22-32 GHz Low-Noise Amplifier Design in 22-nm CMOS-SOI Technology

by

Bolun Cui

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2019

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

This thesis explores the use of a 22-nm CMOS-SOI technology in the design of a two-stage amplifier which targets wide bandwidth, low noise and modest linearity in the 28 GHz band.

A design methodology with a transformer-coupled, noise-matching interstage is presented for minimizing the noise factor of the two-stage amplifier. Furthermore, benefits of interstage noise matching are discussed. Next, a transistor layout for minimizing noise and maintaining sufficient electromigration reliability is described. It is followed by an analysis of transformer configurations and a transformer layout example is depicted.

To verify the design methodology, two amplifier prototypes with noise-matching interstage were fabricated. Measurement shows that the first design achieves a peak gain of 20.7 dB and better-than-10-dB input and output return losses within a frequency range of 22.5 to 32.2 GHz. The lowest noise figure of 1.81 dB is achieved within the frequency range. Input IP_3 of -13.4 dBm is achieved with the cost of 17.3 mW DC power consumption. When the bias at the back-gate is lowered from 2 V to 0.62 V, the power consumption is decreased to 5.6 mW and the peak gain drops down to 17.9 dB. Minimum noise figure increases from 1.81 to 2.13 dB and input IP_3 drops to -14.4 dBm.

The folded output stage in the second design improves the input IP_3 to -6.7 dBm at the cost of 35 mW total power consumption. The peak gain of the second design is 20.1 dB, and the lowest noise figure of 1.73 dB within a frequency range of 23.8 to 32.4 GHz. Both designs occupy about 0.05 mm^2 active area.

Acknowledgements

I would like to express my utmost gratitude to my supervisor Dr. John Long for offering me the opportunity to explore the world of RF IC, as well as his invaluable support, encouragement and guidance throughout my Master's studies. I would also like to thank Dr. Peter Levine and Dr. Slim Boumaiza for reading my thesis and providing valuable feedback.

I would like to thank all my colleagues and friends in the RFIC group for their company and support. I would like to thank especially Hassan and Nansen for their technical knowledge and experience of dealing with the PDK and tool setup, Aman for his technical knowledge and discussion of design documents, Shaeera, Yan, Ayman, Yushi and Stanley for their help and valuable discussion.

Finally, I would like to thank my parents for their deepest love and unconditional support over the years. I also extend my honest appreciation to my grandparents for their selfless support and dedication. Special thanks to my cousin for her support.

I wish this thesis can create some inspiration for other designers.

To the B. C. in 2016, who took the leap

To my parents and my grandparents

To J. M. and Y. L., the AitD

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this chapter, a brief introduction to the thesis topic is given and the motivations for the research project are outlined. The semiconductor process used in this work is also introduced. Then, the design objectives are defined quantitatively.

1.1 Motivation

Demand for services delivered over the internet is driving the development of new communication infrastructure. Among the telecommunication sectors, wireless communication is prominent in mobility and flexibility, and is the largest sector in Canada. As reported in [1], wireless services account for 52% of all retail telecommunications service revenues. New genres such as social networking, live streaming, and electronic commerce are becoming irreplaceable in daily life as technology evolves. These applications require enormous data throughput. In Canada, the average data usage per subscriber has reached around 1.2 GB per month (up from 981 MB in 2015) with a 25% compound annual growth rate [1]. It is reasonable to predict that this growth rate will continue with new market drivers such as the internet of things, automotive, and wearable electronics. Therefore, wireless systems with the capability of handling large data throughputs are currently the subject of intense research & development, and is the application domain for the work in this thesis.

The 22FDX® technology developed by GLOBALFOUNDRIES® is a 22-nm, fully-depleted silicon-on-insulator (FD-SOI) CMOS process [2]. It features performance and power consumption comparable to FinFET technologies, but at the cost of a 28-nm planar technology while consuming up to 70% lower power (versus 28 nm, in digital applications) [2]. By fully depleting the transistor channel, supply power leakage is reduced,

which favours digital applications. Non-idealities such as the kink effect seen in partially-depleted SOI processes are resolved in FD-SOI [3]. Shrinking the channel length to 22 nm leads to a high transition frequency (f_T), which enables high-performance RF and millimetre-wave (mm-wave) circuit designs. The ultra-thin buried oxide (BOX) underneath the channel raises the well breakdown voltage, which permits integration of a power amplifier by cascoding [2]. The capability to bias the transistor back-gate offers a way to trade-off performance and power on the fly, which provides more flexibility in a system [2].

The purpose of this thesis is to design an integrated, single-ended amplifier demonstrator in the 22-nm CMOS-SOI technology for radio front-end, and attempt to attain lowest noise figure and highest bandwidth. The novelty of the amplifier design methodology is that it utilizes noise-matching technique at the interstage, which provides lower noise figure than the traditional high-gain interstage. Priority in the design is given to wide bandwidth and low noise, with sufficient linearity, low power consumption and small chip area. A major motivation of this thesis is to benchmark the 22-nm CMOS-SOI technology performance in an emerging RF application.

1.2 Design Objectives and Challenges

Typical specifications for an RF system [4] consist of linearity (i.e., intercept point), noise figure, power consumption, operating frequency, bandwidth, gain and supply voltage.

1.2.1 Operating Frequency and Bandwidth

As mentioned in Section 1.1, the main trend in the development of wireless communication systems is supporting increasing data throughput. The data rate (C) across a communication channel is proportional to the bandwidth (BW), and the logarithm of the signal-to-noise ratio (SNR), as described by Shannon's equation [5],

$$C = BW \times log_2(1 + SNR). \tag{1.1}$$

Therefore, a wide bandwidth is the main design objective for a high-speed radio link. Fractional bandwidth is also commonly used to compare the performance of such systems. It is defined as the absolute bandwidth divided by centre frequency. For a fixed fractional bandwidth, the absolute bandwidth increases with increasing the centre frequency, thus high data throughput can be achieved. On the other hand, the sub-6 GHz spectrum

is crowded with existing communication standards, which prompts industry to move to higher frequency bands. As a result, mm-wave frequency bands are being considered for the deployment of future generations of radio systems.

However, transistor performance metrics such as gain and noise figure deteriorate rapidly as the operating frequency increases. For passive components, leakage through oxide layers, substrate coupling and substrate losses also significantly degrade their RF performance. Moreover, ground and power plane impedances can affect the performance of a single-ended system. All of these factors have to be taken into consideration, which complicates the design and implementation process.

Overall, 28 GHz is selected as the centre frequency of operation because it has some relevance to 5G mm-wave bands. The highest bandwidth and lowest noise figure achievable are targeted around this centre frequency.

1.2.2 Noise Figure

Aside from gain and bandwidth, noise figure (NF) is a design objective with higher priority over other specifications for this design. Noise factor (F) is defined as the SNR at the input divided by the SNR at the output of an amplifier,

$$F = \frac{S_i/N_i}{S_o/N_o} \quad [4]. \tag{1.2}$$

Note that the noise factor is a measure of the noise power added by the system to the input noise power. By convention, 50 Ω is usually chosen as the input source resistance, which determines the input noise power ($N_i = kT_0B$) at a source temperature of T_0 in Kelvin (k is Boltzmann's constant, and B is bandwidth in Hertz). When a bandwidth of 1 Hz is chosen, the NF may be referred to as the spot noise. Noise figure is the decibel representation of the noise factor,

$$NF = 10 \times log_{10}(F). \tag{1.3}$$

Receiver (RX) sensitivity is determined by the SNR (and thus noise figure), and plays a role in determining the upper limit of the communication channel capacity, as seen from Eq. 1.1. For an RX consisting of cascaded circuit blocks (as shown in Fig. 1.1), the overall noise factor may be expressed by Friis' equation, where F_x and N_x denote the noise factor and power gain of each stage, respectively,

$$F = F_{switch} + \frac{F_{LNA} - 1}{G_{switch}} + \frac{F_{PhaseShifter} - 1}{G_{switch}G_{LNA}} + \frac{F_{RFCombiner} - 1}{G_{switch}G_{LNA}G_{RFCombiner}} + ... \quad . \quad (1.4)$$

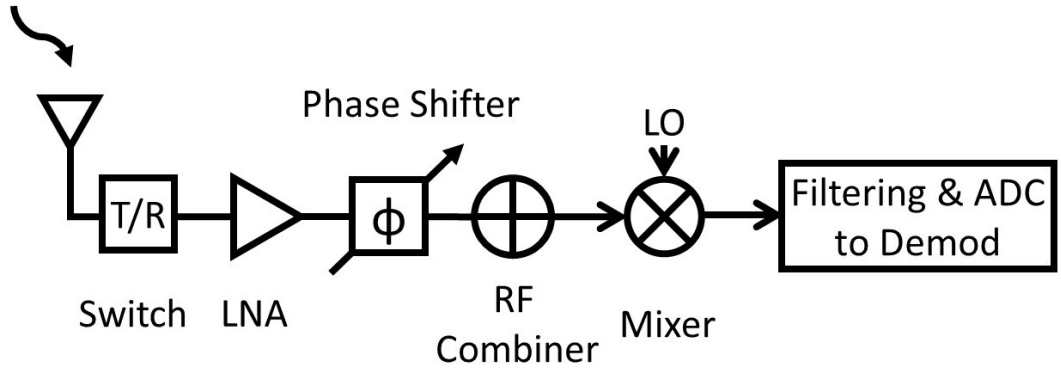


Figure 1.1: A slice of a phased-array RF receiver

For a wideband system, achieving low noise figure across the entire band is difficult because the optimum source impedance for minimum noise added by the transistor varies with frequency. In addition, ohmic losses in passive circuit components contribute noise that is determined partly by the losses of the interconnect metal stack in a given technology.

At mm-wave frequencies, cascading multiple stages with conjugate interstage impedance matching is the most common approach for amplifier design. The reason is that a conjugate match provides high gain for the first stage, which suppresses the noise contributed by later stages (as seen in the Friis equation, 1.4). However, gain decreases as operating frequency increases and transistor gate length decreases (i.e., process scaling) [6]. As a result, noise suppression by the gain of the first stage also decreases. Therefore, the interface between the first and second stages is designed for lowest noise (contributed by the second stage) instead of high-gain (from the first stage) in this work, to optimize the low-noise performance of the amplifier. To the best of author's knowledge, multistage noise-matching has not yet been realized in implementing amplifier operating at 28 GHz.

The target noise figure of this work is set to as low as possible. The upper bound of the noise figure is chosen to be 2 dB after a literature survey on recent low-noise amplifiers (LNAs) in a phased-array receiver application [7]-[9].

1.2.3 Linearity and Power Consumption

Non-linearity of an RF amplifier (e.g., due to gain compression) causes harmonic and intermodulation distortions. Among the different distortion types, third-order intermodulation distortion (IMD_3) is often the most critical in a radio receiver because it lies within the same band as the desired signal. Thus, it cannot be easily filtered out, as shown in Fig. 1.2. As a result, the RX is desensitized to the desired signal, and fidelity suffers (e.g., poor signal quality, higher bit-error-rates, dropped calls, etc.).

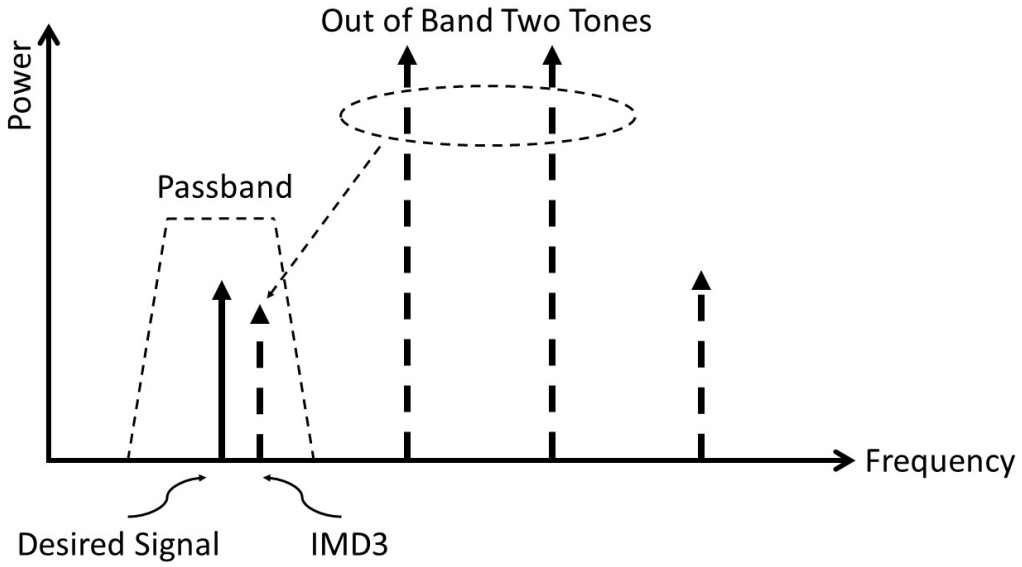


Figure 1.2: Interference due to the third-order intermodulation

To specify this non-linear effect, the third-order intercept point (IP_3) is used. It is defined as the input (or output) power where the extrapolation of the fundamental and the IM_3 powers intercept, as shown in Fig. 1.3. However, mixed products generated by other odd-order distortions may also appear at the same frequency. Thus, this method is only valid assuming that IM_3 is entirely contributed by the third-order component and that the distortion grows exactly three times as fast as the fundamental power. Assuming that P_{tone} represents the output power at tone frequencies, IP_3 and IM_3 at the output are related by

$$OIP_3 = P_{tone} + \frac{P_{tone} - IM_3}{2} \quad [4]. \tag{1.5}$$

Higher IP_3 corresponds to lower IM_3 because the system is more linear. Similar to noise figure, IP_3 can be referred to the output (as OIP_3) or to the input (as IIP_3). The OIP_3 is higher than IIP_3 by the power gain in the system.

