. A center-tapped inductor (each for differential I and Q signal) is used to provide the bias for the $Main$ and the $Peak$ branch. Thus, even though a biasing network should not be necessary for the $Peak$ branch, circuit non-idealities (parasitics and input capacitance of the mixer) necessitate the use of an inductor.

The physical dimensions of the inductor need to be minimized so as to have a compact layout. If the inductor is big (i.e., occupies a large area), then the mixer cannot be placed close to the PPF. As a result, the interconnects required between the PPF output and the input of the mixer become too long. This is detrimental to the performance of the entire system as it introduces parasitics. However, these parasitics are the reason why the center-tapped inductor can be small. In this design, an inductor of about $50x36\mu m$ each was required in the I and the Q path. EM simulation results of the inductor are provided in a separate chapter.

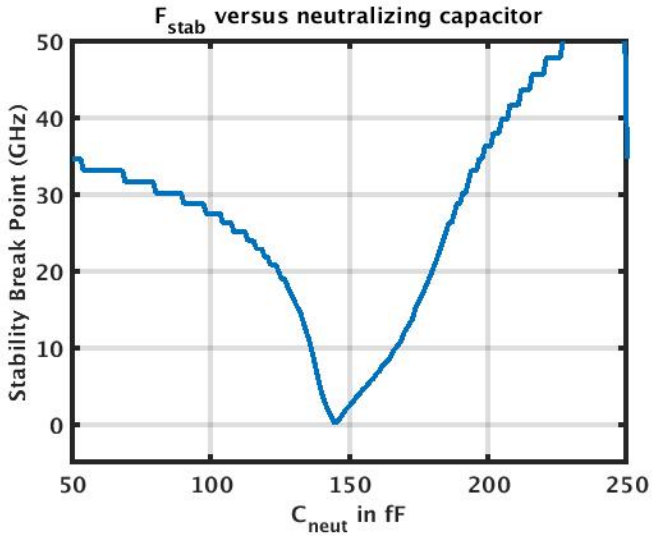
The only issue that remains unaddressed is the problem of the output impedance of the $Main$ branch being modulated by the large drain voltage swing of the $Peak$. Besides this, we also need to consider how to switch between the two branches. As mentioned previously, the transition between the low and the high power region is extremely important. Without a proper transition between the two regimes, the linearity of the system will be affected. The solution to both of these problems can be found by switching the DC bias of the cascode transistor on/off. The DC bias need not be pulled down to 0V. In this work, the DC bias of the cascode has been set to 0.6V for the off state and 1.7V for the on state. In this regard, special attention needs to be paid to the $Main$ branch. When the $Peak$ is on, the drain voltage, on account of its higher swing, can go down to 0.2-0.3V. MOSFET is a symmetrical device hence the node with a higher voltage is always the drain. Ideally, the $Main$ should not conduct when the $Peak$ is on except for the transition period. Assuming that the $Main$ is off



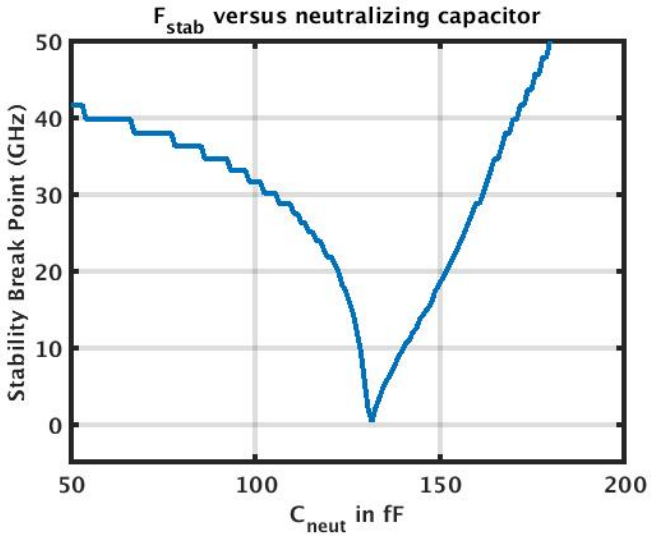
(a) **Main branch**



(b) **Peak branch**



(c) **Stability break point (F_{stab}) for Main branch**



(d) **Stability break point (F_{stab}) for Peak branch**

Figure 3.32: **Differential Stability with capacitive neutralization**

when the $Peak$ is conducting, the voltage at the source node of the cascode (drain of the CS) has to be equal to the VSS of $Main$. In our case, the VSS of $Main$ is connected to 0.8V. Thus, when the drain voltage swings below 0.8V, the source and the drain node of the cascode in the $Main$ branch is interchanged. In order to maintain the non-conducting nature of the $Main$ amplifier, the bias voltage of the cascode is set to 0.6V. This way, the cascode device will never have enough V_{gs} to conduct. The cascode bias voltage has been swept from 0.6-1.7V to enable a smoother transition between the high and the low power regime. This also facilitates the proper operation of the SI transmitter by ensuring that the branch in the off state is not conducting.

For the schematic shown in Figure 3.31c, capacitive neutralization was performed to ensure stability over a wide frequency range. Figure 3.32 shows the plot of the

K-factor and |Δ| for both the $Main$ and $Peak$ amplifier branches after capacitive neutralization. From this plot, it can be observed that the region of instability is also dependent on the value of C_{neut}. One should try to maximize the frequency range of unconditional stability by minimizing the value of F_{stab} (Figures 3.32c & 3.32d). At the optimum value of C_{neut}, i.e., when $C_{neut} = C_{gd}$, the K-factor is maximized and the F_{stab} is minimized. By reducing F_{stab} we increase the frequency range of unconditional stability and hence reduce the frequency range over which the device is conditionally stable.

3.5. Polyphase Filter (PPF)

Since we want to have a complete transmitter solution for 5G applications, we have designed the mixers to convert the baseband signals to RF. Now we need to have a quadrature LO signal. Several different procedures are available to generate quadrature LO signals.

- A divide-by-two circuit can be used, but the problem is that we would then need an LO signal that operates at $2f_o$. Due to the low quality of passives, the phase noise is very high and also the LO signal swing is low for mm-wave operation. So generating very high frequency mm-wave signals with high spectral accuracy and sufficient power is very difficult.

- Quadrature Oscillators are another option. Here, the two coupled oscillators trade off phase noise and tuning range with IQ accuracy [52].

- Quadrature hybrid couplers can also be used. They can provide multi-octave bandwidth of operation with decent amplitude and phase balance. The only disadvantage is that it is not suitable for integrated circuits because of its large size.

- A polyphase filter following a single phase oscillator can be used to generate the quadrature LO signals. Decent IQ accuracy can be obtained by cascading several staggered-tuned stages. This also helps to enlarge the bandwidth of operation [53].

The age old way of generating quadrature signals using PPFs is shown in Figure 3.33a & 3.33b. The configuration in Figure 3.33b is not considered here as it leads to 6dB loss in signal per stage whereas we would experience 3dB loss per stage from the one in Figure 3.33a [54]. The ratio of differential I and Q outputs of the PPF in Figure 3.33a can be expressed by [55] :-

$$H_{const-amp} = \frac{\Delta V_{I,out}}{\Delta V_{Q,out}} = \frac{1 - sRC}{1 + sRC} \tag{3.19}$$

Whereas the ratio of the differential I and Q outputs of the PPF in Figure 3.33b can be expressed by [55] :-

$$H_{const-phase} = \frac{\Delta V_{I,out}}{\Delta V_{Q,out}} = \frac{1}{sRC} \tag{3.20}$$

From Eqn. 3.19, we can state that the amplitude of the I and Q outputs are always the same irrespective of the frequency of operation, the values of R,C and the load impedance. However, the phase difference is $90°$ only at $\omega = \frac{1}{RC}$. By using more number of stages in cascade, the phase can be made equal to $90°$ at each RC pole, i.e., at $\omega_1 = \frac{1}{R_1C_1}$, $\omega_2 = \frac{1}{R_2C_2}$ etc., where the subscript 1 refers to the 1^{st} stage and 2 refers to the 2^{nd} stage and so on.

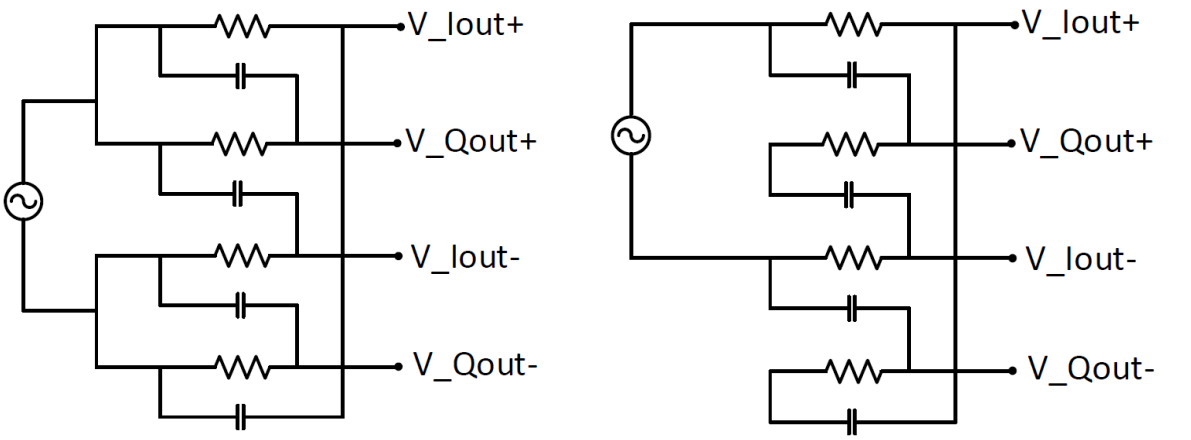
Similarly, from Eqn. 3.20, we can infer that the phase difference between the I and Q outputs is always equal to $90°$ as the ratio is always imaginary. The phase difference is independent of the frequency of operation and the values of R,C, but the amplitude of the I and Q outputs is equal only at $\omega = \frac{1}{RC}$. This again can be resolved by using more stages in cascade. The amplitude can be corrected by using limiting amplifiers [54], but this will degrade the efficiency of the system. So the configuration in Figure 3.33a is a better choice for mm-wave operations. However, it also suffers from a serious drawback.

The long interconnect at the output (Figure 3.33a) makes this design asymmetric. At mm-wave frequencies, the parasitic capacitances associated with such a long interconnect would be detrimental to the performance of the entire transmitter. It is also wise to have less routing as it is severely lossy at mm-wave frequencies. Long interconnects at mm-wave frequencies can have significant inductance and this affects the performance.

A symmetric design has been proposed in [54] that also manages to reduce the length of the long interconnect. Also, since this design is more symmetric, better image rejection ratios (IRR) can be obtained. Two stages of PPF are used to make it more wideband. The different stages are tuned to different frequencies (stagger tuned). Thus the combined effect is that sufficient rejection is available over a range of frequencies. We have used the same PPF design technique that has been illustrated in Figure 3.34. In order to account for the inductance of the interconnects, the value of C_{outer} capacitors have been tuned. Extensive EM simulations are performed to estimate the inductance of the interconnects and thus tune the value of the capacitors to ensure an IRR better than -32dB.

The input impedance of a single stage PPF can be derived as shown below [Figure 3.35]

$$R_{IN} = \left(\frac{1}{j\omega C} + R\right) \Big\| \left(\frac{1}{j\omega C} + R\right)$$
$$R_{IN} = \frac{1}{2}\left(\frac{1}{j\omega C} + R\right) \tag{3.21}$$



(a) **Polyphase Filter with constant amplitude output**

(b) **Polyphase Filter with constant phase output**

Figure 3.33: **Asymmetric Polyphase Filters**

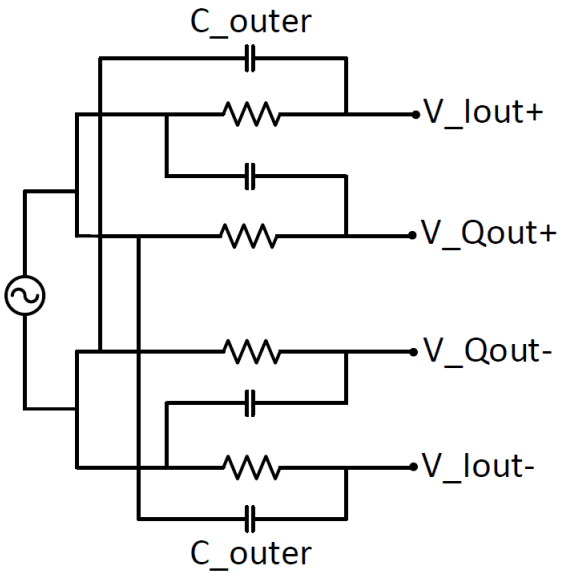
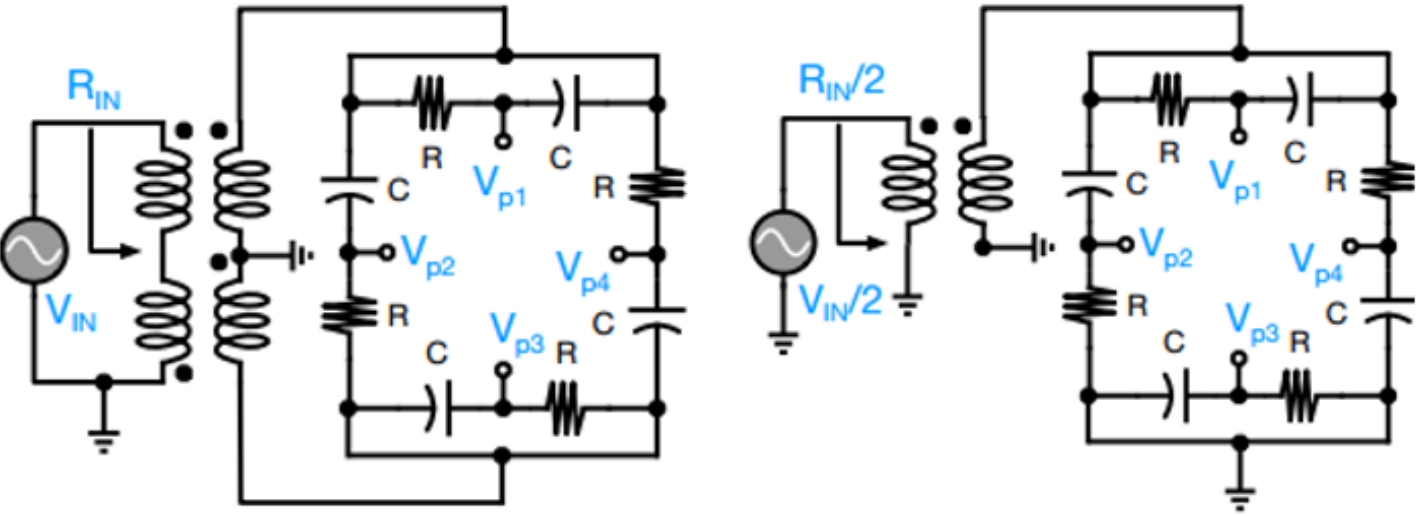


Figure 3.34: **Polyphase Filter with symmetric interconnects**

It can be shown that, if a symmetrical load Z is applied at the four output nodes (Figure 3.36), then the input impedance does not change at all [55], but the output swing is reduced because of voltage division. This is illustrated by Figure 3.36.

Assuming, $V_{P4} = a(-j)V$ and $V_{P1} = aV$, we can apply Kirchhoff's voltage law (KVL)



(a) **Differential** (b) **Single Ended**

Figure 3.35: **Input Impedance of Polyphase Filters**

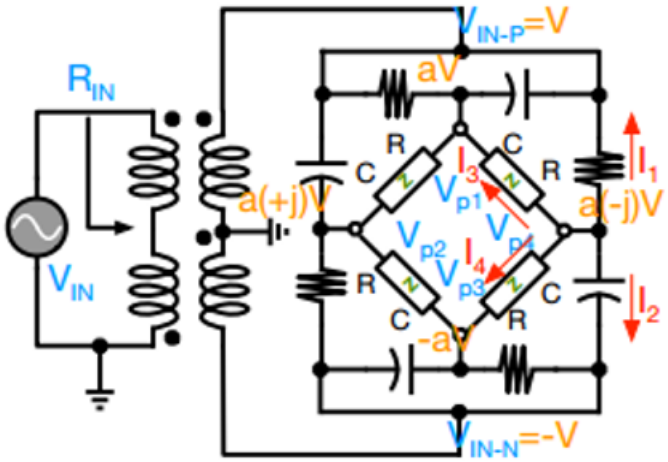


Figure 3.36: **Attenuation of output when loaded with a symmetrical load**

to arrive at the following result

$$
\begin{aligned}
I_1 &= \frac{-jaV - V}{R} \\
I_2 &= \frac{-jaV + V}{-jR} = \frac{aV + jV}{R} \\
I_3 &= \frac{-jaV - aV}{Z} \\
I_4 &= \frac{-jaV + aV}{Z}
\end{aligned}
\tag{3.22}
$$

Since, according to Kirchhoff's current law (KCL), $I_1 + I_2 + I_3 + I_4 = 0$, we can derive that

$$\begin{aligned} \frac{aV(1-j) - V(1-j)}{R} + \frac{-j2aV}{Z} &= 0 \\ or,\ a(Z(1-j) - j2R) &= Z(1-j) \\ or,\ a = \frac{Z}{Z - j(1+j)R} &= \frac{Z}{Z + R(1-j)} \end{aligned} \tag{3.23}$$

If $Z = R(1-j)$, then $a = \frac{1}{2}$. Thus we can state that because of this loading $|V_{P1-4}| = \frac{1}{2}|V_{INP-INN}|$ (Figure 3.36). If $Z = \infty$, then $|V_{IN-P}| = |V_{P1-4}|$ as there is then no loading effect.

