DELFT UNIVERSITY OF TECHNOLOGY

A Medium Power 135-GHz Power Amplifier

Georgios Sarris
4239504

May 27, 2015

A thesis submitted to the Electrical Engineering, Mathematics and Computer Science Faculty of Delft University of Technology in partial fulfillment of the requirements for the degree of Master of Science

Delft University of Technology, The Netherlands

DELFT UNIVERSITY OF TECHNOLOGY

Faculty of Electrical Engineering, Mathematics and Computer Science

The undersigned hereby certify that they have read and recommend to the faculty of Electrical Engineering, Mathematics and Computer Science (EWI) for acceptance a thesis entitled

A Medium Power 135-GHz Power Amplifier

By

Georgios Sarris

Supervisor(s):

_____

Dr. Marco Spirito

Thesis Committee:

_____

Dr. Leonardus C. N. De Vreede

_____

Dr. Michiel Pertijs

_____

Dr. Hao Gao

Table of Contents

Abstract

Recently, a plethora of applications has emerged that can take advantage of the mm – wave frequency band characteristics, i.e. high bandwidth, small form factor etc. Applications at mm-wave frequencies include smart sensors, single chip phased-arrays etc. For many years, III-V technologies were the main driving technologies for circuits operating at high frequencies. However, recent advances in silicon processes have enabled the use of BiCMOS and CMOS technologies as enablers for circuits and systems operating at mm-wave and sub-mm wave frequencies, i.e. frequencies above 100GHz. Hence, high frequency and baseband electronics can be integrated in a single chip, thus reducing cost dramatically. The ultimate goal is to have silicon circuits operation at Terahertz frequencies.

The need for power generation at high frequencies is fundamental for the successful implementation of any application. There is a variety of circuits providing power at frequencies below 100GHz, namely oscillators and power amplifiers. Furthermore, lasers and other optoelectronic devices are capable of providing power above 10THz. Hence, there is a gap in power generation between 100GHz and THz that has to be filled in order for the applications to be accommodated.

This thesis constitutes an effort towards the direction of understanding the challenges and limitation of power generation at frequencies above 100 GHz. Specifically active and passive device limitations are addressed and guidelines are given in order for the best performance possible to be achieved. The acquired knowledge is applied in the design of a medium output power PA at 135 GHz. Specifically, the PA achieves an -1dB compression point output power of 5dBm and a saturated output power of 8 dBm along with a power gain of 17 dB.

Chapter 1. Introduction

1.1 Applications of mm-wave and sub-mm-wave circuits.

Mm-wave and sub-mm wave integrated circuits, i.e. circuits operating at frequencies above 60GHz, provide the workhorse for a plethora of emerging applications. A useful feature of operation at these frequencies is the fact that the wavelength is in the millimeter range and it forms images with millimeter resolution. Furthermore, THz radiation is not ionizing, as its photons are not energetic enough to knock electrons off atoms. Security applications, like concealed weapon detection can harness this feature [2]. Quality control of the surface of various materials also employ mm – wave imaging. Semiconductor wafer surfaces check is one example [3]. Biomedical applications can also benefit, with an example being dental practice [4]. Other applications feature astronomy, automotive radars and various speed and position detection sensors.

Circuits operating at mm-wave frequencies benefit the aforementioned applications. High data rate radios at frequencies above 100 GHz promise increased bandwidth capabilities. Furthermore, when operating at higher frequencies, the antennas assume smaller size leading to complete on-chip solutions. This opens the path for the integration on silicon of complete arrays for even higher data rates (MIMO), beam forming, beam steering or power combining for transmission at higher distances [1]. High output power is one of the key specifications for the above applications to enable transmission at reasonable distances.

One of the reasons why the particular frequency band is the least employed so far, is the lack of circuits able to provide sufficient power at these frequencies. While there are a wide variety of circuits like amplifiers and oscillators at frequencies below 100GHz and solid-state lasers, light emitting diodes and detectors above 10THz, there is a lack of circuits operating in the intermediate frequency range. In [11], a graph of the output power of the available power sources in the THz range is given and the graph is repeated in figure 1.1. From the graph, we note that the output power greatly diminishes at the THz frequency range. Hence, there is a need for the design of circuits covering the frequency band between those “extremes”.

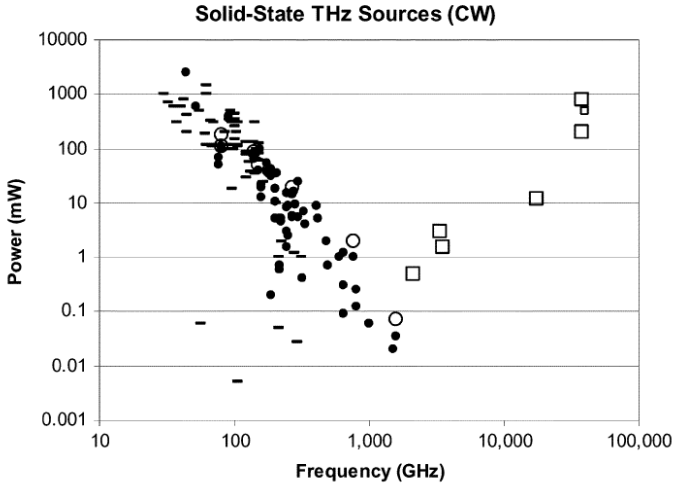


Fig. 1.1. The THz gap (figure taken from [11])

1.2 Enabling technologies for mm-wave and sub-mm-wave circuits.

During the first days of microwave circuit, compound semiconductors were the preferred choice for MMICs, due to their high speed and high power handling capabilities. III – V semiconductors, like GaAs and InP have superior performance compared to their silicon counterparts, due to increased mobility, higher breakdown voltages and the availability to integrate high – Q passives [5]. Record performance have been reported throughout the years for such compound technologies in the mm-wave and sub-mm-wave frequency range, as an example in [6], the authors presented an amplifier with power gain of 10 dB at 550 GHz in a process with f_{MAX} of 1.2 THz. As seen in figure 1.2, the f_{MAX} for CMOS and BiCMOS processes according to the ITRS roadmap [7] is way smaller. Nevertheless, CMOS and BiCMOS technology provide a larger integration platform, with a variety of components (transistors, varactors, capacitances, digital ports, etc.) and an extensive back-end-of-line realized with a high yield allowing to address markets where volume production is required, i.e., automotive and telecom Currently Si processes, like CMOS and SiGe HBTs are pushing to become the technologies of choice for the uprising commercial applications operating in the mm-wave region .

The continuous development of processing technology as resulted in improved high frequency performance of silicon based processes, thanks to the lower foot-print and associated parasitic. In Fig. 1 the ITRS roadmap [7] of silicon based technology f_{MAX} is reported clearly highlighting the increasing speed trend for CMOS and BiCMOS devices. The graph, however, presents some pessimistic values for the BiCMOS processes. As an example, the f_{MAX} of the IHP 130nm process used here, after Metal1 and Metal2 connection on top of the intrinsic devices, is greater than 400GHz. In [1] it is stated that the f_{MAX} of contemporary CMOS anf SiGe HBT devices are comparable. However, according to [1], the SiGe HBT devices offer a plethora of advantages for large signal operation. For comparable fMAX values, the SiGe HBT offer higher current density compared to CMOS. Furthermore, the higher breakdown voltages offer a higher voltage swing. The collector to bulk capacitance assumes a smaller value compared to the gate-drain overlap capacitance of MOS devices. Finally, the increased transconductance of SiGe HBTs leads to superior large signal performance [1], [12].

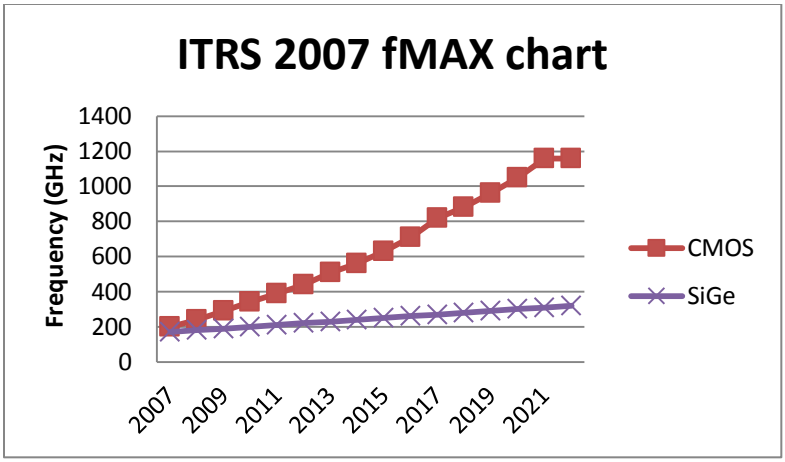


Fig. 1.2 ITRS Roadmap for maximum oscillation frequency.

Another drawback of CMOS is the implication of scaling on the Back End of the Line (BEOL). Since it is a process dedicated to digital circuits, in order for the integration density to be increased, the height of the BEOL shrinks rapidly from one node to the next. This leads to metals being closer to the substrate, which increases the substrate losses of transmission lines [1]. Also the inter-metal distance shrinks, which increases the capacitance to adjacent metals and to the substrate. Finally, the metals become thinner, which is detrimental in terms of electro migration in high current applications.

The higher capacitance associated with passives realized in shrinked BEOLs is detrimental for analog and high frequency circuits. First and foremost the realization of high characteristic impedance lines becomes difficult. Furthermore, the extra parasitic coupling lowers the Self Resonance Frequency (SRF) of the passive components. The reduced SRF in conjunction with the capacitive loading from the active stage makes transformers self-resonate at frequencies below the ones associated with the applications. Finally, the capacitance of the passives add to the output capacitance of the active cells. According to [13] this limits the output power in a way that will be described in detail in Chapter 4.

On the other hand, BiCMOS processes are usually dedicated to high frequency applications and come with BEOLs that are more suitable for mm-wave operation. They include thick metal layers that are away from the substrate compared to their CMOS counterparts. This leads to high – Q passive components and transmission lines.

For the sake of comparison, we give a table with some representative features of two different BEOLs. The first is from a 55nm BiCMOS process from ST Microelectronics which has a digital BEOL. The second one is from a 130nm BiCMOS process from IHP. The characteristics to be compared are distance from bottom metal (ground) to top metal, number of thick metals and thickness of the latter. The IHP process which is dedicated to BiCMOS, offers a clear advantage in terms of the BEOL.

Table 1. Comparison of BEOL characteristics between processes.

Process	M1-TopMetal Distance (um)	No. of thick metals	Metal Thickness (um)
ST BiCMOS 55nm	5.747	2	3 and 0.8
IHP BiCMOS 130nm	9.83	2	3 and 2

Here you should have a motivation section, from what you have highlighted in the previous section you should motivate the reader of this work and mention that the goal is for develop,ent of compact highly efficient mm-wave Pas.

1.3 Motivation and Goal of the thesis

From the above discussion, it turns out that power sources operating in the THz region are necessary. Power amplifiers can be used as power sources that can cover the gap presented in figure 1.1 and hence should be designed for operation at these frequencies. They should provide high output power and certain power gain combined with high efficiency in order for DC power consumption to be minimal. Low power consumption increases the lifetime in case of mobile devices or the cooling costs in the case of large immobile systems.

In this work, we aim for the design of PAs operating at the low end of the unexploited THz range, specifically at 135GHz. The main specification is a high output power in order to fulfill the needs of potential applications. Furthermore, a power gain larger than 15 dB is specified. As this is an experimental work, we are not being strict in the efficiency specification. Nevertheless, we make an effort to achieve efficiency as high as possible.

The output power of a PA depends on the power that the active cells can provide as well as on the power loss introduced by the passive components that are used to provide matching, biasing, coupling between stages etc. Hence, in this thesis we concentrate on the optimization of both. Low loss passive components, that promote high output power and efficient operation, are designed in order for high output power and efficiency to be achieved. Furthermore, the proper design steps regarding the active part should be taken. These include the choice of the proper topology, the correct sizing and biasing of the stages, as well as proper layout in order for the parasitics degrading the performance to be minimized.

1.4 Power Amplifier Requirements and Metrics

Power Amplifier requirements are set from the application layer. An example is wireless system, in which the transmitter should produce an adequately strong signal that will reach the receiver with certain power content despite the attenuation due to path loss. In order for communication to be successful, the signal at the input of the receiver should be higher than the minimum detectable signal [8] which defines the sensitivity of the receiver. The input power of the receiver is given by the link budget equation, which is:

$$P_R = P_T - Path\ Loss \qquad (1.1)$$

where P_R is the input power of the receiver, P_T is the output power of the transmitter and Path Loss includes the loss of the medium during the information transfer. The PA is mainly responsible for the first two terms of the right side of the equation. For robust communication, the PA should provide sufficient output power and large power gain. Furthermore, the more efficient the PA, the lower the cooling costs in case of base stations or the longer the battery lifetime in the case of battery operated mobile systems.

From the above discussion, we conclude that the PA is the most critical part regarding signal transmission. So before entering any design discussions, it is necessary to give the performance metrics of power amplifiers and address some of the design challenges. A simple power amplifier schematic is given below. The inductor L∞ acts as an RF choke and is used to feed the supply voltage to the device. The capacitor C∞ acts as a DC block. The input AC source is the one providing the RF signal as well as the biasing voltage for the transistor.

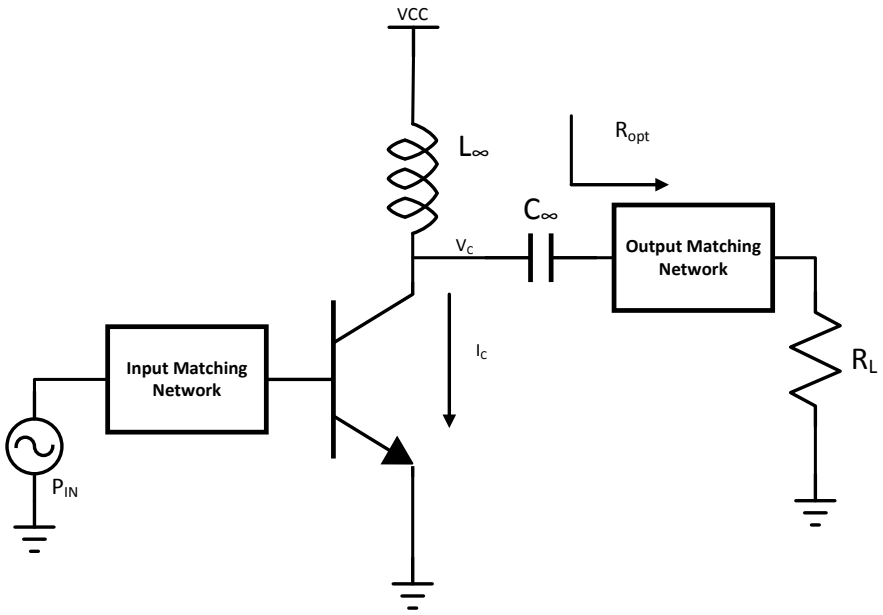


Fig. 1.3 Basic Power Amplifier Schematic

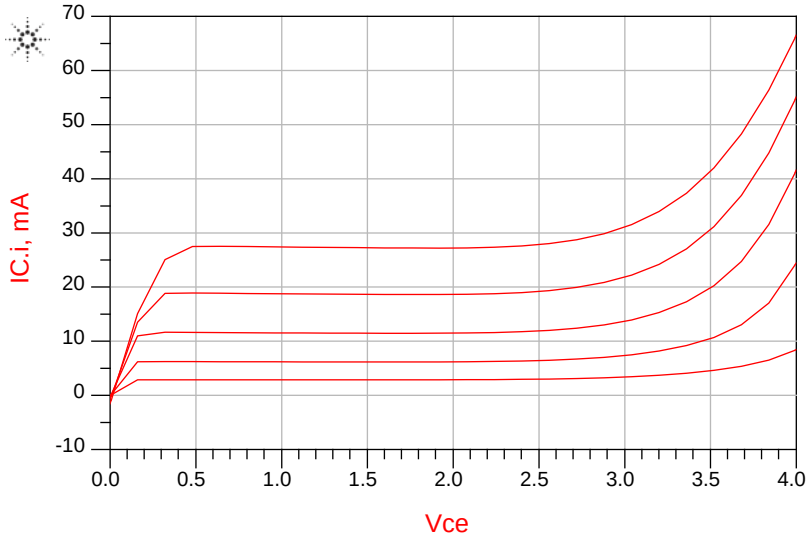


Fig. 1.4 Basic Power Amplifier Schematic

The design of any Power Amplifier starts from the IC-VCE characteristics of the device. These curves indicate the safe operating region of the voltages and currents that the device can handle. Assuming that the output voltage can swing from a value V_{MIN} up to a value V_{MAX} and the current from a minimum value I_{MIN} up to a maximum value I_{MAX}, we can define the power amplifier metrics.

The output power is defined as:

$$P_{OUT} = \frac{V_{MAX}-V_{MIN}}{2\sqrt{2}} \frac{I_{MAX}-I_{MIN}}{2\sqrt{2}} \quad (1.2)$$

And the power gain is defined as:

$$G_P = P_{OUT} - P_{IN} \quad (1.3)$$

where P_{IN} is the power entering the power amplifier.

Furthermore, we can define collector efficiency as

$$n = \frac{P_{OUT}}{P_{DC}} \quad (1.4)$$

where P_{DC} is the power consumed by the supply. In other words it defines the ratio between the power delivered to the load and the power dissipated in the transistor as heat.

Finally, another metric of efficiency is the Power Added Efficiency (PAE), which takes into account the gain of the amplifier.

$$PAE = \frac{P_{OUT}-P_{IN}}{P_{DC}} = n(1 - \frac{1}{G_P}) \quad (1.5)$$

We note that for a power gain higher than 10dB, the PAE approaches n asymptotically.

1.5 Power Amplifier Design Challenges

The design of the PA differs from that of small signal amplifiers, in the sense that the output voltages and currents are not small signals, but they vary between certain limits. The output voltage of the device can theoretically swing between 0 and $2V_{CC}$, due to the presence of the inductor at the output of the device. In practice the minimum output voltage is limited by $V_{CE,SAT}$ for bipolar devices. Furthermore, the maximum output voltage is less than the theoretical V_{CC} due to the finite Q of the inductor. Nevertheless, we should take care that the maximum output voltage is lower than the breakdown limit of the device [9]. The output current can swing from a minimum value, i.e. the current that corresponds to $V_{CE,SAT}$ up to a value of I_{MAX} that corresponds to the onset of high injection effects.

From the I_C-V_{CE} curves for a SiGe HBT common emitter (CE) stage of figure 1.5 with an ideal voltage drive we can create an example. We assume that the curves correspond to acceptable voltage and current range, with $2V_{CC}$=3V, I_{MAX}=28 mA, V_{MIN}=0.3V and I_{MIN}=6 mA. Assuming that we perform a Class – A biasing, i.e. the operating point lies exactly in the middle of the voltage and current range, we get the DC

quantities as V_{DC}=1.35V and I_{DC}=11 mA. We assume sinusoidal quantities. We design at class – A just to give a feel of the signal levels that bound the output power capabilities. More information about different classes of operation can be found at [10].

Finally in order to get the maximum power out of the stage, we have to match the output properly. In order to achieve this, we have to view the stage as a current source that drives a resistor with a certain current which swings between the two limits introduced previously. The voltage across the resistor is also limited between the positive and negative supply. Hence, to avoid clipping and for the output power to be maximum, the resistor should have such a value so that the maximum current flowing through the resistor will not give a voltage higher than V_{MAX}. Also if the supply voltage is enforced to the resistor, the current should not exceed I_{MAX}. From the above considerations, we find that the optimal load resistance is:

$$R_{OPT} = \frac{V_{MAX} - V_{MIN}}{I_{MAX} - I_{MIN}} \quad (1.6)$$

The matching network should transform the load impedance to this optimal value.

The match for maximum output power is not in compliance with maximum operating power gain G_P. To get a higher power gain, we usually have to settle for a smaller amount of output power. To achieve this, we have to design the output matching network accordingly. More details will be given in later design chapters.

1.6 Outline of the thesis

The outline of the thesis is presented in this section. In Chapter 2, we provide a literature review of existing PA implementations in both CMOS and SiGe BiCMOS technologies. We conclude that a differential implementation provides a higher insensitivity to ground parasitics but it requires a single ended to differential conversion interface in order to be compatible with the transmitter antenna or the measurement equipment. In Chapter 3, we address the challenges of implementing passive matching networks as well as balanced to unbalanced (bal-un) conversion structures at frequencies above 100 GHz. The main objectives are low loss structures in order for high output power and efficiency to be achieved. Furthermore, when bal-un operation is desired, a single ended input should be split into two signals of same amplitude but opposite phase. In Chapter 4, various challenges associated with active devices are presented, namely the choice of the proper topology, the proper biasing and sizing of the cells as well as parasitic-minimization-driven layouts. In Chapter 5, the design of a PA IC at 135 GHz in a 130nm SiGe BiCMOS process is presented. Design details as well as the simulated performance are given. Finally, at Chapter 6, conclusions stemming from this work as well as suggestions for further improvements in future work are given.

1.7 References

[1]Ali M. Niknejad and Hossein Hashemi (editors), ”Mm-wave Silicon Technology, 60GHz and beyond”, Springer, 2008.

[2] D. M. Sheen, D. L. McMakin and T. E. Hall, “Three-Dimensional Millimeter-Wave Imaging for Concealed Weapon Detection”, IEEE Transactions on Microwave Theory and Techniques vol. 49, pp. 1581 – 1592, Sept. 2001.

[3]M. Hermann, M. Tani, K. Sakai and R. Fukasawa, “Terahertz Imaging of Silicon Wafers”, Journal of Applied Physics, vol. 91, pp. 1247 – 1250, Feb. 2002.

[4]N. Hoshi, Y. Nikawa, K. Kawai and S. Ebisu, “Application of Microwaves and Millimeter Waves for the Characterization of Teeth for Dental Diagnosis and Treatment”, IEEE Transactions on Microwave Theory and Techniques, vol. 46, pp.834 – 838, June 1998.

[5]Chinh H. Doan, Sohrab Emami, Ali M. Niknejad, Robert W. Broedsen, “Millimeter-Wave CMOS Design”, IEEE Journal of Solid State Circuits, vol. 40, pp.144-155, January 2005.

[6]W. R. Deal, K. Leong, X. B. Mei, S. Sarkozy, V. Radisic, J. Lee, P. H. Liu, W. Yoshida, J. Zhou and M. Lange, “Scaling of InP HEMT Cascode Integrated Circuits to THz Frequencies”, IEEE Compound Semiconductor IC symposium, Monterey, CA, Oct. 2010, pp. 195-198.

[7]ITRS “RF and A/MS Technologies for Wireless Communications”, ITRS 2007 edition. [Online]. Available: http://www.itrs.net/Links/2007ITRS/Home2007.htm

[9]Sorin Voinigescu, “High Frequency Integrated Circuits”, Cambridge University Press, 2013.

[10]Steve C. Cripps, “RF Power Amplifiers for Wireless Communications, Second Edition”, Artech House, 2006.

[11]Thomas W. Crowe et. al., “Opening the Teraherzt Window with Integrated Diode Circuits”, IEEE Journal of Solid State Ciruits, vol. 40, no. 10, pp. 2104-2110, Oct. 2005

[12]John R. Long, “SiGe Radio Frequency ICs for Low-Power Portable Communication”, Proceedings of the IEEE, vol.93, September 2005.

[13] N. Sarmah, P. Chevalier and Ulrich R. Pfeiffer, “160-GHz Power Amplifier Design in Advanced SiGe HBT Technologies with P_{SAT} in excess of 10 dBm”, IEEE Transactions on Microwave Theory and Techniques, vol. 61, pp. 939 – 947, Feb. 2013.

Chapter 2. Literature review

In the previous chapter we went through the benefits of operating at mm-wave and sub-mm wave frequencies, namely increased bandwidth capabilities, a decrease in size etc. We also saw the technologies acting as enablers for circuits operating at frequencies above 100 GHz. We identified CMOS and SiGe BiCMOS as the ones most suitable for integration at these frequencies. A literature survey was conducted regarding previous works aiming at frequencies above 100GHz in both the aforementioned processes. The results of this survey are presented in this chapter. Furthermore, we focus on the advantages of differential circuits compared to single-ended operation.

2.1 Previous Work on SiGe BiCMOS bipolar mm wave power amplifiers.

From the two enabling technologies mentioned above the most promising is SiGe BiCMOS. As described in Chapter 1, despite the comparable f_{MAX} of modern Si CMOS and SiGe HBT transistors, the latter provide a plethora of advantages regarding large signal operation, namely increased current density, increased voltage swing, lower output capacitance and higher transconductance. There are several PA circuits operating above 100GHz. The ones with the best performance are briefly described. Finally a table of comparison is given.

In [1] a 3-stage cascode amplifier operating at 130 GHz is implemented as shown in Fig. 2.1. The process is an IHP 130nm with the f_{MAX} being above 300GHz. Three differential cascode stages are used to provide sufficient power gain. The matching between stages if done by means of transmission lines and transformers. Furthermore, the output transformer is also used as a balun in order to provide a differential to single ended conversion for ease of measurement. This amplifier gives an output power of 7.7 dBm and a gain of 24.3dB.

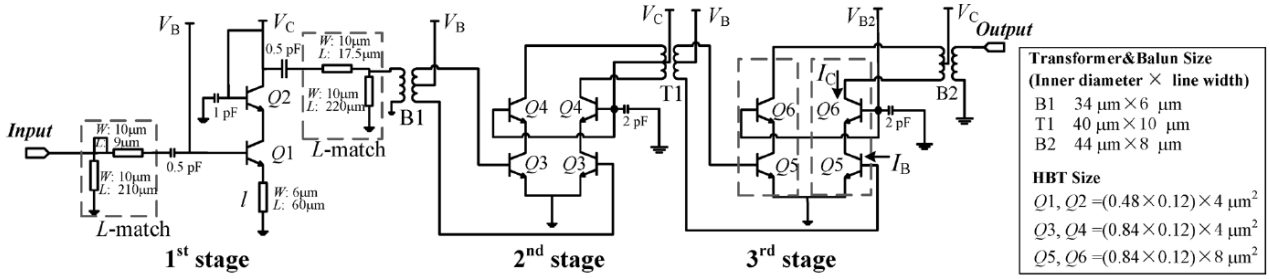


Fig. 2.1 Power Amplifier Schematic from [1]

In [2] a 3-stage cascode amplifier is implemented as shown in Fig. 2.2. The process is an IHP 130nm process with an fmax equal to 450 GHz. Stagger tuning is used in order for a large bandwidth to be achieved. Input and output Marchand bal-uns are used to provide a single ended interface that is necessary for measurements. Finally, the output power is 5-8 dBm from 135-170GHz, while the power gain varies from 14 to 17 dB.

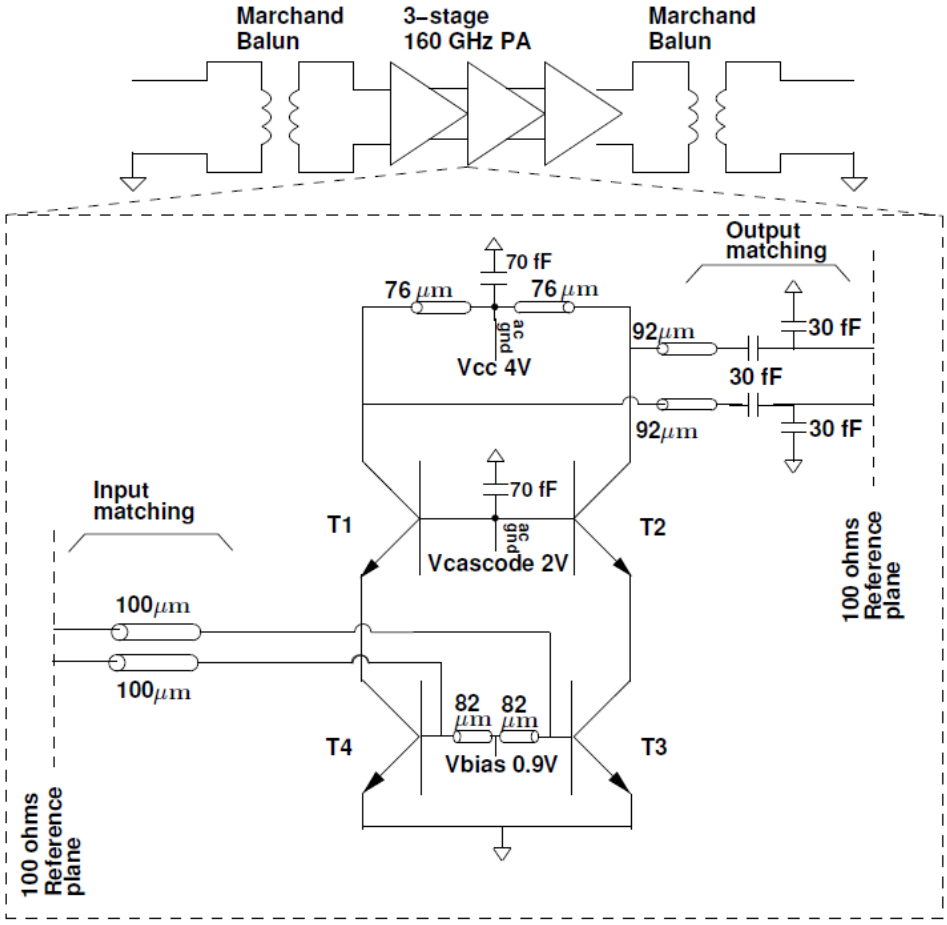


Fig. 2.2 Power Amplifier Schematic from [2]

In [3] a 160GHz PA is implemented in an ST Microelectronics process with an f_{MAX} of 400GHz. Three stages of push-pull cascodes are cascaded. The matching is performed by providing an intermediate match of 100Ω between stages. Marchand baluns are used in order to provide a differential to single ended conversion at these frequencies. The measured output power is 10dBm at saturation and 8.5dBm at -1 dB compression point with a gain of 20-32 dB between 150 and 170 GHz. The schematic is shown in Fig. 2.3. The topologies used in [2] and [3] are identical as they come from the same group. However, as they are among the designs with the highest performance, we present both of them here.

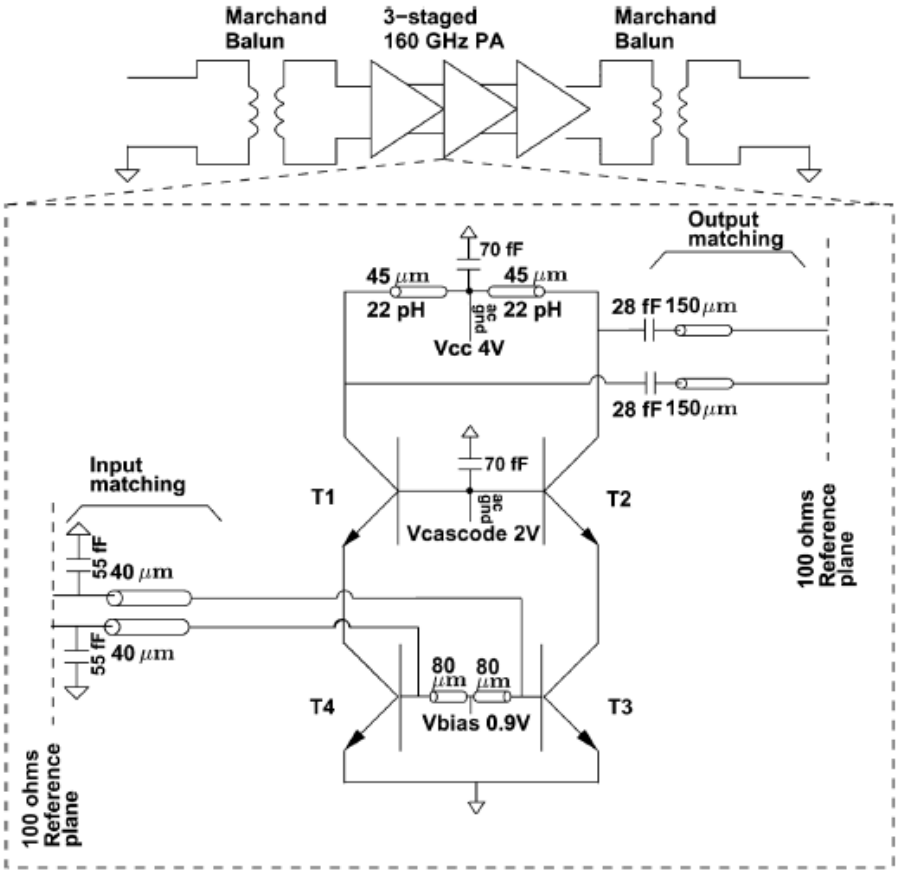


Fig. 2.3 150- 170 GHz Power Amplifier Schematic from [3]

In [4] a 170 GHz PA has been implemented in a 130nm SiGe BiCMOS process with an f_{MAX} of 260-290 GHz. The circuit is single-ended and no bal-uns are necessary at input or output. Five stages are used here. The first three are cascode in order to provide sufficient power gain which is scarce at those frequencies. The last two are Common Emitter (CE) stages which can provide a higher output voltage swing, and hence a higher output power, with a reduced supply voltage compared to the cascodes. This leads to smaller DC power consumption and hence in a higher efficiency. The measured output power is 0 dBm and a power gain close to 15dB.

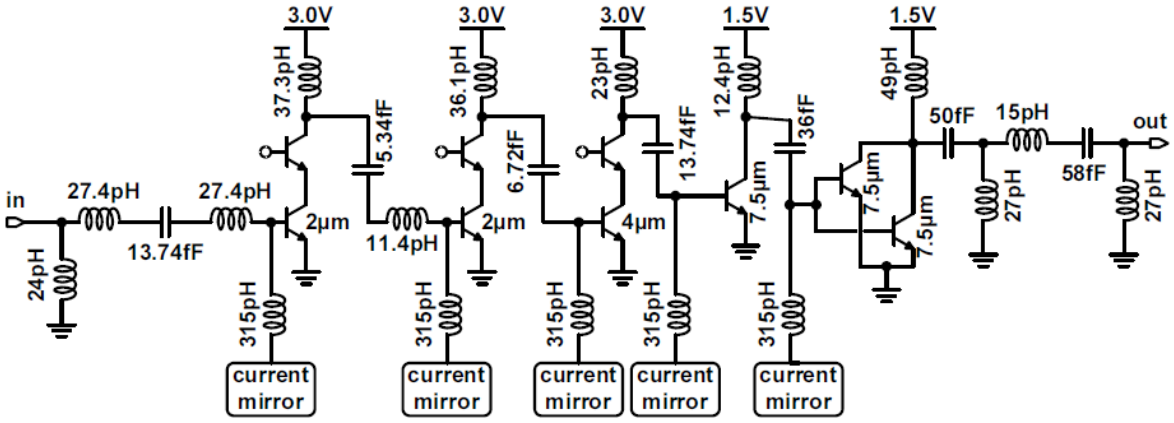


Fig. 2.4 170 GHz Power Amplifier Schematic from [4]

Table 2.1 Comparison of SiGe BiCMOS bipolar PA implementations

Work/Freq	Process	f_{MAX} (GHz)	Pout (dBm)	Gain (dB)	Topology	**Area (mm^2)**
[1]/130GHz	130nm	>300	7.7	24.3	3 stage diff. cascode	0.43x0.7
[2]/135-170GHz	130nm	450	5-8	14-17	3 stage diff. cascode	1.2x0.48
[3]/150-170GHz		400	10@160GHz	20-32	3 stage diff. cascode	1.04x0.374
[4]/170GHz	130nm	260-290	0	15	5 stage single ended. 3 first stages cascodes and 2 final stages CE	0.15x0.19

The above designs are summarized in Table 2.1. By close inspection of the table, we see the direct effect of f_{MAX} on the amplifier power gain and output power capability. In [4] despite the very large number of stages compared to the other works, the gain and output power are poor. This is due to the fact that the f_{MAX} of the particular process is the lowest. Furthermore, the differential topologies promote higher power gain and output power as it will be explained in later sections of this chapter due to their insensitivity to the ground plane parasitics. This can justify the larger area needed in general for a differential implementation.

2.2 Previous Work on CMOS sub-mm wave power amplifiers.

Implementation of PAs in CMOS is desirable because standard CMOS processes are cheaper than BiCMOS which require extra process steps. In this section, we provide some CMOS PA implementations found in the literature.

In [5], a 100GHz PA is implemented in 65nm CMOS is presented with an output power of 10dBm and a Power Added Efficiency (PAE) of 7.3% and a gain of 13dB. Four Common Source amplifiers are cascaded in a single ended fashion. The matching between stages is implemented with Coplanar Waveguide Lines (CPW) and slotted capacitors. The benefit of using CPW lines is that the adjacent ground structures are shielding the line from the rest of the circuit and make EM modeling easier. The amplifier schematic is shown in Fig. 2.5.

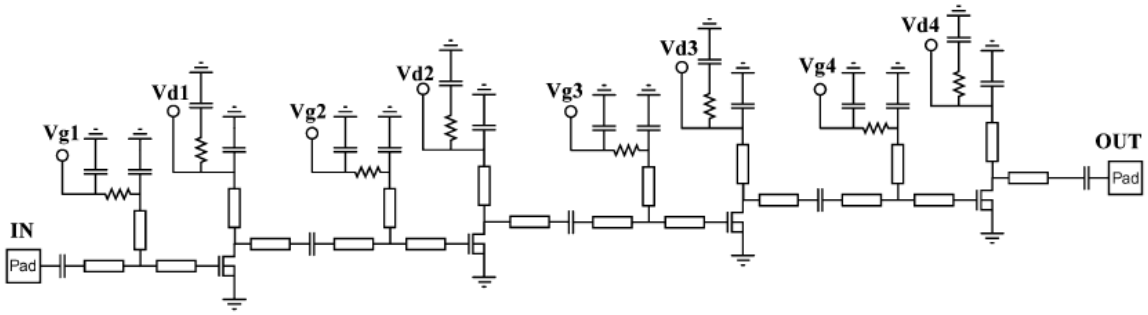


Fig. 2.5 170 GHz Power Amplifier Schematic from [5]

In [6] a PA with an output power of 12dBm consists of 3 cascode stages. The first is single ended and drives a coupler which splits the signal into two anti-phase signals. Each of them drives the 2 stages. Finally, another coupler is used in order to combine the powers coming out of the two paths. The power gain is 17dB and the process is CMOS 90nm with the f_{MAX} being 150GHz.

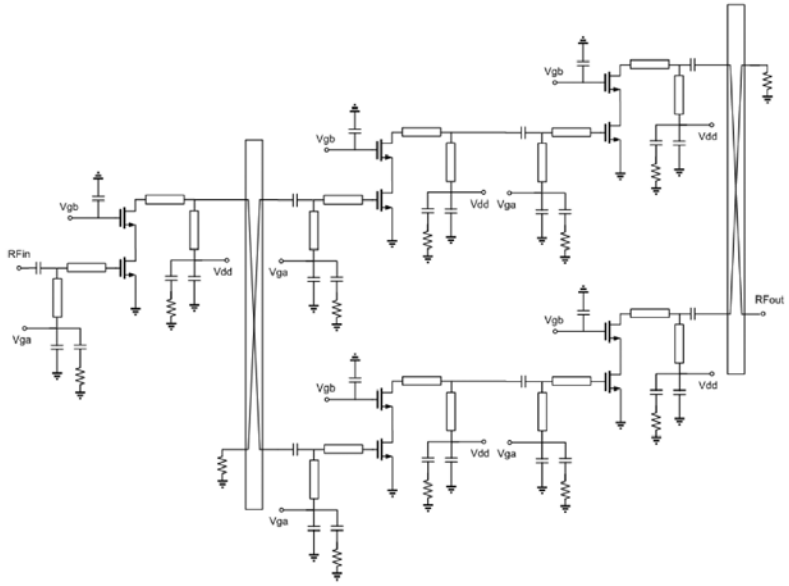


Fig. 2.6 Power Amplifier Schematic from [6] where couplers act as bal-uns

In [7] a 150GHz PA is realized in 65nm CMOS with an f_{MAX} of 280GHz. In this single ended implementation, three single ended Common Source (CS) stages are cascaded. Matching is implemented in terms of Microstrip line matching networks. An output power of 6.3dBm along with a power gain of 8.2dB is achieved. The circuit is shown in Fig. 2.7.