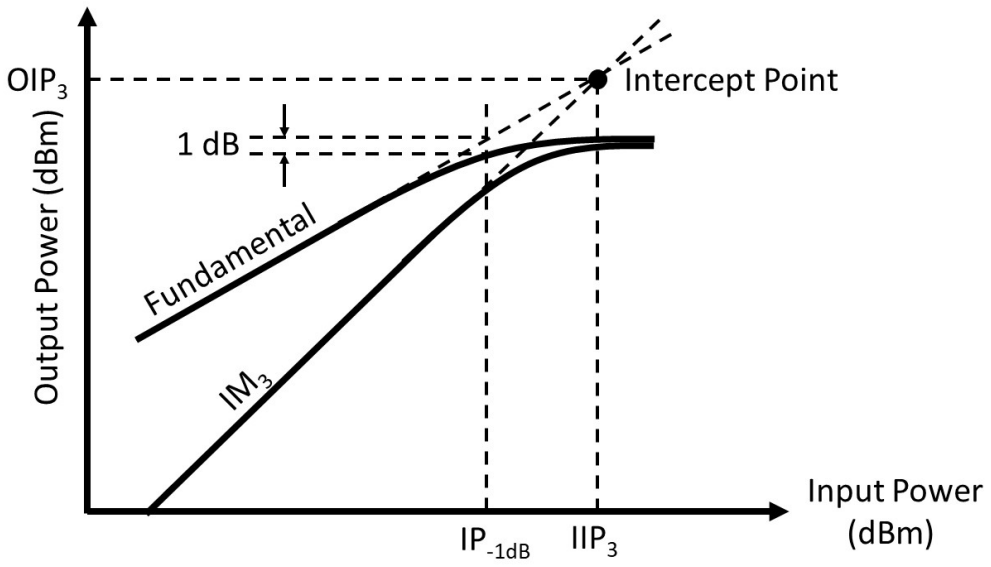


Figure 1.3: Third-order intercept point and 1-dB compression point

Another metric to describe linearity is the 1-dB gain compression point ($P_{\text{-1dB}}$). Amplifiers suffer from output power saturation as the input power increases. The $P_{\text{-1dB}}$ defines the power level where the gain of the amplifier is 1-dB lower than its small-signal gain, as shown in Fig 1.3. The IP_3 and $P_{\text{-1dB}}$ are often related by

$$OIP_3 = OP_{-1dB} + 9.6\,dB \quad [4]. \tag{1.6}$$

If the signal applied to the amplifier input from the antenna has little signal power, then the amplifier is working in the small-signal regime (i.e., the amplifier does not suffer from the gain compression). Thus, the requirement for linearity may not be as high, a priority as bandwidth or noise figure.

Linearity can be improved by Class-A biasing of the output stage. This ensures that the MOS transistor always operates in saturation, and increasing the supply voltage will create more output headroom. However, these approaches increase the DC power consumption (P_{DC}). The 22-nm CMOS-SOI process used in this work features a back-gate bias which

enables dynamic control of the transistor's threshold voltage [10]. This can be used to trade-off noise performance and power.

For low-noise design, transistors are typically biased at the optimum current per transistor width ($\frac{I_{DS}}{W}_{opt}$), which has a unit of mA/μm where minimum noise figure ($\mathrm{NF}_{\mathrm{min}}$) occurs. As process technologies scale down, the transistor width has to be sized larger to realize the total gate area needed for 50-Ω noise matching. This leads to higher DC current and power consumption, which makes low-power design challenging, but generally favours linearity.

Since the design in this work targets hand-held mobile devices, the overall power consumption is set as low as possible with a limit of 40 mW after surveying recent literature on LNAs for similar applications. The goal for linearity is to achieve the best result given the limited power budget.

1.2.4 Gain and Tunability

One of the main motivations for placing the amplifier at the input to the RX chain is to suppress noise from subsequent stages, as justified by equation 1.4. Therefore, sufficient amplifier gain is also necessary. If the amplifier consists of multiple stages, the overall gain is determined by the gain of each stage and loading on each stage. The tunability of the gain may be achieved by biasing the back-gate as it adjusts the threshold voltage and thus transconductance of the transistor. However, changes in threshold voltage also affect the DC bias current and the transistor may no longer be biased at $\frac{I_{DS}}{W}_{opt}$. Noise figure will increase as a consequence.

The minimum gain specification is set to 15 dB from reviewing recent mm-wave LNAs (in a phased-array receiver application [7]-[9], stand-alone: [11]-[15]). The bias voltage at the back-gate can be set to above the nominal V_{DD} of the system. Thus, explicit voltage generation and regulation circuits are required, which leads to some power and area overhead. To sum up, tunability is therefore optional.

1.2.5 Other Objectives

The amplifier in this thesis work is designed to interface with other blocks in the receiver front end, such as a T/R switch and a mixer. From the measurement point of view, setting input and output impedances to be 50 Ω makes the device easier to measure, since test instruments are generally designed for 50 Ω interfaces. Thus, return losses are designed to

be better than 10 dB (i.e., $S_{11},\ S_{22} < -10$ dB) across the amplifier's bandwidth. Meeting this specification is challenging for a wideband system.

1.3 Thesis Organization

This thesis consists of six chapters. In chapter two, a literature survey of previous work in mm-wave low-noise amplifier (LNA) design will be reviewed. Moving on to chapter three, theoretical derivation and a summarized design methodology are presented for two amplifier prototypes with their simulation results. The fourth chapter illustrates the challenges during the layout and implementation of the two designs. Chapter five outlines the test plan and measurement results followed by an analysis of discrepancies between simulation and measurement results. Finally, the last chapter concludes this thesis with a proposed plan for future work.

Chapter 2

Literature Summary on Low-Noise Amplifiers

In this chapter, a literature survey of previous work in RF and mm-wave low-noise amplifier (LNA) design is presented.

2.1 Overview

An LNA can be represented as a two-port network driven at the input by a signal source with a complex impedance $Z_s = R_s + jX_s$, and loaded at the output with an impedance Z_L. There is an optimum source impedance $Z_{opt} = R_{opt} + jX_{opt}$, where the minimum noise factor (F_{min}) occurs for the transistor. The F_{min} of the transistor is determined by the semiconductor process, biasing condition and device dimensions. To minimize the noise factor of the amplifier, one approach is to minimize F_{min} of the transistor through proper biasing and scaling, and to present Z_{opt} to the input of the amplifier by an input matching network.

2.1.1 Biasing Condition for Low-Noise Designs

For a FET, the minimum noise factor can be expressed as [16]

$$F_{min} = 1 + 2r_{gg}g_m\gamma(\frac{w_o}{w_T})^2 + 2\frac{w_o}{w_T}\sqrt{r_{gg}g_m\gamma(1 + r_{gg}g_m\gamma(\frac{w_o}{w_T})^2)}; \tag{2.1}$$

When $r_{gg}g_m\gamma(\frac{w_o}{w_T})^2 \ll 1$ in Eq. 2.1,

$$F_{min} \approx 1 + 2r_{gg}g_m\gamma(\frac{w_o}{w_T})^2 + 2\frac{w_o}{w_T}\sqrt{r_{gg}g_m\gamma}; \tag{2.2}$$

When $r_{gg}g_m\gamma(\frac{w_o}{w_T})^2 \gg 1$ in Eq. 2.1,

$$F_{min} \approx 1 + 4r_{gg}g_m\gamma(\frac{w_o}{w_T})^2, \tag{2.3}$$

where ω_o and ω_T are the operating frequency and transition frequency of the transistor, respectively. The gate resistance is represented as r_{gg}, g_m is the transconductance and γ is the excess (drain) channel noise parameter, which depends upon the semiconductor process. Among these parameters, ω_T, g_m, and r_{gg} are circuit design variables. The ω_T and g_m of the transistor can be adjusted via setting the DC bias point and aspect ratio of the transistor. The gate resistance is given by [4]

$$r_{gg} = \frac{\rho_{sh}W}{L}, \tag{2.4}$$

where ρ_{sh} represents the sheet resistance of the gate material (in $\Omega/\square$), W and L are the width and length of the transistor, respectively. The gate resistance may be reduced by splitting the gate into multiple fingers (M) and applying double-gate contacts. In addition, the distributed R-C effect of the gate resistance and capacitance (which forms a low-pass filter) has to be taken into consideration at radio frequencies. The equation describing r_{gg} becomes [4]

$$r_{gg} = \frac{1}{12}\frac{\rho_{sh}W}{ML}. \tag{2.5}$$

In [17], the optimum current per transistor width ($\frac{I_{DS}}{W}_{opt}$) instead of the drain current (I_{DS}) is used, because the goal here is to find the biasing condition where the minimum F_{min} occurs. The relation between the biasing voltage (V_{eff}) and the drain current in saturation region can be expressed as [18]

$$I_{DS} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{TN})^2[1 + \lambda(V_{DS} - V_{GS} + V_{TN})], \tag{2.6}$$

where $\mu_{\rm n}$, C_{ox}, V_{TN} and λ represent the carrier mobility, gate oxide capacitance per unit area, threshold voltage, and channel length modulation factor, respectively. Neglecting

channel length modulation and assuming the transistor operates in the saturation region, it can be shown that

$$V_{eff} = V_{GS} - V_{TN} \approx \sqrt{\frac{2L}{\mu_n C_{ox}} \frac{I_{DS}}{W}}, \tag{2.7}$$

which indicates that the biasing point is a function of the current per transistor width ($\frac{I_{DS}}{W}_{opt}$) (assuming that μ_n and C_{ox} are constant). The drain-to-source voltage (V_{DS}) from Eq. 2.6 also affects the drain current, which leads to changes in $\frac{I_{DS}}{W}_{opt}$, as confirmed in [17].

To sum up, transistors are typically biased close to $\frac{I_{DS}}{W}_{opt}$ in low-noise designs. Note that biasing the transistor at $\frac{I_{DS}}{W}_{opt}$ does not guarantee the overall noise factor of the amplifier to be at its minimum, because losses from the input-matching network are not included in the analysis.

2.1.2 Input-Matching Condition for Low-Noise Designs

The optimum source impedance ($Z_{opt} = R_{opt} + jX_{opt}$) is given by [16][17]

$$R_{opt} = \frac{\omega_T}{M\omega_o}\sqrt{\frac{r'_{gg}}{g'_m \gamma}} = \frac{1}{\omega_o M(C'_{gs} + C'_{gd})}\sqrt{\frac{r'_{gg} g'_m}{\gamma}}, \tag{2.8}$$

$$X_{opt} = \frac{\omega_T}{M\omega_o g'_m} = \frac{1}{\omega_o M(C'_{gs} + C'_{gd})}, \tag{2.9}$$

where r'_{gg}, C'_{gs}, C'_{gd}, g'_m represents the gate resistance, gate-to-source capacitance, gate-to-drain capacitance, and transconductance of a unit-size transistor, respectively. To reduce noise factor, Z_{opt} should be presented to the input of the amplifier.

Since the source impedance (Z_s) is typically 50 Ω, the transistor can be sized properly by adjusting M and g'_m to match R_{opt} to 50 Ω. Note that for a unit-size transistor, R_{opt} is inversely proportional to the channel length (i.e., as technology scales down, R_{opt} increases). Assuming that the transistor is biased at $\frac{I_{DS}}{W}_{opt}$, it has to be sized very wide in deep sub-micron technology to match R_{opt} to 50 Ω, which increases the drain current. This leads to a trade-off between low-noise and power consumption. Also, a large transistor area may

cause problems such as degradation in gain and noise due to the Miller effect at radio frequencies.

The optimum (noise) source reactance is equal to the input reactance of the FET. Thus, an inductor in series with the source resistance is often used as the input matching network. The ohmic loss of the inductor contributes significantly to the noise factor of the amplifier, and is proportional to the size of the coil. Note that, if the transistor size is large, its input reactance ($\frac{1}{sC}$) becomes low. Thus, a low inductance value is needed, which leads to low ohmic loss, and reduces the noise factor of the amplifier (and vice versa). Therefore, this also leads to the same trade-off (between noise and power consumption) as mentioned in the previous paragraph.

2.2 Previous Work on RF Low-Noise Amplifiers

The common-source NFET with inductive degeneration is probably the most commonly-used topology for LNA designs. As shown in Fig. 2.1,

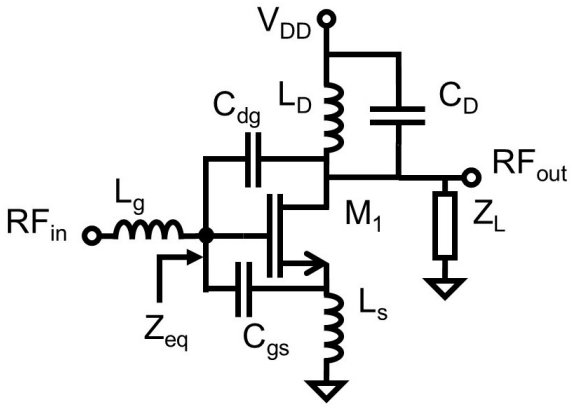


Figure 2.1: Common-source amplifier with inductive degeneration

Neglecting channel length modulation, the equivalent impedance looking into the gate of the transistor (Z_{eq}) can be expressed as [4]

$$Z_{eq}(s) = \frac{(Z_L + \frac{1}{sC_{dg}})(s^2 L_s C_{gs} + g_m L_s + 1)}{s^2 L_s C_{dg} + s(Z_L C_{gs} + g_m L_s) + g_m Z_L + \frac{C_{gs}}{C_{dg}} + 1}. \tag{2.10}$$

Neglecting current fed back through C_{dg} and assuming $Z_L \ll \frac{1}{sC_{dg}}$, the above equation can be approximated to

$$Z_{eq}(s) \approx \frac{g_m L_s}{C_{gs}} + \frac{1}{sC_{gs}} + sL_s, \tag{2.11}$$

where the real part of Z_{eq} becomes $\frac{g_m L_s}{C_{gs}}$. Since L_s and C_{gs} does not generate thermal noise, a noiseless 50 Ω can be realized and used for conjugate input match. The gate inductor L_g may be selected to match the source reactance to X_{opt}, and bring the imaginary part of the input impedance to zero. As a result, the input is simultaneously noise and conjugate matched to 50 Ω.

There are several drawbacks to this topology. As the operating frequency approaches the transition frequency of the transistor, the gain of the transistor rolls off, which leads to insufficient gain at radio frequencies. One solution to counter this problem is to cascade more stages, which would involve complicated interstage interfacing as well as much greater power consumption. If C_{dg} is not neglected in Eq. 2.10 and assuming that the denominator is dominated by $(1 + \frac{C_{gs}}{C_{dg}})$, Z_{eq} drops by a factor of $(1 - \frac{2C_{dg}}{C_{gs}})$, which reduces the input return loss. Another problem created by C_{dg} is the Miller effect, which increases in the equivalent input capacitance due to the voltage gain across the amplifier, thereby limiting the amplifier bandwidth. The equivalent input capacitance due to the (Miller) feedback capacitor (C_{dg}) shown in Fig. 2.1 is given by [18]

$$C_{eq} = C_{dg}(1 + |A_v|), \tag{2.12}$$

where A_v is the voltage gain. The presence of C_{dg} also affects the noise performance. Fig. 2.2 illustrates the common-source amplifier including the drain current noise source for analysis.