The output swing is slightly increased (by a factor of $\sqrt{2}$) if an identical stage is connected in cascade. Figure 3.37 depicts the node voltages. If $Z_{1st} = Z_{2nd} = R(1-j)$, then it can found out that

$$\begin{aligned} V_{P1-4} &= \frac{V_{INP-INN}}{2} = \frac{V_{IN}}{4} \\ V_{P5-8} &= \sqrt{2}V_{P1-4} = \frac{V_{IN}}{2\sqrt{2}} \end{aligned} \tag{3.24}$$

Usually the input capacitance of a differential pair is loading the output stage of a

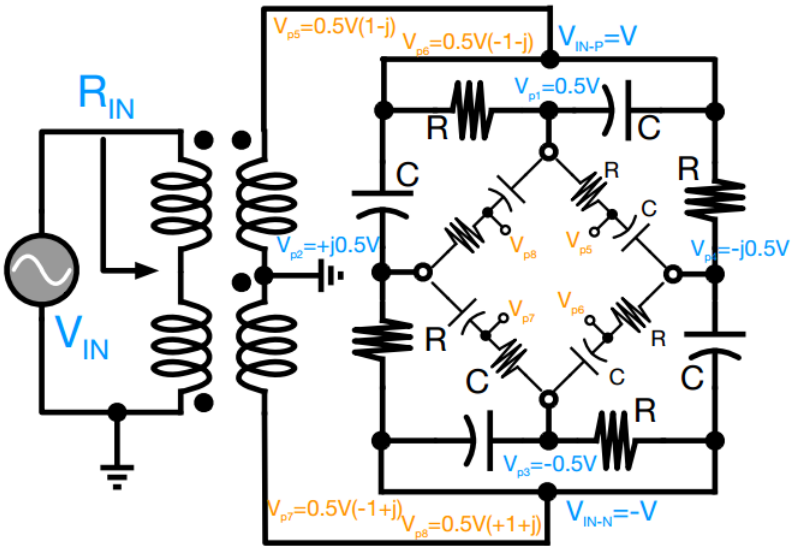


Figure 3.37: **Attenuation of output when loaded with an identical stage**

PPF. Due to voltage division between the polyphase filter capacitor and the grounded load capacitance of the differential pair, losses are incurred [53]. These losses can be quite substantial at mm-wave frequencies where even a small capacitor would significantly load the PPF. As stated in [53] the losses in a cascaded PPF can be limited by using stages with increasingly larger input impedance. This ensures that the succeeding stages load the previous stage less. This helps in reducing the effect of voltage

division. Thus fewer losses will be incurred. However, the value of the resistors cannot be very large as the parasitic capacitances to the substrate would then limit its frequency of operation. However, the impedance of the first stage is also critical and cannot be very low as it would then load the stage driving the PPF. In this work, the PPF is followed by the mixer. The input capacitance of the mixer is 167fF which, at 30GHz, presents a load of 31Ω. Here, we go back to our discussion in Section 3.3.1. It was mentioned that it is possible to resonate out the input capacitance of the mixer by using an inductor at the input of the mixer. Ideally, after resonating out the input capacitance of the mixer, the PPF should see a very high impedance($\approx k\Omega$), but, in reality, the quality factor of the inductor limits the resistance that the PPF sees. At resonance the PPF sees a differential resistance of R_p (Eqn. 3.4). The long interconnects between the output of the PPF and the mixer input also contribute to some inductance. Hence the explicit inductor required at the mixer input is not very high. In this design, a differential inductor of 36pH each was used in the I and Q paths. With a quality factor of ≈ 17 at 30GHz, this leads to a parallel resistance $R_p \approx 116\Omega$. Thus, in reality, the PPF is loaded by this impedance.

Stage	R	C
1^{st}	30Ω	$174\ fF$
2^{nd}	60Ω	$87\ fF$

Table 3.3: Resistance and capacitance values used in the PPF

As already mentioned before, the impedance of each stage needs to be scaled up in order to prevent the loss in signal swing. As a result of this, a resistance of 30Ω and 60Ω has been used in the first and the second stage of the PPF, respectively. Using Eqn. 3.21, we can state that the input impedance of the PPF filter is $(15 - j15)\Omega$. Even though the input capacitance of the mixer does not directly determine the input impedance of the PPF, it plays a major role in determining the value of the resistances to be used in the PPF stages preceding it. The resistance and the capacitance values used in the PPF are mentioned in Table 3.3.

3.6. Conclusion

Compared to other efficiency enhancement techniques, "Supply Interpolation" makes use of multiple supply voltages to provide efficiency peaks at power back-off. This is especially suitable for modern communication systems like 5G where massive MIMO is going to be used. The implementation details of a SI transmitter was provided along with a short overview about the different matching techniques. A passive mixer was implemented in order to minimize the power consumption in the upconversion stage. The analysis of the input impedance of a polyphase filter was provided and a suitable configuration was chosen to minimize parasitic capacitances. A comparative study was provided between the different technologies like CMOS, PDSOI and FDSOI. The benefits of using body biasing were outlined as well and these were supported by simulation results of the three different MOS devices available in 22FDX technology.

4

Layout

The schematic level design and simulations are like the tip of an iceberg. In terms of circuit design, it means that there are many more problems and issues that need to be addressed. The mismatch among different circuit components, parasitic inductances, capacitances, long interconnect, and routing issues are extremely critical and they must be meticulously addressed before signing off the chip. For mm-wave systems, layout plays an essential role, and serious attention needs to be paid in order to optimize it. Let us dive beneath the surface and face the rugged monstrous reality of chip design.

As already illustrated in Figure 3.18, passive mixers are based on pass-gates. An efficient way had to be found to connect all of these numerous transistors together. The P & N MOS devices of a pass gate have been placed together into a unit cell for better matching. Thus the appropriate finger width (W_f) and the number of fingers (N_f) had to be decided but, in order to do that, we need to brush up on a few facts now.

4.1. Basics

The transit frequency (F_t) and the maximum frequency of power gain (F_{max}) are functions of the gate resistance (R_g), gate-source capacitance (C_{GS}), and gate-drain capacitance (C_{GD}) which are reflected in Eqns. 4.1 & 4.2 [56]. All of these factors are dependent on the finger width (W_f) and the number of fingers (N_f). Thus, an optimum value of W_f and N_f has to be found.

$$F_t \approx \frac{g_m}{2.\pi.(C_{GS} + C_{GD})} \tag{4.1}$$

$$F_{max} \approx \frac{F_t}{2.\sqrt{R_g.(g_m.\frac{C_{GD}}{C_{GS}+C_{GD}}) + (R_g + r_{ch} + R_s).g_{DS}}} \tag{4.2}$$

The gate resistance R_g consists of the poly-silicon gate resistance and the resistance of the metal lines used to connect multiple fingers together. For active circuits like power amplifiers, power gain is more important than current gain. Hence, we are more concerned about F_{max} and try to maximize it as much as possible. Our transistor layout

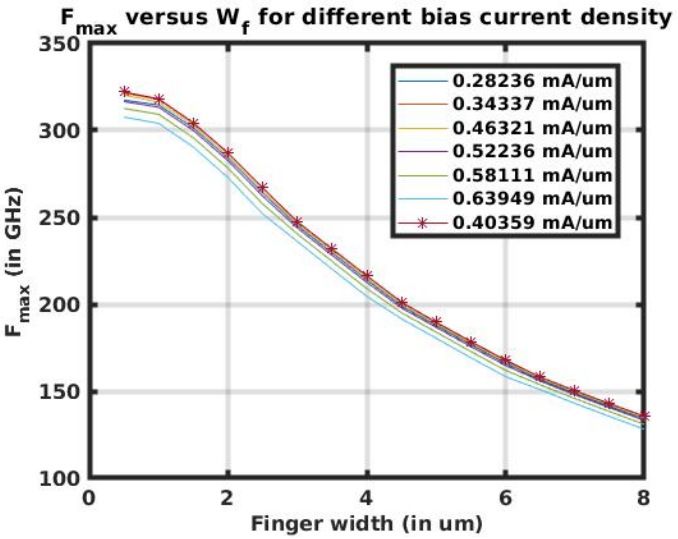
choices are centered around this simple fact. It is evident from Equation 4.2, that F_{max} can be maximized by reducing R_g. As discussed, since R_g depends on both the poly-silicon resistance and the metal interconnect resistance, an optimum value for W_f and N_f must be selected to achieve a balance between the metal and poly resistance. Increasing the number of fingers might help in reducing the poly-silicon resistance for a fixed total width, but the reduction in poly-silicon resistance may be offset by the increase in metal interconnect resistance. Hence, we may not benefit much and the F_{max} might deteriorate too. A similar argument applies for finger width. If the finger width is too wide then the poly-silicon resistance dominates R_g. Increasing the bias current density beyond a certain point is not helpful, either, firstly because it does not help to increase F_{max} and a higher bias current density also increases the DC power consumption. This is not acceptable at mm-wave frequencies because the efficiency is inherently low at such high frequencies.

For the following simulation results, the total width of the device has been kept constant at 16μm, only W_f has been swept. Figure 4.1 & 4.2 shows how F_{max} is affected by the finger widths and the bias current densities. The highest F_{max} is obtained for a bias current density of 0.4mA/μm and 0.25mA/μm for the "Thin-Oxide" and the "Medium-Oxide" device, respectively. The plot corresponding to the highest F_{max} has been highlighted and plotted distinctly in Figures 4.1a & 4.2a. From these figures it can be observed that F_{max} starts to decrease with the increase in finger width. For 65-nm bulk CMOS, maximum F_{max} can be achieved for a current density of 0.8mA/μm [50]. This is almost twice as much current as is required for 22FDX SOI technology. Lower gate resistance (R_g) along with lower overlap capacitances enable high frequency operation. As a result, higher technology nodes are a suitable choice for mm-wave systems. However, the optimum value of finger width ($W_f = 1\mu m$) that gives the best F_{max} is the same as that stated in [50]. The optimum finger width seems to remain unaffected by technology scaling.

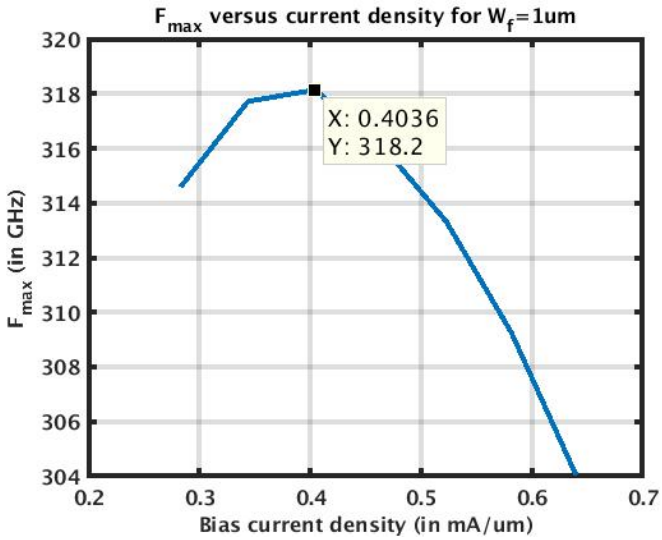
These results indicate the optimum finger width for the CS and the cascode transistor. In order to reduce DC power consumption, the devices have been biased well below the necessary 0.4mA/μm and 0.25mA/μm current density. The maximum available power gain (G_{max}) at 30GHz for both the "Thin-Oxide" and the "Medium-Oxide" transistor is illustrated in Figure 4.3. The variation in G_{max} over bias current density is portrayed. Thus it can be stated that, G_{max} drops by about 1dB for 40% and 33% reduction in bias current density for "Thin-Oxide" and "Medium-Oxide" devices, respectively. Hence, there is some freedom in choosing the bias points. However, as far as the finger width of the devices is concerned, it is better to use a 1μm wide finger for both of these devices. Slightly wider fingers have been used in this work to ease layout efforts and to minimize interconnect lengths.

4.2. Mixers

Figure 4.5 shows the layout of the mixer. This is only a part of the entire mixer shown in Figure 3.18. In Figure 4.4, the NMOS and PMOS transistors have been labelled. The entire mixer for the I and Q paths is placed separately and their differential outputs are taken out separately. The positive side and the negative side of the PA are also laid out separately. Therefore, this makes the routing a bit less complicated. The "I" shaped interconnect in Figure 4.4 is used to connect the sources and the drains of

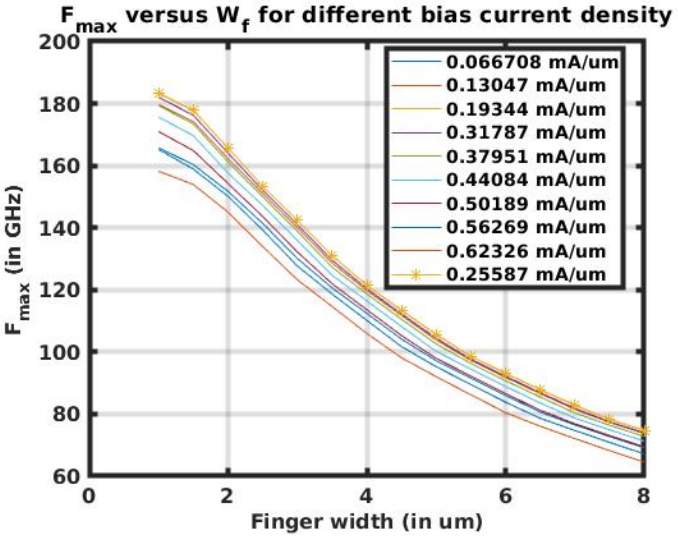


(a) F_{max} **versus Finger Width (W_f) for different Bias current densities**

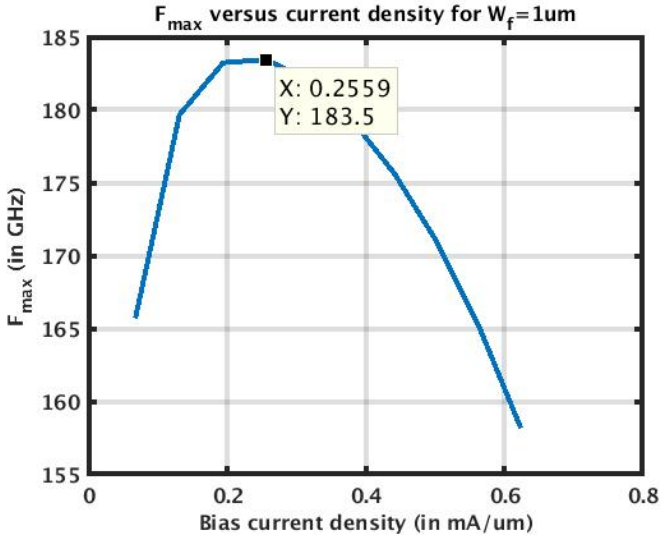


(b) F_{max} **versus Bias current densities for W_f=1μm**

Figure 4.1: **Plot of F_{max} versus Finger Width and Bias current density**



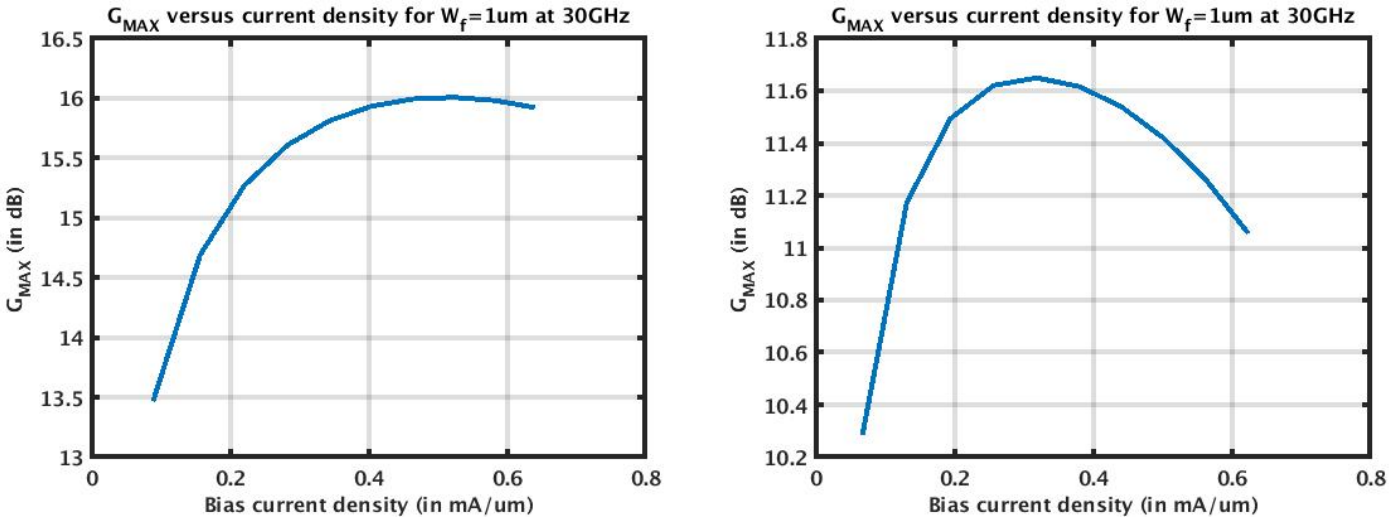
(a) F_{max} **versus Finger Width (W_f) for different Bias current densities**



(b) F_{max} **versus Bias current densities for W_f=1μm**

Figure 4.2: **Plot of F_{max} versus Finger Width and Bias current density for Medium-Oxide NMOS**

the P and the N-MOS devices together. A large number of vias have been used to reduce the overall resistance of the vias and the interconnects as a whole. Multiple metal layers are drawn on top of each other to reduce the resistance of the lines. Metal-4 (M4) is used to feed the baseband signal into the mixer (Fig. 4.5). M7 is used to route the LO signals in to the mixer (Figure 4.4). This is because the mixer



(a) **Thin-Oxide NMOS** (b) **Medium-Oxide NMOS**

Figure 4.3: **Plot of G_{max} versus Bias current density at 30GHz**

is preceded by the polyphase filter and a center-tapped inductor. Since M9 is used in the inductors, it is better to route it using M9 and, then, close to the mixer, the signal is routed through M7 and passed on to M1. Using higher metal layers is beneficial as they are relatively thicker than the lower metal layers. This leads to fewer losses in the metal traces as the resistance is inversely proportional to the metal thickness. Ultimately, the LO signal has to be connected to the gate of the transistor (which is in M1). The purple coloured metal (M8) at the center (Figure 4.5) is used to pick up the RF output from all of the different pass gates. It is connected at the center to maintain symmetry. The LO signals are split up close to the transistor and fed in to the gate from both sides, i.e., from the top and the bottom (Figure 4.4). This helps to lower the gate resistance. From Eqn. 4.2, it can stated that F_{max} can be maximized by reducing the gate resistance R_g.

