UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM MICROELETRÔNICA

ARTHUR LIRANETO TORRES COSTA

Inductorless Balun Low-Noise Amplifier (LNA) for RF Wideband Application to IEEE 802.22

Thesis presented in partial fulfillment of the requirements for the degree of Master of Microeletronics

Advisor: Prof. Dr. Sergio Bampi
Coadvisor: Prof. Dr. Hamilton Duarte Klimach

Porto Alegre
July 2014

*"If I have seen farther than others,
it is because I stood on the shoulders of giants."*
— SIR ISAAC NEWTON

ACKNOWLEDGEMENTS

ABSTRACT

A new 50 MHz - 1 GHz low-noise amplifier circuit with high linearity for IEEE 802.22 wireless regional area network (WRAN) is presented. It was implemented without any inductor and offers a differential output for balun use. Noise cancelling and linearity boosting techniques were used to improve the amplifier performance in a way they can be separately optimized. Linearity was improved using diode-connected transistors. The amplifier was implemented in a 130 nm CMOS process in a compact 136 μm x 71 μm area. Simulations are presented for post-layout schematics for two classes of design: one for best linearity, another for best noise figure (NF). When optimized for best linearity, simulation results achieve a voltage gain > 23.7 dB (power gain > 19.1 dB), a NF < 3.6 dB over the entire band (with 2.4 dB min figure), an input third-order intercept point (IIP3) > 3.3 dBm (7.6 dBm max.) and an input power reflection coefficient S_{11} < -16 dB. When optimized for best NF, it achieves a voltage gain > 24.7 dB (power gain > 19.8 dB), a NF < 2 dB over the entire band, an IIP3 > -0.3 dBm and an S_{11} < -11 dB. Monte Carlo simulation results confirm low sensitivity to process variations. Also a low sensitivity to temperature within the range -55 to 125°C was observed for Gain, NF and S_{11}. Power consumption is 17.6 mA under a 1.2 V supply.

Um Amplificador de Baixo Ruído Banda Larga, sem Indutor, com alta linearidade e 24 dB de Ganho para a banda do padrão IEEE 802.22

RESUMO

Um novo circuito amplificador de 50 MHz - 1 GHz com alta linearidade para o padrão IEEE 802.22 "wireless regional area" (WRAN) é apresentado. Ele foi implementado sem nenhum indutor e oferece uma saída diferencial para ser utilizada como balun. Técnicas de cancelamento de ruído e aumento de linearidade foram usadas para melhorar a performace do amplificador de modo que eles pudessem ser otimizados separadamente. A linearidade foi melhorada utilizando transistores conectados como diodo. O amplificador foi implementado em um processo CMOS 130 nm, em uma área compacta de 136 μm x 71 μm. As simulações são apresentadas para esquemáticos pós-leiaute para duas classes diferentes de projeto: um visando a melhor linearidade e o outro a melhor Figura de Ruído (FR). Quando otimizado para melhor linearidade, os resultados de simulação atingem um ganho de tensão > 23.7 dB (ganho de potência > 19.1 dB), uma figura de ruído < 3.6 dB na banda inteira (com 2.4 dB min), um ponto de intersecção de terceira ordem (IIP3) > 3.3 dBm (7.6 dBm max) e um coeficiente de reflexão de entrada S_{11} < -16 dB. Quando otimizado para melhor figura de ruído, ele atinge um ganho de tensão > 24.7 dB (ganho de potência > 19.8 dB), uma FR < 2 dB na banda inteira, um IIP3 > -0.3 dBm e um S_{11} < -11 dB. Resultados de simulação Monte Carlo confirmam baixa sensibilidade à variabilidade de processo. Além disso, uma baixa sensibilidade com a temperatura na faixa de -55 até 125°C foi observada para Ganho, FR e S_{11}. Consumo de potência é 17.6 mA sob fonte de alimentação de 1.2 V.

Palavras-chave: LNA, amplificador de baixo ruído, receptor, projeto de circuito integrado.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

LNA   Low Noise Amplifier

Balun   Balanced-Unbalanced

CR   Cognitive Radio

RF   Radio Frequency

WRAN   Wireless Regional Area Network

IIP2   Second order input referred interception point

IIP3   Third order input referred interception point

IM2   Second order Intermodulation

IM3   Third order Intermodulation

SNR   Signal-to-Noise-Ratio

dB   decibel

dBm   decibel-miliwatts

NF   Noise Figure

DC   Direct Current

AC   Alternate Current

MHz   MegaHertz

GHz   GigaHertz

BW   BandWidth

WBLNA Wide Band Low Noise Amplifier

ESD   ElectroStatic Discharge

CS   Common-Source

CG   Common-Gate

CD   Common-Drain

I/O   Input-Output

FET   Field Effect Transistor

MOSFET Metal-oxide-semiconductor Field Effect Transistor

CMOS	Complementary metal-oxide-semiconductor
PMOS	P-channel MOSFET
NMOS	N-channel MOSFET
NFET	N-channel FET
PFET	P-channel FET
PCB	Printed-Circuit Board
KCL	Kirchoff's Current Law
PDK	Process Design Kit
IEEE	Institute of Electrical and Electronics Engineers
IBM	International Business Machines Corporation
ICC	IBM Customer Connect
STI	Shallow Trench Isolation
CMP	Chemical Mechanical Polish
MIM	Metal-Insulator-Metal
DRC	Design Rules Check
LVS	Layout-vs-Schematic
VCO	Voltage-Controlled Oscillator
Amp	Amplifier
EMI	Electro-Magnetic Interference
SD	Skin Depth
DUT	Device Under Test

CONTENTS

1 INTRODUCTION

The congestion in pre-allocated parts of the frequency spectrum continues to rise as more users access wireless networks. Cognitive radios (CRs) offer an approach to alleviating the congestion: they continually sense the spectrum and detect and utilize unoccupied channels MITOLA; MAGUIRE G.Q. (1999), HAYKIN (2005). This concept of CR is illustrated in Fig. 1.1. Current efforts in CR design have focused on the TV bands below 1 GHz PARK et al. (2009), KIM et al. (2009), due to the release of the IEEE 802.22 Wireless Regional Area Network (WRAN) standard. This standard focus on the implementation of a CR in the 54 MHz to 862 MHz band.
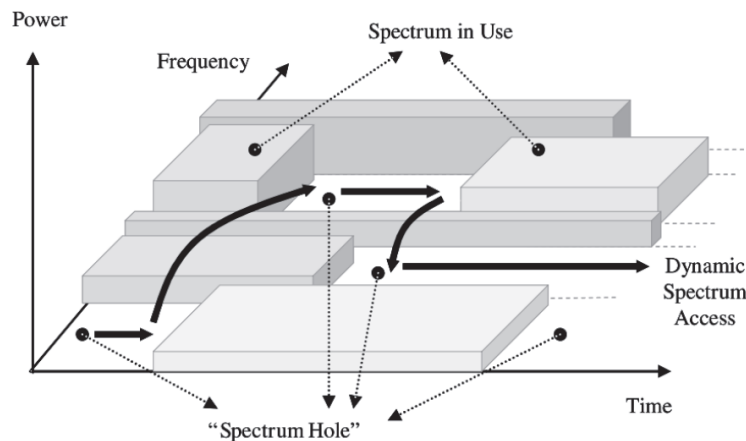
Figure 1.1: Cognitive Radio concept illustration. Picture taken from AKYILDIZ et al. (2006)

As CRs do not target narrowband standards and their allocated narrowbands, it must operate at any frequency in the entire wideband range. Also, CRs are not designed with a priori knowledge of the interfering frequency bands, thus, having to tolerate interferers at any frequency within the wideband (denoted herein by BW_{CR}). Consequently, the mixing spurs and performance parameters such as the third and second order intercept points (IP3 and IP2, respectively), which characterizes the linearity of the circuit due to intermodulation effects, must satisfy more stringent bounds RAZAVI (2009). Refer to appendix A for a detailed description of IP3. Although IP3 is usually the most important linearity parameter in narrowband receivers, in a wideband receiver the IP2 becomes at least as important as the IP3. Fig. 1.2 illustrates the IP2 problem in wideband receivers.

In order to understand the linearity problem, consider the effect of even-order distortion in the signal path in direct conversion narrowband. As illustrated in Fig. 1.2(a), two
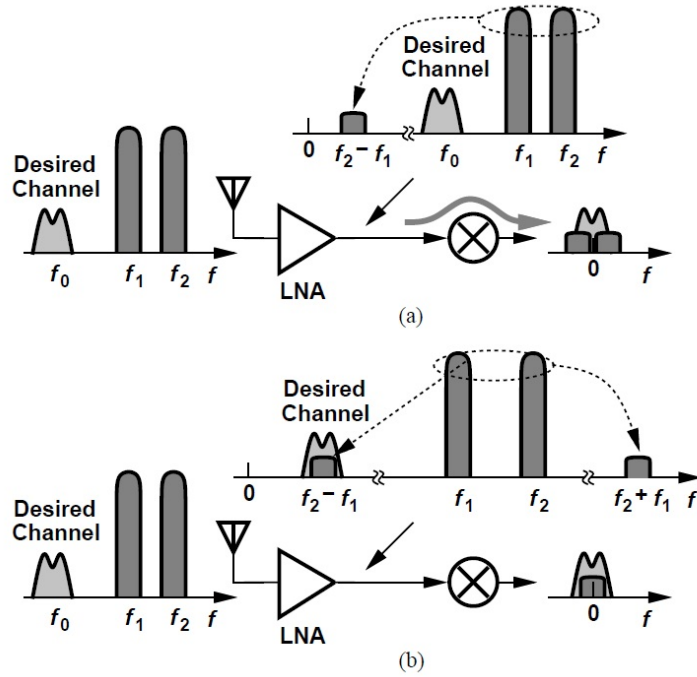
Figure 1.2: (a) Narrowband receiver gets 2nd order intermodulation (IM2) out of band. (b) IM2 non-linearity falls within the wideband of the LNA. Picture taken from RAZAVI (2010)

interferers at f_1 and f_2 generate a beat at $f_2 - f_1$ as they experience even-order distortion in the low-noise amplifier (LNA) and the input stage of the mixer. Due to low isolation, a fraction of this beat leaks to the baseband without frequency translation, corrupting the downconverted signal. In this scenario, only the mixer limits the performance because high-pass characteristic of the LNA output can remove its low-frequency beats. Indeed, the IP2 of most receivers is measured according to this scenario.

The problem of even-order non-linearity assumes new dimensions in cognitive radios. As shown in Fig. 1.2(b), the LNA itself produces components at $f_2 + f_1$ and $f_2 - f_1$, both of which may lie within the wide bandwidth of the cognitive radio (BW_{CR}). That is, the LNA becomes the bottleneck. Differential topologies alleviate this issue considerably, but it is extremely difficult to design low-loss baluns having a wide bandwidth.

A CR receiver front-end must provide a relatively flat gain and a reasonable input return loss across BW_{CR}, putting challenging demands on the radio and its LNA. Refer to appendix B to get a detailed explanation about the input return loss coefficient S_{11}. The wanted frequency span can be reduced to smaller bands, which then can be processed by several dedicated, possibly tuned, LNA circuits. The other extreme is a single LNA, which then obviously needs to have wide bandwidth. In contrast to a multi-LNA solution, the single wideband LNA is flexible and efficient in terms of area, power and costs (Fig. 1.3 and 1.4 illustrates tranceivers following each of these ideas).

The noise figure (NF) of an RF block is the difference in dB of the signal-to-noise-ratio (SNR) at the input and the SNR at the output, which means it is an SNR input-to-output degradation figure. When in the receiver chain, the overall NF is given by the Friis' equation:
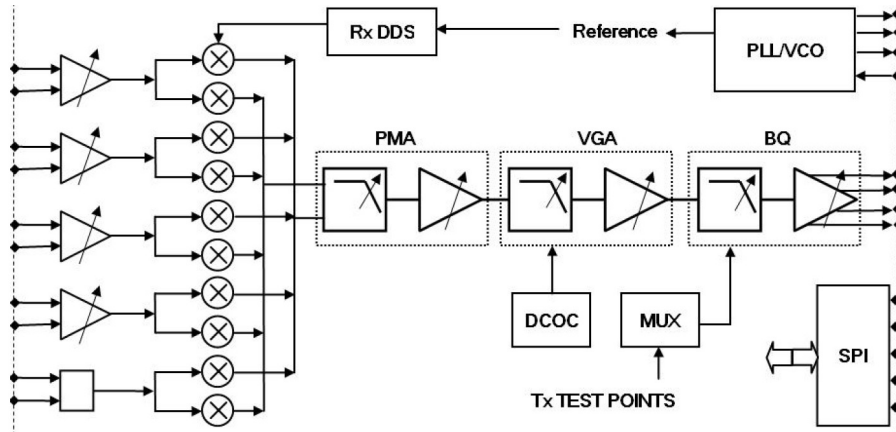
Figure 1.3: Multi-standard tranceiver. Picture taken from CAFARO et al. (2007).
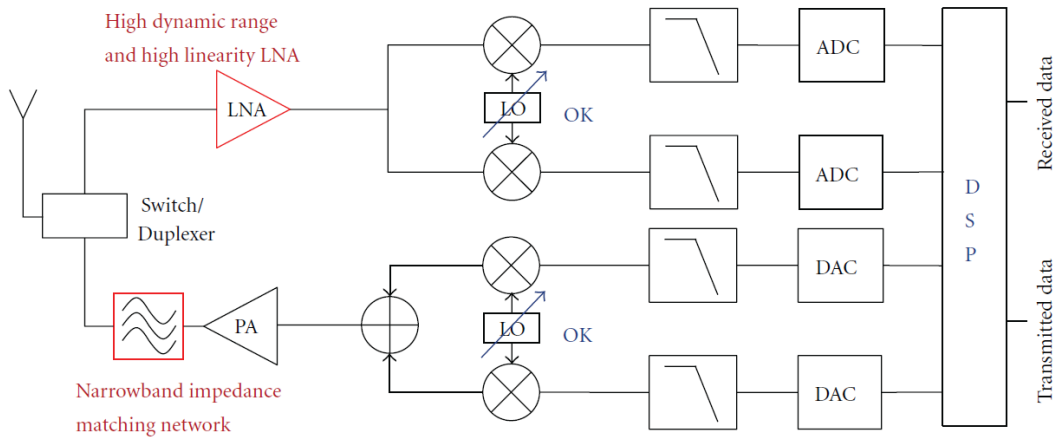
Figure 1.4: Wideband tranceiver. Picture taken from NGUYEN; VILLAIN (2012).

$$F_{total} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + ... + \frac{F_n - 1}{G_1 G_2 ... G_{n-1}} \qquad (1.1)$$

Where F_{total} is called the noise factor, which is simply the NF measured in magnitude instead of in dB, and G are the gains of each block. According to the Friis' equation, the NF of the first block in the cascade, i.e., the LNA, has the highest impact on the overall receiver NF.

In CRs, the NF of the receiver must be as low as 4-6 dB, which is difficult to achieve over a wide bandwidth. However, low NF in a wideband is achieved for very high frequencies above 1 GHz, due to the thermal noise predominance (refer to appendix B for a detailed description about the noise sources considered in this work). In order to have a low NF in low frequencies such as 50 MHz, the overall gain has to be improved, not to mention the need of flicker/thermal noise optimization.

In 2004, BRUCCOLERI; KLUMPERINK; NAUTA (2004) proposed a noise canceling LNA topology to solve the problem of NF in a wideband receiver for CRs. This topology achieved NF < 3 dB and input referred IP3 near 0 dBm. The basic idea is the inclusion of an auxiliary amplifier in parallel with the main amplifier to feed the input noise forward for noise cancellation at the output. Since then, many works tried to deal either with the

linearity as in ZHANG; FAN; SINENCIO (2009), IM; NAM; LEE (2010), CHENG et al. (2012) or the NF limitationsYU; YANG; CHEN (2010), YOUSSEF; ISMAIL; HASLETT (2010), XIMENES; SWART (2011), with some improvement on the overall performance RAZAVI (2010), BLAAKMEER et al. (2008), MOEZZI; BAKHTIAR (2012), KUO; KUO; CHUEH (2009), IM (2013), LIN et al. (2014).

As the traditional noise canceling topologies (either common-source (CS) + common-gate(CG) or resistive feedback CS + CS) have a single auxiliary amplifier, its gain is completely attached to the noise gain of the main amplifier, which makes it difficult to completely decouple from the overall trade-off of NF, input matching, bandwidth, linearity, area and power. This suggests that more degrees of freedom are needed in the auxiliary amplifier in order to improve overall performance of the LNA.

Single-ended input LNAs are preferred to save I/O pins and because antennas and RF filters usually produce single ended signals. On the other hand differential signaling in the receiver chain is preferred in order to reduce second-order distortion and to reject power supply and substrate noise. Thus, at some point in the receiver chain a balun is needed to convert the single-ended RF signal into a differential signal. Off-chip baluns with low losses are typically narrowband so that several baluns would be required in case of wideband operation. On the other hand, wideband passive baluns typically have high loss, degrading the overall noise figure (NF) of a receiver significantly.

Combining the balun and LNA functionality into a single integrated circuit seems an attractive option to realize a wideband low-noise receiver front-end. The target specifications for the LNA of this work are shown in the end of the state of the art chapter.

This dissertation starts with the state of the art review, presenting recent wideband LNAs developed in the literature. This is done in Chapter 2. In Chapter 3, the topology, circuit design and design considerations at the electrical level is done. In this chapter, emphasis is given also to the layout issues of this design, pads and ESD protection and PDK details. The post-layout simulation results are presented and discussed in Chapter 4. In Chapter 5, the expected test-bench and measurements after fabrication, and finally the conclusions in Chapter 6, followed by the references and appendixes. The complementary RF theory such as harmonic distortion, intermodulation, linearity, 1 dB compression point, noise, scattering parameters, stability, required to help understand this dissertation, is included in the appendixes.

2   STATE OF THE ART OF WBLNA

In this chapter, a survey on recent published works on wideband LNAs for CR and their implementation techniques are presented. The feedback technique improves input matching, which improves Noise Figure (NF), while contributing in linearization. Noise canceling can be used to reduce NF by subtracting inverted correlated noise signals, as in CHENG et al. (2011), ANSARI; YAVARI (2011), XIMENES; SWART (2011), WANG et al. (2012), PIMENTEL; BAMPI (2012), BRUCCOLERI; KLUMPERINK; NAUTA (2004), LI; ZHANG (2007), CHEN; LIU (2012). Other linearization method is based on canceling higher order nonlinear terms of the signal of interest by adding a secondary device to subtract currents or voltages, such as LIAO; TANG; MIN (2007), ZHANG; FAN; SINENCIO (2009), CHENG et al. (2012). The key idea is to improve the linearity of a wideband LNA without degrading gain, NF and input impedance matching, and working in a 2-3 decades frequency bandwidth.

2.1   Noise canceling and Negative Feedback

The shunt-feedback technique showed in figure 2.1 is perhaps the simplest way of reducing the LNA noise figure. Neglecting the output conductance, this simple amplifier suffers from a severe trade-off of input impedance matching and noise figure. For low noise figure, it requires $g_{mi}R_i$ and $\frac{R_i}{R_s}$ much larger than 1. Which does not goes with the matching requirement $g_{mi}R_i = 1$ and $\frac{R_i}{R_s} = 1$, where R_S is the source resistance.

In order to break this trade-off, one can add amplifiers cascaded within the feedback loop. This way, one can decouple the gain and input matching, making it possible to lower NF without degrading input matching. However, this strategy might make the design prone to instability.
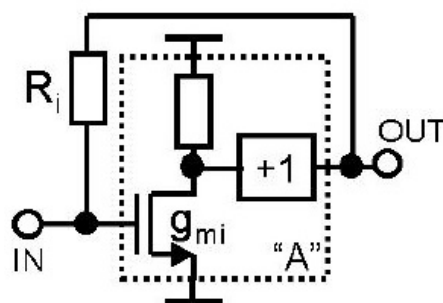
Figure 2.1: Negative feedback LNA. Picture in the right taken from BRUCCOLERI; KLUMPERINK; NAUTA (2004).

The strategy is to break this direct connection of Z_{in}, NF and gain, keeping the feedback advantages. In BRUCCOLERI; KLUMPERINK; NAUTA (2004), the feedback problems and the noise canceling technique are well described and explained. The implementation was made in CMOS 250 nm. Figure 2.2 shows the noise canceling technique for a single-ended amplifier. The idea behind noise cancellation is the addition of another amplifier, which inverts the LNA input signal and noise, generating an output signal that can be added to the direct path signal to subtract noise and enhance gain.
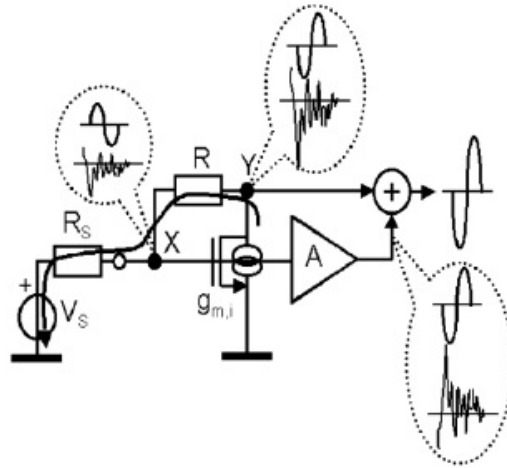
Figure 2.2: Noise canceling technique. Picture taken from BRUCCOLERI; KLUMPERINK; NAUTA (2004).

One possible problem of this technique is the noise generated by the additional amplifier. However, the noise generated from this amplifier is completely decoupled from the noise of the matching device. Thus, it can achieve very low values of NF without degrading input impedance and LNA total gain. Besides, even with mismatch problems, which can shift noise and signal phases, noise is still partially canceled. This technique can achieve a NF below 2.5 in the entire band.

The work in PIMENTEL; BAMPI (2012) uses the feedback + noise cancellation concept previously explained to implement a differential WBLNA in CMOS 130 nm process (Figure 2.3). It incorporates the differential advantages, such as common mode noise rejection and enhanced gain to accomplish the challenge of implementing the idea in CMOS 130 nm. It shows a NF around 3-5 dB.

In WANG et al. (2012), the feedback noise canceling was implemented in a different way. It included the additional canceling amplifier within a cascode WBLNA (figure 2.4). The cancellation occurs in the V_{GS} of the common-gate stage of the cascode. This design also enhances gain and, as a receiver's block, contributes to achieve a total receiver NF of 3 dB, which was mainly due to the LNA. The implementation of the G amplifier of Figure 2.4 was simply a short-circuit (G=1). The actual fabricated circuit was made differential, just mirroring the circuit below.

In XIMENES; SWART (2011), a Common-Gate (CG) differential amplifier was used along with a Common-Source (CS) to implement noise cancellation without resistive feedback. Fig. 2.5 shows the implementation, which achieved 1.4-2.5 dB NF. The input signal enters the CG and CS amplifiers, and then both output signals with different polarities are summed to cancel noise. The main advantage of this implementation is that due to cross-coupling on the input, for a differential input impedance matching, the g_m is two times smaller. Which means half its contribution to the noise factor when comparing
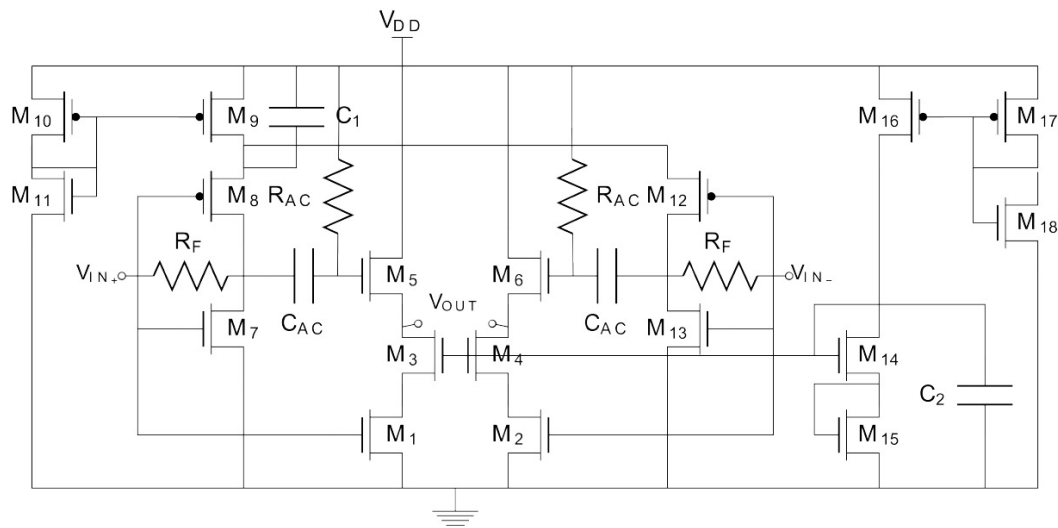
Figure 2.3: Fully differential implementation of noise cancellation technique. Picture taken from PIMENTEL; BAMPI (2012).
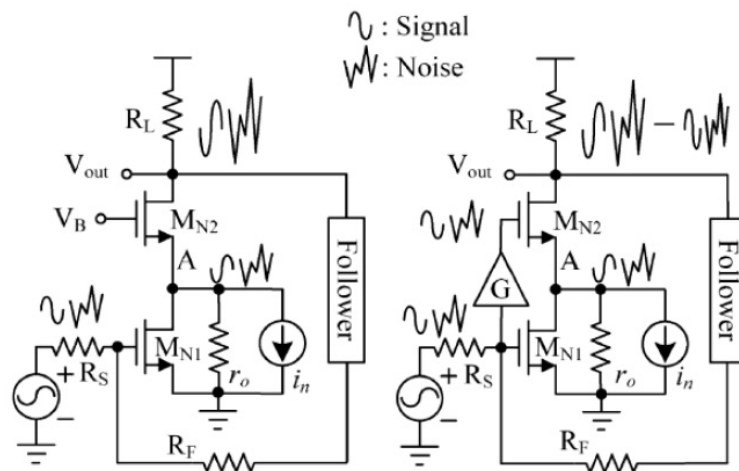
Figure 2.4: Cascode implementation of noise canceling technique. Picture taken from WANG et al. (2012).

with traditional balanced CG circuit. This means that for equal gain, the usage of cross-coupling provides a lower F while requiring half of the power. The drawback is that noise factor is limited to $1+\gamma g_{d0}/2g_m$ and antennas and RF filters are usually single-ended.

Apart from differential implementations, single-ended (such as in BRUCCOLERI; KLUMPERINK; NAUTA (2004)) might be needed, depending on the antenna and RF filter used on the WBLNA input. In ANSARI; YAVARI (2011), a single-ended WBLNA was implemented in CMOS 90 nm. It uses the same idea of XIMENES; SWART (2011), i.e., CG + CS amplifiers for noise cancellation. However, it uses resistive and inductive degeneration to achieve NF as low as 2.3 dB in 90 nm. The circuit is shown in Figure 2.6. The disadvantage is the use of several inductors to improve gain flatness and input impedance matching.

All of the above topologies improved NF of WBLNAs while maintaining gain and input impedance matching. As noise cancellation also cancels distortion, linearity is slightly improved. However, this improvement might not be enough to CR applications.
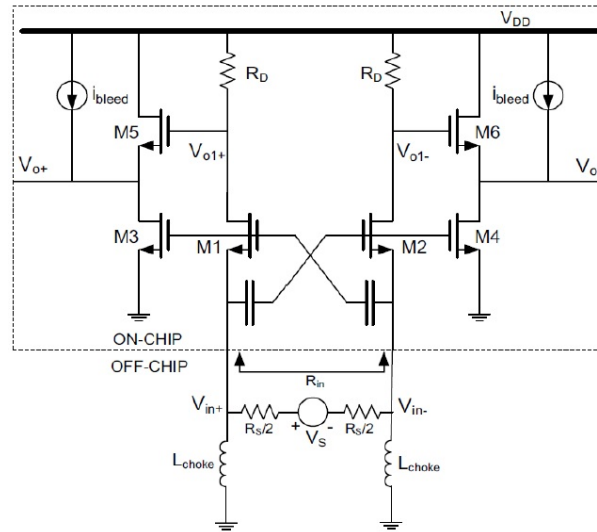
Figure 2.5: Capacitively cross-coupled implementation of noise canceling technique. Picture taken from XIMENES; SWART (2011).
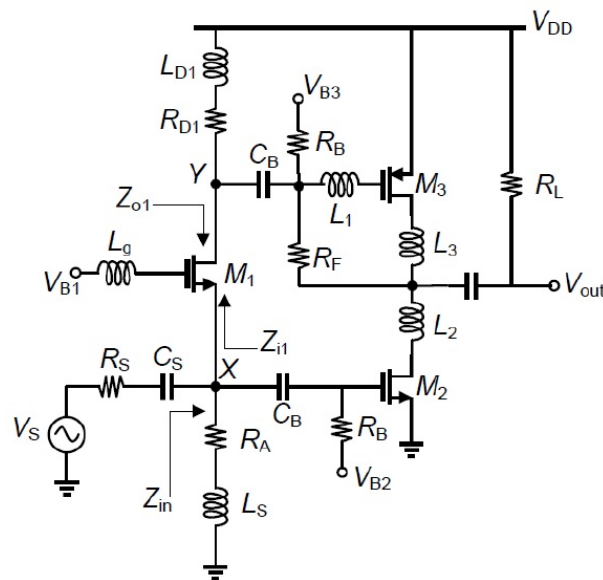
Figure 2.6: Single-ended implementation of noise canceling technique. Picture taken from ANSARI; YAVARI (2011).

The WBLNAs which focus on high linearity improvements are discussed in the linearity section.