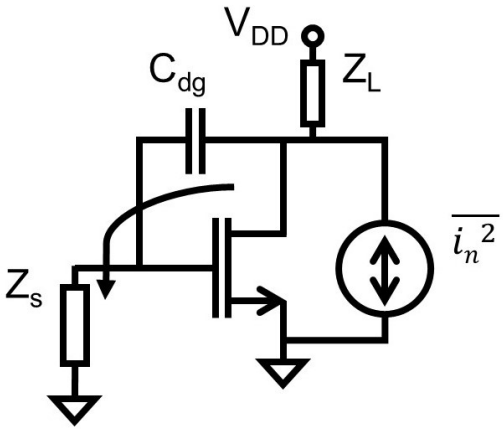


Figure 2.2: Noise fed back through C_{dg} in a common-source amplifier (Simplified)

The presence of C_{dg} and the finite source impedance (Z_s) form a feedback path from the output back to the input for the equivalent noise current source of the transistor. The noise power fed back to the input is given by

$$\overline{v_n^2} = 4kT\gamma g_m Z_L^2 (\frac{Z_s}{Z_s + \frac{1}{sC_{dg}}})^2. \tag{2.13}$$

Therefore, a noise voltage is presented at the input which degrades the noise performance of the common-source amplifier at radio frequencies. Moreover, the feedback also reduces the reverse isolation of the amplifier.

2.2.1 Cascode LNA

To resolve problems such as insufficient gain, bandwidth limitation due to the Miller effect, and poor reverse isolation with the common-source topology, a common-gate amplifier is connected in series with the common-source transistor to form a cascode topology, as depicted in Fig 2.3.

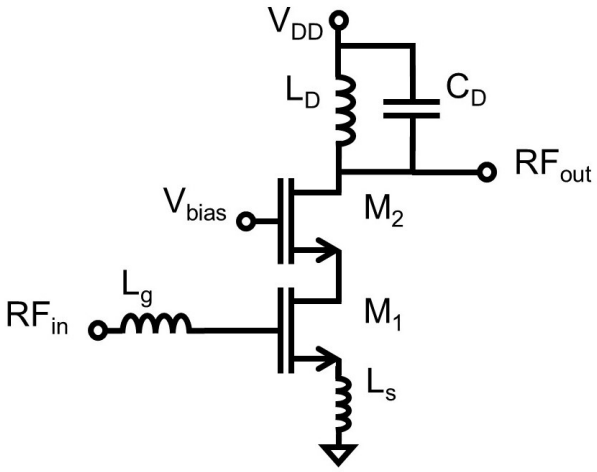


Figure 2.3: Cascode amplifier with inductive degeneration

Compared to the common-source topology, the cascode amplifier can realize higher gain due to the increase in output impedance (note that, the output impedance will be limited by the quality factor of L_D). The low input impedance of the common-gate FET (M_2) driven by the output of the common-source FET (M_1) reduces the voltage gain of the common-source stage. Thus, the Miller effect on M_1 is suppressed which improves the bandwidth. Reverse isolation is also improved due to the high isolation across M_2. However, noise performance is degraded due to noise added by the common-gate stage. Fig. 2.4 shows a simplified small-signal schematic of the common-gate stage with the drain current noise source emphasized.

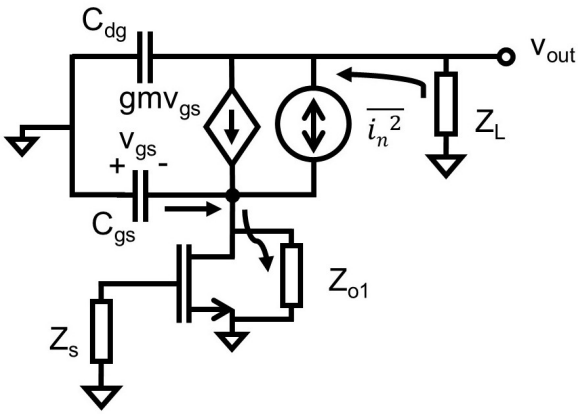


Figure 2.4: Noise of the common-gate stage in a cascode amplifier (Simplified)

Assuming that the output impedance (Z_{o1}) of the common-source stage is finite, there will always be a portion of the noise current flowing through load Z_L which increases the overall noise factor. The noise contributed by the common-gate stage can become more critical as frequency increases, because Z_{o1} of the common-source stage decreases as the frequency increases. Note that the impedance of C_{gs} (i.e., Z_{gs}) of the common-gate stage shown in Fig. 2.4 also affects the noise voltage at the output, as it provides a path to AC ground for the noise current. To optimize the noise performance of the cascode, Z_{o1} of the common-source and Z_{gs} of the common-gate stage must be taken into consideration, especially at mm-wave frequencies.

Although the cascode topology resolves many problems occurred in the common-source amplifier, one noticeable drawback is its linearity. For a fixed supply voltage V_{DD}, adding a transistor M_2 in series with M_1 causes a reduction of the output headroom due to the extra bias voltage required for the common-gate stage. The supply voltage has to be doubled for the same output headroom as a single transistor amplifier, which leads to greater power consumption.

In [19], a cascode LNA is implemented with differential input and output in 180-nm CMOS technology. It achieved a gain of 14.1 dB, the noise figure is 1.8 dB and IIP_3 equals 4.2 dBm for a supply voltage of 1.8 V at 21.6 mW power consumption.

2.2.2 Transformer-Feedback LNA

Another enhancement to the common-source topology to eliminate the unwanted feedback through the Miller capacitor is to apply transformer feedback. In [19], drain and source

terminals of the common-source amplifier are coupled through a transformer.

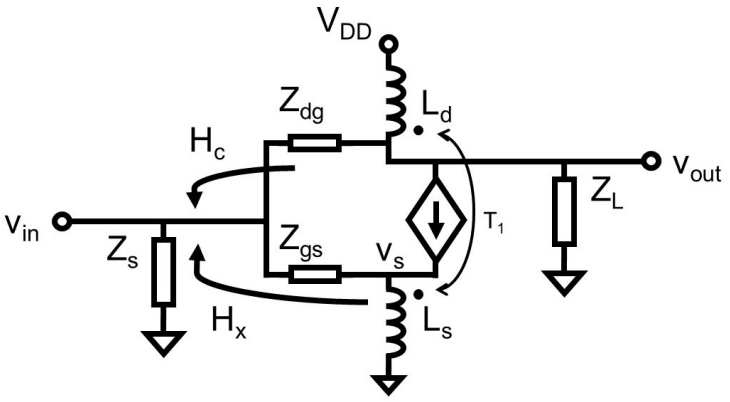


Figure 2.5: Reverse signal flows in drain-source transformer (Ideal) coupled common-source LNA (Simplified)

Neglecting channel length modulation and assuming that the transformer is ideal, the additional output signal fed back to the input (H_x) counteracts the capacitive feedback via C_{dg} (H_c), which are expressed as

$$H_c \approx \left.\frac{v_{in}}{v_{out}}\right|_{v_s=0} = \frac{Z_s||Z_{gs}}{Z_s||Z_{gs}+Z_{dg}}, \tag{2.14}$$

$$\text{and } H_x \approx \left.\frac{v_{in}}{v_{out}}\right|_{H_c=0} = \frac{-\frac{1}{n}(Z_s||Z_{gd})}{Z_s||Z_{gd}+Z_{sg}}. \tag{2.15}$$

Assuming that $C_{gs} = C_{sg}$, $C_{dg} = C_{gd}$ and the transformer is ideal, setting $H_c = -H_x$ leads to

$$\frac{C_{gs}}{C_{dg}} \approx n. \tag{2.16}$$

As a result, the effect of the Miller capacitor is nulled (i.e., C_{dg} is neutralized). Gain and reverse isolation are improved at radio frequencies. Furthermore, noise current through C_{dg} is suppressed and the output headroom is maintained without the use of the cascode topology. Another advantage of using transformer feedback for C_{dg} neutralization is its wide bandwidth. From 2.16, the condition for C_{dg} neutralization only depends on the turns

ratio of the transformer. That is, the bandwidth of C_{dg} neutralization is determined by the bandwidth of the transformer.

In monolithic transformers, the ohmic loss and parasitic capacitances can become critical and limit the bandwidth, which is a potential problem for transformer-coupled amplifiers. The ohmic loss generates thermal noise, which contributes to the amplifier noise factor.

As reported in [19], a differential transformer coupled LNA achieved similar gain as the differential cascode topology (14.2 dB and 14.1 dB, respectively). However, the noise figure has been improved from 1.8 dB to 0.9 dB. With a lower supply voltage of 1 V, a slightly lower IIP_3 of 0.9 dBm is attained with 16 mW power consumption, down from 21.6 mW in the cascode case.

Transformers can also be used to implement wideband matching networks. In [20], a transformer is used at the input of the amplifier to couple gate and source terminals, as shown in Fig. 2.6, where C_{dg} and channel length modulation are neglected to simplify the analysis. The transformer is assumed to be ideal.

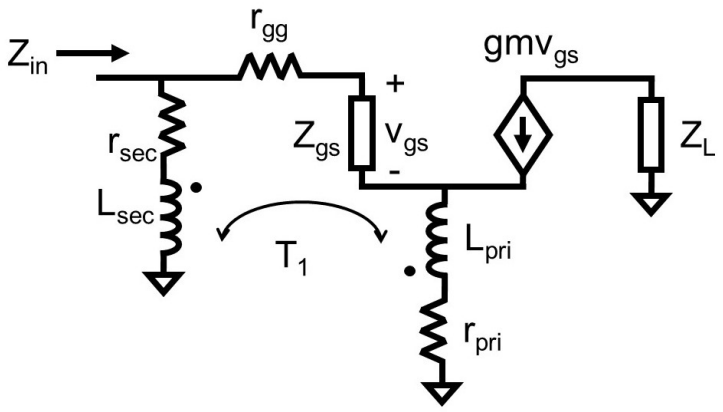


Figure 2.6: Small-signal equivalent circuit of gate-source coupled input stage via ideal transformer (Simplified)

Kirchhoff's current law is applied at the input, the resulted input impedance can be expressed as

$$Z_{in} \approx \left.\frac{v_{in}}{i_{in}}\right|_{Z_{gs}\to\infty} = \frac{n^2}{(n+1)g_m}, \tag{2.17}$$

which implies the approximated input impedance is solely dependent on the transformer turns ratio (n) and the transconductance of the transistor (g_m). Both variables are independent of frequency. Thus, a wideband input match can be achieved.

Although the topology can achieve wideband matching, there are several potential problems associated with it. The ohmic loss in linear transformer windings is in series with the transistor gate resistance, thus it increases the noise factor of the amplifier. The interwinding capacitance bridges the gate and source terminals of the transistor, which increases the input capacitance and degrades the transition frequency. It also leads to a higher minimum noise factor, as explained in Eq. 2.1.

As reported in [20], an input return loss better than 10 dB is achieved across a frequency range from 3.1 to 10.6 GHz, which indicates a wideband input match. Noise figure is around 2.5 dB within the bandwidth (which is higher than 0.9 dB reported in [19] across a narrower bandwidth of 250 MHz).

2.2.3 Performance Summary

Table 2.1 summarizes the performance data of the topologies covered in previous sections. Lower noise figure is observed from the differential cascode and the differential transformer feedback designs in [19] than others. Wideband circuit (i.e., [20]) shows higher noise figure due to the narrow-band nature of noise matching. Linearity generally degrades with lower supply voltage, as shown in the transformer coupled design in [19] and [20]. Single-ended amplifiers show its advantage on power consumption over differential designs (i.e., 9 mW in [20] vs. 16 and 21.6 mW in [19]).

Table 2.1: Typical performance of RF LNAs

Ref.	Technology	Topology	S_{21} (dB)	f_o (GHz)	NF (dB)	S_{11} (dB)	IIP_3 (dBm)	V_{DD}	P_{DC} (mW)
[19]	0.18 CMOS	Differential Cascode	14.1	5.75	1.8	N/A	4.2	1.8 V	21.6
[19]	0.18 CMOS	Differential Transformer	14.2	5.75	0.9	N/A	0.9	1.0 V	16
[20]	0.13 CMOS	Transformer Reactive Feedback	15.1 $\pm$ 1.4	3.1 $\sim$ 10.6	2.2 $\pm$ 0.43	<-10	-8.5 $\sim$ -5.1	1.2 V	9

2.3 Previous Work on Millimetre-Wave Low-Noise Amplifiers

Most mm-wave frequency LNA designs consist of cascaded, cascode amplifiers (e.g., [11]-[15]). Circuit parasitics significantly deteriorate gain, noise figure and bandwidth at these frequencies. Therefore, peaking inductors are commonly utilized to broaden the bandwidth (e.g., as reported in [13]-[15]). Transmission-line-based passive components become attractive for integrated circuits due to the decrease in their electrical length and their simplicity. Also, the area of a spiral coil decreases as the inductance decreases, which may lead to challenges in the layout of inductors (due to coupling from the opposite edge of the coil) and transformers (to realize high magnetic coupling factor). At mm-wave frequencies, coupled transmission lines may be used to implement transformers.

2.3.1 Multistage-Cascode LNA

To counter the problem of a decrease in the gain of single-stage amplifiers as the frequency increases, multiple stages in cascade are used to boost the gain. The multistage, cascode LNAs reported in [11] and [12] operate at K, Q, V and W bands. A block-level diagram of the topology is shown in Fig. 2.7 below.