The finger width of 0.8μm and 1μm has been used for the NMOS and PMOS devices in the mixer, respectively. This is because the mobility of a hole is less than that of an electron. Thus, we usually tend to use a slightly larger device for PMOS. However, in this case, we had to limit its size to ensure that we have a more compact layout.

4.3. Neutralizing Capacitors

Thin-oxide NMOS has been used to implement the neutralizing capacitors. The same transistor has been used in the PA as CS devices. The source and the drain of the NMOS have been shorted together. We also need only half the number of devices we use in the original PA (for the CS) as both the C_{GD} and C_{GS} are tied together. In the PA, we used 12 parallel devices in the "Main" stage, so for C_{neut} we need 6 parallel devices. Figure 4.6 shows the layout of a neutralizing capacitor. M2 (narrow red metal around the device) is used to connect all of the source and drain terminals together. The output is taken out from the extreme right; M7 is used in this case. The thick purple line at the center (M8) is used to connect to the gate terminal. The RF output from the mixer is connected here. The output from the extreme right of C_{neut} goes

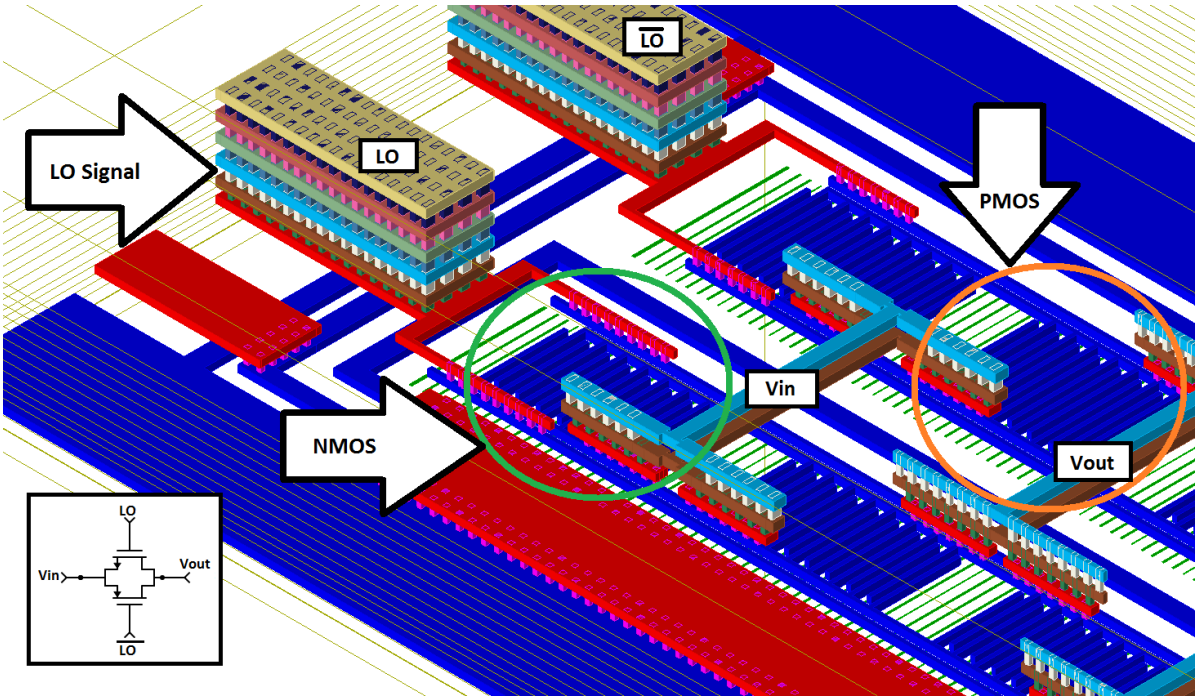


Figure 4.4: **Layout of Mixer unit cell**

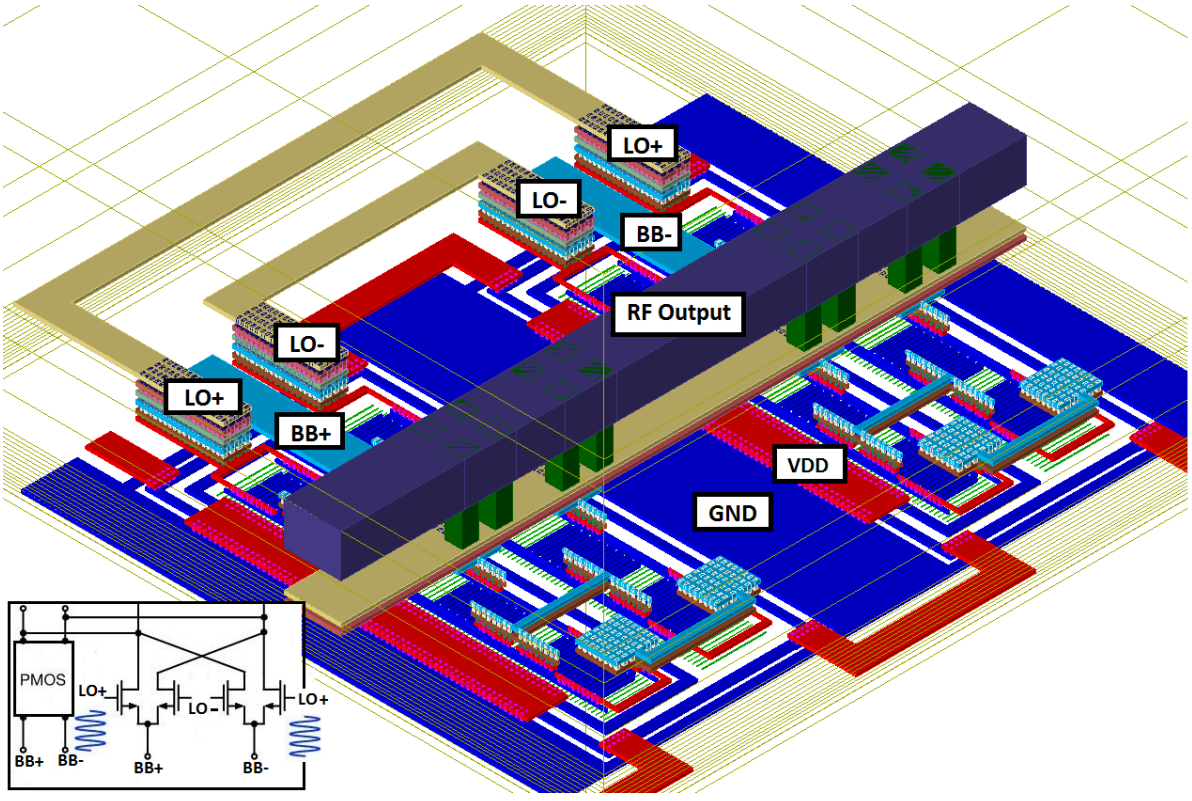


Figure 4.5: **Layout of Passive Mixer**

to the other drain terminal of the differential MOS. The wider M2 (red) lines on the sides are used to connect the ”VDD” signals to the N-well contacts.

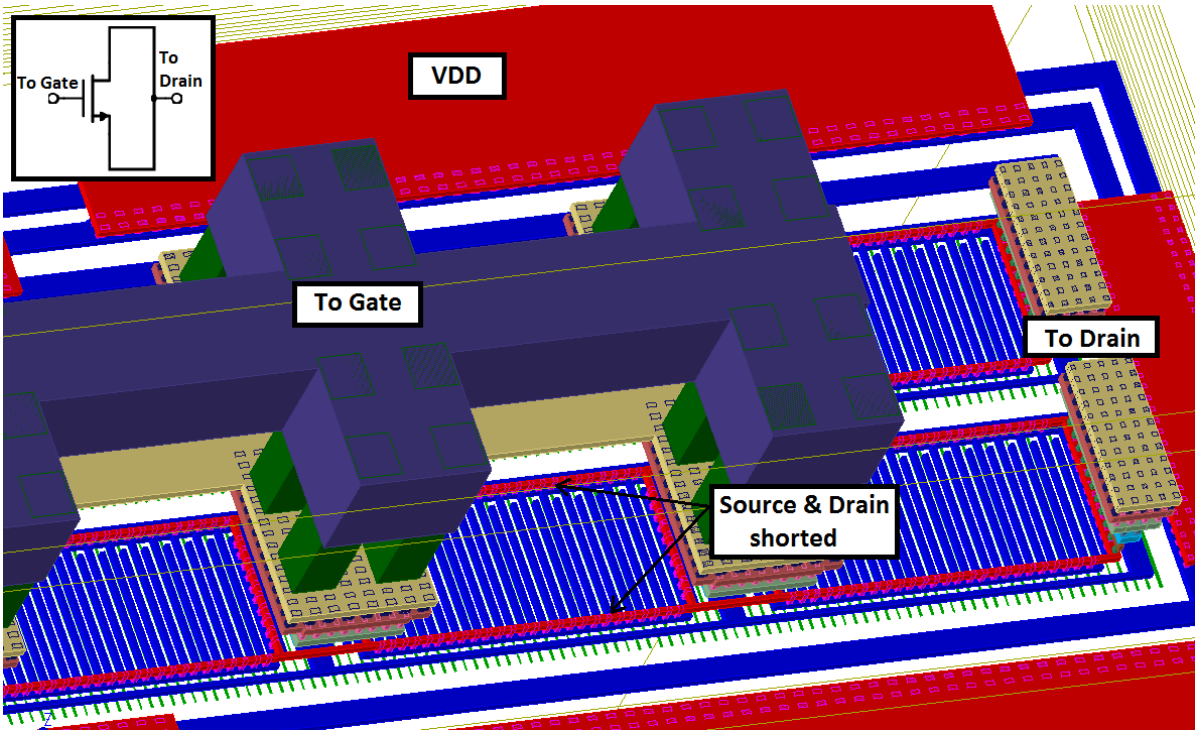


Figure 4.6: **Layout of the Neutralizing Capacitor**

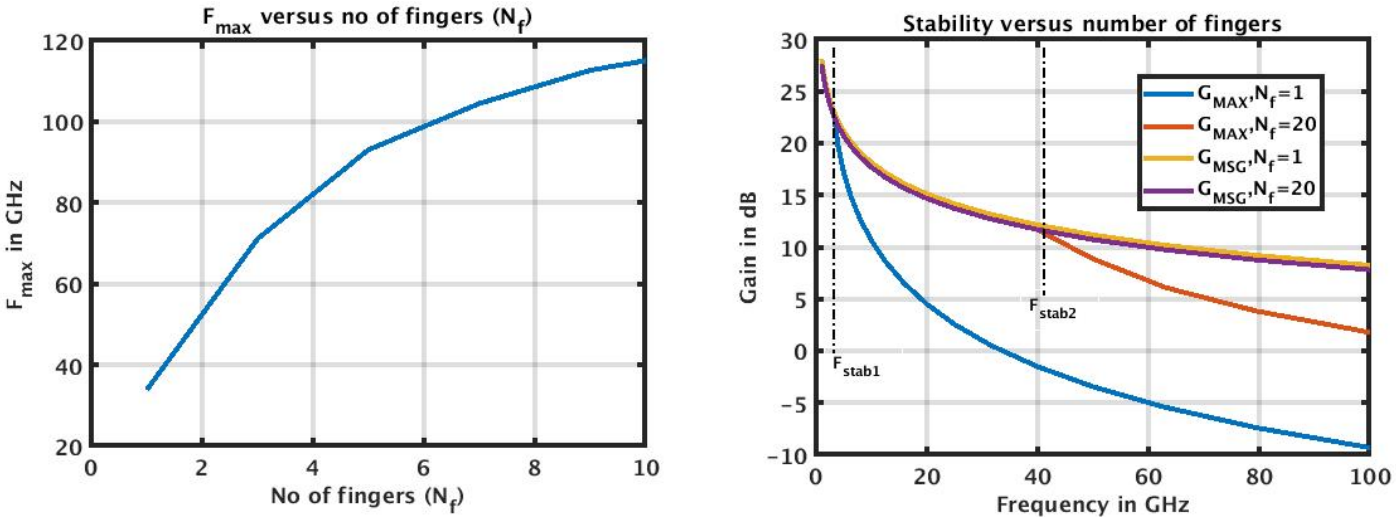
4.4. Power Amplifier

Figure 4.7b shows the stability issue associated with gate resistance. Gate resistance is, in turn, related to the finger width and the number of fingers. For the plot in Figure 4.7a, the total width of the MOS was kept fixed at 16μm. The number of fingers of the device was swept, and the F_{MAX} of the device has been plotted. With the increase in the number of fingers, the finger width of the device decreases (as the total width is fixed). This leads to reduced gate resistance R_g. Figure 4.7b depicts how reduced gate resistance leads to instability. As mentioned before, for $K < 1$, G_{MSG} is defined whereas, for $K > 1$, G_{MAX} or the maximum transducer gain is defined. The plot in Figure 3.21 shows the region of conditional and unconditional stability. The frequency at which G_{MSG} and G_{MAX} diverge from each other is the point of stability break down (Figure 3.21). This frequency (F_{stab}) increases when the F_{MAX} of the transistor increases. This implies that the region of conditional stability also expands. Figure 4.7b illustrates that, when the number of fingers of a transistor is increased, the stability break point (F_{stab}) moves higher up in frequency. F_{stab2} (for $N_f = 20$) marked in the plot is much higher than F_{stab1} (for $N_f = 1$). Hence, the amplifier becomes conditionally stable and is more likely to oscillate over a wider frequency range.

Based on the above analysis, the finger widths of the transistors were set to 1μm and 1.5μm for the CS and the cascode device, respectively of the $Main$ branch. Both the CS and the cascode device have 32 fingers each in order to have a device wide enough to deliver the necessary output current. The finger width of the cascode was set slightly larger because then we can use the same number of parallel devices. This leads to a compact layout. Figure 4.8 shows the layout of the basic unit cell of the

"Main" PA. There are 12 (multiplier) of these in the "Main" PA. The "Gate" of the CS and the cascode are connected on both sides. This helps to reduce the impedance of the gate and maximize its F_{max} (Figure 4.7a). The intermediate node, i.e., the drain of the CS and the source of the cascode are connected together over the "Gate" of the cascode (Figure 3.27).

In both the devices, the gate and the drain terminals are connected on the opposite sides of the transistor. This helps to minimize the gate-drain capacitance which can lead to stability issues at mm-wave frequencies. Similar procedures have been followed to design the "Peak" amplifier branch. Table 4.1 demonstrates the transistor sizes and the multipliers for each PA branch. All the sizes are in nanometers (nm).