2.2 Forward Body Bias

There are LNAs which target low voltage and also wideband operation.The Forward body bias was used, HAO et al. (2012), to improve the cascode implementation of the LNA under low voltage power supply conditions. The proposed topology works with voltages under 1V, using the body bias do decrease V_{th}. Other tricks, such as output inductor load to improve gain flatness and input capacitance to reduce input impedance in high frequencies, were also used. The topology is a simple cascode LNA with output

buffer, achieving 2.7-3.9 NF without any noise cancelation techniques. The circuit is shown in Figure 2.7.
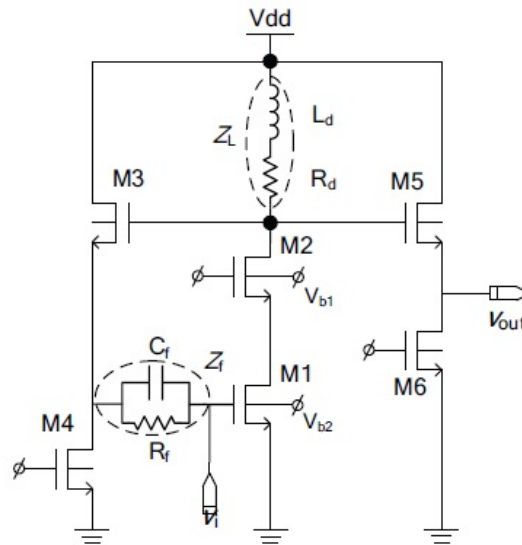
Figure 2.7: Forward body bias implementation to operate under 1V power supply. Picture taken from HAO et al. (2012).

2.3 Balun-LNA

In this section, two main topologies used to implement wideband balun LNAs are analyzed. The first one, is a CG-CS amplifier initially proposed theoretically in BRUC-COLERI; KLUMPERINK; NAUTA (2004) and them implemented by the same group in BLAAKMEER et al. (2008), shown in Fig. 2.8.

This topology is based on the fact that the CG amplifier has the same polarity for signal voltage and difference polarities for noise voltages at input and output nodes. Hence, the noise of the CG transistor can be canceled by adding an auxiliary CS amplifier to feed it forward to the output and subtract from the CG amplifier path. Besides, this topology benefits from the low input impedance of the CG amplifier, having a higher than 3 dB NF though. It achieves 15 dB gain, NF < 3.5 dB, IIP2 > 20 dBm and IIP3 > 0 dBm. The best performance is achieved in the 300 MHz - 3.5 GHz band.

The second balun topology is the work in CHENG et al. (2011), which uses the same 3 cascaded CS amplifiers with differential output of RAZAVI (2010), but with source RC degeneration in the last CS stage. This improves matching between the noise canceling nodes and linearity at the same time. Source degeneration is applied (RC parallel branch) in the third stage to lower the gain, as well as providing better matching between nodes X and Y, besides improving also the linearity. However, the source degeneration resistor R_{DEG} degrades the high frequency input matching. Thus, the degeneration capacitor C_{DEG} is added to boost up the bandwidth of the balun-LNA, helping as well th input impedance matching at high frequency. The output signal is obtained from the pseudo-differential sensing, i.e., subtracting v_Y from v_X. Based on such a pseudo-differential sensing the even-order linearity (IIP2) can be improved due to partial signal cancellation. However, the non-linearity cancellation relies on phase and gain matching nodes X and Y. To minimize the difference between X and Y, the gain of the third stage is therefore
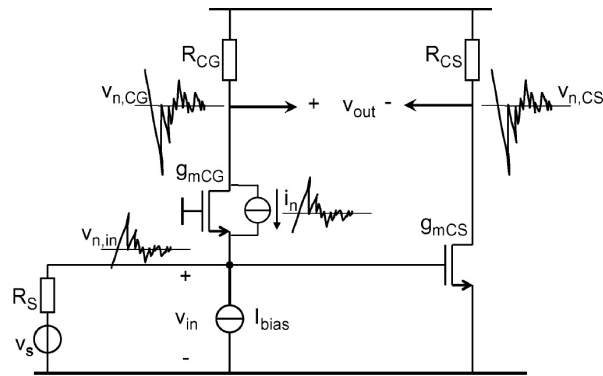
Figure 2.8: Balun-LNA with CG-CS amplifiers. Picture taken from BLAAKMEER et al. (2008).

sized to be close to 0 dB. The circuit of CHENG et al. (2011) is showed in Figure 2.9. It achieves 2.7-3.6 dB NF in 65 nm. In RAZAVI (2010) the third stage of the LNA is a simple CS amplifier which provides a relatively large gain, thus leading to a degraded non-linearity cancellation.
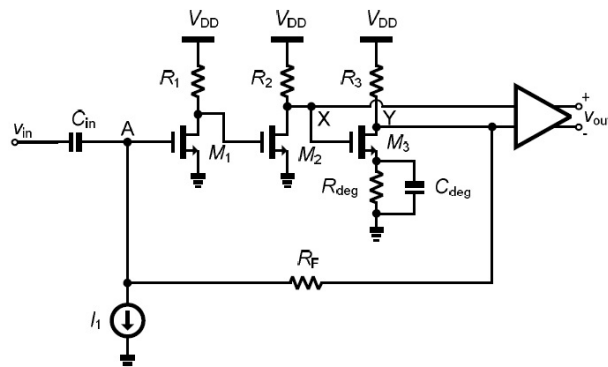
Figure 2.9: Balun-LNA with cascaded amplifiers. Picture taken from CHENG et al. (2011).

2.4 Linearization Techniques

The feedback + noise cancellation technique achieves IIP3 around 0 dBm. This is because its main purpose is to cancel noise without degrading input impedance matching and gain.

In CR, however, higher IIP3 values might be required, regarding its wide bandwidth. In order to solve this linearity problem, some new techniques have been developed. A great result is achieved in ZHANG; FAN; SINENCIO (2009), where the cascode CG WBLNA has its IIP3 improved by +11.7 to +14.1 dBm, without degrading gain and NF in CMOS 130 nm. The circuit is shown in Figure 2.10.

The CG stage is used to improve input impedance matching and the cascode is used to improve gain and bandwidth. The inductor L_d increase bandwidth and inductor L_c cancels parasitic effects of the cascode transistors. The amplifier at the right of the design is a buffer, used to interface the measurement equipment and also emulates the input
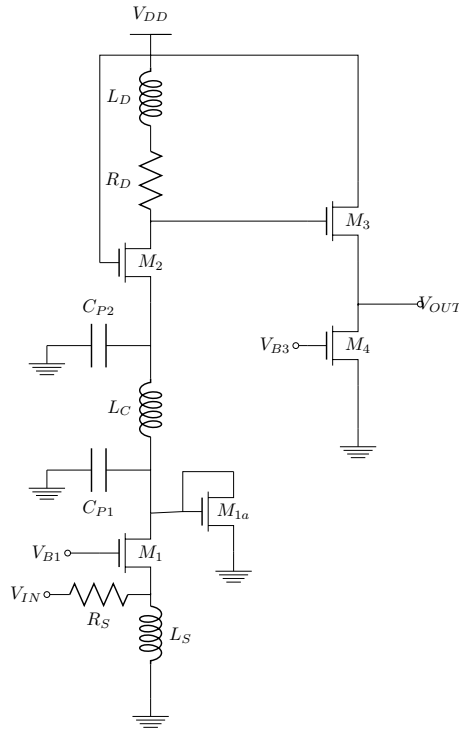
Figure 2.10: Single-ended implementation with 3rd order non-linearity cancellation. Picture taken from ZHANG; FAN; SINENCIO (2009).

impedance of a mixer. The key component of this design is the diode-connected transistor at the drain of the CG input stage. This transistor taps the drain voltage, replicating the CG current. The output current which goes to the cascode transistor is the subtraction of both currents, partially canceling second and third order nonlinear terms. Although this technique also cancels the linear term, it does not degrade gain/NF because the diode bias is much less than that of the CG transistor. The complete analysis of the non-linearity of this topology is showed in ZHANG; FAN; SINENCIO (2009) with Volterra series expansion. The linearization technique is designed for 1.5 GHz to 8.1 GHz.

The above linearity analysis considered the CG input transistor non-linearity as dominant, neglecting the output conductance and cascode transistor non-linearities. This approach is well suited to long channel devices and high supply voltages. In deep submicron technologies, the transistor output resistance is typically low, increasing the distortion of the cascode transistor. Besides, the low voltage operation might push the cascode transistor out of deep saturation, which also increases the distortion of the circuit. The work in CHENG et al. (2012) makes a rigorous analysis of the differential resistive feedback cascode WBLNA, considering the transconductor, output resistance and cascode transistor non-linearities. A method of intermodulation-product of third-order (IM3) cancellation is proposed using negative impedance connected in the middle of the cascode, as shown in Figure 2.11. The 4 transistors plus the capacitor in the middle implements the negative impedance.

The idea of the negative impedance is to generate a degree of freedom in cancelling the nonlinear terms of all the components considered. The nonlinearity symbolic analysis was made based on harmonic balance to arrive at a given expression for the current of the cascode in terms of vgs, vds and the negative impedance added. The exact expression was not the focus, what matters is the required value on the negative impedance to partially
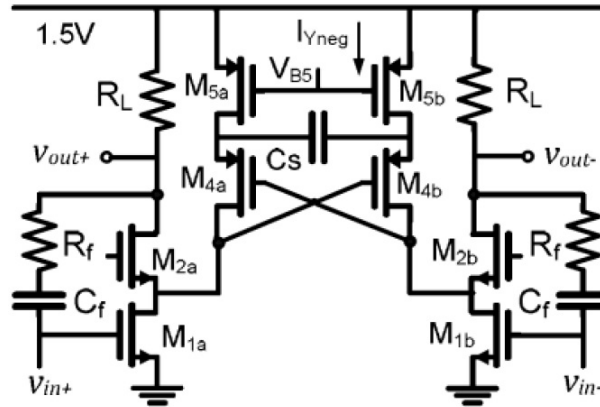
Figure 2.11: Negative impedance used to cancel terms of IM3 equation. Picture taken from CHENG et al. (2012).

or fully cancel the IM3 contributions of both the cascode transistors. The thermal noise addition of the negative impedance does not degrade the performance, since the noises of all the other components are lowered by the negative impedance effect. This topology without inductors achieves IIP3 improvements of 6.3 dB to 10 dB in 0.1 Ghz to 1 Ghz, while maintaining low NF, high gain and input impedance matching (using feedback).

2.5 Promising WBLNA Designs

Designing WBLNA for cognitive radios is very challenging due to the 2-3 decades of frequency band and high IIP3 requirement. As shown in the previous sections, many research around the world has been carried on targeting linearity and noise figure improvements.

In this chapter, a selection for the best suited WBLNA for each application focus is presented. Here the recent research in WBLNAs are grouped in the following categories: best circuit area, best bandwidth, best linearity, best power consumption, best noise figure and finally best trade-off. The best trade-off category groups several designs who attempted to get to the optimum point of all requirements.

2.5.1 Best Area Occupied

The best area occupied circuit, CUSTODIO et al. (2010), is shown in figure 2.12. This design uses the core of Bruccoleri's idea, which is the negative feedback inverter circuit, and used 2 additional inverters to implement the noise cancellation. The first auxiliar inverter is cascaded with the input core and the second one senses the LNA's input. The output is taken differentially from the additional inverters outputs.

One might notice, the cascaded inverter is needed in order to to have in phase noise and 180° out of phase signal at the output to allow noise subtraction and signal enhancement in the differential output. As the design is inductorless and uses only 3 inverters, it achieves only 153 μm x 45 μm area.

Also, this circuit has the advantage of being a balun-lna, which is well suited for single-ended antennas.
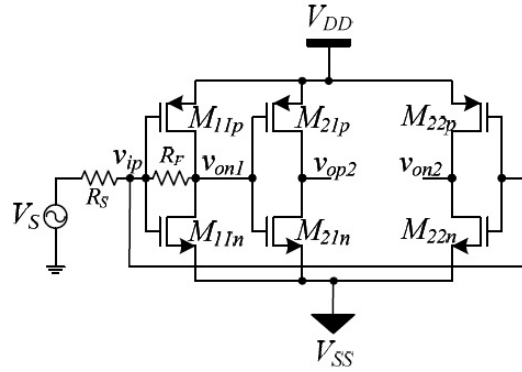
Figure 2.12: The best area occupied circuit. Picture taken from CUSTODIO et al. (2010).

Table 2.1: Best area occupied circuit.

V. Gain (dB)	Bandwidth (GHz)	NF (dB)	IIP2 (dBm)	IIP3 (dBm)	Area (mm^2)	Power (mW)	Tech (nm)
17.7	0.2-1	4-4.2	-	-9.8	0.0069	7.35	130

2.5.2 Best Bandwidth

The best bandwidth design is the one of Figure 2.9. The bandwidth here is improved by the 65 nm technology node and the cascaded amplifiers within the feedback. It can achieve bandwidth of 50 MHz - 10 GHz with very high power gain.

Table 2.2: Best bandwidth circuit.

V. Gain (dB)	Bandwidth (GHz)	NF (dB)	IIP2 (dBm)	IIP3 (dBm)	Area (mm^2)	Power (mW)	Tech (nm)
24-25	0.05-10	2.7-3.6	10.2-32.4	-10-(-2)	-	21.7	65

2.5.3 Best Linearity

The circuit which achieved the best linearity was the one in 2.10. The linearization technique using a diode can achieve IIP3 = 11.7 to 14.1 dBm.

2.5.4 Best Power Consumption

The design which achieved the best power consumption, KIHARA; MATSUOKA; TANIGUCHI (2008), was the one in Figure 2.13. This design features a single transistor and a single on chip transformer, which performs the noise canceling. Fig. 2.13 show the behavior of noise canceling using the transformer. The CG amplifier have output and input noise voltages 180º out of phase and the input and output signal in phase. In such conditions, noise can be canceled if one takes the input signal + noise, invert it and then add to the CG output. This job is implemented by the transformer coils. One coil has current going in and the other has current going out of it, which induces inverted voltages in each other. The output induced noise voltage is 180º out of phase of the output noise voltage. As these 2 voltages add in the output, noise canceling is performed.

2.5.5 Best Noise Figure

The best noise figure is achieved in the design, YOUSSEF; ISMAIL; HASLETT (2010), shown in Figures 2.14 and 2.15.

Table 2.3: Best linearity circuit.

P. Gain (dB)	Bandwidth (GHz)	NF (dB)	IIP2 (dBm)	IIP3 (dBm)	Area (mm^2)	Power (mW)	Tech (nm)
8.6-11.7	1.5-8.1	3.6-6	7.6-23	11.7-14.1	0.58	2.62(core)	130
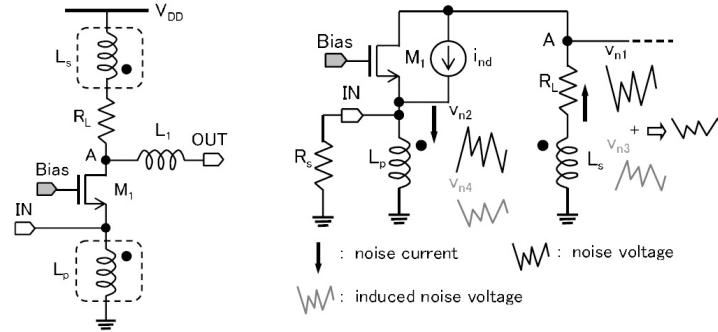
Figure 2.13: Noise canceling LNA using a transformer. Picture taken from KIHARA; MATSUOKA; TANIGUCHI (2008).
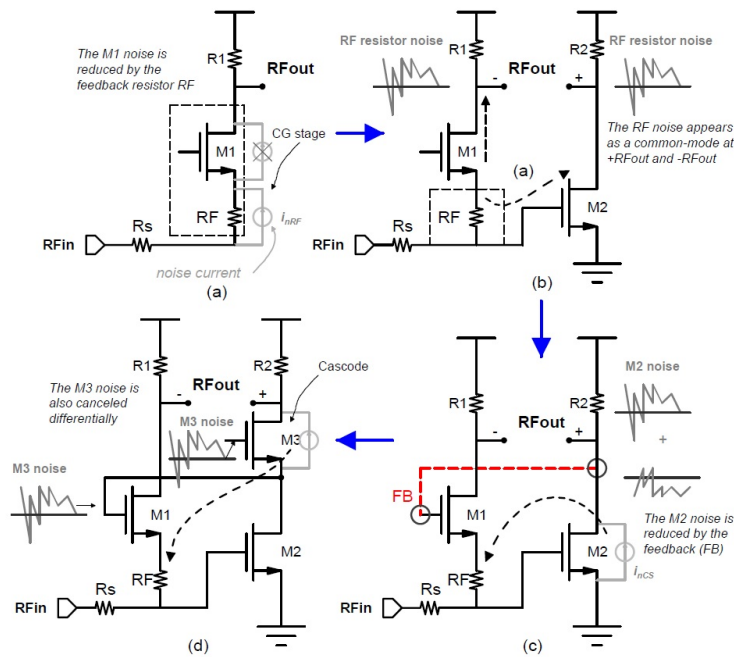
Figure 2.14: The circuit evolution with noise canceling highlighted. Picture taken from YOUSSEF; ISMAIL; HASLETT (2010).

This topology uses a resistor RF to boost IIP3. Thus, the NF of the CG stage is dominated by the noise of RF. Then, it uses CG-CS noise canceling technique, to cancel RF's resistor noise.

A cascode transistor is added to provide feedback to the CG stage. This feedback reduces the noise of M_2 and makes the noise of M_3 to be neglected. This cascode feedback scheme can improve noise cancellation achieving sub-2dB NF. IIP3 is -11 dBm though. The branches of the circuit are not well balanced in non-linearity, which makes IIP3 negative.

Table 2.4: Best power consumption circuit.

P. Gain (dB)	Bandwidth (GHz)	NF (dB)	IIP2 (dBm)	IIP3 (dBm)	Area (mm^2)	Power (mW)	Tech (nm)
7.8-12.3	3.1-13.9	2.7-3.3	-	-6.4	0.1	2.5	90
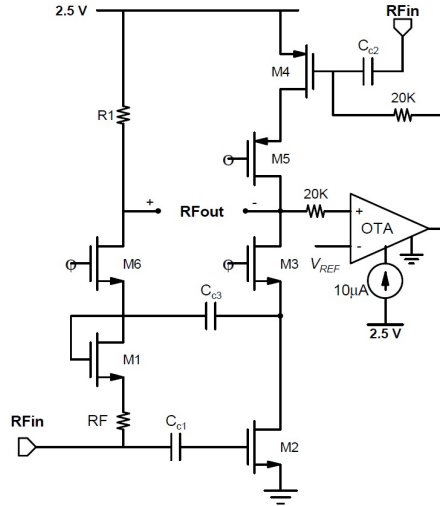
Figure 2.15: The best noise figure circuit schematic. Picture taken from YOUSSEF; ISMAIL; HASLETT (2010).

Table 2.5: Best noise figure circuit.

V. Gain (dB)	Bandwidth (GHz)	NF (dB)	IIP2 (dBm)	IIP3 (dBm)	Area (mm^2)	Power (mW)	Tech (nm)
36.5	0.17-0.9	1.6	-	-11	-	18	65

Table 2.6 shows the final comparison of all topologies discussed above. The * indicates power gain, the others are voltage gain. Based on what has been published in the literature, the aim of this work is to achieve or get close to the following specifications: Voltage Gain > 15 dB, Noise Figure < 3 dB, IIP3 > 0 dBm, Balun, Inductorless, Bandwidth of 50 MHz - 1 GHz, which contains the bandwidth of the IEEE 802.22 standard, S_{11} < -10 dB, Power: as low as possible to achieve the other specifications and Area: as small as possible to achieve the other specifications.

Table 2.6: State of the art comparison table.

Ref.	Gain (dB)	Bandwidth (GHz)	NF (dB)	IIP2 (dBm)	IIP3 (dBm)	Area (mm^2)	Power (mW)	Tech (nm)
Custodio(2010)	36.5	0.17-0.9	1.6	-	-11	-	18	65
Cheng(2011)	7.8-12.3	3.1-13.9	2.7-3.3	-	-6.4	0.1	2.5	90
Sinencio(2009)	8.6-11.7	1.5-8.1	3.6-6	7.6-23	11.7-14.1	0.58	2.62(core)	130
Taniguchi(2008)	24-25	0.05-10	2.7-3.6	10.2-32.4	-10-(-2)	-	21.7	65
Haslett(2010)	17.7	0.2-1	4-4.2	-	-9.8	0.0069	7.35	130

3 THE LOW NOISE AMPLIFIER DESIGN

As can be noticed by the previous chapter, all noise-canceling LNA topologies had the auxiliary amplifier gain attached to the main amplifier. Also, if either IIP3 or NF was targeted to be improved, one had to choose a different topology, because the decoupled parameters for optimization were different. Thus, in order to get any desired performance improvement, without having to change the topology, one has to start a new topology to get more degrees of freedom in design, to decouple all parameters, such as gain, NF, S_{11}, IIP3 from each other.

The IEEE 802.22 WRAN standard was targeted to demonstrate the application of the topology developed in this work. However, this topology is suitable for any wideband application up to 1 GHz or few GHz.

The next section shows how this new topology was derived and section 2 presents its main characteristics. Section 3 if focused on the layout implementation of the topology in IBM 130 nm and section 4 shows all simulation results for each LNA, the best NF and best IIP3 LNA. These are the main performance parameters of an LNA (as its name calls out for Low-Noise and wideband for high linearity), putting gain and S_{11} as secondary parameters of performance, being power consumption and area the last ones to be cared about in such an environment. This means, that the secondary parameters can be on their worst case limit (e.g., S_{11} could be sacrificed to even -8 dB if an incredible IIP3 and NF were achieved), while the main parameters should always be the best possible.

3.1 Topology design

Looking carefully at the noise-canceling LNA topologies, one can notice that the gain of the auxiliary amplifier (e.g., $-1 - \frac{R_F}{R_S}$ in resistive-feedback topology) is totally dependent of the main amplifier noise gain ($1 + \frac{R_F}{R_S}$), as shown in Fig. 3.1. That is exactly what makes the noise canceling principle work, the auxiliary amplifier gain is adjusted to the main amplifier noise gain (which is different from the signal gain) and subtracted/added at the output to cancel noise and distortion.

Thus, as this auxiliary amplifier is usually implemented as single transistor amplifier, there's no degrees of freedom to adjust the gain of this amplifier to get a better trade-off with other parameters. As the auxiliary amplifier gain affects both NF and IIP3, if one could decouple it from the main amplifier's noise gain, it would be possible to improve overall performance of NF and IIP3. This lead to the initial topology of this work, the addition of more degrees of freedom in the auxiliary amplifier by cascading two CS amplifiers, decoupling each transistor gain from the main amplifier gain. This is described in the following subsection.
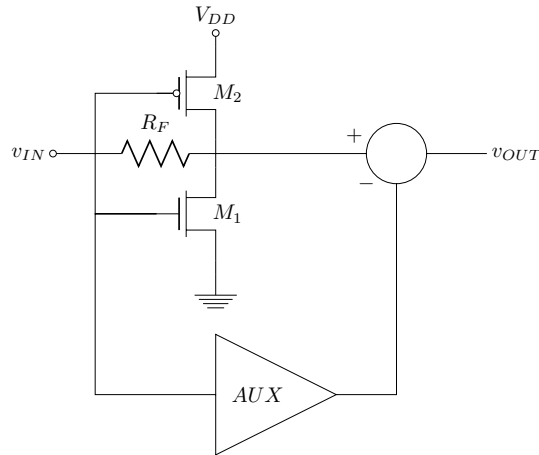
Figure 3.1: The resistive-feedback noise canceling topology. If the auxiliary amplifier is implemented using a single transistor, its gain is completely attached to that of the main amplifier.

3.1.1 Initial Topology

The initial topology created is shown in Fig. 3.2. It is based on the resistive-feedback + CS amplifier noise-canceling, however the auxiliary amplifier is composed by a cascade of 2 CS amplifiers instead of one. A resistive load in each auxiliary stage was initially used in order not to add flicker noise of transistors, if they were used as load. Thus, only thermal noise was present on the auxiliary amplifier loads.
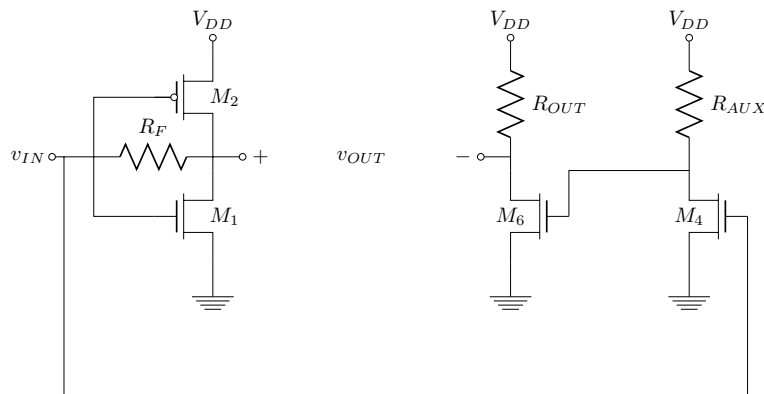
Figure 3.2: Initial topology with cascaded auxiliary amplifier.

This topology works well, with improved NF, however the cascaded non-linearity of both CS amplifiers in the auxiliary branch produces a very low IIP3 at the output, which is undesired. The branches have unbalanced non-linearity orders. Thus, another degree of freedom was needed to improve overall IIP3 and equate, as much as possible, the non-linearity order of the output voltage of each stage. This problem was solved and is shown in the next subsection.

3.1.2 Second Topology with Improved Auxiliary Amplifier IIP3

The noise-canceling topology also helps canceling distortion at the output. However, one must have the same non-linearity order (output current or voltage) in both branches in order to get a good distortion cancellation at the output. The initial topology had two

cascaded CS amplifiers which generated a non-linearity of higher order than the main amplifier.

In order to solve this problem, a diode-connected NMOS transistor was used as a load (instead of a resistor) in the first amplifier to make it very linear. For example, if the transistor quadratic model is considered, a 2nd order current is generated and turned into voltage by a square root relation in the diode as load, ideally, making this amplifier perfect linear. This behavior can be explained by considering the I-V curve of the transistor as a function f and the load V-I curve as f^{-1}. However, as the voltage-current curve of the transistor is not perfectly quadratic, nor the current-voltage curve of the diode is perfectly square root, this amplifier is not completely linear. Although having the best linearity possible among the single transistor amplifiers. The low gain of this amplifier (given the diode load) also contributes to having a high linearity. Fig. 3.3 shows the second topology derived in this work.
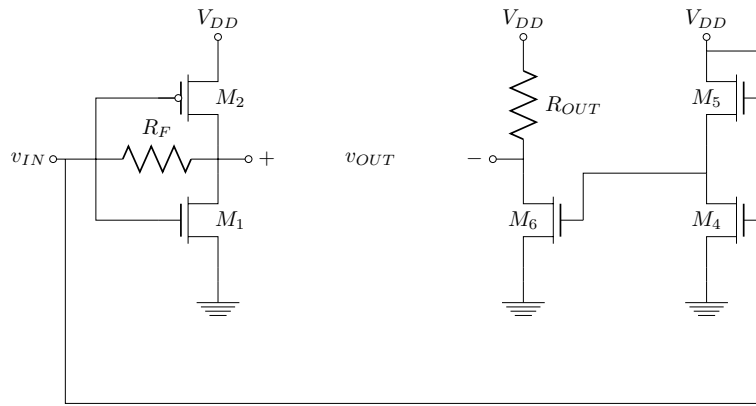
Figure 3.3: Second topology with a diode as a load of the first amplifier in the auxiliary branch.

There is a small inconvenience coming along with the diode load, which is the Flicker noise, not present in resistive loads. So a Flicker/thermal noise optimization was need in the first amplifier of the auxiliary cascade. It can be done by increasing the transistor area (WL), while keeping the aspect ratio (W/L) to maintain the same g_m. The transistor area can be increased until the corner frequency goes lower than the minimum frequency in the band of the amplifier, but at the expense of larger parasitic capacitances, requiring a trade-off adjust between noise and gain flatness. The flicker noise of M_4 and M_6 can also be reduced in the same way. This optimization was implemented so as to achieve low NF in frequencies as low as 50 MHz.

Making the first amplifier in the auxiliary cascade highly linear improved IIP3 by 10 dBm, reaching 0 dBm, up from -10 dBm. Which means, it has almost equal non-linearity order on both branches, improving the distortion cancellation at the output. However, this distortion cancellation is not very well controlled and still a more certain adjustment in IIP3 was needed to guarantee the best IIP3 possible. This problem was solved in the final topology described in the next subsection. Besides all that, putting a diode-connected transistor as a load in the first amplifier of the auxiliary cascade made it consume the highest power in this topology. This is mainly because this amplifier has its best linearity with high currents, in the mA order. Also, it is difficult to increase gain in this topology, given an increase in the V_{GS} of the gain transistor, increases the output current and the load transistor V_{GS} is also increased, not providing the expected gain increase. As only a small gain was needed at this stage, this idea worked well.

3.1.3 Final Topology with Improved Main Amplifier IIP3

In order to have a certain adjustment dedicated to improve IIP3 and provide a controlled adjustment, a diode-connected NMOS transistor was placed in the output node to subtract the output current higher order non-linear terms. This was intended in order to allow higher gains in each branch, compensating the difference in non-linearity order by adjusting the diode aspect ratio. If the diode is not used, the non-linearity of the main amplifier is not decoupled from NF and input impedance matching. This final topology makes the IIP3, NF, input impedance matching and gain all decoupled from each other, providing design flexibility.

In order to demonstrate this flexibility and the decoupling characteristics in all parameters of the proposed topology, in this work, 3 LNAs were implemented, a first design, which, due to MOSIS schedule, was submitted to fabrication before the idea of focusing on best-NF or best-IIP3 came out. Thus it was designed on an point between best-NF and best-IIP3, a second amplifier maximizing IIP3 with a reasonable NF and a third design minimizing NF to the limit of < 2 dB, while having reasonable IIP3. This shows the designer the topology flexibility available to different applications.

The current bias of this diode in the main amplifier output is 2-3% of that of the main transistor, which reduces its impact on gain. The final topology is shown in Fig. 3.4, with each part highlighted.
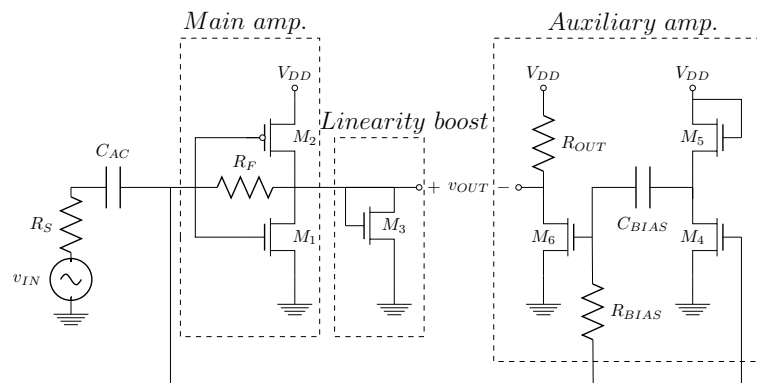
Figure 3.4: Final topology with diode at main amplifier's output to better control IIP3.