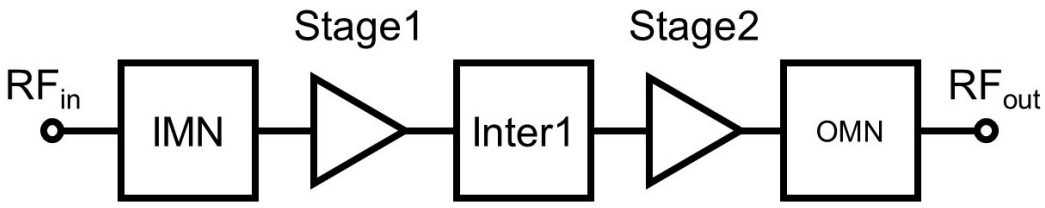


Figure 2.7: Block-level diagram of the multistage-cascode LNA

The input matching network (IMN) matches the first stage (Z_{in1}) to the source impedance, and synthesizes the source impedance for optimum noise factor for the first stage (Z_{opt1}). The output matching network (OMN) matches the impedance of the second stage output (Z_{out2}) to the amplifier load. In [11], the interstage is designed such that 50 Ω is presented to both the output of the first stage and the input of the second stage (i.e., the first and second stages are conjugate-matched to 50 Ω) to maximize the voltage gain of the first stage [22].

This leads to a general discussion of interfacing stages from microwave and analogue viewpoints. For microwave designs (especially at the print-circuit-board level, where transmission-line effects dominate), the conjugate match is implemented to achieve original integrity. From the analogue point of view, maximizing the load impedance for a transconductor maximizes the gain.

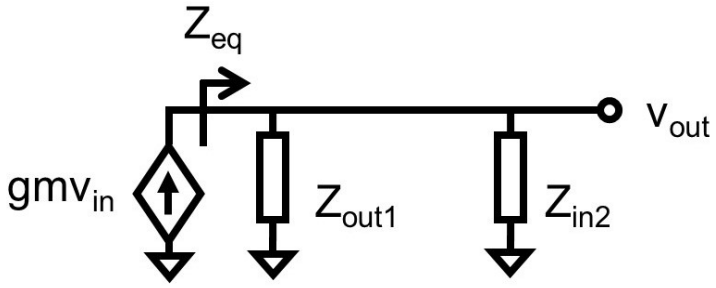


Figure 2.8: Simplified small-signal interstages model

Mathematically, the signal source (with output impedance $Z_{out1} = R_{out1} + jX_{out1}$) delivered to a load Z_{in2} equal to $(R_{in2} + jX_{in2})$ for a conjugate match ($i.e., R_{out1} = R_{in2}$ *and* $X_{out1} = -X_{in2}$) can be expressed as

$$\begin{aligned} v_{out,conj} = g_m v_{in}(Z_{out1}||Z_{in2}) &= \frac{1}{2} g_m v_{in}(R_{in2} + \frac{X_{in2}^2}{R_{in2}}) \\ &= \frac{1}{2} g_m v_{in} R_{in2}(1 + \tan^2\theta) = \frac{1}{2} g_m v_{in} \frac{Z_{in2}}{\cos\theta}, \end{aligned} \tag{2.18}$$

where θ is the angle of Z_{in2} in complex form. For the analogue approach to maximize gain, $v_{out,analogue} = g_m v_{in} Z_{in2}$ for Z_{out1} approaching infinity. Therefore, an inequality arises between $v_{out,conj}$ and $v_{out,analogue}$. When θ is less than 60°, the analogue approach provides a better output voltage swing. When θ is greater than 60°, performing conjugate match results in higher voltage gain. At mm-wave frequencies, realizing a high impedance for Z_{out1} may be challenging since the drain-to-source capacitance degrades the output impedance of the transistor. Furthermore, the channel conductance (g_{ds}) of the transistor increases as the process scales down, which worsens the problem. The above considerations make conjugate matching attractive at mm-wave frequencies to achieve greater voltage swing at interstages (i.e., high voltage gain), but the load (Z_{in2}) has to be taken into account

when designing the interstage. Another advantage of the 50-Ω interface is that at the system level, stages can be easily swapped by another block with 50-Ω interface without modifying the matching network. However, the conjugate interstage match may not always be intended. The increase in gain may lead to stability issue, as explained in A (i.e., the interstage mismatch is one way to improve stability). From power viewpoint, conjugate match generates real power at interstage (due to the presence of a real load), which leads to unnecessary heat dissipation (i.e., a high reactance load does not produce any real power).

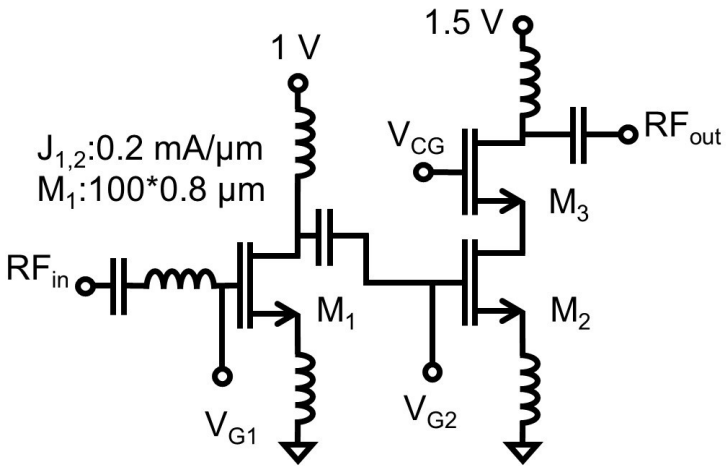


Figure 2.9: Two-stage K-band low-noise amplifier

The DC bias points of both stages are chosen to be 0.2 mA/μm, where F_{min} of the transistor is at its minimum [11] (i.e., both stages are biased for minimum noise). To further reduce noise, the input transistor is laid out as 4 discrete instances of 25 fingers, where the width of each finger is 0.8 μm [11]. The supply voltage of the second stage is set above nominal to improve the output headroom, which increases DC power consumption. The Q- and V-band designs in [12] use transmission-line-based inductors for interstage and output matching. The common-gate transistor gates are biased at supply voltage through resistors. The use of resistors eliminates the need of the voltage bias for the common-gate stage and adds the flexibility for tuning the output impedance (it can be taken as a part of the output matching network). However, the voltage at the source of the common-gate transistor is not guaranteed to be at the middle of the supply voltage, which leads to uneven distribution of the voltage drop across the common-source and the common-gate stages. As a result, the linearity may be affected.

The gain of 19.5 dB is achieved from 16 to 24 GHz by cascading a common-source

stage with a cascode stage with 45-nm CMOS-SOI technology, as reported in [11]. As the frequency increases, the gain is expected to decrease, which correlates with designs mentioned in [12] on the same process. The amplifier designed for Q-band reports a gain of 15 dB and drops down to 12.5 dB at V-band [12]. At around 80 GHz, a three-stage, cascaded common-source amplifier realizes a gain of 14.7 dB in [12]. Noise figure is expected to increase along with the increase in operating frequency ratio, as explained in Eq. 2.1. The noise figure for amplifiers at K, Q, V, W bands are 2 dB, 3.3 dB, 4 dB and 5.7 dB, respectively, which agrees with this expectation. Interestingly, power consumptions reported for these designs have a falling trend (from 32.5 mW to 13.5 mW). This can be explained using Eq. 2.8 and Eq.2.9. Assuming that transistors are biased at the same $(\frac{I_D}{W})_{opt}$ for the lowest transistor F_{min}, the multiplicity (M) required to perform noise matching in Eq. 2.8 and Eq. 2.9 decreases as operating frequency increases. As a result, drain current decreases as well as power consumption.

2.3.2 Inductive Peaking for the Cascode Topology

To determine the bandwidth of a regular cascode amplifier (as shown in Fig. 2.10), its effective time constant (τ) is given by

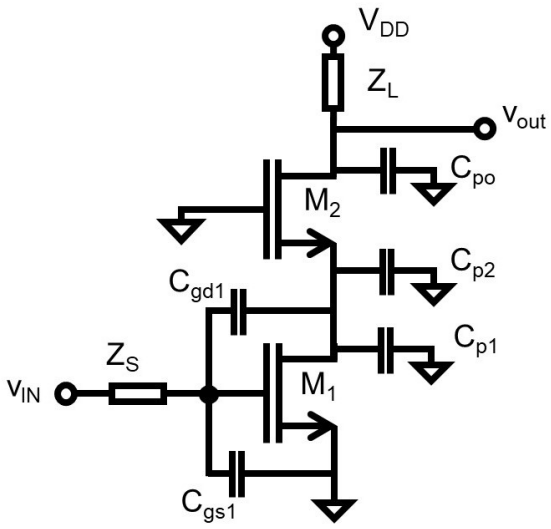


Figure 2.10: Cascode amplifier with critical capacitances related to bandwidth (Simplified)

$$\tau = R_s[C_{gs1} + C_{dg1}(1 + g_{m1}R_{d1})] + R_{d1}(C_{p1} + C_{p2}) + R_L C_{po}, \tag{2.19}$$

assuming Z_s equals to R_s, Z_L equals to R_L and neglecting channel length modulation. The impedance looking out from the drain of M_1 is assumed to be purely resistive (R_{d1}). The drain-to-source and drain-to-bulk capacitances of M_1 are together represented as C_{p1}, while C_{p2} contains the source-to-gate and source-to-bulk capacitances of M_2. The parasitics at the drain of M_2 is shown as C_{po}. The first term in Eq. 2.19 represents the pole at the input of the amplifier, while the later two terms represent the poles at the internode (between the CS and CG stages) and the output, respectively. As cascode amplifier suppresses the Miller effect by reducing the gain of the CS stage, the dominant pole has moved from the input to the output and the internode. Parasitic capacitances at these nodes become the bottleneck of the bandwidth of the amplifier. Thus, inductive peaking may be used to extend the bandwidth of the amplifier. Note that, the series and shunt peaking inductor in [23] are realized with coplanar waveguides (CPW), which occupies large chip area. To keep the active area small, peaking elements in this thesis are implemented with spiral inductors.

2.3.2.1 Series and Shunt Peaking

A simplified small-signal equivalent circuit is utilized to illustrate series peaking (as depicted in Fig. 2.11).

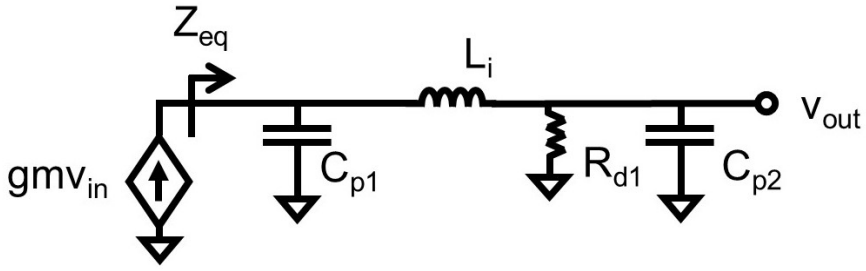


Figure 2.11: Small-signal model of common-source output with series peaking (Simplified)

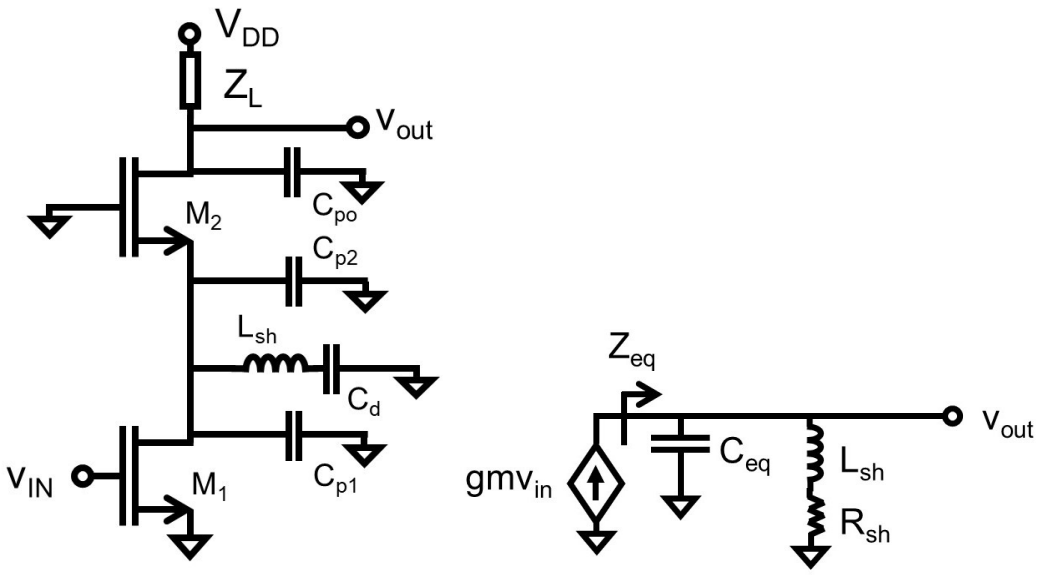
Neglecting channel length modulation, without L_i, the internode pole locates at $\frac{1}{R_{d1}(C_{p1}+C_{p2})}$ (rad/s). With L_i, the equivalent output impedance $Z_{eq}(s)$ can be expressed as [24]

$$Z_{eq}(s) = \frac{R_{d1}}{1 + sR_{d1}(C_{p1} + C_{p2}) + s^2 L_i C_{p1} + s^3 L_i R_{d1} C_{p1} C_{p2}} = \frac{R_{d1}}{1 + \frac{s}{\omega_u} + (\frac{s}{\omega_u})^2 mn + (\frac{s}{\omega_u})^3 mn(1-n)}, \tag{2.20}$$

where $m = \frac{L_i}{R_{d1}^2(C_{p1}+C_{p2})}$ is the level of compensation (unitless), $n = \frac{C_{p1}}{C_{p1}+C_{p2}}$ is dependent on the process (unitless), and $\omega_u = \frac{1}{R_{d1}(C_{p1}+C_{p2})}$ is the normalized 3-dB bandwidth for the uncompensated common-source stage with $L_i = 0$ [24]. By properly choosing m, the bandwidth can be extended beyond ω_u.

As described previously in Section 2.2.1, the noise contribution of M_2 is suppressed by increasing the output impedance of M_1 in a cascode amplifier. Placing a series inductor at the drain of M_1 improves the impedance, thus leads to lower noise performance. The disadvantage of adding a series inductor is that the ohmic losses of the inductor degrade the noise performance of the amplifier, which may overcome the benefit depending on the quality factor of the inductor.