(a) **Variation of F_{MAX} with number of fingers** (b) **Stability analysis for multiple fingers**

Figure 4.7: **Effect of number of fingers on stability**

Branch	Common Source W/L	Cascode W/L	Multiplier	V_{DD}
Main	1200/18	1500/100	12	0.9
Peak	800/18	800/100	18	1.7

Table 4.1: Transistor size in each PA branch

4.5. Polyphase filter

In order to design the polyphase filter(PPF), the procedure outlined in [54],[57] was followed. First, the R and C (nominal values) were laid out and the interconnects were separately simulated in Momentum. Then, depending on the value of the interconnects, the capacitance values were adjusted. Since center-tapped inductors were used at the output of the PPF to tune out the input capacitance of the mixers, the routing became longer and lossy. Thus, EM simulations were used to select the right

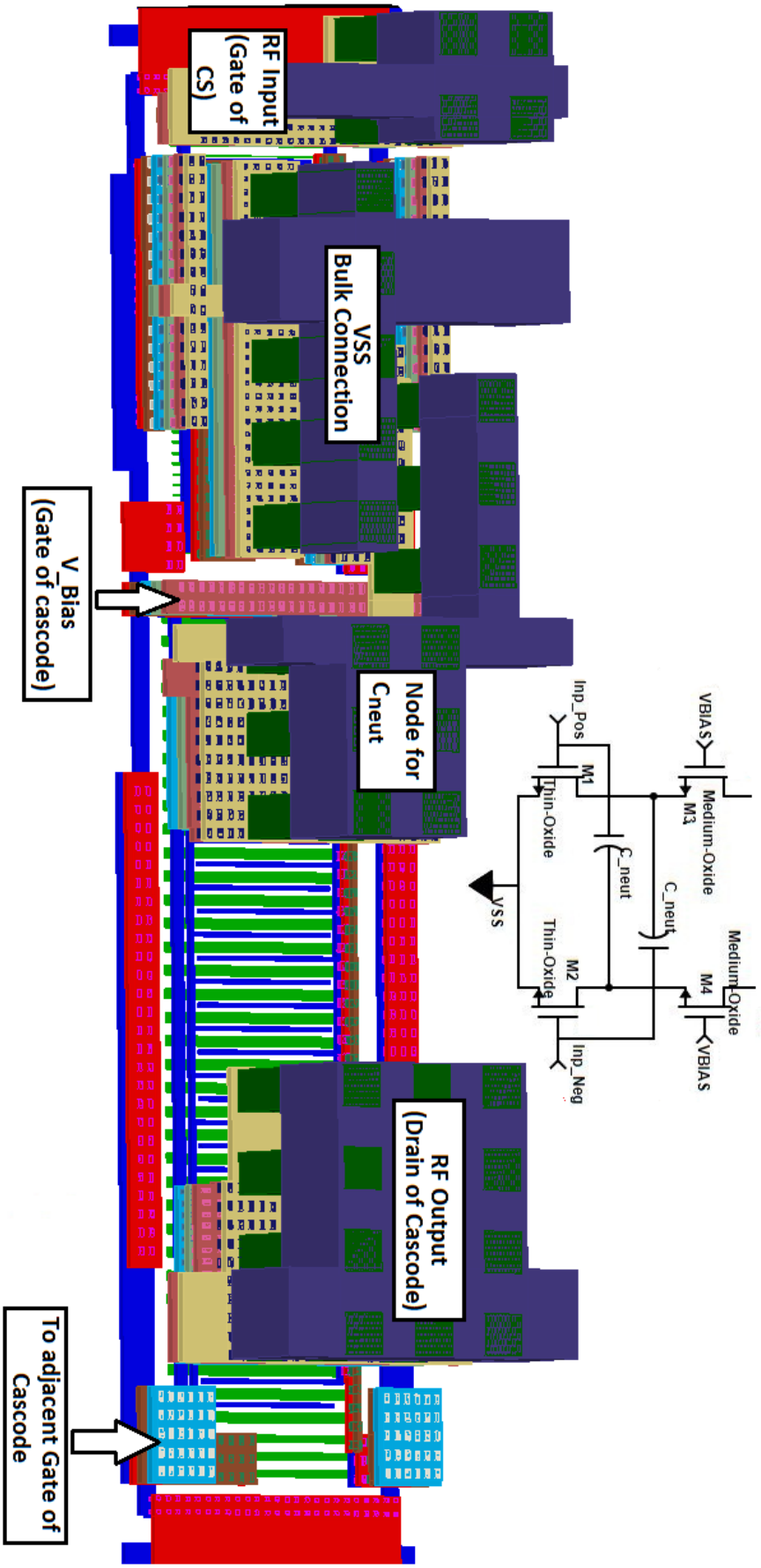


Figure 4.8: **Main Power Amplifier**

inductor size. The input balun that converts a single-ended LO carrier signal to a differential signal has been simulated in EM.

In order to reduce the simulation time, the vias were approximated by drawing a big rectangle with the same via material. This helped to drastically reduce the time of simulation without affecting the accuracy so much. Another approximation was performed where the lower metal layers from M1-M6 were removed, and the multi-layer substrate was replaced by a substrate with an equivalent dielectric constant [58]. This does not affect the result as the lower metal layers were not being used, and simplifying the substrate layer also led to faster simulations and thus more iterations were possible to optimize the layout. Figure 4.9 shows the final optimized layout of the PPF. The LO signals are marked on the left. The differential LO signal is generated by the input balun. As already mentioned before, two stages are used in the PPF to have sufficient image rejection over a wide bandwidth. Dummy capacitors and resistors have been used in order to minimize lithography edge effects.

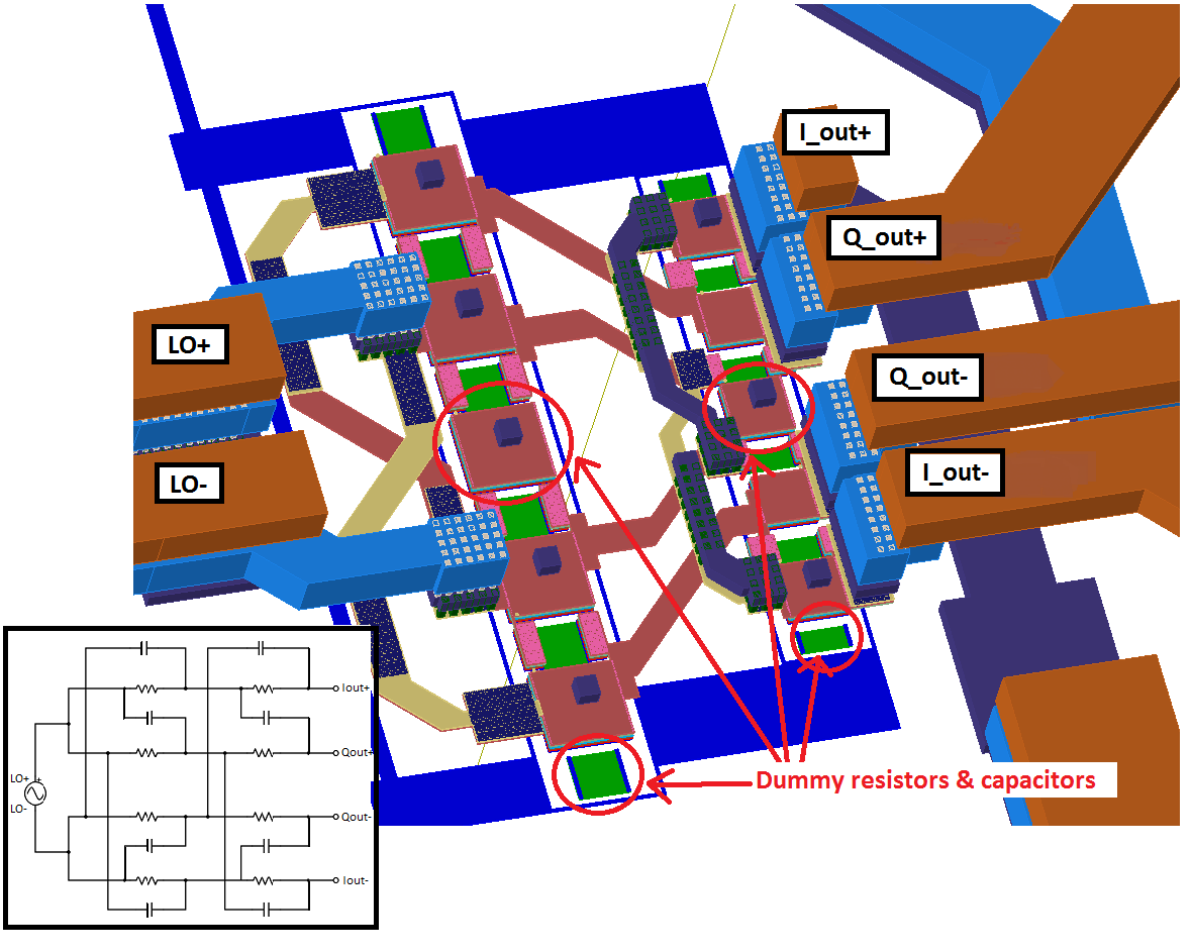


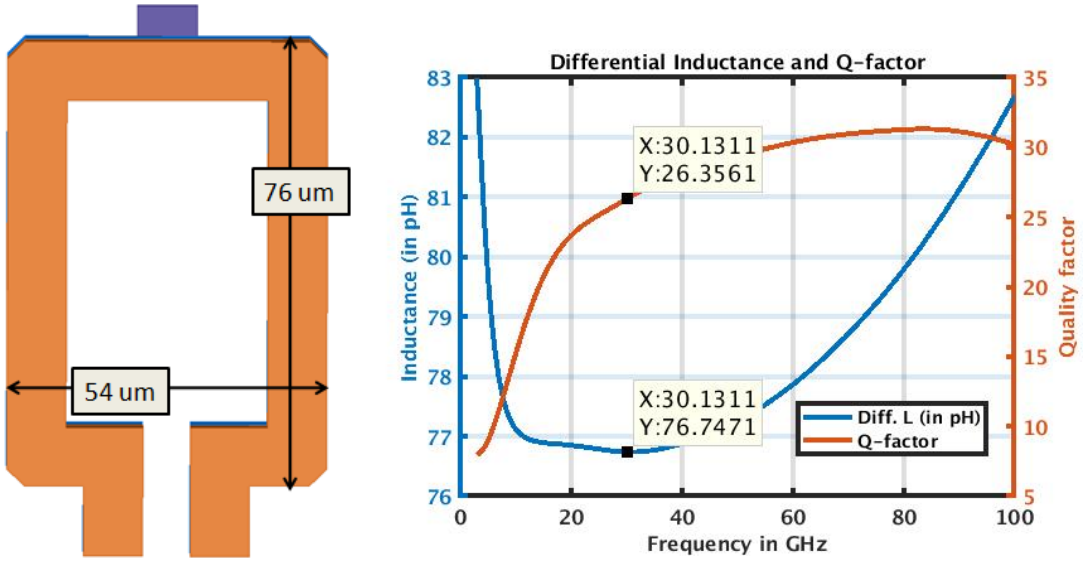
Figure 4.9: **Layout of Polyphase filter**

The routing at the output of the PPF has been optimized as much as possible. Several iterations have been performed in order to arrive at a routing scheme that minimizes the coupling capacitance between the outputs, i.e. between I_{out+}, I_{out-}, Q_{out+} and Q_{out-}. The output of the PPF is not symmetric. This is the result of

the symmetry within the PPF. This issue has already been highlighted in the previous chapter. Numerous EM simulations have been performed on the routings at the output of the PPF in order to ensure symmetry. This is essential to ensure amplitude and phase balance in all of the four quadrature outputs.

4.6. Balun and Inductors

The balun and the inductors were laid out in Cadence and then extensive EM simulations were performed to optimize the coupling coefficient, transformer efficiency, area occupied, etc. The center tapped inductors were simulated extensively along with the interconnects of the polyphase filter. The inductor size was reduced by increasing the number of turns. Later on, after the interconnect to the mixer was added, the inductor size had to be drastically reduced. This is due to the fact, that the interconnect from the input of the inductor to the input of the mixer contributed some inductance and other parasitics. Hence, ultimately, only a single turn center-tapped inductor was used. The metal track width is $10\mu m$, and the spacing between the "positive" and the "negative" side of the coil is $8\mu m$. M8 and M9 have been used to reduce the series resistance of the coil and thus minimize the losses. A BFMOAT layer is drawn to prevent p-well implants and creates a high resistive substrate. It is usually used to shield critical RF and analog lines from the substrate. The substrate losses of an inductor can be reduced by using a BFMOAT layer below the inductor. A BFMOAT layer has been used in designing the center-tapped inductor to improve the quality factor (Q) of the inductor. Figure 4.10 illustrates the center-tapped inductor along with its differential inductance and Q-factor.



(a) **Layout of the center-tapped inductor**

(b) **Differential Inductance and Q-factor**

Figure 4.10: **Properties of the Center-tapped Inductor**

An initial Balun from the "RF" library of the PDK was laid out in Cadence. The "GDS" file was exported to Momentum. EM simulations were performed in Momentum, and

the coupling coefficient (k_m) and the transformation efficiency were plotted. It was found that the k_m and the efficiency of the PDK model were not optimum. The PDK model only provided a k_m of 0.5, and the transformation efficiency was about 60%. Thus, the balun was flattened, and its size was altered. The spacing between the primary and the secondary coil was reduced. The spacing between the "positive" and the "negative" side of the primary coil at the input was also reduced. Because of DRC rules, a certain distance had to be maintained between the coils. The final balun occupies an area of 80x70 μm^2. The layout of the balun is illustrated in Figure 4.11. The metal track width is $5\mu m$, and the spacing between the coils has been kept at $1.8\mu m$. Reducing the spacing between the "positive" and the "negative" side of the input (primary coil) also enhanced the coupling and the overall transformer efficiency. This spacing has been kept at $4\mu m$. Two metal layers, M8 and M9, have been used for both of the coils. This helps to reduce the losses in the coil and thus enhances the transformer efficiency. However, the parasitic capacitance between the coils is more and, as a result, the self-resonance frequency decreases. This can be seen in the inductance plots of Figure 4.12a & 4.12b. Figure 4.12 depicts the inductance, Q-factor, coupling coefficient, and the transformation efficiency. The balun or transformation efficiency is defined by [59]:

$$\eta_{Balun} = \frac{\frac{R_L}{n^2}}{\left(\frac{\frac{\omega L_1}{Q_2} + \frac{R_L}{n^2}}{\omega k_m L_1}\right)^2 . \frac{\omega L_1}{Q_1} + \frac{\omega L_1}{Q_2} + \frac{R_L}{n^2}} \tag{4.3}$$

where L_1 is the primary winding inductance, Q_1 and Q_2 are the quality factors of the primary and the secondary coils, k_m is the coupling coefficient and n is the turns ratio. A BFMOAT layer has been drawn below the balun to minimize substrate losses and to enhance the Q factor.

After the EM simulations, the "Broadband Spice Model Generator" feature of ADS was used to generate a spice model of the balun. The spice model file was used in Cadence to perform transient simulations along with the rest of the circuit. Optimizations were made based on the results of the transient simulation.

The interconnects are a slightly different story. The spice model of the interconnects was found to be inaccurate initially. Thus, the S-parameter file was directly saved in ADS as a Touchstone file. This was then used in Cadence for performing transient simulations with the "NPORT" component available in "analogLib" of Cadence. However, if more numbers of points are used in the frequency range for EM simulation, then the spice model can closely match the S-parameter data. Nevertheless, warnings about violations of passivity and causality appear in the transient simulation. These S-parameter datas were used to simulate the SI transmitter as a whole. The simulation results depicting the overall drain and power-added efficiency has been provided in the next chapter. The efficacy of the polyphase filter in generating in-phase and quadrature phase signals has also been simulated with these S-parameter files and the results are provided in Chpater-5.

The layout of the polyphase filter along with its interconnections to the passive mixer and the center-tapped inductor has been done. Both the PA branches have been laid out, however, the interconnects to the output balun has not been simulated

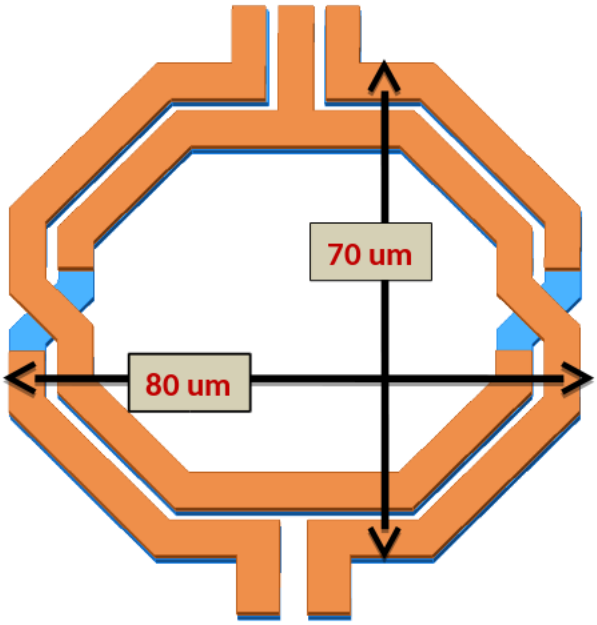


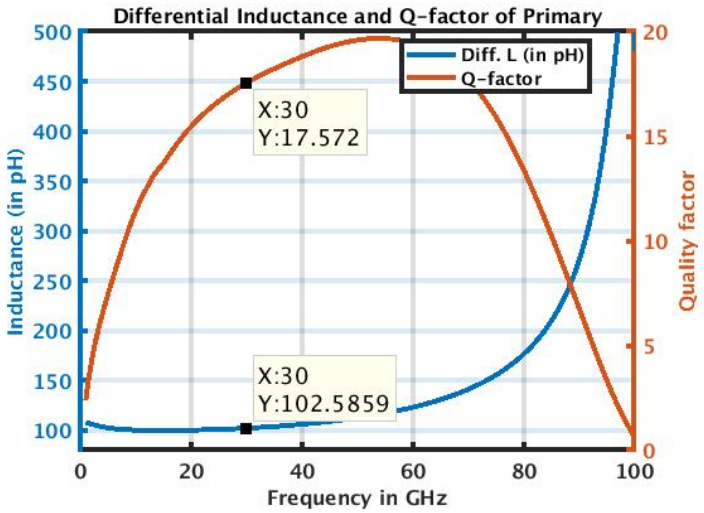
Figure 4.11: **Output Balun**

yet. As a result, the balun at the output node is not shown in Figure 4.13. Both the output and the input baluns have been laid out and EM simulations performed to verify its functionality. The driver amplifier and the power splitter has not been designed yet. The layout in its current state is illustrated in Figure 4.13.