Also, a decoupling capacitor was inserted between the amplifiers of the auxiliary cascade, so that the bias of the output amplifier could come from M_1, within the negative feedback loop of the main amplifier. This configuration sets all amplifiers biases based on a feedback, which gives lower sensitivity to process variations.

Monte Carlo simulations were run and shown in the results chapter to demonstrate the low sensitivity to process variations and mismatch. Also, the impact of temperature variation was simulated also presented in the results chapter. All this low sensitivity to process and temperature is due to the bias being based on the feedback of the main amplifier.

In the following section all quantitative characteristics of the final topology are described and analyzed.

3.2 Circuit Analysis

Circuit equations are presented here in this section.

3.2.1 Input Matching

The input impedance is directly connected (through R_F) to the output load. This will cause different sizing for power gain and voltage gain (different loads). In this work, the LNA was designed to have voltage gain with the load specified in the simulation section (R_{load} is 20x higher than 50 Ω, so it's neglected in this analysis, as it is in parallel with R_F). The input impedance matching is measured as the S_{11} parameter (refer to appendix A for details about S_{11}), which is simply (considering $Z_L = Z_0$):

$$S_{11} = \frac{Z_{in} - R_S}{Z_{in} + R_S} \tag{3.1}$$

where Z_{in} is the amplifier input impedance and R_S the source output resistance. Usually, this is only valid if the output is matched. However, in modern RF design, this condition can be removed and the above expression still represents a matching figure RAZAVI (2011). It is usually measured in dB, and intended to be less than -10 dB, which means that the mismatch between Z_{in} and R_S is lower than 10%.

The input impedance of the LNA is set by the parallel association of the input impedance of the main amplifier and the first amplifier of the auxiliary cascade, $Z_{in_{main}}$ and $Z_{in_{AUX1}}$, respectively. Neglecting the r_{ds} of the transistors, the small signal model shown in Fig. 3.5 is for the main amplifier input impedance, $Z_{in_{main}}$, neglecting the influence of M_3 for the moment.
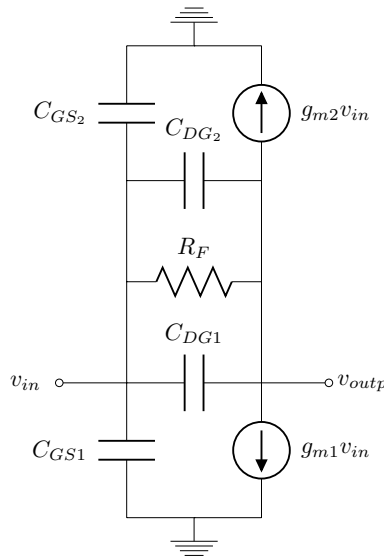
Figure 3.5: Small signal of main amplifier, neglecting transistor r_{ds}.

Setting up the Kirchoff's Current Law (KCL) equations for the input and output nodes, one gets:

$$(v_{in} - v_{outp})(sC_{DG1} + \frac{1}{R_F} + sC_{DG2}) + v_{in}(sC_{GS1} + sC_{GS2}) - i_{in} = 0 \tag{3.2}$$

$$(v_{outp} - v_{in})(sC_{DG1} + \frac{1}{R_F} + sC_{DG2}) + g_{m2}v_{in} + g_{m1}v_{in} = 0 \tag{3.3}$$

Summing both equations above gives:

$$v_{in}(sC_{GS1} + sC_{GS2}) + (g_{m1} + g_{m2})v_{in} = i_{in} \tag{3.4}$$

$$Z_{in_{main}} = \frac{v_{in}}{i_{in}} = \frac{1}{(g_{m1} + g_{m2}) + s(C_{GS1} + C_{GS2})} \tag{3.5}$$

The above equations show a low-pass filter behavior, which means, the parasitic capacitance C_{GS} of both transistors will limit the bandwidth.

The small signal model of the first amplifier in the auxiliary cascade is shown in Fig. 3.6. The effect of the second amplifier in the auxiliary cascade is neglected.
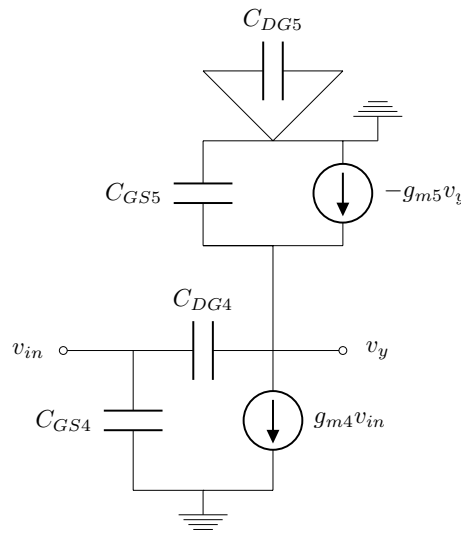
Figure 3.6: Small signal of the first amplifier in the auxiliary cascade, neglecting transistor r_{ds}.

KCL equations for the input impedance gives:

$$
\begin{aligned}
(v_{in} - v_y)sC_{DG4} + v_{in}sC_{GS4} - i_{in} &= 0 \\
(v_y - v_{in})sC_{DG4} + g_{m4}v_{in} + g_{m5}v_y + v_y sC_{GS5} &= 0
\end{aligned}
\tag{3.6}
$$

Which gives (v_y is the output voltage of the first amplifier of the auxiliary cascade):

$$v_y(sC_{DG4} + g_{m5} + sC_{GS5}) = (sC_{DG4} - g_{m4})v_{in} \tag{3.7}$$

$$v_y = \left(\frac{sC_{DG4} - g_{m4}}{sC_{DG4} + g_{m5} + sC_{GS5}} \right) v_{in} \tag{3.8}$$

$$v_y = \alpha v_{in} \tag{3.9}$$

$$\tag{3.10}$$

Where α is used to simplify notation. Substituting into 3.6, it gives:

$$(v_{in} - \alpha v_{in})sC_{DG4} + v_{in}sC_{GS4} - i_{in} = 0 \tag{3.11}$$

$$(sC_{DG4} - \alpha sC_{DG4} + sC_{GS4})v_{in} = i_{in} \tag{3.12}$$

$$Z_{in_{aux}} = \frac{v_{in}}{i_{in}} = \frac{1}{s(C_{DG4}(1 - \alpha) + C_{GS4})} \tag{3.13}$$

Thus, $Z_{IN} = Z_{in_{main}}//Z_{in_{AUX1}}$, giving:

$$Z_{IN} = \frac{1}{g_m + s(C + C_{DG4}(1 - \alpha) + C_{GS4})} \tag{3.14}$$

Where $g_m = g_{m1}+g_{m2}$, $C = C_{GS1}+C_{GS2}$ and:

$$\alpha = \frac{-g_{m4} + sC_{DG4}}{g_{m5} + s(C_{DG4} + C_{GS5})}. \tag{3.15}$$

This means that the influence of the first amplifier in the auxiliary cascade will be mostly in the bandwidth. The biggest influence on input resistance matching will be given by the main amplifier.

With this idea in mind, now it is possible to neglect the first amplifier in the auxiliary cascade influence and estimate the impact of M_3 on the input impedance. Fig. 3.7 shows the small signal model for the next calculus. Neglecting the parasitic capacitances and transistor r_{ds}, the input impedance is given by the expression:

$$R_{IN} = \frac{1 + g_{m_3}R_F}{g_{m_1} + g_{m_2} + g_{m_3}} \approx \frac{1}{g_{m_1} + g_{m_2}} \tag{3.16}$$
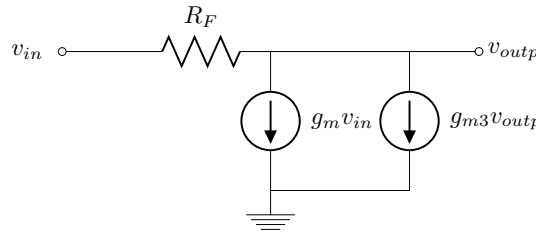
Figure 3.7: Small signal of the main amplifier, considering M_3, neglecting transistor r_{ds}.

Where $g_m = g_{m1}+g_{m2}$. For a first design, as g_{m3} is 2-3% of g_{m1} and g_{m2}, it can be neglected, but it is going to be considered afterwards in the optimization subsection. The input impedance was matched to a source impedance of $R_S = 50$ Ohms, leading to a preliminary value of $g_{m_1} + g_{m_2} = 20$ mS.

3.2.2 DC Gain

The calculation of the total gain A_V of the LNA can be divided into 3 parts: $A_{V_{MAIN}}$, $A_{V_{AUX1}}$ and $A_{V_{AUX2}}$. Each one of these is explained in sequence in the following subsections.

3.2.2.1 Main Amplifier

The small signal model for the gain calculation of the main amplifier is shown in Fig. 3.8. For the sake of simplicity it is considered only C_{GS} of each transistor (PMOS and NMOS). The C_{DG} effect can be included in C_{GS} using the Miller theorem.

In the KCL equations for this amplifier, if C_{GS} is not considered, there is a series connection of R_S and R_F, which makes possible to equate the current which passes through both resistors to the current which gets out of them, i.e., the sum of the controlled sources
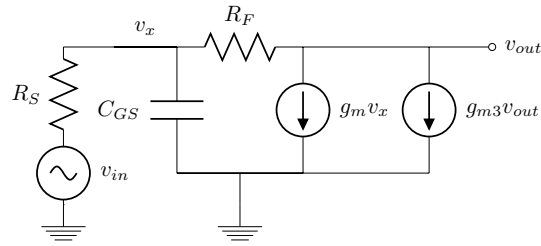
Figure 3.8: Small signal of the main amplifier, considering M_3, neglecting transistor r_{ds}.

currents. Thus, the expression for $A_{V_{MAIN}}$, neglecting g_{m3}, C_{GS} and assuming $g_{m1} + g_{m2}$ = $1/R_S$, is simplified to:

$$A_{V_{MAIN}} = \frac{1}{2}(1 - \frac{R_F}{R_S}) \qquad (3.17)$$

which is not achieved if one takes the other KCL equations considering C_{GS}, which adds its current in the middle of R_F and R_S, and in the end of the calculations one makes s=0, g_{m3}=0, $g_{m1} + g_{m2}$ = $1/R_S$ on it. In this case one gets

$$A_{V_{MAIN}} = -\frac{R_F}{R_S} \qquad (3.18)$$

instead, due to the inclusion of C_{GS}. However, both equations considering or not C_{GS}, when simplified, lead to the same dependance on R_F and R_S. Thus, neglecting parasitic capacitances, r_{ds} and making $g_m = g_{m1} + g_{m2}$, the KCL equations are:

$$\frac{v_{in} - v_{out}}{R_S + R_F} = g_m v_g + g_{m3} v_{out}$$
$$\frac{v_{in} - v_x}{R_S} = \frac{v_x - v_{out}}{R_F} \qquad (3.19)$$

which leads to:

$$v_{in} R_F - v_x R_F = R_S v_x - R_S v_{out}$$
$$v_x = \frac{v_{in} R_F + v_{out} R_S}{R_F + R_S} \qquad (3.20)$$

substituting into 3.19, one gets:

$$\frac{v_{in} - v_{out}}{R_S + R_F} = g_m \frac{v_{in} R_F + v_{out} R_S}{R_F + R_S} + g_{m3} v_{out}$$
$$A_{V_{MAIN}} = \frac{1 - (g_{m_1} + g_{m_2})R_F}{1 + (g_{m_1} + g_{m_2})R_S + g_{m_3}(R_F + R_S)} \qquad (3.21)$$

neglecting g_{m3}, considering $R_F \gg R_S$ and assuming input impedance matching ($R_S = \frac{1}{g_{m_1} + g_{m_2}}$),

$$A_{V_{MAIN}} = \frac{1}{2}(1 - \frac{R_F}{R_S}) \approx -\frac{1}{2}\frac{R_F}{R_S}. \tag{3.22}$$

In order to see the effect of the gate parasitic capacitance $C_{GS} = C_{GSP} + C_{GSN}$, one has to remake the KCL equations as follows:

$$v_x - v_{in} + v_x sC_{GS} + \frac{v_x - v_{out}}{R_F} = 0$$

$$\frac{v_{out} - v_x}{R_F} + g_m v_{in} + g_{m3} v_{out} = 0 \tag{3.23}$$

isolating v_x in both equations of 3.23:

$$v_x = R_F g_m v_{in} + (g_{m3}R_F + 1)v_{out} \tag{3.24}$$

$$v_x = \frac{v_{in}R_F + v_{out}R_S}{R_S + R_F + sR_F R_S C_{GS}} \tag{3.25}$$

equating v_x of both equations:

$$\frac{v_{in}R_F + v_{out}R_S}{R_S + R_F + sR_F R_S C_{GS}} = R_F g_m v_{in} + (g_{m3}R_F + 1)v_{out}$$

$$v_{in}R_F + v_{out}R_S - R_S R_F g_m v_{in} = R_S g_{m3}R_F v_{out} + R_F R_S v_{out} +$$

$$+R_F^2 g_m v_{in} + R_F^2 g_{m3}v_{out} + R_F v_{out} + sR_F^2 R_S C_{GS} g_m v_{in} +$$

$$+sR_F^2 R_S C_{GS}g_{m3}v_{out} + sR_F R_S C_{GS}v_{out} \tag{3.26}$$

isolating v_{in} and v_{out}:

$$A_{V_{MAIN}} = \frac{v_{out}}{v_{in}} = \frac{R_F(1 - g_m(R_S + R_F)) + sR_F R_S C_{GS}g_m}{R_F + (R_S R_F + R_F^2)g_{m3} + s(R_F R_S C_{GS} + (R_F^2 R_S C_{GS})g_{m3})} \tag{3.27}$$

Now, if one makes s = 0, $g_{m3} = 0$, $g_m = 1/R_S$, it simplifies to

$$A_{V_{MAIN}} = -\frac{R_F}{R_S}. \tag{3.28}$$

The effect of the parasitic capacitance C_{GS} and also C_{DG} included only limits the bandwidth of the gain as shown in Eq. 3.27. So the designer only needs to care about the size of M_1 and M_2, since making them too large will impact in bandwidth and also input impedance matching.

3.2.2.2 Auxiliary amplifier

The auxiliary amplifier is composed by M_4, M_5, M_6 and R_{OUT}, implementing two cascaded CS amplifiers. The first CS amplifier has a diode-connected NMOS load (M_5) in order to improve its linearity, while penalizing with the addition of flicker noise.

The small signal model for the gain calculation of the first auxiliary amplifier is shown in Fig. 3.9. For the sake of simplicity it is considered only C_{GS} of each transistor (M_4, M_5, M_6). The C_{DG} effect can be included in C_{GS} using the Miller theorem.
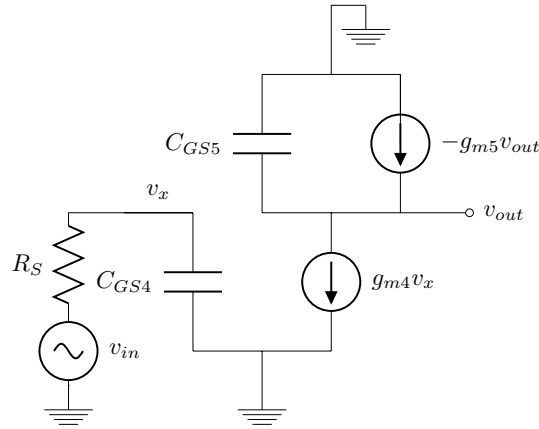
Figure 3.9: Small signal model of the auxiliary amplifier input stage, neglecting transistor r_{ds}.

The KCL equations are:

$$
\begin{aligned}
\frac{v_{in} - v_x}{R_S} &= v_x s C_{GS4} \\
g_{m4} v_x + g_{m5} v_{out} + v_{out} s C_{GS5} &= 0
\end{aligned}
\tag{3.29}
$$

isolating v_x in the first equation and substituting into the second:

$$
\begin{aligned}
v_x &= \frac{v_{in}}{s R_S C_{GS4} + 1} \\
\frac{v_{in} g_{m4}}{s R_S C_{GS4} + 1} &= v_{out}(-g_{m5} + s C_{GS5})
\end{aligned}
\tag{3.30}
$$

then,

$$
A_{V_{AUX1}} = \frac{-g_{m4}}{g_{m5} + s(C_{GS5} - R_S C_{GS4} g_{m5}) + s^2 R_S C_{GS4} C_{GS5}}
\tag{3.31}
$$

which simplifies to

$$
A_{V_{AUX1}} = -\frac{g_{m4}}{g_{m5}}
\tag{3.32}
$$

if s = 0. In order to achieve a good linearity (high IIP3) in this stage, it requires a high power consumption, or high DC current. This first stage consumes half of the entire power consumption of the LNA.

The second stage of the auxiliary cascade is a simple CS amplifier with resistive load. As low frequencies as low as 50 MHz were targeted, an active load was avoided here so as to have lowest Flicker noise possible at the output and highest bandwidth. The small signal model is shown in Fig. 3.10. For the sake of simplicity it is again considered only C_{GS} of the transistor. The C_{DG} effect can be included in C_{GS} using the Miller theorem.

The KCL equations for this amplifier are (R_S for this amplifier is the R_{AUX1}, coming from the first stage):
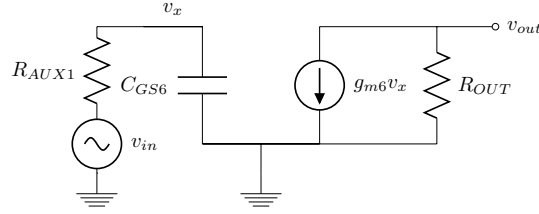
Figure 3.10: Small signal of the second stage of the auxiliary amplifier, considering M_3, neglecting transistor r_{ds}.

$$v_x = \frac{v_{in}}{sR_{AUX1}C_{GS6} + 1} \tag{3.33}$$

$$g_{m6}v_x + \frac{v_{out}}{R_{out}} = 0 \tag{3.34}$$

which gives:

$$A_{V_{AUX2}} = -\frac{g_{m6}R_{out}}{sR_{AUX1}C_{GS6} + 1} \tag{3.35}$$

which simplifies to (DC gain):

$$A_{V_{AUX2}} = -g_{m_6}R_{OUT}. \tag{3.36}$$

if s = 0. Thus, the total again of the LNA is $A_V = A_{V_{MAIN}}$ - $A_{V_{AUX1}} * A_{V_{AUX2}}$. The total expression, assuming the previously explained simplifications, is:

$$A_{V_{TOTAL}} = -\frac{R_F}{R_S} - \frac{g_{m_4}g_{m_6}R_{OUT}}{g_{m_5}} \tag{3.37}$$

Each branch of the LNA was designed to have 8-9 voltage gain in order to have $A_{V_{TOTAL}} > 16$ (24 dB), with $A_{V_{AUX1}}$ and $A_{V_{AUX2}}$ having 2.5-3 V/V each. This was the highest gain achieved, maintaining the IIP3 higher than 0 dBm.

3.2.3 Noise Figure

The NF of the LNA was divided into the main amplifier's NF and the auxiliary amplifier's NF. As they are decoupled from each other in this design, they could be separately optimized. The thermal noise depended on the resistances values R_F and R_{OUT} and the g_m of the transistors. The higher the gain of each amplifier, the lower the NF and the lower the IIP3 too (so one must care about increasing gain too much). Due to the feedback, the higher the value of R_F (the higher the gain), the lower the NF of the main amplifier. Also, as R_{OUT} was directly connected to the output with no feedback, the lower its value, the lower the NF of the auxiliary cascade. The derivation of the NF equations are detailed as follows.

The NF of a generic amplifier is given by RAZAVI (2011):

$$NF_{amp} = 1 + \frac{\overline{V_{n_{OUT}}^2}}{|\alpha|^2 A_V^2} \frac{1}{V_{n_{RS}}^2} \tag{3.38}$$

where $\overline{V_{n_{OUT}}^2}$ is the amplifier noise at the output, α is 1/2 for perfect input matching, $\overline{V_{n_{RS}}}$ is the source impedance noise and A_V the amplifier voltage gain.

For the main amplifier, the $\overline{V_{n_{OUT}}^2} = \overline{I_{n_{OUT}}^2} * R_{OUT_{MAIN}}^2$, where $\overline{I_{n_{OUT}}^2}$ is its noise current. The expression for $R_{OUT_{MAIN}}$, neglecting M_3, is:

$$
\begin{aligned}
R_{OUT_{MAIN}} &= \left[\frac{1}{g_{m1} + g_{m2}} \left(1 + \frac{R_F}{R_S} \right) \right] // (R_F + R_S) \\
R_{OUT_{MAIN}} &= \frac{1}{2}(R_F + R_S)
\end{aligned}
\tag{3.39}
$$

and the expression for $\overline{I_{n_{OUT}}^2}$ is RAZAVI (2001):

$$
\overline{I_{n_{OUT}}^2} = 4kT\gamma(g_{m1} + g_{m2})
\tag{3.40}
$$

where γ is not the body factor of the MOSFET. It actually comes from the inversion charge model used for strong inversion and saturation of a MOSFET noise current model TSIVIDIS (1987), known to be 2/3 in long-channel devices. Considering the R_F noise as $4kTR_F$, the final expression for $\overline{V_{n_{OUT}}^2}$:

$$
\overline{V_{n_{OUT}}^2} = 4kTR_F + 4kT\gamma(g_{m1} + g_{m2})\frac{(R_F + R_S)^2}{4}
\tag{3.41}
$$

whose terms are the feedback resistor and transistor noises contributions at the output.

Substituting $\overline{V_{n_{OUT}}^2}$ in the NF expression, considering α=1/2, $A_{V_{MAIN}} = -\frac{R_F}{R_S}$, and $\overline{V_{n_{RS}}^2}$ = 4kTR_S:

$$
\begin{aligned}
NF_{MAIN} &= 1 + \frac{4R_F}{R_S \left(-\frac{R_F}{R_S} \right)^2} + \frac{\gamma(g_{m1} + g_{m2})(R_F + R_S)^2}{R_S \left(-\frac{R_F}{R_S} \right)^2} \\
&\approx 1 + \frac{4R_S}{R_F} + \gamma(g_{m1} + g_{m2})R_S \\
&\approx 1 + \frac{4R_S}{R_F} + \gamma
\end{aligned}
\tag{3.42}
$$

if $(g_{m1} + g_{m2}) = \frac{1}{R_S}$. The last term is canceled by the auxiliary amplifier at the output, using the noise canceling technique. Adding the noise of the auxiliary amplifier, the total NF equation (considering only thermal noise) for the LNA is:

$$
\begin{aligned}
NF_{TOTAL} &= 1 + \frac{4R_S}{R_F} + \frac{A_{V_{AUX_{total}}}^2 \overline{v_{n_{AUX}}^2}}{\left(\frac{1}{2} \right)^2 \left(-\frac{R_F}{R_S} \right)^2 4kTR_S} \\
NF_{TOTAL} &= 1 + \frac{4R_S}{R_F} + \frac{A_{V_{AUX_{total}}}^2 \overline{v_{n_{AUX}}^2} R_S}{kTR_F^2}
\end{aligned}
\tag{3.43}
$$

where k is the boltzmann constant, T is temperature, $A_{V_{AUX_{total}}} = A_{V_{AUX1}} * A_{V_{AUX2}}$ and $\overline{v_{n_{AUX}}^2}$ is the input referred noise of the auxiliary amplifier given by:

$$\overline{v_{n_{AUX}}^2} = 4KT \left[\gamma \left(\frac{g_{m4} + g_{m5}}{g_{m4}^2} + \frac{g_{m5}^2}{g_{m4}^2 g_{m6}} \right) + \frac{g_{m5}^2}{R_{OUT} g_{m4}^2 g_{m6}^2} \right] \tag{3.44}$$

Although R_{OUT} is in the denominator of the above equation, an increase on its value does not decrease the value of $v_{n_{AUX}}^2$, given an increase in the value of R_{OUT}, reduces M_6 bias current, also reducing g_{m6}. Thus, $v_{n_{AUX}}^2$ is reduced linearly by R_{OUT}, but increased quadratically by the reduction of g_{m6}, which leads to an overall increase in $v_{n_{AUX}}^2$.

Then, R_{OUT} was designed to be as low as possible in order to optimize thermal noise at the output, while also contributing to provide power gain (having output resistance not far from input 50 Ω). The feedback resistor R_F was designed to be 680 Ω (1081Ω on sub-2dB LNA). It could not be too high, because it started to degrade input impedance matching. R_{OUT} was designed to be 142 Ω (156 Ω on sub-2dB LNA), reducing thermal noise and allowing power gain if needed (being not too far from input 50 Ω).

3.2.4 Linearity

The noise-canceling topology proposed by BRUCCOLERI; KLUMPERINK; NAUTA (2004) also provides non-linear terms cancellation, achieving an IIP3 not far from 0 dBm, without any additional circuitry to improve IIP3. Thus, many people around the world started doing research on linearity improvement techniques to apply them to the noise-canceling topology as additional circuitry. All of these linearity techniques, although slightly changing the original topology, still kept the single transistor auxiliary amplifier idea, which gives no freedom in design to decouple the gain, NF and IIP3 trade-off of the auxiliary amplifier.

This work introduces a new topology which is one of the firsts to use a cascaded auxiliary amplifier. The main reason for this is the additional degrees of freedom to optimize linearity, decoupling it from other trade-offs like NF and Gain. The main problem of the traditional noise-canceling technique with single transistor as the auxiliary amplifier is that the gain of this auxiliary amplifier is attached to the main amplifier's noise gain. Which means, in order to cancel the main amplifier's noise, the auxiliary amplifier has no freedom in gain design. As linearity is directly related to the gain of the amplifier, this is a limitation for the designer. Therefore, this explains why the NF of all LNAs which use linearity improvement techniques are quite high. As the designer did not have any degree of freedom to change the gain of the auxiliary branch related to the main amplifier, the designer was forced to put additional circuitry, thus, increasing overall NF.

Using an auxiliary cascade gives the designer the freedom to divide the gain of the auxiliary cascade into 2 amplifiers and profit from all the advantages of cascading amplifiers. One advantage comes from the LNA concept or the Frii's equation, which means the first amplifier in the cascade has to have a low NF (LNA concept itself). Then, the topology developed in this work has a small LNA inside the final LNA. Actually, the first amplifier in the auxiliary cascade has the highest DC current of all amplifiers in the topology, providing the least noise possible and the highest linearity possible. In order to achieve this, a CS amplifier with diode load was used.

The use of a CS amplifier with diode load was intended because it is the most linear amplifier among the single transistors amplifiers. The drawback of using cascade amplifiers in the auxiliary branch is the non-linearity order increase at the output as more and more amplifiers are cascaded. Considering the quadratic model for the MOSFET transistor for simplicity and assuming a cascade of single transistor amplifiers, which are capable

of providing gain (i.e., not considering the common-drain amplifier), they all have second order currents or quadratic I-V curves. If one cascades two of these amplifiers, it will provide a fourth order current at the output. Thus, as the main amplifier has a second order current at the output, the noise-canceling topology does not have an efficient non-linear term cancellation. If two CS amplifiers are cascaded, with the fourth order current at the output, the IIP3 of the noise-canceling topology goes to -10 dBm, which is undesirable.

Thus, as explained in section 3.1.2, a diode-connected transistor was used as a load in order to linearize the first stage of the auxiliary cascade in exchange of adding the flicker noise of M_5.

Reducing flicker noise by swapping the position of M_5 and R_{OUT} was not implemented, since the total linearity of the auxiliary amplifier was taken as a higher design priority. The following equation (taken from RAZAVI (2011)) describes the total IIP3 of the auxiliary amplifier.

There are two possibilities to reduce the left side term of this equation (increase total IIP3). Both try to minimize the right side of the equation (increase IIP3 of both stages).

The first way is to increase A_{IIP3_1} as much as possible making the first term of the right side small. In order to achieve a high A_{IIP3_1}, it will usually imply on a small gain $A_{V_{AUX1}}$, which also contributes to reduce the second term of the right side of the equation. The value of A_{IIP3_2} can have more margin to be little small, but not too far from A_{IIP3_1}. This approach was used in this work. So the first amplifier stage in the auxiliary cascade was designed to be very linear with very small gain, while the second amplifier stage had a similar gain but not very linear. In summary, it was chosen to make the first term in the right side of the equation smaller than the second term.

The second possibility to have a high total IIP3 is to make the second term of the right side of this equation smaller than the first term. In order to achieve this, the second amplifier stage has to have a higher IIP3 than the first amplifier stage. This gives more margin for the first amplifier stage to have a higher gain and lower IIP3. The value of A_{IIP3_2} must be as high as the first amplifier stage gain, in order to cancel its effect.

Thus, in the strategy used in this work, the IIP3 of the first amplifier has a bigger impact than the second, i.e., the first amplifier has to have a better IIP3 and it uses a diode as a load to achieve that.

$$\frac{1}{A_{IIP3_{TOTAL}}^2} = \frac{1}{A_{IIP3_1}^2} + \frac{A_{V_{AUX1}}}{A_{IIP3_2}^2} \tag{3.45}$$

As the CS with diode load had the best linearity possible among the single transistors amplifiers (see section 3.1.2), a high IIP3 was achieved on the overall LNA, with both branches (main and auxiliary) having almost the same order of non-linearity at the output current/voltage.

The added flicker noise of M_5 can not be removed, but can be reduced by the increase of the transistor area (WL), while keeping the aspect ratio (W/L) to maintain the same g_m. The transistor area can be increased until the corner frequency goes lower than the minimum band frequency of the amplifier but at the expense of larger parasitic capacitances, requiring a trade-off adjust between noise and gain flatness. The flicker noise of M_4 and M_6 can also be reduced in the same way, and R_{OUT} is designed to be low in order to reduce thermal noise, while providing enough power gain.