Another approach to push the internode pole to higher frequency is to use a shunt inductor, as depicted in Fig. 2.12,



(a) Schematic (b) Small-signal equivalent

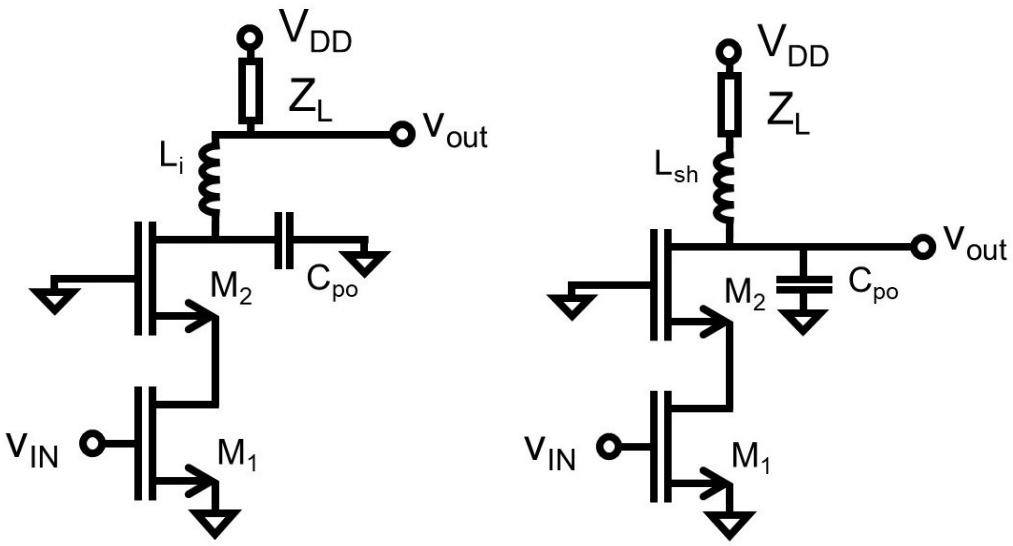
Figure 2.12: Simplified cascode amplifier with shunt peaking inductor

where C_{eq} is the shunt equivalent of C_{p1} and C_{p2} and R_{sh} represents the ohmic losses of the shunt peaking inductor. Neglecting channel length modulation, the equivalent output impedance $Z_{eq}(s)$ can be expressed as

$$Z_{eq}(s) = \frac{R_{sh} + sL_{sh}}{1 + sR_{sh}C_{eq} + s^2 L_{sh}C_{eq}}, \tag{2.21}$$

which transforms the one-pole response (without L_{sh}) to a transfer function with two poles and a zero. The peaked response is determined by the ratio between the zero (formed by L_{sh} and R_{sh}) and the pole (formed by R_{sh} and C_{eq}). Similar to series peaking, the ohmic losses of the inductor degrades the noise performance of the amplifier.

Similarly, the output pole may also be pushed to a higher frequency by performing series or shunt peaking, as depicted in Fig. 2.13. Eq. 2.20 and Eq. 2.21 may be used to express the bandwidth-enhancement ratio.



(a) Series peaking (b) Shunt peaking

Figure 2.13: Inductive peaking at the output of cascode (Simplified)

With 65-nm CMOS technology, a two-stage cascode with series peaking inductors at internode achieves a peak gain of 16 dB at 58 GHz with 4.5 dB noise figure across a bandwidth of 14 GHz [13]. With single-stage triple-cascode, the gain reaches 14.3 dB at 38 GHz with 3.8 dB noise figure across a bandwidth of 6 GHz in 130-nm CMOS [13]. Although the performance is lower than the ones in [12], it is worthy to mention that the work is done in a much older bulk process rather than the advanced 45-nm CMOS-SOI (used in [12]).

2.3.2.2 Inductive Peaking at the Gate of the Common-Gate Stage

Besides series and shunt peakings, an inductor may be placed at the gate node of the common-gate stage to enhance the bandwidth, as depicted in Fig. 2.14.

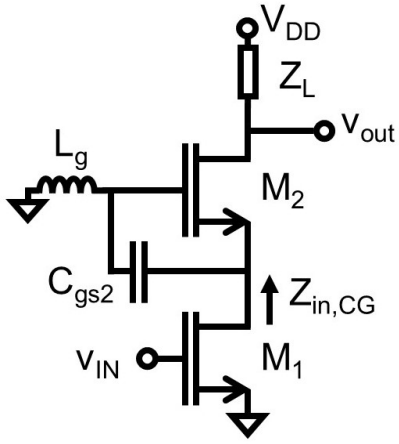


Figure 2.14: Inductive peaking at the gate of the common-gate stage

Neglecting channel length modulation and including the effect of C_{gs2}, the voltage gain of a cascode amplifier may be expressed as

$$A_{v,woL_g} \approx -g_{m1}Z_L \frac{1}{s\frac{C_{gs2}}{g_{m2}} + 1}, \tag{2.22}$$

where L_g is excluded. Taking the effect of L_g and channel length modulation of M_1 into account, the gain expression becomes

$$A_{v,wL_g} \approx -g_{m1}Z_L \frac{1}{s^2\frac{L_gC_{gs2}}{g_{m2}ro_1} + s\frac{C_{gs2}}{g_{m2}} + 1}. \tag{2.23}$$

The frequency response has changed from a one-pole response (with a pole at $s = -\frac{g_{m2}}{C_{gs2}}$) to a two-pole response (with the new pole frequency adjusted by L_g). Thus, bandwidth peaking may be realized.

A significant impact of adding the gate inductor is stability as it introduces positive feedback. Qualitatively, the impedance of the gate inductor ramps up as frequency increases, which makes the gate of the CG stage floating. The signal fed back through the drain-to-gate capacitance of M_2 is forced to flow through the source and eventually back to the input of the amplifier. Therefore, the isolation is deteriorated (i.e., the amplifier is less stable, as explained in Appendix A). Quantitatively, with channel length modulation neglected, the real part of the input impedance of the CG stage ($R_{in,CG}$) is approximately

$$R_{in,CG}(s) \approx \frac{1 + s^2 L_g C_{gs2}}{g_{m2} + sC_{gs2}}. \tag{2.24}$$

When $sC_{gs2} \ll g_{m2}$ in Eq. 2.24,

$$R_{in,CG}(s) \approx \frac{1}{g_{m2}}(1 + s^2 L_g C_{gs2}), \tag{2.25}$$

which will become negative if the frequency is beyond the resonant frequency of L_g and C_{gs2} (i.e., $\frac{1}{\sqrt{L_g C_{gs2}}}\ rad/s$). As a result, the amplifier becomes unstable due to this positive feedback. Thus, the value of L_g has to be chosen properly to ensure the unstable region is far beyond the operating frequency.

The three-stage cascode with inductive peaking (at the gate of CG) utilizing 65-nm CMOS in [14] reports a gain of 20.6 dB at 60 GHz across a $-3dB$ bandwidth of 14.1 GHz. The in-band minimum noise figure is 4.9 dB at 58 GHz. With a supply voltage of 1.2 V, the output $-1dB$ compression point is at -8.4 dBm. The second design in [14] utilizes spiral inductor instead of transmission-line based passives. It achieves a gain of 18 dB across a bandwidth of 12.2 dB. The in-band minimum noise figure improves to 4 dB at 61 GHz. The reduction in bandwidth and improvement in gain could be due to the spiral inductor has a higher quality factor (i.e., with a high quality factor, ohmic losses are improved but bandwidth shrinks). With the same supply voltage, the output $-1dB$ compression point is at -5 dBm. The design in [15] uses inductive peaking at both internode (series peaking) and the gate of CG. It realizes a gain of 20 dB across a bandwidth from 24 to 48 GHz. The in-band minimum noise figure is 3.1 dB. With a supply voltage of 1.2 V, the output $-1dB$ compression point is at -3 dBm.

2.3.3 Performance Summary

Table 2.2 summarizes the specifications of the topologies covered in previous sections. At mm-wave frequencies, cascading multiple stages is a common way to realize high gain, as seen from all the entries. With the same process, lower noise figure is observed from lower operating frequency, which matches with the trend predicted by Eq. 2.1 (2, 3.3, 4.0, and 5.7 dB noise figure at K, Q, V, W bands from [11] and [12], respectively). Compared between [12], [13] and [14], more advanced process technology seems to generally have a lower noise figure, as the transition frequency of transistors becoming higher. Lower supply voltage generally leads to lower IIP_3 (as seen from [12]). Thus, to mitigate this problem,

the DC bias of the output stage can be set to above nominal to provide more voltage headroom to improve linearity (as seen from [11]). High input return loss is difficult to be maintained across the -3 dB bandwidth, as seen from the table. Thus, a wideband input matching network may be considered, as S_{11} is a critical specification of an LNA. Lastly, if the transistor is biased around $(\frac{I_{DS}}{W}_{opt})$ (for lowest device noise), the required bias current is predicted to decrease as operating frequency increases, which leads to low power consumption. This trend is confirmed in [12].

Table 2.2: Typical performance of mm-wave LNAs

Ref.	Technology	Topology	S_{21} (dB)	f_o (GHz)	NF (dB)	S_{11} (dB)	IIP_3 (dBm)	V_{DC}	P_{DC} (mW)
[11]	45-nm CMOS SOI	2-stage cascode	19.5	16 ~ 24	2.0	<-5[1]	-8	1.5 V	32.5
[12] Work1	45-nm CMOS SOI	2-stage cascode	15	40 ~ 53	3.3	<-8[1]	-13.5	1.3 V	20.8
[12] Work2	45-nm CMOS SOI	2-stage cascode	12.5	60 ~ 73	4.0	<-13[1]	-14.5	1.3 V	15
[12] Work3	45-nm CMOS SOI	3-stage common-source	14.7	76 ~ 88	5.7	<-6[1]	-16.2	1.0 V	13.5
[13] Work1	65-nm CMOS	2-stage Series Peaking	16	48 ~ 62[1]	4.5	<-7[1]	N/A	N/A	10
[13] Work2	130-nm CMOS	Triple cascode Series Peaking	14.3	36 ~ 42[1]	3.8	<-7[1]	N/A	N/A	28.8
[14] Work1	65-nm CMOS	3-stage CG GatePeaking	20.6	54 ~ 68.1[1]	4.9	<-5[1]	-19.4	1.2 V	33.6
[14] Work2	65-nm CMOS	3-stage CG GatePeaking	18	56 ~ 68.2[1]	4	<-6[1]	-13.4	1.2 V	28.8
[15] Work1	90-nm CMOS	3-stage Series & CG Gate Peaking	20	24 ~ 48	3.1	<-5[1]	-13.4	1.2 V	21.1

[1]Estimated from plots
[2]Estimated from P_{-1dB}

Chapter 3

Design

In the previous chapter, it is shown that cascading multiple cascode stages typically achieves a noise figure around 3 dB. Applying noise-matching technique at the interstage is one option to realize sub-2 dB noise figure. In this chapter, greater than 0.13 dB improvement in noise figure is shown after analyzing the cascade noise factor of noise-matching versus conjugate-matching interstages. Design examples of a two-stage amplifier with transformer-coupled, noise-matching interstage are discussed in detail. After considering device selection for the lowest noise factor, the choice of interstage transformer parameters is presented. Lastly, a design methodology for the low-noise amplifier is summarized.

3.1 Design A: a Two-Stage Amplifier with Noise-Matching Interstage

The amplifier is single-ended to lower power consumption and to save chip area, at the expense of lower isolation.

3.1.1 Proposed Circuit Topology

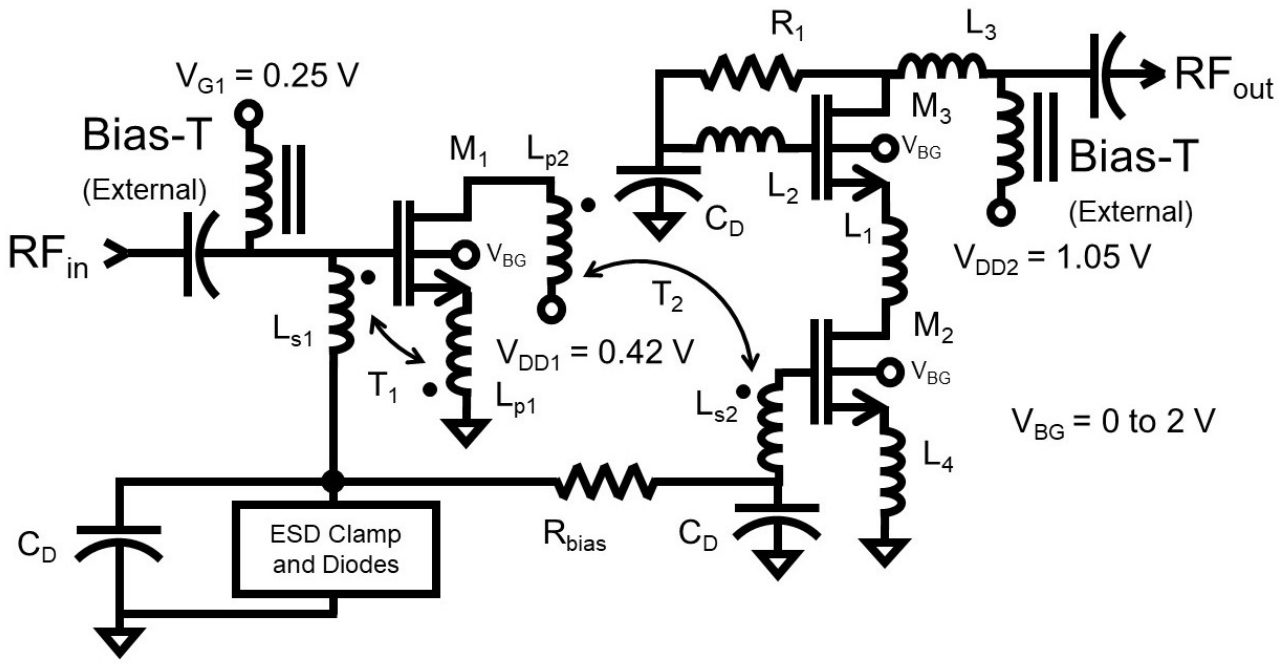


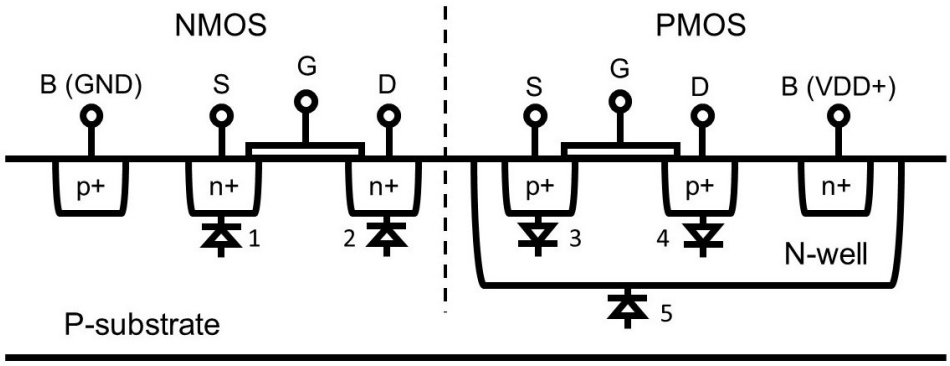
Figure 3.1: Topology of a two-stage amplifier with interstage noise matching

The overall schematic is shown in Fig. 3.1 above. The first stage consists of a common-source amplifier that is biased at $\frac{I_{DS}}{W}_{opt}$. Transistor M_1 is sized such that its R_{opt} is close to 50 Ω. Impedance transformation via T_1 is used to realize the input conjugate and noise match simultaneously. The supply voltage of the first stage (V_{DD1}) is set at 0.42 V to reduce DC power consumption (at the cost of some noise and linearity degradation). The second stage consists of a cascode amplifier (M_2/M_3) with M_2 resistively biased at the same gate voltage as M_1 via R_{bias}. Thus, M_2 also operates at $\frac{I_{DS}}{W}_{opt}$ for lowest device noise. The interstage transformer T_2 transforms the output impedance of the first stage to the optimum source impedance for minimum noise ($Z_{n,opt2}$) from the second stage. As a result, both stages are biased and matched for lowest noise. Inductor L_1 is used to peak the response and thereby suppress the noise contributed by M_3. Inductor L_2 is used to further extend the bandwidth, but the value is selected carefully so that the amplifier remains stable. Resistor R_1 is used to bias M_3, as well as to pad the output impedance. The total output reactance of the second stage is nulled by L_3 to give a wideband output return loss.