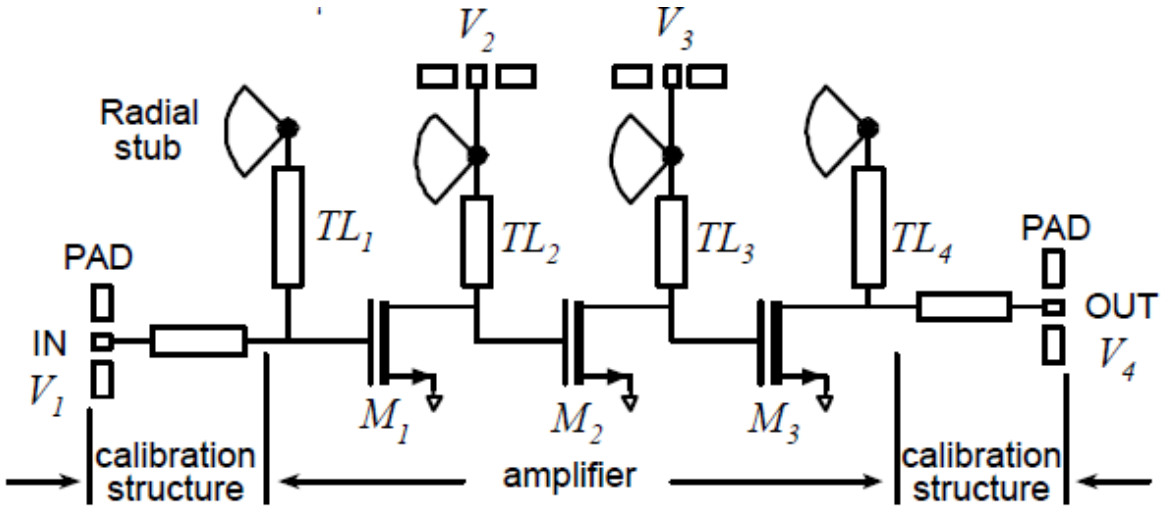


Fig. 2.7 Power Amplifier Schematic from [7].

The above designs are summarized in Table 2.2.

Table 2.2 Comparison of CMOS bipolar PA implementations

Work/Freq.	Process	f_{MAX} (GHz)	Pout (dBm)	Gain (dB)	Topology	**Area (mm^2)**
[5]/170GHz	65nm		10	13	5 stage single ended CS	0.775x0.42
[6]/103GHz	90nm	150	12	17	3 cascode stages, 1st is single ended. Coupler as balun and drives 2 differential cascode stages. Balun converts to single ended output.	0.74x0.54
[7]/150GHz	65nm	280	6.3	8.2	3 single ended CS stages	0.64x0.64

2.3 CMOS vs BiCMOS technologies in terms of Power Amplifier Gain and Output Power capabilities

From the above tables, we can conclude that there is a distinct advantage of the BiCMOS process compared to the CMOS in terms of f_{MAX} and power gain. For the frequency range of interest, PA design is getting easier when using BiCMOS in terms of power gain because the operating frequency is a larger fraction of f_{MAX}. This also eases the burden on the passives which are lossy in this frequency range, as

mentioned in chapter 1. Hence, more gain can be lost without severe consequences on the total gain performance of the amplifier.

The output power of the amplifier is not dependent on the frequency of operation, although for frequencies close to f_{MAX} the devices may be unable to provide power amplification due to the reduced power gain. In chapter 1, we saw from equation (1.2) and the related discussion that the output power depends on (1) the breakdown voltage of the device, (2) the maximum current that it can handle and (3) the transformer load impedance that is presented to the collector/drain of the output stage, i.e. the impedance transformation capabilities of the output matching network. All these factors are independent of f_{MAX}.

Another characteristic of the SiGe HBT bipolar devices which may advocate their use as the preferred active cells is the breakdown voltage. In MOS technology, the device will fail if the applied voltage exceeds the dielectric strength of the oxide. As the technologies scale down, the breakdown voltages also diminish. In bipolar technologies, breakdown voltage is limited by the collector – emitter breakdown voltage (BV_{CEO}) when there is a high impedance path from the base to ground. The infinite impedance does not allow the impact ionization generated carriers to flow out of the base of the transistor, thus causing breakdown. The breakdown voltage of a device with a finite resistor at the base (BV_{CER}), which is a more practical situation, can be two to three times higher than BV_{CEO} [8]. Thus SiGe devices are more suitable for high output power PAs, since their higher breakdown voltages permit the use of higher supply voltages and hence a larger output voltage swing. As an example, in our 130nm IHP BiCMOS process, the drain to source breakdown voltage for the NMOS device is 2.7V. For the bipolar NPN deice, the BV_{CEO} is 1.7V, but the BV_{CBO} is 4.8V. Hence we can allow for a higher collector to emitter voltage which also enables a higher voltage swing.

2.4 Differential vs Single Ended

During the design of a PA, the designer should keep in mind the measurement plan. In other words, he or she should take into consideration the laboratory infrastructure and design the circuit in a way compatible with the measurement equipment. At these frequencies, the latter provides a single ended interface for accurate measurements (Marco maybe you can provide some reference that further supports this). Hence, the amplifier should provide single ended inputs and outputs.

On the other hand, a differential operation provides higher insensitivity to the supply line ground plane parasitics due to the virtual ground that it creates along the axis of symmetry of the circuit. This is extremely important at high frequencies, since even the smallest interconnect can introduce a considerable amount of parasitic inductance and resistance. The bond wires, which are necessary to connect the chip to the outside word introduce high impedances which can make the on-ship ground vary considerably. Furthermore, the ground plane which is distributed all over the chip from the circuit ground connections to the pad introduces considerable parasitics. The ground plane can be simply modeled by a distributed inductance - resistance. In figure 2.8 we model the ground plane with a single R-L section. This is distributed all over the chip and makes the on-chip ground non-ideal since there is always and voltage drop and phase shift along the ground.

For a single ended circuit, the RF signal path is depicted in the left part of the next figure. The return path consists of the ground plane and all its parasitics; hence the signal is contaminated by them and this leads to drastic reduction of the gain due to inductive degeneration. It can also create undesired feedback

between stages in the case of a multi-stage configuration. To go over this, we should decouple as close to the transistor collector as possible with a large enough capacitor. The capacitor connects from the supply to the on-chip ground. In this way we provide a low impedance return path which will make the circuit less sensitive to the ground plane parasitics. This situation is depicted on the right side of the figure. However the amount of capacitance that is needed in order to provide sufficiently low impedance may be not easy to implement due to the limited area. Metal on metal (MOM) capacitors can provide capacitance densities in the order of $1fF/um^2$. At 130GHz, to provide 1 Ohm impedance the required capacitance is 1.2 pF, a value not easy to implement when using MOM capacitors. Another solution would be the use of a dedicated MIM capacitor layer. MIM capacitors provide very high capacitance densities, but they require extra process steps, thus increasing the cost of the chip. Nevertheless, the amount of the capacitance depends on the impedance that the ground plane introduces. If the ground plane has significant impedance there is no point in making the capacitance impedance extremely low.

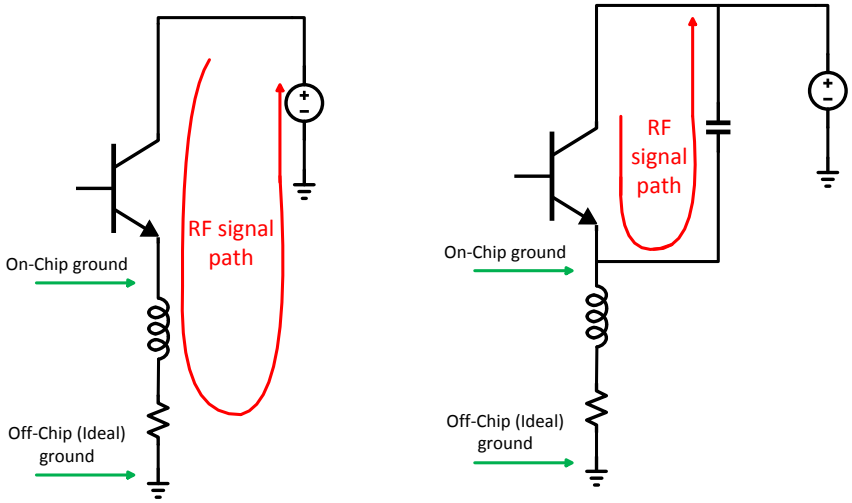


Fig. 2.8 RF signal flow in single ended circuits

We perform some simulations below in order to show the need for decoupling in the case of single ended stages. The lines connecting to the supply pads are thick since they are implemented in the top metals for enhanced resistance to electromigration. Furthermore they are designed wide in order to further suppress the latter effect. Hence their inductance is much less compared to the ground plane inductance, which consists of the thin bottom metals, and is not taken into account. The different scenarios to be simulated are depicted in figure 2.9. The left part of the figure gives the intrinsic stage without any ground parasitics. The emitter is connected to the ideal ground. The inductance depicted in the middle figure models the inductance of the signal path from the on-chip ground to the off-chip ideal ground. The right part contains the decoupling capacitance. The metric used to quantify the effect of parasitics is the maximum available gain, which is the gain after simultaneous conjugate match of both input and output when the amplifier is unconditionally stable [11]. Simulation results are presented in figure 2.10.

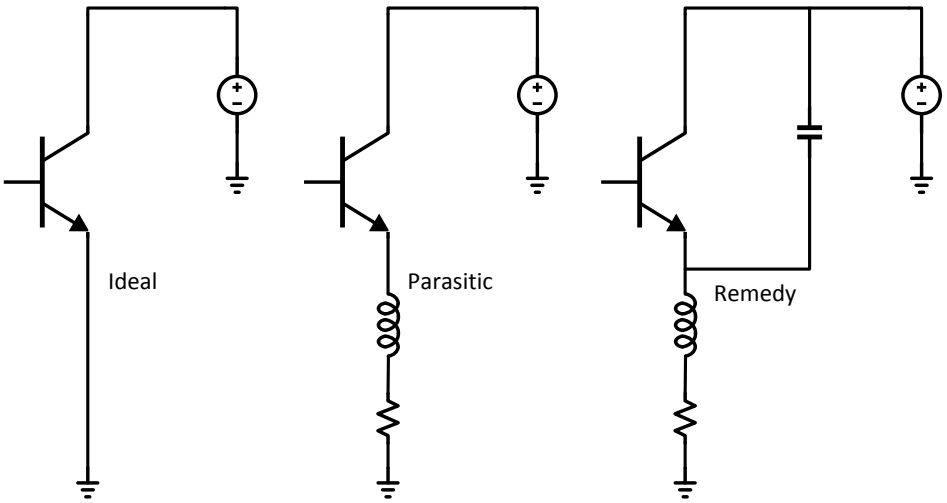


Fig. 2.9 RF signal flow in single ended circuits

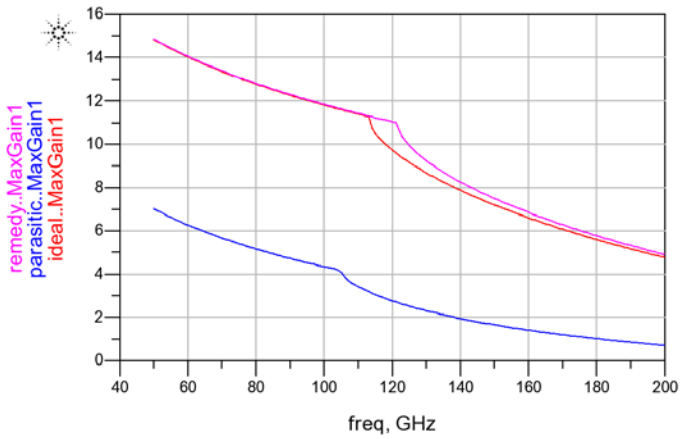


Fig. 2.10 Gain of elemental stage with ideal ground (ideal), ground parasitics into account (parasitic) and decoupled version (remedy).

We notice a dramatic reduction in the gain of the simple common emitter stage in the case that there is no decoupling. Proper decoupling will bring the gain back. However, the value of the capacitance needed may be quite large to fit in the available chip area. Apart from the sufficiently low impedance necessary, we should note from the schematic that effectively there is a capacitor connected from the emitter of the transistor to ground. This may lead to a negative resistance at the base of the transistor which may trigger oscillations [12]. The larger the capacitance, the better the stability will be. From simulations, we note that we need a capacitance larger than 1 pF for stability in the 130 – 140 GHz range. In the figure for a 1pF of capacitance we already see a small extension of the unstable region.

A differential circuit can ease the decoupling requirements, due to the virtual ground that it provides along its axis of symmetry. In figure 2.11 due to the potential distribution under differential operation, we see the current signal flowing from the emitter of one transistor to the emitter of the other, without entering the ground plane and hence getting contaminated by the parasitics. If the differential operation is perfect, there is no need for decoupling. Furthermore, the signal is confined to the stage

itself and there is a reduced possibility for feedback between the stages due to signals from stages flowing into the ground plane.

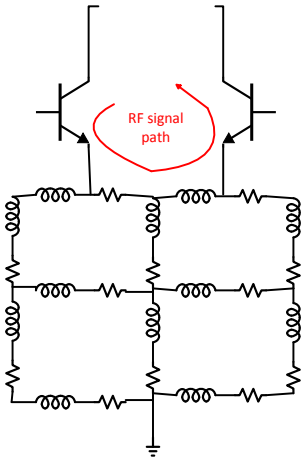


Fig. 2.11 RF signal flow in a differential circuit

In order for the virtual ground concept to hold, we have to ensure pure differential excitation. Hence the two signals should have exactly the same amplitude and 180 degrees phase difference. If this is not the case, there is a common mode component in the excitation which will disturb the balance and the effect of parasitics comes partly again into the picture. We perform simulations in order to prove this. The quantity used here is the operating power gain, i.e. the ratio of power delivered to the network to the input power of the circuit [11]. The devices are modeled as ideal voltage controlled current sources in order to limit the complexity of the structure to be simulated. Next, we give the circuit schematic to be simulated in figure 2.12 and a table with the power gain under various excitations. Under perfect differential operation, the effect of the parasitics is absent. For a phase imbalance of 10 degrees, we notice a slight reduction in gain of ~ 0.7 dB. For amplitude imbalance of ~1 dB the reduction in gain is more and gets close to 2 dB. Hence the amplitude imbalance is more severe regarding sensitivity to parasitics.

This may not seem that severe; however except from the excitation all other aspects of the circuit are highly idealized. Furthermore, in frequencies above 100 GHz, the MaxGain of active cells is in the order of 7-8 dB. Hence a 1 of 2 dB reduction due to imbalance may be a no-go for implementation of an amplifier stage. Finally, the exact effect of ground parasitics is hard to be simulated due to an excessively large computation time that is required. Hence apart from the gain reduction, other unpredictable consequences might go unnoticed during the design and simulation phase due to modeling inaccuracies.

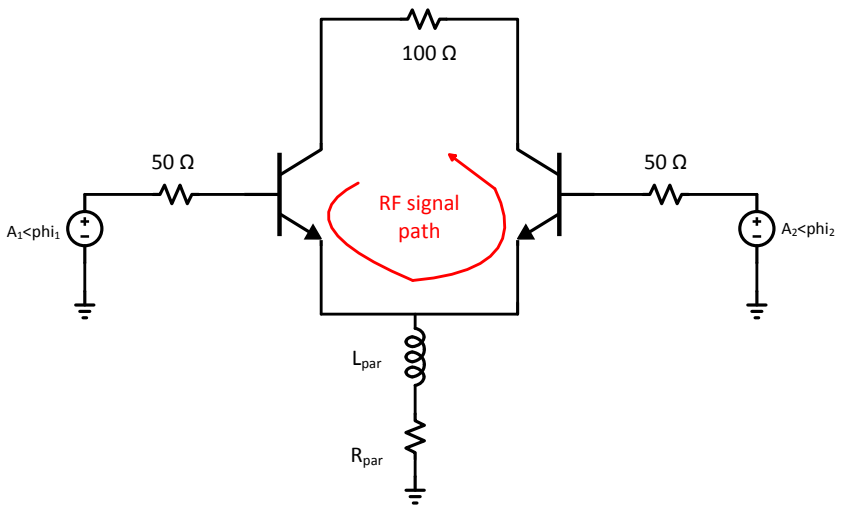


Fig. 2.12 RF signal flow in a differential circuit

Table 2.3 Gain of the circuit depicted in figure 2.12 for different excitations

Situation	Gain (dB)
A_1=1 phi_1=0 A_2=1 phi_2=180 L_{par}=0 R_{par}=0	26.7
A_1=1 phi_1=0 A_2=1 phi_2=180 L_{par}=300pH R_{par}=5 Ω	26.7
A_1=1 phi_1=0 A_2=1 phi_2=170 L_{par}=300pH R_{par}=5 Ω	26
A_1=1 phi_1=0 A_2=0.9 phi_2=180 L_{par}=300pH R_{par}=5 Ω	25

The use of differential circuits brings up some issues that require our attention. The first is the issue of common mode stability. We should ensure that the common mode path has very limited gain so that the onset of oscillations is not possible. The second is the necessity for a differential to single ended conversion. Either we would like to just measure the chip for test purposes or to use it in an actual circuit driving an antenna, the output should be single ended. Hence the design of balanced to unbalanced signal conversion structures (baluns) is necessary. If this structure is unable to provide the conversion of a single ended signal to perfectly balanced signals, i.e. of the same amplitude and 180 degrees out of phase, the differential operation of the circuit is compromised. Hence the challenge is shifted from the necessity of decoupling to the design of efficient and balanced balun structures.

2.5 Conclusions

From the above discussion, we conclude that BiCMOS technology is superior compared to CMOS when it comes to PA design. First and foremost, the increased f_{MAX} allows higher gain values to be achieved. Furthermore, the increased breakdown voltage allows higher output voltage swings and hence increased output power capabilities. For the PA designed in this thesis, we are going to use a BiCMOS 130nm process from IHP, which can achieve an f_{MAX} of about 450GHz. We would also line to employ the advantages of differential operation, so a differential implementation is chosen. In order for measurements of the structure to be possible, a single ended interface is provided at input and output. Baluns perform the single ended to differential and differential to single ended

conversion. The challenges associated with the design of quality baluns are the subject of the next chapter.

2.6 References

[1]Hou et. al. ”A D-band cascode Amplifier with 24.3 dB Gain and 7.7dBm Output Power in 0.13um SiGe BiCMOS Technology”, MWCL,2012.

[2] N. Sarmah, Heinemann and Ulrich R. Pfeiffer, “A 135-170 GHz Power Amplifier in an advanced SiGe HBT technology”, RFIC, 2013.

[3]N. Sarmah, P. Chevalier and Ulrich R. Pfeiffer, “160-GHz Power Amplifier Design in Advanced SiGe HBT Technologies with P_{SAT} in excess of 10 dBm”, IEEE Transactions on Microwave Theory and Techniques, vol. 61, pp. 939 – 947, Feb. 2013.

[4]E. Laskin et. al., “170 – GHz Transceiver with On-Chip Antennas in SiGe Technology”, in 2008 IEEE Radio Frequency Integrated Circuits Symposium, pp.637 – 640.

[5]Sandstrom D. et. al., “W-band CMOS amplifiers achieving +10dBm saturated output power and 7.5dB NF” , IEEE International Solid State Circuit Conference, ISSCC, 2009.

[6]Yu – Sian Jiang, Jeng – Han Tsai, Huei Wang, “A W band medium power amplifiers in 90nm CMOS”, IEEE Wireless and Microwave Component Letters, vol. 18, no. 12, pp.818-820, Dec. 2008.

[7]M. Seo et. al., **“**A 150GHz Amplifier with 8 dB gain and 6dBm in Digital 65nm CMOS using Dummy-Prefilled Microstrip Lines”, IEEE International Solid State Circuit Conference, ISSCC, 2009.

[8] John R. Long, “SiGe Radio Frequency ICs for Low-Power Portable Communication”, Proceedings of the IEEE, vol.93, September 2005.

[9]Sorin Voinigescu, “High Frequency Integrated Circuits”, Cambridge University Press, 2013.

[10]Ali M. Niknejad and Hossein Hashemi (editors), ”Mm-wave Silicon Technology, 60GHz and beyond”, Springer, 2008.

[11]Guillermo Gonzalez, “Microwave Transistor Amplifiers, Analysis and Design, 2nd edition”, Prentice Hall, 1997.

[12]Ali M. Niknejad, “Electromagnetics for High – Speed Analog and Digital Communication Circuits”, Cambridge University Press, 2007.

Chapter 3. Passive Component Design Challenges at mm-wave frequencies

3.1. Introduction

As the center frequency of the design shifts towards higher frequencies, the design efforts are shifted from optimizing the active circuitry towards minimizing the impact of the passive elements. As mentioned in [1] at high frequencies, the impedance levels are kept low in order for high bandwidths to be achieved. Hence, transconductance defines the "active" gain at RF. The difference between this active gain and MSG/MAG comes from the impedance mismatch at input and output of the device. Hence, by proper impedance matching we can get this extra "passive" gain and optimize/improve the performance of the deisgned circuits. This is graphically shown in figure 3.1 taken from [1] for a 0.2um BJT biased at peak f_T.

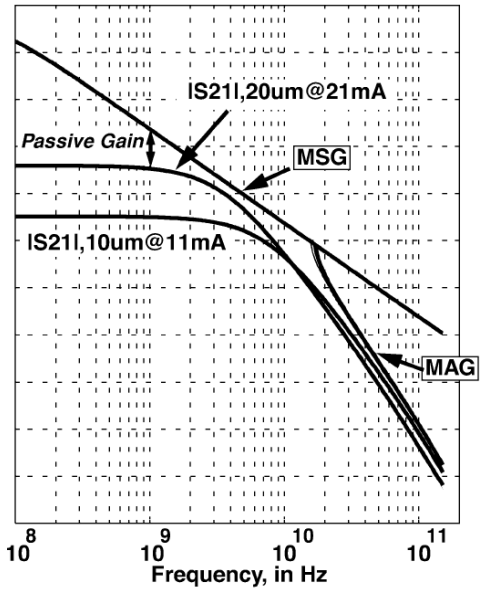


Figure 3.1. Gain versus frequency for a 0,2um BJT biased at peak f_T.

From the above paragraph and the impedance level provided at the higher mm-wave frequency band of even very small parasitics that can alter the terminal impedance of the active devices drastically, see Fig. 3.2, we can understand that the performance of high frequency circuits depends heavily on the passive components that complement the active circuitry to form the total circuit. These passive sections have the role of transforming the source, load or some intermediate impedance of interest to some desired impedance level at the device terminals in order to maximize gain and power transfer or to minimize noise.

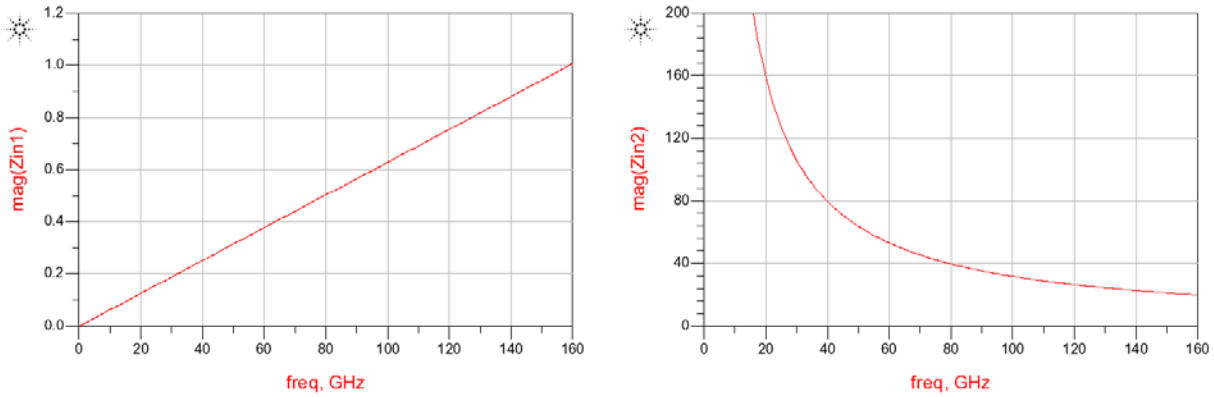


Figure 3.2. Impedance levels for a 1 pH inductor (left) and 50 fF capacitor (right)

These matching networks should be low loss and as compact as possible in order to maintain a relatively small form factor. Furthermore, they should be in compliance with other requirements, such as RF coupling and DC isolation between stages. Finally, in case of differential circuits, we would like them to provide differential to single ended conversion in order to get a single ended interface, which is necessary either for measurement purposes or for driving an antenna during actual system operation.

Typical impedance matching networks for sub-mm wave frequencies may consist of transmission lines or transformers. The typical structure of the resulting circuits is shown in figure 3.3 for transmission lines and in figure 3.4 for transformers. The DC feed inductors provide the DC biasing needed for the active stages and do not allow the RF signal to enter the DC supplies. The coupling capacitors present a low impedance for RF signals, allowing the AC connection between the RF stages and at the same time they act as DC blocks, keeping the biasing of subsequent stages isolated. The balun structures can be implemented in various ways as will be described later in this chapter. The transformer matching networks lead to a much more compact circuit implementation as no extra components are needed for DC feed and RF coupling. The transformer provides DC isolation as its windings are electrically separated. Furthermore, the DC supply can be fed through the center taps which are at zero AC potential due to the virtual ground of differential circuits. Finally, the transformers can act as baluns to interface with single ended inputs and outputs.

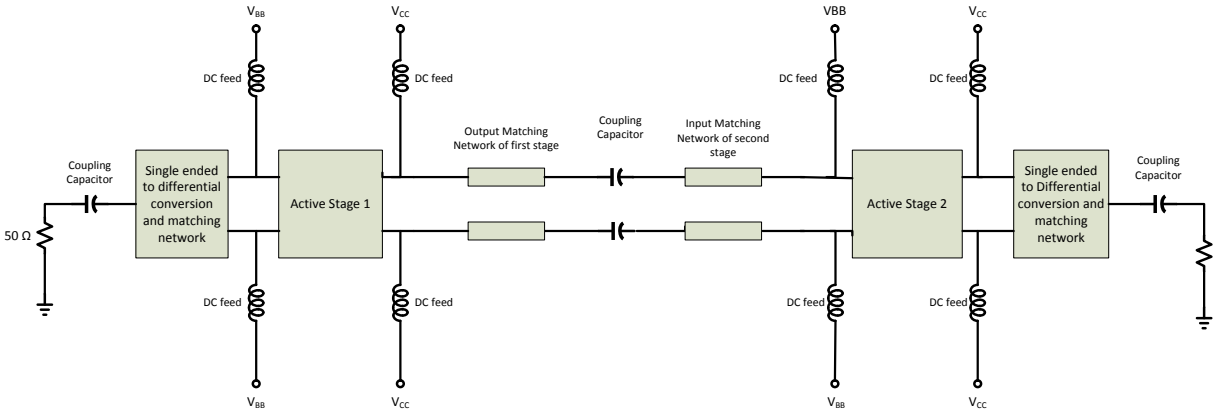


Figure 3.3. Typical circuit structure for transmission line matching networks

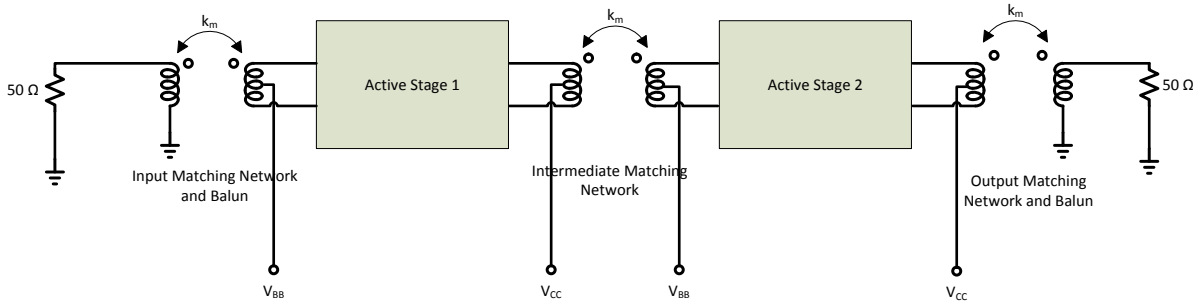


Figure 3.4. Typical structure for transformer matching networks

In this chapter, we first discuss the power loss mechanisms of two major matching network categories, namely lumped element L matching networks and transformers. Transmission line matching networks are considered undesirable in general due to their large dimensions. At 130GHz, the guided wavelength for a silicon back end with SiO_2 inter-metal dielectric (ε_r=4), is λ=1.2mm. Hence a $\lambda/4$ line, often used in matching and biasing feed networks, is 290um long, which is in contrast with compactness considerations. Afterwards we discuss optimization techniques for transformer structures in order to minimize the power losses and obtain the required impedance levels. To quantify the power loss, the quantity of efficiency is used [2]. This quantity can be defined as:

$$n = \frac{P_{OUT}}{P_{IN}} = \frac{P_{OUT}}{P_{OUT}+P_{DISS}} \tag{3.1}$$

Finally, the challenges of operating transformers as baluns are presented and a solution is proposed for alleviating the issue of imbalance between the differential terminals.

3.2 Power Loss Analysis for L matching networks.

All lumped L- matching networks include an inductor and a capacitor. Both of these elements are generally lossy. However, the inductor tends to introduce larger losses than the capacitor, due to the higher impact of conducting losses in finite conductivity metal lines. The losses of a lumped element can be quantified by its quality factor Q. At mm-wave frequencies, the Q of the inductors is usually in the order of 10 or less, while capacitor Qs are higher. Hence we can safely assume that the inductor losses are the dominant ones and the capacitor losses can be neglected. The lossy inductor can be represented by an equivalent series or parallel equivalent circuit. We can convert from one configuration to the other by using the equations (3.2) to (3.4).

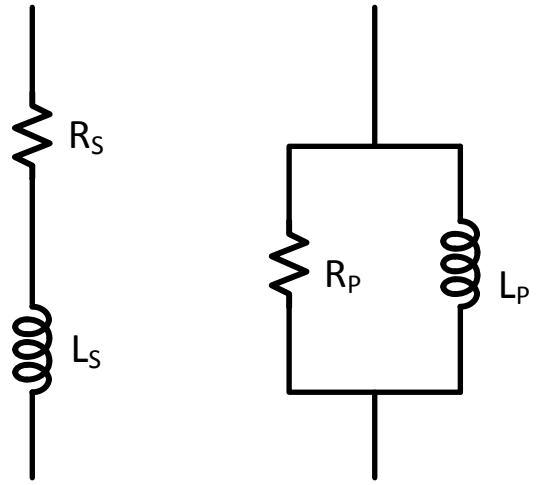


Fig. 3.5 Parallel and series lossy inductor models

$$Q_L = \frac{\omega L_s}{R_s} \tag{3.2}$$

$$R_p = R_s * (1 + Q_L^2) \tag{3.3}$$

$$L_p = L_s * \left(\frac{Q_L^2+1}{Q_L^2}\right) \approx L_s \tag{3.4}$$

In the matching network of figure 3.2, the inductor loss has been modeled with a parallel resistor R_P. At the resonance frequency, L and C resonate with each other and the Q of the network is given by equation (3.4) where Q_L is the quality factor of the inductor.

$$Q_{network} = \frac{R_L // R_P}{\omega L_p} = \frac{R_L}{\omega L_P + \frac{R_L}{Q_L}} \tag{3.5}$$

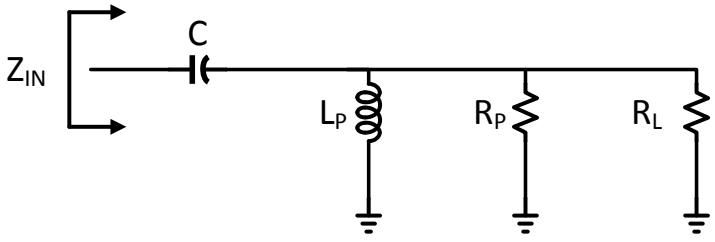


Fig. 3.6 Parallel and series lossy inductor models

Since this a down-converting matching network, the real part of the input impedance is given by equation (3.6) [3].

$$R_{in} = \frac{R_L//R_P}{1+Q_{network}^2} = \frac{1}{1+Q_{network}^2}\frac{R_L}{1+\frac{R_L}{\omega L_P Q_L}} \tag{3.6}$$

From the above equation, we can find the impedance transformation ratio (ITR), which is:

$$ITR = \frac{R_L}{R_{in}} = \frac{R_L}{R_L//R_P}(1+Q_{network}^2) = (1+\frac{R_L}{\omega L_P Q_L})(1+Q_{network}^2) \tag{3.7}$$

From equation (3.7) we find out that as the impedance transformation ratio increases, so does the Q factor of the network. The efficiency of the matching network at resonance is given by:

$$n = \frac{P_{out}}{P_{in}} = \frac{{|V_i|^2}/{2R_L}}{{|V_i|^2}/{2(R_L//R_P)}} \tag{3.8}$$

or

$$n = \frac{\frac{1}{R_L}}{\frac{1}{R_L}+\frac{1}{R_P}} = \frac{1}{1+\frac{R_L}{\omega L_P Q_L}} \tag{3.9}$$

By substituting (3.9) into (3.7) we get the following expression of the inductance required to achieve a certain ITR which is:

$$\omega L_P = \frac{R_P}{Q_L} = \frac{2(Q_L+{1}/{Q_L})}{ITR-2+\sqrt{ITR^2+4Q_L^2\,(ITR-1)}}R_L \tag{3.10}$$

Finally, by substituting (3.10) into (3.9) we are going to get an expression of efficiency in terms of ITR and the inductor Q.

$$n = \frac{1+Q_L^2}{Q_L^2+\frac{ITR+\sqrt{ITR^2+4Q_L^2(ITR-1)}}{2}} \approx \frac{1}{1+\frac{ITR}{Q_L^2}} \tag{3.11}$$

From equation (3.11), we see that the efficiency is inversely proportional to ITR with a proportionality factor of $1/Q_L^2$. This means that for lossier inductors, larger ITRs lead to more loss. This can be explained qualitatively if we model the losses of the inductor with a series resistance and from the fact that the operation of the above matching network is based on a parallel resonance. In the parallel resonance, there is a current circulating between the inductor and capacitor with a value of $Q_{network}$*I where I is the current

through the load. This current flows through the series resistance of the inductor, thus causing losses. The lossier the inductance is, the higher the series resistance and the higher the power dissipation will be. The next figure verifies the above conclusions.

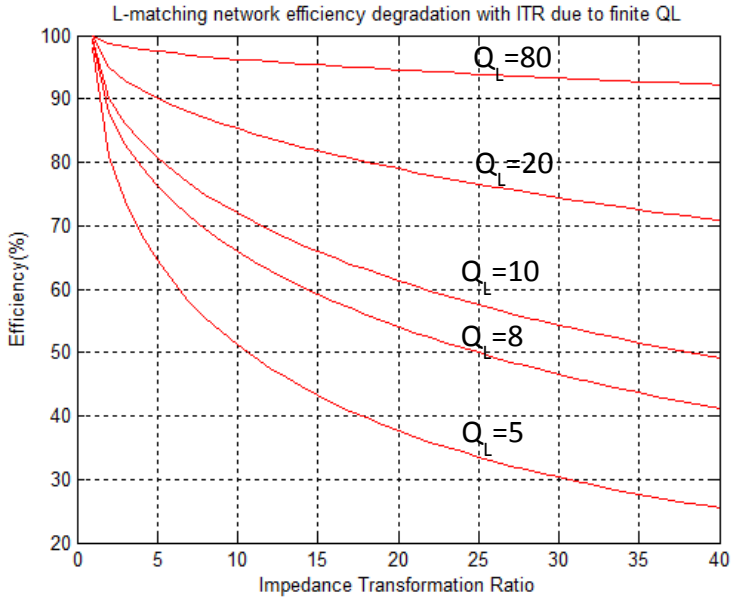


Fig. 3.7 Parallel and series lossy inductor models

From figure 3.3, we see that the lower the Q of the inductor, the higher the impact of the ITR on the efficiency. For the typical case of Q=10 at high frequencies, the efficiency degrades rapidly. For power amplifiers this is extremely important. In order to get a certain output power, we should size the output device accordingly. As we size up the device, the real part of the output impedance goes down due to the higher current and device parasitics [4]. From [5] we know that in order to get a 1 W output power the resistance at the terminals of the device should be ~0.27Ω for a 90/65nm CMOS process and ~1.16Ω for a SiGe HBT process with an f_{MAX} in the order of 230GHz. The characteristics of the two processes are shown in Table 3.1. The lower load impedance of the CMOS active cell stems from the lower V_{MAX} which limits the supply voltage V_{DD}.

Table 3.1. Characteristics of process candidates for Power Amplifier design

Process	**V_{MAX} (V)**	**f_{MAX} (GHz)**	**P_{SAT} (mW/um)**	**R_{Lopt} (Ω)**
65nm/90nm CMOS	1.5	240	0.11 (per um of width)	0.27 (for W=9mm)
SiGe HBT	3	230	1.5 (per um of emitter length)	1.16 (for l_E=0.67mm)

Would we like to transform a 50 Ω load to these impedance values, we would get an ITR=185(!) for the CMOS process and an ITR=43 for the SiGe process. For the CMOS process, this impedance transformation ratio is unrealistic, while for the SiGe process the ITR yields an efficiency of ~50%. This

means that half the power is lost at the matching network, something that makes the matching effort nearly pointless.

3.3 Power Loss Analysis for transformers.

Impedance transformers provide, in first approximation, an insertion loss which is independent of the transformation ratio, as will be proven by eq. 3.30 of section 3.2. This quality makes them more suitable than the L-matching networks when a high impedance transformation ratio is desired. Furthermore, they provide ease of biasing, in differential circuits, through the center taps which realizes a virtual ground, when symmetrical layouts are employed. They also provide DC isolation between subsequent stages, since the energy coupling is only occurring in AC operation. Finally, they can be used as balanced-unbalanced (bal-uns) structures when one of the terminals is grounded.

In order to quantify the power loss, we need a lumped circuit model for the transformer. The one used is shown in figure 3.8. The model is derived from the terminal equations of the transformer.

$$k = \frac{M}{\sqrt{L_1 L_2}} \tag{3.12}$$

k represents the coupling factor of the transformer with values ranging from 0 (no coupling) to 1 (perfect coupling), M is the mutual inductance and L1 and L2 are the self-inductances of the primary and secondary windings, respectively. Inductance kL_P models the portion of the magnetic flux that couples from the primary to the secondary winding, while inductances $(1-k)L_P$ and $(1-k)L_S$ model the flux that is scattered in the environment and does not contribute to the transformer action. Resistors R_P and R_S model the losses of the primary and secondary windings respectively. C_L and R_L comprise the load of the transformer. They can model either the source/load resistance and pad capacitance or the input impedance of the next transistor stage.