At this point one should notice that the main amplifier has not been touched, keeping the same characteristics of the common noise-canceling topology. However, it is

not enough to optimize the IIP3 of the auxiliary amplifier, because the IIP3 of parallel connected amplifiers is dominated by the amplifier with worst IIP3. Therefore, an IIP3 improvement on the main amplifier was also needed, not to mention the main amplifier itself does not have a mechanism of controlling the non-linear term cancellation at the output.

The diode-connected at the main amplifier's output was used to improve IIP3 and decouple it from the gain and NF design of the main amplifier. This linearization technique was used in ZHANG; FAN; SINENCIO (2009) on a CG amplifier. In this work, it is shown that it is also possible to use this linearization technique on a CS amplifier as follows. First assume v_{g1} and v_{g3} are the gate and drain voltages of M_1 and M_3, respectively. The small signal drain currents that result from these gate voltages can be expanded in power series to the cubic power ZHANG; FAN; SINENCIO (2009) as:

$$i_{d1} = g_{m_1}v_{g1} + g_{2_1}v_{g1}^2 + g_{3_1}v_{g1}^3 \tag{3.46}$$

$$i_{d3} = g_{m_3}v_{g3} + g_{2_3}v_{g3}^2 + g_{3_3}v_{g3}^3 \tag{3.47}$$

Since v_{g3} is also the drain voltage of M_1, it can be related to v_{g1} by:

$$v_{g3} = b_1 v_{g1} + b_2 v_{g1}^2 + b_3 v_{g1}^3 \tag{3.48}$$

where b_1-b_3 are in general frequency dependent and can be estimated from simulation. Since gate and drain of M_3 are tited at the same node, the two non-linear small signal currents i_{d1} and i_{d3} are subtracted to define the output current i_{out}. Keeping only the terms to the cubic power results in $i_{out} = i_{d1}$ - i_{d3}, where:

$$
\begin{aligned}
i_{out} &= (g_{m_1} - b_1 g_{m_3})v_{g1} + (g_{2_1} - b_1^2 g_{2_3} - b_2 g_{m_3})v_{g1}^2 \\
&+ (g_{3_1} - b_1^3 g_{3_3} - g_{m_3}b_3 - 2g_{2_3}b_1 b_2)v_{g1}^3
\end{aligned}
\tag{3.49}
$$

From 3.49, one can see that the third and second order distortion terms of output current are reduced by the i_{d3} non-linearity factors. Although M_3 partially cancels the linear term as well, it does not appreciably degrade the amplifier gain because its transconductance is much lower than that of M_1, since its aspect ratio is smaller.

The IIP3 adjustment is made by changing the ratio of M_1 and M_3 sizes. Since it's difficult to model I-V curves in modern short-channel transistors, the designer can optimize this relation via electrical simulations with models provided by the foundry.

This topology allows the designer to adjust the third order cancellation through simulation.

3.2.5 Biasing

Transistors M_1-M_2 and resistor R_F compose the self-biased main amplifier of this topology. We introduce an extra load: transistor M_3 operates as a diode and is responsible for the linearity improvement. Its bias current is chosen to be 2%-3% of the current of M_1, in order not to interfere significantly in the gain of the main amplifier ($g_{m3} \ll (g_{m1}+g_{m2})$).

This bias has low sensitivity to process variation, since it's controlled by the feedback loop. Thus, this bias was used to bias all other amplifiers in the LNA circuit to guarantee a low sensitivity to process variations all over the entire LNA.

Therefore, the V_{GS} of all transistors that are responsible for amplification are the same, which is good, having all of them in the same inversion level gives similar non-linear behavior, improving the non-linear cancellation at the output.

The DC operating point of all layout components are shown in Fig. 3.11. The values shown are for a single multiplier only. So one has to multiply each device current value shown by the number of multipliers in each device. The main amplifier transistors have 20 multipliers, the main amplifier diode is a single transistor, the transistors of the first stage of the auxiliary amplifier have 10 multipliers each and, finally, the transistor of the second stage of the auxiliary amplifier has 2 multipliers. So, according to Fig. 3.11, the calculations for power consumption for the best-NF LNA are:

1. Main amp.: 155.3 μA*20 = 3.1 mA;
2. Aux. amp.1: 1.13 mA*10 = 11.3 mA;
3. Aux. amp.2: 1.3 mA*2 = 2.6 mA;
4. Total: 3.1 mA + 11.3 mA + 2.6 mA = 17 mA;

The best-IIP3 consumes 17.62 mA, with the same distribution among the transistors.
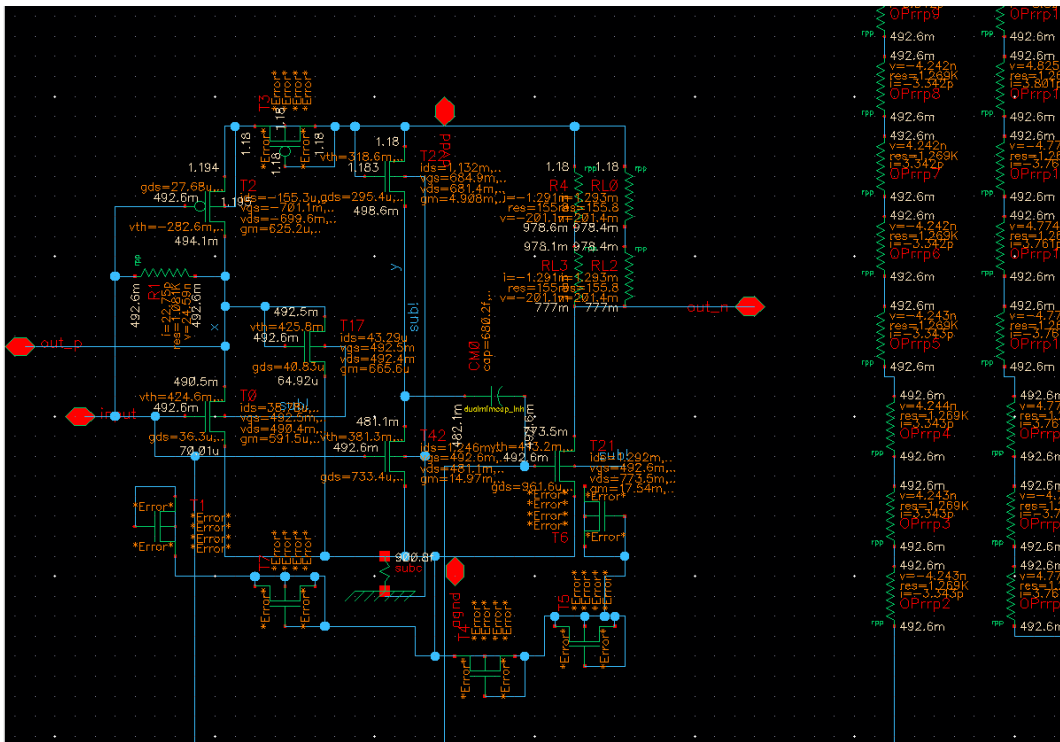
Figure 3.11: DC simulation results for the best-NF LNA.

3.2.6 Transistor and resistor sizing

The transistor sizing was done using all minimum L at first, then optimized via parametric simulation to have the best result possible. All g_ms were designed to get the lowest NF possible, best input impedance matching and a gain of 24 dB at the differential output (highest achieved maintaining IIP3 above 3.3 dBm in the entire wideband). The design started using the smallest transistor size possible and then increasing size as needed. A large transistor was needed mainly on the first amplifier in the auxiliary cascade (M_4), since it's hard to give a moderate gain on that configuration as explained in the gain sec-

tion. The main amplifier is not as large, being around 1/3 of the first stage of the auxiliary cascade, due to the impedance matching requirement (set by its g_m).

The resistor sizing was done using minimum size at one dimension (say width) and then putting all the resistance on the other dimension (length). This made possible to brake it into multipliers so as to have the best layout possible.

TABLE 3.1 shows the final transistor and resistor sizes after optimization, which is described next.

Table 3.1: Device sizing.

MOSFETs	LNA-best-IIP3			LNA-best-NF		
	W(μm)	L(μm)	g_m(mS)	W(μm)	L(μm)	g_m(mS)
M_1	194.88	0.12	58.3	162.4	0.12	12.18
M_2	210.24	0.44	14.85	175.2	0.44	12.5
M_3	4	0.12	1.1	2.4	0.12	0.66
M_4	339.84	0.15	116.08	424.8	0.15	149.7
M_5	125.12	0.4	39.67	156.4	0.4	49.08
M_6	91.2	0.12	33.24	91.2	0.12	34.4
Resistors	W(μm)	L(μm)	R(Ω)	W(μm)	L(μm)	R(Ω)
R_F	1	2.5	681	1	4.2	1081
R_{OUT}	5	2.7	142	5	3	155.8

3.2.7 Optimization
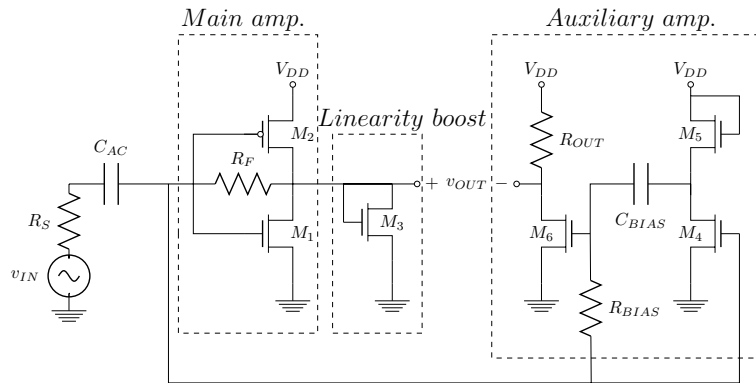
Figure 3.12: Final topology with diode at main amplifier's output to better control IIP3.

The LNA topology described here can be designed to achieve a given set of performance specifications parameters depending on the application requirements. In order to adjust its overall performance, the amplifier blocks that are highlighted by dashed rectangles in Fig. 3.12 can be optimized according to their influence in the amplifier parameters, as listed below.

- Main amplifier: bias, input impedance matching, Noise figure, Gain;
- Linearity boost diode: IIP3, NF;
- Cascaded auxiliary amplifier: NF, IIP3, Gain;

The equations used in overall optimization are repeated here for convenience:

$$R_{IN} = \frac{1 + g_{m_3} R_F}{g_{m_1} + g_{m_2} + g_{m_3}} \tag{3.50}$$

$$A_{V_{TOTAL}} = \frac{1 - (g_{m_1} + g_{m_2})R_F}{1 + (g_{m_1} + g_{m_2})R_S + g_{m_3}(R_F + R_S)} - \frac{g_{m_4}g_{m_6}R_{OUT}}{g_{m_5}} \quad (3.51)$$

$$NF_{TOTAL} = 1 + \frac{4R_S}{RF} + \frac{\overline{v_{n_{AUX}}^2}R_S}{kTR_F^2} \quad (3.52)$$

$$\begin{aligned} i_{MAIN_{OUT}} &= i_{out} = i_{d1} - i_{d3} \\ i_{MAIN_{OUT}} &= (g_{m_1} - b_1 g_{m_3})v_{g1} + (g_{2_1} - b_1^2 g_{2_3} - b_2 g_{m_3})v_{g1}^2 \\ &+ (g_{3_1} - b_1^3 g_{3_3} - g_{m_3}b_3 - 2g_{2_3}b_1 b_2)v_{g1}^3 \end{aligned} \quad (3.53)$$

where $i_{MAIN_{OUT}}$ is the main amplifier output current.

In order to explore the topology possibilities, two optimization directions can be taken, resulting in two different LNA versions (demonstrating the topology flexibility). One improving NF and the other improving IIP3. The two versions are shown below with the adjustment strategies shortly described.

1. Best NF: R_F must be increased, increasing gain; increase $g_{m_{1,2}}$ to compensate the S_{11} change; increase W of M_3 to compensate the IIP3 loss due to the gain increase and increase R_{OUT} to balance the output.

2. Best IIP3: increase W of M_3 until the maximum IIP3 is achieved; decrease R_F until the NF reach a given limit; decrease R_{OUT} to keep the output balanced.

The designer must observe that R_F can not be too high, since it must be counter-balanced by $g_{m_{1,2}}$ to keep input impedance matching, reflecting in power consumption increase. Also R_{OUT} should be low for lower thermal noise and allowing power gain if needed (being not too far from 50 Ω for good matching with the following block).

3.3 Layout design

The layout of the LNA was made using the IBM 130 nm process design kit (PDK), CMOS 8RF-DM (CMRF8SF, option DM). Several layout techniques were used and are explained in the subsections which follow. First, the PDK characteristics are presented.

3.3.1 The IBM 130 nm CMRF8SF-DM Process Characteristics

The CMOS8RF design kit comes in four versions that correspond to the technology last metal options: LM that has a Cu last metal. AM that has a Al last metal. DM that has a aluminum layer LY, a copper layer E1 followed by a aluminum top layer MA and OL that has a copper layer OL followed by a aluminum top layer LD. They offer designer a wide variety of metal options to choose from.

CMOS8RF is the IBM marketing name for this 0.13 μm technology that is used on the IBM Customer Connect (ICC). The design kit name is also CMRF8SF which is used in the PDK. Its short designation is 8RF.

IBM's CMOS8RF technology starting wafer is doped p-type. The lithography node is 130 nm. The technology utilizes shallow trench isolation (STI), into the silicon, to provide a dense isolation between devices. All diffusion and *polysilicon* are silicided for

low resistivity unless the silicide formation is intentionally blocked to form resistors with a higher specific sheet resistance.

IBM technologies use chemical mechanical polish (CMP) to planarize the surface of the inter-metal dielectric layers throughout processing. The surface planarity allows better lithographic resolution than could be obtained with topography present, enabling smaller layout ground rules as well as stacked contacts and vias for better wiring density. Automated fill routines used during mask data preparation create uniform active region (RX) and *polysilicon* (PC) pattern density for robust manufacturing.

Final chip passivation is formed by a sequence of oxide, nitride and *polyimide* films. The nitride serves as an ionic contamination barrier while the *polyimide* provides mechanical protection.

There are many devices available on this PDK, but only the ones used in this work will be discussed in this text.

3.3.1.1 Resistors

The technology offers an n-well resistor, a silicided *polysilicon* resistor, three types of unsilicided *polysilicon* resistors, an n-type junction resistor and a tantalum nitride resistor.

The resistors used to design R_F and R_{OUT} were all *oprppres*, RP poly over isolation. This layer has an 8% tolerance on its sheet resistance, which makes it a high precision resistor in comparison to other layers in this 130 nm CMOS process. Normally this type of resistors are not offered in cheaper PDKs. It was chosen because it had the sheet resistance near the resistance values needed for this work, occupying less area possible.

Only rectangular resistors are allowed. Resistors may not be of a "dog bone" configuration or have bends. Only two sets of contacts are permitted. Two resistors can not share a single n-well, RX, *polysilicon*, or *Kx* (TaN) shape by adding a third set of contacts in the middle of the resistor. Series resistors with separate shapes must be used instead.

3.3.1.2 Capacitors

The options associated with the MIM (Metal-Insulator-Metal) capacitor are dependent on the last metal option of the PDK.

The MIM capacitor uses an optional mask to define the capacitor metal top plate which is separated from the LY aluminum metal level bottom plate by a thin silicon nitride layer. The top capacitor plate is aluminum as well. The capacitance can be doubled if dual MIM capacitors are used.

The MIM capacitor plates must be tied to a silicon diffusion once the capacitor is fully formed. The *mimcap* pcell has two ground plane options. The NW option places the capacitor over an nwell. The SUB option is used when the capacitor is placed over p-well regions, and for cases where the MIM is over mixed n- and p-wells. The NW option can aid in the reduction of noise coupling to and from the substrate by using the well as a shield.

In this work, the capacitor used was a dual nitride MIM cap, to have double the capacitance density and reduce area. The SUB option was used as ground plane.

3.3.1.3 Transistors

The transistor used throughout this work was the regular FET, with thin oxide. The thin oxide FETs have a gate oxide thickness of 22 Å and a minimum drawn channel length of 0.12 μm (0.092 μm effective channel length). The typical and nominal threshold of this

device is 355 mV.

FETs are formed in either n-wells or p-wells created by a sequence of ion implants. Well implants are customized using the mask levels. The active region is defined by the RX mask level; STI isolation is everywhere outside of RX. Areas without DG will be 22 Ågate oxide for 1.2/1.5 V FETs. After the gate is formed, device junction characteristics are determined by the halo and extension implants. These are customized by derived layers BH/PH for 1.2 V NFET and PFET.

These FETs have a large variation of V_T with transistor length (L). Depending on the L value (from minimum to 1 um), V_T can vary from 250 mV to 350 mV or even 400 mV. So, using the quadratic model for the MOSFET DC current in this case was just a rough initial guess of transistor inversion level. After the initial guess, a parametric design based on BSIM4 simulated curves was needed in this work.

3.3.2   Device Matching

Device matching was not a concern in the LNA design in our case. As there was no current mirror, no differential pair, or similar circuits, no devices needed to be matched to each other.

However, dummy devices, such as transistors and resistors, were used to provide a equal behavior of the devices in the border and the ones located more to the center.

3.3.3   RF isolation

In order to improve RF isolation, guard-rings were used in every transistor of this design.

3.3.4   Pads

The pads were used when this work was submitted to fabrication. The pcell bondpad was used in the border and in the center of the chip. The LNA was connected directly to the bondpads provided by the PDK, with ESD protection also provided by the PDK.

Wire bond pads were available through the bondpad pcell with type set to WB.

Pads may be placed over BFMOAT, RX(active), or PC(polysilicon) ground planes. The BFMOAT ground plane is analogous to the BFMOAT region under an inductor. BFMOAT blocks the p-well implant in the region below the pad such that the parasitic capacitance to the substrate appears in series with the resistance of the un-implanted substrate. The ground plane terminal is connected to an RF ground using the BFMOAT IND dummy shape. The BFMOAT option provides the lowest parasitic capacitance of the three ground planes. And this was the option used in this work.

For accurate bond pad modeling, the layout of the region under the pad must be strictly controlled. The bondpad pcell contains a shape on the PAD DEV design level to trigger the extraction of the bond pad as a bondpad device for simulation and invoke model-related DRC rules. This shape prohibits placement of any devices or wires under the pad in order to ensure model accuracy.

The bondpad model and circuit symbol contains three terminals: the pad, the ground plane, and the substrate. The substrate terminal is the substrate region under the bond pad. None of the ground plane options supported in CMOS8RF couple significant energy into the substrate terminal, but the terminal was retained to allow for commonality and re-use of design kit elements among different IBM technologies. In the CMOS8RF bondpad model, the substrate terminal is decoupled from the pad and ground plane terminals using

a very large resistor. The terminal still needs to be connected for LVS purposes.

3.3.5 ESD Protection

A double diode ESD protection was used at the input and outputs of the LNA. A power clamp was also used between V_{DD} and GND. Fig. 3.13 shows each of these.
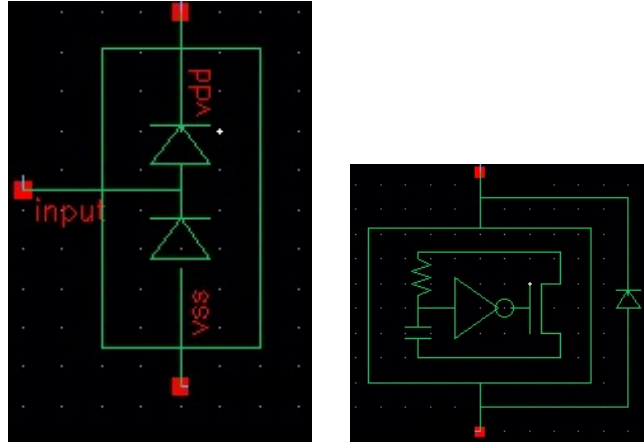
Figure 3.13: ESD protection used for fabrication.

3.3.6 Layout sent to fabrication

There were 2 preliminary versions of this LNA sent to fabrication. The topology is essentially the same, having only the bias of M_6 different from each other. This idea came up after the circuit was sent to fabrication. Also kx resistors were used in the version sent to fabrication.

There is one version which is going to be stimulated only through pads and another with ESD protection. This is going to show the effect of ESD protection on the LNA. The layouts sent to fabrication are shown in Fig. 3.14.

The fabricated chip is shown in Fig. 3.15, with both versions (one with PAD stimulated and another including ESD protection).

3.3.7 Final Layout

After the circuit was sent to fabrication, many improvements were made, including the idea of breaking it into 2 LNAs, one focusing in best-NF and the other on best-IIP3. Then, other 2 layouts were made to accomplish each of these ideas. The final layout for the best-NF LNA and the best-IIP3 are shown in Fig. 3.16 and 3.17. The final area achieved is 135 μm x 72 μm or 0.0096 mm^2 for each LNA.
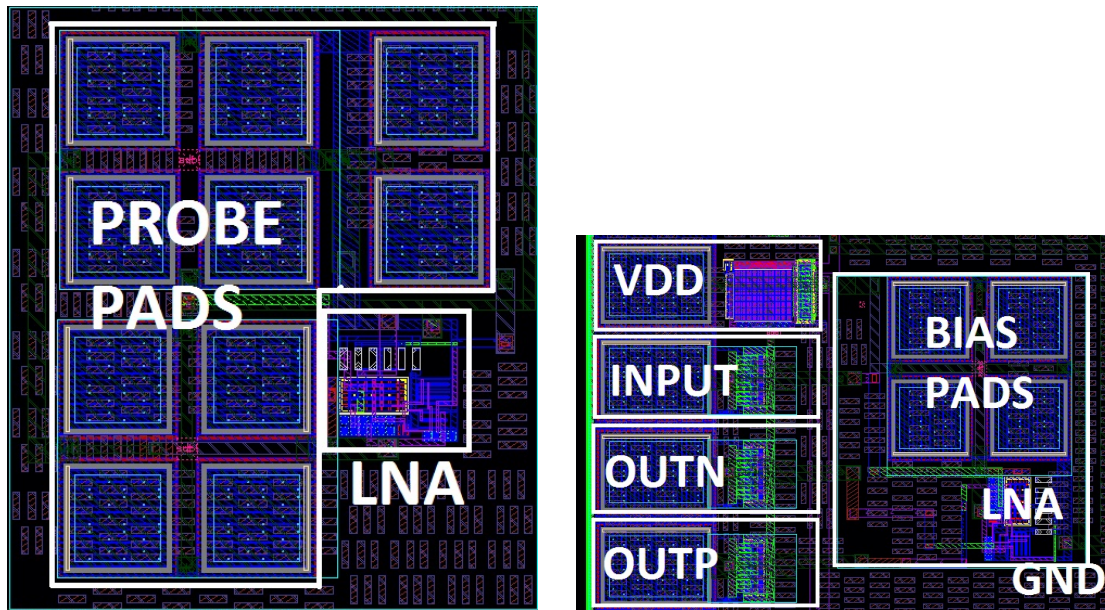
Figure 3.14: Pad-only LNA on the left and the other with ESD protection on the right. The small rectangles throughout the layout are the required metal fill features.
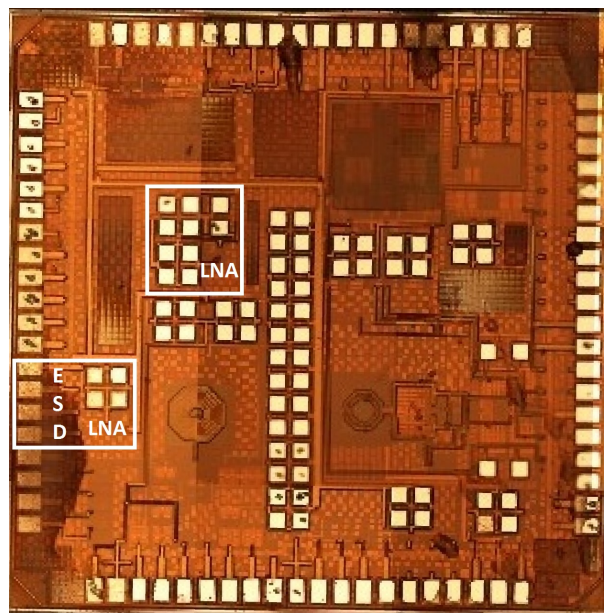
Figure 3.15: Picture of the chip fabricated, containing the LNAs (highlighted) and other designs by 7 graduate students from UFRGS.
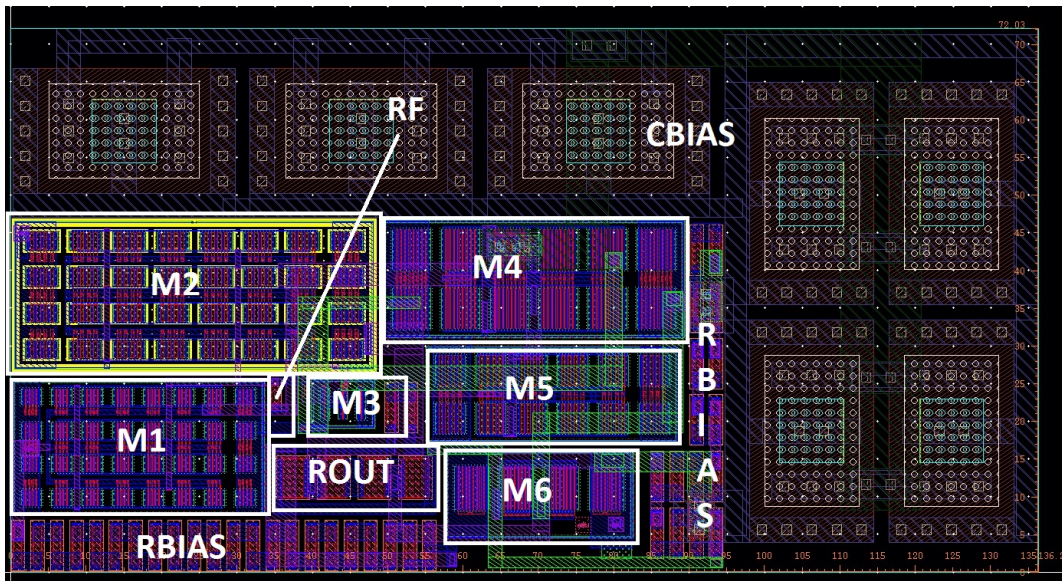
Figure 3.16: Layout of the LNA-best-IIP3.
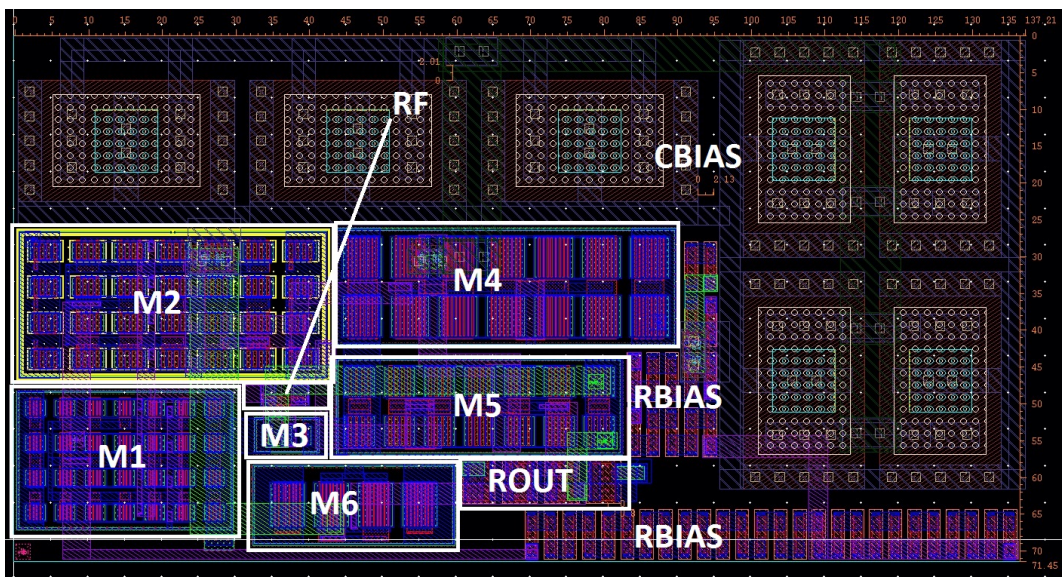
Figure 3.17: Layout of the LNA-best-NF.

4 POST-LAYOUT SIMULATION RESULTS

After layout and parasitic extraction, both versions of the LNA were simulated in Cadence Virtuoso Spectre, using the IBM 130 nm technology PDK. The simulation testbench is shown in Fig. 4.1 and includes estimated bondwire inductances (L_{BOND} = 2 nH), PAD capacitances (C_{PAD} = 1 pF) and external decoupling capacitors (C_{AC} = 350 pF). A load composed by R_L = 1.5kΩ and C_L = 100 fF was also considered and a good trade-off between R_{BIAS} and C_{BIAS} was achieved using 50 kΩ (total of 40 multipliers) and 33 pF (total of 7 multipliers), respectively. R_S = 50 Ω and V_{DD} = 1.2 V.

Package parasitic L_{BOND} and C_{PAD} were taken into account for the simulations, degrading the performance of the LNA. In case the LNA is designed to be used as an internal block of an application circuit, like in a RF front-end, results can be improved further. It was not included the effect of ESD protection for these versions. The simulation results for the circuit submitted for fabrication are not shown here, but their values are between the LNA-best-NF and LNA-best-IIP3 results. The focus of this work was changed from designing one LNA to the design of two LNAs, each focusing in different parameters such as NF and IIP3. Thus, the LNA sent to fabrication (before this decision was made) will be a proof of concept only but not used to estimate the circuit performance.

Unless otherwise specified, all nominal results are for the 27°C temperature. The simulations executed were DC and AC analysis, QPSS for IIP3 and 1dB-Compression Point, Transient, Noise and SP. All in Cadence virtuoso analog design environment. The layout extraction was RC with 1 μm resolution.
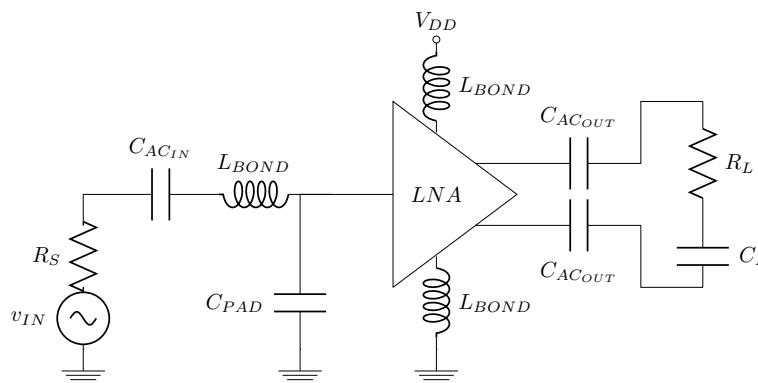
Figure 4.1: Test-bench used for post-layout simulations. The output was taken differentially at the output of each decoupling capacitor.