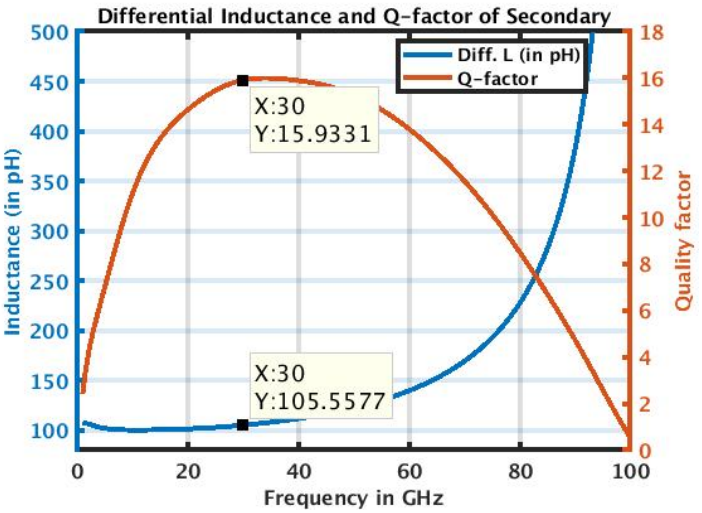
The block diagram of the chip is demonstrated in Figure 4.14. Except for the Driver amplifier (DA), power splitter, and the bond pads, the rest of the circuit blocks have been implemented and laid out. All simulation results presented in the next chapter include the parasitic capacitances of the PAs, the passive mixers and the two stage PPF extracted from the layout of each block. The parasitics contributed by the interconnects between the polyphase filter and the mixer have also been included in the simulations.

4.7. Conclusion

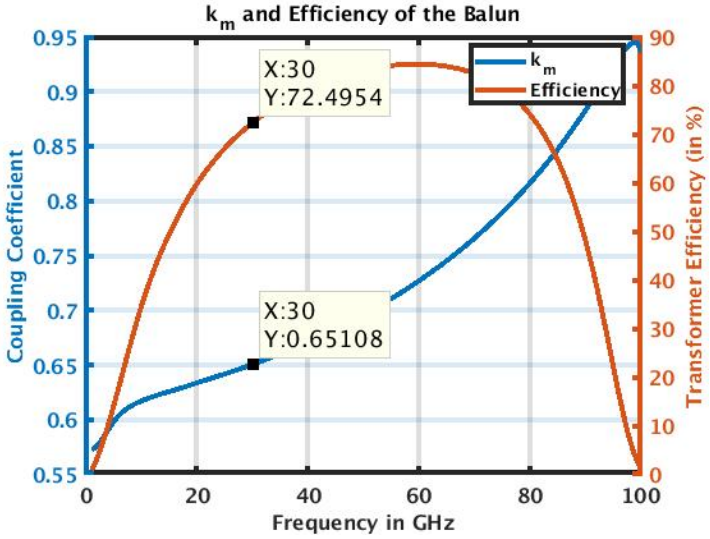
The effect of finger width, number of fingers and bias current on the maximum frequency of power gain was analyzed. The maximum transducer gain of a thin-oxide NMOS was compared to that of a medium-oxide NMOS. Based on this analysis, a finger width of 1μm was found to provide the maximum F_{max}. A brief discussion about the layout of the passive mixer has also been provided. This was followed by the layout of the neutralization capacitor. Both stability and F_{max} were taken into account in order to determine the suitable transistor size for the PA. Although, it is beneficial to maximize F_{max}, however, this also increases the tendency of the amplifier to oscillate as it becomes conditionally stable over a larger frequency range. As a result, a compromise was made between the power gain of the amplifier and its stability. The complete layout of an unit cell of the $Main$ branch has been provided and methods to reduce its gate resistance has been pointed out. The layout considerations of a two-stage polyphase filter has been discussed as well. Finally, a center-tapped inductor and a



(a) **Differential Inductance and Q-factor of the Primary coil**



(b) **Differential Inductance and Q-factor of the Secondary coil**



(c) **Coupling coefficient and efficiency of the Balun**

Figure 4.12: **Properties of the Balun**

balun was designed and simulated in Momentum to optimize its performance. Ideally a coupling coefficient of 0.707 is desired in order to provide the optimum load to the PA. A coupling coefficient of 0.65 was achieved along with a transformation efficiency of about 72.5% for the balun. A higher quality factor of the passives is greatly desired at mm-wave operation. This minimizes the losses incurred in the passive network. As a result, considerable efforts have been made to maximize the quality factor of the inductor and the transformation efficiency of the balun.

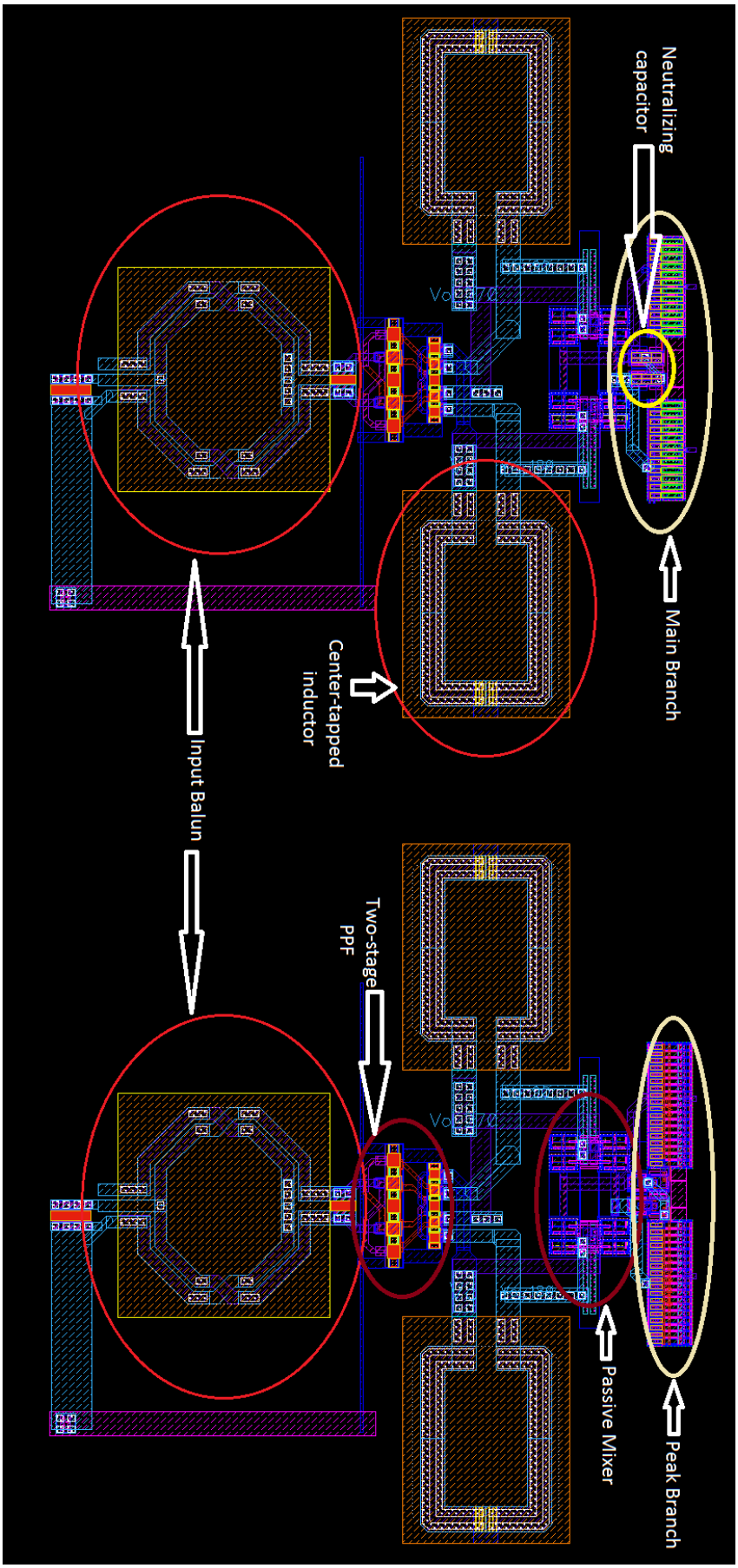


Figure 4.13: **Layout of the chip**

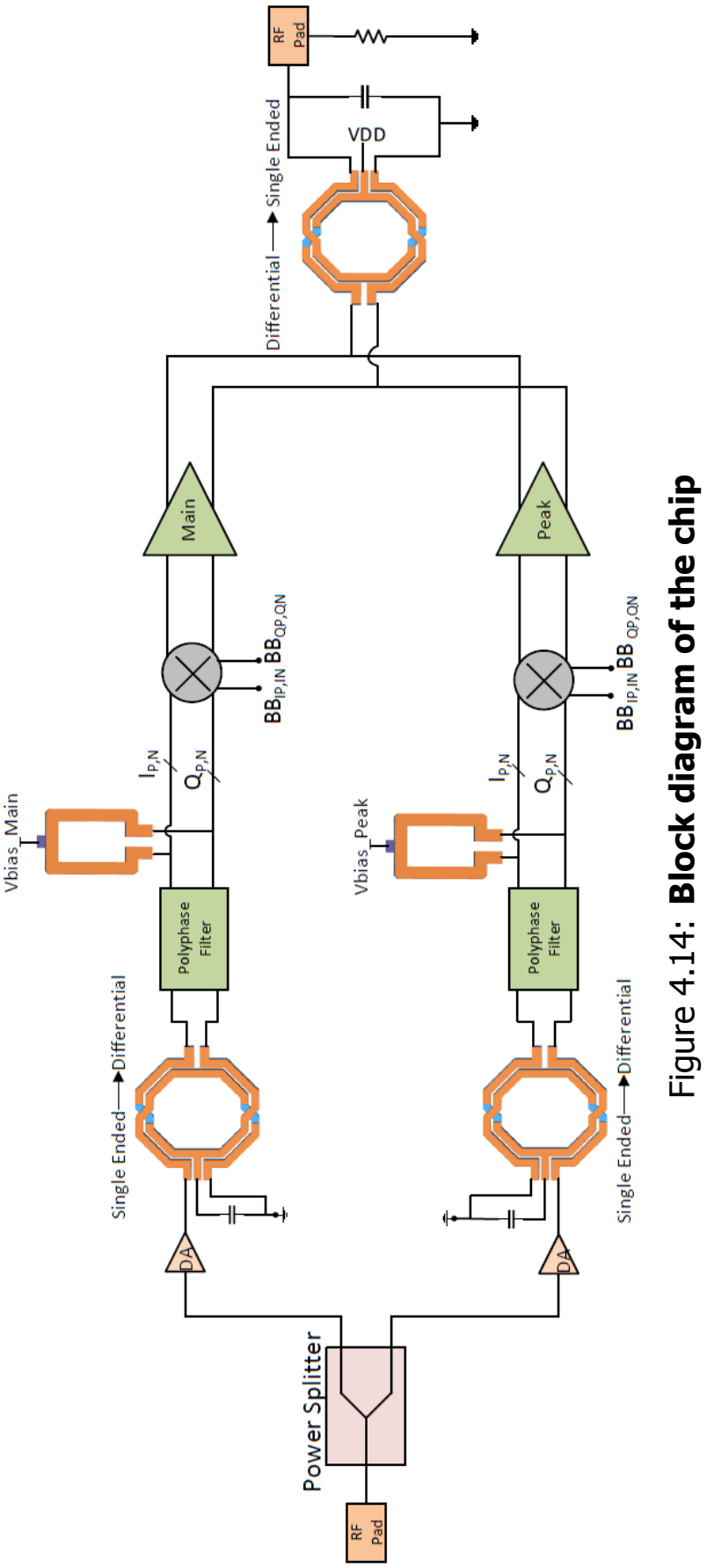


Figure 4.14: **Block diagram of the chip**

5

Simulation Results

This chapter presents the simulation results of the Supply Interpolation transmitter. The properties of individual circuit components like the inductor and the balun have already been presented in the last chapter. In this chapter, the performance results of the two-stage polyphase filter have been provided. This is followed by the output power and efficiency plots of the amplifier branches. A brief description has also been provided about two different switching schemes. All simulation results include the parasitic capacitances extracted after the layout of each block.

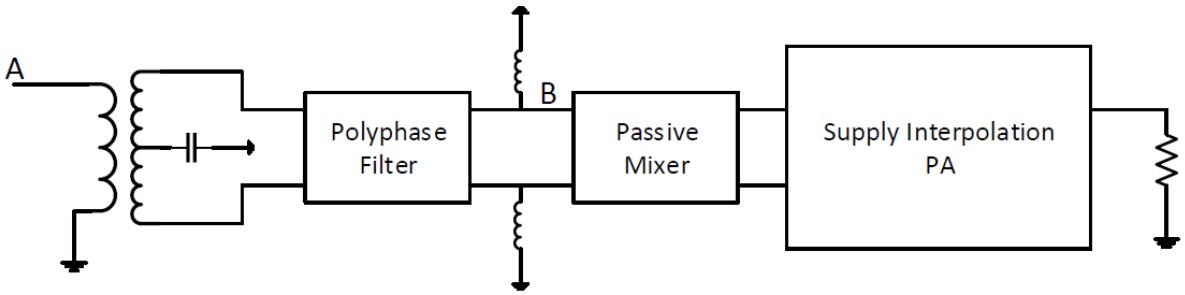
5.1. Polyphase Filter

The amount of phase and amplitude imbalance introduced by a PPF is quantized by the image rejection ratio (IRR). The IRR is defined as the ratio of the power of the image to that of the desired tone and can be expressed in terms of the amplitude and phase imbalance by [60]

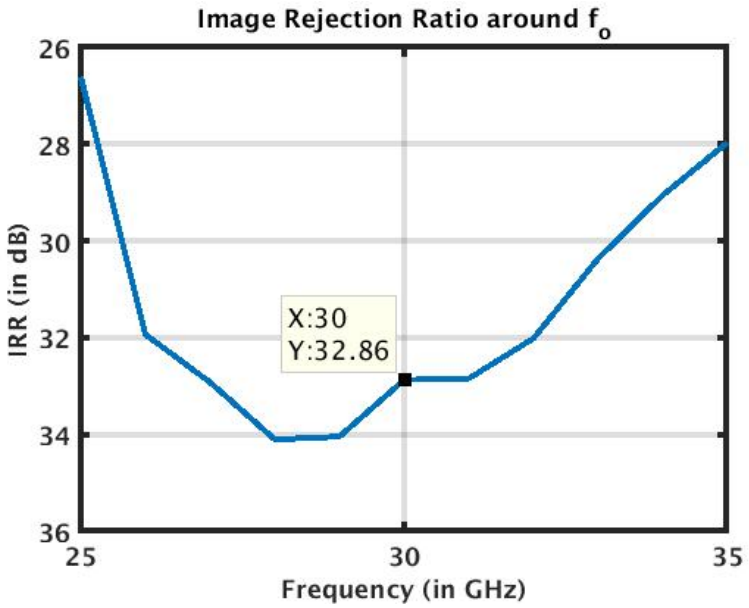
$$IRR = \frac{1 + 2A_{bal}\cos\Delta\theta + A_{bal}^2}{1 - 2A_{bal}\cos\Delta\theta + A_{bal}^2} \tag{5.1}$$

where A_{bal} represents the amplitude of the I and Q signals and $\Delta\theta$ represents the phase deviation between the I and Q branches. As already mentioned, a two stage PPF has been implemented in this work. Each stage is tuned to a different frequency which helps to achieve proper IRR over a wide bandwidth. Figure 5.1b depicts the IRR over a range of frequencies. It is evident that a two stage PPF can provide decent IRR over almost 6 GHz. The simulation temperature has been set at 90^{o}C and the parasitics of the PPF, and of the interconnects connecting it to the mixer have been included in this simulation. The parasitic capacitances of the mixer have also been taken into account.

Figure 5.1a shows a simplified structure of the SI Tx along with the nodes A and B. The signal power loss starting from the input (i.e., the single ended clock at Node A) provided before the balun to the input of the IQ passive mixer (Node B) is about $\approx$ 3.9dB. The power of the differential I and Q signals has been included in this calculation.



(a) **Simple structure of SI Transmitter**



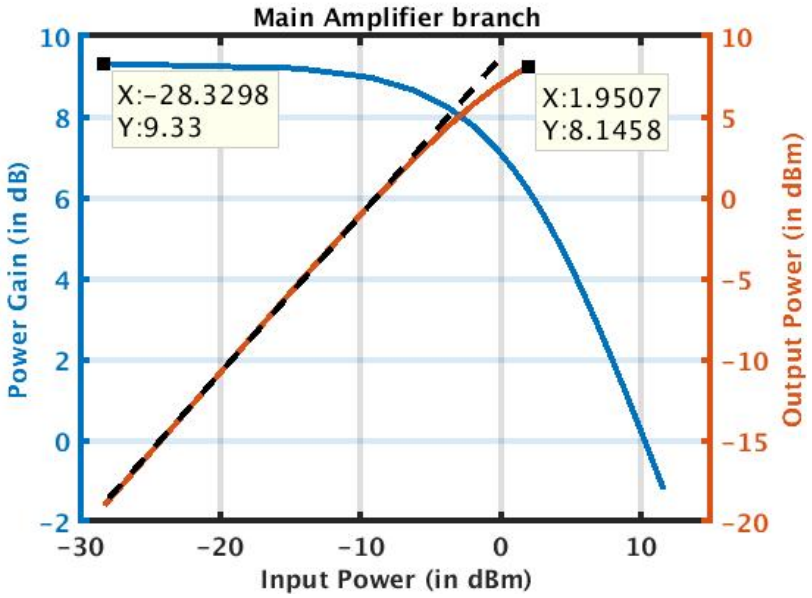
(b) **IRR at different frequency of operation for a 2-stage PPF**

Figure 5.1: **SI Transmitter and IRR of PPF**

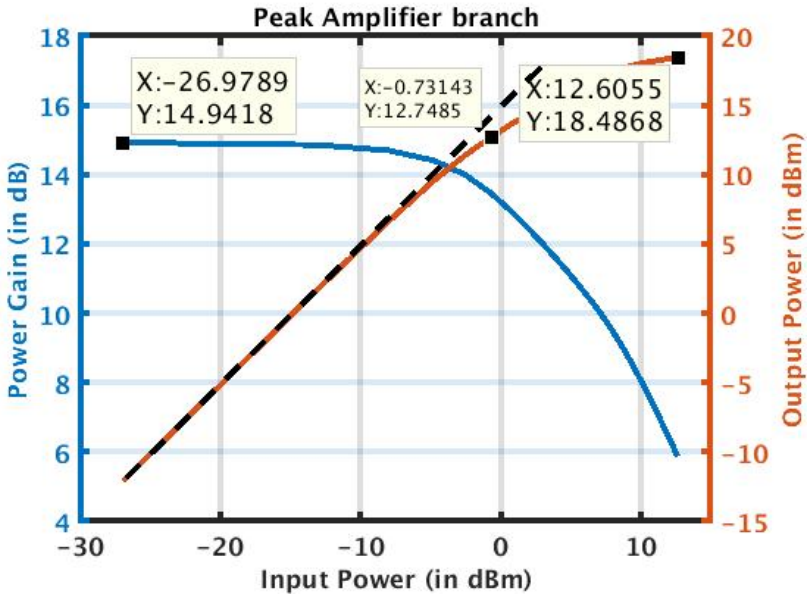
5.2. Supply Interpolating PA

The efficiency and output power are the most important parameters that determine the efficacy of a power amplifier. The P_{1dB} point is used to determine the maximum "linear" output power available from a PA. A SI transmitter consists of two amplifier branches, i.e., the $Main$ and the $Peak$ branch. The $Main$ branch is used to generate output powers less than 8dBm while the $Peak$ branch is operational in the range of 8-18dBm of output power. The P_{1dB} for the $Main$ branch lies at 7dBm while that of $Peak$ branch lies at 13dBm. Figure 5.2 shows the Power gain and output power for the $Main$ and $Peak$ amplifier branch, respectively.