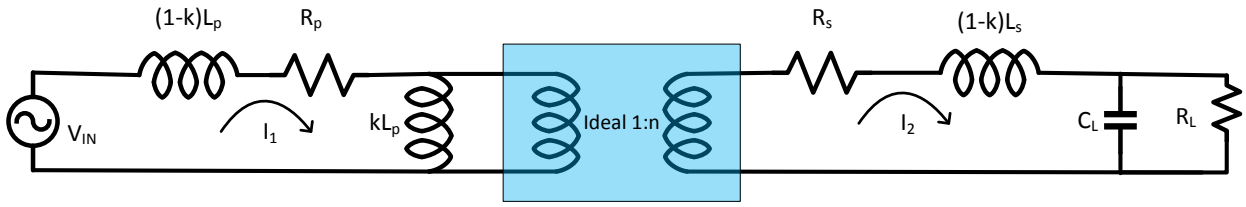


Fig. 3.8 Model used for the transformer power loss analysis

To simplify the above circuit, we can bring the secondary impedance to the primary and we get the following circuit. The situation is depicted below.

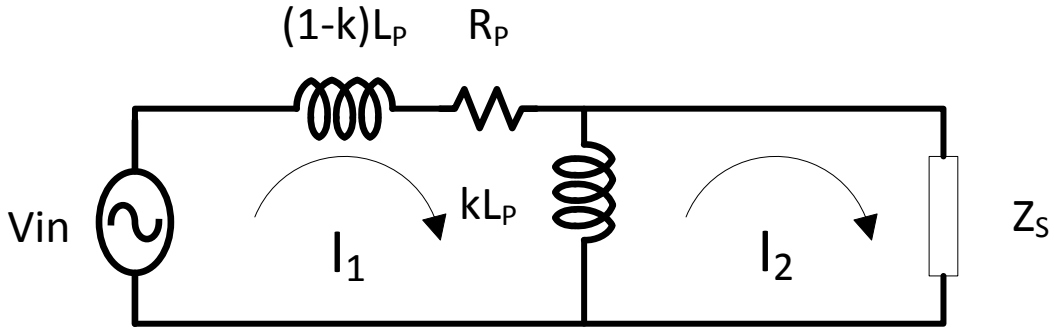


Fig. 3.9 Simplified AC-only transformer model

$$Z_S = \frac{j\omega(1-k)L_S + R_S + R_{eq} + \frac{1}{j\omega C_{eq}}}{n^2} \tag{3.13}$$

where the series combination of the load resistance and pad capacitance are converted from parallel to series and are given by equations 3.14 and 3.15:

$$R_{eq} = \frac{R_L}{1+(\omega R_L C_L)^2} \tag{3.14}$$

$$C_{eq} = \frac{1+(\omega R_L C_L)^2}{\omega^2 R_L^2 C_L} \tag{3.15}$$

The current loop equations are given by equation 3.16 and 3.17:

$$V_{in} = I_1(R_P + j\omega(1-k)L_P + (I_1 - I_2)j\omega\kappa L_P) \tag{3.16}$$

$$(I_2 - I_1)j\omega k L_P + I_2 Z_s = 0 \tag{3.17}$$

From (3.17) we readily obtain (3.18)

$$I_2 = I_1 \frac{j\omega k L_P}{Z_S + j\omega k L_P} \tag{3.18}$$

By substituting (3.18) into (3.16) we get an expression of V_{in} in terms of I_1 alone. The expression is going to be as follows:

$$V_{in} = I_1 R_P + j\omega L_P I_1 + I_1 \left|\frac{j\omega k L_P}{Z_S + j\omega k L_P}\right|^2 \left[\frac{R_S + R_{eq} + \frac{1}{j\omega C_{eq}} + j\omega(1-k)L_S}{n^2} - j\omega k L_P\right] \tag{3.19}$$

We multiply with the conjugate of I1 and we get the following expression:

$$V_{in}I_1^* = |I_1|^2 R_P + j\omega L_P|I_1|^2 + |I_1|^2 \left|\frac{j\omega k L_P}{Z_S + j\omega k L_P}\right|^2 \left[\frac{R_S + R_{eq} + \frac{1}{j\omega C_{eq}} + j\omega(1-k)L_S}{n^2} - j\omega k L_P\right] \quad (3.20)$$

The input power is given as:

$$P_{IN} = Re\{V_1 I_1^*\} = |I_1|^2 R_P + |I_1|^2 \left|\frac{j\omega k L_P}{Z_s + j\omega k L_P}\right|^2 \left[\frac{R_S + R_{eq}}{n^2}\right] \quad (3.21)$$

and the power delivered to the load resistance is given as:

$$P_L = Re\{V_{LOAD} I_2^*\} = |I_1|^2 \left|\frac{j\omega k L_P}{Z_s + j\omega k L_P}\right|^2 \left[\frac{R_{eq}}{n^2}\right] \quad (3.22)$$

The efficiency is readily obtained be dividing (3.22) by (3.21) and we get the following:

$$n = \frac{P_{LOAD}}{P_{IN}} = \frac{\frac{R_{eq}}{n^2}\left|\frac{j\omega k L_P}{Z_S + j\omega k L_P}\right|^2}{R_P + \left[\frac{R_S + R_{eq}}{n^2}\right]\left|\frac{j\omega k L_P}{Z_S + j\omega k L_P}\right|^2} \quad (3.23)$$

Finally, by dividing both the numerator and the denominator with the absolute term and substituting R_{eq} as in equation (3.14) we are going to get the following.

$$n = \frac{\frac{1}{1+(\omega R_L C_L)^2}\frac{R_L}{n^2}}{\frac{R_L}{n^2}\frac{1}{1+(\omega R_L C_L)^2} + \frac{R_S}{n^2} + R_P \frac{(\frac{R_S}{n^2} + \frac{R_L}{n^2}\frac{1}{1+(\omega R_L C_L)^2})^2 + (k\omega L_P + \omega(1-k)\frac{L_S}{n^2} - \frac{1}{\omega C_{eq} n^2})^2}{(k\omega L_P)^2}} \quad (3.24)$$

By recognizing that $L_P = {L_S}/{n^2}$ and in order to maximize efficiency, it is easily seen from (3.24) that we should set

$$(k\omega L_P + \omega(1-k)\frac{L_S}{n^2} - \frac{1}{\omega C_{eq} n^2})^2 = 0$$

$$\omega L_S = \frac{1}{\omega C_{eq}} \quad (3.25)$$

Qualitatively, we can explain (3.25) as follows: $j\omega k L_P$ and $j\omega(1-k)L_s/n^2$ resonate with $1/j\omega C_{eq}n^2$. This is a parallel resonance; hence the current flowing through Rp is minimized, thus minimizing the losses.

After substituting (3.25) into (3.24) we get

$$n = \frac{\frac{1}{1+(\omega R_L C_L)^2}\frac{R_L}{n^2}}{\frac{R_L}{n^2}\frac{1}{1+(\omega R_L C_L)^2}+\frac{R_S}{n^2}+R_P\frac{(\frac{R_S}{n^2}+\frac{R_L}{n^2}\frac{1}{1+(\omega R_L C_L)^2})^2}{(k\omega L_P)^2}} \quad (3.26)$$

and by setting $R'_{eq} = \frac{R_L}{n^2}\frac{1}{1+(\omega R_L C_L)^2}$ we get the following expression:

$$n = \frac{R'_{eq}}{R'_{eq}+\frac{\omega L_P}{Q_S}+\frac{\omega L_P}{Q_P}\left(\frac{1}{kQ_S}+\frac{R'_{eq}}{k\omega L_P}\right)^2} \quad (3.27)$$

where $Q_P=\omega L_P/R_P$ and $Q_S=\omega L_S/R_S$.

The primary inductance is an extra degree of freedom for efficiency maximization. If we differentiate (3.27) with respect to ωL_P we get according to [6]:

$$\omega L_P = \frac{a}{1+a^2}\frac{R_L}{n^2} \quad (3.28)$$

Where

$$a = \frac{1}{\sqrt{\frac{1}{Q_S^2}+\frac{Q_P}{Q_S}k^2}} \quad (3.29)$$

Finally, the expression for the maximum efficiency, by choosing the value of pad capacitance and primary inductance accordingly, is going to be:

$$\eta_{max} = \frac{1}{1+\frac{2}{Q_P Q_S k^2}+2\sqrt{\frac{1}{Q_P Q_S k^2}(1+\frac{1}{Q_P Q_S k^2})}} \quad (3.30)$$

We note that the maximum efficiency depends only of the factor k of the transformer and the Q factor of the windings. In contrast to the L matching networks, we don't see any dependency on the ITR. However a larger transformation ratio may require a large inductance value. So after some point the problem shifts to the design of high Q large or small inductance transformers. Furthermore, the large inductors needed for high ITRs may have an insufficiently high SRF.

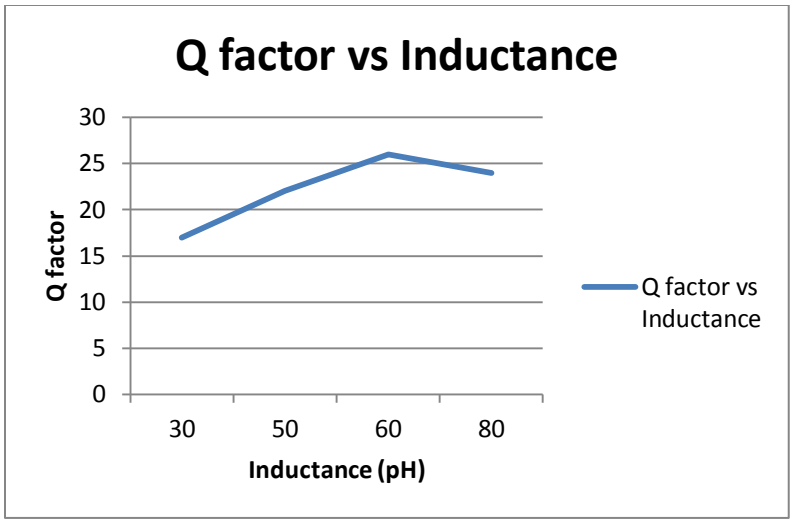


Figure 3.10. Inductor Q factor as a function the inductance value

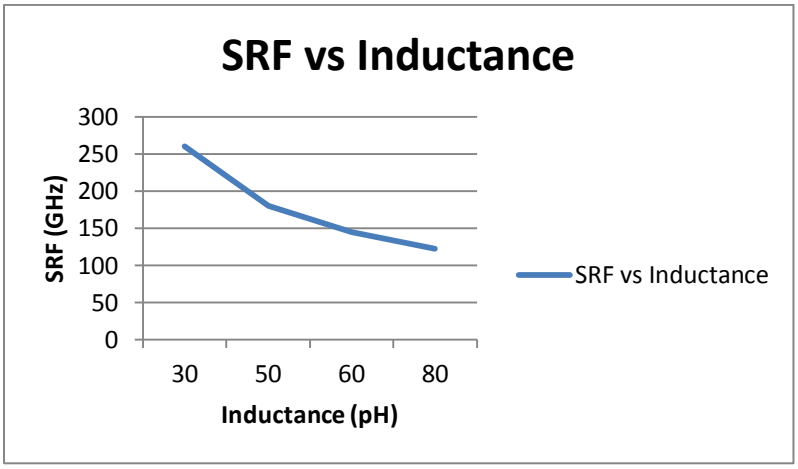


Figure 3.11. Inductor SRF as a function of the inductance value

Figure 3.10 depicts the dependence of the Q factor on the inductance value, while figure 3.11 the dependence of the SRF on the inductance value for an IHP 130nm SiGe BiCMOS process for inductors implemented on the top metal layer which is 3um thick. We find a large Q in the order of 20 for the inductors. The relatively high values stem from the fact that the metal on which the inductors are implemented is thick and 10um away from the conductive substrate. For increasing values of inductance, the SRF becomes the main limiting factor since the Q factor still assumes high values. Hence, transformation ratios that require large inductance values are not easily implemented with transformers at high frequencies. Nonetheless, for a variety of moderate transformation ratios, transformers are still the matching networks of choice since they also offer ease in DC biasing and RF coupling. A table of comparison between L matching networks and transformers concludes the section.

Table 3.2 Comparison of L-matching network and transformer matching network properties.

Property	L network	Transformer
Loss Vs ITR	Poor	Good
DC Basing	Extra Components needed	Via center tap
RF coupling	Extra Components needed	Via winding coupling
Component Availability	Components available in library	Custom design required
Compactness	Moderate	Good
Isolation from adjacent structures	None	Good for high coupling factors

3.4 Transformer Design and Optimization

In this chapter, we provide some guidelines for the realization of efficient transformers. The presented flow will start from the choice of the coupling factor, and a discussion regarding the optimal geometry and topology. In order to optimize the operation of the structure for the desired frequency, the inductances of the primary and secondary windings have to be chosen accordingly. Finally, by loading the transformer with the appropriate capacitances, we can further optimize its efficiency.

3.4.1. Choice of coupling factor

From (3.30) we can construct our first transformer design aid. Assuming the use of the proper primary self-inductance and pad capacitance, we can see the effect of winding quality factors and the coupling factor on transformer efficiency. In figure 3.12 we provide several graphs that depict the efficiency in terms of the coupling factor and the windings quality factors. From these curves, conclusions regarding the transformer design can be made.

For a limited winding coupling factor, the efficiency of the transformer degrades. Based on this argument, we have to aim for a coupling factor as high as possible for efficient operation. However, the efficiency also depends on the quality factors of the windings and the initial argument would hold only if infinite Q windings would exist. From figure 3.12, it is obvious that when the quality factors are relatively low, there is a point beyond which the increase of coupling factor does not increase the efficiency considerably and does not worth the extra design effort. Furthermore, as we will see later, transformers operating at higher frequencies do not allow high coupling factors due to the increased inter-winding capacitance that limits the SRF of the device below the desired frequency.

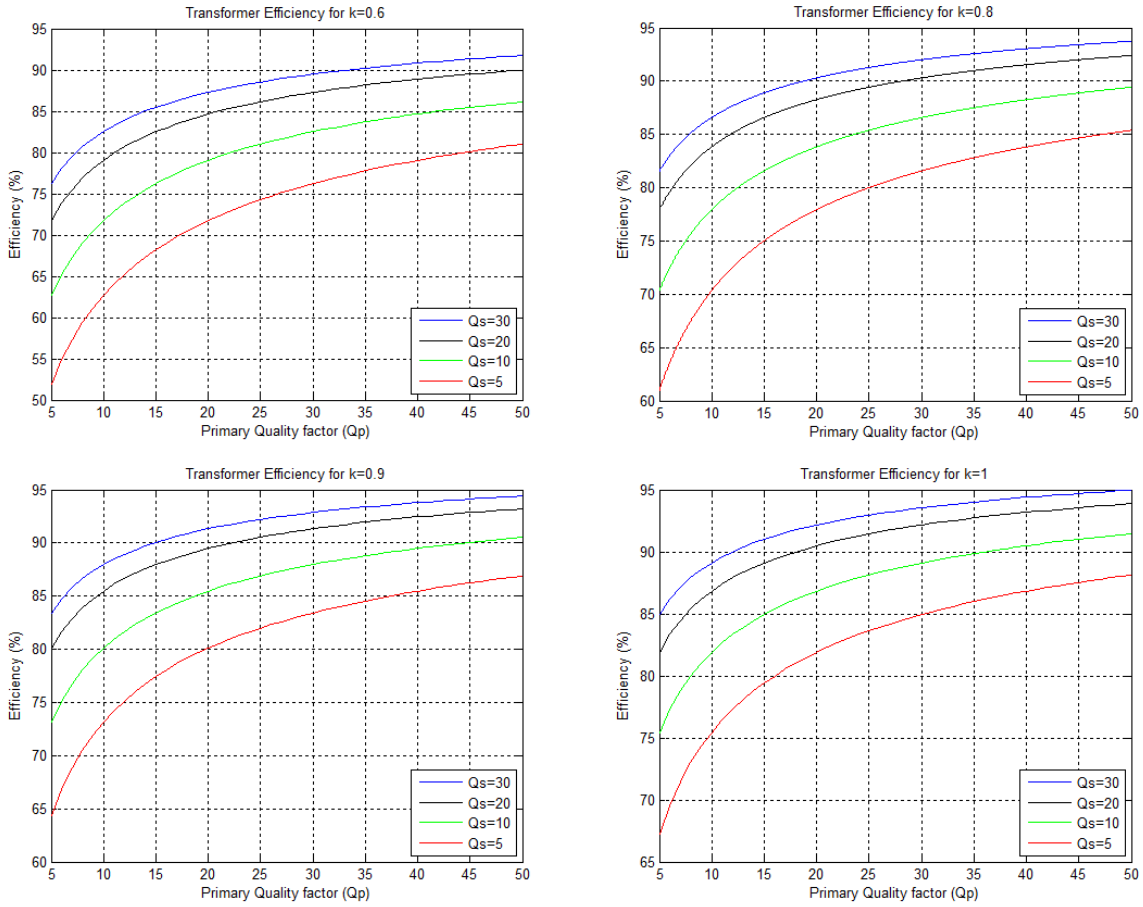


Fig. 3.12 Transformer efficiency in terms of winding quality factors and transformer coupling factor

Next, by assuming winding quality factors of about 20, we construct a plot of efficiency vs coupling factor in order to see the extent of efficiency degradation as the latter decreases. The graph is given in figure 3.13. For values of coupling factor above 0.5, the efficiency does not vary significantly. However, for coupling factors below 0.5, the deterioration is rapid.

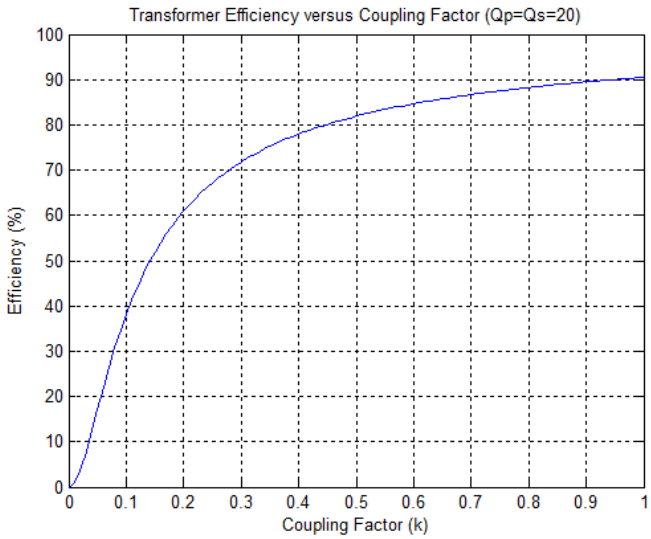


Fig. 3.13 Transformer efficiency versus coupling factor

Finally, we would like to see how the coupling factor k varies, by increasing the size of the windings. By employing as windings the inductors whose performance was evaluated in the figures 3.10 and 3.11, we plot the coupling factor as a function of the inductance value. The results are depicted in figure 3.14. It is seen that the coupling factor increases for higher inductance values. This is due to the fact that for larger winding diameters, the area where coupling happens increases. The increase is almost linear. Hence a passive with larger dimensions will give a better coupling factor.

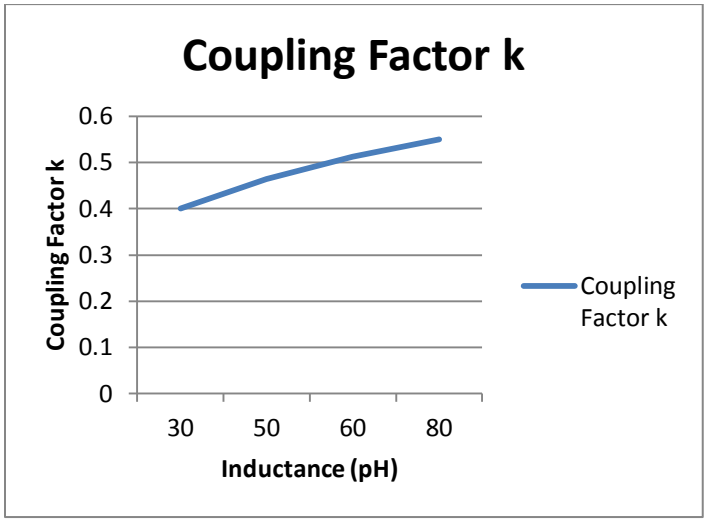


Fig. 3.14 Coupling Factor k versus winding inductance

3.4.2. Transformer Winding Diameter

Based on the transformer model presented in figure 3.8, we can qualitatively draw some conclusions about the optimum diameter of the windings, which in turn affect the primary inductance of the transformer [8]. In figure 3.15 the transformer model is augmented by the parasitics from the windings to the silicon substrate [7], [8]. Based on this model, the limitations regarding the size of the transformer winding diameter can be understood.

For low winding diameters, the value of the coupling factor k is small. This causes kL_P be also small) and the power is shorted to ground. On the other hand, for large diameters, the $(1-k)L_P$ component becomes large and in conjunction with the parasitic capacitance to substrate, they form a low pass network which attenuates the signal as it propagates from input to output. Furthermore, the quality factors of the windings degrade for higher diameters. These two mechanisms cause an increased loss for larger dimension transformers despite the increased coupling factor.

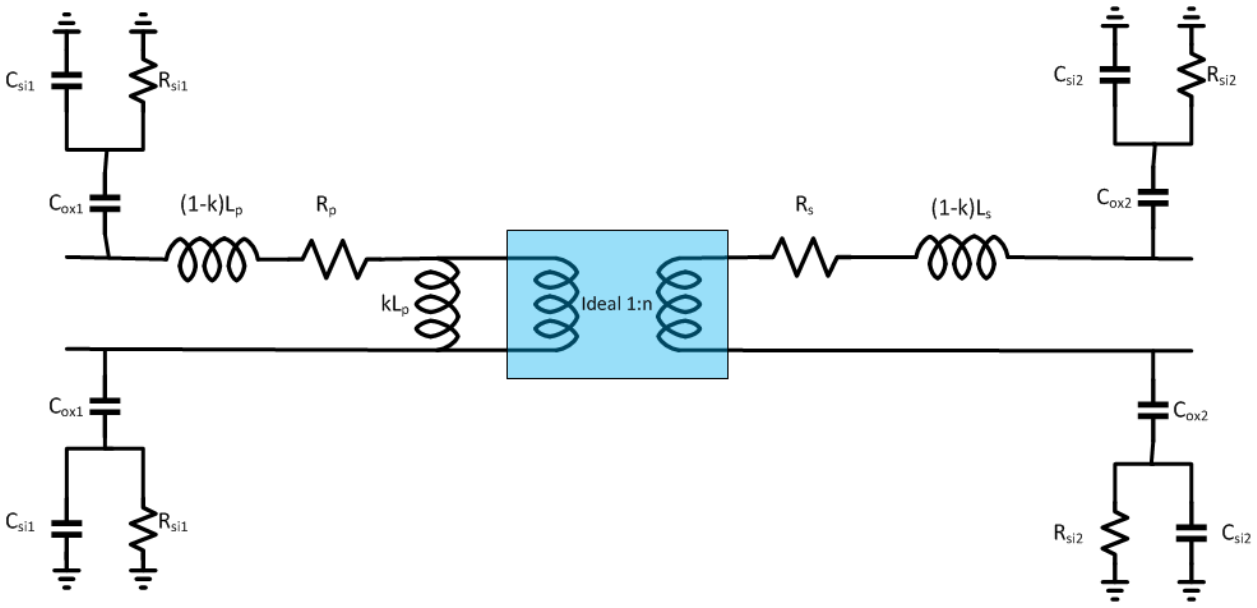


Fig. 3.15 Transformer model including parasitics from the windings to substrate

The graph that is presented in figure 3.16 shows simulation results for the Insertion Loss (IL) of the transformer as a function of the inner diameter of the transformer windings. The simulations were performed for an ST Microelectronics 55nm BiCMOS process BEOL. The top two metals are respectively 0.8um and 3um thick with the distance between them being 1.5um.

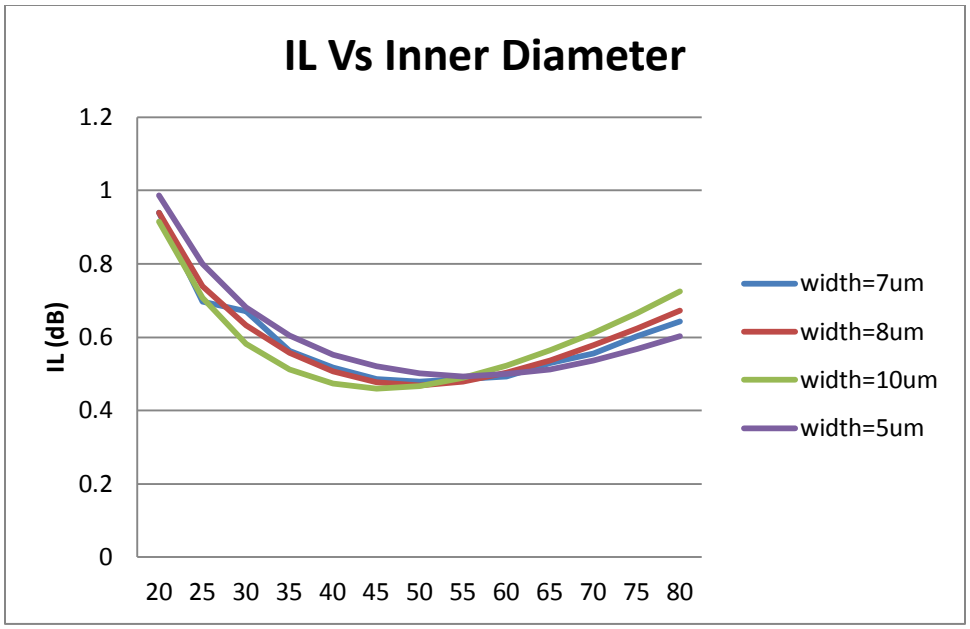


Fig. 3.16 IL of transformer as a function of inner diameter and winding width

The arguments stated above are verified by the simulation results. There is a quite wide range of winding diameters that give the optimum insertion loss. For diameters from 40um to 60um we get the minimum IL with small variation on the optimal inner diameter versus line width of the winding.

The winding width modulates both the series resistance (i.e., smaller widths provide a higher resistance) as well as the capacitance to ground (i.e., larger widths provide a larger capacitance). It is interesting to note a switch in the loss mechanisms when moving from lower diameters to large ones. The phenomenon can be summarized as follows.

- For small diameters the losses increase with a reduction of the winding width. This indicates that the loss mechanism is dominated by the series resistance.
- For larger diameters the losses increase with the increase of the winding width. This indicates that the loss mechanism is dominated by the shunt capacitance.

3.4.3 Winding Topology and Orientation

As we move towards higher frequencies, the role played by the inter-winding capacitance of the transformer becomes significant. The complete transformer model is given in figure 3.17. The inter-winding capacitance is modeled by elements C_{iw1} and C_{iw2} and is distributed between the transformer windings.

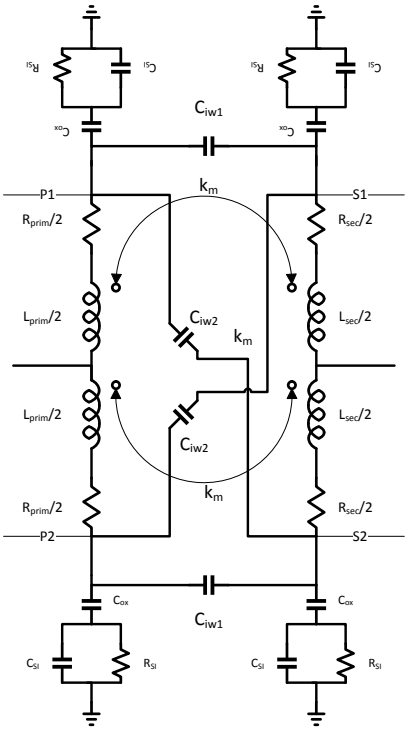


Fig. 3.17 Full lumped transformer model

The coupling between the windings is realized via two mechanisms. The first is via the magnetic field and is the classical coupling mechanism that dominates at lower frequencies. The second one is via the electric field and is due to the presence of the inter-winding capacitance and becomes the dominant coupling mechanism as we move to higher frequencies. How these two mechanisms contribute to the power transfer depends on the orientation of the windings.

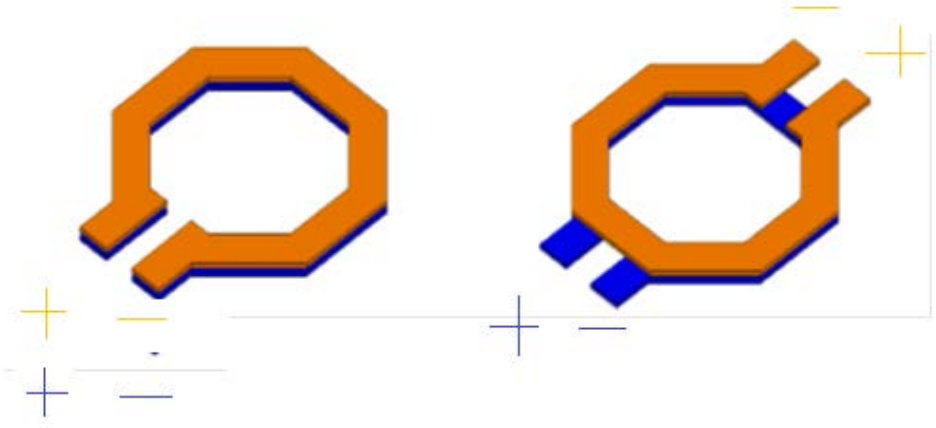


Fig. 3.18 Non-inverting (left) and inverting (right) transformer topologies (figure taken from [9])

In figure 3.18, the non-inverting (left) and inverting (right) transformer topologies are depicted. The classification stems from the polarity of the voltage across the load connected to the secondary. Let's assume that we always drive the primary (blue color) with a voltage source and the current flows in a clock-wise direction. By Lenz's law, we can find the current flowing in the secondary. In the non-inverting configuration, the current flows through the load in such a way so that the voltage across it is the same polarity as in the primary. In the inverting configuration the voltage has the opposite polarity, hence the name inverting.

What is more important is the difference between these topologies in terms of power transfer. In figure 3.19, a high frequency transformer model is presented taken from [8], where all the parasitics are moved to the secondary winding. C_P is the capacitance associated with the primary winding, C_S is the capacitance associated with the secondary and C_O is the inter-winging capacitance. n is the turns ratio and is positive in the case of a non-inverting configuration and negative in the case of an inverting configuration for the dot convention shown in the model.

For a non-inverting configuration, i.e. n>0, the Co/n and Ls forn a bandpass filter with a transmission zero in the passband due to the parallel action of the capacitor and the inductor. This zero causes a notch in the frequency response which impedes power transfer. In the case of the inverting configuration, i.e. n<0, the capacitor has a positive reactance which decreases for increasing frequency. Hence, the structure resembles a low pass filter with a cutoff frequency which is higher than the transmission zero of the non-inverting configuration.

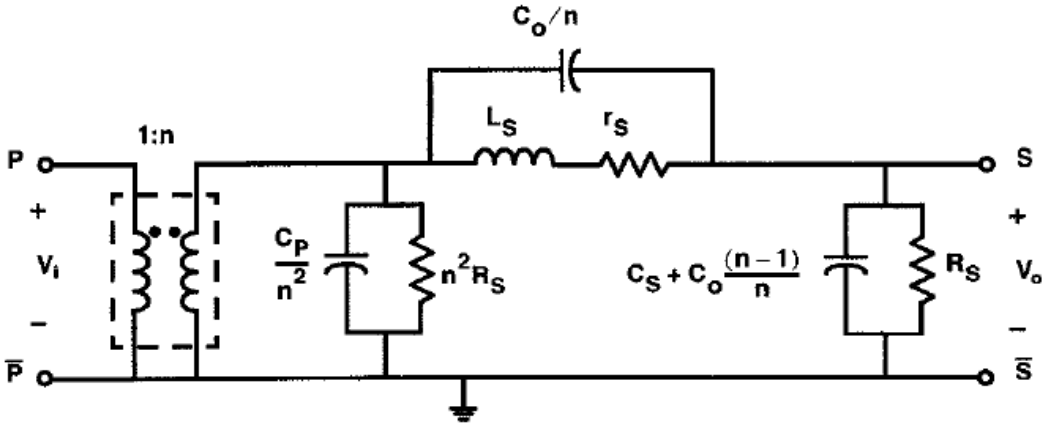


Fig. 3.19 High frequency transformer model with the primary parasitics referred to the secondary (taken from [8]).

Finally, according to the relative position of the windings, we can identify interleaved and stacked transformers as shown in figure 3.20. According to [7] the stacked configuration is preferable in terms of power transfer due to the bigger mutual area that it gives. The advantage of inter-leaved transformers is apparent when there is only one thick metal layer in the BEOL of the process. This topology may also be used when the different metal layers are very far apart vetically. However, at higher frequencies, in order to get a sufficiently high SRF, the windings should be separated by a certain distance in order for the inter-winding capacitance to be reduced and a sufficiently high SRF to be achieved. It turns out that when both topologies are designed in order to have a high SRF, the stacked one presents a higher coupling factor, hence it is the configuration of preference.

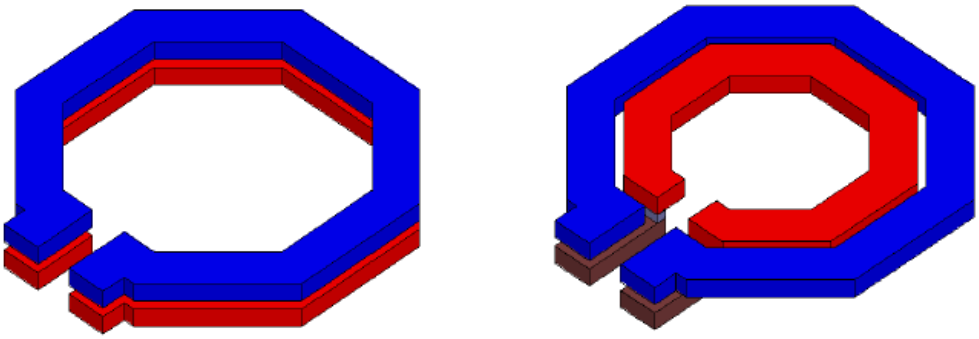


Fig. 3.20 Stacked (right) and interleaved (left) transformers (figure taken from [9]).

3.4.4. Transformer design guidelines summary

Based on the above discussion, we can summarize some basic guidelines that will help us in the design of integrated transformers.

- The transformer efficiency depends on the coupling factor and the windings quality factors. The winding quality factors are determined from the process and can be found from EM simulations. For relatively low winding quality factors, there is a value for the coupling factor, beyond which the gains in efficiency are limited. Hence, any further design effort to increase the coupling factor will not necessarily result in a higher efficiency.
- There is a range of optimum winding diameters, which will give the minimum losses. This optimum is wide and various inductance values can be chosen. For example the designer might choose an inductance value which resonates with a certain device capacitance in the circuit.
- An inverting configuration is preferable in terms of power transfer since there is no notch in the transfer as in the case of the non-inverting configuration.
- In the case where more than one thick metal exist, the stacked transformer configuration is the one of choice since it gives a higher coupling factor compared to the interleaved one.

3.5 Transformer Design Challenges for mm-wave transformers

In the previous sections of the chapter, we gave guidelines for the design of efficient transformers. However, there are cases where these guidelines cannot be followed due to limitations that the environment of the circuit poses. Inability to follow these guidelines closely, leads to increased losses. We address these issues in this section.

3.5.1 Loading Capacitance of the Secondary Winding

From equations (3.24) and (3.25) it is shown that the efficiency of a transformer is maximized when the secondary inductance resonates with the load capacitance of the secondary winding. However, when this frequency is too high and the inductance of the secondary is not low enough, there are cases where this may not be possible.

In our power amplifier design case, there are two possible places for the transfomers. The first possible position is for a transformer to be between active stages acting as an inter-srage matching network. The second case is for the transformer to be at the periphery of the circuit, i.e. as a balun providing a single ended interface between an active stage and the pads. In the first case, the load of the secondary winding is the input capacitance of the next stage. In the second case, the load is the pad capacitance. Hence, these capacitances should be chosen accordingly in order to resonate with the secondary inductance.

In the case of a transformer acting as an inter-stage matching network, the capacitance of the next stage is of interest and should be chosen in order to maximize the efficiency by resonating at the frequency of operation. There are usually other considerations that dictate the choice of the particular capacitance. The capacitance can be scaled by altering the size of the device. In the case of a PA, the size of the device is sized in order to get a certain amount of output power. In the case of an LNA, the size affects the real part of the optimum noise impedance. The design of transformer for matching of these particular impedance levels might not be compatible with low loss operation.

In the case of a transformer being on the periphery of the circuit, the pad capacitance is of importance. The pad capacitance is directly related to pad dimensions. The minimum pad dimensions are dictated by the measurement equipment in the case of a stand-alone PA that needs to be measured. For a GSG (Ground – Signal – Ground) pad, the minimum Signal pad opening for reliable probe landing is 35-40um. The lower metals of the back-end will be used to shield the pad from the underlying substrate. Furthermore, the distance between the Signal and Ground pads are restricted by the pitch of the probe, which is 75um-100um center-to-center for measurements above 100GHz. For a GSG configuration with minimum signal opening, a 75um center to center pitch and 10um distance from the bottom metals, the capacitance is around 27fF. If optimization points to a lower value that this, the particular value is not practical. Hence the maximum efficiency given by equation (3.30) is not achievable.

At this part of the section, in order to obtain a clear view of the challenges of designing high performance transformers and explore the design space, we give capacitance values of various active cells and pad configurations. Furthermore, we simulate some passive structures and we find the optimal capacitances for efficiency maximization. These optimal values are compared to the available active cell and pad capacitances. Finally, the deviation of the actual capacitance values from the optimal ones and the associated degradation in power transfer is given.

In table 3.3 various GSG pads are simulated and the capacitance values are given. The dimensions of the signal pad are varied; the pad center-to-center pitch remains the same and equal to 100um and the ground pads are fixed to 100x100um^2. The bottom two metals, i.e. Metal1 and Metal2, are shunted and used to shield the pad and act as the ground plane.

Table 3.3 Pad Capacitance for different signal (S) pad dimensions

Pad Dimensions	Pad Capacitance (fF)
40x40	17.5
50x50	25
60x60	40
70x70	71
80x80	117

In table 4.4, the capacitances associated with standard active cells are given. In our final power amplifier design, the last stage consists of five parallel devices each employing an emitter area of 0.13x1.1 um^2. The driver stage consists of three parallel devices with the same emitter area.

Table 3.4 Input and Output Capacitance for various active cells

Device Size	Input Capacitance (fF)	Output Capacitance (fF)
2x0.13x1.1	8.5	2.7
3x0.13x1.1	12.43	4
4x0.13x1.1	16.55	5.15
5x0.13x1.1	20.52	6.27

For a transformer with a diameter of 50um, like the one we used as the output matching network for our PA, the pad capacitance that optimizes the power transfer of the transformer is 11 fF. By terminating the secondary with this capacitance value, we get a loss of 1.9 dB. However, the pad capacitance of the minimum pad reliable for measurement is 25 fF. This will lead to a loss of 2.5 dB, which is 0.6 dB more than the optimum with the latter being impossible to be achieved. By making the transformer smaller, we can reduce the loss of the matching network at the expense of output power from the active stage. It turns out that this approach leads to inferior overall performance; hence we accept the extra losses of the transformer.