After a long time using this test-bench, it was realized that a π load would result in a better test-bench. It is recommended to proceed with π load configuration for future

works based on this one. Due to the lacking of time after realizing this, it could not be implemented in this work.

4.1 LNA-best-NF

The LNA-best-NF was designed to have sub-2 dB NF in the frequency band of 0.1 GHz - 1 GHz. The nominal results for post-layout achieved this goal, however, in Monte Carlo yield results for the same layout, considering 1000 samples, 91.8% of samples stayed < 3.7 dB, with 61.7% staying < 3 dB. This result was the best achieved after several redesigns.

4.1.1 Nominal results

Fig. 4.2 and 4.3 show that the topology can achieve NF < 2 dB in the 0.1 GHz - 1 GHz frequency band. The IIP3 goes into the vicinity of 0 dBm, given M_3 can not be large to cancel more nonlinear terms. Also, it results in a higher gain imbalance than the LNA-best-IIP3 at the output as shown in Fig. 4.3. This gain imbalance is due to layout imperfections. Several redesigns were done to achieve the least gain imbalance. The S_{11} also increased being $<$-11 dB, still having a margin below -10 dB. The gain of this LNA-best-NF stays > 24.7 dB in the entire band.

This sub-2dB NF is achieved, having a near 0 dBm IIP3, considering the PAD capacitances, bondwire inductances and external decoupling capacitances. Usually in other topologies, NF is higher than 3 dB for 0 dBm or positive IIP3s. The IIP2 is highly positive due to differential output.
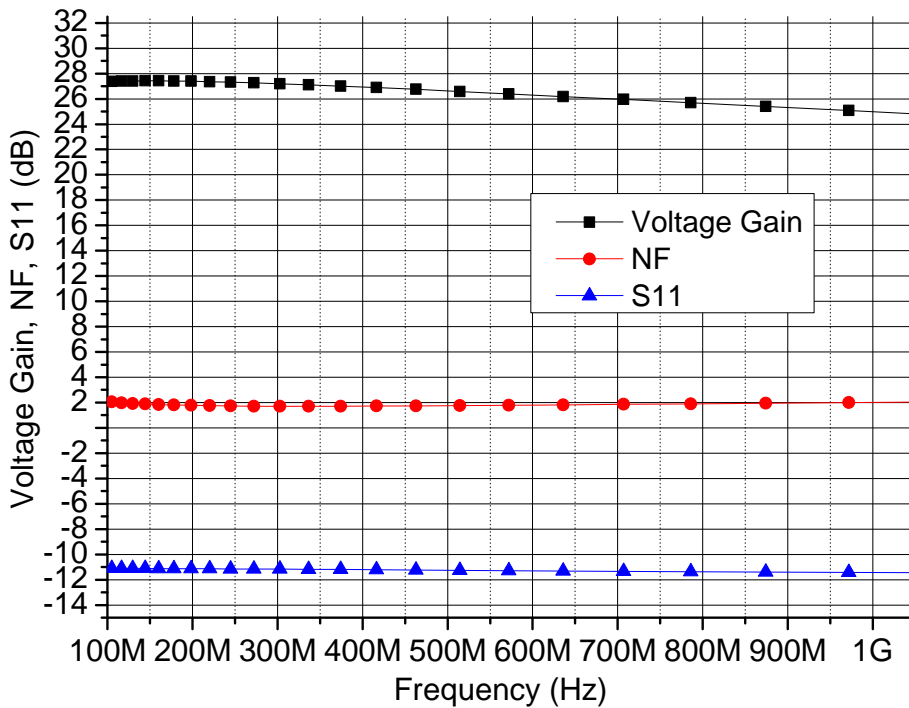
Figure 4.2: Post-layout voltage gain, NF and S11 simulation results for the sub-2 dB LNA.

The 1 dB compression point was -20 dB as shown in Fig. 4.4. This compression point does not follow the usual 9.6 dB distance from the IIP3, because there is a non-linear
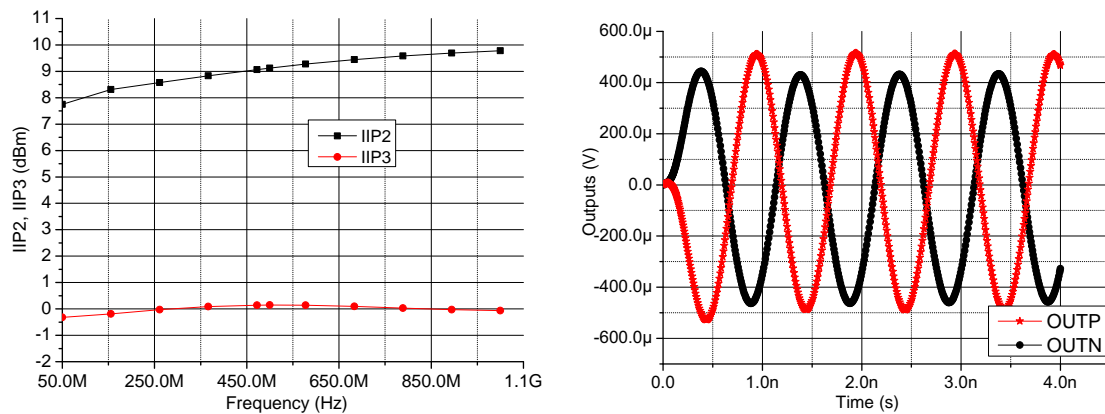
Figure 4.3: Post-layout IIP3 and IIP2 simulation results of LNA-best-NF on the left and transient simulation results showing the outputs on the right.

terms cancellation scheme in this circuit. The IIP3 is an effect due to intermodulation, while 1 dB compression point relies on the circuit characteristic of gain compression. So it is perfectly possible, through non-linear terms cancellation schemes, to have a high IIP3 with low 1 dB compression point. As both are related to different phenomena, one does not harm the other. Having a low 1 dB compression point only limits the input power for the LNA.
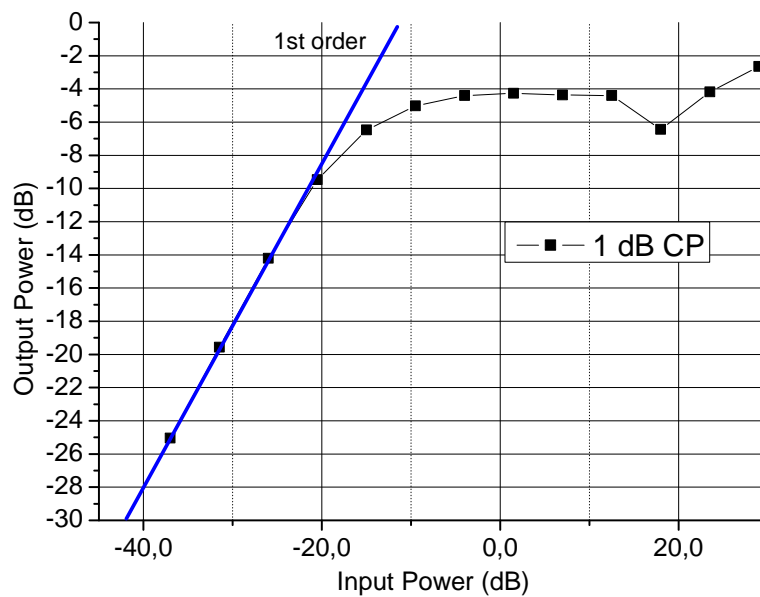
Figure 4.4: 1 dB compression point at -20 dB for the LNA-best-NF.

The K-factor was higher than 1 in the entire band as shown in Fig. 4.12. The LNA of this work is designed to give voltage gain, so if one needs it to give power gain, it's necessary to build up an output network for output impedance matching. As the output is not far from 50 Ω (it varies from 20Ω to 120Ω), this should not be difficult.

4.1.2 Temperature sensitivity

LNA-best-NF has presented low sensitivity to temperature variations in the military range for gain, S_{11} and NF, as shown in Fig. 4.6. The IIP3 goes to a minimum of -7.5 dBm
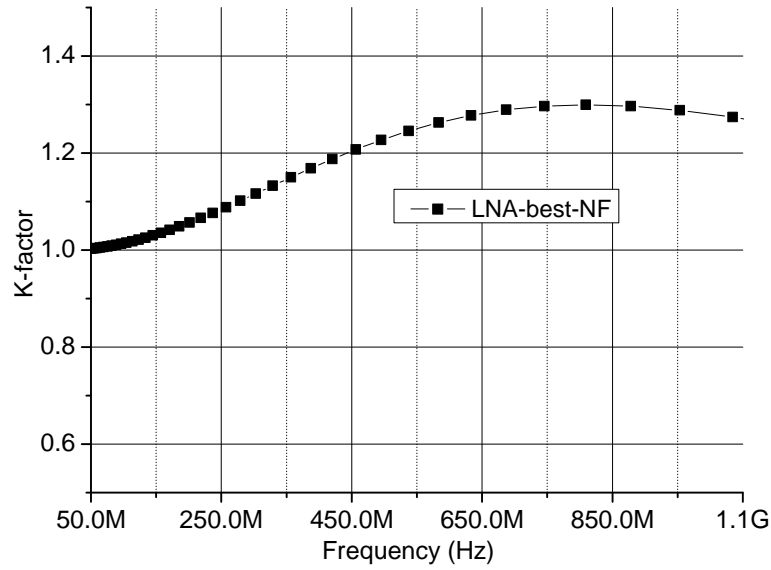
Figure 4.5: Although not having a meaningful value due to output impedance mismatching, the K-factor remains higher than 1.

at the -55°C temperature, being near or above 0 dBm in the range of 27-80°C temperatures as shown in Fig. 4.6. This LNA-best-NF is not suited for negative temperatures operation.
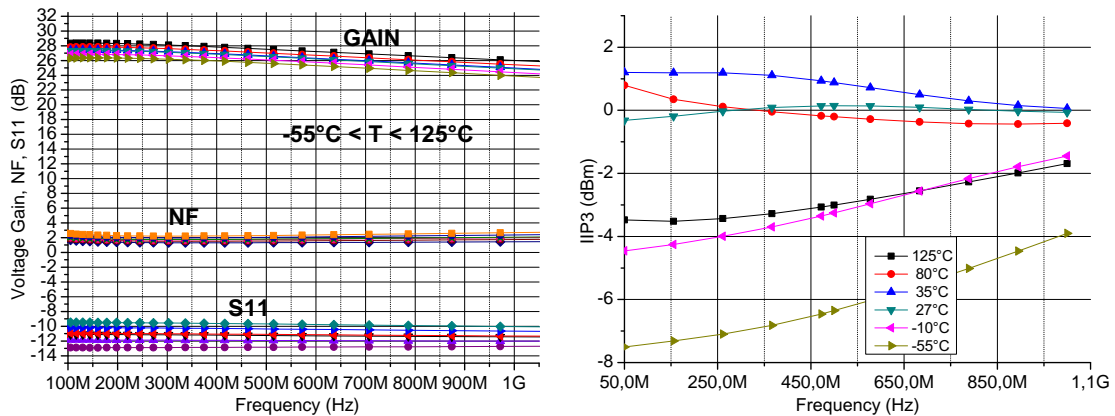
Figure 4.6: Temperature variation impact on gain, NF and S_{11} of the LNA-best-NF on the left and the impact on IIP3 on the right.

4.1.3 Fabrication process sensitivity

Monte Carlo simulations were done for the LNA-best-NF with 1000 samples to verify the impact of process variations and mismatch. Fig. 4.7 shows the result for gain, S_{11} and NF on the left and worst case NF (@50 MHz) on the right. Fig. 4.8 shows the results for worst case IIP2 (left) and IIP3 (right), which is @1GHz.

Monte Carlo simulation results show very low sensitivity to process variations and mismatch. It was achieved a $\mu = 3.03$ dB, with $\sigma = 0.54$ dB for the worst case NF (@50 MHz). 91.8% of 1000 samples stayed < 3.7 dB, with 61.7% staying < 3 dB. The worst case IIP3 stayed above 0 dBm for 98.8% of 1000 samples. The worst case IIP2 stayed > 5.6 dBm for 100% of 1000 samples. Also, for 100% of samples, the gain was > 22.5 dB

and $S_{11} < $ -11 dB. One can notice, the IIP3 of this version of the LNA does not go higher than 4 dBm.
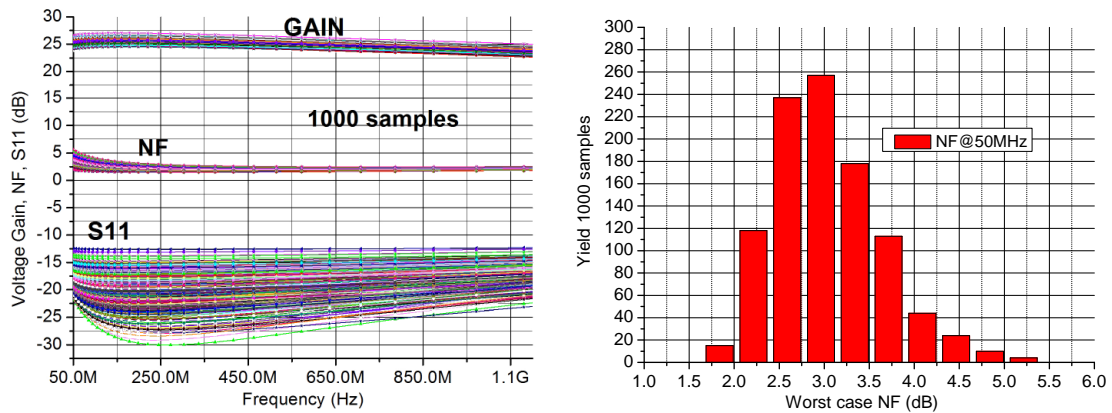
Figure 4.7: Monte Carlo simulation results for 1000 samples. The gain, NF and S_{11} of the LNA-best-NF on the left and the worst case NF on the right for the same amplifier.

Figure 4.8: Monte Carlo simulation results for 1000 samples. The worst case IIP2 of the LNA-best-NF on the left and the worst case IIP3 on the right for the same amplifier.

4.2   LNA-best-IIP3

The LNA-best-IIP3 was designed to have the best IIP3 that this topology could achieve on IBM 130 nm technology. It achieved an IIP3 $>$ 3.3 dBm in the entire 0.05 GHz to 1 GHz frequency band.

4.2.1   Nominal results

The results of the LNA designed for best IIP3 are shown in Fig. 4.9 and 4.10. The gain, NF and S_{11} are nicely flat in the 0.05 GHz - 1 GHz frequency band.

The S_{11} was designed to be less than -16 dB so as to have a good margin for bondwire inductance variation. The topology achieves a voltage gain above 23.7 dB, NF $<$ 3.6 dB in the entire frequency band. The IIP3 was above 3.3 dBm in the entire frequency band. The IIP2 is above 5.5 dBm.
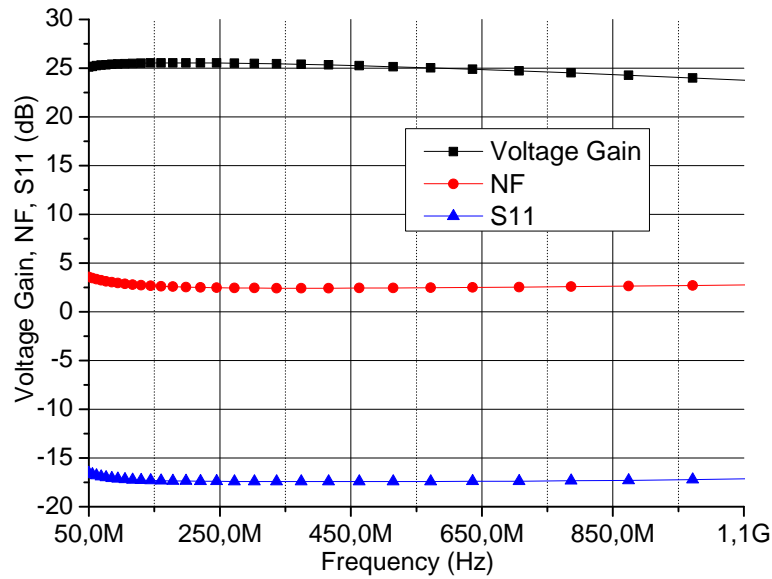
Figure 4.9: Post-layout voltage gain, NF and S11 simulation results of LNA-best-IIP3.
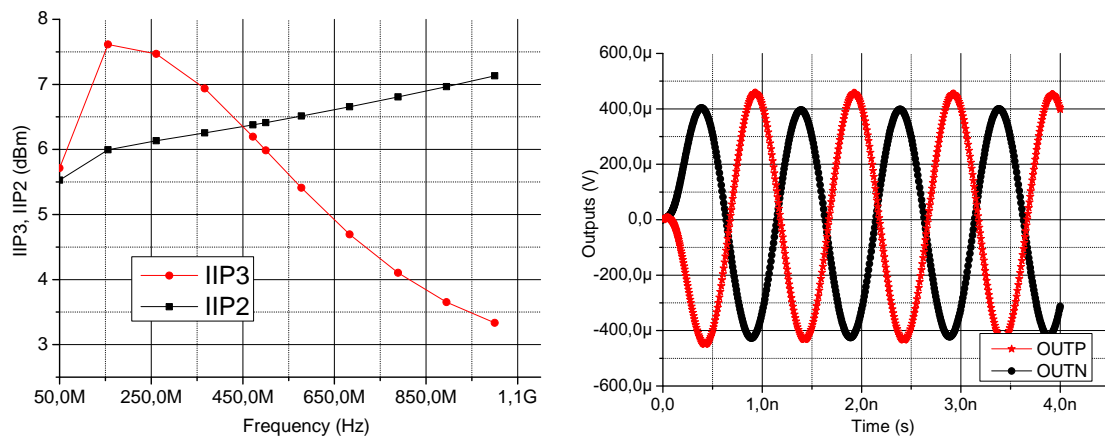
Figure 4.10: Post-layout IIP3 and IIP2 simulation results of LNA-best-IIP3 on the left and transient simulation results showing the outputs on the right.

The 1 dB compression point was -19 dB as shown in Fig. 4.11. This compression point does not follow the usual 9.6 dB distance from the IIP3, because there is a non-linear terms cancellation scheme in this circuit. The IIP3 is an effect due to intermodulation, while 1 dB compression point relies on the circuit characteristic of gain compression. So it is perfectly possible, through non-linear terms cancellation schemes, to have a high IIP3 with low 1 dB compression point. As both are related to different fenomena, one does not harm the other. Having a low 1 dB compression point only limits the input power for the LNA.

The K-factor was higher than 1 in the entire band as shown in Fig. 4.12.

4.2.2 Temperature sensitivity

The highly linear LNA has presented low sensitivity to temperature variations for gain, S_{11} and NF, as shown in Fig. 4.13. The IIP3 goes to a minimum of -3 dBm only in the -55°C temperature, being above 0 dBm in all other temperatures.
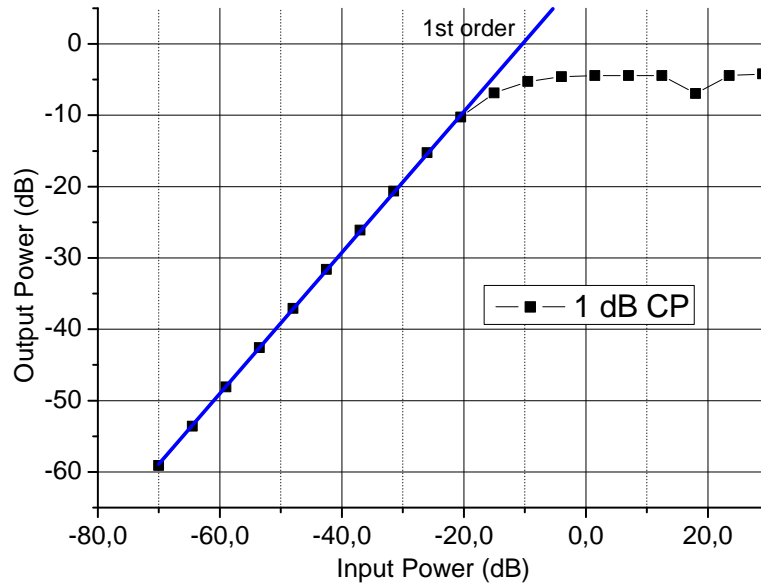
Figure 4.11: 1 dB compression point at -19 dB for the LNA-best-IIP3.
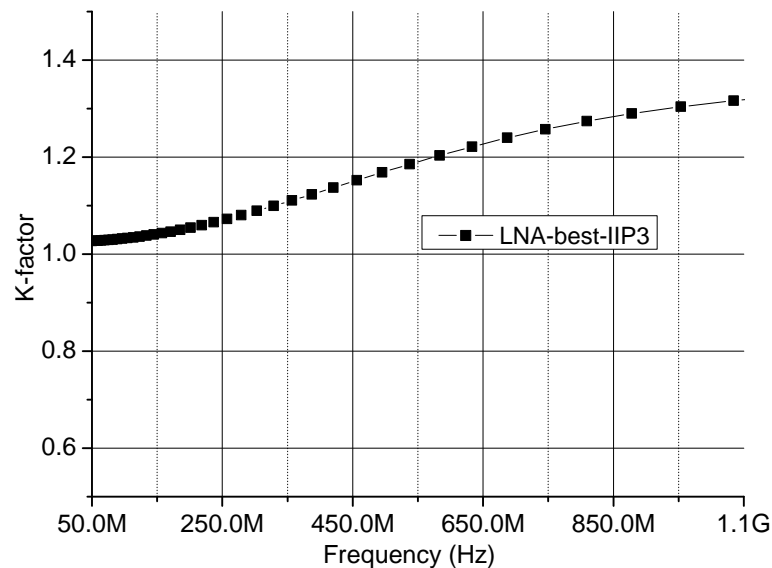
Figure 4.12: Although not having a meaningful value due to output impedance mismatching, the K-factor remains higher than 1.

The gain stays well above 23.7 dBm and the NF stays < 4 dB. The S_{11} keeps < -14 dB.

4.2.3 Fabrication process sensitivity

Monte Carlo simulation was done for the LNA-best-NF with 1000 samples to verify the impact of process variations and mismatch. Fig. 4.14 shows the result for gain, S_{11} and NF on the left and worst case NF (@50 MHz) on the right. Fig. 4.15 shows the results for worst case IIP2 (left) and IIP3 (right).

Monte Carlo simulation results show very low sensitivity to process variations and mismatch. It was achieved a $\mu = 3.7$ dB, with $\sigma = 0.7$ dB for the worst case NF (@50 MHz). Having 96% of 1000 samples with NF < 4.7 dB. The worst case IIP3 stayed above
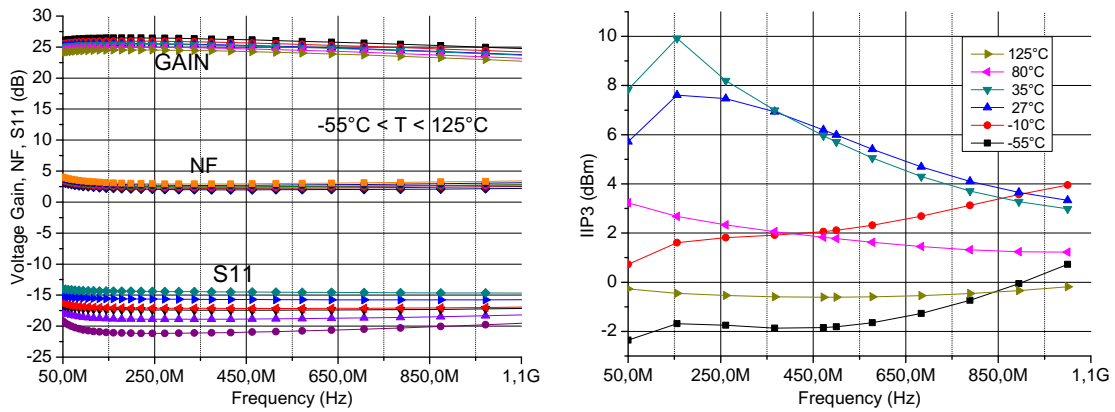
Figure 4.13: Temperature variation impact on gain, NF and S_{11} of the LNA-best-IIP3 on the left and IIP3 on the right.

0 dBm for 98.2% of 1000 samples. The worst case IIP2 had μ = 7.2 dB and σ = 0.67 dB. For 100 % of samples, the gain was > 22.7 dB and S_{11} < -11.7 dB. The IIP3 of this version achieves values higher than 4 dBm.
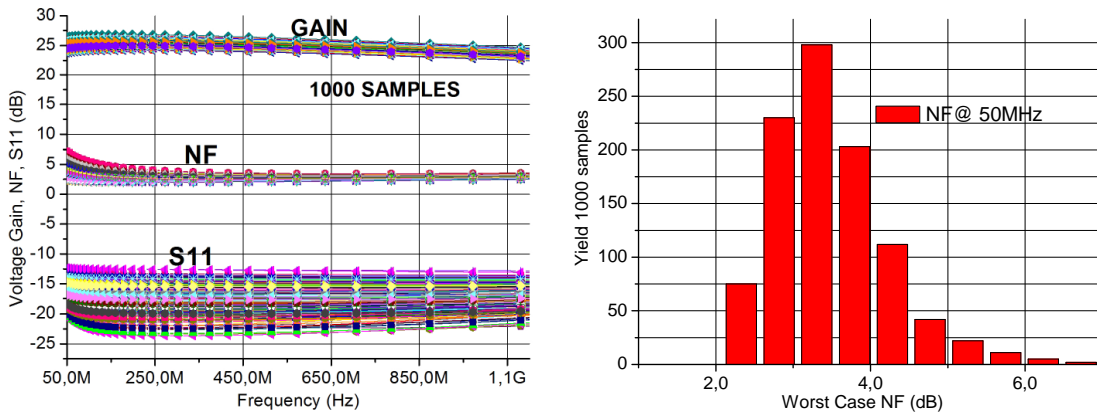
Figure 4.14: Monte Carlo simulation results for 1000 samples. The gain, NF and S_{11} of the LNA-best-IIP3 on the left and the worst case NF on the right for the same amplifier.

4.3 Comparison with State of the Art LNAs

Table 4.1 presents a comparison with 12 other LNAs from the literature, restricting those designed and fabricated in 130 nm down to 65 nm process technologies.

It is noticeable that the LNA-best-IIP3 has the second best S_{11} margin. Only WANG et al. (2012) has better S_{11} margin, however, their design has a 10 dB smaller gain and it's not a balun. This is significant in view of expected variations in the bondwire inductances. The LNA-best-IIP3 has the highest gain and lowest NF among the positive IIP3 LNAs, which relaxes the requirements for the mixer connected at the output. This work and references, CHENG et al. (2012) and BLAAKMEER et al. (2008) have the smallest silicon area. This topology achieves a high IIP3 even at 50 MHz, for which is difficult to have both high linearity and low NF. The transient simulation results show a good gain
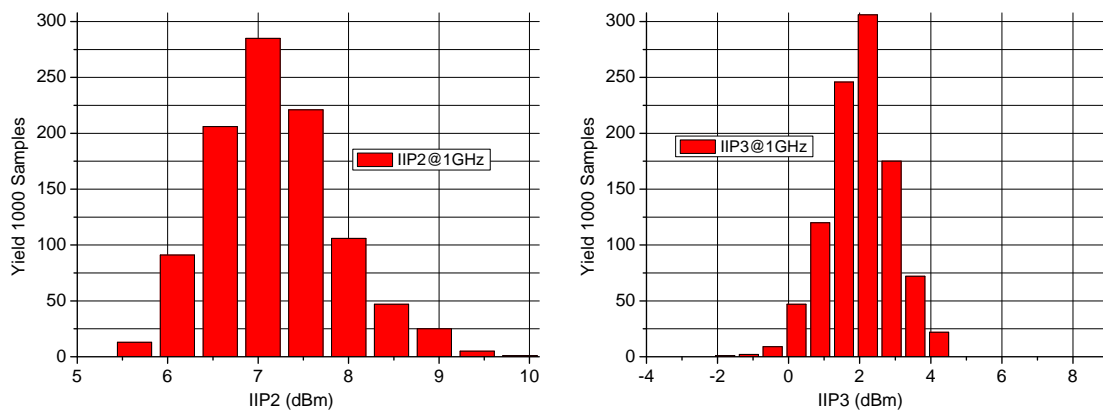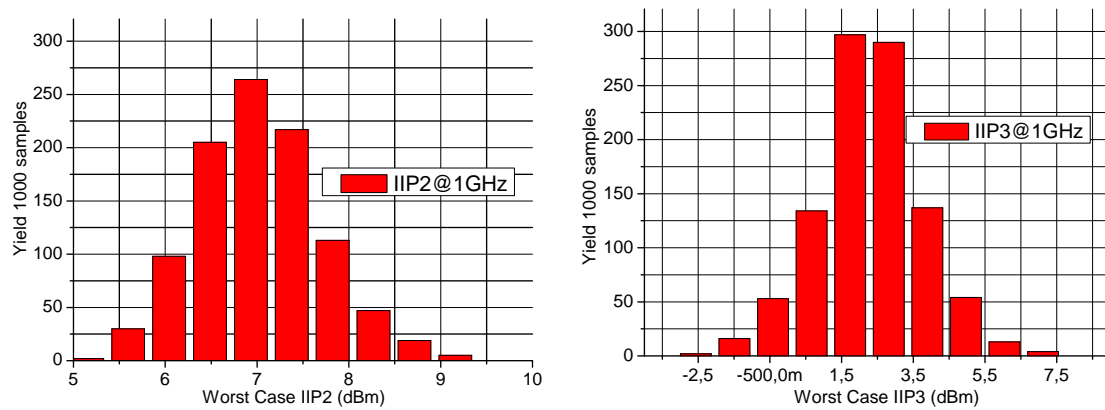
Figure 4.15: Monte Carlo simulation results for 1000 samples. The worst case IIP2 of the LNA-best-IIP3 on the left and the worst case IIP3 on the right for the same amplifier.

balance for the LNA-best-IIP3, making it a good balun.