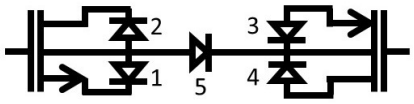
Inductor L_4 provides degeneration, which improves stability. The supply voltage of the second stage (V_{DD2}) is chosen above the nominal voltage to provide more output headroom for better linearity, at the cost of some transistor reliability degradation (during turn-on, the drain-to-source voltage across M_3 may be higher than nominal V_{DS}). The design is ESD protected by a network consisting of an RC-clamp and a double-diode connected from the bottom of L_{s1} to ground.

3.1.2 Body-Bias and the Flip-Well Structure

Body-bias trimming of an FD-SOI FET may be used to optimize the threshold-voltage (V_{TH}) of transistors, which may be advantageous as the gate dielectric thickness becomes challenging to be scaled down [25]. Forward body-bias (i.e., applying a positive relative voltage on the body of NFETs or negative relative voltage on the body of PFETs) can lower transistor V_{TH} to allow higher drive and faster switching capabilities. Reverse body-bias (i.e., applying a negative relative voltage on the body of NFETs (or positive relative voltage on the body of PFETs) raises V_{TH} to allow lower leakage. However, in a traditional bulk CMOS process, the range of forward body-bias is limited by the threshold voltage of parasitic diodes between wells, as depicted in Fig. 3.2.



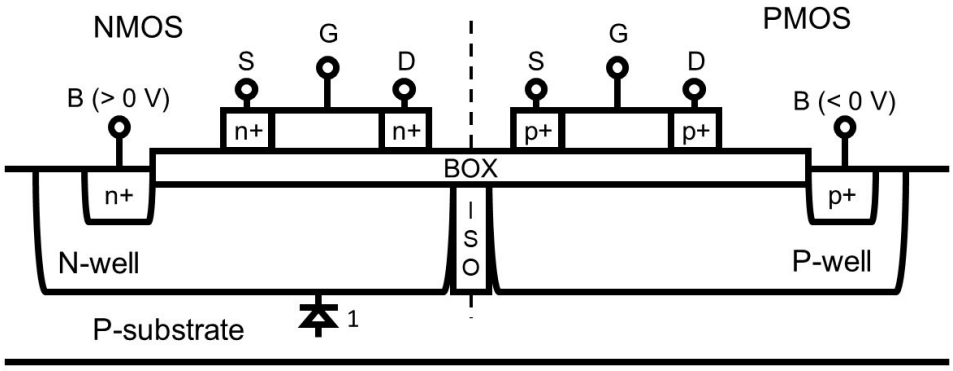
(a) Crosssection view



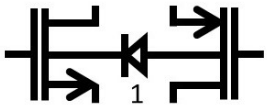
(b) Schematic view

Figure 3.2: Forward body-bias limitations in a bulk CMOS process

The 22-nm CMOS-SOI process has a buried oxide layer between the source/drain and substrate which enables a flip-well device, where NFETs are fabricated in N-wells instead of P-wells (and PFETs are fabricated in P-wells instead of N-wells). With the flip-well scheme, most of the parasitic diodes shown in Fig. 3.2 no longer exist, which leads to a large voltage range of forward body-bias. The body-bias limit becomes the threshold voltage of the parasitic diode between the P-well and N-well, which will only limit reverse body-bias operations, as shown in Fig. 3.3.



(a) Crosssection view



(b) Schematic view

Figure 3.3: Body-bias limitations in the 22-nm CMOS-SOI process

3.1.3 Characterization of Active Devices in Simulation

Transconductance (g_m), transition frequency (f_T), minimum noise figure (NF_{min}), transistor unit width and optimum source resistance for minimum noise (R_{opt}) are transistor metrics simulated for designing for lowest noise. By simulating f_T and NF_{min} versus current per unit transistor width ($\frac{I_{DS}}{W}$), the DC bias point may be selected around where lowest NF_{min} occurs. Then, the optimum unit transistor width is found by simulating NF_{min} versus different finger width. Lastly, the total width of the transistor is selected based on the simulation of R_{opt} to realize noise matching.

The 22-nm CMOS-SOI process design kit (PDK) offers two types of transistor (SLVT and LVT) which use the flip-well structure. The PDK also provides two back-end-of-line (BEOL) options. The raw transistor model represents the option of transistor metalization up to the first metal layer (M_1), and the other option goes up to the fifth metal layer (M_5). The difference between them is the additional parasitic capacitances between transistor terminals caused by metal interconnections. Common-source NFETs[1] with the same physical dimensions (transistor total width of $10 \times W = 1$ μm) are used to conduct a fair comparison between SLVT and LVT transistors. Note that the bias at the back-gate terminal also affects the drain-to-source current of the transistor through changes in the threshold voltage. For simulations versus the drain current per unit transistor width ($\frac{I_{DS}}{W}$), the voltage at the drain, gate, and back-gate terminals are swept.

3.1.3.1 Threshold Voltage

To benchmark the effect of the back-gate bias on the threshold voltage, V_{TH} of LVT and SLVT transistors are extracted using the second-derivative method introduced in [26]. The second-derivative method assumes the transistor turns on at the threshold voltage and then the drain current grows linearly as the gate-to-source voltage increases. Thus, the first derivative of the drain current may be regarded as a step function. Furthermore, the second derivative of the drain current becomes a Dirac delta function. Therefore, the threshold voltage may be found at the peak of the second derivative occurs.

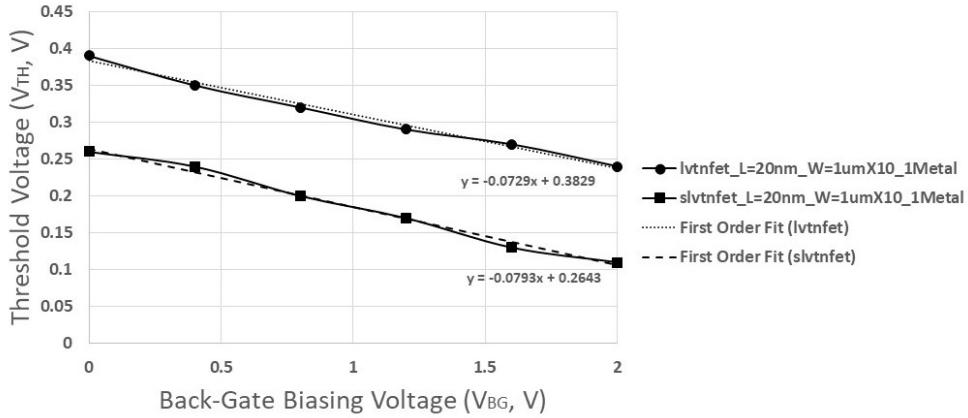


Figure 3.4: Threshold voltages for 20-nm LVT and SLVT NFETs

[1]The transistor notation used in legends is in the format of [type]_[gate length]_[gate finger width×number of gate fingers]_[misc. conditions]

(Cadence® Spectre® simulator uses the constant-current method [27] to generate DC operating point. The threshold voltage is defined as the gate-to-source voltage at an arbitrary constant drain current, typically very tiny. The results from the second-derivative method are matched with the DC operating point). Note that the transistor is biased in the deep triode region (i.e., $V_{DS} = 10$ mV), as the threshold voltage defines the boundary between transistor cut-off and triode regions.

As explained in the previous section, forward body-bias improves transistor performance by reducing the threshold voltage. As shown in Fig. 3.4, with no biasing at the back-gate, LVT and SLVT report threshold voltages around 390 mV and 260 mV, respectively. When biasing at 2 V, the threshold voltages drop to around 240 mV and 110 mV, with an approximately linear rate of -73 mV/V and -79 mV/V, respectively.

3.1.3.2 Transconductance

Next, transconductance (g_m) of LVT and SLVT NFETs are simulated versus the drain current for devices normalized per unit unit width (i.e., vs. $\frac{I_{DS}}{W}$ in mA/μm). Due to the large size of data set, the simulation is broken into two parts to illustrate the effect of back-gate (V_{BG}) and drain-to-source (V_{DS}) voltages on g_m, respectively.

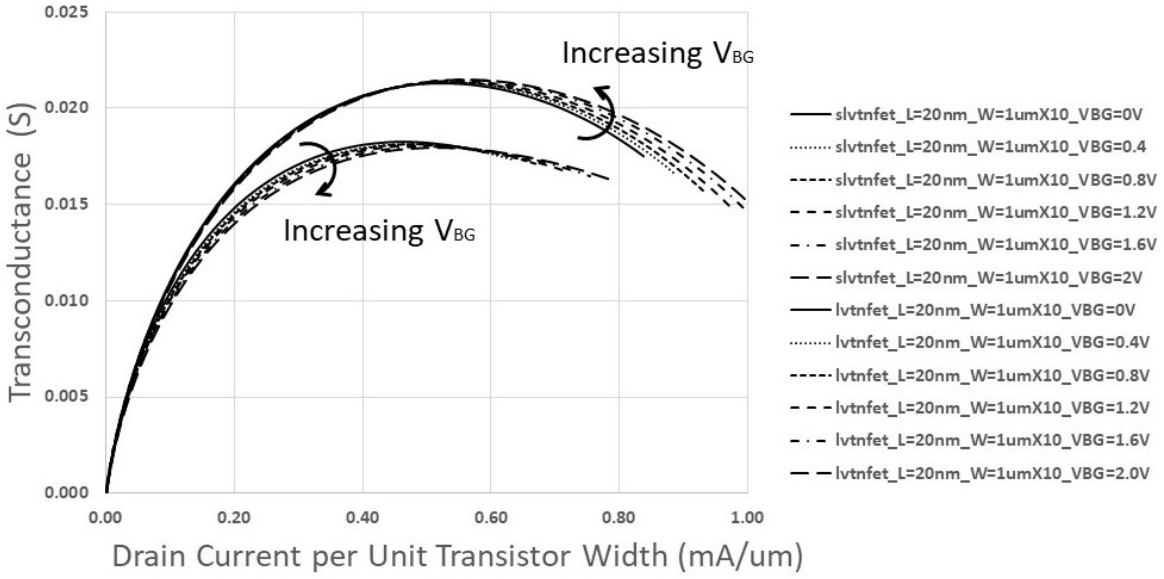


Figure 3.5: Transconductance for 20-nm LVT and SLVT NFETs at nominal V_{DS}

Firstly, the gate-to-source voltage (V_{GS}) and V_{BG} are swept (with V_{DS} is set to 0.8 V) to illustrate the effect of V_{BG} on g_m. As depicted in Fig. 3.5, SLVT NFETs provide higher g_m than LVT NFETs across various $\frac{I_{DS}}{W}$ conditions due to lower V_{TH}. Interestingly, biasing at high V_{BG} leads to slightly lower g_m efficiency at low $\frac{I_{DS}}{W}$. This trend is found in both FET variants, but it is more prominent in LVT NFETs. Higher V_{BG} provides better g_m at high $\frac{I_{DS}}{W}$, however, this is not beneficial because the device has already passed the bias point when the g_m is peak, which just results in excessive power consumption.

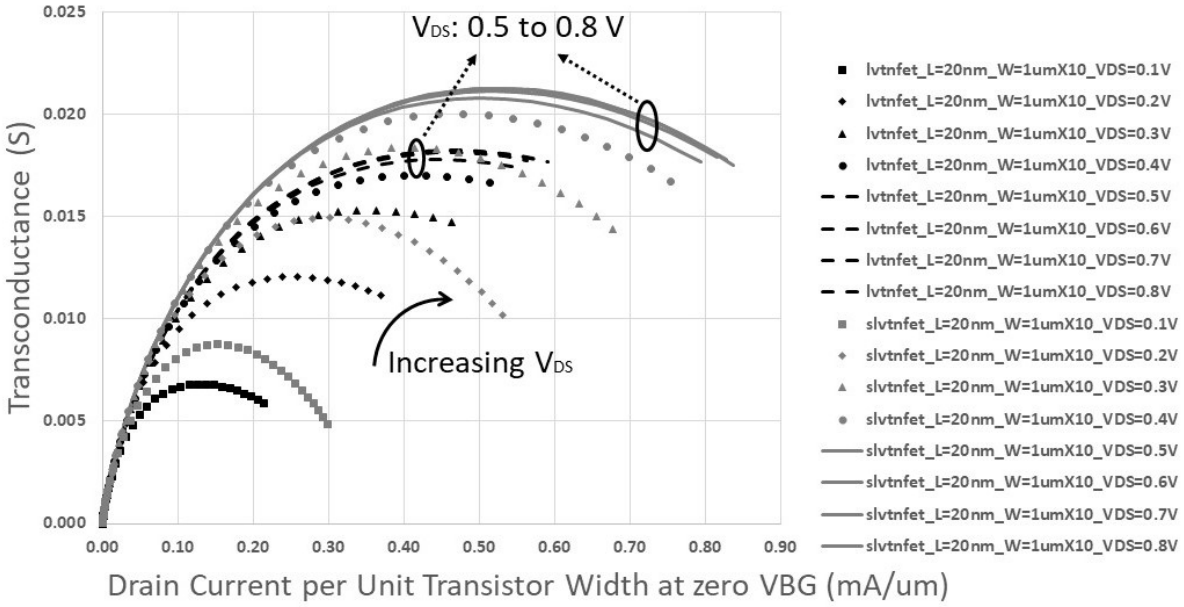


Figure 3.6: Transconductance for 20-nm LVT and SLVT NFETs at zero V_{BG}

As depicted in Fig. 3.6, when V_{DS} is above 0.5 V, increasing V_{DS} further does not cause much change in g_m for both LVT and SLVT transistors, which indicates that when devices operate deep in the saturation region, the g_m (and thus drain current) is almost independent of V_{DS} (i.e., low channel conductance). This is beneficial, as a certain g_m may be realized with a supply voltage lower than nominal, which leads to lower DC power consumption.