The purpose of the inter-stage transformer is to convert the input impedance of the final active stage to the conjugate of the output impedance of the driver stage. The design is done based on this observation and does not take into account minimum loss operation. Hence, the reflection losses will be minimized but the intrinsic losses of the passive can be considerably increased. For example, for our inter-stage matching network, the optimum capacitance value is 8 fF and results in a loss of 1.82 dB. When loading the particular transformer with the 20 fF input capacitance of the last stage, we get a loss of 2.8 dB. Of course if we don't perform conjugate match, the losses are bigger due to the reflections. Hence, we have to live with this extra 1 dB of power loss in the passive.

3.5.2 Primary Inductance for Impedance Matching

From previous discussions, we saw that there is a particular value for primary inductance that leads to maximum efficiency. This value of inductance is not always practical in the case we would like to design the self-inductance of the primary winding in order to resonate with some device capacitance at some particular frequency, which is usually the case. Hence more deviation from the optimum occurs. The way to circumvent this is to size the device in order to provide capacitive impedance that resonates with the optimum inductance of the winding at the desired frequency. However, this is not always possible especially in cases when we size the device for a certain output power, as in the case of a PA, or when we would like to set the real part of the input impedance of an LNA.

3.5.3 Inter-Winding Capacitance, Self – Resonance Frequency and Coupling factor

Another culprit when operating integrated transformers at high frequencies is the inter-winding capacitance. We have already mentioned that, by choosing an inverting transformer configuration, we can avoid the transmission zero at the passband and hence achieve more power transfer and hence efficiency. However, the inter-winding capacitance has the effect of lowering the self-resonance frequency of the structure. Specifically, this is the main reason for the low SFR of transformers with a high coupling factor.

In our ST 55nm process, we performed some simulations to verify this. In order to study the effect of inter-winding capacitance, we compared stacked transformers implemented in different metals. The first transformer implementation was done in Metal8 (top metal) and the Alu-cap layer. The distance between these layers is 0.64um. At 100 GHz, we got a coupling factor of 0.8 which is very strong but also an SFR at 140GHz. This is not suitable for operation at 100-130GHz because after the addition of the device and pad capacitances, the SRF will be even lower. The second implementation was performed at metal 8 and metal 7. The distance between the windings is 1.5um, i.e. more than double compared to the previous case. Now the coupling factor has dropped to 0.6 but SFR has risen to 200 GHz. The results are summarized at table 3.1 below.

Table 3.5. Comparison of transformers with different winding distance

Metals	Inter-metal distance (um)	Coupling Factor k	SRF (GHz)
M8-Alu	0.64	0.8	140
M7-M8	1.5	0.6	200

From the above table we note the real drawback of the inter-winding capacitance, which is the fact that we have to accept a limited coupling factor in order to achieve a high SRF. We cannot afford a higher factor k due to the fact that the SFR shifts to lower frequencies and operation at relatively high frequencies is not possible. From 3.2 we see that a reduced coupling factor has a big impact on transformer efficiency. Hence high performance transformers at sub-mm wave frequencies are not easy to construct.

In the previous section, we wrote down the capacitance values of the pads as well as the capacitances associated with the active cells. By loading the transformer with these capacitances the resonance of the total structure should happen in the frequency band of interest. By loading the transformers simulated in the previous sections with the capacitances of the pad and the active cells, we can find the range of transformer diameters best suited for operation around 135 GHz. A pad capacitance of 25fF and a In this group of transformers if there is one that can perform the required impedance transformation, it is the one of choice. If not, matching networks other than 1:1 transformers should be sought.

In the first column of Table 3.6, we give the diameter of the passive component. In the second column, the intrinsic SRF of the passive is given. In the third column the capacitances that load the passive are given. Finally, the frequency range across which the loaded passive gives acceptable loss is given. The acceptable is given as the minimum loss, at some particular frequency, plus/minus 0.5 dB. The study is repeated for the intermediate and input stages. The loading chosen is the capacitive loading of the pads and the active stages. The dimensions of the latter ones have been chosen based on the discussion in Chapter 4. The passives were simulated in the configurations that are used in the final circuit. In Tables 3.6 and 3.8 the transformers are configured as baluns, to provide the single-ended interface, by grounding one terminal of the winding that is connected to the pad. The intermediate stage, in Table 3.7, is configured in a fully differential manner.

Table 3.6. SRF of various transformer structures and Resonance Frequency bands after pad and active device capacitive loading. Case study for the output stage (differential to signle-ended).

Diameter	Intrinsic SRF (GHz)	Loading Capacitance	Loaded RF (GHz)
30um	180	C_{PAD}=25fF C_{outo}=6.2fF	93-140
40um	170	C_{PAD}=25fF C_{outo}=6.2fF	87.1-148
50um	153	C_{PAD}=25fF C_{outo}=6.2fF	93-145
60um	115	C_{PAD}=25fF C_{outo}=6.2fF	64-130

Table 3.7. SRF degradation after loading of the transformers with the pad and device capacitances. Case Study for the intermediate stage (differential to differential).

Diameter	Intrinsic SRF (GHz)	Loading Capacitance	Loaded RF (GHz)
30um	231	C_{outd}=20.5fF C_{ino}=4fF	106-135
40um	178	C_{outd}=20.5fF C_{ino}=4fF	94-140
50um	160	C_{outd}=20.5fF C_{ino}=4fF	96-140
60um	122	C_{outd}=20.5fF C_{ino}=4fF	85-125

Table 3.8. SRF degradation after loading of the transformers with the pad and device capacitances. Case Study for the input stage.

Diameter	Intrinsic SRF (GHz)	Loading Capacitance	Loaded RF (GHz)
30um	180	C_{PAD}=25fF C_{ind}=12.43fF	127-177
40um	170	C_{PAD}=25fF C_{ind}=12.43fF	88-153
50um	153	C_{PAD}=25fF C_{ind}=12.43fF	93-140
60um	115	C_{PAD}=25fF C_{ind}=12.43fF	65-130

We note that the fully differential structure has a slightly higher SRF than the differential – single ended one. This is due to the inter-winding capacitance. The different voltage distribution across the windings, due to the forced ground at one terminal, modulates the inter-winding capacitance differently, i.e. there is a variation in voltage difference across the capacitance, compared to the case where the structure is fully differential. The net result is a slightly lower SRF for the case where one of the two windings is excited in a single-ended fashion. In figure 3.16, we can see the different distribution of voltages.

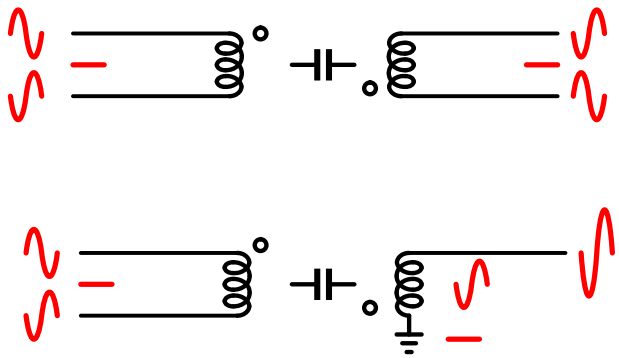


Fig. 3.21 Different voltage distribution between fully differential and balun operation.

3.6 Balanced to Unbalanced (Balun) Conversion at mm wave frequencies

In the previous chapters we noted the necessity for differential to single ended conversion in order to employ the benefits of differential operation and at the same time provide the so much needed single ended interface for measurements or antenna drive. Here we present the baluns used most at high frequencies, namely the transformer balun and the Marchand balun.

3.6.1 Marchand Balun

The Marchand balun [10], [11], [12] is a transmission line based structure, which is sketched in terms of the lines electrical lengths in figure 3.22. It consists of a λ/2 line and two λ/4 lines. The operation of the Marchand balun is based on coupled lines couplers with a reflective (short) load on the isolated port to provide a 180 degrees of phase shift and add the signals constructively. The signal flow paths from the single ended input to the differential outputs are shown in figure 3.18 and the associated phase shifts are annotated in the same figure. We can distinguish four signal paths, which are in phase pair-wise.

The impedance transformation ratio and the balance between ports are given by equations (3.31) and (3.32) respectively.

$$ITR = \frac{Z_{in+}}{R_L} = \frac{Z_{in-}}{R_L} = \frac{1}{{R_L}^2}\left(\frac{Z_{oo}Z_{oe}}{Z_{oe}-Z_{oo}}\right)^2 \tag{3.31}$$

$$Imbalance = \frac{S_{21}}{S_{31}} = -\frac{Z_{oe}-Z_{oo}}{Z_{oe}+Z_{oo}} \tag{3.32}$$

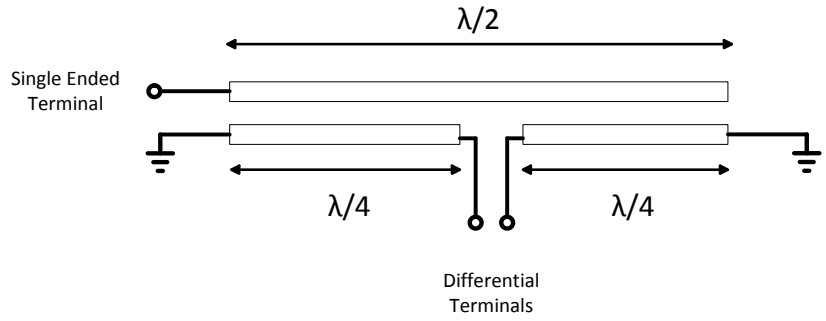


Fig. 3.22 Marchand Balun. Right terminal of λ/2 line can be open or shorted.

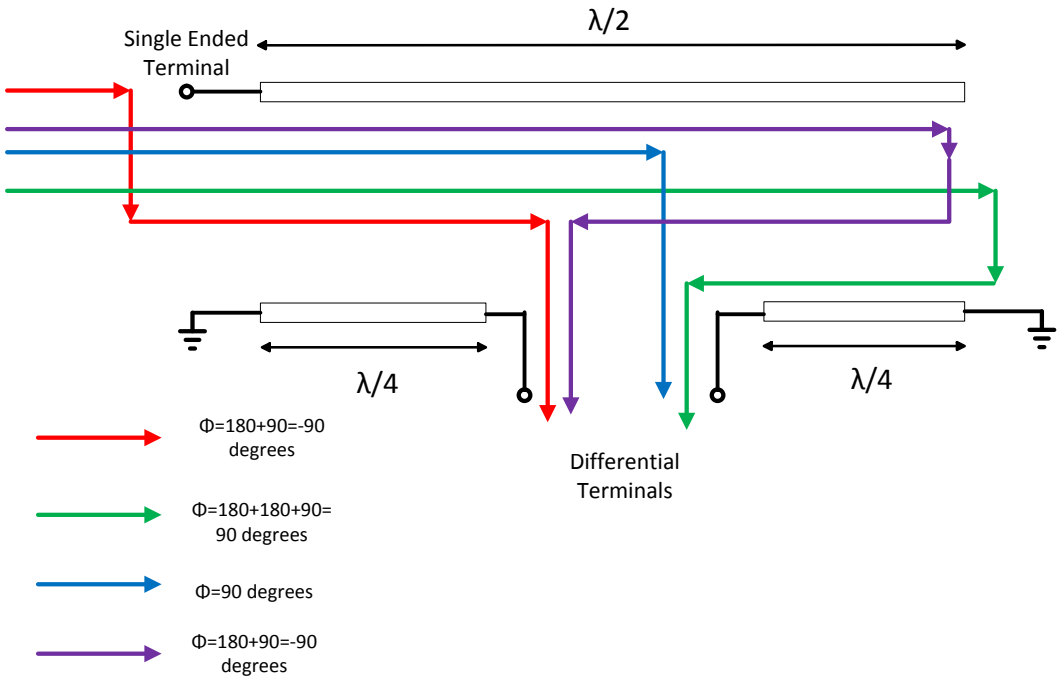


Fig. 3.23 Phase shift along different signal paths of the Marchand balun.

In order to get sufficient balance from the structure, a high ratio between even and odd mode impedance is necessary, namely the even mode impedance should be high, while the odd mode impedance should be low. The odd mode impedance is set by the capacitance between the adjacent lines. Ideally, the Imbalance should have a value of -1. From coupled line theory [13], we know that in order to minimize Z_{oo}, we have to achieve a strong coupling between the lines. This is limited by the minimum spacing of adjacent conductors in the case of edge-coupled lines and by dielectric thickness in the case of broadside coupled lines. Furthermore, to increase the even mode impedance, many sections have to be cascaded. This leads to really bulky structures. Even a single section Marchand balun is bulky. At 100GHz, λ=1.3mm and hence λ/2 is 650um.

A plot of the Marchand balun balance is depicted in figure 3.24. The odd mode impedance is set to 25 Ohms and the even mode impedance stems from multiplying the odd mode impedance by a factor N. The balance of the structure is plotted as a function of the multiplication factor N. The balance gets better for higher even to odd mode impedance ratios. The even mode impedance is defined by the capacitance of the structure to adjacent ground structures. In order to achieve a relatively high even mode impedance, we have to ensure that the adjacent ground structures are far apart from the Marchand balun. This will create a gap in the layout of the structure. The absence of particular metals in large layout areas can lead to DRC density errors which are hard to pass. Dummy metal structures may be used to pass DRC check, but the total structure needs to be re-simulated to ensure proper operation.

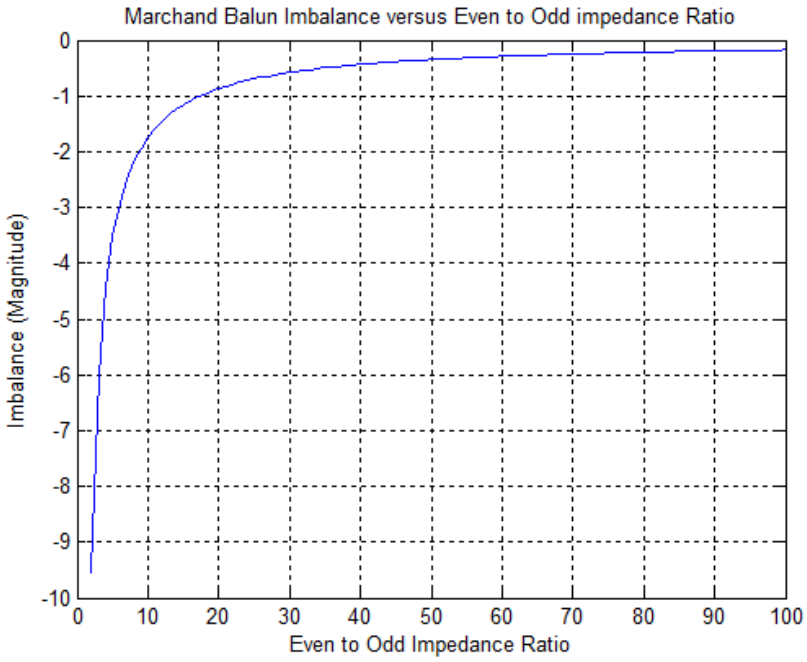


Fig. 3.24 Amplitude Balance of Marchand Balun versus even to odd mode impedance ratio

3.6.2 Transformer Baluns Design Challenges

The transformer has been chosen as the preferable matching network thus far. This choice was based on discussions made in previous sections. It would thus be very compact if we could use it as a balun. The challenges associated with usage of transformers as baluns are addressed in this section. The main mechanism that makes transformer baluns at these frequencies difficult to design is the inter-winding capacitance between the windings which affects the balance, i.e. same amplitude and 180 degrees phase difference at the differential terminals.

The way the inter-winding capacitance manifests as imbalance is given in this paragraph. In figure 3.25 a simple model is used for the transformer including only the inductances and the inter-winding capacitance. Under perfect differential operation, i.e. $Z_3=Z_4$, the voltages of the two secondary terminals, due to magnetic coupling, are the same because the magnetically induced current flows through the same loads. The inter-winding capacitors have the same voltage swing across them and thus the effective capacitances at the input terminals are multiplied by the same factor, i.e. Miller multiplication. Hence the impedances seen by both terminals are the same. In the case of a balun, Z_4 is zero, i.e. the terminal is grounded, and the current flow produces unequal voltages at the two terminals of the secondary windings. Now the capacitances seen by the two terminals of the primary winding are different due to the different voltage swings across them and the different Miller multiplication. Hence the impedances of the two primary terminals are different and so is their frequency behavior, leading to an imbalance at the input. Hence, we can conclude that if there was no inter-winding capacitance; there would be no impedance imbalance.

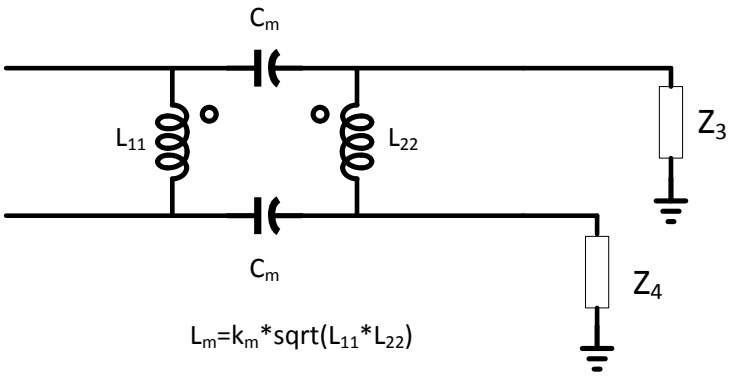


Fig. 3.25 Lumped Element Model of the transformer.

A severe consequence of the imbalance between the terminals is the disturbance of the virtual ground of the circuit. This compromises the superiority of the differential implementation compared to the single ended one, since a portion of the current flows through the ground plane again and the parasitics begin to play a significant role on the circuit performance. A model that shows how the virtual ground is disturbed and also indicates a way for imbalance improvement is shown in figure 3.26.

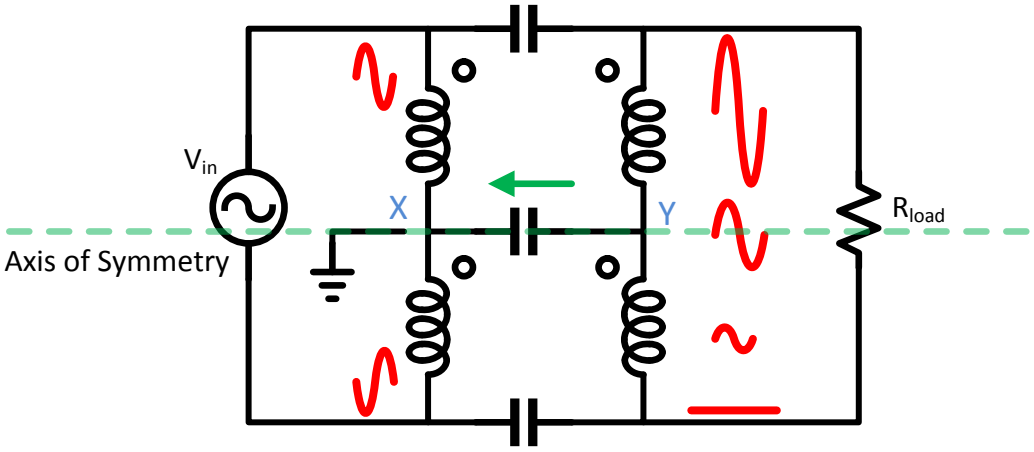


Fig. 3.26 Signal Amplitudes along the transformer balun.

The output voltage swing varies along the secondary winding, starting from twice as large as the input signal at each differential terminal and reaching zero at the other side of the winding. The inter-windings capacitance is distributed along the windings and couples them electrically. At the axis of symmetry of the transformer, point Y of the secondary has a non-zero voltage swing. This voltage swing is coupled to the primary winding at the point X through the inter-winding capacitance. If the impedance connected from X to ground is not zero, a current flows through that impedance and the non-zero voltage swing is sustained. Hence point X is no longer a virtual ground. If the impedance from X to ground is zero, any current imbalance at the primary is equalized with the excess current flowing through the center tap connection. In this case however, no voltage drop exists and point X is still at zero voltage.

In practice there will always be some impedance at that point due to the center-tap connection to feed the DC voltages for the active devices. This connection will introduce some inductance and some

resistance. By adding some capacitance to the center tap, we can series resonate the inductance and hence reduce the voltage swing across X. We can verify this by simulating two transformers in our ST BiCMOS process. Initially, we put no center tap connection at the transformer axis of symmetry and use an ideal ground. In the second setup, we draw a small metal line as the center tap connection and see the effect on balance. Finally, by using an appropriate capacitance we tune out the inductance and we manage to get some of the balance back. Figures 3.27 to 3.29 depict the EM simulation results.

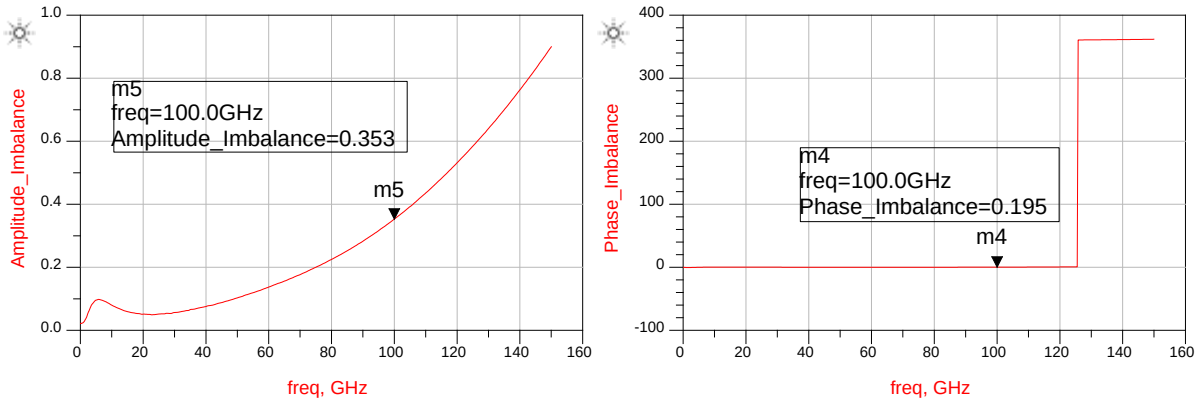


Figure 3.27 Balance of transformer balun with an ideal ground connected at the center tap.

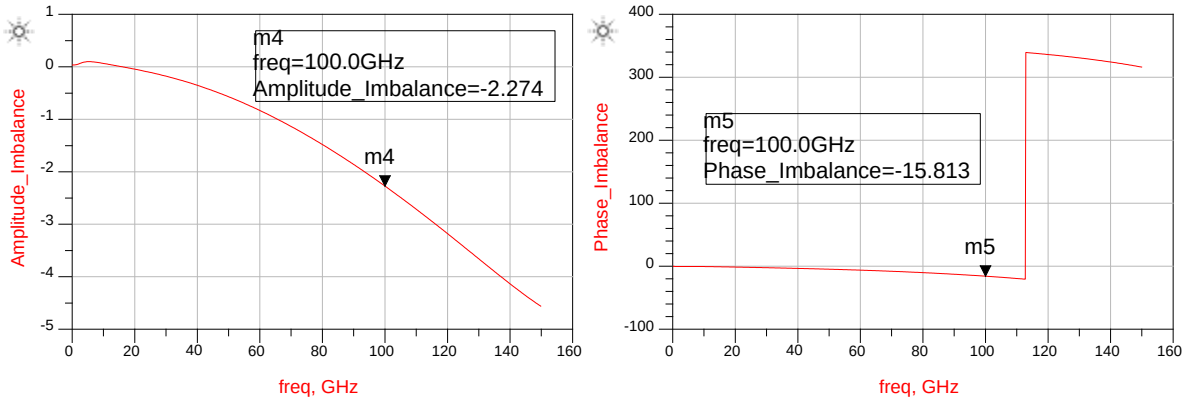


Figure 328 Balance of transformer balun with an ideal ground connected at the center tap.

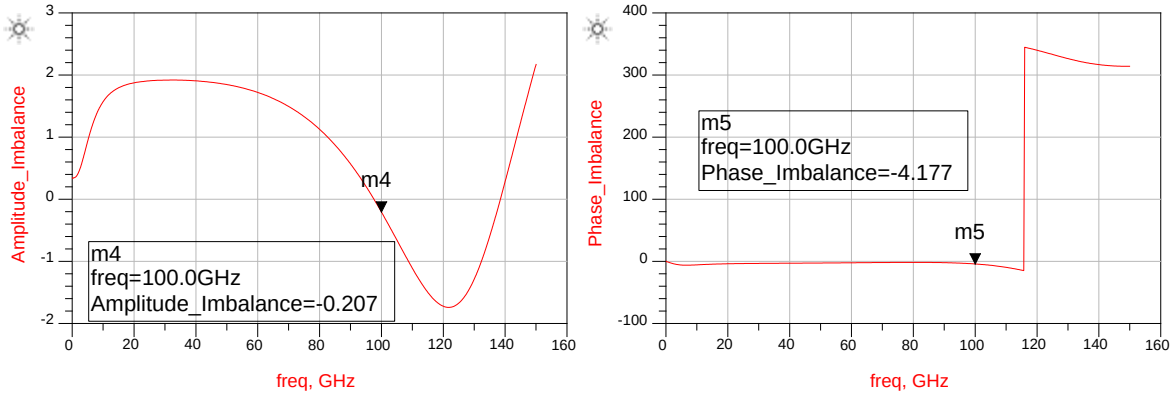


Figure 3.29 Balance of transformer balun with an ideal ground connected at the center tap.

The imbalance of the structure is greatly affected by the impedance of the center tap. An added capacitance at the center tap can alleviate the problem up to some point. However, the solution is not very broadband and applies to frequencies around the resonance frequency. Furthermore, as the inductance of the center tap increases, the required capacitance value becomes smaller and smaller especially at high frequencies. Hence, the impedance of the center tap connection should be kept as low as possible in order for a realistic capacitance to be of use. This technique is used in Chapter 5 for the design of a magnetically coupled two stage PA at 130GHz in an IHP 130nm BiCMOS process.

3.7 Transformer and Marchand Balun Comparison

In order to verify the accuracy of simulations, a test chip was constructed which is shown in figure 3.30. A Marchand balun which will be measured back to back is shown at the top side of the figure. At the bottom side we see two transformers connected to the pads with some transmission lines to be de embedded. The transformer center taps are directly connected to ground and no capacitance is used to tune out the inductance. The process used is the ST Microelectronics BiCMOS 55nm. The ground planes are not visible due to the zoom required to see the whole chip. By measuring the S parameters for the 4 ports of each transformer and by using mixed-mode S parameters, we can calculate the differential and common mode behavior of the structures.

In order to compare the balance of the structures, we will assume that at the center tap of the transformers the appropriate capacitors are added. These were found from simulations. The comparison between the structures is given in Table 3.9.

From the comparison, we see that the transformer is favorable in terms of insertion loss. This is expected since the Marchand balun is very bulky compared to the transformer and this takes a high toll on loss since at high frequencies transmission lines are very lossy. Also after the addition of the appropriate capacitance, the imbalance of the transformer is very good. Furthermore, the inductive nature of the transformer winding gives a readily available inductance that can be used to tune out capacitive parasitics of the active devices. Finally, the ease of biasing and compactness associated with the use of transformers makes them even more attractive for use at high frequencies.

However, the use of transformers is not culprit-free. It turns out that the balance is very sensitive to the value of capacitance used in the center tap. Furthermore, the inductance of the center tap should be sufficiently small in order for the capacitance value to be acceptably high. This causes the center tap connection to be physically close to the transformer, leading to increased coupling between the transformer and the center tap connection. Since we use the center tap to feed the bias to the active devices, the AC coupling may disturb the operating point of these devices.

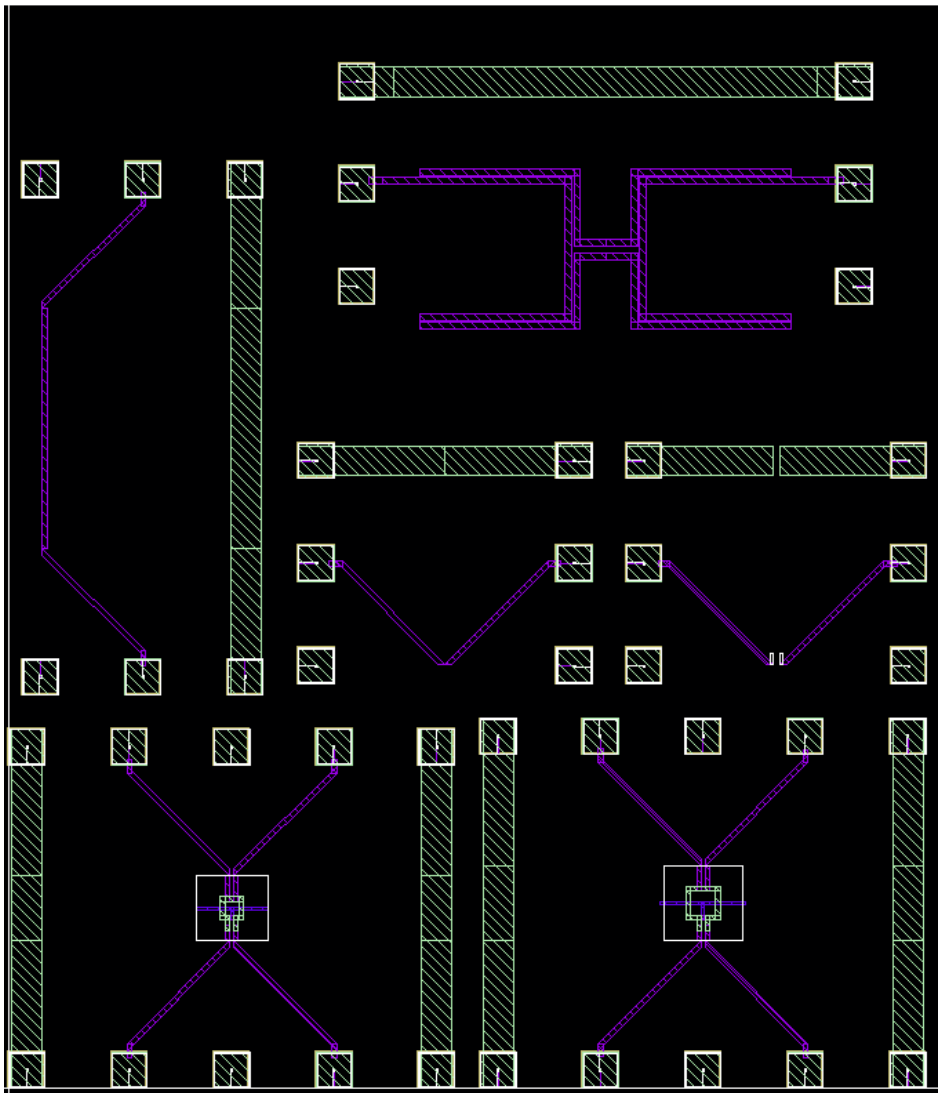


Figure 3.30 Chip consisting of transformer and Marchand balun structures

Table 3.9 Transformers and Marchand Balun Comparison

	Transformers	Marchand Balun
Insertion Loss (dB)	0.5	1.679
Amplitude Error (dB)	<0.3	0.075
Phase Error (degrees)	1	3.3
Area (um^2)	120x120	350x240

3.8. Conclusions

After the analysis of power loss performed in this section and after facing practical issues regarding the implementation of transformers at high frequencies, it is obvious that the transformer efficiency cannot reach the optimum value. However, the benefits of using transformers are so many that we still choose it as intermediate matching networks for our subsequent design. The balance of the transformer balun is hampered by the inter-winding capacitance. However, with the addition of some extra capacitance at the center tap to tune out the inductance associated with the center tap connection, the balance can be improved. Hence, the transformer balun is the one of choice for our PA.

3.9. References

[1]J. R. Long, “SiGe Radio Frequency ICs for Low-Power Portable Communication”, Proceedings of the IEEE, vol.93, Sep. 2005.

[2]Ali M. Niknejad, “Electromagnetics for High-Speed Analog and Digital Communication Circuits”, Cambridge University Press, 2007.

[3]Thomas H. Lee, “The Design of CMOS Radio-Frequency Integrated Circuits”, Cambridge University Press, 2004.

[4]N. Sarmah, P. Chevalier and Ulrich R. Pfeiffer, “160-GHz Power Amplifier Design in Advanced SiGe HBT Technologies with P_{SAT} in excess of 10 dBm”, IEEE Transactions on Microwave Theory and Techniques, vol. 61, pp. 939 – 947, Feb. 2013.

[5] Sorin Voinigescu, “High Frequency Integrated Circuits”, Cambridge University Press, 2013.

[6]Ichiro Aoki, ”Distributed Active Transformer-a new power-combining and impedance-transformation technique”, vol.50, pp. 316-331, Jan. 2002.

[7] T. O. Dickson, M. LaCroix, S. Boret, D. Gloria, R. Beerkens and S. P. Voinigescu, “30-100-GHz inductors and transformers for millimeter-wave (Bi) CMOS integrated circuits”, IEEE Transactions on Microwave Theory and Techniques, vol. 53, no. 1, pp. 123-133, Jan. 2005.

[8]J. R. Long, “Monolithic Transformers for Silicon RFIC design”, IEEE Journal of Solid State Circuits, vol. 35, no. 9, pp. 1368-1382, Sep. 2000.

[9]B. Leite, “Design and Modeling of mm-wave integrated transformers in CMOS and BiCMOS technologies”, PhD Dissertation, University of Bordeaux, November 2011.

[10] G. I. Zysman and A. K. Johnson, “Coupled transmission line networks in an inhomogeneous dielectric medium,” IEEE Transactions on Microwave Theory and Techniques, vol. 17, no. 10, pp. 753–759, Oct. 1969.

[11] Z. Y. Zhang, Y. X. Guo, L. C. Ong, and M. Y. W. Chia, “A new planar Marchand balun,” in Proc. of IEEE-MTT-S, Jun. 2005.

[12] K. S. Ang, Y. C. Leong, and C. H. Lee, “Multi-section impedance-transforming coupled-line baluns, ”IEEE Transactions on Microwave Theory and Techniques, vol. 51, no. 2, pp. 536–541, Feb. 2003.

[13] D. M. Pozar, “Microwave Engineering, 4th edition”, John Wiley and Sons Inc., 2011.

Chapter 4. Active Device Operation at sub-mm wave frequencies

In the previous chapter the design challenges of passive components to be employed at mm-wave frequencies were discussed. Moreover, a flow for the optimization of integrated transformers/balun was indicated. In this chapter, the design choices associated with the active devices are analyzed and guidelines regarding topology selection are given. Focus is placed on biasing and device sizing for a given output power and acceptable performance in terms of power gain and stability. in .

The minimization of layout parasitics of the active devices is deemed as critical and it is addressed by making use of a 2.5D electro-magnetic (EM) simulator, i.e., Keysight MoMemntu. Parasitics have a big impact on both large and small signal characteristics.

As a result of the device analysis and layout optimization, three main transistor configurations (CB, CE and cascode) are analyzed to identify the topology that offers, in the context of mm-wave operation (i.e., operating frequency higher than one third of the transistor f_t/f_{max}), the best compromise between output power, power gain and stable operation.

The summary of this chapter is as follows. First the aspects and implication of achieving reliable operation from a thermal failure and a breakdown-voltage limit standpoint are discussed. Afterwards, guidelines on biasing and device sizing are given, based on output power and speed requirements. Finally, layout techniques that aim towards the minimization of hazardous parasitics, in terms of output power loss, power gain loss and stability reduction, are proposed.

4.1. Voltage limits during bipolar device operation

For power applications, the first issue to be addressed is that of reliable operation for the SiGe BiCMOS bipolar transistors to be used. The maximum voltage swing that does not lead to immetiate failure or long-term quality degradation of the devices is dictated by avalanche breakdown and thermal considerations. The various topologies are compared in terms of their behavior as far as the above mechanisms are concerned [1].

4.1.1. Review of avalanche breakdown phenomena

Avalanche breakdown is a result ofthe impact ionization generated carriers caused by the strong fields that occur across the reverse-biased base collector junction. These carriers will increase the collector current as well as generate a negative base current, i.e., flowing out of the base. When a high impedance path, for the bias line, is provided at the base the carriers re-enter the base of the transistor, leading to an increase in collector current. Moreover, the carriers generated in the inversely biased base-collector junction will again contribute to impact generation process making the process of increased current abrupt (i.e., avalanche). This effect is briefly depicted in in figure 4.1

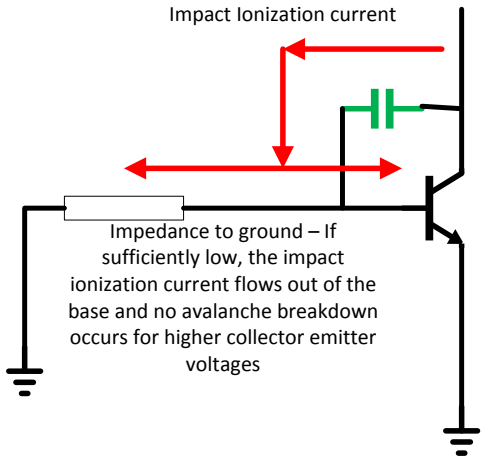


Figure 4.1 Avalanche breakdown mechanisms

To quantify these phenomena, some metrics have been devised in order to characterize various bipolar processes [1]. The first one is BV_{CEO}, which assumes that infinite impedance exists from base to ground. Hence, all impact ionization current flows into the base of the transistor, causing avalanche breakdown for lower collector to emitter voltages. The second one is BV_{CBO}, which assumes that a perfect short exists from base to ground, while the emitter of the device is open. BV_{CBO} is several volts higher than BV_{CEO} [3]. Due to the low impedance path to ground provided to the impact ionization carriers, avalanche breakdown is further delayed until higher collector-emitter voltages are obtained. Although the conditions under which these quantities are calculated are not realistic (open/short of base or emitter), we reach an important conclusion. A low impedance path should be provided from the base of the device to ground in order for avalanche breakdown to happen for higher collector – emitter voltages.

4.1.2. Review of Thermal Stability Issues

The mechanism responsible for the thermal stability of bipolar devices is the negative temperature coefficient of V_{BE}. When the temperature is increased, the value of V_{BE} that gives a certain collector current is reduced. Thus, by keeping V_{BE} constant and elevating the temperature, the collector current keeps increasing. This in turn further increases the temperature and a positive feedback mechanism is created which eventually destroys the device [4]. Even if the device is not destroyed, hot spots will be created and the current will be conducted only by a small area of the device (current hogging). On the other hand, the current gain factor β has a very weak or negative thermal coefficient [5]. Hence, a current driven bipolar device is stable with temperature, while a voltage driven is generally not. In [6] emitter and base ballasting schemes have been proposed in order to alleviate the effect of thermal instability.

4.1.3. Design for Extended Breakdown Voltage and Thermal Stability

As can be easily derived from the small review regarding breakdown and thermal effects ,a design conflict exists between optimizing the device behavior for avalanche breakdown or thermal behavior:

- To extend the breakdown voltage (i.e., operation close to BVCBO), which is required in very low breakdown devices to achieve sufficient voltage swing, thus output power, we need to provide a low base impedance in order for the impact ionization current to be shorted to ground.
- To provide a thermally stable bias the employment of a resistance at the base is preferable. This resistance poses a ballasting role which compensates for the increased current associated with the termerature increase.

In order to satisfy both conditions differential topologies can be employed. In this case, ballasting may be employed which will regulate the temperature behavior of the device and any ballasting resistors will not be visible during the AC operation of the circuit because of the virtual ground at the axis of symmetry of the differential circuit. Furthermore, the virtual ground can be used in order for the DC signals to be fed.

4.1.4 Voltage limits simulations

The I_C-V_{CE} curves of various topologies are compared here. Modern HICUM and VBIC models incorporate both avalanche breakdown and electro-thermal (ET) effects due to self-heating. For HICUM models, the latter effects can be turned on and off by setting the *flsh* parameter to 1 and 0 respectively. In this way, we can distinguish between the behaviors caused by avalanche breakdown and ET effects. The process used for the simulations in this section is an IFX SiGe BiCMOS 130nm from Infineon technologies. The models correspond to a fixed size device and a collector current of 15mA gives the maximum f_{MAX}.