The second design, LNA-best-NF, and YOUSSEF; ISMAIL; HASLETT (2010) are the only ones achieving NF < 2 dB in the entire 100 MHz - 1GHz band. Our design LNA-best-NF features a 10 dBm higher IIP3 than YOUSSEF; ISMAIL; HASLETT (2010), hence it is the only sub-2dB LNA to achieve a IIP3 >-0.3 dBm in the entire band.

This LNA circuit has potential to improve bandwidth even more if denser technology nodes, with higher f_T transistors, such as 65 nm, are used.

For a better comparison, it was used a Figure of Merit (FoM) based on the ITRS 2011 System drivers chapter. The ITRS was adapted to include the area of the circuit. The smaller the area, the higher the FoM. It is showed below:

$$FoM = \frac{G * IIP3 * BW}{(F - 1) * P * AREA} \tag{4.1}$$

Where G is the gain, denoted VFoM when considering Voltage Gain and PFoM when considering Power Gain. IIP3 is the input referred third order intercept point in Watts, BW is the bandwidth in Hertz, F is the noise factor, P is power consumption in Watts and Area in metters. The comparison is showed in table 4.2.

62

Table 4.1: Comparison with state of the art wideband LNAs. The * means Power gain values, [2] means measured and [3] post-layout simulation.

Ref.	BW(GHz)	Gain(dB)	NF(dB)	$S_{11}(dB)$	IIP2(dBm)	IIP3(dBm)	Power(mW)	Area(mm^2)	Tech(nm)	BALUN
Sinencio[2](2009)	1.5-8.1	8.6-11.7*	3.6-6	<-9	7.6-23	11.7-14.1	2.62(core)	0.58	130	NO
Lee[2](2010)	0.072-0.85	14.5*	3.6	-	38	2.5	11.52	0.08	130	NO
Nauta[2](2012)	0.1-1	11-18	2.8-4	<-10	-	1-12	12.3	0.003	160	NO
Y−JChen[2](2010)	0.04-1.2	16.4*	2.1-3.4	<-10	-	0	14.4	0.036	180	NO
Haslett[2](2010)	0.17-0.9	36.5	1.5	-	-	-11	18	-	65	YES
Swart[3](2011)	0.05-5	12*	1.4-2.4	<-10	-	-7.89	15	0.08	130	NO
Razavi[2](2010)	0.05-10	18-20	2.9-5.9	<-10	14-19.5	-11.2-(-7)	22	-	65	YES
Nauta[2](2008)	0.2-5.2	13-15.6	<3.5	<-11	>20	0-4.8	21	0.009	65	YES
Bakhtiar[2](2012)	0.32-1	18-23.5	2.2-2.7	<-10	-	0@600MHz	15.3	0.1	180	NO
Chueh[2](2009)	0.15-1	22-24(14.3*)	2.5-2.9	<-10	10	-0.5	3	0.06	130	YES
DongguIm[2](2013)	0.5-1.1	10*	2.9-3.1	-	12.5	7.5	18	0.07	180	NO
C−CChen[2](2014)	3-10	13.7*	2.3	<-10.7	-	-0.2	18	0.39	180	NO
Zhigong[2](2012)	0.1-8.5	16.3	2.7-3.9	-	-	-4	10	-	90	NO
Zhang[2](2007)	2-9.6	11	3.6-4.8	<-8.3	-	-7.2	19	0.05	130	NO
Andreani[3](2010)	0.2-1	17.7	4-4.2	<-16	-	-9.8	7.35	0.007	130	YES
Min[2](2007)	0.05-0.86	15	2.3-2.9	<-24	-	8.3@500MHz	7.2	-	180	NO
Taniguchi[2](2008)	3.1-13.9	7.8-12.3*	2.7-3.3	-	-	-6.4	2.5	0.1	180	NO
Liu(2012)	0.05-10	10.5*	2.7-3.3	<-11	-	-3.5	13.7	0.02	65	NO
Ourbest−IIP3	0.05-1	**23.7-25.5(19.1*)**	2.4-3.6	**<-16**	5.5-7.1	**3.3-7.6**	21	**0.0096**	130	**YES**
Ourbest−NF	0.1-1	**24.7-27(19.7*)**	**<2**	<-11	7.8-9.7	**-0.3-0.1**	20	**0.0097**	130	**YES**

Table 4.2: Figure of Merit (FoM) table. VFoM considers Voltage gain and PFoM considers Power gain.

Ref.	VFoM	PFoM	BALUN
Sinencio[2](2009)	-	12.8	NO
Lee[2](2010)	-	0.33	NO
Nauta[2](2012)	2.82	-	NO
Y−JChen[2](2010)	-	1.39	NO
Swart[3](2011)	-	0.28	NO
Nauta[2](2008)	5.32	-	YES
Bakhtiar[2](2012)	0.1	-	NO
Chueh[2](2009)	0.86	1.45	YES
DongguIm[2](2013)	-	0.28	NO
C−CChen[2](2014)	-	0.32	NO
Zhang[2](2007)	0.04	-	NO
Andreani[3](2010)	0.08	-	YES
Taniguchi[2](2008)	-	1.94	NO
Liu(2012)	-	2.11	NO
Ourbest−IIP3	**8.08**	**34.8**	YES
Ourbest−NF	**2.13**	**8.88**	YES

5 TEST AND MEASUREMENTS

In order to test and measure the LNA circuits previously described, designed and sent to fabrication, a PCB is being designed. As the circuit is going to be packaged, there will be additional parasitic devices due to the package itself. This is going to change the S_{11}, which might leave the circuit unmatched to the 50 Ω input impedance. The LNA circuit has a margin for S_{11}, and an external matching network can be easily designed if needed.

At the highest operating frequency, i.e., 1 GHz, the wavelength is 30 cm, which is of the order of the PCB distances (2-10cm). Thus, one has to match the PCB tracks to the chip and to any other high frequency device or equipment involved in the signal path.

If the transmission line behavior is very small, even if the package changes the S_{11}, it might still work without a perfect matched input impedance. It will depend on how long the PCB tracks will be and how much the package resistance and capacitances are going to be. One must know that transmission line behavior should not be neglected in PCBs for MHz-GHz frequency signals, because it will steal power from the signal, degrading gain, NF and IIP3 measurements. Even if the degradation is small, one has to remember that there will already be other degradations from parasitic devices from package, PCB, connectors, etc. So, it is better to avoid every degradation possible, to measure the best performance possible of the LNA.

5.1 PCB Design

In this section, the PCB design is discussed taking into account RF effects, such as crosstalk, reflections, skin effect, etc. A careful design must be done, paying attention to board dielectric constant, power bus decoupling, minimum space between traces, etc.

5.1.1 Basic Concepts

The design of the PCB has to be done very carefully, since the LNA goes into the RF frequencies (up to 1 GHz). This requires specific layout techniques and strategies in order to minimize signal attenuation and power bus coupling.

The RF layout encompasses the design of analog based circuits in the range of hundreds of MHz to many GHz. The frequency range is roughly in the 500 MHz - 2 GHz band. A Microwave layout can be considered to be above 2 GHz. The LNA designed in this work is in the RF range. RF signals are very sensitive to noise, ringing and reflections and must be treated with great care. The PCB needs complete impedance matching (to a given impedance $Z_0 = 50$ Ω) from the input, through the trace line to the output. All should be matched in order to avoid reflections and hence signal attenuation.

A transmission line is any pair of wires or conductors used to move energy from point

A to point B. The impedance Z_0 is a function of the signal conductor width, thickness and dielectric constant of the PCB material. All signal return currents follow the path of least impedance (least inductance in RF). Whenever one neglects to provide low impedance return path, the signals will find a path on their own, which might not be the desired path.

The signal wavelength in free space is given by:

$$\lambda = \frac{c}{f} \tag{5.1}$$

where c is the speed of light and f is the frequency. Thus, λ = 30 cm at 1 GHz in free space. However, if the signal travels on a dielectric, it yields:

$$\lambda = \frac{c}{f}\frac{1}{\sqrt{\epsilon}} \tag{5.2}$$

which is 15.6 cm for the material RO 4003 from Rogers corporation, with $\epsilon = 3.7$. Thus, the signal critical length, which is how long the PCB trace can be before paying attention to impedance control, is:

$$L_{crit} = \frac{c}{f}\frac{1}{\sqrt{\epsilon}}\frac{1}{16} \tag{5.3}$$

which is 9.7 mm at 1 GHz for the RO 4003 material. In this work, the RO 4003 material was chosen because it has a controlled dielectric constant along the PCB. Materials such as FR4 should be avoided for RF signals traversing over the PCB.

5.1.2 Layout Techniques and Strategies

The first thing to do when designing a PCB for RF or Microwave is to divide the PCB into sections or groups of circuits (VCO, Amps, Mixers, etc). One should always start by placing the high frequency components first, in order to minimize trace lengths of each RF route. The highest frequency components should be as near as possible to connectors and pins.

As the trace impedance Z_0 is a critical factor in the effort to control reflections, always match driver and load. Traces shorter than 1/20th λ long can neglect impedance matching (as they are not considered transmission lines for the frequency of interest). Inductors, if any, should be placed far from each other or perpendicular to each other. All trace routes within a group should not be routed into adjoining groups.

An extremely important detail is to design the chip pin-out, already with the PCB design in mind. In this work, the top level of the chip had 7 Master student works and a simple floorplaning was made, putting one design next to the other, resulting in a bad pin assignment for the PCB design. For example, a Master student had 13 pins in a row dedicated to DC measurements of voltage and current references, with the LNA pins of this work all next to each other in the same row. If a PCB-oriented design was intended, the LNA pins should be placed 2 pins away from each other. As when measuring the LNA, the pins of the other Master student work would be grounded, the LNA pins would be spaced on a signal-ground-signal pattern, ideal for RF measurements! However, for the moment, it has a signal-signal-signal pattern, which is incredibly challenging to design the PCB.

Another extremely important strategy is to place a ground plane in the layer below the RF traces. This drastically reduces signal coupling between traces. These traces should be kept away from each other by minimum distances or surrounded by ground lines as shielding to avoid crosstalk. In this work, the PCB is going to have double layer, in which one is going to have the RF traces and the other a ground plane. The RF traces layer unused areas are going to be poured with ground. Both ground planes are going to be connected to each other with many vias 1/20th λ apart from each other.

Long traces can be an Antenna for radiation or EMI (Electro-Magnetic Interference) or even reception of noise, as the ideal antenna trace is 1/4 λ. The corners of traces should also have a given shape, in order to reduce signal attenuation by corners. Fig. 5.1 shows how they should be done.
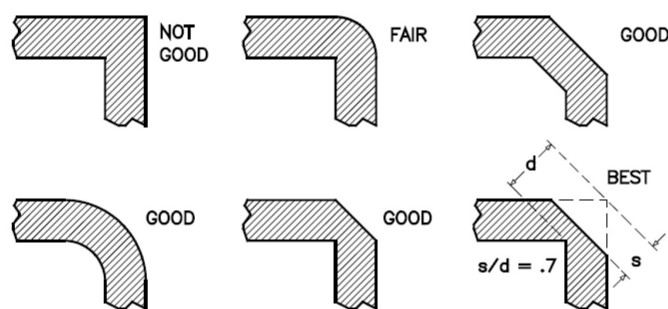
Figure 5.1: How corners shape should be on an RF PCB.

5.1.3 Power Bus

If possible the power traces should be routed on a separate plane between ground planes on a 4-layer PCB. However, in this work, as a double layer PCB is going to be used in the test set up, power is going to be routed on the same layer as the RF traces. This requires more attention to how these traces are going to block interference from the power lines.

The power decoupling consists of a low-pass filter with several capacitors to cover a broad range of frequencies and currents. Fig. 5.2 shows how it is supposed to be done. The smallest capacitor should have a resonating frequency equal or above the highest frequency of operation. This ensures a capacitance behavior inside the band. If this is not accomplished, the capacitors are going to become inductors in high frequencies and reverse their utility. RF capacitors should be bought, they are more expensive than standard ones. However, as only the small ones needs to have such a high resonating frequency, all the others can be standard SMD ceramic capacitors. In this work, the capacitors array is starting at 10 pF and going up to 10 uF with 10x steps. Only the 10 pF, 100 pF capacitors are RF capacitors (to cover 1 GHz, 100 MHz), all the others are standard (usually cover 10 MHz and below). The largest value should carry the highest current required by the IC. The smallest capacitors should be places as close as possible to the IC power pin. Power capacitors should have wide traces connecting them. Use many vias if necessary.

5.1.4 Skin Effect

The skin effect of traces can attenuate signals, given the effective area is reduced at high frequencies (signal is rather going through borders than in the middle). The trace
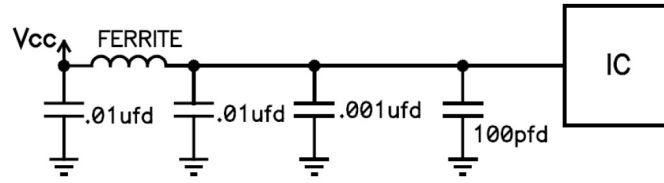
Figure 5.2: RF PCB power bus scheme.

resistance is altered by the skin depth (SD) which reduces with increasing frequency, reducing the effective area as show below:

$$
\begin{aligned}
R &= \rho \frac{length}{Area_{eff}} \\
Area_{eff} &= 2(w+t)SD \\
SD &= \frac{66}{\sqrt{f}}
\end{aligned}
\tag{5.4}
$$

where, w is the trace width, t is the thickness and SD is given in mm. For a microstrip line (most common transmission line trace), R should be multiplied by an adjusting factor, which is 1.7. The attenuation in Volts is then given by R*I, where I = $\frac{V_{driver}}{V_{load}}$. For the worst case of this work (at 1 GHz), the signal attenuation can be, considering copper resistivity, a trace of 10 mm x 1 mm and 1 mV_{peak} amplitude:

$$
\begin{aligned}
I &= \frac{10^{-3}}{50} = 2 \times 10^{-5} \\
SD &= \frac{66}{\sqrt{10^9}} = 0.0021 \\
Area_{eff} &= 2(10^{-3} + 0.07 \times 10^{-3})0.0021 = 4.5 \times 10^{-6} \\
R &= 16.78 \times 10^{-7} \frac{10^{-2}}{4.5 \times 10^{-6}} = 0.004 \\
Atten &= 2 \times 10^{-5} \times 0.004 = 80nV
\end{aligned}
\tag{5.5}
$$

One can notice that the resistance R should be 0.002 if there was no skin effect. Thus, it doubled the trace resistance in this case. If longer traces, such as 100 mm, are to be used, 800 nV of signal attenuation is expected. As the LNA of this work has a high gain (24 dB) the signal amplitude is expected to be near 1 mV differentially.

5.1.5 Designed PCB in Agilent ADS

As our chip have multiple circuits (of 7 Master Students), this PCB is supposed to test the LNA and the Mixer, which are the RF circuits of the chip. The PCB was designed using the software Agilent Advanced Design Systems (ADS) in order to be able to simulate S-parameters. According to this simulation results, the designer is able to estimate coupling between traces, which is important in RF design. The designed PCB is shown in Fig. 5.3 and Fig. 5.4.

The PCB was designed with double ground plane (top and bottom layers). A Balun was used to convert the LNA differential output into single. All routing was supposed to be done in the top layer, however, due to the bad pin assignment of our chip (as explained
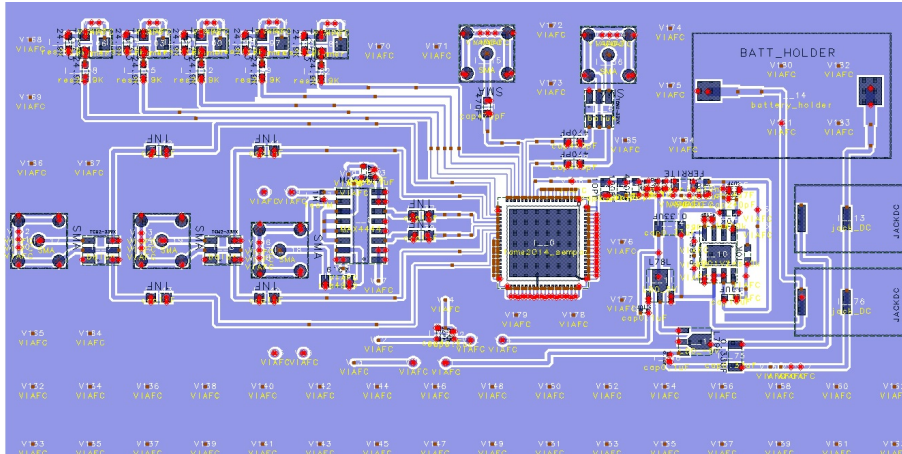
Figure 5.3: Designed PCB in ADS (top layer view).
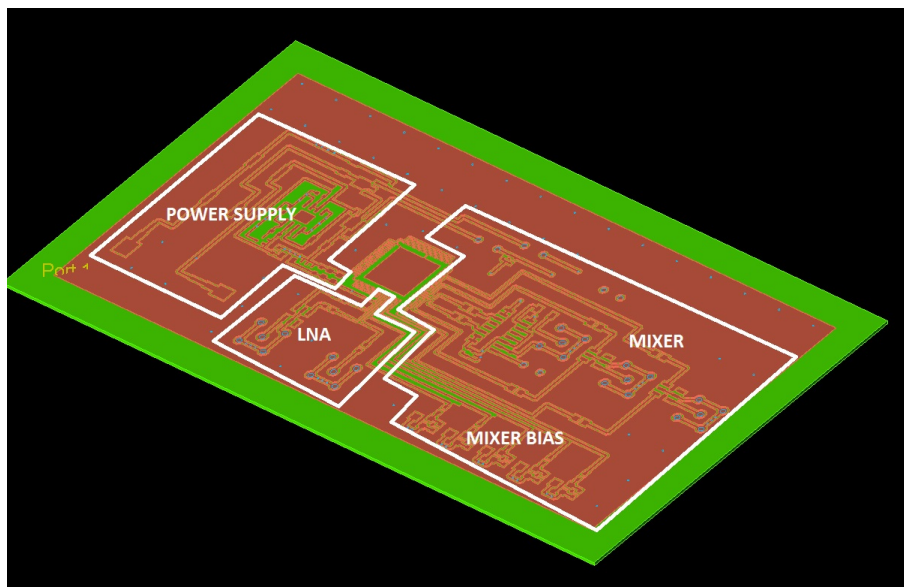
Figure 5.4: 3D view of the designed PCB.

in section 5.1.2), a few traces had to be in the bottom layer. Also, as many vias as possible were inserted spaced 1/20 of wavelength from each other. The traces (LNA input and output) supposed to have 50 Ω, for input impedance matching, were done with increasing width from the chip pins to the Balun, before getting to the SMA connector. The worst-case coupling between traces is located near the LNA pins of our chip, where 5 traces (4 from the LNA and 1 from the Mixer) are next to each other. Electromagnetic (EM) simulation was run and the result is shown in Fig. 5.5.

The worst-case coupling is -16 dB to -18 dB. Up to the moment of the writing of this dissertation, the PCB gerber files were ready to send to fabrication and due to the lack of time, measurements are supposed to be done after this dissertation is defended.

5.2 Test-bench circuits and Measurements

In this section, all test-benches and measurements details are shown. This is the big picture of what is expected by the designer. Small details are not shown here, such as how
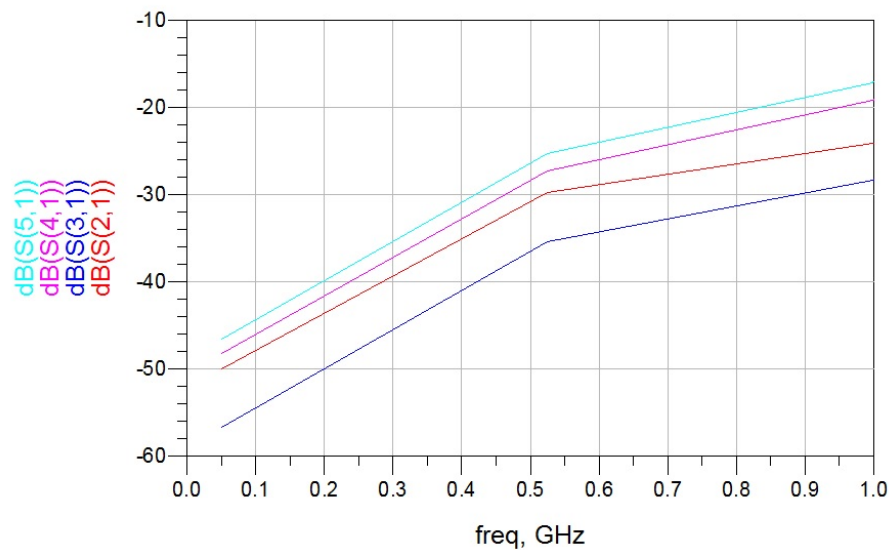
Figure 5.5: S-parameters simulation in ADS to estimate the worst case coupling of 4 traces coming out of the LNA plus 1 of the Mixer.

to configure and setup each equipment and how they are connected (connectors, cable types, PCB material, etc) to the chip.

5.2.1 S-parameters

The S-parameters measurements are made using a network analyzer. It would be better if a vector network analyzer (VNA) is available, since it can measure the phase of each S-parameters. Otherwise, the scalar network analyzer is going to measure only amplitude properties, which does not allow the designer to know imaginary parts of each parameter.

The test-bench for this measurement is shown in Fig. 5.6. The PCB is going to have both input and output connected to the network analyzer and it is going to be fed by an external voltage source. The input and output of the PCB has to be matched to the equipment impedance, which might be 50 or 75 Ω, to give true results of S-parameters. Refer to appendix A to see how S-parameters are defined.

5.2.2 Third order intercept point and P1dB

The test-benches for IIP3 and 1 dB compression point (P1dB) are different, since to measure IIP3 one needs two RF power sources. In this case, it will also be necessary to use an RF power combiner so as to sum both signals into one, which is fed into the PCB connector.

For IIP3 measurement, one has to increase the power of both signal sources equally until the intermodulation products show up within the band at the spectrum analyzer. When the intermodulation products are seen at the spectrum analyzer, the corresponding input power is the IIP3.

A slightly different approach goes for the 1 dB compression point, which has only one RF power source. One should increase the power of the signal source until the output power start dropping. When it achieves a 1 dB drop, seen at the spectrum analyzer, the respective input power is the 1 dB compression point.
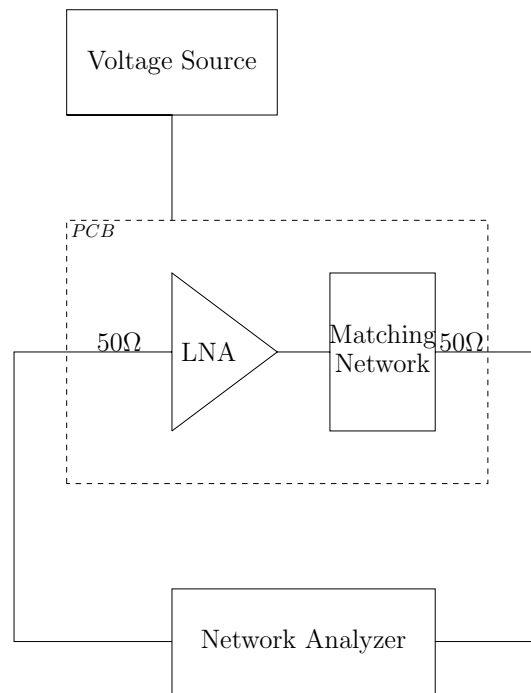
Figure 5.6: S-parameters measurement test-bench.
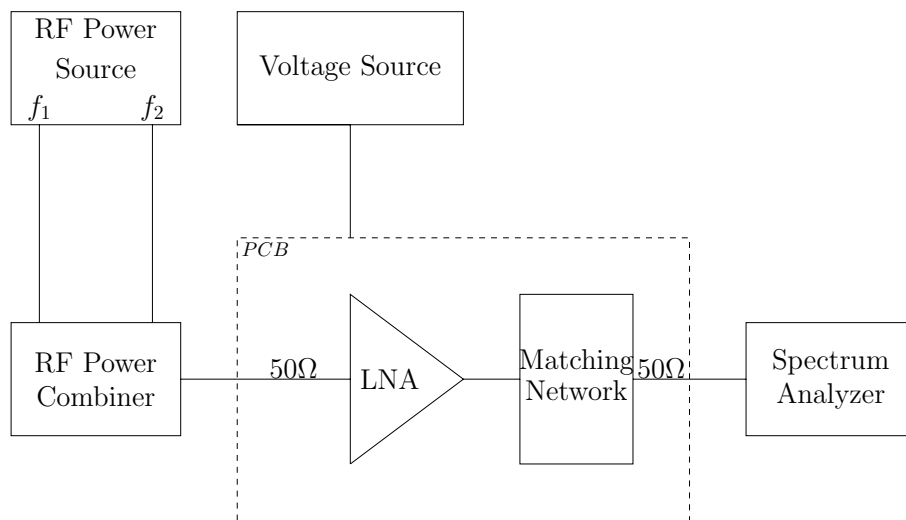
Figure 5.7: IIP3 measurement test-bench.

5.2.3 Noise Figure

There are two common techniques used to measure noise figure. One uses a specific expensive equipment called Noise Figure Analyzer (NFA), and the other uses a method called the Y-factor.

The test-bench for the first technique, using the noise figure analyzer, is shown in Fig. 5.9. This is straight forward to understand. Basically, the NFA uses a known noise source connected to the input of the Device-under-test (DUT) and measures the output power. As it knows all parameters of the noise source, it makes all calculations internally and shows up at the display the NF of the DUT.

The second method implements the internal calculations made inside a NFA. First, one has to know the noise figure theory. Therefore, using a noise source proportional to
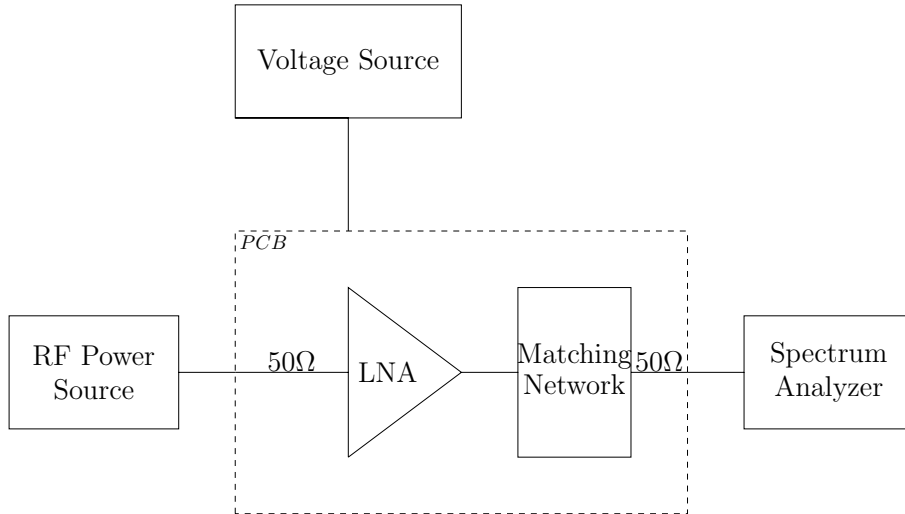
Figure 5.8: 1 dB compression point test-bench.
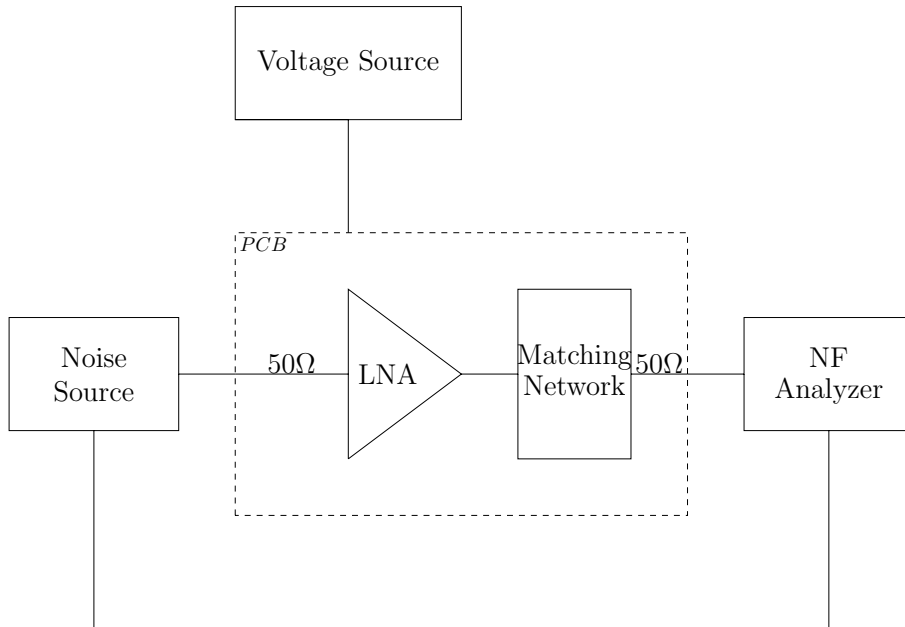
Figure 5.9: NFA method test-bench.

kT (thermal noise) as the input of the amplifier, the following equations for the output of the noise source and the output of the amplifier are valid:

$$
\begin{aligned}
Noise_{OUT_{SOURCE}} &= kT_0BA_V \\
Noise_{OUT_{AMP}} &= kT_0BA_V + noise_{AMP}
\end{aligned}
\tag{5.6}
$$

where A_V is the voltage gain of the amplifier, B bandwidth, T_0 a known temperature and k the boltzmann constant. The noise factor and noise figure are given by:

$$
\begin{aligned}
F &= \frac{kT_0BA_V + noise_{AMP}}{kT_0BA_V} \\
NF &= 10log(F)
\end{aligned}
\tag{5.7}
$$

Thus, if one uses the noise source at two different temperatures, one gets the following noise power at the output of the amplifier:

$$
\begin{aligned}
Noise_1 &= kT_0BA_V + noise_{AMP} \\
Noise_2 &= kT_hBA_V + noise_{AMP}
\end{aligned}
\tag{5.8}
$$

If one equates $noise_{AMP}$ in the above equations, it will yield:

$$
\begin{aligned}
Noise_1 - Noise_2 &= kBA_V(T_0 - T_h) \\
\frac{\Delta Noise}{\Delta T} &= kBA_V
\end{aligned}
\tag{5.9}
$$

Setting up the noise factor equation, then:

$$
\begin{aligned}
F &= \frac{kT_0BA_V + noise_{AMP}}{kT_0BA_V} = 1 + \frac{noise_{AMP}}{kT_0BA_V} \\
F &= 1 + \frac{noise_{AMP}}{T_0\left(\frac{\Delta Noise}{\Delta T}\right)}
\end{aligned}
\tag{5.10}
$$

Thus, one only needs to know the temperature and output noise power differences measured at a spectrum analyzer. For the purpose of the noise source, there are commercial noise sources which comes with given temperature difference (in the form of ENR = excess noise ratio = $\frac{T_{EX}}{T_0}$). They work by switching it on and off (T_h hot and T_c cold) to get both temperatures. The ratio $\frac{Noise_2}{Noise_1}$ is known to be the Y-factor. Thus, one can prove that, based on the above equations, a final expression for the noise factor is:

$$
F = \frac{\frac{T_{EX}}{T_0} - Y\left(\frac{T_c}{T_0} - 1\right)}{Y - 1}
\tag{5.11}
$$

where $T_{EX} = T_h - T_0$, T_c is the cold temperature (noise source off). The ENR $\frac{T_{EX}}{T_0}$ is valid if everything is matched to 50 Ω. Again, only the output powers and ENR are needed to measure the noise figure. Fig. 5.10 shows the test-bench for the Y-factor technique.