Figure 5.3 demonstrates the peak drain efficiency and power-added-efficiency for the "Supply Interpolating" transmitter. The dashed red curve denotes the drain efficiency without any efficiency enhancement technique. As a result, it can be concluded that using "Supply Interpolation" yields more than 10% higher PAE and almost 15%



(a) **Main Branch**



(b) **Peak Branch**

Figure 5.2: **Output power and Power Gain of SI Transmitter**

higher drain efficiency compared to a conventional system without any efficiency enhancement technique. Efficiency enhancement has been achieved at ≈ 8.6dB power back-off. The slightly inferior power gain of the $Main$ branch leads to the difference in the PAE and drain efficiency curve at power back-off. This is mainly because of the under-estimation of the parasitic capacitances of the $Main$ branch. Resizing the $Main$ branch alleviates this issue, and significant improvement in power gain has been observed. Enhancing the power gain for the $Main$ branch also leads to an improvement in the PAE curve which then closely follows the peak drain efficiency curve. As a result, the efficiency enhancement at power back-off can be improved by about 15%.

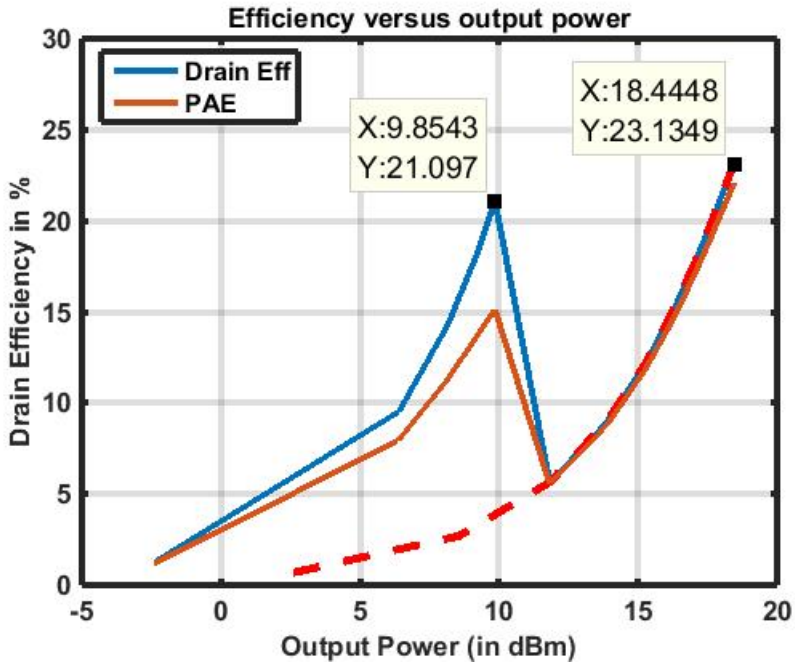


Figure 5.3: **Drain Efficiency and PAE of SI Tx**

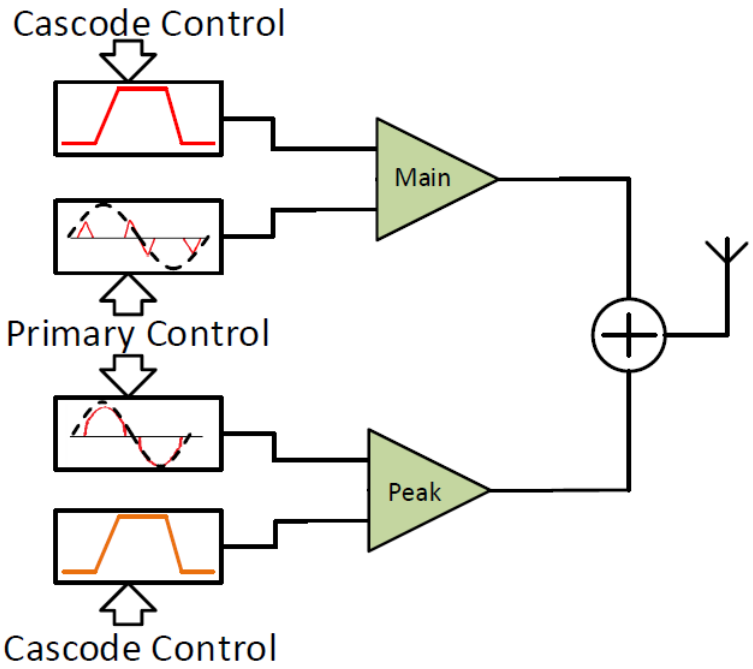
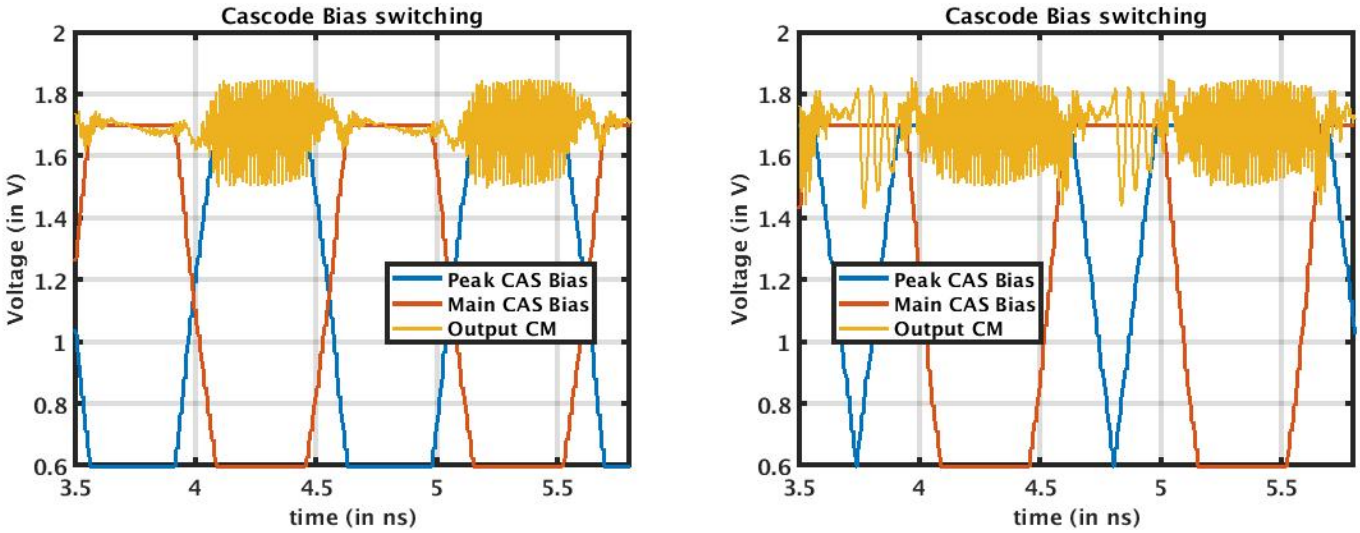


Figure 5.4: **SI Tx with control signals**

A MATLAB script is used to generate the control signal that automatically turns on and off the appropriate amplifier branch depending on the envelope signal. Figure 5.4 demonstrates the different control signals available in SI Tx. Both the bias of the cascode and the baseband input signal is controlled to ensure a smooth transition between the two amplifier branches. Due to large output common mode transients

during the switching, the bias of the cascode has been turned on and off steadily. This smoothens out the output common mode transients. However, a compromise between output common mode oscillations and linearity has been made. Two types of control profile has been investigated. The symmetric control profile has equal rise and fall times for the cascode bias switching (Figure 5.5a). On the othe hand, the overlap control profile has a higher on time for the $Peak$ branch. The bias of the cascode device in the $Peak$ branch starts to decrease only after the bias voltage of the cascode in the $Main$ branch has reached its maximum value (Figure 5.5b).



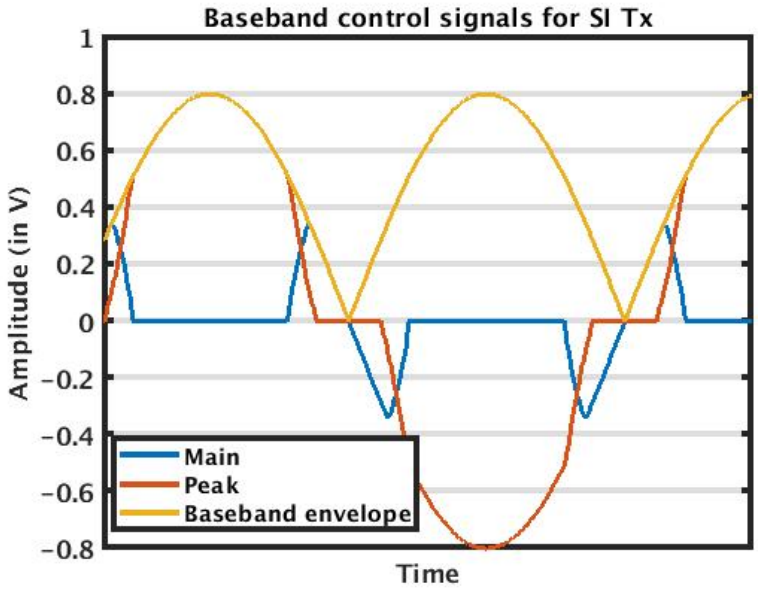
(a) **Symmetric Control Profile** (b) **Overlap Control Profile**

Figure 5.5: **Different control profiles**

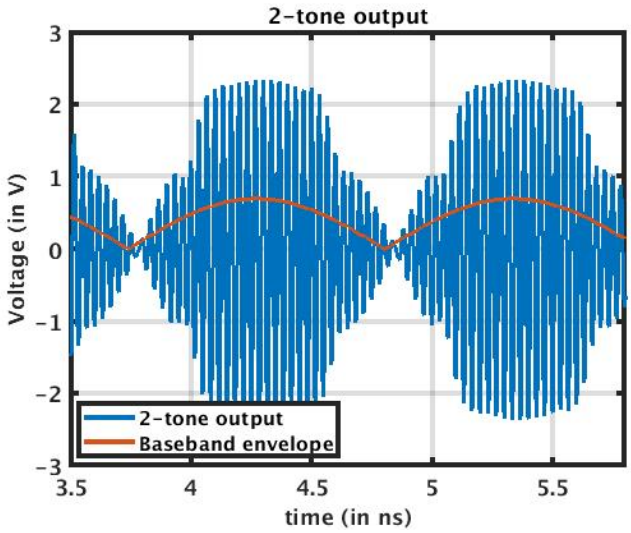
The baseband signal is split into two parts, one for each branch. Figure 5.6a demonstrates the envelope signal and the baseband signal for $Main$ and $Peak$. As already stated, the $Main$ is operated when the input is small (<0.35V). Beyond that, the $Peak$ branch is used. An overlap between the baseband signals for the $Main$ and the $Peak$ has been maintained to ensure smooth transition between the two branches.

The output common mode voltage oscillations can be suppressed effectively by using a symmetric control for the cascode bias. However, because the cascode bias drops off equally for both the $Peak$ and the $Main$ branch, the output generated during this transition period changes rapidly. This leads to non-linearity in the output which can be observed as steep changes in the output voltage during the transition period (Figure 5.6b). The power spectral density of the output signal also contains a lot of intermodulation products. However, when the transition between the $Main$ and the $Peak$ branch is more gradual, the output common mode oscillations are more pronounced as both the branches are working simultaneously from two different supply voltages (Figure 5.5b). Nevertheless, this ensures that the output doesn't change drastically during the transition period. Figure 5.6c illustrates the smooth transition between the branches when using a overlap controlling profile. The rise and fall time of the control signal is about 146.5ps for $Main$, while it is 185ps for $Peak$.

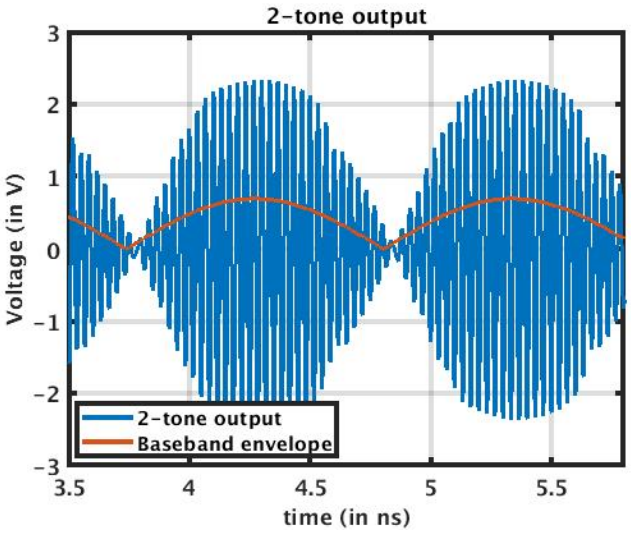
The superior behaviour of the overlap control profile can be ascertained from the frequency domain plot of the output signal. From Figure 5.7, we can state that the



(a) **Baseband signal**



(b) **Output with symmetric control**

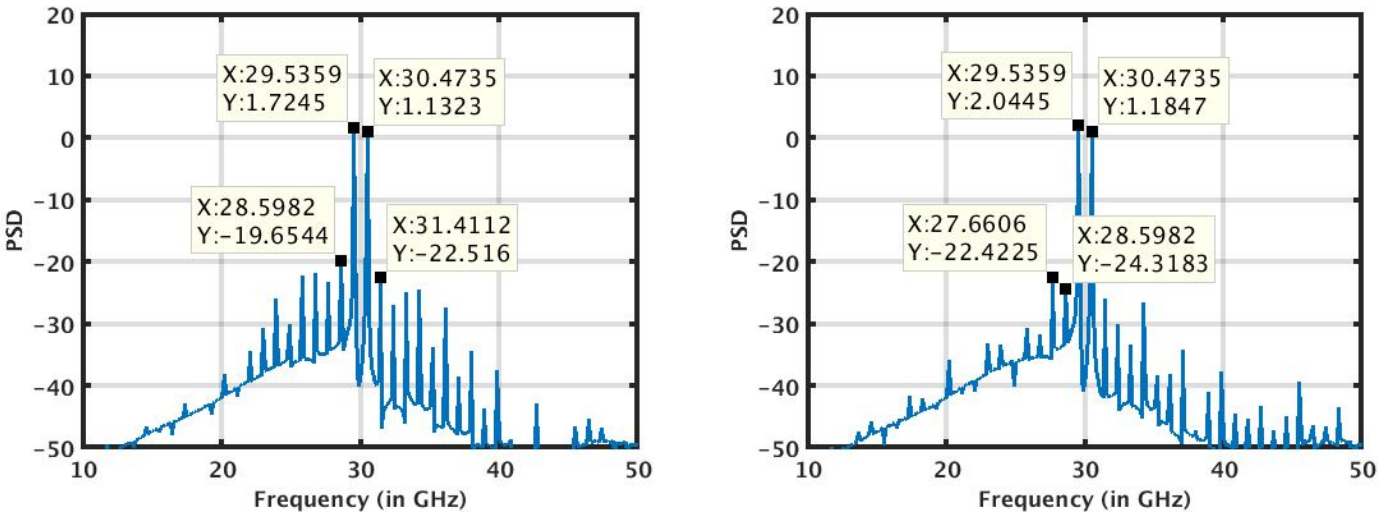


(c) **Output with overlap control**

Figure 5.6: **Different control profiles**

overlap signalling leads to a more linear operation. There are less intermodulation products when overlap signalling is used. The third harmonic is -30dBc below the fundamental. More than 25dB of IM3 (Figure 5.7b) can be obtained when using overlap signalling. However, this is only about 20dB when the symmetric control profile is used (Figure 5.7a).