Voltage Driven CE configuration

First, a voltage driven Common Emitter (CE) configuration is examined for the extreme case of zero source impedance as shown in figure 4.2. Initially, ET effects are disabled and the results are presented in figure 4.3. For a current of ~15mA, the maximum V_{CE} before the onset of avalanche breakdown is approximately 3V. This voltage corresponds to the BV_{CBO} limit.

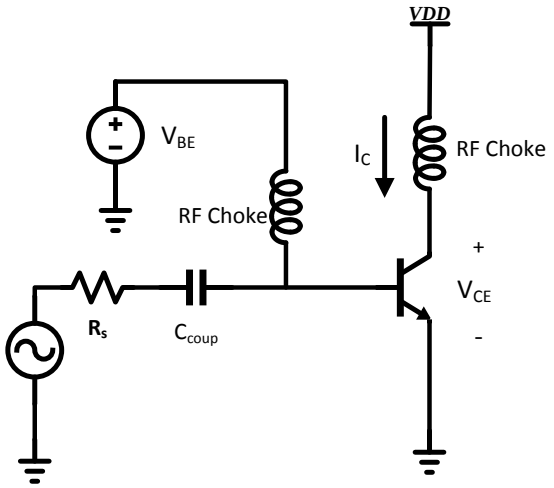


Figure 4.2 Common Emitter Configuration with ideal voltage drive

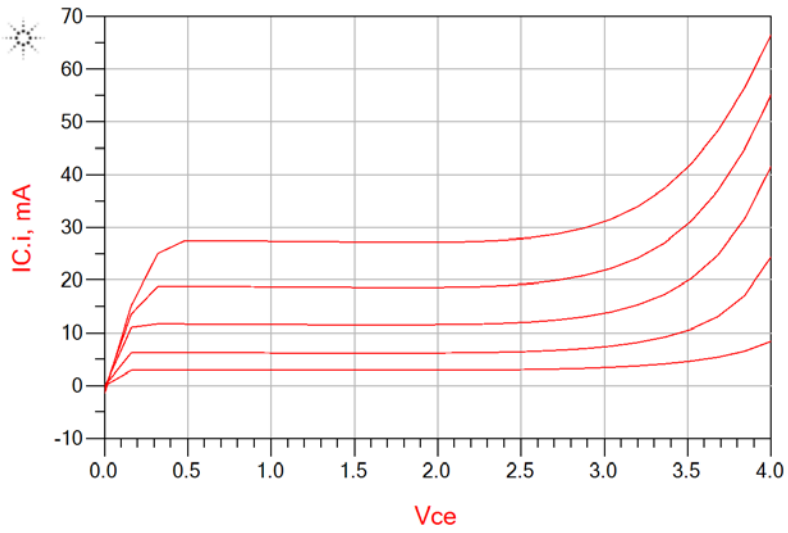


Figure 4.3 Common Emitter with ideal voltage drive DC curves with ET effects disabled

The curves including ET effects are given in figure 4.4. The thermal instability discussed in the previous section can now be seen. The thermal instability causes a strong variation in the collector current even for low values of the collector current. This makes the CE configuration with ideal voltage drive unsuitable for large signal swing operation. Emitter and/or base ballasting can be used to regulate the collector current. In the same figure, the behavior of a CE stage with 500 Ω base ballasting is shown. The variation related to the thermal effects is less severe at the cost of a reduced breakdown voltage due to the high base impedance. However, as we will see, by choosing a different topology, their use can be avoided.

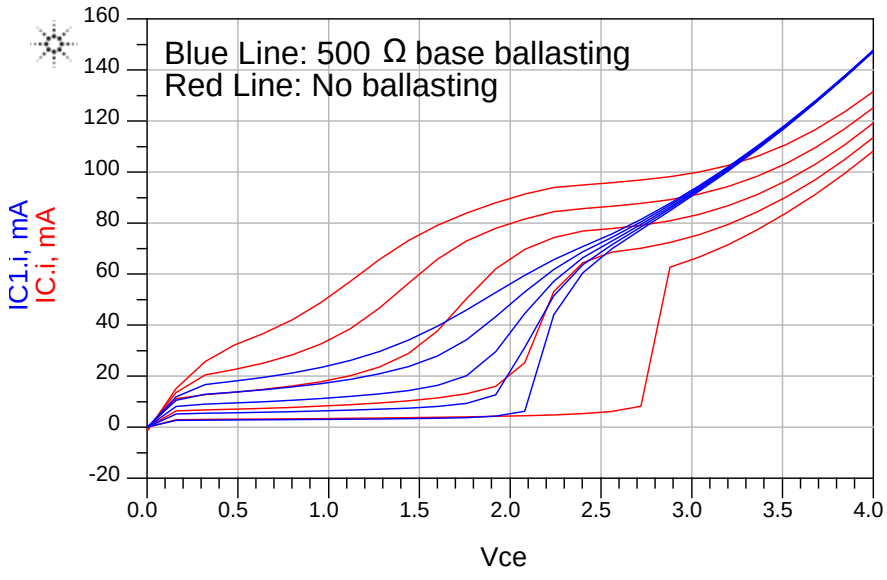


Figure 4.4 Common Emitter with voltage drive DC curves with ET effects enabled. The red curce corresponds to ideal voltage drive, while the blue curve emerges after the addition of 500 Ω base ballasting.

Current Driven CE configuration

An ideal current source corresponds to the other extreme case of a CE stage, the one with infinite base impedance. This situation is depicted on figure 4.5. Now, the onset of avalanche breakdown phenomena happens for a lower VCE. However, we expect more thermally stable behavior. These statements are verified by figures 4.6 and 4.7 in which ET effects are disabled and enabled respectively.

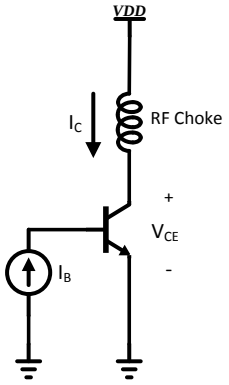


Figure 4.5 Common Emitter Configuration with ideal current drive

Due to the high base impedance, the voltage at which avalanche breakdown begins is now approximately at 2V. Hence the output voltage swing and thus power is reduced significantly. However, when ET effects are taken into consideration, the collector current variation is less significant (see Fig. 4.7), due to the weak temperature coefficient of transistor beta. However, we notice a bending of the IC curves, which corresponds to lower output impedance. This implies a reduction in the power gain for higher collector currents and collector – emitter voltages.

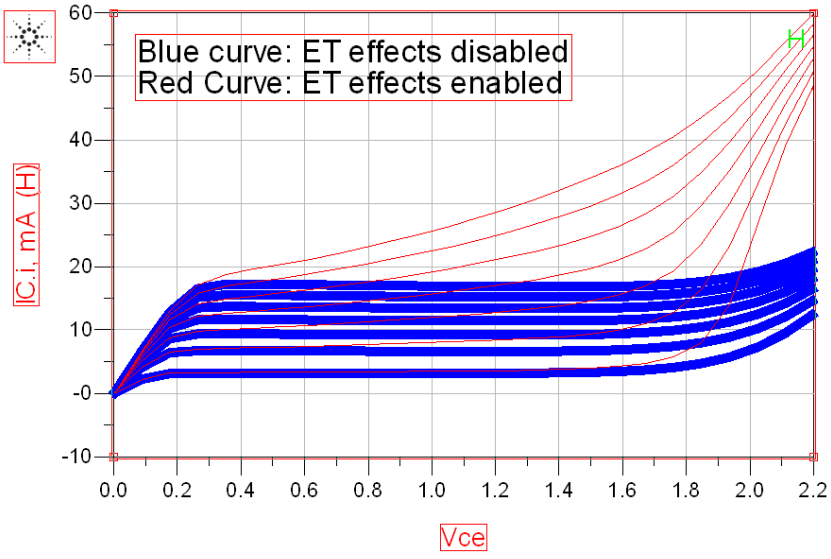


Figure 4.6 Common Emitter with ideal current drive DC curves with ET effects enabled (red curve) and ET effects disabled (blue curve).

Voltage Driven CB configuration

The common base configuration is presented below. The DC operation is identical to the behavior of the voltage driven CE stage of section 4.1.4.1. Throughout the literature, the common base stage is said to be preferred because the avalanche breakdown voltage is now BV_{CBO}, which is higher than BV_{CEO}, thus enabling larger voltage swings. However, the constant V_{BE} character of the biasing triggers thermal effects that prohibit the use of the particular topology without special measures of the resistive ballasting nature.

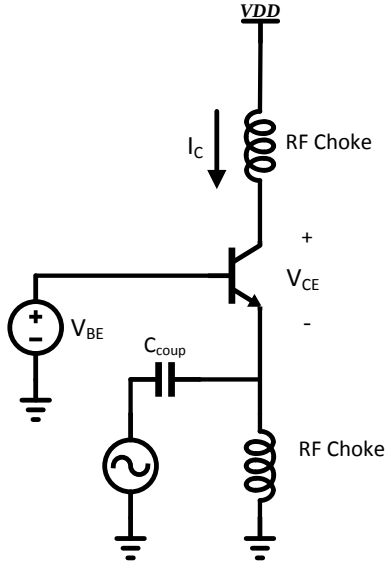


Figure 4.7 Common Base configuration

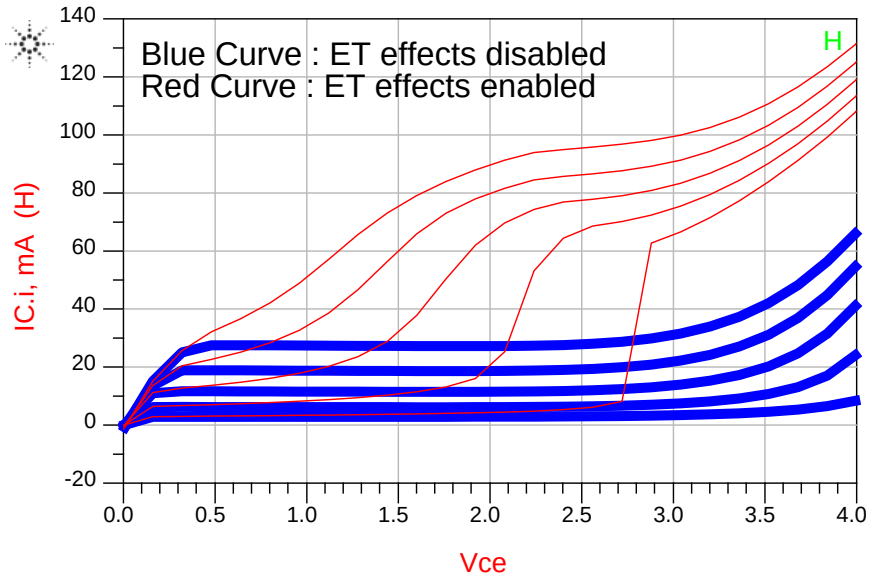


Figure 4.8. Voltage Biased CB I-V curves without ET effects (blue curve) and including ET effects (red curve).

Emitter Current Biased CB/CE configuration

A thermally stable topology is one with a DC current source at the emitter. This topology is given in figure 4.9. Although, signal-wise, we can drive the device from both the base and the emitter, in the picture we present a CB configuration (i.e., emitter driven). From figure 4.10 we note that the avalanche breakdown voltage isextended due to the low impedance provided on the base node.. Furthermore, the nature of the biasing, i.e. the forced bias current, does not allow any significant variation due to temperature and self-heating effects. This is why the behaviors with enabled and disabled ETs are practically identical. Nevertheless, this behavior requires a high quality current source, i.e., providing an high output impedance [6].

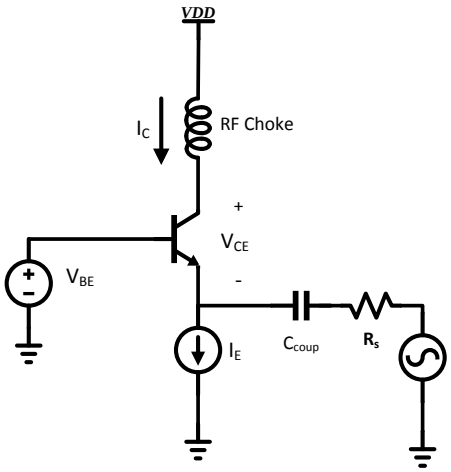


Figure 4.9 Current Driven Common Base configuration

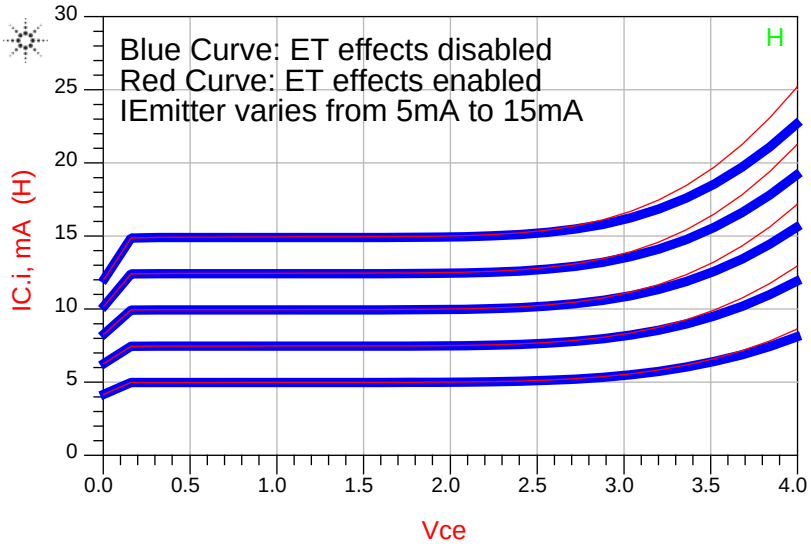


Figure 4.10 Common Emitter with ideal current drive DC curves with ET effects disabled (blue curve) as well as ET effects enabled (red curve).

Cascode configuration

The previous analysis indicated that a common base configuration with a current and signal drive at the emitter would provide a thermally stable situation. When we consider the classical cascode configuration, we can see that the CE stage does provide to the CB stage indeed a current drive as well as a signal from the emitter node, see figure 4.11. Moreover, the extra CE stage provides increased gain and results in an higher overall output impedance which are both extremely useful at mm-wave frequencies where the gain per stage is limited and the output impedance is strongly reduced by the layout parasitics.

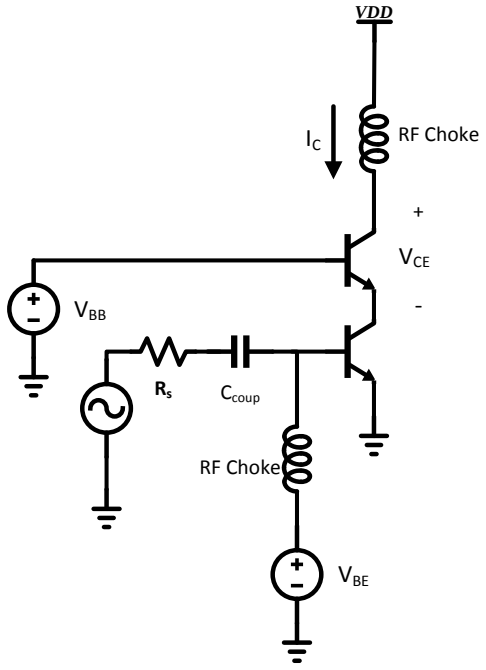


Figure 4.11 Cascode Configuration

In figure 4.11, V_{BB} should be chosen sufficiently high so that the CE device remains in the forward active region for linear operation by keeping the CB junction under reverse bias. The required non-zero voltage at the collector of the CE device limits the voltage swing at the collector of the CB device by dictating a lower voltage bound above which both devices are operating in the forward active region. In order to maintain the voltage swing and hence the output power comparable to the ones of the CE stage, we have to increase the supply voltage. This leads to increased power dissipation. In order for the supply voltage increase to be as low as possible, we can allow only for a small V_{CE} voltage for the CE stage, by choosing V_{BB} (see figure 4.11) as low as possible. However, under this regime of operation, the device may enter saturation in some part of its operation, which degrades linearity. Finally, the cascode topology requires one extra supply to bias the CB transistor. In figures 2.13 the output characteristics of the cascode topology are presented. The red curves correspond to the behavior including ET effects, while the blue curves correspond to the behavior not including them. Now we have two devices and a higher V_{CE} voltage is required in order for both transistors to be in the forward active region. Hence, for lower VCE values, the I-V curves should be ignored.

For the currents of interest, i.e. close to 15mA, the ET effects are not changing the picture much compared to the case where no ET effects are included. For currents above 20mA, the ET effects start changing the behavior quite significantly. Furthermore, we can safely operate the output device up to approximately V_{CE} = 3V without problems. Note also that the onset of avalanche breakdown is delayed due to the grounded base of the common base device.

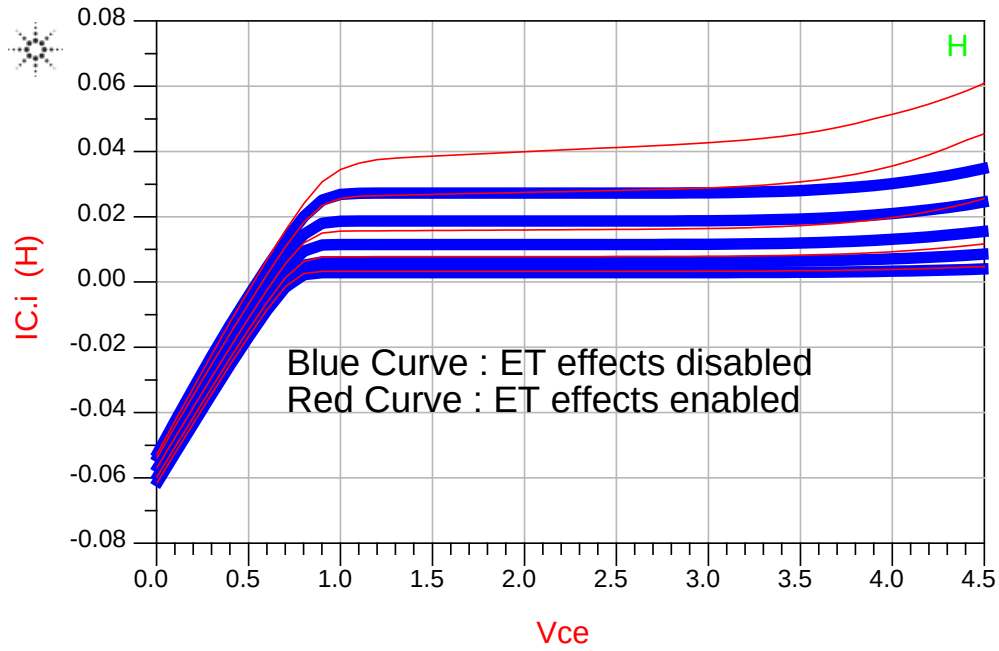


Figure 4.12 Cascode Configuration DC curves

Selection of Topology based on voltage limits

From the above discussion, a current forced at the emitter leads to the most thermally stable behavior. The cascode configuration can emulate this condition with the CE device acting as a current source. This fact, in conjunction with the increased power gain that this topology offers compared to the CE and CB, makes the cascode topology favorable for gain stages at mm-wave frequencies or in general at frequencies that are a fraction of f_T/f_{max}. The downside of the cascode configuration is the reduced output swing due to the voltage that the CE device consumes. Nevertheless, if power consumption and thus efficiency are not the main requirements (as is the case in gain limited stages), we can increase the power supply and restore the output swing.

4.2. Bias Current Selection

The bias current for the amplifying stage is selected in order to maximize the speed of the transistor, which in turns maximizes the available gain. To illustrate this point, it is useful to analyze the graphs presenting f_{MAX} as a function of the base-emitter voltage V_{BE}. This latter quantity can be seen as equivalent to the current density of the device. The graph below is based on the IHP 130nm SiGe

BiCMOS process. The peak f_{MAX} is found to be 350GHz for a V_{BE} equal to 0.91V. For a single cell device with an area of 0.13x1.10um^2, this corresponds to a current of, approximately, 2mA.

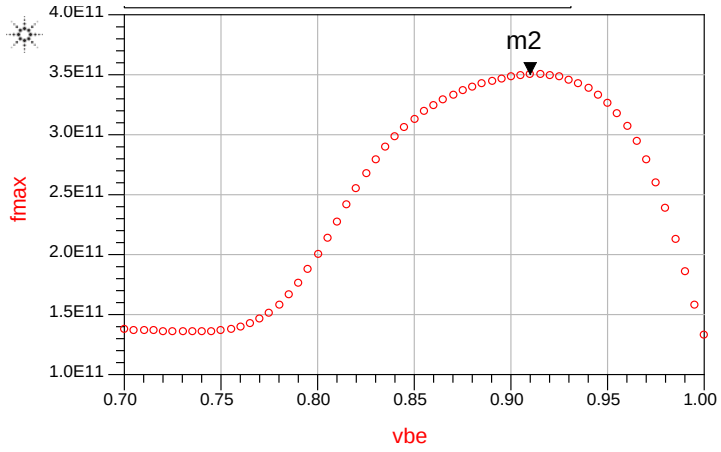


Figure 4.13 f_{MAX} as a function of Base – Emitter voltage for an IHP SiGe 130nm bipolar device

In general we do not want to operate the device at peak f_{MAX} for the following reasons:

- backing up slightly in terms of current, we can move from a pure class A operation to a class AB mode, thus providing a more efficient operation without real power gain penalties.
- the signal swing makes the operating point move along the curve of figure 4.12 . This situation is depicted in figure 4.13 where the green line depicts a possible trajectory of the current value during large signal operation. In order to maximize the speed of the device in large signal operation we would like to remain in the left side of the curve 4.14 (in respect to $f_{max\ peak}$) for the whole signal excursion. If we move to the right part of the graph, the slope is steeper and thus f_{MAX} degrades rapidly leading to slower device operation. In commercial applications, an operating point even at half the peak f_{MAX} can be chosen so that the device operation does not degrade under any condition (temperature, different corners etc.).

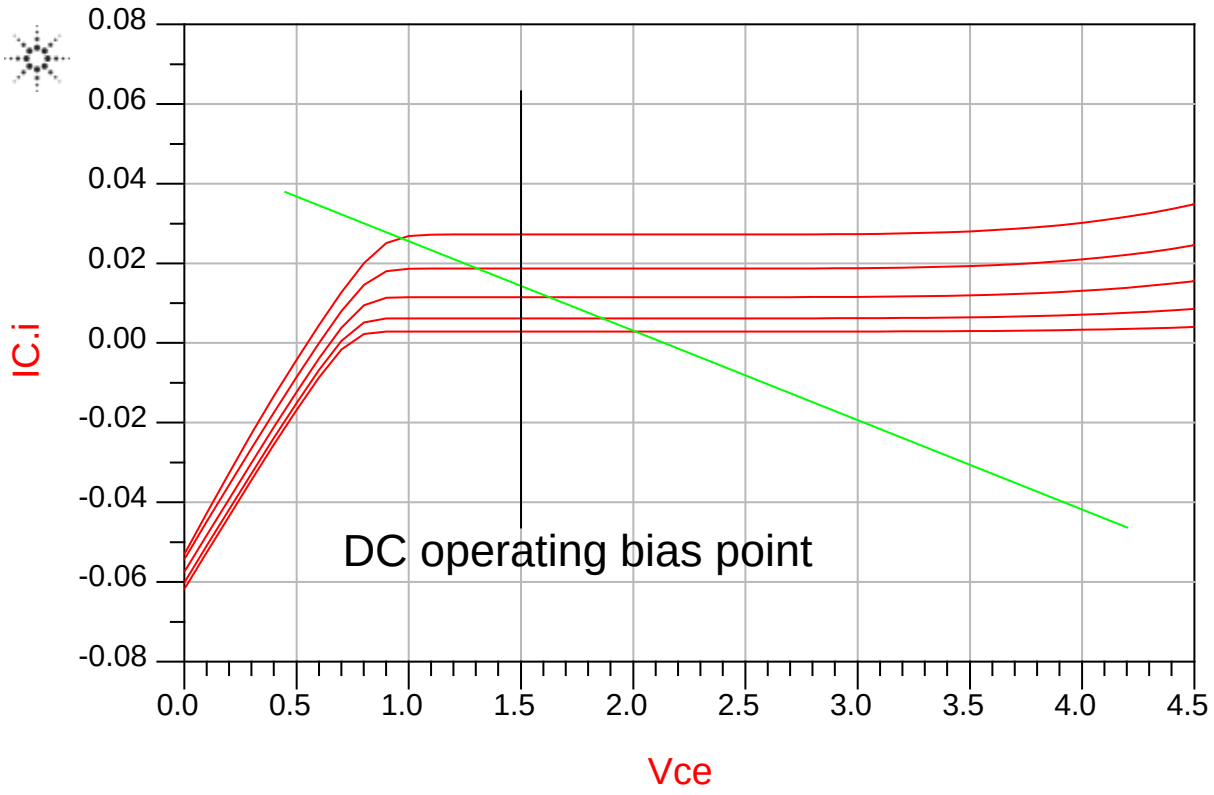


Figure 4.14. Current Variation during amplifier operation.

4.3. Device Sizing

The selection of the device size is based on output power considerations. By placing devices in parallel or by increasing the periphery of the active device (i.e., emitter area) we essentially sum the output currents of the collectors, thus increasing the output power. As we will see in the next chapter, load-pull simulations [10] can guide us in choosing an appropriate size in order to achieve the required power.

A question emerges regarding how far we can go in terms of size increase in order to get more output power [7]. In order to get insight into the answer of this question, we have to examine how power is produced at the output of the active device. For this purpose we can model the transistor as a current source with output impedance consisting of a resistance and a capacitance. The resistance corresponds to the finite slope of the DC curves, while the capacitance consists of the active device parasitics. The model is depicted in figure 4.15.

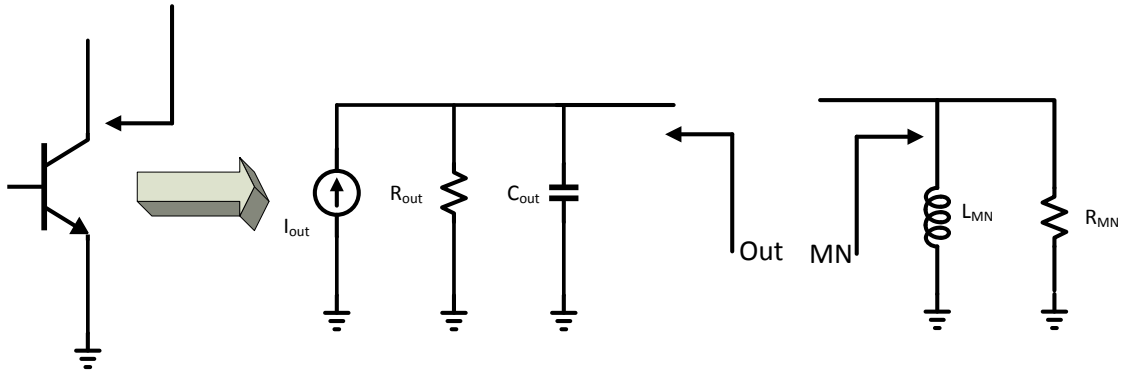


Figure 4.15 Simplified model of power transistor output and matching network impedance.

For typical class – A operation, the goal is to tune out the transistor capacitance by designing the output matching network in order to provide the appropriate inductance at the output terminal. The matching network should also provide the appropriate real part so that the equivalent resistance is the one needed to provide the required voltage and current swing in order to get the specified output power. The equivalent resistance is the parallel combination $R_{out}//R_{MN}$.

For the inductance provided by the matching network, the following relationship has to hold:

$$f_{op} = \frac{1}{2\pi\sqrt{L_{MN}C_{out}}} \tag{4.1}$$

where f_{op} is the operating frequency. When we keep increasing the device size to get more power, the capacitance also increases. At very high frequencies (i.e., mm-wave) scaling up the device size imposes that the required inductance value in order for (4.1) to hold is too small to be realized or provided by any matching network. This makes the design of high power amplifiers very difficult at high frequencies.

Even if there was a solution for the realization of small inductances, there are issues regarding the real part. Under normal operating conditions, the output impedance of the active device R_{out} is sufficiently large and R_{opt}, which is the parallel combination of Rout and R_{MN} is approximately equal to R_{MN}. For large device sizes, the parasitics cause R_{out} to reduce and may even be lower than the optimal Ropt [7]. Hence, if $R_{opt}>R_{out}$, then there is no way that R_{opt} can be achieved by any matching network since it is bigger than the smallest of the two parallel resistors. Furthermore, a low output impedance from the device causes most of the power to be dissipated in the device and not delivered to the load. In this situation, the maximum power is achieved by conjugate match, which allows half of theavailable power to be delivered to the load. Based on this argument, the cascode topology would be preferable due to its higher output impedance, which is more likely to be sufficiently higher than R_{opt}.

4.4. Cascode Stability Analysis

In this section, we analyze the cascode topology in terms of stability. The first stability problem that may arise is due the capacitive nature of the CE stage output impedance. This situation is depicted in figure 4.16. For high frequencies, the bottom transistor can be in first approximation modeled by a capacitance. In the figure, C_{out1} represents the output capacitance of the CE device as well as the parasitic

capacitance to ground originating from the layout connections. When calculating the input impedance by looking through the base of the transistor, we find the real part is given by equation (4.2), where gm is the transconductance of the device and C_{be} the base-emitter capacitance of the CB device. The term is negative, which means that energy is provided to the circuit that is connected to the left of the base node. Since there is already a negative resistance and a capacitance at the base node, a small inductance completes the list of ingredients for an oscillator to be formed. This inductance is inevitable due to the parasitic of the base inter-connect.

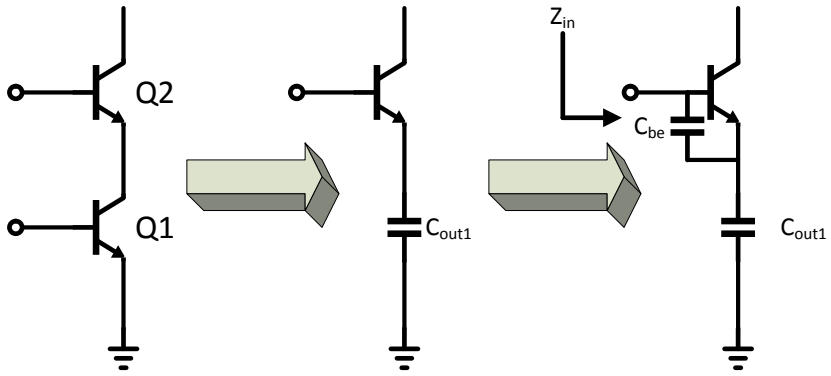


Figure 4.16 Source of instability due to emitter capacitance

$$Re\{Z_{in}\} = \frac{-g_m}{C_{out1}C_{be}\omega^2} \tag{4.2}$$

For an amount of inductance from the base to ground, the real part of the impedance seen looking towards the emitter of the CB device is found to be:

$$Re\{Z_{in}\} = \frac{1-\omega^2 C_{be}L_B}{g_m + j\omega C_{be}} \tag{4.3}$$

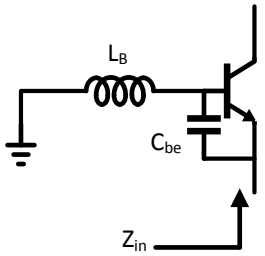


Figure 4.17 Source of instability due to base inductance

From equation (4.3) we see that larger devices tend to present a negative resistance for a smaller base inductance due to the increased C_{be}. Hence, in the case of a cascode device, or a common base stage, when we aim for high output power, we need to pay extra attention to the stability of larger devices.

In order to minimize the effect of the base inductance, we can adopt certain measures. First and foremost, we should provide a relatively large capacitance as close as possible to the base of the transistor. Some resistance should also be put there in order to de-Q the network and alleviate peaking.

This resistance should be high enough to damp any occurring peaking. The parasitic resistance of the capacitor, due to its finite Q, can also help in damping any oscillations. However, its value should be low enough in order not to limit the efficacy of the capacitor.

In this section, some simulations are given that illustrate these instability issues. First and foremost, the influence of the base inductance on stability is illustrated. In figure 4.19, the cascode topology along with some parasitic inductance annotated at the base is given. An inductance of 5pH is sufficient to lead to oscillations. In figure 4.20, the same quantities are depicted, but this time a 10Ω resistor is added in series. We note that the maximum inductance value that can lead to oscillations is increased to 10pH. Nonetheless, the structure is still very sensitive to parasitics inductance and effort should be made for its minimization during the layout, either by reducing the inductance itself, or by inserting resistance in series with the base. The capacitance does not play a significant role there, since it cannot assume large values due to the limited area. The MIM module will allow us to achieve very high capacitance densities, however we do not use it here, due to its extra cost.

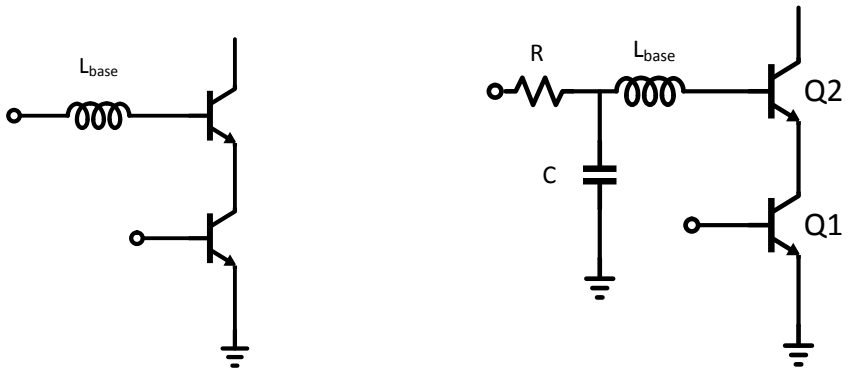


Figure 4.18 Cascode Structure with annotated base inductance (left). Cascode structure with base resistance and capacitance to ground (right).

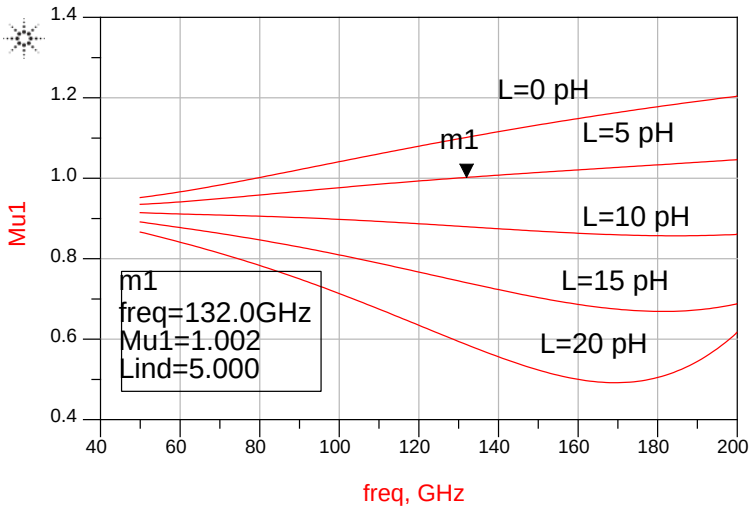


Figure 4.19. Influence of base capacitance on stability of the cascode structure.

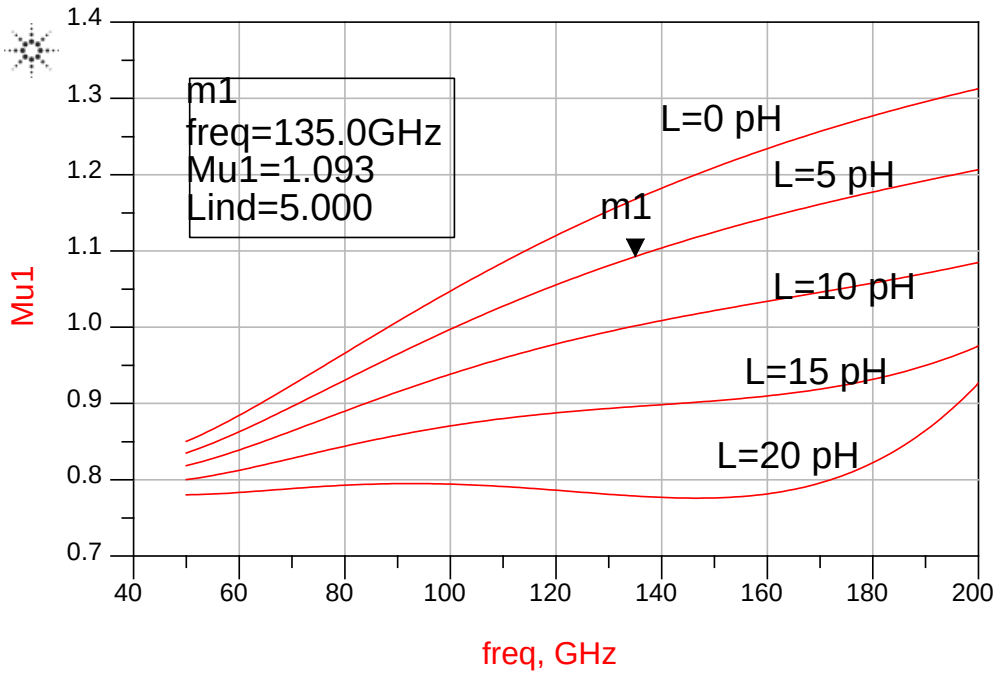


Figure 4.20. Influence of base capacitance on stability of the cascode structure after the addition of the base resistance and capacitance.

In the next section, a layout technique including a conductor sandwiched between the ground planes is described that achieves the above goals. Finally, the use of a differential configuration is preferred, since in that case L_B is just the inductance of the connection from base to the virtual ground. Hence the decoupling requirements are greatly reduced compared to the single ended case, where the whole inductance from the base to the supply pad has to be taken into account.

4.5. Cascode Gain Reduction

After addressing the parasitics affecting stability, we also have to address the parasitics that cause power gain reduction in the cascode device. Figure XX depicts the cascode topology with various parasitics annotated. The effects of the parasitics have to be examined in order to find a layout strategy that ensures optimum performance.

In order to get insight regarding power gain performance, f_T and f_{MAX} are the appropriate metrics. Expressions for these quantities can be found in [11] and are reproduced here in equations (4.4) and (4.5). These equations hold for a single device. In equation (4.4) τ_F is the base-collector transit time, which quantifies the time for a carrier to travel from base to collector. J_c is the collector-current density, A_E is the emitter area and R_c, R_b, R_e are the parasitic resistances of the collector, base and emitter respectively.

$$\frac{1}{2\pi f_T} = \tau_{EC}(I_C) = \tau_F(J_C A_E) + \frac{kT}{qJ_C A_E}(C_{be} + C_{bc}) + (R_e + R_c)C_{bc} \tag{4.4}$$

$$f_{MAX} = \sqrt{\frac{f_T}{8\pi R_b C_{bc}}} \tag{4.5}$$

For moderate current densities, which is the region of operation for peak f_T/f_{MAX}, the second term of the right hand side of eq. 4.4 is small. The last term has a weak dependency on the current density and can be considered constant. Hence from these equations, C_{bc} has the greatest impact on f_T/f_{MAX}. Moreover, as can be seen by eq. 4.5, R_b reduces f_{MAX}. Resistive-inductance parasitics in the emitter introduce negative feedback, thus dramatically reducing the gain. Even 1pH of inductance is shown to reduce the gain by 0.6dB.. Differential operation is advantageous in this respect, since the emitter parasitics that affect the AC performance are the ones from the intrinsic device up to the virtual ground of the structure instea of the whole inductance from the emitter to the pad. From the above analysis and simulations, it is seen that the main parasitics that affect the gain are the ones in series with the base of the emitter and the capacitance C_{bc}. This can be seen from the equations 4.4 and 4.5. The emitter and base parasitics introduce local feedback to the stage which reduces the power gain. The collector parasitics have a minor effect on gain, however in case there are large collector currents, the collector series resistance should be minimized because even small values can sustain a large voltage drop which leads to power dissipation and efficiency reduction.