Voltage Source

PCB

Noise
Source
ON/OFF

50Ω

LNA

Matching
Network

50Ω

Spectrum
Analyzer
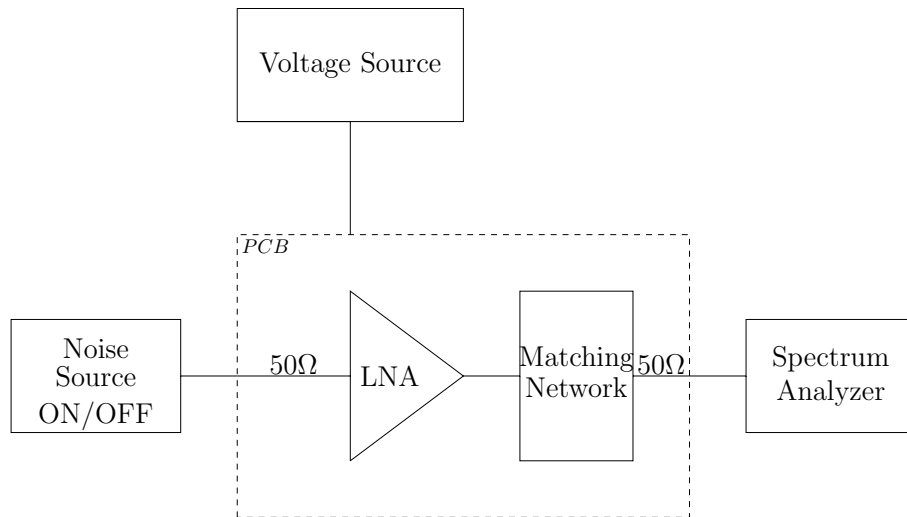
Figure 5.10: Y-factor method test-bench.

6 CONCLUSION

A new Wideband Balun Low Noise Amplifier topology suited for IEEE 802.22 frequency band was introduced. A noise-canceling topology with a new auxiliary amplifier and linearity improving techniques were used. The key idea was a cascaded auxiliary amplifier and the use of diode-connected transistors to completely decouple all design parameters, such as gain, NF, IIP3 and input impedance matching. The concept of using a cascaded auxiliary amplifier was described, demonstrating that the topology gives the designer the flexibility of having a NF < 2 dB (LNA-best-NF) or an IIP3 > 3.3 dBm (LNA-best-IIP3) in the entire band of 50 MHz - 1 GHz, without degrading the other parameters. Which means, the designer can use the same topology for applications with different requirements.

All simulation results were from post-layout, considering bondwire inductances, PAD capacitances and decoupling capacitors on the test-bench. Better values for these (i.e., lower inductance bondwires, lower PAD capacitances, higher decoupling capacitors) can improve overall performance. Temperature sensitivity simulations showed that this topology is suited to operate in the military range from -55°C to 125°C, having IIP3 near or above 0 dBm and all other parameters with very low sensitivity to temperature for the LNA-best-IIP3. The LNA-best-NF achieves a low sensitivity to temperature in the range of 0-80°C.

Fabrication process sensitivity was estimated through Monte Carlo simulation results (1000 samples) which show a very low spread in all parameters of this topology due to the feedback bias used in all transistors V_{GS}. The worst case results of NF and IIP3 for both amplifiers are near the nominal values, which is a good figure of merit for the designer who intends to fabricate this circuit.

Overall, this topology can achieve challenging specifications, i.e., low frequency operation (@50MHz) with extreme low NF and moderate IIP3 or high IIP3 with moderate NF. Besides, curves of parameters such as gain, NF, S_{11} have a great flatness in the entire band, without using any inductor. This makes this circuit very compact, consuming only 17.6 mA at 1.2 V power supply, which compares favorably with similar designs in the same CMOS technology node.

REFERENCES

AKYILDIZ, I. F. et al. NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey. **Computer Networks**, [S.l.], v.50, n.13, p.2127 – 2159, 2006.

ANSARI, A.; YAVARI, M. A very wideband low noise amplifier for cognitive radios. In: ELECTRONICS, CIRCUITS AND SYSTEMS (ICECS), 2011 18TH IEEE INTERNATIONAL CONFERENCE ON, 2011. **Anais...** [S.l.: s.n.], 2011. p.623–626.

BLAAKMEER, S. et al. Wideband Balun-LNA With Simultaneous Output Balancing, Noise-Canceling and Distortion-Canceling. **Solid-State Circuits, IEEE Journal of**, [S.l.], v.43, n.6, p.1341–1350, June 2008.

BRUCCOLERI, F.; KLUMPERINK, E.; NAUTA, B. Wide-band CMOS low-noise amplifier exploiting thermal noise canceling. **Solid-State Circuits, IEEE Journal of**, [S.l.], v.39, n.2, p.275–282, 2004.

CAFARO, G. et al. A 100 MHz 2.5 GHz Direct Conversion CMOS Transceiver for SDR Applications. In: RADIO FREQUENCY INTEGRATED CIRCUITS (RFIC) SYMPOSIUM, 2007 IEEE, 2007. **Anais...** [S.l.: s.n.], 2007. p.189–192.

CHEN, K.-H.; LIU, S.-I. Inductorless Wideband CMOS Low-Noise Amplifiers Using Noise-Canceling Technique. **Circuits and Systems I: Regular Papers, IEEE Transactions on**, [S.l.], v.59, n.2, p.305–314, Feb 2012.

CHENG, W. et al. A wideband IM3 cancellation technique using negative impedance for LNAs with cascode topology. In: RADIO FREQUENCY INTEGRATED CIRCUITS SYMPOSIUM (RFIC), 2012 IEEE, 2012. **Anais...** [S.l.: s.n.], 2012. p.13–16.

CHENG, W.-F. et al. A highly-linear ultra-wideband balun-LNA for cognitive radios. In: EUROCON - INTERNATIONAL CONFERENCE ON COMPUTER AS A TOOL (EUROCON), 2011 IEEE, 2011. **Anais...** [S.l.: s.n.], 2011. p.1–4.

CUSTODIO, J. et al. A small-area self-biased wideband CMOS balun LNA with noise cancelling and gain enhancement. In: NORCHIP, 2010, 2010. **Anais...** [S.l.: s.n.], 2010. p.1–4.

HAO, Z. et al. A 0.1-8.5 GHz wideband CMOS LNA using forward body bias technology for SDR applications. In: MICROWAVE AND MILLIMETER WAVE TECHNOLOGY (ICMMT), 2012 INTERNATIONAL CONFERENCE ON, 2012. **Anais...** [S.l.: s.n.], 2012. v.3, p.1–4.

HAYKIN, S. Cognitive radio: brain-empowered wireless communications. **Selected Areas in Communications, IEEE Journal on**, [S.l.], v.23, n.2, p.201–220, 2005.

IM, D. A +9-dBm Output P1db Active Feedback CMOS Wideband LNA for SAW-Less Receivers. **Circuits and Systems II: Express Briefs, IEEE Transactions on**, [S.l.], v.60, n.7, p.377–381, July 2013.

IM, D.; NAM, I.; LEE, K. A Low Power Broadband Differential Low Noise Amplifier Employing Noise and IM3 Distortion Cancellation for Mobile Broadcast Receivers. **Microwave and Wireless Components Letters, IEEE**, [S.l.], v.20, n.10, p.566–568, Oct 2010.

KIHARA, T.; MATSUOKA, T.; TANIGUCHI, K. A 1.0 V, 2.5 mW, transformer noise-canceling UWB CMOS LNA. In: RADIO FREQUENCY INTEGRATED CIRCUITS SYMPOSIUM, 2008. RFIC 2008. IEEE, 2008. **Anais...** [S.l.: s.n.], 2008. p.493–496.

KIM, J. et al. A 54 -862 MHz CMOS direct conversion transceiver for IEEE 802.22 cognitive radio applications. In: CUSTOM INTEGRATED CIRCUITS CONFERENCE, 2009. CICC '09. IEEE, 2009. **Anais...** [S.l.: s.n.], 2009. p.255–258.

KLIMACH, H. et al. Resistorless switched-capacitor bandgap voltage reference with low sensitivity to process variations. **Electronics Letters**, [S.l.], v.49, n.23, p.1448–1449, Nov 2013.

KLIMACH, H. et al. A resistorless switched bandgap reference topology. In: CIRCUITS AND SYSTEMS (LASCAS), 2013 IEEE FOURTH LATIN AMERICAN SYMPOSIUM ON, 2013. **Anais...** [S.l.: s.n.], 2013. p.1–4.

KLIMACH, H. et al. A resistorless switched bandgap voltage reference with offset cancellation. In: INTEGRATED CIRCUITS AND SYSTEMS DESIGN (SBCCI), 2013 26TH SYMPOSIUM ON, 2013. **Anais...** [S.l.: s.n.], 2013. p.1–5.

KUO, M.-C.; KUO, C.-N.; CHUEH, T.-C. Wideband LNA Compatible for Differential and Single-Ended Inputs. **Microwave and Wireless Components Letters, IEEE**, [S.l.], v.19, n.7, p.482–484, July 2009.

LI, Q.; ZHANG, Y. P. A 1.5-V 2 -;9.6-GHz Inductorless Low-Noise Amplifier in 0.13-nm CMOS. **Microwave Theory and Techniques, IEEE Transactions on**, [S.l.], v.55, n.10, p.2015–2023, Oct 2007.

LIAO, Y.; TANG, Z.; MIN, H. A CMOS wide-band low-noise amplifier with balun-based noise-canceling technique. In: SOLID-STATE CIRCUITS CONFERENCE, 2007. ASSCC '07. IEEE ASIAN, 2007. **Anais...** [S.l.: s.n.], 2007. p.91–94.

LIN, Y.-S. et al. High-Performance Wideband Low-Noise Amplifier Using Enhanced π -Match Input Network. **Microwave and Wireless Components Letters, IEEE**, [S.l.], v.24, n.3, p.200–202, March 2014.

MITOLA, J.; MAGUIRE G.Q., J. Cognitive radio: making software radios more personal. **Personal Communications, IEEE**, [S.l.], v.6, n.4, p.13–18, 1999.

MOEZZI, M.; BAKHTIAR, M. Wideband LNA Using Active Inductor With Multiple Feed-Forward Noise Reduction Paths. **Microwave Theory and Techniques, IEEE Transactions on**, [S.l.], v.60, n.4, p.1069–1078, April 2012.

NGUYEN, V. T.; VILLAIN, F. Cognitive Radio RF: overview and challenges. **VLSI Design**, [S.l.], v.2012, 2012.

PARK, J. et al. A Fully Integrated UHF-Band CMOS Receiver With Multi-Resolution Spectrum Sensing (MRSS) Functionality for IEEE 802.22 Cognitive Radio Applications. **Solid-State Circuits, IEEE Journal of**, [S.l.], v.44, n.1, p.258–268, Jan 2009.

PIMENTEL, H.; BAMPI, S. A 50MHz x2013;lGHz wideband low noise amplifier in 130nm CMOS technology. In: INTEGRATED CIRCUITS AND SYSTEMS DESIGN (SBCCI), 2012 25TH SYMPOSIUM ON, 2012. **Anais...** [S.l.: s.n.], 2012. p.1–6.

RAZAVI, B. **Design of Analog CMOS Integrated Circuits**. International.ed. [S.l.]: McGRAW-HILLl, 2001.

RAZAVI, B. Challenges in the design of cognitive radios1. In: CUSTOM INTEGRATED CIRCUITS CONFERENCE, 2009. CICC '09. IEEE, 2009. **Anais...** [S.l.: s.n.], 2009. p.391–398.

RAZAVI, B. Cognitive Radio Design Challenges and Techniques. **Solid-State Circuits, IEEE Journal of**, [S.l.], v.45, n.8, p.1542–1553, 2010.

RAPPAPORT, T. S. (Ed.). **RF Microeletronics**. 2.ed. Upper Sadle River, NJ-US: Prentice Hall, 2011.

TSIVIDIS, Y. **Operation and Modeling of The MOS Transistor**. 2.ed. [S.l.]: Oxford University Pressl, 1987.

WANG, X. et al. 0.6 x2013;3-GHz Wideband Receiver RF Front-End With a Feedforward Noise and Distortion Cancellation Resistive-Feedback LNA. **Microwave Theory and Techniques, IEEE Transactions on**, [S.l.], v.60, n.2, p.387–392, 2012.

XIMENES, A.; SWART, J. A wideband noise canceling low-noise amplifier for 50MHz x2013;5GHz wireless receivers in CMOS technology. In: NEW CIRCUITS AND SYSTEMS CONFERENCE (NEWCAS), 2011 IEEE 9TH INTERNATIONAL, 2011. **Anais...** [S.l.: s.n.], 2011. p.197–200.

YOUSSEF, A.; ISMAIL, A.; HASLETT, J. A sub - 2 dB noise figure wideband LNA in 65 nm CMOS for mobile TV applications. In: SILICON MONOLITHIC INTEGRATED CIRCUITS IN RF SYSTEMS (SIRF), 2010 TOPICAL MEETING ON, 2010. **Anais...** [S.l.: s.n.], 2010. p.164–167.

YU, Y.-H.; YANG, Y.-S.; CHEN, Y.-J. A Compact Wideband CMOS Low Noise Amplifier With Gain Flatness Enhancement. **Solid-State Circuits, IEEE Journal of**, [S.l.], v.45, n.3, p.502–509, March 2010.

ZHANG, H.; FAN, X.; SINENCIO, E. A Low-Power, Linearized, Ultra-Wideband LNA Design Technique. **Solid-State Circuits, IEEE Journal of**, [S.l.], v.44, n.2, p.320–330, 2009.

LIST OF PUBLICATIONS

Prior to defining the LNA theme for this Master Thesis, the author also worked on a new DC voltage reference design, namely a with switched-capacitor bandgap reference. The topology developed was new and generated 3 publications (LASCAS2013, SBCCI2013 and Electronic Letters 2013). The main feature was the use of a single bipolar transistor and no resistors in this BGR design. Instead, capacitors were used. Results showed a reduced spread in Monte Carlo simulations for Temperature Coefficient (TC) and output voltage, when compared with a traditional topology. The publications are listed in the references KLIMACH et al. (2013), KLIMACH et al. (2013) and KLIMACH et al. (2013) and repeated below for convenience.

The publications regarding the LNA design appeared at the SIM2013 Seminar and another submitted to IEEE transactions on circuits and systems part 2 - express briefs (TCASII) in May 2014 (waiting for acceptance notification) also listed below.

1. Costa, A. L. T., Klimach. H. Bampi, S., "A Survey of Wideband Low Noise Amplifiers Design Techniques for Cognitive Radios", Simposio Sul de Microeletronica (SIM), 2013;

2. Klimach, H., Monteiro, M. F. C., Costa, A. L. T., Bampi, S., "A Resistorless Switched Bandgap Reference Topology", IEEE Fourth Latin American Symposium on Circuits and Systems (LASCAS), 2013;

3. Klimach, H., Costa, A. L. T., Monteiro, M. F. C., Bampi, S., "A Resistorless Switched Bandgap Voltage Reference with Offset Cancellation", 26th Symposium on Integrated Circuits and Systems Design (SBCCI), 2013;

4. Klimach, H., Costa, A. L. T., Monteiro, M. F. C., Bampi, S., "Resistorless Switched-Capacitor Bandgap Voltage Reference with Low Sensitivity to Process Variations", Electronics Letters, November 25, 2013;

5. Costa, A. L. T., Klimach. H. Bampi, S., "High Linearity 24 dB Gain Wideband Inductorless Balun Low-Noise Amplifier for IEEE 802.22 Band", IEEE Transactions on Circuits and Systems part II - Express Briefs (submitted in May, 2014).

APPENDIX A   SUMMARY IN PORTUGUESE

UM AMPLIFICADOR DE BAIXO RUÍDO BANDA LARGA, SEM INDUTOR, COM ALTA LINEARIDADE E 24 dB DE GANHO PARA A BANDA DO PADRÃO IEEE 802.22

A.1 INTRODUÇÃO

À medida que mais e mais usuários acessam redes sem fio, o congestionamento do espectro de frequências continua a aumentar. Rádios Cognitivos (RC) oferecem uma maneira de aliviar esse congestionamento: eles monitoram continuamente o espectro e detectam e utilizam canais não ocupados MITOLA; MAGUIRE G.Q. (1999), HAYKIN (2005). Esse conceito de RC é ilustrado na Fig. 1.1. Os esforços atuais no projeto de RC tem focado nas bandas de TV abaixo de 1 GHz PARK et al. (2009), KIM et al. (2009), devido ao lançamento do padrão IEEE 802.22 Wireless Regional Area Network (WRAN). Esse padrão foca na implementação de um RC na banda de 54 MHz até 862 MHz.

Como os RCs não focam nenhum padrão e suas bandas estreitas alocadas, ele deve operar em qualquer frequência na banda larga inteira. Além disso, RCs não são projetados com informações, à priori, das bandas dos interferentes, tendo que suportar interferentes em qualquer frequência em toda a banda larga de operação. (denotado aqui por BW_{CR}. Consequentemente, os espúrios de mixagem e parâmetros de performance como os pontos de intersecção de terceira e segunda ordem (IP2 e IP3, respectivamente), que caracterizam a linearidade do circuito devido à efeitos de intermodulação, devem satisfazer requisitos mais restritos. Uma descrição detalhada de IP3 está no apêndice B. A Fig. 1.2 ilustra o problema do IP2 em receptores de banda larga.

Para entender o problema da linearidade, considere o efeito de distorção de segunda ordem no caminho do sinal em uma banda estreita para conversão direta. Como ilustrado na Fig. 1.2(a), dois interferentes em f_1 e f_2 geram um batimento em $f_2 - f_1$, à medida que eles sofrem distorção de segunda ordem no amplificador de baixo ruído (LNA em inglês) e no estágio de entrada do mixer. Devido à baixa isolação, uma fração desse batimento vaza para a banda base sem translação de frequência, corrompendo o sinal convertido para baixo. Neste cenário, somente o mixer limita a performance, porque a característica passa-altas da saída do LNA pode remover os batimentos de baixa frequência. De fato, o IP2 de muitos receptores é medido de acordo com esse cenário.

O problema de não-linearidade de segunda ordem assume novas dimensões em rádios cognitivos. Como mostrado na Fig. 1.2(b), o LNA produz componentes em $f_2 + f_1$ e $f_2 - f_1$, onde ambos podem estar dentro da banda larga do rádio cognitivo (BW_{CR}). Isto é, o LNA se torna o elemento crítico. Topologias diferenciais aliviam esta questão consideravelmente, porém é extremamente difícil projetar balun passa-baixas com banda larga.

Um front-end do receptor do RC deve prover um ganho relativamente plano e um casamento de impedância razoável em toda a BW_{CR}, colocando demandas desafiadoras no radio e no seu LNA. O apêndice C contém explicação detalhada sobre o coeficiente de reflexão S_{11}. A faixa de frequência desejada pode ser reduzida para menores bandas, onde assim podem ser processadas por vários circuitos de LNA sintonizados dedicados. O outro extremo é um único LNA, onde obviamente precisa ter uma banda larga. Em contraste com a solução multi-LNA, o LNA banda larga único é flexível e eficiente em termos de área, energia e custos (Fig. 1.3 e 1.4 ilustram transceptores seguindo cada uma dessas ideias).

A figura de ruído (FR) de um bloco de RF é a diferença em dB da relação sinal ruído (SNR em inglês) na entrada e o SNR na saída, o que significa que é uma figura de degradação de SNR da entrada para a saída. Quando na cadeia de blocos do receptor, o FR total é dado pela equação de Friis na equação 1.1.

Onde F_{total} é chamado de fator de ruído, que é simplesmente a FR medida em mag-

nitude, ao invés de dB, e G são os ganhos de cada bloco. De acordo com a equação de Friis, a FR do primeiro bloco da cascata, i.e., o LNA, tem o maior impacto na FR total do receptor.

Em RC, a FR do receptor deve ser tão baixa quanto 4-6 dB, o que é difícil atingir em uma banda larga. Porém, baixa FR em banda larga é atingível para altas frequências acima de 1 GHz, devido à predominância do ruído térmico (o apêndice C tem uma descrição detalhada sobre as fontes de ruído consideradas neste trabalho). Para ter uma baixa FR em frequências mais baixas, como 50 MHz, o ganho total tem que ser melhorado, como também deve haver uma otimização dos ruídos Flicker e térmico.

Em 2004, BRUCCOLERI; KLUMPERINK; NAUTA (2004) propuseram uma topologia de LNA com cancelamento de ruído para solucionar o problema de FR em banda larga para receptores de RCs. Esta topologia atingiu FR < 3 dB e IIP3 perto de 0 dBm. A ideia básica é a inclusão de um amplificador auxiliar em paralelo com o amplificador principal para levar o ruído de entrada até a saída para cancelamento de ruído. Desde então, muitos trabalhos tentaram lidar ou com as limitações de linearidade como em ZHANG; FAN; SINENCIO (2009), IM; NAM; LEE (2010), CHENG et al. (2012) ou com as de FR como em YU; YANG; CHEN (2010), YOUSSEF; ISMAIL; HASLETT (2010), XIMENES; SWART (2011), com alguns focando na otimização da performance geral dos parâmetros como em RAZAVI (2010), BLAAKMEER et al. (2008), MOEZZI; BAKHTIAR (2012), KUO; KUO; CHUEH (2009), IM (2013), LIN et al. (2014).

Como as topologias tradicionais de cancelamento de ruído (ou fonte-commum (FC) + porta-commum (PC) ou realimentação resistiva + FC) tem um amplificador auxiliar único, seu ganho é completamente atrelado ao ganho de ruído do amplificador principal, o que dificulta o desacoplamento entre os vários parâmetros de projeto como FR, casamento de impedância, banda, linearidade, area e energia. Isso sugere que mais graus de liberdade são necessários no amplificador auxiliar de forma a melhorar a performance total do LNA.

LNAs com entradas simples são preferidos para economizar pinos de entrada e saída e porque antenas e filtros de RF normalmente produzem sinais de conexão simples. Por outro lado, sinais diferenciais na cadeia do receptor são preferidos para reduzir a distorção de segunda ordem e para rejeitar ruído da fonte de alimentação e do substrato. Então, em algum ponto na cadeia do receptor um balun é necessário para converter o sinal simples para diferencial. Baluns fora do chip com baixas perdas são tipicamente banda estreita, de modo que alguns baluns seríam necessários para implementar operação em banda larga. Por outro lado, um balun passivo banda larga tipicamente tem altas perdas, degradando a FR total do receptor significativamente.

Combinar as funciondalidades do balun e do LNA em um único circuito integrado parece ser uma opção atrativa para realizar um LNA banda larga para um receptor. As especificações para o LNA deste trabalho estão mostradas no final do capítulo de estado da arte.

Esta dissertação começa com a revisão do estado da arte, apresentando LNAs recentes desenvolvidos na literatura. Isso é feito no capítulo 2. No capítulo 3, a topologia, projeto do circuito e considerações de projeto em nível elétrico é feita. Neste capítulo, é dada mais ênfase aos problemas de layout do projeto, pads e proteções ESD e detalhes do PDK. Os resultados de simulações pós-layout são apresentados e discutidos no capítulo 4. No capítulo 5, o plano de testes e as medidas após fabricação esperadas são discutidas, e finalmente as conclusões no capítulo 6, seguida das referências e apêndices. A teoria complementar de RF como distorção harmônica, intermodulação, linearidade, ponto compressão de 1ª ordem, ruído, parâmetros de espalhamento, estabilidade, necessárias

para ajudar à entender essa dissertação é incluída nos apêndices.

A.2 PROJETO DO LNA

O projeto do LNA deste trabalho começou pela criação da topologia. Baseado no que já existia, surgiu a ideia de montar um balun utilizando o núcleo da topologia do Bruccoleri de 2004. Dessa forma, a intenção foi criar um novo amplificador auxiliar para melhorar os compromissos dos parâmetros de projeto (ganho, FR, IIP3, casamento de impedância, etc).

Na primeira seção do capítulo 3, há uma descrição de como a topologia foi montada. Inicialmente se pensou em um amplificador principal com um inversor CMOS realimentado resistivamente e um amplificador auxiliar com dois inversores em série para montar o balun (entrada simples e saída diferencial) e em seguida diodos foram conectados em ambos os ramos pra melhorar a linearidade e desacoplar os parâmetros de projeto entre si.

Dada a topologia de forma qualitativa, as equações e descrições quantitativas estão na seção 2 do capítulo 3. Todas as equações para definir ganho, casamento de impedância e figura de ruído foram apresentadas para cada bloco do circuito de forma simplificada para baixas frequências. O projeto de altas frequências teve como base o de baixas frequências, tentando melhorar o projeto ao lidar com os parasitas por simulação. Nesta seção também é mostrado o dimensionamento dos transistores. para cada LNA projetado (LNA-best-NF e LNA-best-IIP3). Após o primeiro projeto, uma etapa de otimização foi feita. A maneira como foi otimizado o projeto é descrita abaixo de acordo com a influência que cada bloco tem no circuito:

- Amplificador principal: polarização, casamento de impedância de entrada, FR, Ganho;
- Amplificador auxiliar: FR, IIP3 e Ganho;
- Diodo de linearização M_3: IIP3 e FR;

Neste trabalho, foram projetados dois LNAs, um focando na menor FR possível e o outro no maior IIP3 possível. Abaixo segue a descrição de como o circuito foi otimizado para cada projeto:

1. Melhor FR: o valor de R_F deve ser incrementado, aumentando o ganho; aumentar $g_{m1,2}$ para compensar a mudança no S_{11}; aumentar o W de M_3 para compensar a perda de IIP3 devido ao aumento de ganho e aumentar R_{OUT} para balancear a saída.
2. Melhor IIP3: Aumentar W de M_3 até que o IIP3 máximo seja atingido; diminuir o valor de R_F até que a FR atinja uma limite dado; diminuir R_{OUT} para balancear a saída.

O projetista deve observar que R_F não pode ser muito grande, visto que ele deve ser contrabalanceado por $g_{m1.2}$ para manter o casamento de impedância de entrada, o que se reflete em um maior consumo de energia. Além disso, R_{OUT} deve ser baixo para um baixo ruído térmico na saída e para permitir ganho de potência (caso desejado impedância de saída de 50 Ω).

Na seção 3 do capítulo 3, está descrito como foi feito o projeto do layout, incluindo as especificações do PDK utilizado. Foram utilizados resistores de polisilício, que possuem maior resistência por unidade de área, quando comparado com outros resistores do PDK. Os capacitores utilizados foram do tipo MIM (Metal-Isolante-Metal) com plano de terra

do tipo SUB. Os transtistores utilizados foram os FET regulares de 1.2 V, com V_{TH} típico de 355 mV.

Casamento entre dispositivos não foi um problema neste circuito, então, somente transistores dummy foram utilizados para manter a borda de todos os dispositivos ok. Para uma melhor isolação de RF, anéis de guarda foram utilizados em todos os transistores desse projeto.

Os PADs utilizados foram os fornecidos pelo PDK com tamanho próximo ao mínimo e proteções ESD foram inseridas no projeto do LNA que foi fabricado.

O capítulo 4 contém os resultados de simulação pós-layout e a descrição do test-bench utilizado. Neste test-bench, foram consideradas capacitâncias de PAD, capacitâncias de desacoplamento, indutâncias de bondwires e a carga de um Mixer que está sendo desenvolvido no mesmo grupo de microeletrônica analógica da UFRGS.

Os resultados estão mostrados nas figuras do capítulo 4, com destaque para a FR do LNA-best-NF ser abaixo de 2 dB em toda a banda de 50 MHz - 1 GHz. Outro destaque é o IIP3 do LNA-best-IIP3 ser maior do que 3.3 dBm em toda a banda de 50 MHz - 1 GHz. Em ambos os circuitos, houve um pequeno desbalanceamento da tensão de saída devido às imperfeições no layout. Ambos os circuitos mostraram ótima estabilidade com a temperatura na faixa de -55 até 125°C e baixa variabilidade com processo e mistmach.

A comparação com outros projetos de LNAs banda larga da literatura foi feita em duas tabelas. Uma contendo todos os parâmetros de projeto e outra utilizando a Figura de Mérito (FoM em inglês) descrita a seguir (baseada na Figura de Mérito da ITRS 2011):

$$FoM = \frac{G * IIP3 * BW}{(F - 1) * P * AREA} \tag{A.1}$$

Onde G é o ganho, BW é a banda do circuito, F é o fator de ruído, P é a potência consumida. Todos os valores em magnitude. O circuito LNA-best-IIP3 projetado neste trabalho mostrou ser melhor do que todos os comparados.