A switching threshold of 0.3V has been used. When the maximum value of the envelope is $<= 0.3$V, then, only the $Main$ branch is in operation. The bias of the cascode in the $Peak$ is set to low (0.6V). Both the branches are in operation when the envelope is greater than the switching threshold of 0.3V. The threshold value is chosen such that the $Main$ branch is not under compression. An IM3 of about 21dB



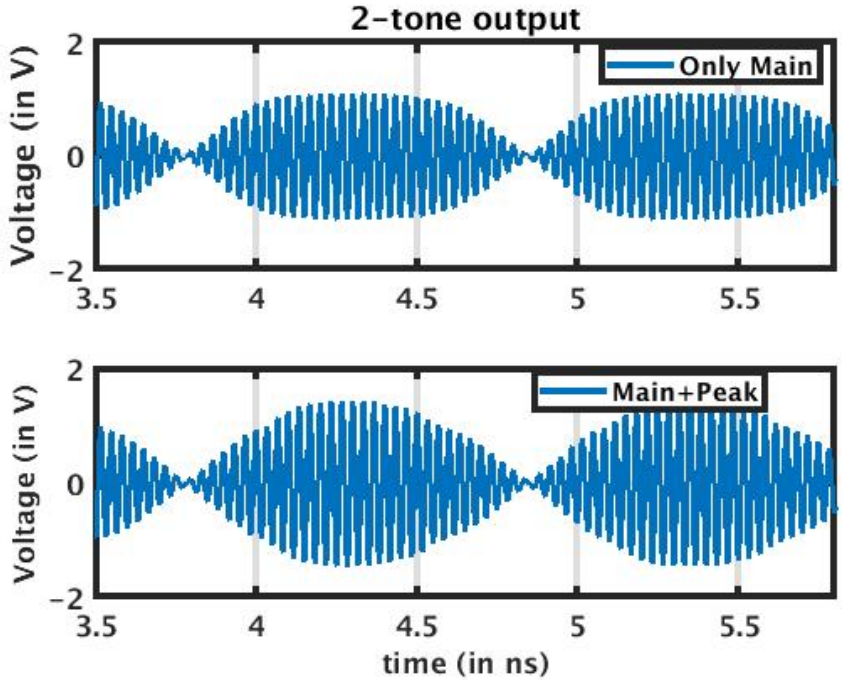
(a) **Output with symmetric control** (b) **Output with overlap control**

Figure 5.7: **Output PSD with different controls**

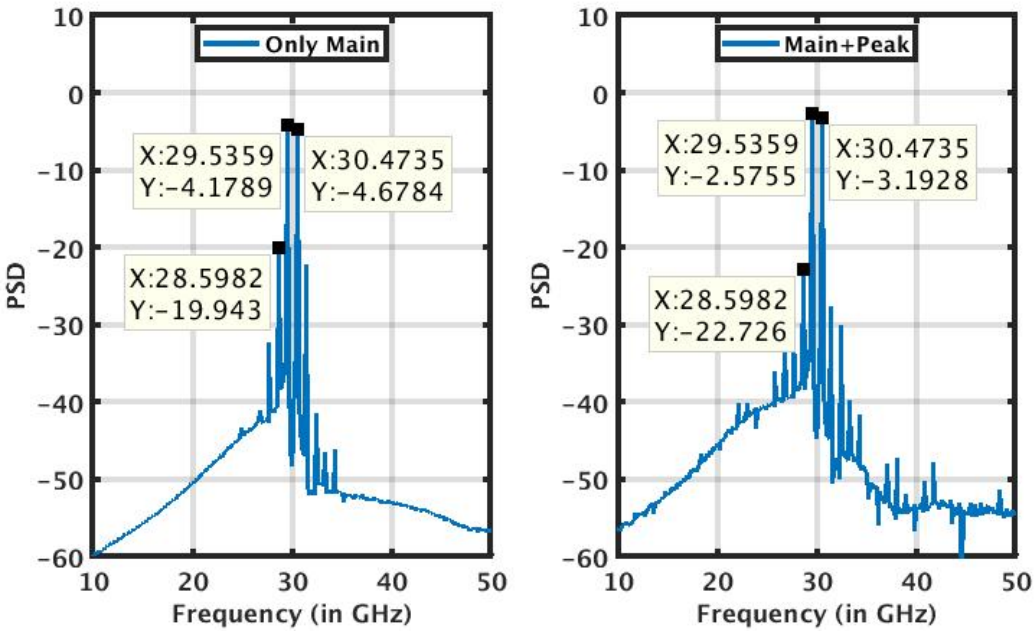
is observed for an envelope $<= 0.3$V. Any attempt to increase the threshold value degrades the linearity of the overall system. This is mainly because of the $Main$ PA working in saturation.

Figure 5.8 demonstrates how the linearity of the system gets degraded when the $Main$ is operating in saturation for a higher switching threshold of 0.4V. For this simulation, a baseband signal with an amplitude (or envelope) of 0.4V was provided. The switching threshold was set to 0.4V initially. Since the envelope lies within the threshold, only the $Main$ branch is used. The $Peak$ is in off state. The top subplot of Figure 5.8a depicts the output signal when only the $Main$ branch is operational. The saturated behaviour of the $Main$ branch can be observed. Next, the switching threshold was reduced to 0.3V, such that both the $Main$ and the $Peak$ branch is used. The bottom subplot of Figure 5.8a illustrates the output waveform for this case. The frequency domain plots also convey the same information (Figure 5.8b). Due to compression, the IM3 of the system is only about 15dB when only the $Main$ branch is used for a baseband signal with an envelope of 0.4V. However, by using the $Main$ and the $Peak$ branches judiciously, to provide output in the low and high power region, respectively, >19.5dB of IM3 can be achieved.

A 16-QAM signal with a bandwidth of 1.1GHz, has been used to evaluate the spectral purity of the SI TX; this has been illustrated in Figure 5.10. The spectral purity is better than 24dB (Figure 5.10a). The EVM is better than -18dB for peak output power of 18.25dBm with a PAPR of 7.8dB. Figure 5.10b demonstrates the constellation diagram for this scenario. At 4dB PBO, the spectral purity is better than 29dB (Figure 5.11a) and exhibits an EVM of -20.59dBc (Figure 5.11b). The output balun rejects the 2^{nd} harmonic efficiently, which can be confirmed from Figure 5.9. The 2^{nd} harmonic is 60dB below the fundamental. The same plot also illustrates that the 3^{rd} harmonic is more than 40dB below the fundamental. Figure 5.12 elucidates the relationship between EVM and PBO. The EVM decreases with the increase in the peak-to-average power ratio.



(a) **Time-domain output for different switching thresholds**



(b) **Frequency-domain output for different switching thresholds**

Figure 5.8: **Linearity for different thresholds**

Table 5.1 summarizes the performance of the proposed "Supply Interpolating Transmitter". This work can be compared to [2], [9] and [5]. Compared to this work, all the other works use a multi-stage design. A multiple stage proves to be more beneficial in terms of overall efficiency and P_{1dB}. Although [5] can handle slightly larger RF bandwidth, it requires the use of a bulky power combiner. More than 20dB of

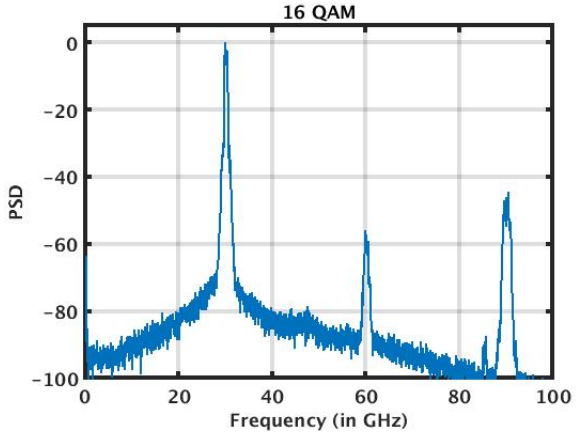
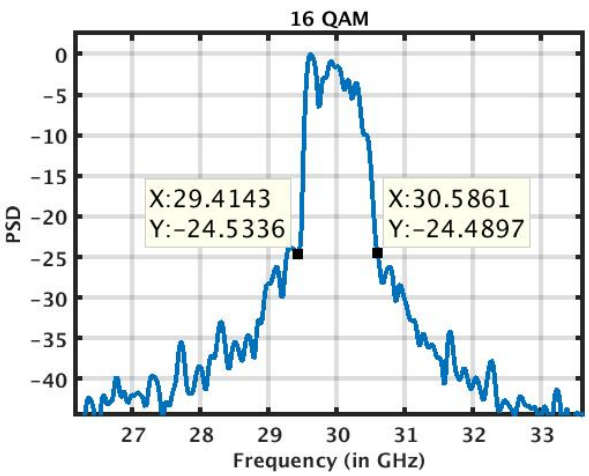
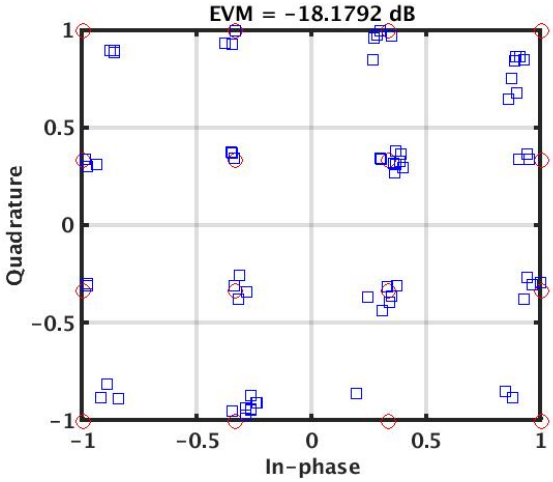


Figure 5.9: **Frequency-domain output for 16-QAM**



(a) **Frequency-domain output for 16-QAM**



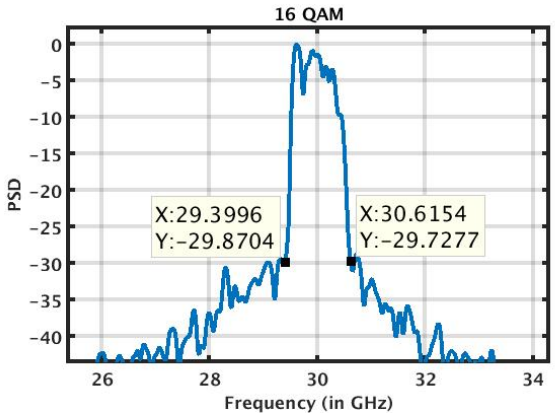
(b) **EVM for 16-QAM at full power**

Figure 5.10: **Linearity of SI Tx with 16-QAM signal at full power**

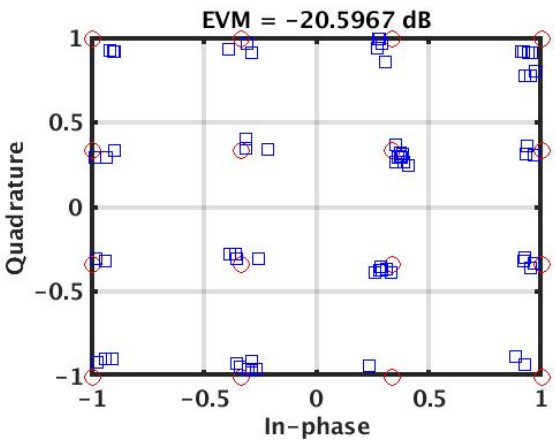
power gain has been reported by the other works, however, this is the power gain of the entire system (i.e., of the two/three stages, depending on the design). Compared to that, the power gain reported in this, from a single stage is about 15dB. The difference of 2dB in the P_{1dB} point can be attributed to the multi-stage designs followed by the other works. No VGAs and power combining networks are necessary in this work.

5.3. Conclusion

In this chapter, the post layout simulation results of the major circuit blocks have been provided. At first, the image rejection ratio of the two-stage polyphase filter was demonstrated. An IRR greater than 30dB has been achieved in the 26-33GHz range. Next, the power gain and the AM-AM plots of both the $Main$ and the $Peak$ branch were illustrated. The maximum power gain of 9.3dB and 14.9dB has been achieved along with P_{1dB} of 7dBm and 12.7dBm for the $Main$ and the $Peak$ branch respectively. About 15% enhancement in PAE has been achieved at 8.6dB power



(a) **Frequency-domain output for 16-QAM after zooming in**



(b) **EVM for 16-QAM at 4dB PBO**

Figure 5.11: **Linearity of SI Tx with 16-QAM signal at 4dB PBO**

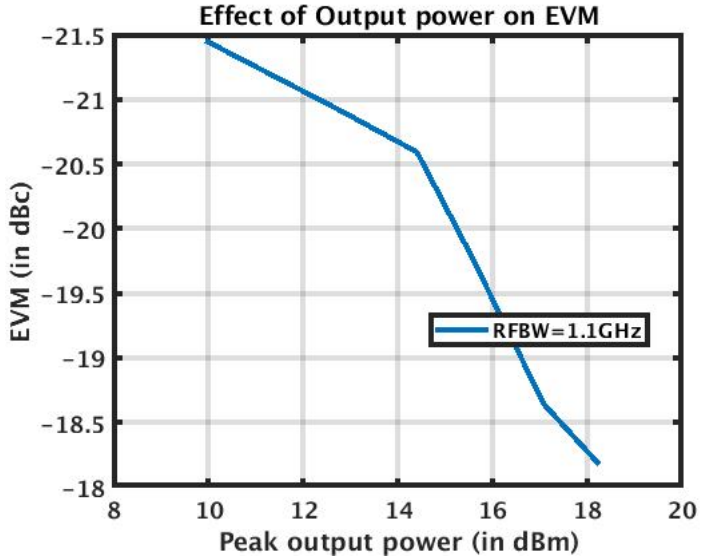


Figure 5.12: **Effect of Power Back-off on EVM**

back-off. A peak drain efficiency of 23% can be achieved at 18.4dBm output power. The PAE is about 15% at PBO and this is mainly due to the slightly inferior power gain of the $Main$ branch. Two different control profiles have been studied in this work. The overlap control profile demonstrates significant improvement in linearity as compared to that of the symmetric control profile. An IM3 of 25dB (Figure 5.7b) can be achieved with the overlap controlling. A suitable threshold has also been set up in order to prevent unnecessary switching between the branches at low power and also to avoid compressing the $Main$ branch. Simulation results have been provided to verify this fact. For an envelope with a maximum of 0.4V, the IM3 is almost 5dB better when both the $Main$ and the $Peak$ branch is used as compared to when only the $Main$ is used. This is because the $Main$ starts to compress at that power and as a result it is prudent to use the $Peak$ branch when the envelope is greater than the threshold voltage. Hence, a threshold voltage of 0.3V has been chosen to maintain

a balance between unnecessary switching at low power and the compression of the $Main$ branch. An EVM of -20.59dBc has been achieved for a 16-QAM signal with a RFBW of 1.1GHz, delivering a peak output power of 14.39dB with a PAPR of 9.18dB.

	This Work	[1]	[2]	[9]	[5]	[61]
Technology	22nm FDSOI	28nm CMOS	40nm CMOS	130nm SiGe	40nm CMOS	22nm FDSOI
Supply [V]	0.9, 1.7	1.15	1.1	1.5	1	1.3
Frequency (GHz)	30	30	27	28 / 37 / 39	27	76-81
RF BW (MHz)	1100	250	800	500	1200	5000
No. of stages	1	2	3	2	2	2
Gain [dB]	9.33, 14.9	16.3	22.4	18.2 / 17.1 / 16.6	20.5	NA
P_{1dB} [dBm]	7, 13	14.3	13.7	15.2 / 15.5 / 15.4	16.8	NA
P_{SAT} [dBm]	9.85, 18.44	15.3	15.1	16.8 / 17.1 / 16.6	18.1	9.5
Drain Eff [%]	21, 23	NA	NA	29.4 / 27.6 / 28.2	NA	NA
PAE_{1dB} [%]	15, 22	35.8	31.1	19.5 / 21.6 / 20.7	37.5	10^{a}
EVM^{b} [dBc]	-20.59^{c}	-25	-25	-27 / -30.3 / -28.7	-25	NA
P_{out} (dBm) @ EVM	5.22	5.3	6.7	9.2 / 9.5 / 9.3	8.4	NA
PAE [%] @ EVM	7	9.6	11	NA	8.8	NA
PA Topology	Neutralized cascode with Main and Peak branch	Neutralized CS	Neutralized CS with a VGA as 1^{st} stage	Doherty with transformer based power combiner	Neutralized CS with power combiner	Injection locked PA followed by a cascode PA

[a] transmitter efficiency [b] normalized to reference RMS power [c] 16-QAM

Table 5.1: Comparison of mm-wave transmitters

6

Conclusion

Modern communication systems call for significant changes in the transmitter architecture. In order to address hundreds of MHz of bandwidth, a linear transmitter is necessary. The power efficiency of the system needs to be improved as well, so as not to quickly drain the device's battery. Tens of milliwatts need to be generated in order to fulfil the link budget requirements and, finally, parameters like adjacent channel leakage and EVM need to be considered as well. Due to the incursion of massive MIMO, load modulation techniques will slowly lose their appeal. As a result, a novel efficiency enhancement technique has been investigated in this work.

A "Supply Interpolating" wideband transmitter has been designed in this work. Two different amplifier branches have been used, however, the application of this concept is not limited by the number of amplifier branches. In theory, multiple amplifier branches can be added to generate several efficiency peaks at power back-off.