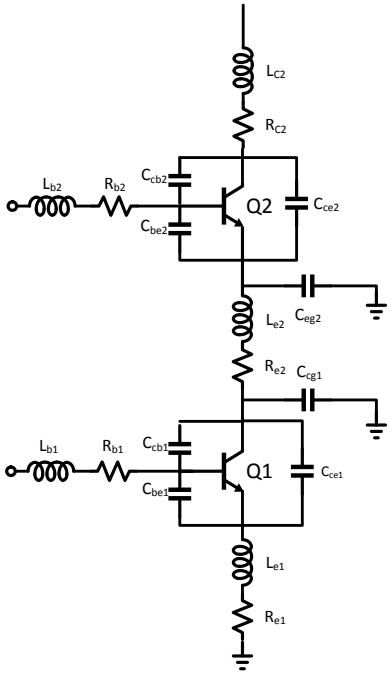


Figure 4.21 Cascode Stage with Annotated Parasitics

Based on the above discussions, we should make an effort during layout to minimize certain parasitics. Based on the discussion of section 4.4, By minimizing the capacitance betweeb the CE and CB stage, as well as the base inductances of the CB stage, we enhance the stability of the structure. Furthermore, from equations 4.4 and 4.5, by minimizing the emitter and base parasitics, especially the former, we avoid

major reduction in power gain. Extra collector-base capacitance should also be kept to a minimum. Collector resistance should be kept to a minimum to avoid output power dissipation.

4.6. Device Layout and Topologies comparison

In this section, layout techniques are proposed to optimize the active device cells, following the parameter optimization depicted in the previous section.

The "golden" objective of the layouting process is to a gain reduction, compared to the pure schamtic simulation, *as small as possible* , as well allow unconditional stability for a frequency range from close to DC up to transistor's ft.the trans. The minimization of the capacitive parasitic should still respect the minimum track widths for the output cells to prevent electromigration, making the task even more complex.

In order to find the proper layout strategy, we have to initially observe the layout of the devices, as given in the PDK. In the scope of this work, the process is the IHP 130nm SiGe BiCMOS. Metallization consists of five thin metals and two thick ones. The layouts of a bipolar transistor unit cell as well as of a cell consisting of five unit cells are shown in Fig. 4.22. We observe two levels of metallization. The light blue thin lines that appear at the core of the devices denote connections on Metal1. The bottom thin light blue line is the Base (B) connection, while the upper light blue line denotes the collector (C) connection. The thick white line in between denotes a connection on Metal2 and is the Emitter (E) connection. The drawn dimensions of the unit cells are 0.13x1.1 um^2 and 5x0.13x1.1 um^2 respectively. The Momentum 2.5D EM simulator from Keysight Technologies [12] will be used to simulate the planar layout structures. Next, these structures are going to be co-simulated with the active cells in order to get the final performance of the structure.

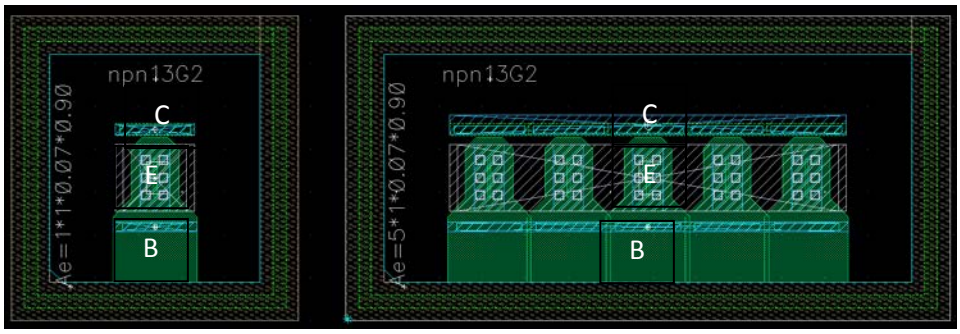


Figure 4.22 Layout of NPN13G2 cells: single cell (left) and5 cells structure (right).

First, a Common Emitter (CE) stage will be layout and simulated. The layout of the differential CE stage is given in figure 4.23. The final cell of the proposed PA will consist of five parallel unit cells. The guard ring is also taken into account to properly model all the cpacitances that ae surrounding the transistor. The guard ring is going to be connected to the ground plane, which is going to run throughout the whole chip. Since this topology requires a grounded emitter, it is reasonable to connect the emitter to the guard ring. The ground plane around the device comprises of multiple metals shunted together in order to minimize parasitics. However, for a differential operation, the ground parasitics will have a minimum effect on differential performance.

The way the cell is laid out can be justified as follow:

- *emitter connection*: all metals are used from metal 1 up to TopMetal1. This provides inductance and resistance minimization. Furthermore, electromigration risks are minimized. For the base.
- *AManhattan* structure is used on Metal1 for the base connection as is shown in Figure 4.23. This structure will give the best compromise between capacitance and inductance minimization. The area of overlap between Metal2 and Metal1 is reduced. Furthermore, the line is getting wider as we move towards the cell, thus minimizing the inductance. For further inductance and capacitance minimization, we use two fingers for the base connection.
- *the collector* is taken out on Metal4 in order to reduce the capacitance to the guard ring/ground plane. By employing a higher metal, the capacitance reduction is not significant.

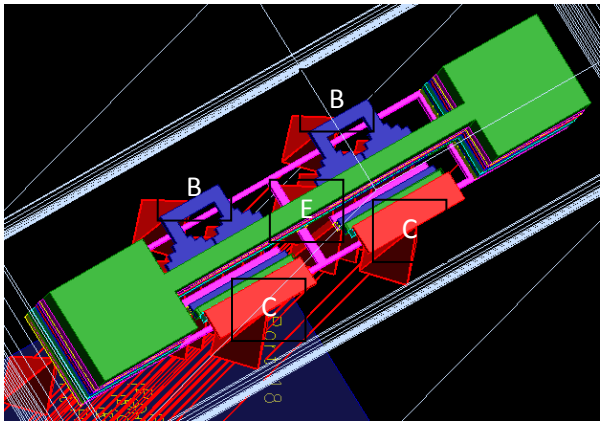


Figure 4.23 Layout of NPN13G2 cells

Next, co-simulations of the EM modeled planar structures and the active devices are going to be performed. Metrics for characterizing the performance of the various active cells are defined in the paragraph. MAG and MSG are the maximum gains when the device is unconditionally and conditionally stable respectively [13]. These quantities are incorporated in the MaxGain quantity in the Keysight ADS environment. As a stability measure, we are going to use Mu. This quantity has the advantage that it can be used alone allow us to draw conclusions about the stability of the system [14]. Finally S12 is simulated, in order for the reverse isolation to be assessed. Higher reverse isolation promotes stability and also eases input and output matching due to the isolation between the input and output ports.

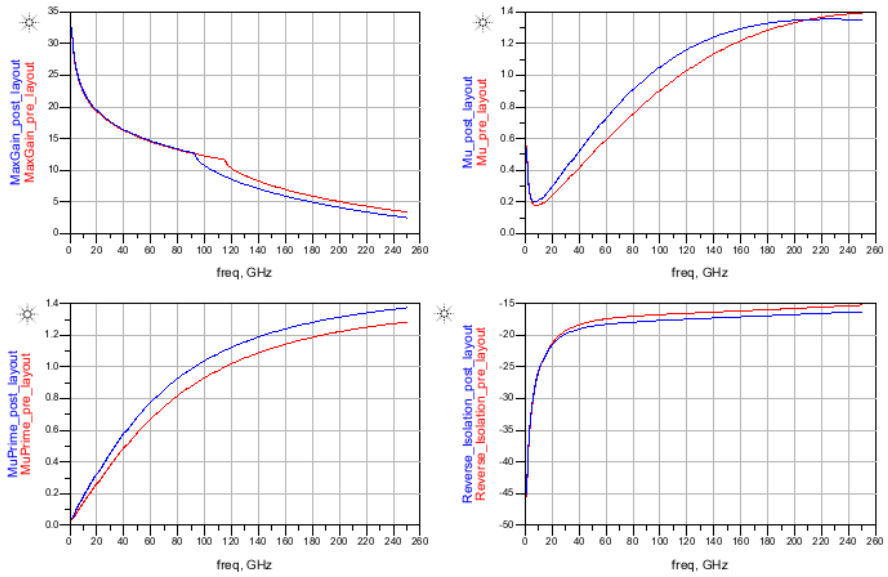


Figure 4.22 Small Signal RF Performance of a CE stage before (red curve) and after (blue curve) parasitics. MaxGain and reverse isolation are in dB.

In figure 4.22, the post-layout small signal performance of the CE stage is depicted. The device is biased to a point close to the peak of the f_{MAX} curve. For the frequency of interest, i.e. 135 GHz, the CE topology has a very small MaxGain, namely 7.4 dB. For the last stage of a PA, we are going to end up with a smaller value, due to the fact that we match for output power and not for power gain. Hence the output will be deliberately mismatched and the power gain will be reduced. The stage is potentially unstable up to 90 GHz. The reverse isolation is quite poor.

In figure 4.23 the layout of a CB stage is depicted. The biasing conditions are the same as in the case of the CE stage. Narrow metals can be used for the base lead, since the currents are low and there is no risk for electromigration. In section 4.5 we showed that in order for stability to be ensured, the inductance of the base lead to ground should be minimal and then it should be decoupled as soon as possible. From simulations, we saw that increasing the width of the base interconnect reduces the inductance and improves stability. Hence, a wide line is chosen.

Additionally, another approach is used here to increase stability and the associated layout is shown in figure 4.24. Here, the base interconnects are “sandwiched” between the ground plane which consists of all the available metallization up to TopMetal1. The capacitance due to the close proximity of the base interconnects and the ground plane offers some decoupling very close to the device. Furthermore, a grounded line on Metal5 runs on top of the base interconnects. By providing a return path very close to a conductor, we minimize the current loop and hence its inductance [16]. From simulations, the inductance is reduced to about 1/3 of the original value. Also the resistance of the conductor is also increased due to the proximity effect, which reduces the Q of the network.

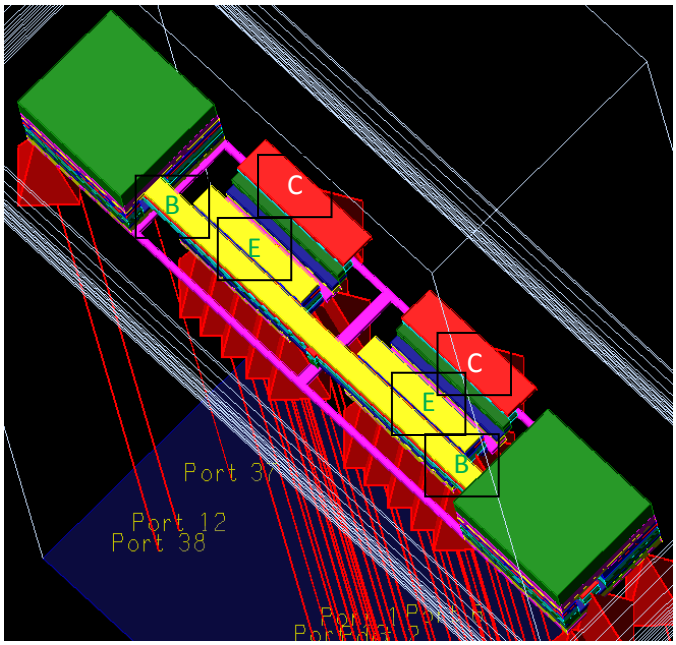


Figure 4.23 CB layout

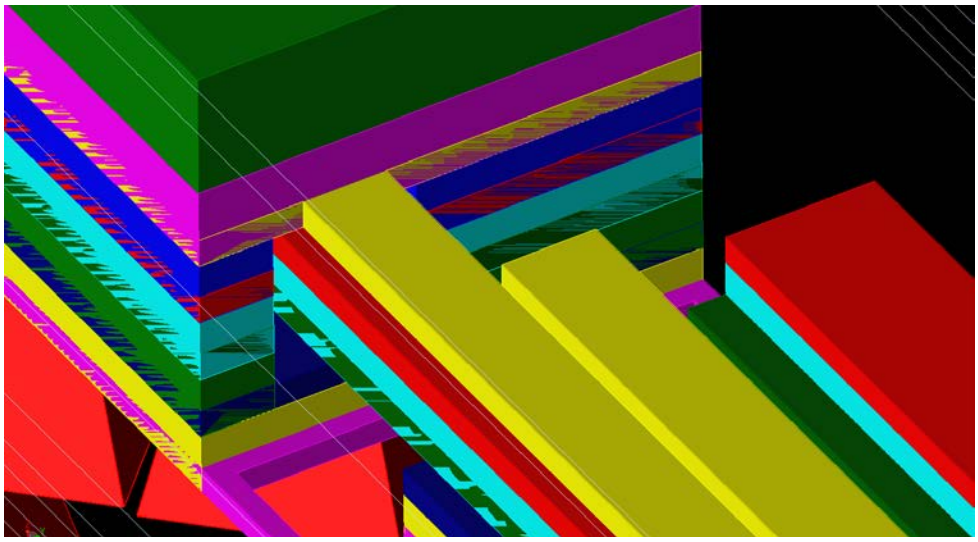


Figure 4.24 CB layouts to reduce base inductance and introduce capacitance

Despite the measures taken above, the stability problem is not alleviated, as it can be seen from figure 4.25. The instability comes from the inductance in the base and causes problems even before the inclusion of layout parasitics. Also the capacitance between emitter and collector reduces the isolation between input and output. Neutralization techniques can be used to solve the latter issue [6]. However, we did not use them here due to the parasitics they introduce and the inadequate modeling experience at such high frequencies. Despite the instability, we can still design with this cell but only by ensuring that the source and load impedances are in the stable regions of the Smith Chart for all frequencies. Finally, the gain is

also moderate for this stage. All the above observations make this topology rather unattractive for operation at high frequencies.

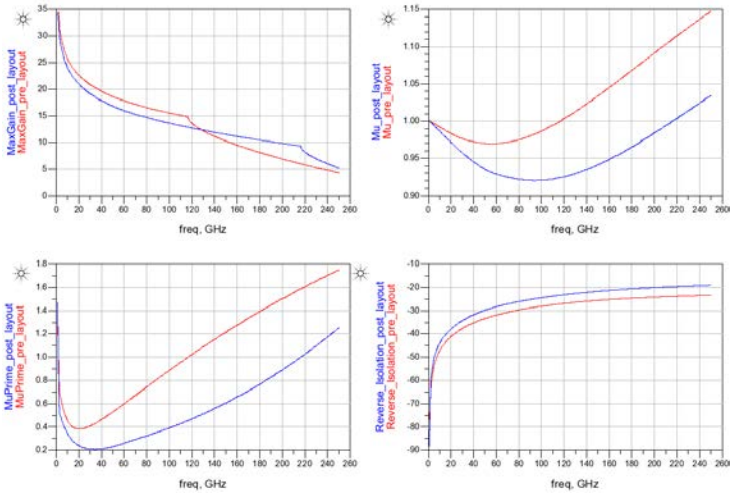


Figure 4.25 Small Signal RF Performance of a CB stage before (red curve) after (blue curve) parasitics. MaxGain and reverse isolation are in dB.

Finally, the cascode configuration is going to be examined. The layout is a combination of the ones described above and it is depicted in figure 4.26. By watching closely, one can realize that is consists of a CE and CB layout, in which the collector of the first is connected to the emitter of the second. The layouts details of the base interconnect are exactly like the ones of the CB stage described above. The connection of the CE and CB stages is implemented on TopMetal1. Also the collectors of the CB devices are taken out on TopMetal1.

In figure 4.27, the small signal characteristics of the cascode topology are depicted. We note improved stability compared to the CB amplifier. This can be attributed to the increased isolation. Furthermore, the MaxGain is very high at 17dB. Hence, this topology seems to meet all requirements for a PA stage at least from a small signal point of view. Even if some large signal characteristics like output power and efficiency are better for the other configurations, the limited power gain at these frequencies as well as the high losses of the passive components, make the cascode the topology of choice.

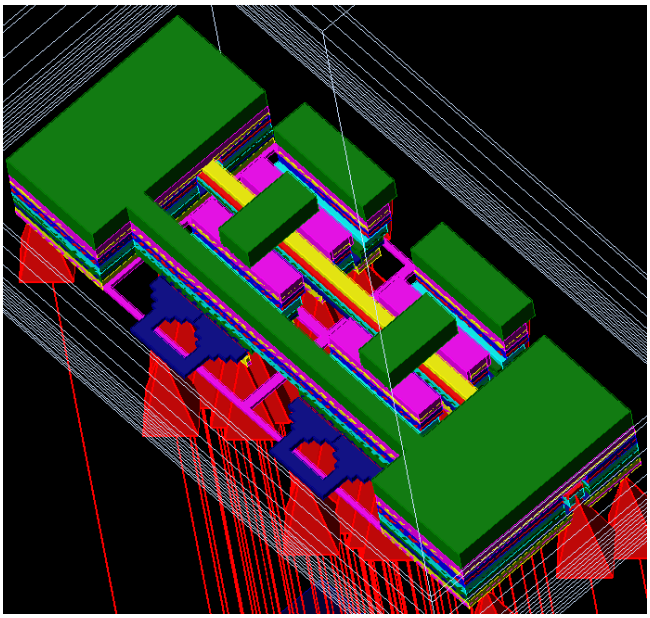


Figure 4.26 Small Signal RF Performance of a CB stage after parasitics.

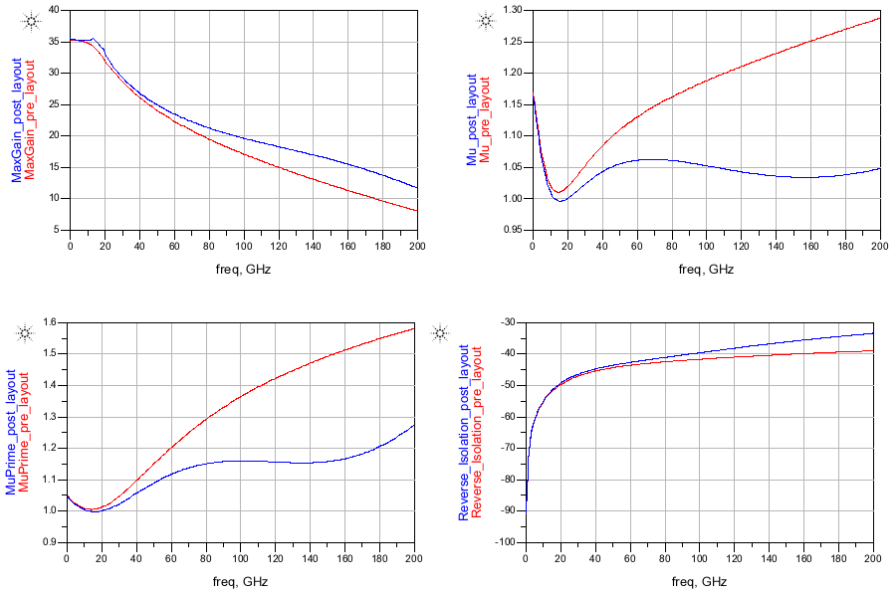


Figure 4.27 Small Signal RF Performance of a Cascode stage before (red curve) and after (blue curve) parasitics. MaxGain and reverse isolation are in dB.

4.7. Conclusions

In this section, we addressed various design challenges regarding the active cells for high frequency PA design. Initially, the thermal stability was analyzed and the cascode turned out to be the most preferable topology especially if extra resistor ballasting schemes are to be avoided. Next, we noted that

interconnects should all assume a certain minimum width due to electromigration risks. Furthermore, the problems of biasing and sizing of the cells were addressed. Finally, the effect of layout on the cell performance was described. Based on the observations of this section, we proposed layouts for the PA active cells. Post-layout simulations point towards the use of the cascode topology, since it is the only one that can provide sufficient power gain along with a sufficient degree of stability.

4.8. References

[1]J. R. Long, “SiGe Radio Frequency ICs for Low-Power Portable Communication”, Proceedings of the IEEE, vol.93, Sep. 2005.

[2]Alan Hastings, “The Art of Analog Layout, 2nd edition”, Pearson Education, Inc., 2006.

[3]J. D. Cressler, “Emerging SiGe HBT reliability issues for mixed-signal circuit applications”, IEEE Transactions on Device and Materials reliability, vol. 4 no. 2, pp. 222-236, Jun. 2004.

[4]A. J. Joseph, J. D. Cressler, D. M. Richey, R. C. Jaeger, D. L. Harame, “Neutral Base Recombination and its Influence on the Temperature Dependency of Early Voltage and Current Gain – Early Voltage Product in UHV/CVD SiGe Heterojunction Bipolar Transistors”, IEEE Transactions on Electron Devices, vol. 44, no.3, pp. 404 – 413, Mar. 1997.

[5]Jinshu Zhang, Hongyong Jia, Pei-Hsin Tsien, Tai-Chin Lo, “Emitter-Ballasting-resistor free SiGe Microwave Power Heterojunction Bipolar Transistors”, IEEE Electron Device Letters, pp. 326-328, Jul. 1999.

[6]Yi Zhao, “High-Performance mm-wave and wideband Large-Signal Amplifiers”, PhD. Dissertation, TU Delft, 2013.

[7]N. Sarmah, P. Chevalier and Ulrich R. Pfeiffer, “160-GHz Power Amplifier Design in Advanced SiGe HBT Technologies with P_{SAT} in excess of 10 dBm”, IEEE Transactions on Microwave Theory and Techniques, vol. 61, pp. 939 – 947, Feb. 2013.

[8]Tak Shun Dickson Cheung, “A Multistage, Parallel-Path Power Amplifier in SiGe Bipolar Technology for mm-Wave Wireless Applications”, PhD Dissertation, University of Toronto, 2007

[9] Jens Lienig, “Electromigration and its impact on Physical Design in Future Technologies”, Proceedings of the Int. Symposium on Physical Design (ISPD) 2013, pp. 33–44, March 2013

[10]Steve C. Cripps, “RF Power Amplifiers for Wireless Communications, Second Edition”, Artech House, 2006.

[11]Sorin Voinigescu, “High Frequency Integrated Circuits”, Cambridge University Press, 2013.

[12]Online: http://www.keysight.com/en/pc-1475688/keysight-eesof-electronic-design-automation-EDA-software?nid=-34360.0&cc=NL&lc=dut

[13]Guillermo Gonzalez, “Microwave Transistor Amplifiers, Analysis and Design, 2nd edition”, Prentice Hall, 1997.

[14]M. L. Edwards and J. H. Sinsky, “A new criterion for linear 2-port stability using a single geometrically derived parameter”, IEEE Transactions on Microwave Theory and Techniques, vol. 40, no.12, pp. 2303-2311, Dec. 1992.

[15]J. Rollet, “Stability and power-gain invariants of linear two-ports”, IRE Transactions on Circuit Theory, vol.9, no.1, pp. 29-32, Mar. 1962.

[16]Eric Bogatin, “Signal and Power Intergrity Simplified, 2nd edition”. Prentice Hall, 2009.

Chapter 5. Power Amplifier Design at mm-wave frequencies

In this chapter, we present the design of a PA operating at a center frequency of 135 GHz to be implemented in an IHP SiGe BiCMOS 130nm process (SG13G2 process). The design concept, topology analysis and the building blocks developed in the previous chapters are here employed to realize the implementation of a medium output power PA, i.e., ~5-8 dBm. The design consists of two active stages to provide a power gain above 15 dB. The matching networks are based on integrated transformers [2], [3], [4], described in chapter 3. These components provide impedance transformation, biasing through the center taps and coupling between stages. Since a single ended to differential interface is required at the input and the dual, at the output, the first and last transformers are configured as balanced to unbalanced structures [5].. The details of biasing and sizing as well as the design of the passive components are presented in this chapter. Finally, some post-layout simulations of the final PA structure are given in order to evaluate the final expected performance.

5.1. Mm-wave frequency Power Amplifier Design Plan

As noted in Chapter 1, the basic specifications for a PA are output power, power gain and efficiency. There are various degrees of freedom and design choices and each one of them affects all three specifications. In order not to get lost in the design process, it is imperative that the specifications are strictly prioritized. In our case, we aimed for a medium output power PA, with the output power being in the range of 5-8 dBm. We would like to achieve this along with a power gain greater than 15dB. In this work, we do not put a strict specification on collector efficiency or PAE in order to maintain the problem complecity into certain limits. Hence the design is not optimal in terms of efficiency.

The output power specification, is the main measure to size the last stage, i.e., such that is capable of providing the required power level. The breakdown limit of the device defines the supply voltage of the device. For a Class-A amplifier, from the output power specification and the supply voltage, we can find the bias current needed fot the specified power to be achieved. To find the size of the device (number of parallel devices) we have divide the total current by the current density chosen for the transistors. The latter is chosen in order to provide optimal operation with respect to speed. The load of the PA is the resistor that stems from the ratio of output voltage and output current. The method described above uses DC curves and is theoretically independent of frequency. It is used to extract the highest amount of output power that a certain device is capable of providing.

However, at higher frequencies certain parasitics can change the device behavior [4], i.e. the output impedance and the transconductance. This may have a direct impact on the value of the optimum load resistance. Furthermore, we might want to choose an output power lower than the maximum one, as output power and power gain are contradictory specifications. Thus, by sacrificing output power, we are able to achieve a higher power gain. In order to capture these phenomena Load-Pull simulations [6] can be performed and the results can be presented on the output reflection coefficient plane on the Smith Chart. During a Load-Pull simulation, the load impedance of the PA is varied in both real and imaginary parts. The results are constant output power contours, i.e. contours of load impedances that give a certain amount of power. In the same Smith Chart, we can depict Operating Power Gain Circles as well as the Output Stability Circle [1]. In this way, we can choose the reflection coefficient that can, at the same time, give adequate power, adequate power gain as well as stable operation.

The amount of gain obtained from the last stage and the driver stage will be a first indicator of the number of stages that are going to be needed. The last stage operates in compression, where gain compression exists. The driver stage is driven by a lower input power, hence its gain will be higher as it operates in the un-compressed region. In order to get sufficient power gain, which is scarce at these frequencies, all stages are biased close to the peak f_{MAX} value by choosing the collector current accordingly. The second indicator will be the power losses of the matching networks. It turns out that two stages are adequate in order to achieve the desired power gain specification, i.e. 15 dB, when cascode active cells are used.

Regarding the matching networks, the output stage is designed in order to provide to the collector of the device the right impedance value that provides a certain output power, power gain and stability. Also the output matching network should present low power loss, since any loss from it directly penalizes output power, power gain and efficiency. Furthermore, it should provide differential to single ended conversion. The intermediate matching network will be used to provide a conjugate match between the two stages in order to achieve a high power gain. Since the interface here is differential, the center tap will be at virtual ground. The input matching network is critical only for balance, since it does not have a direct impact on output power or power gain.

5.2. Sizing of the output stage and determination of optimal load impedance

As stated in Chapter 4, in this process, the basic active cell consists of a 0.13x1.1 um^2 emitter area.In order to achieve the desired output power level, from Load-Pull simulations, five parallel devices are sufficient in order to provide an output power of 5-8 dBm. Apart from otuput power, a sufficient power gain is also needed. Output power and power gain are contradictory specifications [6]. In order to get a certain amount of output power, the amplifier should be terminated with the suitable impedance that stems from Load-Pull simulations. In order to maximize the power gain, conjugate match should be performed. Hence a compromise should be made in order to achieve the desired amount of power along with a sufficient power gain. Towards this direction, we can get great help from the graphical aid depicted in figure 5.1. The blue set of curves is constant output power contours ranging from 5.2 to 9.2 dBm. We see that for a higher output power, the power gain reduces. The 5.2 - 7.2 dBm power contours lie between the 12 and 14 dB power gain circles. Hence the desired reflection coefficient lies in this part of the Smith Chart. The output matching network should provide to the active cell, impedances in this range.

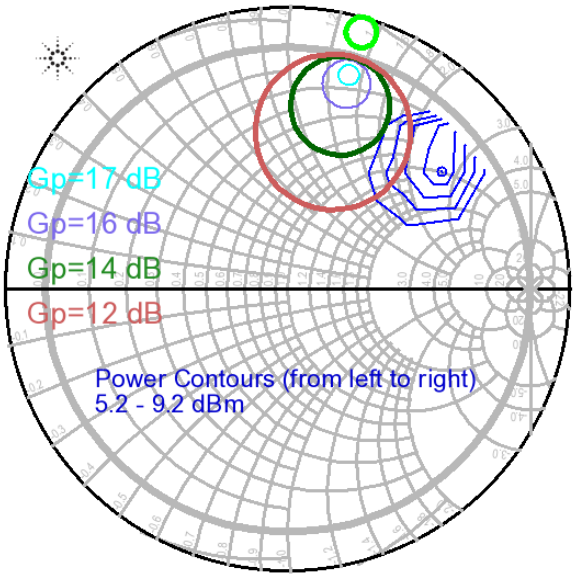


Figure 5.1 Graphical aids for the sizing of the output stage and selection of the load reflection coefficient

The layout and the associated results for the device are the same as the ones given in Chapter 4.

5.3 Output matching network design

In order to quantify the power loss of the output passive device, the operating power gain is considered, see eq. 5.1 [1].. Note that in the power gain definition, no input mismatch is accounted for Using this expression we correctly account for the losses of the output matching network since this specific passive is not designed for conjugate match, due to the non linear operation of the last cells. It is important to note that the gain reduction due to a non-optimal impedance level is already accounted by the large signal contour circles of 5.1.

$$G_P = \frac{1}{1-|\Gamma_{IN}|^2}|S_{21}|^2\frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2} \quad (5.1)$$

As shown in Fig. 5.2, a floating patterned shield is used under the passives. This was done mainly in order to have defined conductor boundary around the passive in order for the simulation to be more accurate. Furthermore, a shield can alleviate losses due to the conductive substrate. However, after the shield was added no reduction in the power loss was observed. Hence substrate losses are not the dominant ones. By leaving the passive unshielded, we expose it to the underlying silicon substrate. The properties of the latter depend on various process variation factors, for example substrate resistivity. Furthermore, the total substrate thickness, which is usually 300um after CMP, cannot be accurately simulated by Momentum which presents convergence problems even for substrates as high as 80um. To minimize the uncertainty due to these factors, we chose to put a floating shield under the passives. The use of a patterned shield will also help us pass the DRC metal density rules. Unfortunately, the presence of the shield seems to increase the loss of the passive by ~0.6 dB. The loss increase is not that dramatic, hence we chose to shield it

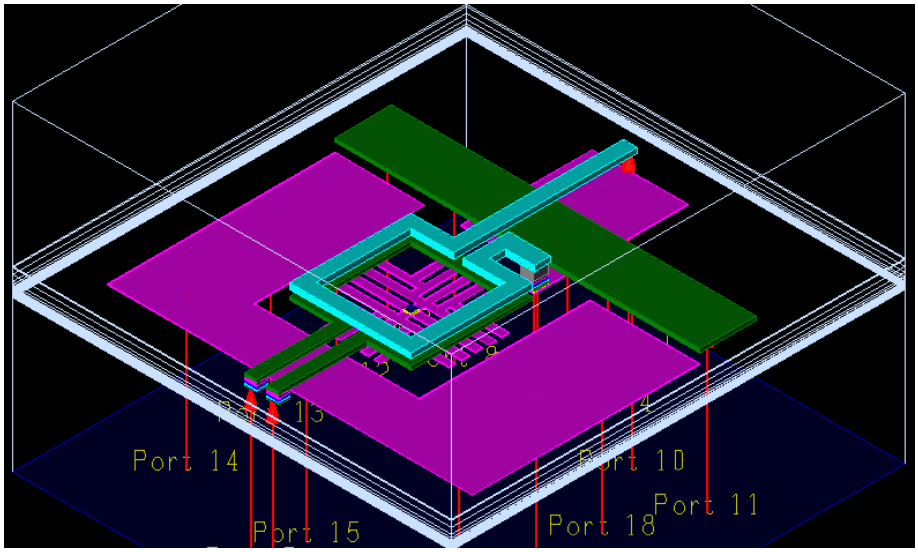


Figure 5.2 Output Matching Network.

The structure to be simulated should be identical to the one used in the final layout in order for the EM simulation model to capture accurately the behavior of the physical structure. This is why the structure of figure 5.2 contains all the adjacent metals and connections as they are implemented in the final layout of the structure. The primary winding is implemented on TopMetal1 (2 um thick) depicted in green and the secondary is implemented on TopMetal2 (3 um thick) depicted in light blue. The ground plane surrounding the structure (purple), the shield (purple) as well as via stacks connecting the primary to the collector of the active stage and the secondary to ground are included in the EM simulation.

The center tap connection is implemented as a very wide line to minimize inductance. As mentioned in Chapter 3, a capacitance is required to resonate out the center tap inductance at the center frequency in order to achieve a balanced behavior when the structure is going to be used as a balun.

The performance of the passive is simulated and the SRF is ~ 160 GHz. Given the fact that the capacitive loading from the device and the pad are small, the passive can still be used at 135GHz. The coupling factor is around 0.55. This is due to the fact that the distance between the windings is large and equal to 3um. The loose coupling is a price we have to pay in order to ensure a high SRF. A stacked configuration was used for the transformer. The interleaved transformer was found to provide even lower coupling for proximities leading to sufficiently high SRFs.

The simulation result is shown in figure 5.3. For the chosen frequency range, the loss is close to 4 dB, which is considerable. In order to compensate for this, we can fine-tune our matching network in order to move towards higher power impedance levels in the graph of figure 5.1. Under this regime, the devices outputs more power and the total output power is higher at the expense of power gain.

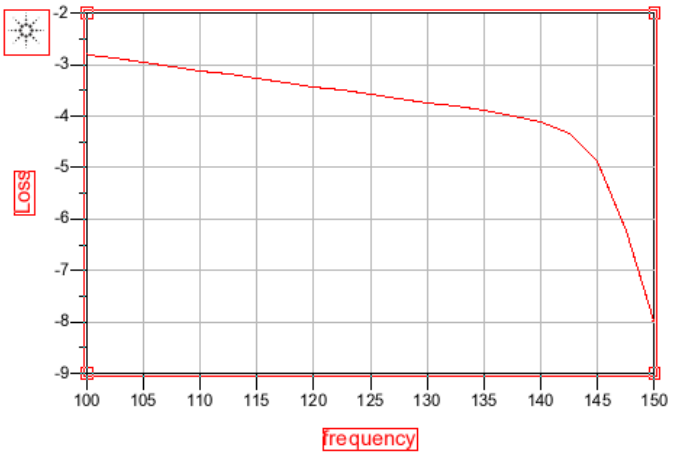


Figure 5.3 Loss of the output transformer balun

Finally, the balance of the structure needs to be evaluated. By placing a certain amount of capacitance at the center tap, we can achieve an improvement in the balance of the structure. The structure as drawn in the final layout is given in figure 5.4. Results concerning the balance are given in figure 5.4... From simulations a total value of approximately 240 fF is estimated for the balancing capacitance. This value is smaller than the capacitance of the line connecting to the pads, as is in the order of the parasitic capacitance of many building block, for this reason accurate modeling and EM simulationof the actual capacitance (laid out) and the entire structure should should be performed.

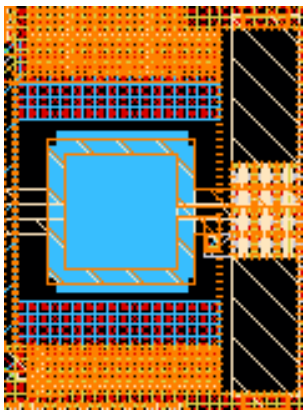


Figure 5.4 Output transformer bal-un with capacitance at the center tap

The results before and after the addition of the proper capacitance at the center tap are shown in figure 5.5. In the passive without any capacitance at the center tap, the amplitude imbalance is close to 3 dB. After the addition of 240fF of center tap capacitance, the amplitude imbalance becomes is reduced to only 0.47 dB, while the phase imbalance is reduced from close to 15 degrees to 2.5 degrees, allowing to minimize loss of energy in the common mode..

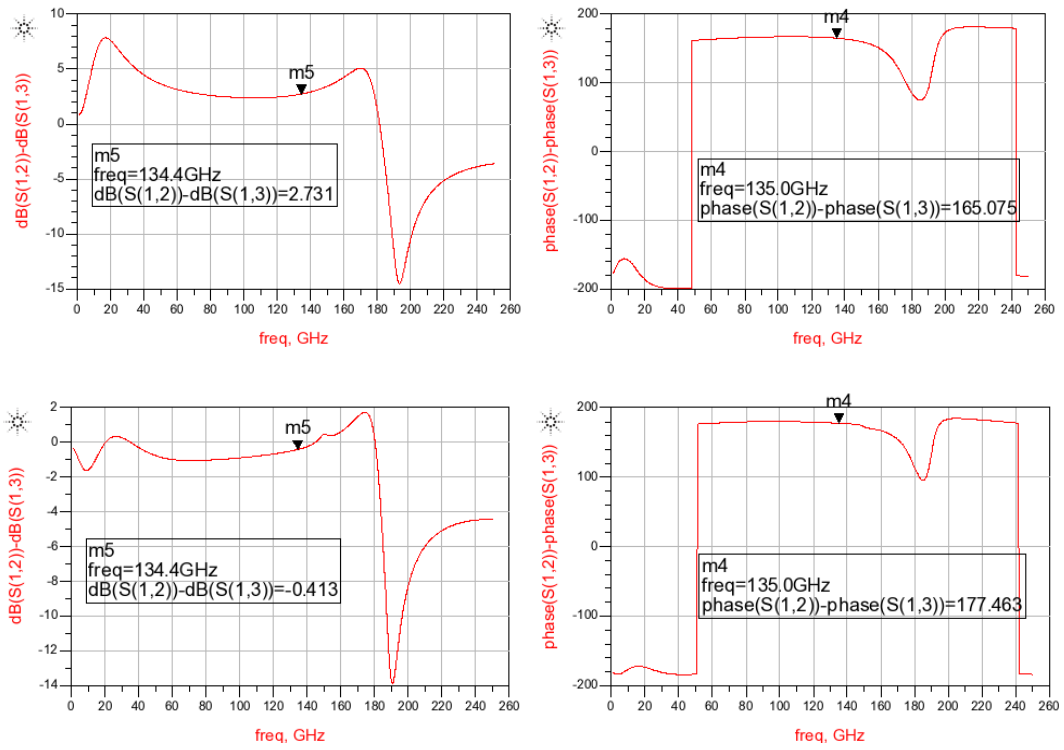


Figure 5.5 Balance of the output bal-un before (upper graphs) and after (lower graphs) the addition of the center tap capacitance.

5.4. Sizing of the driver stage

The driver stage should provide power in order to drive the last stage in saturation [6]. After the output stage loading condition for the prescribed output power is determined, the power gain of this stage can be simulated. The driver stage should provide the additional gain in order to reach the power gain specification.. In our design we sized the driver device with a ratio of 3:5 compared to the output one.
The chosen driver to output stage ration can be justified by the following considerations:

- Increasing the area of the driver stage would lead to a large bias current also for the driver stage (i.e., comparable to the output stage) leading to a loss in the PA efficiency.
- Reducing the ration to 2:4 (often used in literature) would results in a lower capacitance of the driver cascade stage. Note that this capacitance is resonated out from the leakage inductor of the inter-stage transformer, thus a small capacitance would require a larger inductance to be resonated out for the same frequency. This results in an overall loss of performance in the transformer as was shown in chapter XX, where the area optimization of the transfer based on the SRF was given.