O capítulo 5 descreve como foi projetada uma Placa de Circuito Impresso (PCI) para medir o chip contendo o LNA que foi submetido à fabricação (prova de conceito somente, pois não é nem o LNA-best-NF nem o LNA-best-IIP3). A ideia de fazer dois LNAs surgiu após ter submetido um para fabricação, bem como a ideia de desacoplar o primeiro estágio do amplificador auxiliar do segundo.

Esse capítulo descreve principalmente técnicas de RF para obter o melhor projeto de PCI possível, tentando utilizar trilhas o mais afastadas entre si possível, o conjunto de capacitores da linha de alimentação próximo ao chip, cuidados com a largura das trilhas que contém sinais de RF, plano de terra na camada inferior e superior da placa. Além disso, VIAs foram colocadas espaçadas 1/20th do comprimento de onda do pior caso de sinal de RF para reduzir acoplamento entre os sinais.

Ao final do capítulo 5, são discutidos os test-benches de medição. Basicamente, como a PCI será conectada aos equipamentos de medição: analisador de espectro, analisador de rede, analisador de figura de ruído.

A.3   CONCLUSÃO

Uma nova topologia de Amplificador de Baixo Ruído balun e banda larga adequada para a banda de operação do padrão IEEE 802.22 foi introduzida. Uma topologia de cancelamento de ruído com um novo amplificador auxiliar e melhorias de linearidade foram

usadas. A ideia principal foi a cascata de amplificadores auxiliares e o uso de transistores conectados como diodos para desacoplar completamente os parâmetros de projeto, como ganho, FR, IIP3, e casamento de impedância. O conceito da utilização do amplificador auxiliar em cascata foi descrito, demonstrando que a topologia permite ao projetista ter a flexibilidade de ter FR < 2 dB (LNA-best-NF) ou um IIP3 > 3.3 dBm (LNA-best-IIP3) em toda a banda de 50 MHz - 1 GHz, sem degradar os outros parâmetros. O que significa que o projetista pode usar a mesma topologia para aplicações com diferentes requisitos de projeto.

Todos os resultados de simulação foram feitas após extração de parasitas do layout, considerando indutâncias de bondwires, capacitâncias de PAD e de desacoplamento no circuito de teste. Melhores valores desses parâmetros pode melhorar a performance do circuito. As simulações de sensibilidade à temperatura mostraram que essa topologia é adequada para operar na faixa militar de -55°C à 125°C, com IIP3 próximo ou acima de 0 dBm e todos os outros parâmetros com baixa sensibilidade com a temperatura para o LNA-best-IIP3. O LNA-best-NF atingiu uma baixa sensibilidade com a temperatura na faixa de 0-80°C.

A sensibilidade ao processo de fabricação foi estimada de acordo com resultados de simulações Monte Carlo (1000 amostras), que mostrou um baixo espalhamento em todos os parâmetros desta topologia devido à polarização do circuito ter uma realimentação segurando o ponto de operação, com todos os transistores com o mesmo V_{GS}. Os resultados de pior caso para FR e IIP3 para ambos os amplificadores estão próximos aos valores nominais, o que é uma boa figura de mérito para o projetista que pretende fabricar este circuito.

No geral, esta topologia pode atingir especificações desafiadoras, i.e., operar em baixa frequência (@ 50 MHz) com FR muito baixas e IIP3 moderado ou alto IIP3 e FR moderada. Além disso, curvas dos parâmetros como ganho, FR, S_{11} tem uma ótima planaridade na banda inteira, sem usar nenhum indutor. Isso faz este circuito muito compacto, consumindo somente 17.6 mA com fonte de alimentação de 1.2 V, o que se compara favoravelmente com projetos similares do mesmo nó tecnológico CMOS.

APPENDIX B   RF BASIC CONCEPTS

Most of the basic concepts described here, including pictures, are taken from the book RF Microelectronics 2nd Edition of Behzad Razavi.

B.1   Effects of Nonlinearity

In this section, phenomena that are not predicted by small signal models are described. For a memoryless system whose input/output characteristic can be approximated by

$$y(t) \approx \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t) \tag{B.1}$$

This should not be considered as Taylor's series expansion, but rather a fit across the signal swings of interest. The nonlinearity effects described here are mainly due to the third order (odd order). The second order (even order) also manifests itself, however they are generally negligible, being taken into account only in certain types of receiver architectures such as direct conversion.

B.1.1   1 dB Compression Point

Ifa sinusoid is applied to a nonlinear system, the output generally exhibits frequency components that are integer multiples of the input frequency. In B.1, if x(t) = Acosωt, then

$$
\begin{aligned}
y(t) &= \alpha_1 A cos\omega t + \alpha_2 A^2 cos^2\omega t + \alpha_3 A^3 cos^3\omega t \\
&= \alpha_1 A cos\omega t + \frac{\alpha_2 A^2}{2}(1 + cos2\omega t) + \frac{\alpha_3 A^3}{4}(3cos\omega t + cos3\omega t) \\
&= \frac{\alpha_2 A^2}{2} + \left(\alpha_1 A + \frac{3\alpha_3 A^3}{4}\right)cos\omega t + \frac{\alpha_2 A^2}{2}cos2\omega t + \frac{\alpha_3 A^3}{4}cos3\omega t \quad \text{(B.2)}
\end{aligned}
$$

The small-signal gain of circuits is usually obtained with the assumption that harmonics are negligible. However, a formulation of harmonics, as expressed by B.2, indicates that the gain experienced by Acosωt is equal to $a_1 + 3a_3 A^2/4$ and hence varies appreciably as A becomes larger.

Returning to the third-order polynomial in B.1, one notes that if $a_1 a_3 > 0$, then $a_1 x + a_3 x^3$ overwhelms $a_2 x^2$ for large x regardless of the sign of a_2, yielding an "expansive" characteristic (Fig. B.1). For example, an ideal bipolar transistor operating in the forward active region produces a collector current in proportion to $\exp(V_{BE}/V_T)$, exhibiting expansive behavior. On the other hand, if $a_1 a_3 < 0$, the term $a_3 x^3$ "bends" the characteristic

for sufficiently large x (Fig. B.1), leading to "compressive" behavior, i.e., a decreasing gain as the input amplitude increases. For example, the differential pair suffers from compression. Since most RF circuits of interest are compressive, this text focus on this type.
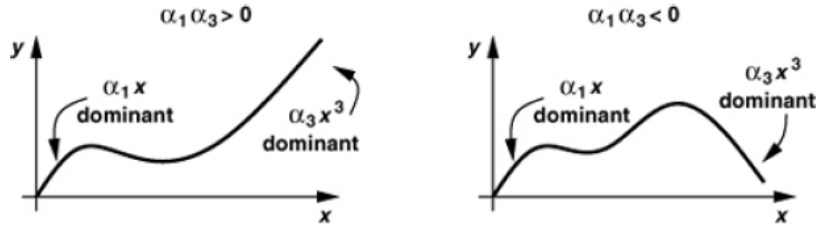

Figure B.1: Expansive and compressive characteristics.

With $a_1 a_3 < 0$, the gain experienced by $A cos \omega t$ in B.2 falls as A rises. This effect is quantified by the "1-dB compression point", defined as the input signal level that causes the gain to drop by 1 dB. If plotted on a log-log scale as a function of the input level, the output level, A_{out}, falls below its ideal value by 1 dB at the 1-dB compression point, $A_{in,1dB}$ (Fig. B.2). Note that A_{in} and $A_{out,1dB}$ are voltage quantities here, but compression can also be expressed in terms of power quantities. The 1-dB compression may also be specified in terms of the output level at which it occurs, $A_{out,1dB}$. The input and output compression points typically prove relevant in the receive path and the transmit path, respectively.

To calculate the input 1-dB compression point, one can equate the compressed gain, $\alpha_1 + (3\alpha_3/4)A_{in,1dB}^2$, to 1 dB less than the ideal gain, a_1:

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

It follows that:

$$A_{in,1dB} = \sqrt{0.145 \left| \frac{\alpha_1}{\alpha_3} \right|} \tag{B.4}$$

Note that B.4 gives the peak value (rather than the peak-to-peak value) of the input. Also denoted by P_{1dB}, the 1-dB compression point is typically in the range of -20 to -25 dBm (63.2 to 35.6 mV_{pp} in 50 Ω system)at the input of RF receivers.
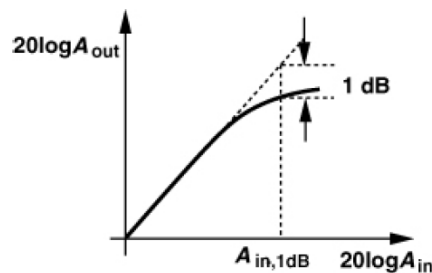

Figure B.2: Definition of 1-dB compression point.

B.1.2 Harmonic Distortion

From Fourier series theory, a periodic signal can be expanded as a sum of sines and cosines:

$$f(t) = b_0 + \sum_{n=1}^{\infty} b_n cos(n\omega t) + c_n sin(n\omega t) \tag{B.5}$$

If the function f(t) is even (f(t) = f(-t)), then, $c_n = 0$ for all n. Considering an ordinary transfer function V_{sa} f(v_{in}), it can be turned into an even function if $v_{in} = V_{in,A}cos\omega t$. This leads to a simplification in the Fourier series, and the output is given by:

$$V_{sa} = b_0 + b_1 cos\omega t + b_2 cos(2\omega t) + ... \tag{B.6}$$

Based on this, harmonic distortion can be defined. The harmonic distortion of nth order is the ratio of the nth order harmonic amplitude and the fundamental amplitude:

$$HD_n = \frac{|b_n|}{|b_1|}, n \geqslant 2 \tag{B.7}$$

Also, total harmonic distortion, THD, is defined as:

$$THD = \frac{\sqrt{\sum_{n=2}^{\infty} b_n^2}}{|b_1|} = \sqrt{\sum_{n=2}^{\infty} HD_n^2} \tag{B.8}$$

B.1.3 Intermodulation

If two interferers at ω_1 and ω_2 are applied to a nonlinear system, the output generally exhibits components that are not harmonics of these frequencies. Called "intermodulation" (IM), this phenomenon arises from "mixing" (multiplication) of the two components as their sum is raised to a power greater than unity. To understand how Eq. B.1 leads to intermodulation, assume x(t) = A_1cos ω_1t. Thus,

$$\begin{aligned} y(t) &= \alpha_1(A_1 cos\omega_1 t + A_2 cos\omega_2 t) + \alpha_2(A_1 cos\omega_1 t + A_2 cos\omega_2 t)^2 \tag{B.9} \\ &+ \alpha_3(A_1 cos\omega_1 t + A_2 cos\omega_2 t)^3 \tag{B.10} \end{aligned}$$

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

$$\begin{aligned} \omega &= 2\omega_1 \pm \omega_2 : \frac{3\alpha_3 A_1^2 A_2}{4}cos(2\omega_1 + \omega_2)t + \frac{3\alpha_3 A_1^2 A_2}{4}cos(2\omega_1 - \omega_2)t \tag{B.11} \\ \omega &= 2\omega_2 \pm \omega_1 : \frac{3\alpha_3 A_1 A_2^2}{4}cos(2\omega_2 + \omega_1)t + \frac{3\alpha_3 A_1 A_2^2}{4}cos(2\omega_2 - \omega_1)t \tag{B.12} \end{aligned}$$

$$\tag{B.13}$$

and these fundamental components:

$$\omega = \omega_1, \omega_2 : (\alpha_1 A_1 + \frac{3}{4}\alpha_3 A_1^3 + \frac{3}{2}\alpha_3 A_1 A_2^2)cos\omega_1 t \tag{B.14}$$

$$+ \ (\alpha_1 A_2 + \frac{3}{4}\alpha_3 A_2^3 + \frac{3}{2}\alpha_3 A_2 A_1^2)cos\omega_2 t \tag{B.15}$$

Fig. B.3 illustrates the results. Among these, the third-order IM products at $2\omega_1$-ω_2 and $2\omega_2$-ω_1 are of particular interest. This is because, if ω_1 and ω_2 are close to each other, then $2\omega_1$-ω_2 and $2\omega_2$-ω_1 appear in the vicinity of ω_1 and ω_2.
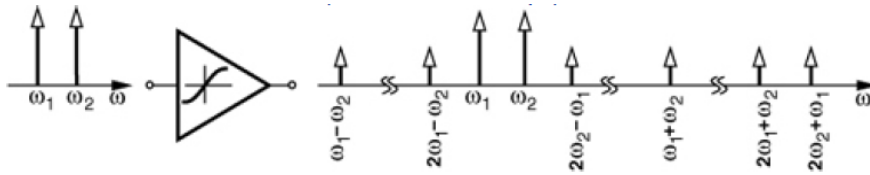

Figure B.3: Generation of various intermodulation components in a two-tone test.

B.1.4   Third order intercept point - IIP3

A common method of IM characterization is the "two tone" test, whereby two pure sinusoids of equal amplitudes are applied to the input. The amplitude of the output IM products is then normalized to that of the fundamentals at the output. Denoting the peak amplitude of each tone by A, we can write the result as:

$$RelativeIM = 20log\left(\frac{3}{4}\frac{\alpha_3}{\alpha_1}A^2\right) dBc \tag{B.16}$$

where the unit dBc denotes decibels with respect to the "carrier" to emphasize the normalization. Note that, if the amplitude of each input tone increases by 6 dB (a factor of two), the amplitude of the IM products ($\propto A^3$) rises by 18 dB and hence the relative IM by 12 dB (it is assumed that no compression occurs so that the output fundamental tones also rise by 6 dB).

The principal difficulty in specifying the relative IM for a circuit is that it is meaningful only if the value of A is given. From a practical point of view, it is preferred a single measure that captures the intermodulation behavior of the circuit with no need to know the input level at which the two-tone test is carried out. Fortunately, such a measure exists and is called the "third-order intercept point" (IP3).

The concept of IP3 originates from the earlier observation that, if the amplitude of each tone rises, that of the output IM products increases more sharply ($\propto A^3$). Thus, if we continue to raise A, the amplitude of the IM products eventually becomes equal to that of the fundamental tones at the output. As illustrated in Fig. B.4 on a log-log scale, the input level at which this occurs is called the "input third-order intercept point" (IIP3). Similarly, the corresponding output is represented by OIP3. In subsequent derivations, the input amplitude is going to be denoted as A_{IIP3}.

To determine the IIP3, one simply equate the fundamental and IM3 amplitudes:

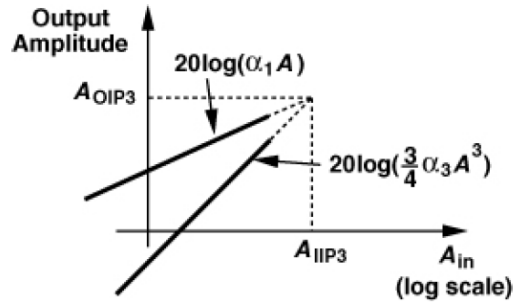$$|\alpha_1 A_{IIP3}| = \left|\frac{3}{4}\alpha_3 A_{IIP3}^3\right| \tag{B.17}$$

Figure B.4: Definition of IP3 (for voltage quantities).

obtaining

$$A_{IIP3} = \sqrt{\frac{3}{4}\left|\frac{\alpha_1}{\alpha_3}\right|} \tag{B.18}$$

Interestingly,

$$\frac{A_{IIP3}}{A_{1dB}} = \sqrt{\frac{4}{0.435}} \approx 9.6dB \tag{B.19}$$

This ratio proves helpful as a sanity check in simulations and measurements (note that this relationship holds for a third-order system and not necessarily if higher-order terms manifest themselves).

B.1.4.1   Cascaded Nonlinear Stages

Since in RF systems, signals are processsed by cascaded stages, it is important to know how the nonlinearity of each stage is referred to the input of the cascade. For the sake of brevity, the input IP3 is denoted by A_{IP3} unless otherwise noted.

Consider two nonlinear stages in cascade (Fig. B.5). If the input/output characteristics of the two stages are expressed, respectively, as
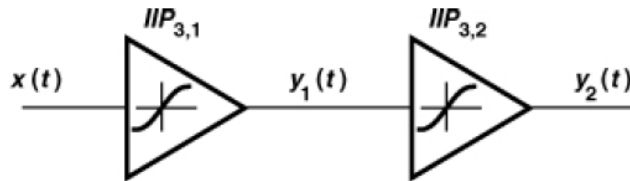

Figure B.5: Cascaded nonlinear stages.

$$y_1(t) = \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t) \tag{B.20}$$
$$y_2(t) = \beta_1 y_1(t) + \beta_2 y_1^2(t) + \beta_3 y_1^3(t) \tag{B.21}$$

then,

$$
\begin{aligned}
y_2(t) &= \beta_1[\alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t)] + \beta_2[\alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t)]^2 \\
&+ \beta_3[\alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t)]^3
\end{aligned} \tag{B.22}
$$

Considering only the first and third-order terms,

$$
y_2(t) = \alpha_1 \beta_1 x(t) + (\alpha_3 \beta_1 + 2\alpha_1 \alpha_2 \beta_2 + \alpha_1^3 \beta_3)x^3(t) + ... \tag{B.23}
$$

Thus, from B.18,

$$
A_{IP3} = \sqrt{\frac{4}{3}\left| \frac{\alpha_1 \beta_1}{\alpha_3 \beta_1 + 2\alpha_1 \alpha_2 \beta_2 + \alpha_1^3 \beta_3} \right|} \tag{B.24}
$$

Equation B.24 leads to more intuitive results if its two sides are squared and inverted:

$$
\begin{aligned}
\frac{1}{A_{IP3}^2} &= \frac{3}{4}\left| \frac{\alpha_3 \beta_1 + 2\alpha_1 \alpha_2 \beta_2 + \alpha_1^3 \beta_3}{\alpha_1 \beta_1} \right| \\
&= \frac{3}{4}\left| \frac{\alpha_3}{\alpha_1} + \frac{2\alpha_2 \beta_2}{\beta_1} + \frac{\alpha_1^2 \beta_3}{\beta_1} \right| \\
&= \left| \frac{1}{A_{IP3,1}^2} + \frac{3\alpha_2 \beta_2}{2\beta_1} + \frac{\alpha_1^2}{A_{IP3,2}^2} \right|
\end{aligned} \tag{B.25}
$$

where $A_{IP3,1}$ and $A_{IP3,2}$ represent the input IP3's of the first and second stages, respectively. Note that A_{IP3}, $A_{IP3,1}$ and $A_{IP3,2}$ are voltage quantities.

The key obsevation in B.25 is that to "refer" the IP3 of the second stage to the input of the cascade, one must divide it by α_1. Thus, the higher the gain of the first stage, the more nonlinearity is contributed by the second stage.

B.2 Noise

Noise is a signal whose instantaneous value is unpredictable. It comes from the probabilistic phenomena of charge carriers within electronic circuits. Its quantitative analysis is made through the use of statistics, considering most of the time its squared mean value or RMS value.

For electronic circuits purposes a good approximation is to consider the noise sources uncorrelated and the circuit linear, so that the superposition principle can be applied. Thus, one can evaluate the impact of each noise source separately at the output and sum all the effects to get the final result.

B.2.1 Noise sources

There are four main noise sources in eletronic circuits. Their circuit model is usually taken as electrical currents, which can also be turned into voltage, as shown below:

1. Shot noise $\rightarrow \overline{i_{noise}^2} = 2qI_{BIAS}\Delta\text{f}$;
2. Thermal noise $\rightarrow \overline{i_{noise}^2} = 4k\text{T}\frac{1}{R}\Delta\text{f}$;
3. Flicker noise $\rightarrow \overline{i_{noise}^2} = K_1 \frac{I_{BIAS}}{f^b}\Delta\text{f}$;
4. Popcorn $\rightarrow \overline{i_{noise}^2} = K_2 \frac{I_{BIAS}}{1+\left(\frac{f}{f_c}\right)^2}\Delta\text{f}$;

Where T is temperature, k is the boltzmann constant, Δf is the frequency band of interest, q is the electron's charge, R is the resistance and K is a constant. In this work, all transistors used were MOSFET, so thermal and flicker noises were dominant and the others neglected. For MOSFET flicker noise expression refer to appendix B.

B.3 Scattering Parameters

Microwave theory deals mostly with power quantities rather than voltage or current quantities. Two reasons can explain this approach. First, traditional microwave design is based on transfer of power from one stage to the next. Second, the measurement of high-frequency voltages and currents in the laboratory proves very difficult, whereas that of average power is more straightforward. Microwave theory therefore models devices, circuits and systems by parameters that can be obtained through the measurement of power quantities. They are called "scattering parameters" (S-parameters).
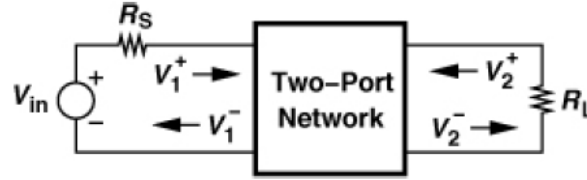

Figure B.6: Illustration of incident and reflected waves at the input and output.

For a given 2-port network (can be generalized to n-port network), Fig. B.6, the incident and reflected waves at the input port are denoted by V_1^+ and V_1^-, respectively. Similar waves are denoted by V_2^+ and V_2^-, respectively, at the output. Note that V_1^+ denotes a wave generated by v_{in} as if the input impedance of the circuit were equal to R_S (source resistance). Since that may not be the case, it's included the reflected wave, V_1^-, so that the actual voltage measured at the input is equal to $V_1^+ + V_1^-$. Also, V_2^+ denotes the incident wave traveling into the output port or, equivalently, the wave reflected from R_L (load resistance). These four quantities are uniquely related to one another through the S-parameters of the network:

$$V_1^- = S_{11}V_1^+ + S_{12}V_2^+ \tag{B.26}$$
$$V_2^- = S_{21}V_1^+ + S_{22}V_2^+ \tag{B.27}$$

For S_{11}, according to Fig. B.7,

$$S_{11} = \frac{V_1^-}{V_1^+}\big|_{V_2^+=0} \tag{B.28}$$

Thus, S_{11} is the ratio of the reflected and incident waves at the input port when the reflection from R_L is zero. This parameter represents the accuracy of the input matching.

For S_{12},

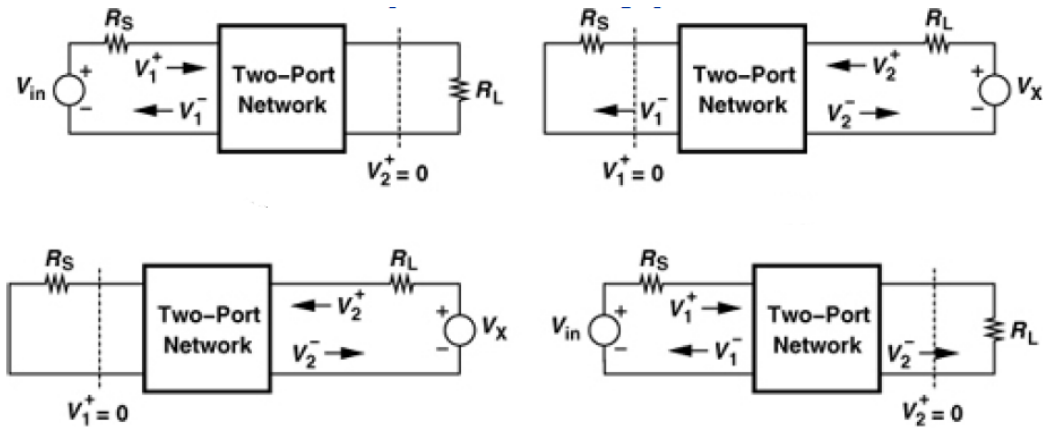$$S_{12} = \frac{V_1^-}{V_2^+}\big|_{V_1^+=0} \tag{B.29}$$

Figure B.7: Illustration of four S-parameters.

Thus, S_{12} is the ratio of the reflected wave at the input port to the incident wave into the output port when the input port is matched. In this case, the output port is driven by the signal source. This parameter characterizes the "reverse isolation" of the circuit.

For S_{22},

$$S_{12} = \frac{V_2^-}{V_2^+}\Big|_{V_1^+=0} \tag{B.30}$$

Thus, S_{22} is the ratio of reflected and incident waves at the output when the reflection from R_S is zero. This parameter represents the accuracy of the output matching.

For S_{21},

$$S_{12} = \frac{V_2^-}{V_1^+}\Big|_{V_2^+=0} \tag{B.31}$$

Thus, S_{21} is the ratio of the wave incident on the load to that going to the input when the reflection from R_L is zero. This parameter represents the gain of the circuit.

The condition $V_2^+ = 0$ requires that the reflection from R_L be zero, but it does not mean that the output port of the circuit must be conjugate matched to R_L. This condition simply means that if, hypothetically, a transmission line having a characteristic impedance equal to R_S carries the output signal to R_L, then no wave is reflected from R_L. A similar note applies to the requirement $V_1^+ = 0$. The conditions $V_1^+ = 0$ at the input or $V_2^+ = 0$ at the output facilitate high-frequency measurements while creating issues in modern RF design. Which typically does not strive for matching between stages. Thus, if S_{11} of the first stage must be measured with $R_L = R_S$ at its output, then its value may not represent the S_{11} of the cascade.

In modern RF design, S_{11} is the most commonly used S-parameter as it quantifies the accuracy of impedance matching at the input of receivers. Consider the arrangement shown in Fig. B.8, where the receiver exhibits an input impedance of Z_{in}. The incident wave V_1^+ is given by $V_{in}/2$ (with $Z_{in} = R_S$). Moreover, the total voltage at the receiver input is equal to $V_{in}Z_{in}/(Z_{in} + R_S)$, which is also equal to $V_1^+ + V_1^-$. Thus,

$$V_1^- = V_{in}\frac{Z_{in}}{Z_{in}+R_S} - \frac{V_{in}}{2} \tag{B.32}$$

$$= \frac{Z_{in}-R_S}{2(Z_{in}+R_S)}V_{in} \tag{B.33}$$

It follows that

$$\frac{V_1^-}{V_1^+} = \frac{Z_{in}-R_S}{Z_{in}+R_S} \tag{B.34}$$

Called the "input reflection coefficient", this quantity can also be considered to be S_{11} if the condition $V_2^+ = 0$ is removed.
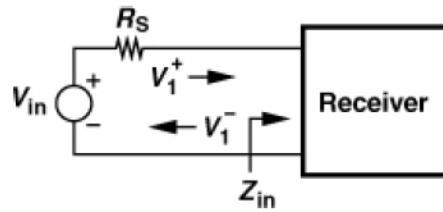

Figure B.8: Receiver with incident and reflected waves.

B.4 Stability

The LNA must remain stable for all source impedances at all frequencies. One may think that the LNA must operate properly only in the frequency band of interest and not necessarily at other frequencies, but if the LNA begins to oscillate at any frequency, it becomes highly nonlinear and its gain is very heavily compressed.

In the presence of a front-end band-select filter, the LNA sees smaller changes in the source impedance. A parameter often used to characterize the stability of circuits is the "Stern stability factor", defined as

$$K = \frac{1+|\Delta|^2 - |S_{11}|^2 - |S_{22}|^2}{2|S_{21}||S_{12}|} \tag{B.35}$$

where $\Delta = S_{11}S_{22} - S_{12}S_{21}$. If K > 1 and $\Delta<1$, then the circuit is unconditionally stable, i.e., it does not oscillate with any combination of source and load impedances. In modern RF design , on the other hand, the load impedance of the LNA (the input impedance of the on-chip mixer) is relatively well-controlled, making K a pessimistic measure of stability. Also, since the LNA output is typically not matched to the input of the mixer, S_{22} is not a meaningful quantity in such an environment.

APPENDIX C   MOS TRANSISTOR SMALL SIGNAL MODELS

In this chapter, the transistor models and noise sources considered in hand calculations are presented. The whole project of the LNA was initially calculated by hand and then adjusted in simulations using Cadence Virtuoso environment.

C.1   Low Frequency and High Frequency Models

The low frequency model used in hand calculations is shown in Fig. C.1. This model was used to calculate small signal gains of each transistor at low frequencies.
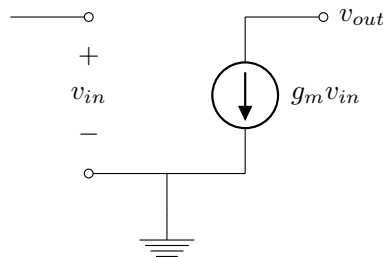

Figure C.1: Low frequency model used for hand calculations.

The large signal model has the same shape, putting a quadratic current source instead of a linear one. The r_{ds} was not taken into account because it was not necessary. As most hand calculations were done using approximated expressions, the aim was to figure out how to minimize noise figure with the best IIP3 possible. The inclusion of r_{ds} would not change the main results, leading only to more complicated calculations.
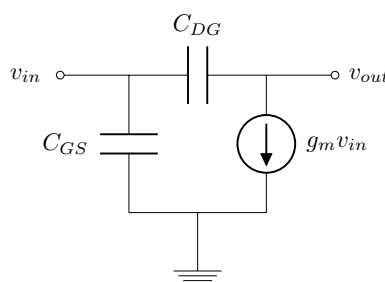

Figure C.2: High frequency model used for hand calculations.

The high frequency model used in hand calculations is shown in Fig. C.2. The difference from the low frequency model is the inclusion of the parasitic capacitances C_{GS} and

C_{DG}. This is also a simple model, however this was enough for the purpose of this work.

C.2 Noise Sources

Only the main noise sources of the MOSFET were considered, i.e., the flicker and thermal noise. This work aimed to cancel thermal noise of the main amplifier and optimized the flicker noise of the first stage in the cascaded auxiliary amplifier. It is showed below the power spectral density for each type of noise.

C.2.1 Flicker Noise

The flicker noise equation used in hand calculations was:

$$\overline{i^2_{n,flicker}} = \frac{K g_m^2}{C_{ox} W L} \tag{C.1}$$

This equation shows the stronger dependance on g_m than on W*L.

C.2.2 Channel Thermal Noise

The channcel thermal noise equation used in this work was:

$$\overline{i^2_{n,thermal}} = 4kT\gamma g_m \tag{C.2}$$

Where k is the boltzmann constant, T is temperature, γ is 2/3 for long-channel transistors and may need to be replaced by a larger value for submicron MOSFETs. The theoretical determination of γ is still under active research.