6.1. Thesis Outcome

Polar architecture is widely used in transmitters around the world. However, with the advent of the 5G communication system, such an architecture is bound to become obsolete. Polar transmitters lead to huge bandwidth expansion due to the nonlinear operation of converting from cartesian to polar coordinates. 5G transmitters need to handle about 800MHz of instantaneous bandwidth, thus bandwidth expansion due to polar operation degrades the linearity of the system and leads to a poor adjacent channel leakage ratio (ACLR). Consequently, an IQ operation is preferred. Therefore, an IQ transmitter has been proposed in this work. In order to limit the workload that is suitable for a Masters' thesis project, the baseband IQ signal has been assumed to be generated off-chip. However, the IQ LO signal has been generated on-chip with the help of a two-stage polyphase filter. The PPF architecture proposed in [54] has been adopted in this work due to its symmetric layout and reduced parasitic inductances that are associated with the interconnects. IRR better than 30dB has been achieved in the 26-33GHz range. The interconnects were all EM simulated in order to resonate out any parasitic inductances associated with the interconnects. The center-tapped inductor has been used to resonate out the parasitic inductances and also to provide DC bias to the PA input.

The upconversion of the baseband signal to the carrier frequency has been performed with a passive mixer. The benefit of using a passive mixer is that it consumes very less power (due to the parasitic capacitors). As a result, it is suitable for mm-wave applications when power efficiency is inadvertently low. However, instead of using a single MOS based switch in the passive mixer, a pass gate based passive mixer has been implemented. The benefits of using pass gate has already been outlined. The input capacitance of the passive mixer also puts a certain limit on the inductor value that is used to resonate it out. The losses in the inductor can be modelled by a parallel resistance. This resistance thus appears at the input of the mixer. As already stated in Chapter 3, in order to minimize the signal loss through the PPF, the resistance of each succeeding stage has been doubled. Consequently, the input impedance of the two-stage PPF is low ($15 - 15j$). This necessitates the use of a driver amplifier and possibly a buffer in order to drive the PPF.

Both the amplifier branches have been laid out and simulated with parasitic capacitances. A peak PAE $> 22\%$ and 15% has been achieved at 18.4dBm and 9.8dBm output power. By the virtue of "Supply Interpolation", efficiency peaks at power back-off of 8.6dB have also been demonstrated. Small signal power gains of 9.33dB and 14.9dB have been reported along with P_{1dB} of 7dbm and 13dBm for the $Main$ and the $Peak$ amplifier branches, respectively.

The overlap control profile allows for a smoother handover between the amplifier branches. An IM3 of 25dB and IM5 of 23.5dB (Figure 5.7b) can be achieved with such a controlling scheme. A threshold voltage of 0.3V has been chosen to minimize unnecessary switching in the low power region. The spectral purity of > 24dB has been achieved with a 16-QAM signal delivering the peak output power of 18.25dBm. An EVM of -20.59dBc has been achieved with a 1.1GHz wide 16-QAM signal. The peak output power delivered in this case is 14.39dB with a PAPR of 9.18dB.

The input power splitter, LO driver amplifier, IO pads and the electrical connection from the drain of the PAs to the input of the output balun have not been implemented yet. The design will be tape-out ready after the aforementioned blocks are laid out.

6.2. Suggestions for Future Work

In order to keep the workload manageable, the baseband signal generation has not been looked into in this work. Future prototypes can address this shortcoming. The losses in the interconnects and the low input impedance of the PPF was not estimated correctly at the onset. Hence, future efforts can be directed in this direction to optimize the design such that a driver amplifier becomes redundant. Another possibility for future work lies in using three or more amplifier branches to enhance the efficiency at deep power back-off. The power back-off requirement continues to rise due to the use of highly spectral efficient modulation techniques. As a result, the use of multiple amplifier branches will provide more efficiency in deep power back-off regions. The input power splitter has not been implemented in this design. The power gain of the $Main$ branch is slightly low and also falls off rapidly. Design optimizations are necessary to ensure proper power gain and to improve the PAE at power back-off. Because of massive MIMO, the antenna load impedance will vary dynamically. Although, an SI transmitter is inherently insensitive to load modulation, additional antenna tuning circuitry must be employed for large VSWR (3:1) scenarios. Such a tuning circuit can

translate the load reflection coefficients (Γ_L) to the origin of the Smith chart [44]. This also provides for an interesting research question.

Bibliography

[1] S. Shakib, H. Park, J. Dunworth, V. Aparin, and K. Entesari, *A 28ghz efficient linear power amplifier for 5g phased arrays in 28nm bulk cmos,* in *2016 IEEE International Solid-State Circuits Conference (ISSCC)* (2016) pp. 352–353.

[2] S. Shakib, M. Elkholy, J. Dunworth, V. Aparin, and K. Entesari, *A wideband 28ghz power amplifier supporting 8×100mhz carrier aggregation for 5g in 40nm cmos,* in *2017 IEEE International Solid-State Circuits Conference (ISSCC)* (2017) pp. 44–45.

[3] A. Sarkar and B. Floyd, *A 28-ghz class-j power amplifier with 18-dbm output power and 35% peak pae in 120-nm sige bicmos,* in *2014 IEEE 14th Topical Meeting on Silicon Monolithic Integrated Circuits in Rf Systems* (2014) pp. 71–73.

[4] K. Dasgupta, S. Daneshgar, C. Thakkar, K. Datta, J. Jaussi, and B. Casper, *A 25 gb/s 60 ghz digital power amplifier in 28nm cmos,* in *ESSCIRC 2017 - 43rd IEEE European Solid State Circuits Conference* (2017) pp. 207–210.

[5] Y. Zhang and P. Reynaert, *A high-efficiency linear power amplifier for 28ghz mobile communications in 40nm cmos,* in *2017 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)* (2017) pp. 33–36.

[6] C. Florian, T. Cappello, R. P. Paganelli, D. Niessen, and F. Filicori, *Envelope tracking of an rf high power amplifier with an 8-level digitally controlled gan-on-si supply modulator,* IEEE Transactions on Microwave Theory and Techniques **63**, 2589 (2015).

[7] C. Lu, H. Wang, C. Peng, A. Goel, S. Son, P. Liang, A. Niknejad, H. Hwang, and G. Chien, *A 24.7dbm all-digital rf transmitter for multimode broadband applications in 40nm cmos,* in *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers* (2013) pp. 332–333.

[8] R. Bhat, A. Chakrabarti, and H. Krishnaswamy, *Large-scale power combining and mixed-signal linearizing architectures for watt-class mmwave cmos power amplifiers,* IEEE Transactions on Microwave Theory and Techniques **63**, 703 (2015).

[9] S. Hu, F. Wang, and H. Wang, *A 28ghz/37ghz/39ghz multiband linear doherty power amplifier for 5g massive mimo applications,* in *2017 IEEE International Solid-State Circuits Conference (ISSCC)* (2017) pp. 32–33.

[10] D. Zhao and P. Reynaert, *A 60-ghz dual-mode class ab power amplifier in 40-nm cmos,* IEEE Journal of Solid-State Circuits **48**, 2323 (2013).

[11] D. Zhao and P. Reynaert, *A 40-nm cmos e-band 4-way power amplifier with neutralized bootstrapped cascode amplifier and optimum passive circuits,* IEEE Transactions on Microwave Theory and Techniques **63**, 4083 (2015).

[12] R. R. Britt, *Milky way's age narrowed down,* (2004), [Online; posted 17-August-2004].

[13] B. H. Project, *Humans,* [Online; accessed 1-August-2018].

[14] R. P. Mitra and D. P. Agrawal, *5G mobile technology: A survey,* ICT Express (2016), [Online].

[15] A. Morgado, K. M. S. Huq, S. Mumtaz, and J. Rodriguez, *A survey of 5G technologies: regulatory, standardization and industrial perspectives,* Digital Communications and Networks (2017), [Online].

[16] sdx central, *The importance of 5g research,* [Online; accessed 1-August-2018].

[17] *Ieee standard letter designations for radar-frequency bands,* [Online; accessed 1-August-2018].

[18] *Spectrum for wireless broadband – 5g,* [Online; accessed 1-August-2018].

[19] J. A. Kowaleski, T. Truex, D. Dever, D. Ament, W. Anderson, L. Bair, S. Bakke, D. Bertucci, R. Castelino, D. Clay, J. Clouser, A. DiPace, V. Germini, R. Hokinson, C. Houghton, H. Kolk, B. Miller, G. Moyer, R. O. Mueller, N. O'Neill, D. A. Ramey, Y. Seok, J. Sun, G. Zelic, and V. Zlatkovic, *Implementation of an alpha microprocessor in soi,* in *2003 IEEE International Solid-State Circuits Conference, 2003. Digest of Technical Papers. ISSCC.* (2003) pp. 248–491 vol.1.

[20] J. Goo, R. Q. Williams, G. O. Workman, Q. Chen, S. Lee, and E. J. Nowak, *Compact modeling and simulation of pd-soi mosfets: Current status and challenges,* in *2008 IEEE Custom Integrated Circuits Conference* (2008) pp. 265–272.

[21] G. Foundries, *Design kit and technology training book,* [Confidential; Document].

[22] D. Jacquet, F. Hasbani, P. Flatresse, R. Wilson, F. Arnaud, G. Cesana, T. D. Gilio, C. Lecocq, T. Roy, A. Chhabra, C. Grover, O. Minez, J. Uginet, G. Durieu, C. Adobati, D. Casalotto, F. Nyer, P. Menut, A. Cathelin, I. Vongsavady, and P. Magarshack, *A 3 ghz dual core processor arm cortex tm -a9 in 28 nm utbb fd-soi cmos with ultra-wide voltage range and energy efficiency optimization,* IEEE Journal of Solid-State Circuits **49**, 812 (2014).

[23] A. Wentzel, C. Meliani, and W. Heinrich, *Rf class-s power amplifiers: State-of-the-art results and potential,* in *2010 IEEE MTT-S International Microwave Symposium* (2010) pp. 812–815.

[24] A. M. Niknejad, *Electromagnetics for High-Speed Analog and Digital Communication Circuits* (Cambridge University Press, 2007).

[25] Y. Zhao, *HIGH-PERFORMANCE MM-WAVE AND WIDEBAND LARGE-SIGNAL AMPLIFIERS*, PhD dissertation, Technische Universiteit Delft (2013).

[26] N. Reynders and W. Dehaene, *Sub-threshold operation: Theory and challenges,* in *Ultra-Low-Voltage Design of Energy-Efficient Digital Circuits* (Springer International Publishing, Cham, 2015) pp. 19–45.

[27] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, *Leakage current mechanisms and leakage reduction techniques in deep-submicrometer cmos circuits,* Proceedings of the IEEE **91**, 305 (2003).

[28] T. Hanna, N. Deltimple, and S. Frégonèse, *A class-j power amplifier for 5g applications in 28nm cmos fd-soi technology,* in *2017 30th Symposium on Integrated Circuits and Systems Design (SBCCI)* (2017) pp. 110–113.

[29] *Ieee 802.11ay,* [Online; accessed 18-Sept-2018].

[30] D. Lie, J. 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 **2018** (2018).

[31] Y. Li, J. Lopez, P. Wu, W. Hu, R. Wu, and D. Y. C. Lie, *A sige envelope-tracking power amplifier with an integrated cmos envelope modulator for mobile wimax/3gpp lte transmitters,* IEEE Transactions on Microwave Theory and Techniques **59**, 2525 (2011).

[32] J. C. Mayeda, D. Y. C. Lie, and J. Lopez, *A highly efficient and linear 15 ghz gan power amplifier design for 5g communications,* in *2017 Texas Symposium on Wireless and Microwave Circuits and Systems (WMCS)* (2017) pp. 1–4.

[33] A. Grebennikov and J. Wong, *A dual-band parallel doherty power amplifier for wireless applications,* IEEE Transactions on Microwave Theory and Techniques **60**, 3214 (2012).

[34] *Ccdf measurements,* [Online; accessed 4-August-2018].

[35] D. A. Peterson, *Amplifier for interpolating the power supply from multiple supply voltages,* US Patent **6,078,169** (2000).

[36] J. . Choi and J. G. Fossum, *Analysis and control of floating-body bipolar effects in fully depleted submicrometer soi mosfet's,* IEEE Transactions on Electron Devices **38**, 1384 (1991).

[37] A. Marshall and S. Natarajan, *Pd-soi and fd-soi: a comparison of circuit performance,* in *9th International Conference on Electronics, Circuits and Systems*, Vol. 1 (2002) pp. 25–28 vol.1.

[38] M. Narayanan, H. Al-Nashash, B. Mazhari, D. Pal, , and M. Chandra, *Analysis of kink reduction in soi mosfet using selective back oxide structure,* Active and Passive Electronic Components **38**, 1 (2012).

[39] S. Consortium, *Fully depleted (fd) vs. partially depleted (pd) soi,* (2008), [Online; posted 14-May-2008].

[40] T. Takayasu Sakurai, Akira Matsuzawa, *Fully-depleted soi cmos circuits and technology for ultralow-power applications,* (2007), [Online].

[41] M. Youssef, A. Zolfaghari, B. Mohammadi, H. Darabi, and A. A. Abidi, *A low-power gsm/edge/wcdma polar transmitter in 65-nm cmos,* IEEE Journal of Solid-State Circuits **46**, 3061 (2011).

[42] F. H. Raab, *Intermodulation distortion in kahn-technique transmitters,* IEEE Transactions on Microwave Theory and Techniques **44**, 2273 (1996).

[43] B. Razavi, *RF Microelectronics (2Nd Edition) (Prentice Hall Communications Engineering and Emerging Technologies Series)*, 2nd ed. (Prentice Hall Press, Upper Saddle River, NJ, USA, 2011).

[44] S. M. Alavi, *All-Digital I/Q RF-DAC*, PhD dissertation, Technische Universiteit Delft (2014).

[45] D. A. Frickey, *Conversions between s, z, y, h, abcd, and t parameters which are valid for complex source and load impedances,* IEEE Transactions on Microwave Theory and Techniques **42**, 205 (1994).

[46] S. C. Cripps, *RF Power Amplifiers for Wireless Communications, Second Edition (Artech House Microwave Library (Hardcover))* (Artech House, Inc., Norwood, MA, USA, 2006).

[47] S. Eid and G. A. Blum, *Neutralization techniques for differential low noise amplifiers,* US Patent **7,256,646 B2** (2007).

[48] N. Deferm and P. Reynaert, *Differential and common mode stability analysis of differential mm-wave cmos amplifiers with capacitive neutralization,* Analog Integrated Circuits and Signal Processing **80**, 1 (2014).

[49] D. Chowdhury, P. Reynaert, and A. M. Niknejad, *Design considerations for 60 ghz transformer-coupled cmos power amplifiers,* IEEE Journal of Solid-State Circuits **44**, 2733 (2009).

[50] N. Deferm and P. Reynaert, *Cmos at millimeter wave frequencies,* in *CMOS Front Ends for Millimeter Wave Wireless Communication Systems* (Springer International Publishing, Cham, 2015) pp. 13–31.

[51] K. H. An, O. Lee, H. Kim, D. H. Lee, J. Han, K. S. Yang, Y. Kim, J. J. Chang, W. Woo, C. Lee, H. Kim, and J. Laskar, *Power-combining transformer techniques for fully-integrated cmos power amplifiers,* IEEE Journal of Solid-State Circuits **43**, 1064 (2008).

[52] K. Scheir, S. Bronckers, J. Borremans, P. Wambacq, and Y. Rolain, *A 52 ghz phased-array receiver front-end in 90 nm digital cmos,* IEEE Journal of Solid-State Circuits **43**, 2651 (2008).

[53] F. Behbahani, Y. Kishigami, J. Leete, and A. A. Abidi, *Cmos mixers and polyphase filters for large image rejection,* IEEE Journal of Solid-State Circuits **36**, 873 (2001).

[54] S. Kulkarni, D. Zhao, and P. Reynaert, *Design of an optimal layout polyphase filter for millimeter-wave quadrature lo generation,* IEEE Transactions on Circuits and Systems II: Express Briefs **60**, 202 (2013).

[55] J. Kaukovuori, K. Stadius, J. Ryynanen, and K. A. I. Halonen, *Analysis and design of passive polyphase filters,* IEEE Transactions on Circuits and Systems I: Regular Papers **55**, 3023 (2008).

[56] A. M. Niknejad, S. Emami, B. Heydari, M. Bohsali, and E. Adabi, *Nanoscale cmos for mm-wave applications,* in *2007 IEEE Compound Semiconductor Integrated Circuits Symposium* (2007) pp. 1–4.

[57] D. Zhao and P. Reynaert, *A 40 nm cmos e-band transmitter with compact and symmetrical layout floor-plans,* IEEE Journal of Solid-State Circuits **50**, 2560 (2015).

[58] W. K. W. Ali and S. H. Al-Charchafchi, *Using equivalent dielectric constant to simplify the analysis of patch microstrip antenna with multi-layer substrates,* in *IEEE Antennas and Propagation Society International Symposium. 1998 Digest. Antennas: Gateways to the Global Network. Held in conjunction with: USNC/URSI National Radio Science Meeting (Cat. No.98CH36*, Vol. 2 (1998) pp. 676–679 vol.2.

[59] I. Aoki, S. D. Kee, D. B. Rutledge, and A. Hajimiri, *Distributed active transformer-a new power-combining and impedance-transformation technique,* IEEE Transactions on Microwave Theory and Techniques **50**, 316 (2002).

[60] D. E. Norgaard, *The phase-shift method of single-sideband signal reception,* Proceedings of the IRE **44**, 1735 (1956).

[61] S. T. Lee, A. Bellaouar, and S. Embabi, *A low-power, compact 76-81ghz fmcw transmitter for automotive radar in 22nm fdsoi,* in *ESSCIRC 2018 - 44rd IEEE European Solid State Circuits Conference* (2018) pp. 186–189.