The layout and the small signal performance of the driver stage are given in figures 5.6 and 5.7 respectively. The results are similar to the ones for the five device cell. Mu factor points to the edge of instability. More detailed stability analyses are performed on a later section.

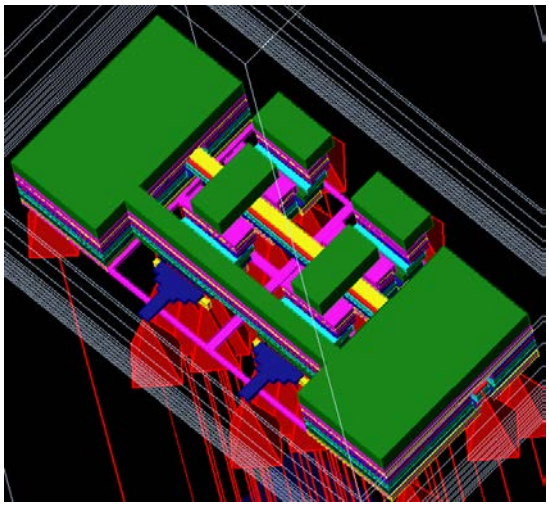


Figure 5.6 Layout of cascode devices

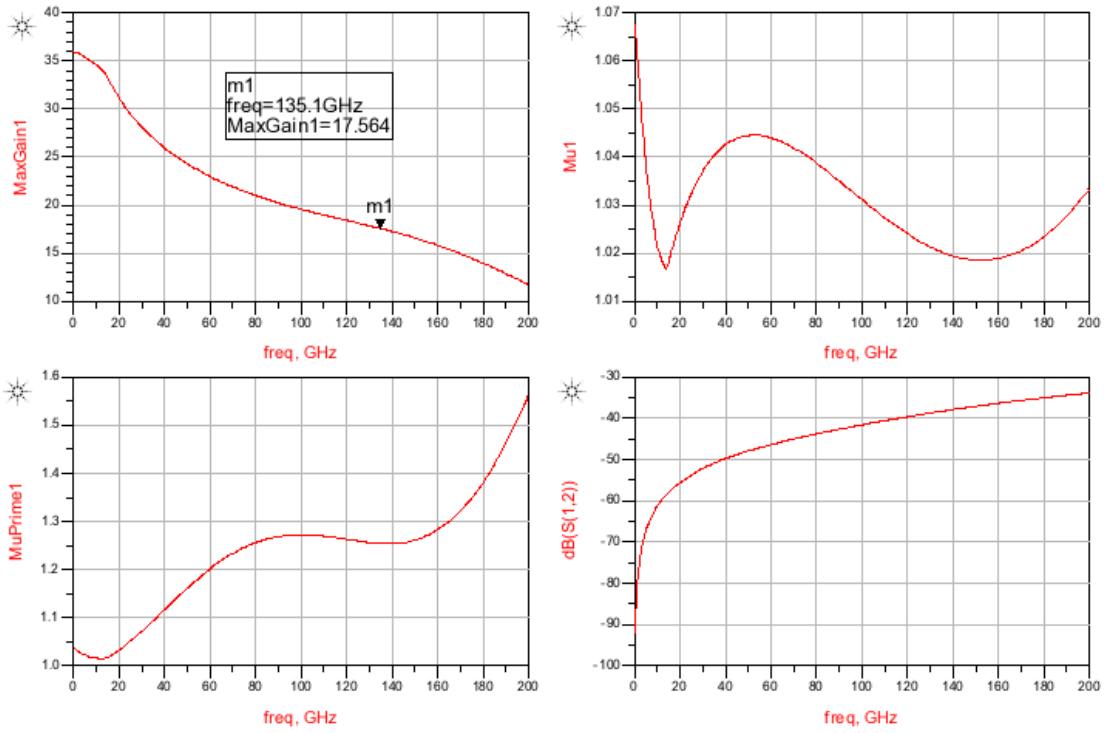


Figure 5.7 Post-Layout Simulations for the output stage of the PA.

5.5. Intermediate Matching Network

The intermediate matching network transforms the input impedance of the output stage to the complex conjugate output impedance of the driver stage. Conjugate match instead of output power match is preferred here in order to maximize the power gain. A floating shield was put under the passive in order to shield it from the underlying substrate and circumvent uncertainties stemming from the substrate inaccurate modeling as was previously described.

Certain design choices were made associated with the intermediate matching network. First of all, one of the windings is implemented on Metal5, while the other is implemented on TopMetal2. Metal5 is thin, but no electromigration risk exists since it is used to feed the DC current to the base of the device, which is in the order of tens of microns. Furthermore, there is a large distance between Metal5 and TopMetal2, leading to a lower coupling coefficient, thusincreasing the leakage inductance of the transformer which is $(1-k)L_p$ as shown in the transformer model of Chapter 3. This extra inductance is required to resonate with the output capacitance of the driver stage at the operating frequency, which as mentioned before is larger than the output stage one. However, the transformer size (set by the inductance required to resonate out the output stage capacitance) is relatively small requiring a large extra inductance to resonate the capacitance of the driver stage. For this reason,a transformer-inductor co-design leading to minimum loss can be achieved after fine tuning through extensive EM simulations. The final structure consists of a transformer with an inner diameter of 40um and an extra meandered line, with inner area of 27.5x20 um^2 at the input, providing the extra inductor implementation. The coupling factor is calculated to be 0.45 at the frequency band of operation.

The power loss of the network is quantified by using the transducer power gain [1]. Here the reflection losses should be taken into account apart from the power losses. The loss is presented in figure 5.9 and it is around 4.5 dB. This will affect the power gain, but not the output power as soon as the output stage is driven to a full swing.

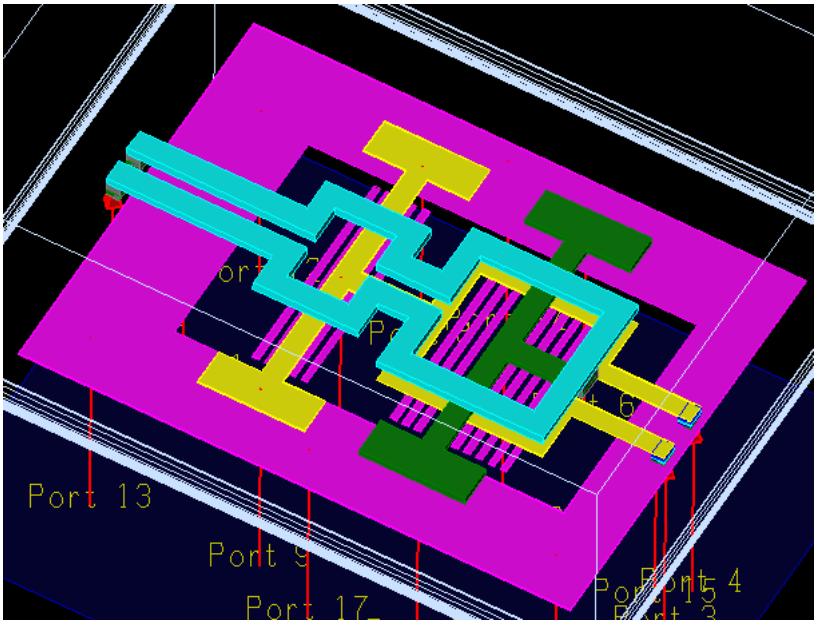


Figure 5.8 Layout of the intermediate matching network

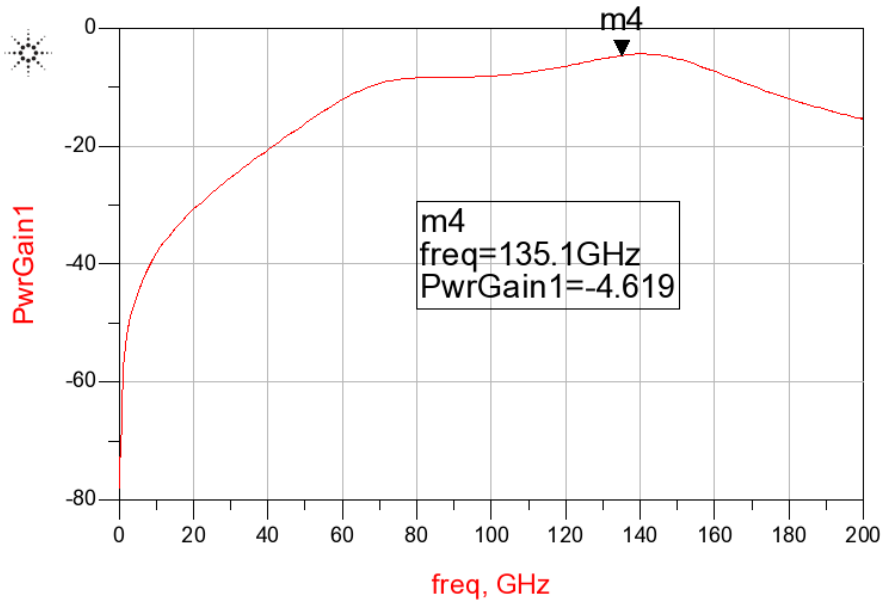


Figure 5.9 Power Loss of the intermediate matching network

5.6. Input Matching Network

The input matching network is used to interface the single ended input pad to the differential input of the driver stage Since the input power to the whole amplifier can be increased by increasing the output power of the power source, the loss of this network is not critical to the performance of the amplifier. The coupling factor is simulated to be 0.5. However, the loss appears to be increased by ~4dB in comparison to the output matching network, despite the similar coupling coefficient. This difference stems from the reflection losses due to mismatch rather than the small coupling factor. The inner diameter of the transformer is 40um. The capacitance used for the achievement of balance is again roughly 240 fF, since the center tap structure is the same as the output transformer. The structure used for the EM simulation is shown in figure 5.12 and it includes all the adjacent structures. The transducer power gain is used to quantify the power loss and the simulation result for the power loss is shown in figure 5.10. Note that the resulting high losses are mostly deriving from impedance mis-match to the 50 Ohm source.The relations between amplitude and phase of the differential terminals are shown in figure 5.12. There is a drastic improvement after the addition of the capacitance at the center tap of the structure. This section completes the design of the PA building blocks.

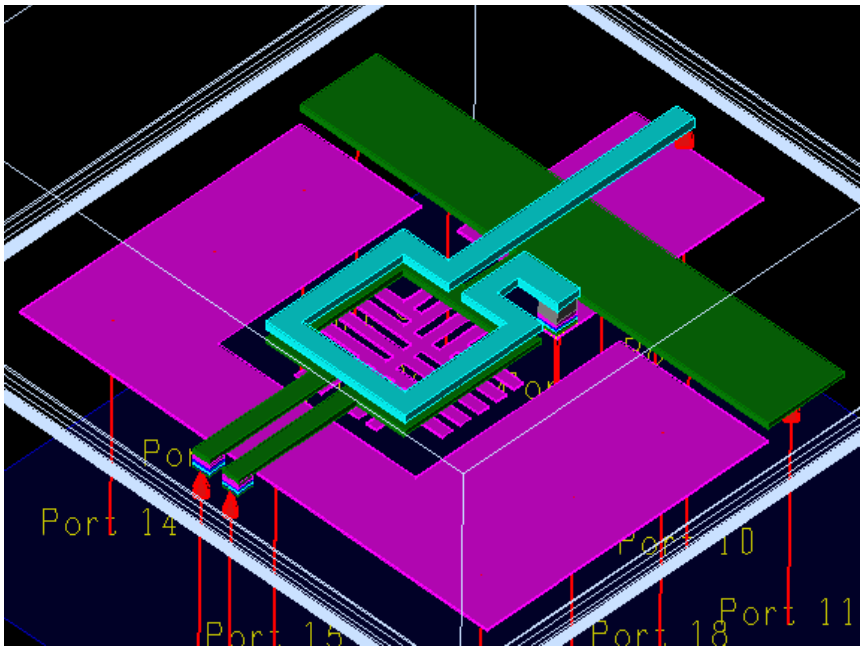


Figure 5.10 Layout of input matching network

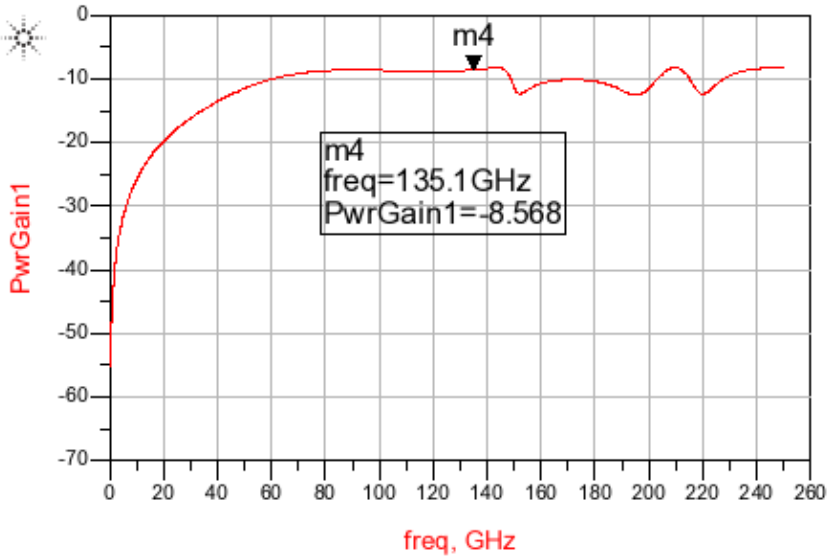


Figure 5.11 Power Loss of the input matching network

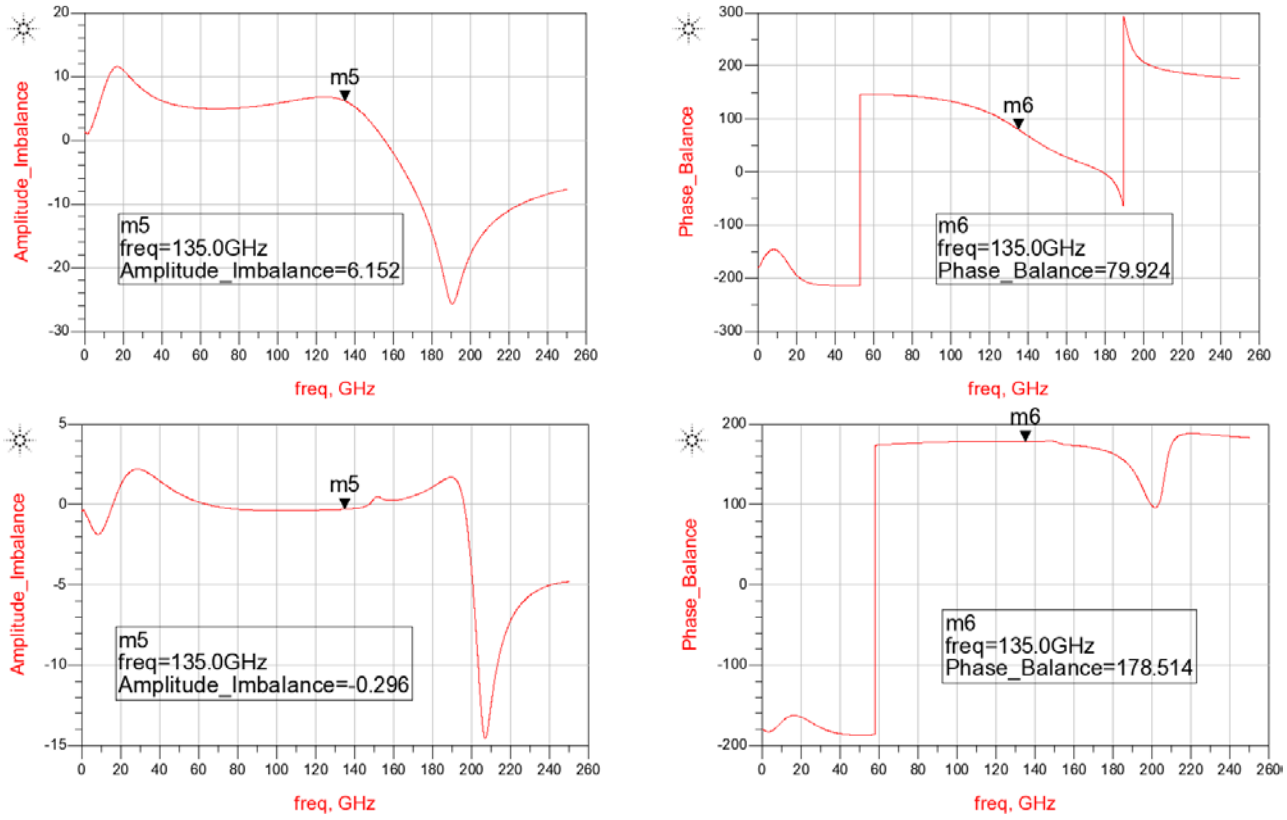


Figure 5.12 Amplitude and Phase relation of differential terminals before (upper graphs) and after (bottom graphs) the addition of the capacitance at the center tap of the structure.

5.7. Total Power Amplifier (PA)

The final structure of the complete PA is described in this section. The dimensions of the active cells and the passives are also annotated on the schematic. Also the key characteristics of each individual block are given in Table 5.1. In Table 5.2, the small-signal impedance levels at the ports of the active and the passive devices are given. C_{PAD} is the capacitance of the input and output pads as predicted by the EM simulations. The pads are shielded from the underlying conductive substrate to minimize the losses due to capacitive coupling and also to alleviate any inaccuracies and uncertainties stemming from the substrate modeling. M1 and M2 are used as the ground plane and hence they are used as the shield. The pad capacitance is simulated to be around 25 fF.

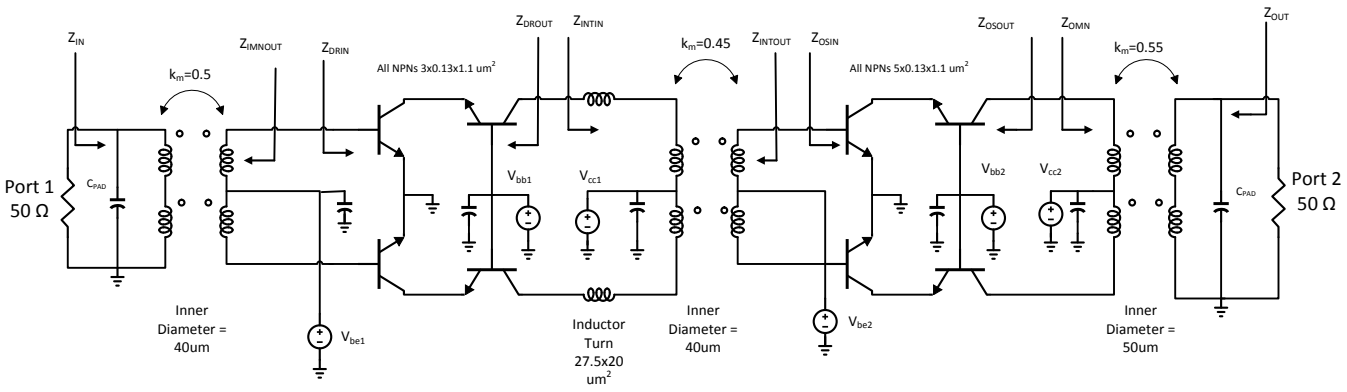


Figure 5.13 Final PA schematic

Table 5.1. Key characteristics of each individual building blocks

Block	Important Specifications
Output Matching Network	Power Loss, Balance
Output Stage	Output Power, Power Gain, Stability
Intermediate Matching Network	Power Loss
Driver Stage	Power Gain, Stability
Input Matching Network	Balance

Table 5.2. Small Signal Impedance Levels for the various cells of the amplifier at 135 GHz

Block	Small Signal Impedance (Ω)
Output Impedance (Z_{OUT})	30+j1
Output Matching Network Input Impedance (Z_{OMN})	36.4+j*117
Output Stage Output Impedance (Z_{OSOUT})	10-j*136
Output Stage Input Impedance (Z_{OSIN})	19-j*30
Intermediate Matching Network Output Impedance (Z_{INTOUT})	25+j*50
Intermediate Matching Network Input Impedance (Z_{INTIN})	40+j*232
Driver Stage Output Impedance (Z_{DROUT})	20-j*214
Driver Stage Input Impedance (Z_{DRIN})	40-j*45
Input Matching Network Output Impedance (Z_{INOUT})	22+j*90
Input Matching Network Input Impedance (Z_{IN})	9-j*30

5.8. Layout Details

In this section, various implementation details of the layout are given.. The final layout of the PA is shown in figure 5.14. The three passives can be clearly seen as well as the active cells in between. GSG pads are used in order to interface the circuit with the measurement equipment.

The configuration of the PAD depends on the probe type and pitch, as shown in figure 5.15, and on the capacitance needed, from pad to ground, in order to resonate with the winding inductance at the frequency of interest to minimize transformer losses, as mentioned in Chapter 3. In our case, the pitch of the probe is 100um and this is the distance between the ground and the signal pad on the chip. The signal pad is the main contributor of capacitance to ground, since the distance to ground is only 10um, while the distance from the signal the ground pads is ~40um. Hence, it should be made sufficiently small in order to provide a sufficiently small capacitance. However, the minimum size depends on the minimum pad opening for reliable measurement. If the opening is too small, we may not be able to land the probe on the pad accurately. Here we choose an opening width of 40 um which is sufficient for reliable measurement. The length of the pad cannot be made too small because the probe requires to “skate” on the pad, i.e., scratch the native aluminum oxide. A length of 50um is chosen as the pad length. The ground pad dimensions are not that critical and the dimensions chosen are 100x115 um^2. The pad capacitance from signal to ground is simulated to be 25 fF.

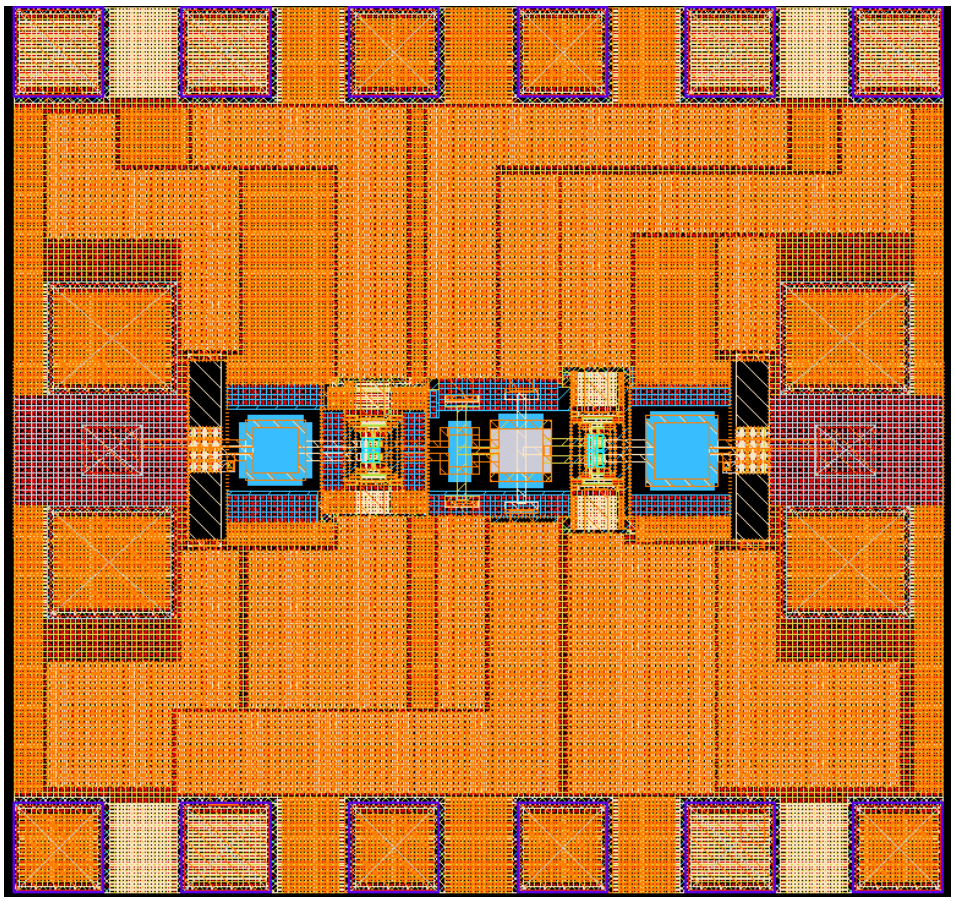


Figure 5.14 Final PA layout.

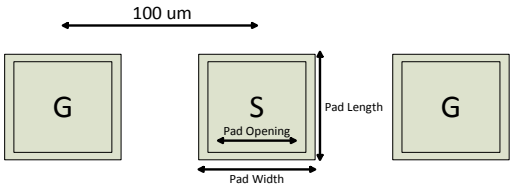


Figure 5.15 GSG pad configuration

The ground plane has a significant effect on the performance of high-frequency circuits since its parasitics can be significant at high frequencies. This is due to the fact that the bottom metals usually employed to create the ground plane are in the order of 0.35um each. Hence, even shunting multiple metals together cannot reduce the inductance and resistance to very low values. Furthermore, the use of continuous metal planes is prohibited due to metal density rules for enhanced reliability. This will further increase the inductance and resistance of the ground plane.

In figure 5.16, we present the basic cell element which is repeated to form the ground plane. The blue cross is Metal 1 and the white square is Metal2. These two layers are connected with vias. The thickness of the cross and square lines can be changed in order for the metal density rules to be satisfied. The red squares belong to the "Poly" layer, since there is a density rule for polysilicon. Finally the green region is "Active", which corresponds to diffusions associated with active devices, and density limits exist for this layer also. More metals can be added to the cell forming subsequent crosses and squares until all the desired metals are included. The total ground plane is shown in figure 5.17. The ground plane extends over the entire chip and it is continuous. Cutting the ground plane is not recommended since the return current flows through it. Cutting the path of return currents can force them to follow unpredictable paths that cannot be EM simulated, hence making the behavior of the circuit unpredictable. Furthermore, narrow paths in the ground plane can cause excessive amounts of ground parasitics and even worse they can cause the destruction of the ground plane metals due to electromigration.

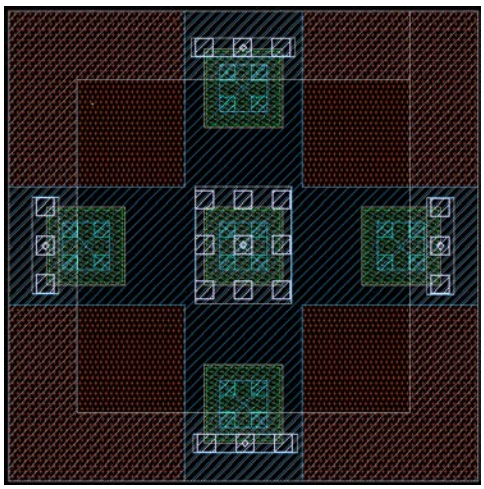


Figure 5.16 Ground Plane Cell

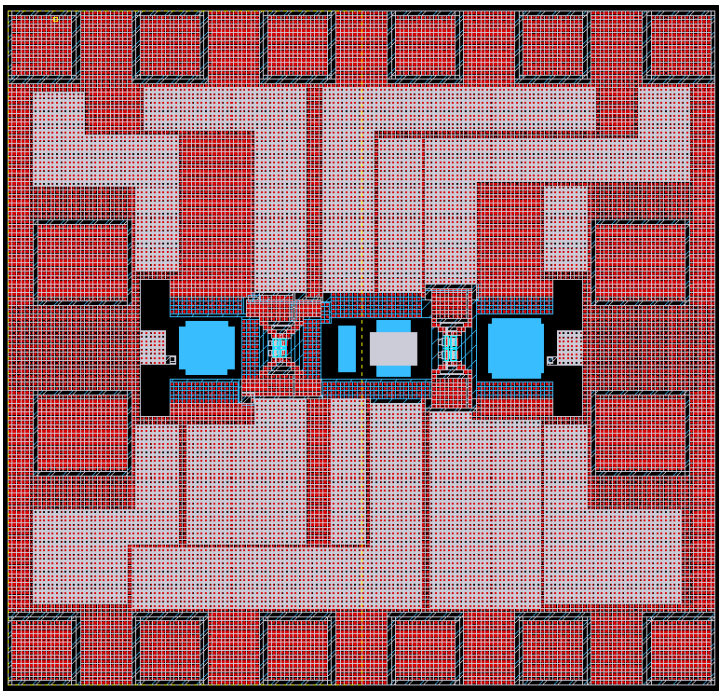


Figure 5.17 Total chip ground plane

Finally, the decoupling scheme requires some attention. Since the circuit is differential, the need for decoupling is relatively relaxed for the differential operation. For the common mode regime, we should introduce some capacitance in order for the critical nodes to be of low impedance. For perfectly symmetric layouts and perfect balance in the balun structure, no common mode operation will exist. However, in practice there will always be some common mode component due to noise or coupling from the environment. Hence, the supply lines have to be decoupled.

In order for the decoupling network to operate properly, the amount of capacitance inserted should provide relatively low impedance for the frequency band of interest and all frequencies below it. The decoupling schemes on the chip were designed in the following way:

- The supply lines are made as wide as possible on the thickest metal of the process, i.e. TopMetal2, in order to minimize the inductance.
- To simulate a decoupled line, the inductance of the line is first calculated by EM simulations and afterwards, we tap it to equal amounts of inductance and add the decoupling capacitance blocks.

The impedance of a typical decoupled line used on the chip is simulated and the result is shown in figure 5.18. The magnitude of the impedance is at 4.5Ω at the frequency band of interest and remains lower than 10Ω down to very low frequencies. A resonance occurs around 5GHz. This resonance peak can be dealt off –chip using larger capacitance and resistances to provide the appropriate low frequency behavior.

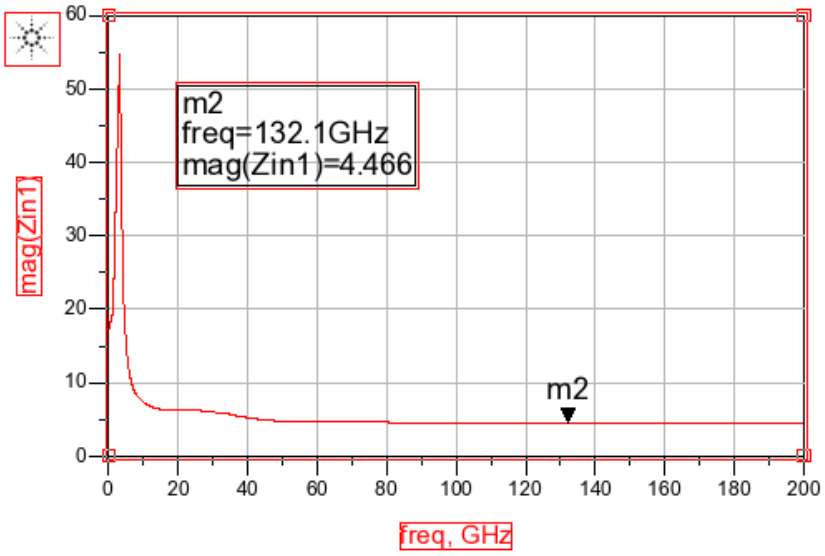


Figure 5.18 Input impedance of supply line decoupling network

The decoupling capacitance cells are shown in figure 5.19. 3-D views are given in figure 5.20. Metal1 and Metal2 form the ground plane and at the same time the bottom plate of the capacitance. The rest of the metals are inter-wound and connected properly in order to make the second plate of the capacitor. The capacitance value is simulated and found to be around 20 fF with a series resistance of ~6 Ohms. Higher capacitance values are not feasible due to the metal density rules that the process dictates. The use of thicker metals results in a bigger area filled with metal and DRC errors occur.

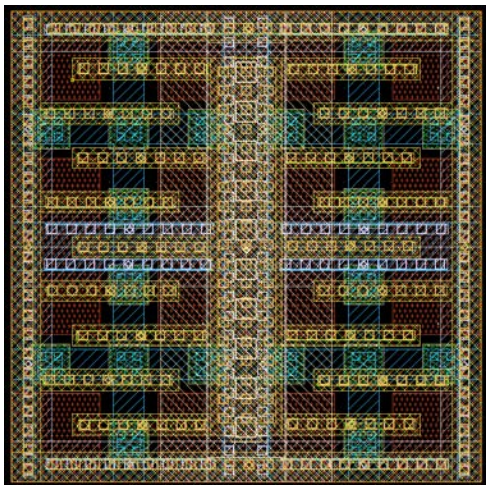


Figure 5.19 Top view of the capacitor layout

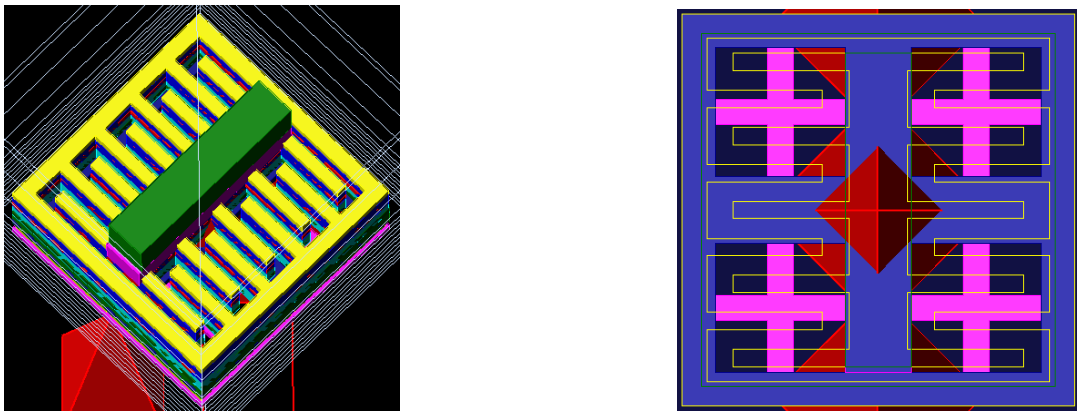


Figure 5.20 3D views of the decoupling capacitor. View of the fingers (left) and view of the bottom metals used as a ground plane (right).

5.9. Stability

After the design and layout details have been presented, we have the final circuit schematic with the active devices co-simulated with the layout planar structures, the passive structures and the modeled capacitive decoupled lines. Before delving into the simulations of large and small signal operation, we have to ensure that the structure is stable. Unconditional stability is desired, since under this regime no variations in loading conditions can cause undesired oscillations. Furthermore, since this is a differential structure, attention should be given to both differential and common mode stability [4].

Since this is a multi-stage amplifier, we have to examine the stability of the individual stages apart from the stability performance of the whole amplifier with all the stages included. First the two stages are examined separately. One way to judge stability is to look at the input impedance of each port, input or output, while the other port is terminated with the impedance which loads it in the actual circuit. If the real part becomes negative, the circuit is potentially unstable at the associated frequency [1].

The ultimate stability check for an amplifier is a transient analysis. However, in our case this is extremely difficult as lumped models have to be derived for all the layout structures from DC up to the frequency where the device presents power gain higher than unity. However, we can reliably judge the stability of the network by using the Mu factor [1]. This stability measure is the most useful, since stability can be judged by only calculating the value of one parameter, i.e. no extra parameters needs to be calculated as in the case of the Rollet stability factor. For values of Mu above unity, the two-port is unconditionally stable. Furthermore, the value of Mu quantifies the degree of stability, i.e. the higher the Mu factor the more stable the device. When calculating the Mu factor, if the frequency step is fine enough, reliable conclusions regarding stability can be drawn.

5.9.1 Differential Stability

We start from the stability study of the single stages. We will use the negative impedance criterion for stability assessment. This means that we can terminate the port with any impedance without risk of unwanted oscillations. For each stage, two simulation setups need to be set. The first will allow us to find the impedance of the input port when the output port is loaded exactly as in the final circuit. The second

one will let us find the impedance when looking towards the output port, while the input port is loaded as in the actual circuit.

The stability setups for the final stage are shown in figure 5.21. In the top figure, we drive the input with a 100Ω port and calculate the input impedance as described above. The same is done for the output port as shown at the bottom part of the figure. The results are shown in figure 5.22.

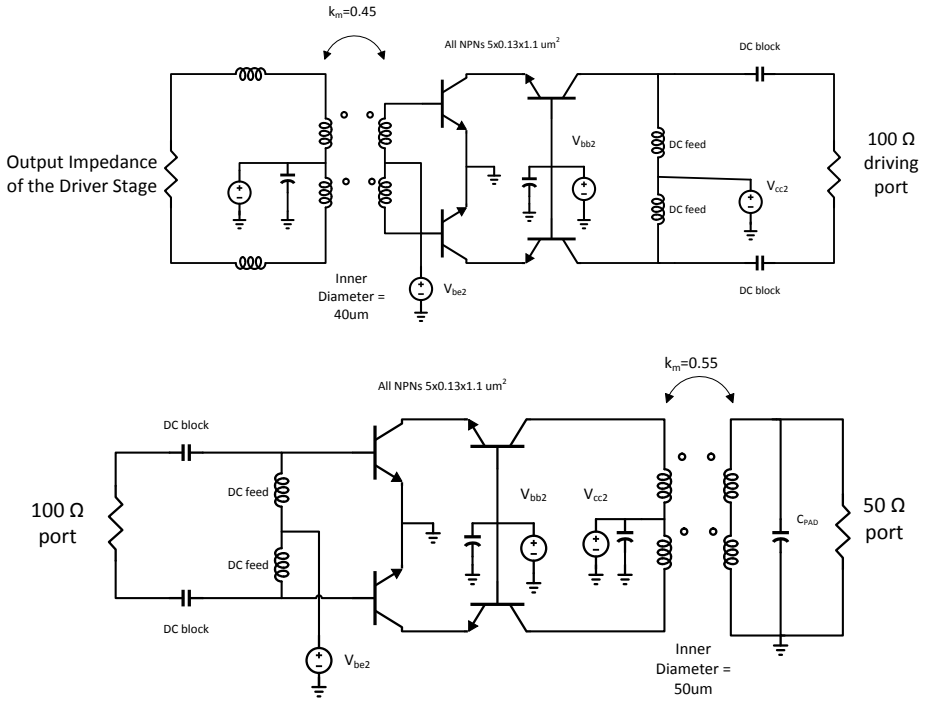


Figure 5.21 Stability assessment of the output stage. Simulation setup for the input impedance (top figure) and output impedance (bottom figure).

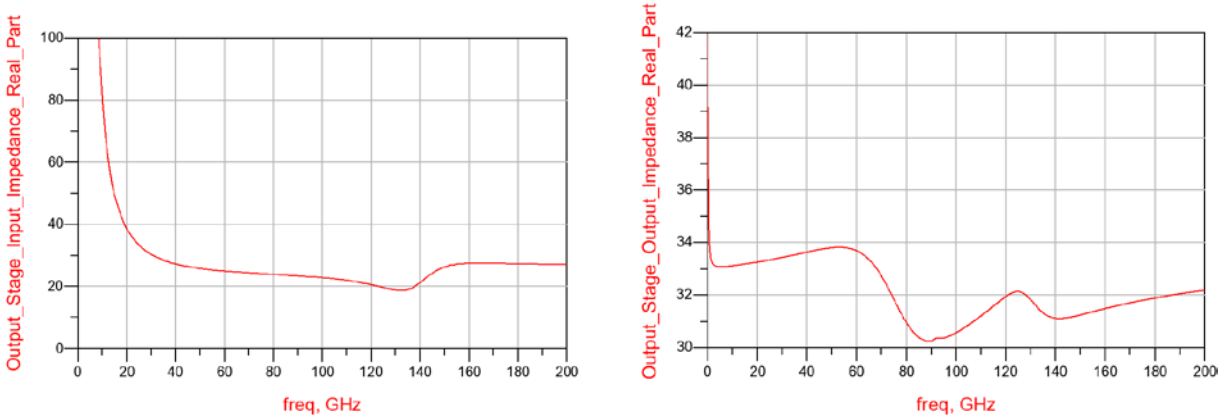


Figure 5.22 Stability assessment of the output stage. Simulation of the input impedance real part (left figure) and output impedance real part (right figure).