3.1.3.3 Transition Frequency

To realize a low noise factor, the transistor has to be biased for sufficient gain-bandwidth and operates at a low RF to transition frequency ratio, ω_o/ω_T, as described by Eq. 2.1. Since f_T is directly related to g_m, similar trends seen in Fig. 3.5 and Fig. 3.6 appear.

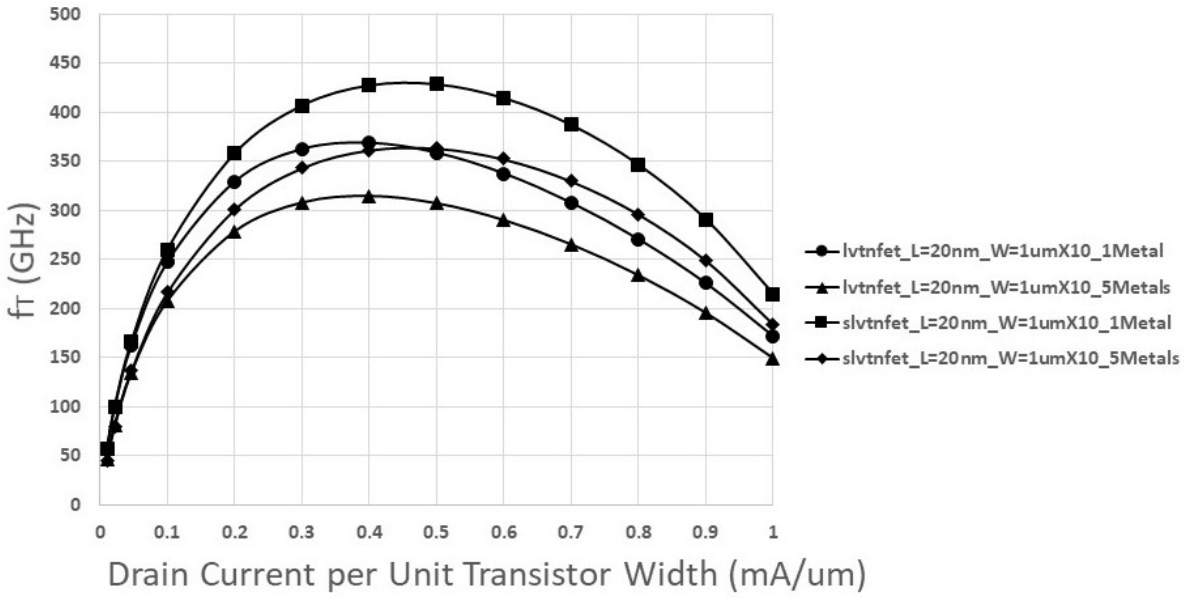


Figure 3.7: Transition frequency for 20-nm LVT and SLVT NFETs

Fig. 3.7 shows the simulated result for f_T versus normalized drain current for constant V_{DS} (0.8 V), $\frac{I_{DS}}{W}$, for both LVT and SLVT NFETs. Peak f_T occurs around $\frac{I_{DS}}{W}$ of 0.4 mA/μm for both FET variants. The f_T decreases beyond 0.4 mA/μm which may be due to mobility (i.e., transconuductance) degradation at high current densities [28]. Transition frequencies of 360 GHz and 310 GHz are observed around 0.4 mA/μm for SLVT and LVT NFETs, respectively. The higher f_T observed for SLVTs may be due to SLVT NFETs having better transconductance than LVT NFETs. With additional metalization (i.e., 5 metals versus 1 metal in Fig. 3.7), f_T of both FETs decrease. Therefore, the SLVT NFET is a good candidate for designing for lowest noise performance.

3.1.3.4 Minimum Noise Figure

As explained in Section 2.1.1, an optimum current unit per transistor width ($\frac{I_{DS}}{W}_{opt}$) exists where the noise factor of the device is at its minimum. The minimum noise figure (NF_{min}) is therefore simulated versus $\frac{I_{DS}}{W}$ to locate the optimum biasing condition for the transistor.

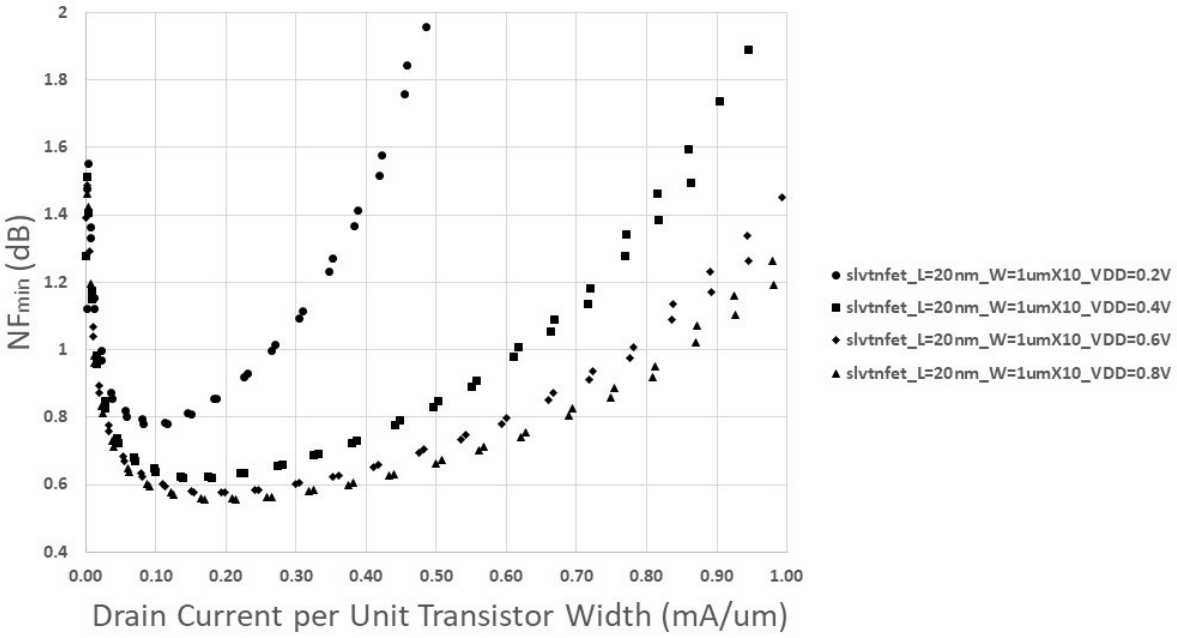


Figure 3.8: Minimum noise figure for 20-nm SLVT NFETs

Fig. 3.8 shows the simulated result for NF_{min} versus $\frac{I_{DS}}{W}$ for SLVT NFETs. The lowest NF_{min} of approximate 0.56 dB occurs around 0.2 mA/μm (i.e., $\frac{I_{DS}}{W}_{opt} \approx 0.2$ mA/μm), but NF_{min} less than 0.6 dB can be seen between 0.1 and 0.35 mA/μm. Next, V_{BG} has negligible effect on NF_{min} in the low $\frac{I_{DS}}{W}$ region, which indicates that NF_{min} is independent of biasing at the back-gate. However, varying V_{DS} does have an effect on NF_{min}. At $\frac{I_{DS}}{W}_{opt}$, NF_{min} degrades roughly 1.6% when V_{DS} changes from 0.4 V to 0.8 V (e.g. from 0.56 dB at 0.4 V and 0.63 dB at 0.8 V, respectively), but the change is most prominent when V_{DS} is below 0.4 V. Therefore, power consumption may be reduced by lowering the supply voltage, with tolerable degradation in noise. However, this will significantly decrease the linearity due to the reduced voltage headroom.

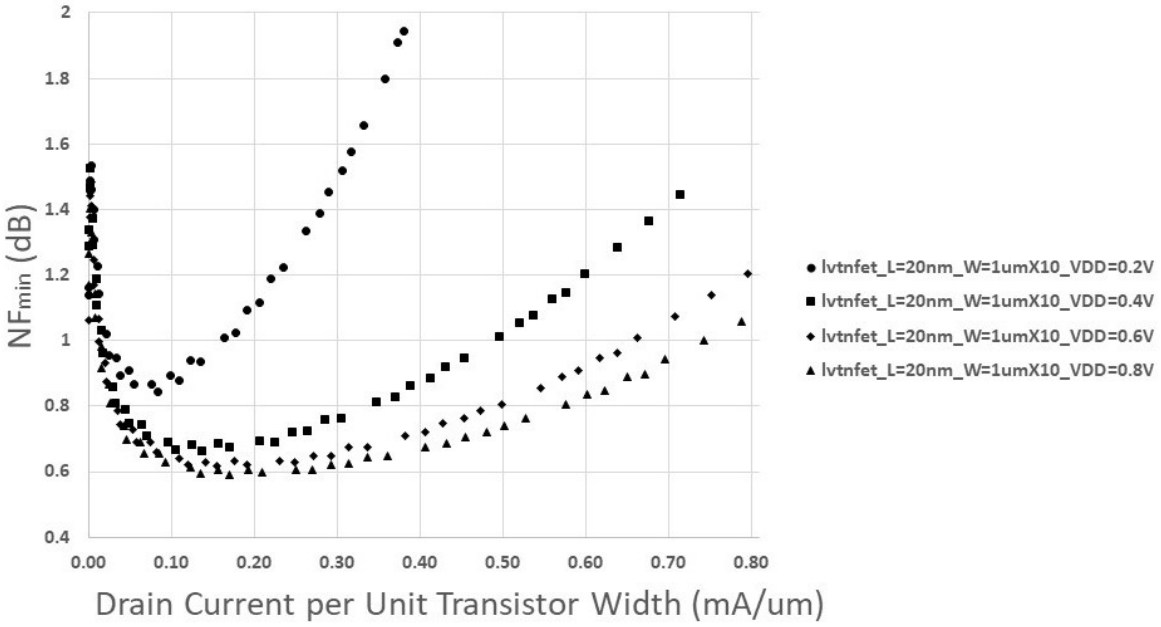


Figure 3.9: Minimum noise figure for 20-nm LVT NFETs

A similar trend can also be seen for LVT NFETs, as depicted in Fig. 3.9. Near $\frac{I_{DS}}{W}$ of 0.2 mA/μm the lowest NF_{min} is around 0.6 dB. As V_{DS} changes from 0.8 V to 0.4 V, the degradation in NF_{min} is around 2.6%, which is higher seen for than the SLVT.

Furthermore, NF_{min} is plotted for both transistors operating at low power consumption per transistor width (i.e., below 0.05 mA/μW) in Fig. 3.10.

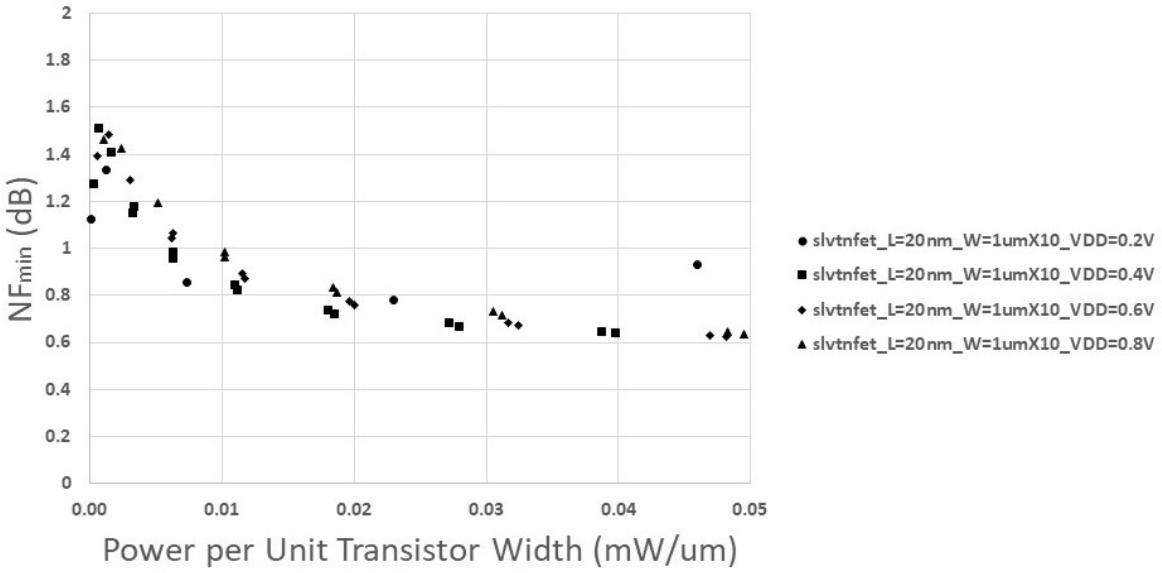


Figure 3.10: Minimum noise figure in low-power region for 20-nm SLVT NFETs

Interestingly, as the supply voltage decreases from nominal, a trend of decreasing NF_{min} can be observed, which is opposite to biasing at $\frac{I_{DS}}{W}_{opt}$ for lowest noise. This feature may be beneficial for the system in which power consumption has a higher priority than noise. A similar trend may also be found in LVT NFETs.

3.1.3.5 Transistor Unit Width

For a fixed transistor width, finger width and the number of fingers affect gate resistance and parasitics, which leads to changes in the minimum noise figure. The effect of finger width on minimum noise figure is analyzed in this section. A unit SLVT NFET with a total width of 16 μm is used for all simulations.