The stability setups for the input stage are shown in figure 5.23. They are set based on the same arguments as in the case of the output stage. The results are shown in figure 5.24.

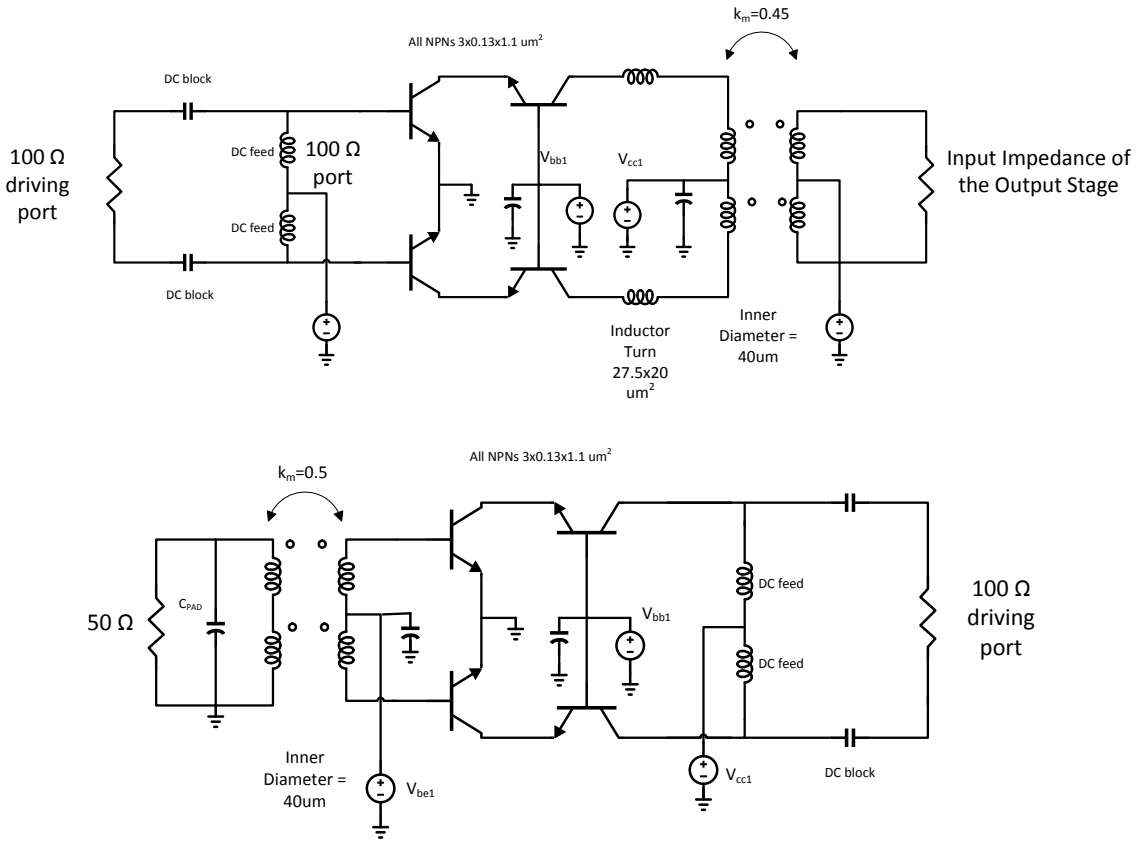


Figure 5.23 Stability assessment of the driver stage. Simulation setup for the input impedance (top figure) and output impedance (bottom figure).

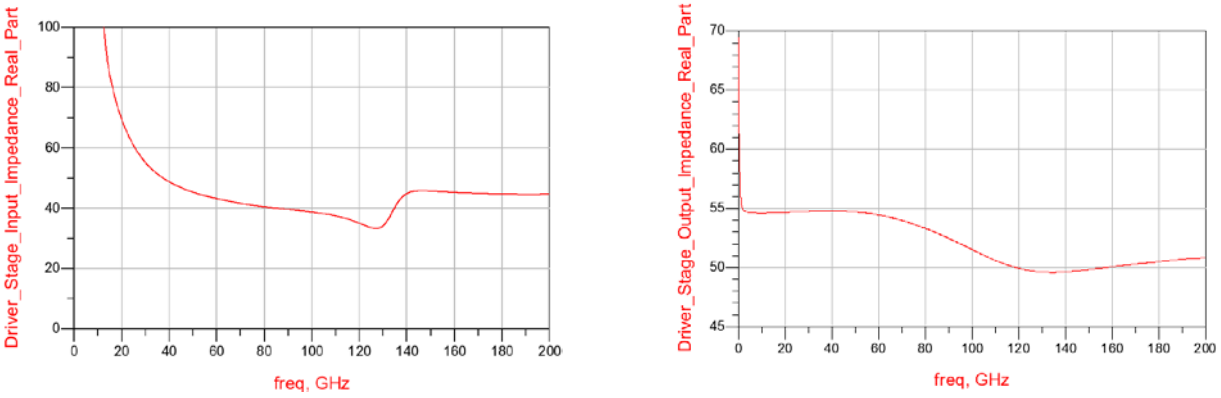


Figure 5.24 Stability assessment of the driver stage. Simulation results for the input impedance real part (left figure) and output impedance real part (right figure).

The differential stability section concludes with the simulation of the whole amplifier. The schematic of the total amplifier is shown in figure 5.13. The Mu factor is used here to assess the stability of the total circuit. The stability simulation results are shown in figure 5.25. For all frequencies, Mu is higher than one and hence no potential instability problem is apparent.

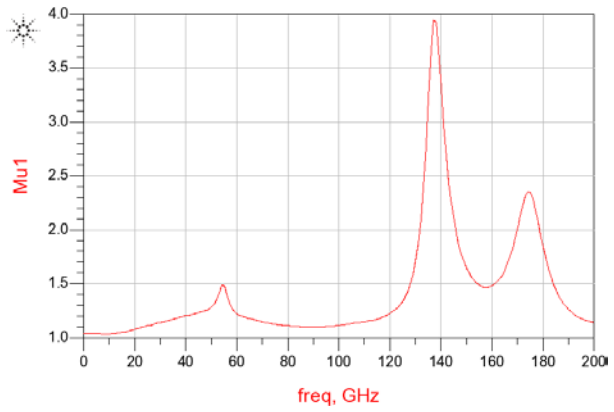


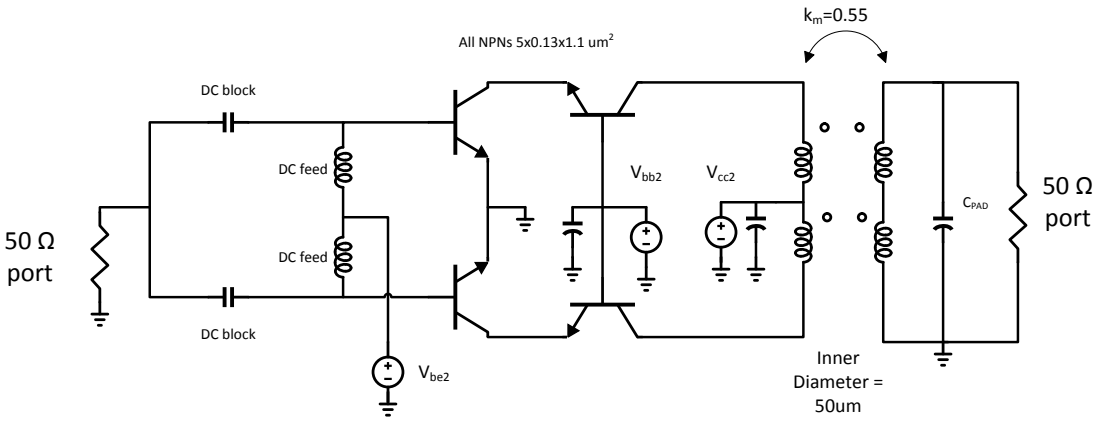
Figure 5.25 Stability assessment of the total amplifier. The Mu factor is used to quantify the stability.

5.9.2 Common Mode Stability

When using a differential circuit, the common mode stability should be also checked. In the case where the circuit is potentially unstable when excited in a common mode fashion, there is a common mode path having enough gain. This is undesired and ways should be sought in order for the gain of the common mode path to be suppressed. This is necessary because in the case of parasitically coupled common mode signals, common mode signal growth and common mode oscillations might occur. As we will see in this chapter, proper design of decoupling capacitor Q can help the common mode stability.

In order to assess the common mode stability of the circuit, the circuit should be excited in a common mode fashion. The stability setups for the individual stages are shown in this section.

Starting from the output stage, the common mode stability setup is shown in figure 5.26. The results are given in figure 5.27.



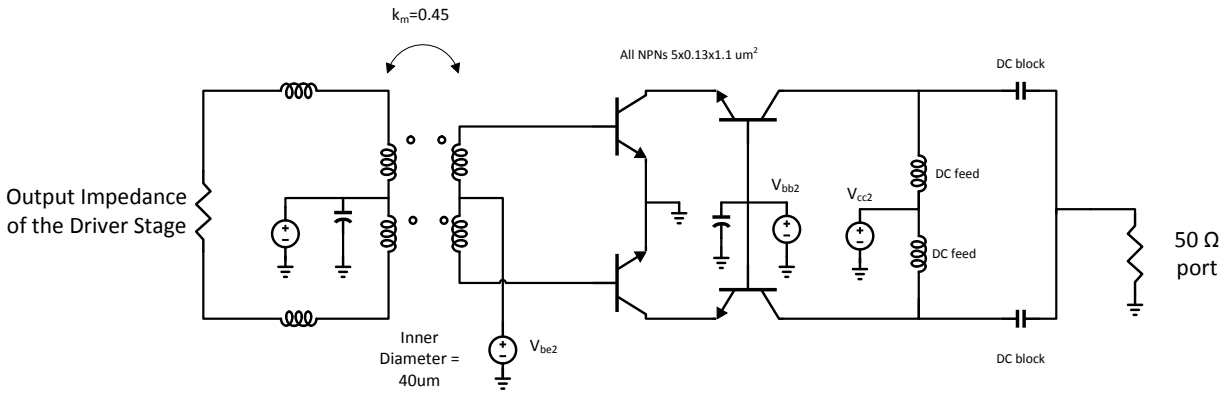


Figure 5.26 Common Mode Stability assessments of the output stage of the power amplifier; simulation setups for determining the input common mode impedance (top) and the output common mode impedance (bottom).

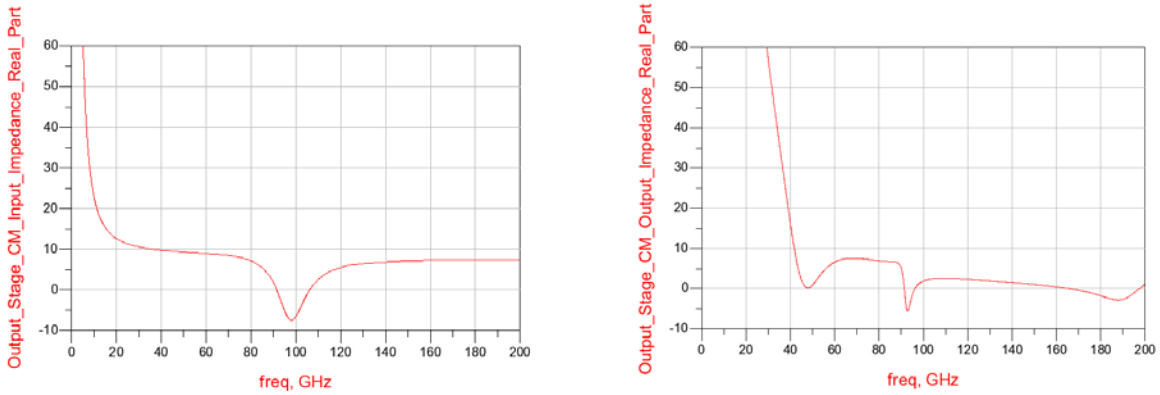


Figure 5.27 Common Mode stability simulation results for the output stage; simulation results for the real part of the input impedance (left) and the real part of the output impedance (right).

We note some instability across the frequency range between 90 GHz and 100 GHz. This is non-acceptable but it can be handled by the parasitic resistance of the decoupling capacitors. These are included in the common mode feedback loop, which closes via the ground plane, and reduce the gain of the loop due to their dissipative nature. The situation is schematically shown in figure 5.28. The common mode stability test is repeated in figure 5.29 after the decoupling capacitors and their parasitic resistances are included. The parasitic resistance of the capacitors elevates the resistance levels of the loop making the negative resistance points positive and thus ensuring unconditional stability.

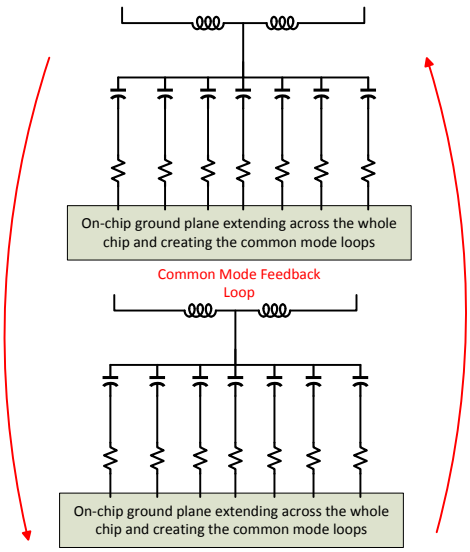


Figure 5.28 Common Mode stability simulation results for the output stage; simulation results for the real part of the input impedance (left) and the real part of the output impedance (right).

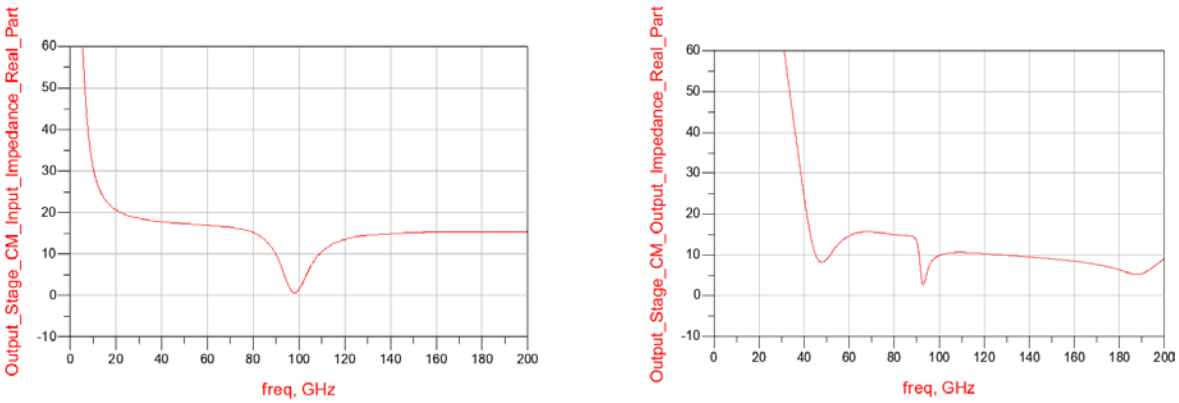


Figure 5.29 Common Mode stability simulation results for the output stage after the inclusion of the decoupling capacitance network; simulation results for the real part of the input impedance (left) and the real part of the output impedance (right).

We perform the same simulations for the driver stage of the PA. The simulation setups and the simulation results are shown in figure 5.30 and 5.31 respectively. We note again that the real part of the output impedance of the driver stage becomes negative around 90 GHz. However, after the addition of the decoupling capacitors in the simulation, the stage is unconditionally stable under common mode excitations. The common mode stability including the decoupling capacitance is shown in figure 5.32.

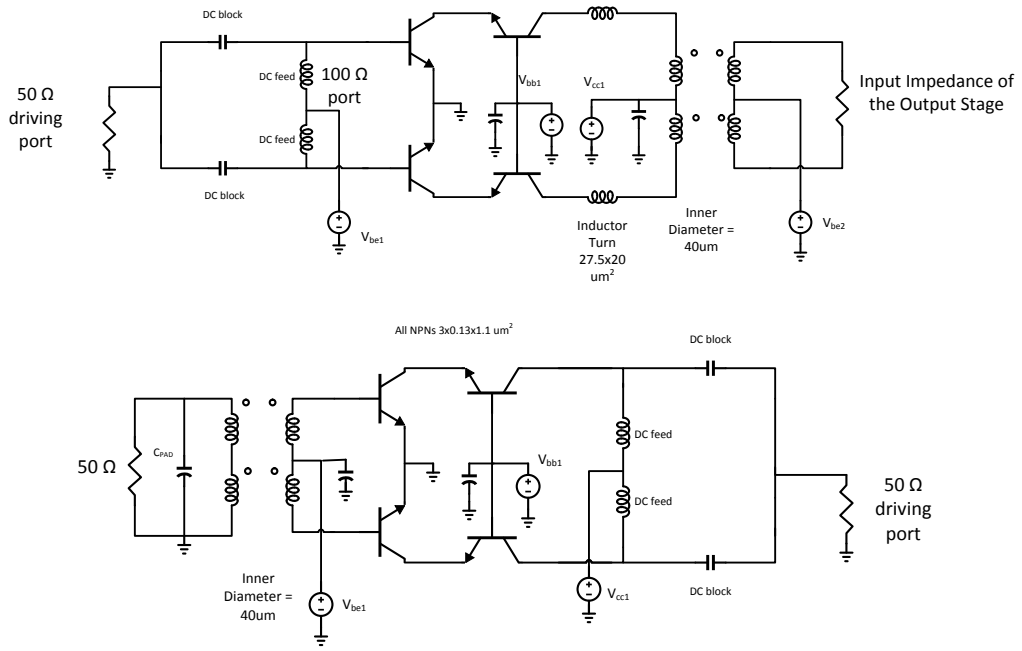


Figure 5.30 Common Mode Stability assessments of the driver stage of the power amplifier; simulation setups for determining the input common mode impedance (top) and the output common mode impedance (bottom).

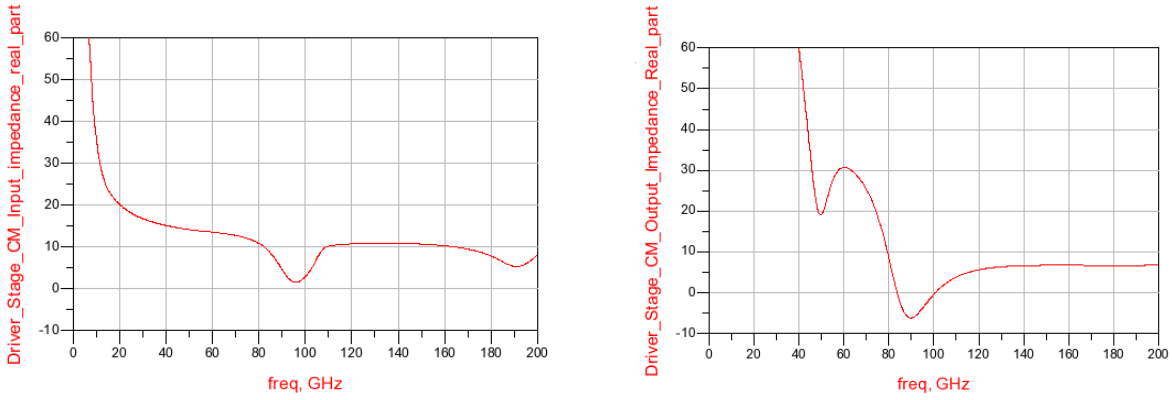


Figure 5.31 Common Mode stability simulation results for the output stage after the inclusion of the decoupling capacitance network; simulation results for the real part of the input impedance (left) and the real part of the output impedance (right).

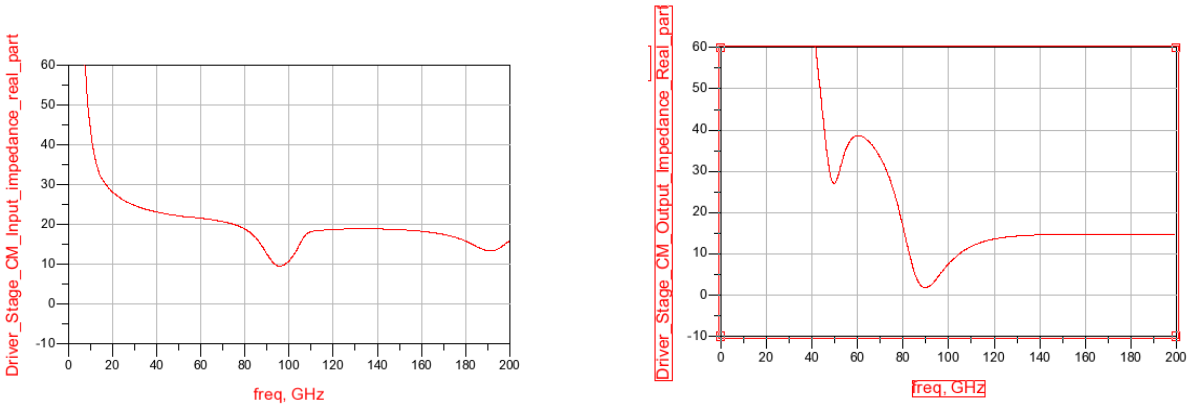


Figure 5.32 Common Mode stability simulation results for the output stage after the inclusion of the decoupling capacitance network; simulation results for the real part of the input impedance (left) and the real part of the output impedance (right).

5.10. Small Signal Performance

The amplifier small signal simulation results are going to be presented in this section. In previous sections, we found that the maximum available gain of one cascode stage, after layout parasitics were found to be ~17dB. The losses of the passives, i.e. reflective and intrinsic losses, are almost equal to that amount. This verifies the requirement of a second stage to ensure sufficient power gain. The schematic used for the calculation of the small signal power gain is repeated here for completeness. The small signal gain calculation can be quantified via the parameter S (2, 1) since both ports of the amplifier are terminated to 50 Ohms ports. The gain across all frequencies from DC up to 200 GHz is shown in figure 5.34. In the same figure, we include the stability factors Mu and MuPrime. Finally the S (1, 1) and S (2, 2) parameters are included. In figure 5.35, we zoom into the band of interest and see the exact gain behavior there.

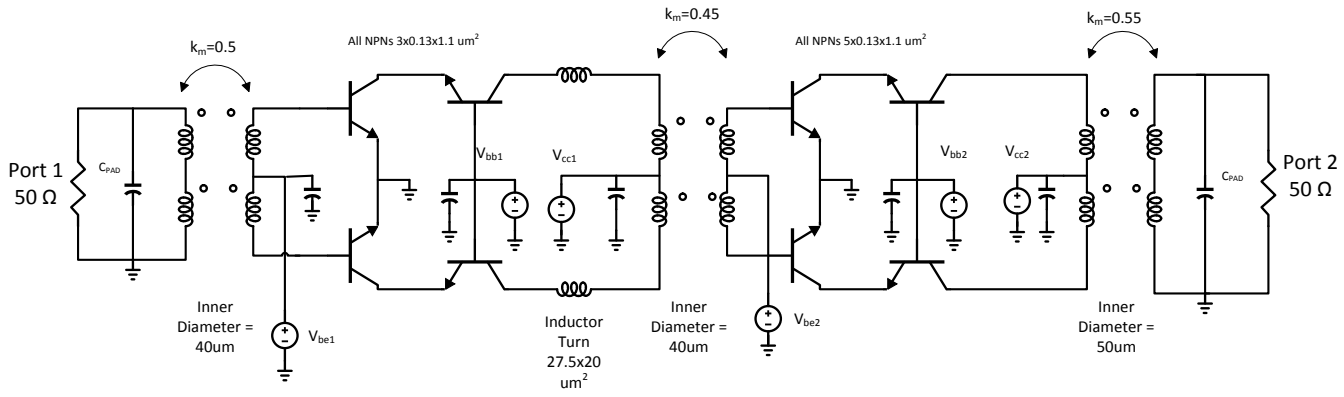


Figure 5.33 Final Power Amplifier Schematic

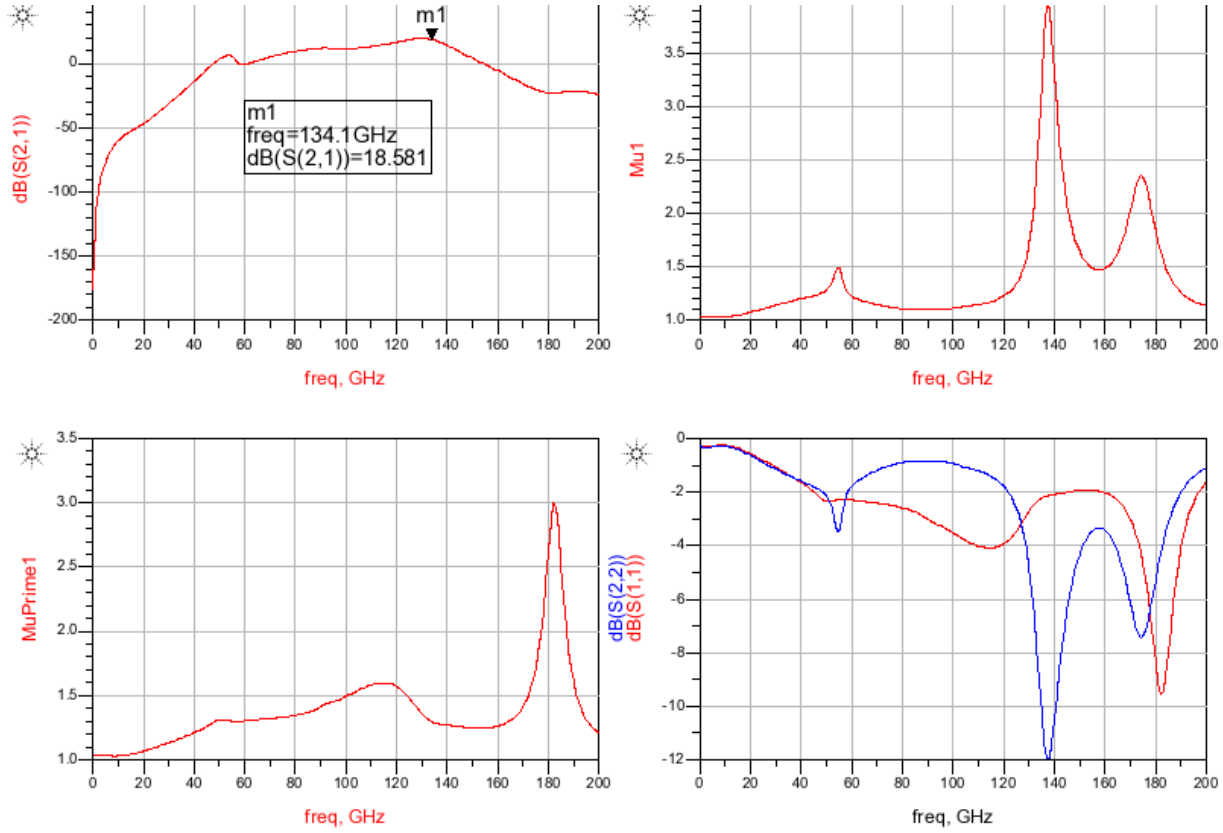


Figure 5.34 Final Power Amplifier Small Signal behavior.

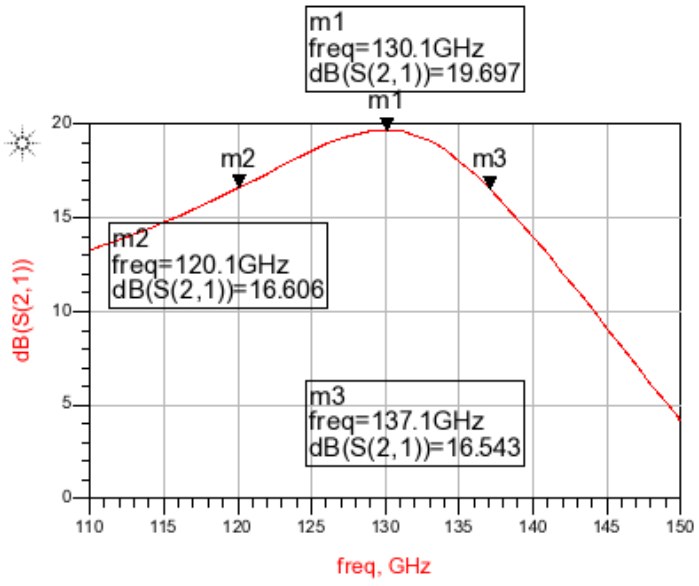


Figure 5.35 Final Power Amplifier Small Signal gain behavior.

From figure 5.35, we see the gain attaining its peak value at 130 GHz instead of 135 GHz. This is expected since at the PA frequency of operation, the device is matched for output power and not conjugate matched in order for the power gain to be maximized. However, the gain at 135 GHz is only ~1 dB lower. From the same figure we can also see the -3dB bandwidth of the PA, which is around 17 GHz, i.e. a 13% bandwidth.

5.11. Large signal performance

Finally, the large signal performance of the amplifier should be assessed. Harmonic Balance simulations in the Keysight ADS environment are performed in order for the large signal performance to be determined. The simulation setup used for the simulation of the large signal performance is given in figure 5.36. A power source P_{in} excites the circuit with a variable amplitude 135GHz sinusoidal signal. Power meters are put at the input and output to measure the input and output RF power respectively. Furthermore, the DC power consumption of the active stages has to be calculated. DC power is drawn only from the collector DC supplies. The rest of the DC supplies do not contribute significantly to the DC power consumption because they are connected to the bases of the transistors. Base currents are in the orders of micro-amps and can be considered negligible in the overall power consumption. The large signal simulation results are shown in figure 5.37. The horizontal axis represents the power that is input to the PA as determined by the input power meter and not the source available power.

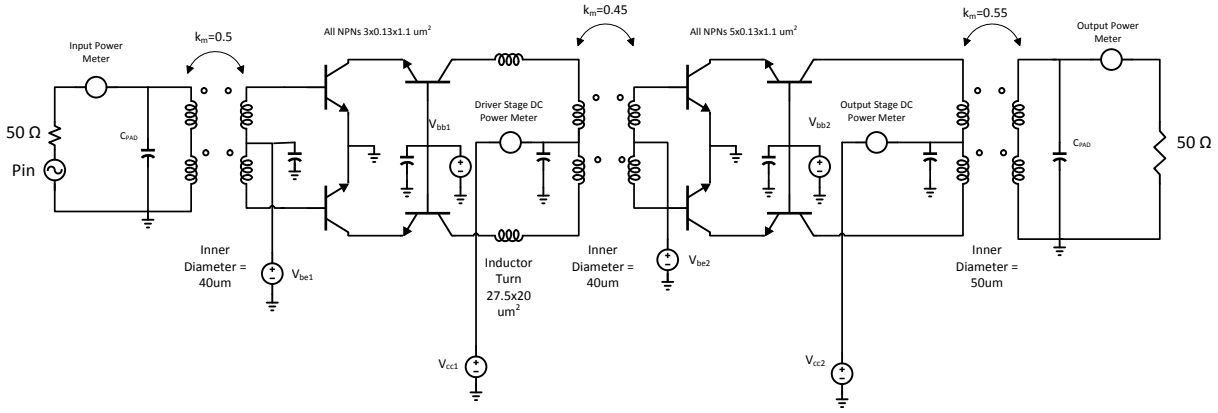


Figure 5.36 Final PA Large Signal Simulation Setup.

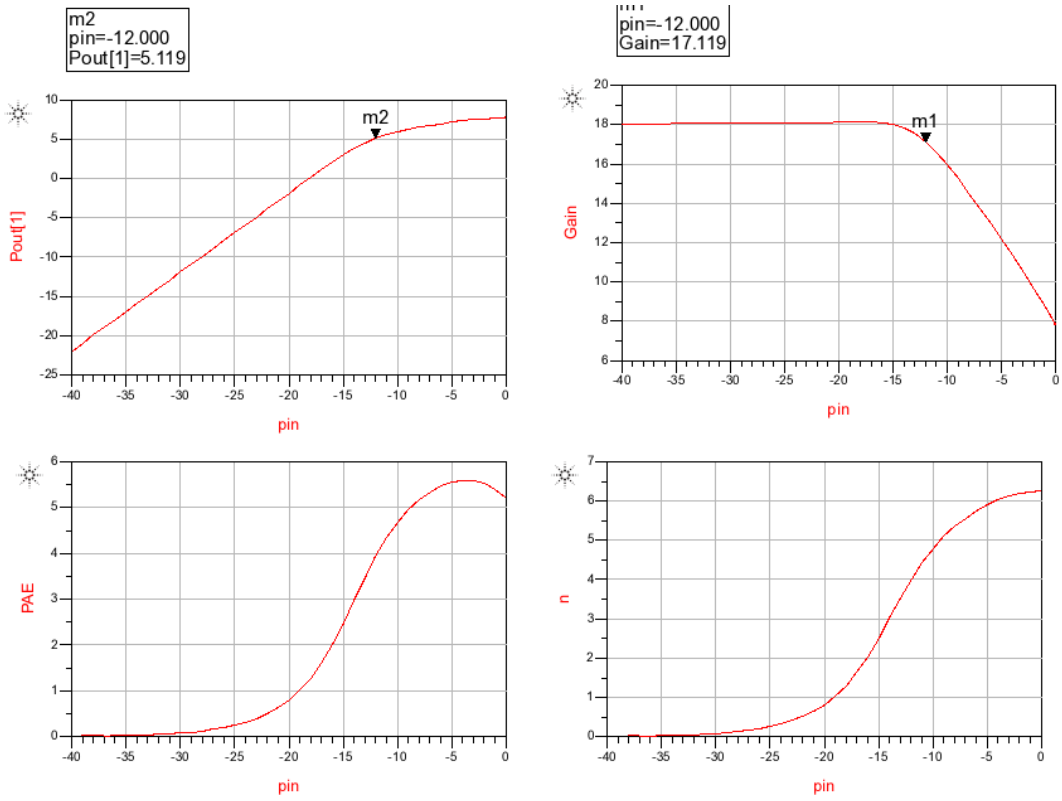


Figure 5.37 Final PA Large Signal Simulation Results. Output power (top left), power gain (top right), Power Added Efficiency (bottom left) and efficiency (bottom right).

From the Output Power graph, i.e. top left, we can see the output power versus the power that is input to the PA. Of particular interest is the output power at the -1dB compression point, P_{-1dB}, which is found to be 5dBm. The saturated output power P_{SAT} assumes a value of 8dBm. The power gain has the value of the small signal gain for low input power levels at 135GHz. The PAE efficiency peaks at 5.5% for an input power level of 6 dB. At -1dB compression point, the PAE value is close to 4%. The efficiency of the total PA is ~1% higher than the PAE. Nevertheless, the difference is not that high due to the fact that the power gain is high enough, with the value being approximately 10 dB.

5.12 Comparison with the state of the art

In this section, the performance of the PA is compared to the amplifiers presented in Table 2.1. The output power is comparable to the ones found in the literature. Furthermore, the power gain is comparable despite the fact that we are using 2 stages instead of 3. This stems from the careful design and optimization of the passive components and the parasitics tolerant layout of the device. The chip area is quite large due to the reason that out amplifier is of medium output power. Due to the latter, the device sizes are small and so is the associated capacitance. Large inductance is needed in order to resonate with capacitances at 135 GHz. This led to the use of large transformer diameters. This is area inefficient but also ensures lower loss operation as shown in Chapter 3.

Work/Freq	Process	f_{MAX} (GHz)	Pout (dBm)	Gain (dB)	Topology	**Area (mm^2)**
[7]/130GHz	130nm	>300	7.7	24.3	3 stage diff. cascode	0.43x0.7
[8]/135-170GHz	130nm	450	5-8	14-17	3 stage diff. cascode	1.2x0.48
[9]/150-170GHz		400	10@160GHz	20-32	3 stage diff. cascode	1.04x0.374
[10]/170GHz	130nm	260-290	0	15	5 stage single ended. 3 first stages cascodes and 2 final stages CE	0.15x0.19
This work	130nm	450	8@135GHz	17	2 stage diff. cascode with transformer matching networks	0.83x0.79

5.12. References

[1] Guillermo Gonzalez, “Microwave Transistor Amplifiers, Analysis and Design, 2nd edition”, Prentice Hall, 1997.

[2] J. R. Long, “Monolithic Transformers for Silicon RFIC design”, IEEE Journal of Solid State Circuits, vol. 35, no. 9, pp. 1368-1382, Sep. 2000.

[3] Ali M. Niknejad, “Electromagnetics for High-Speed Analog and Digital Communication Circuits”, Cambridge University Press, 2007.

[4] Sorin Voinigescu, “High Frequency Integrated Circuits”, Cambridge University Press, 2013.

[5]Yi Zhao, “High-Performance mm-wave and wideband Large-Signal Amplifiers”, PhD. Dissertation, TU Delft, 2013.

[6] Steve C. Cripps, “RF Power Amplifiers for Wireless Communications, Second Edition”, Artech House, 2006.

[7] Hou et. al. ”A D-band cascode Amplifier with 24.3 dB Gain and 7.7dBm Output Power in 0.13um SiGe BiCMOS Technology”, MWCL,2012.

[8] N. Sarmah, Heinemann and Ulrich R. Pfeiffer, “A 135-170 GHz Power Amplifier in an advanced SiGe HBT technology”, RFIC, 2013.

[9] N. Sarmah, P. Chevalier and Ulrich R. Pfeiffer, “160-GHz Power Amplifier Design in Advanced SiGe HBT Technologies with P_{SAT} in excess of 10 dBm”, IEEE Transactions on Microwave Theory and Techniques, vol. 61, pp. 939 – 947, Feb. 2013.

[10] E. Laskin et. al., “170 – GHz Transceiver with On-Chip Antennas in SiGe Technology”, in 2008 IEEE Radio Frequency Integrated Circuits Symposium, pp.637 – 640.

Chapter 6. Future Work

The integrated PA presented in Chapter 5 is going to be fabricated. Hence, the first future assignment is the measurement of the chip.

6.1 Frequency Multiplication and Power Combining

The power generation strategy followed in this thesis is called fundamental power generation. The circuitry designed is going to generate power at the frequency of the application. The higher the frequency, the smaller the amount of power that can be generated is. Another approach to the power generation frequency is frequency multiplication. This includes power generation at a certain frequency and then using a frequency doubler or tripler in order to multiply this frequency by two or three respectively. In this way, we can achieve output power at frequencies above 300GHz.

The loss of the doubler at frequencies above 10GHz will not be negligible [1] hence power combining can be used in order to achieve an even higher output power [2].

The steps following this work, are to feed the output of the PA of a doubler in order to achieve a frequency of 260GHz and the use of a novel waveguide power combiner in order to achieve sufficient power generation at this frequency.

6.2 References

[1] George D. Vendelin, Anthony M. Pavio and Ulrich L. Rhode, “Microwave Circuit Design using Linear and Nonlinear Techniques, 2nd edition”, Wiley Inter-Science, 2005.

[2] Ali M. Niknejad, “Electromagnetics for High-Speed Analog and Digital Communication Circuits”, Cambridge University Press, 2007.