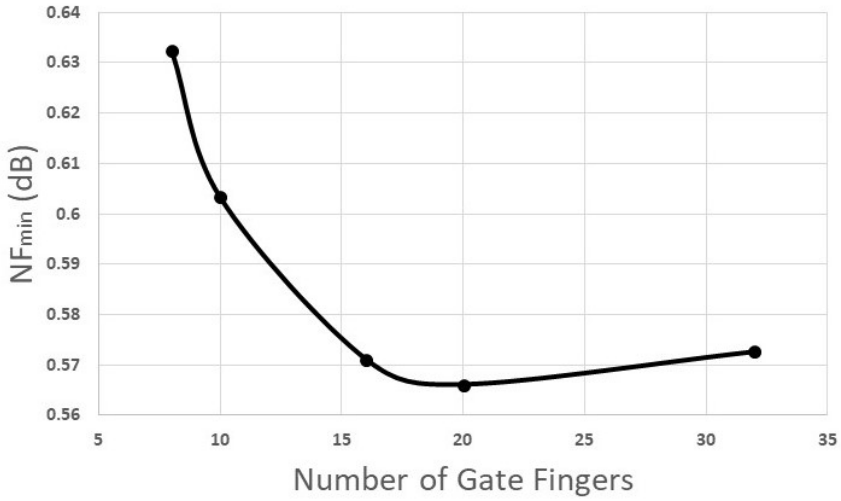


Figure 3.11: Minimum noise figure versus number of fingers for a 16-μm total-gate-width SLVT NFET

As depicted in Fig. 3.11, finger width between 640 nm and 1 μm (i.e., total gate width of 16 μm = 640 nm×25=1 μm×16) outputs low NF_{min}. For finger width greater than 1 μm, NF_{min} increases rapidly.

3.1.3.6 Optimum Source Resistance

Lastly, the transistor has to be sized such that its optimum source resistance (R_{opt}) for minimum noise factor equals 50 Ω for input noise matching. The optimum noise resistance versus device width is simulated for both NFET variants (with 1 μm finger width) and is plotted in Fig. 3.12.

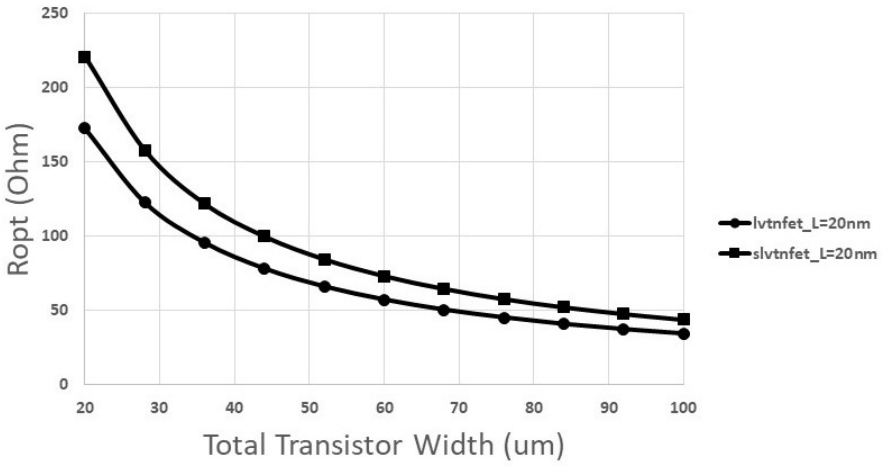


Figure 3.12: Optimum noise resistance for 20-nm LVT and SLVT NFETs

The LVT transistor has a lower R_{opt} than the SLVT device. This can be explained by Eq. 2.8 (g_m of the LVT transistor is lower than SLVT, thus lower R_{opt}). From the figure above, an 80-μm wide SLVT transistor has an R_{opt} around 50 Ω and 70-μm for the LVT case. Therefore, to realize noise matching, SLVT consumes more DC power than LVT variants when the transistor is biased around $\frac{I_{DS}}{W}_{opt}$.

3.1.3.7 Summary of Device Selection

The discussions in previous sections lead to a trade-off between noise and power consumption. As the $\frac{I_{DS}}{W}_{opt}$ of SLVT and LVT NFETs are both approximately 0.2 mA/μm, the required width of SLVT transistors is greater than for LVT transistors to realize 50-Ω noise matching at the input. This leads to higher drain current and power consumption. However, SLVT devices have superior transconductance over LVT transistors, which results in higher f_T and lower NF_{min}. This thesis targets the lowest noise performance, thus the SLVT NFET is chosen for the amplifier. Furthermore, the supply voltage is set to 0.42 V to reduce power consumption at the cost of some additional noise. The unit finger width is chosen as 800 nm, and the total device width is set to 80 μm (post-layout). The gate of the transistor is biased at 0.43 V with $V_{BG} = 0$ V, or 0.25 V with $V_{BG} = 2$ V, as V_{BG} has a negligible effect on noise performance.

3.1.4 Noise Analysis of Cascaded Amplifiers with Transformer Coupled Interstage

Friis' equation (i.e., Eq. 1.4) predicts the noise factor of a cascaded chain of components in a system. It assumes that the interface between the two stages is 50 Ω. However, this constraint does not apply to monolithic designs. Alternatively, the noise factor may be calculated in terms of voltage gain. For a two-stage amplifier (as seen in Fig. 3.13), it is assumed that the unloaded voltage gain, input impedance, output impedance and circuit added noise power are given by $A_{v[x]}$, $Z_{in[x]}$, $Z_{out[x]}$ and $\overline{v_{n[x]}^2}$, respectively.

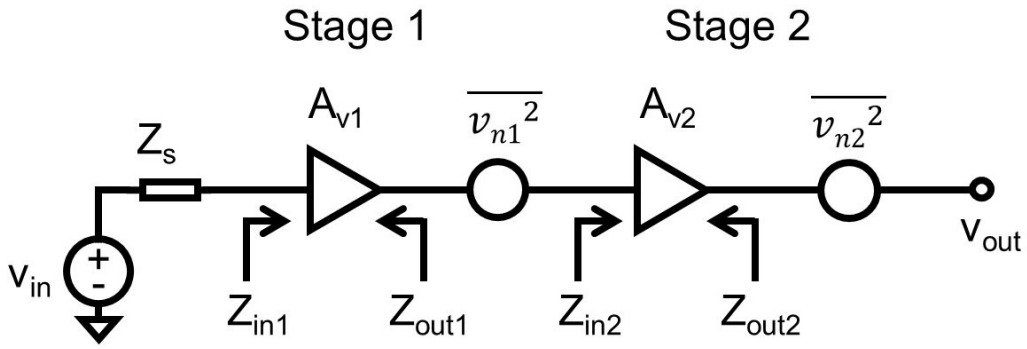


Figure 3.13: Simplified two-stage amplifier for noise factor calculation

Assuming that Z_s is 50 Ω (i.e., $Z_s = R_s = 50\Omega$), the noise factor of the two-stage chain is given by

$$F_{cascade} = 1 + \frac{\overline{v_{n1}^2}}{(\frac{Z_{in1}}{Z_{in1}+R_s})^2 A_{v1}^2} \frac{1}{4kTR_s} + \frac{\overline{v_{n2}^2}}{(\frac{Z_{in1}}{Z_{in1}+R_s})^2 A_{v1}^2 (\frac{Z_{in2}}{Z_{in2}+Z_{out1}})^2 A_{v2}^2} \frac{1}{4kTR_s} \quad [4]. \tag{3.1}$$

In Eq. 3.1, the first two terms represent the noise factor of the first stage (F_1) with respect to a source impedance, R_s. The noise factor of the second stage with respect to a source impedance given by Z_{out1} is

$$F_2 = 1 + \frac{\overline{v_{n2}^2}}{(\frac{Z_{in2}}{Z_{in2}+Z_{out1}})^2 A_{v2}^2} \frac{1}{4kTZ_{out1}}. \tag{3.2}$$

The noise factor of the chain may therefore be rewritten as

$$F_{cascade} = F_1 + \frac{\overline{v_{n2}^2}}{(\frac{Z_{in1}}{Z_{in1}+R_s})^2 A_{v1}^2 (\frac{Z_{in2}}{Z_{in2}+Z_{out1}})^2 A_{v2}^2} \frac{1}{4kTR_s}. \tag{3.3}$$

Substituting Eq. 3.2 into Eq. 3.3 yields,

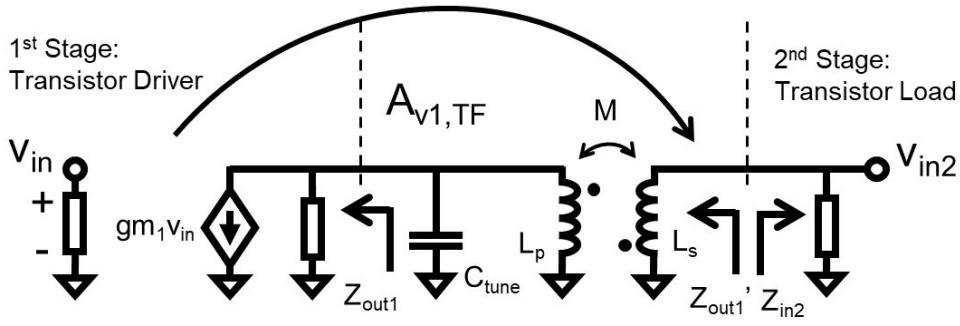
$$F_{cascade} = F_1 + \frac{F_2 - 1}{(\frac{Z_{in1}}{Z_{in1}+R_s})^2 A_{v1}^2 \frac{R_s}{Z_{out1}}}. \tag{3.4}$$

Assuming that the input is noise and conjugate matched to 50 Ω (i.e., $R_s = Z_{in1} = Z_{n,opt1} =$ 50 Ω, $F_1 = F_{min1}$), the cascaded noise factor only depends on the voltage gain of the first stage and interstage (A_{v1}), the output impedance of the first stage (Z_{out1}) and the noise of the second stage (F_2).

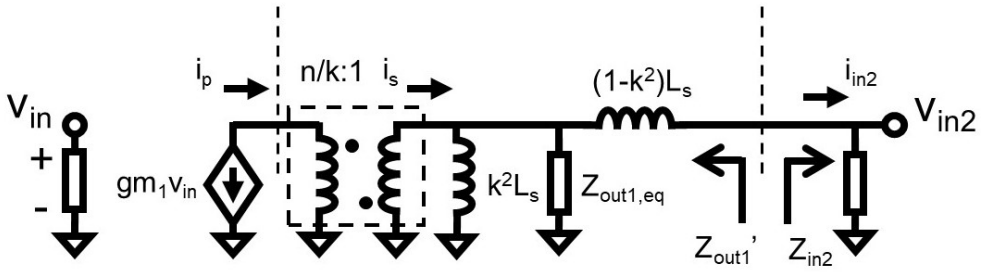
3.1.4.1 Voltage Gain of the First Stage with Transformer-Coupled Interstage

A passive L-section matching network has a bandwidth that is determined by the quality factor (Q-factor) of the components. For an ideal L-C matching network, the loss may be small, but at the same time, the frequency response becomes narrowband. For wideband applications, the Q-factor has to be reduced which increases losses.

A transformer may be used to realize interstage coupling and impedance transformation (as depicted in Fig. 3.14). The bandwidth of the transformer is determined by the magnetic coupling factor (k) and the insertion loss is determined by k and losses in the windings [29]. Furthermore, the primary winding may be connected to the supply and the secondary winding may be connected to the bias voltage for the second stage, which realizes DC biasing and DC isolation between stages (note that biasing is not shown in Fig. 3.14).



(a) with linear transformer



(b) with leakage shifted to the secondary

Figure 3.14: Transformer-coupled interstage

With the interstage transformer (C_{tune} of Fig. 3.14a excluded), the voltage gain of the first stage (A_{v1}^2 in Eq. 3.3 and Eq. 3.4) is given by

$$A_{v1,TF}^2 = \frac{v_{in2,unloaded}^2}{v_{in}^2} = \frac{(i_{in2}Z'_{out1})^2}{v_{in}^2} = (\frac{n}{k}g_{m1}Z'_{out1}\frac{sk^2L_s||[Z_{out1}/(\frac{n}{k})^2]}{s(1-k^2)L_s + sk^2L_s||[Z_{out1}/(\frac{n}{k})^2]})^2. \tag{3.5}$$

When the transformer is ideal (i.e., $k = 1$ and $sL_s \gg Z_{out1,eq}$ in Fig. 3.14),

$$A_{v1,idealTF}^2 \approx (ng_{m1}Z'_{out1})^2 = (ng_{m1}\frac{Z_{out1}}{n^2})^2 = (\frac{g_{m1}Z_{out1}}{n})^2. \tag{3.6}$$

3.1.4.2 Analysis of Two Interstage Interfacing Approaches

It is desired that the amplifier input is conjugate and noise matched to 50 Ω, and F_1 in Eq. 3.3 and Eq. 3.4 may be approximated by F_{min1} of the device (assuming that the passive

matching network is noiseless). From Eq. 3.3, it can be seen that the cascaded noise factor may be optimized by maximizing the gain from the input of the first stage to the input of the second stage (i.e., A_{v1}). The cascaded noise factor may also be improved by optimizing the noise factor of the second stage (F_2). These two approaches to interstage design are summarized as follows:

1. Maximize $A_{v1}^2(\frac{Z_{in2}}{Z_{in2}+Z_{out1}})^2$ in the denominator of Eq. 3.3 such that $F_{cascade}$ is minimized (i.e., maximize the gain of the first stage to suppress noise of the second stage);
2. Optimize $\frac{F_2-1}{A_{v1}^2/Z_{out1}}$ in Eq. 3.4 such that $F_{cascade}$ overall is minimized (i.e., set F_2 to F_{min2}).

Assuming $A_{v1} = g_{m1}Z_{out1}$ (i.e., unloaded voltage gain of stage 1), the first approach brings back the discussion of maximizing voltage gain using the microwave and analog viewpoints (covered in Section 2.3.1). Instead of maximizing gain, Eq. 3.4 suggests minimizing F_2 (in the numerator of Eq. 3.4). This may be realized by scaling and biasing the transistor used in the second stage, and designing the first stage to present the optimum source impedance for minimum noise to the input of the second stage (i.e., $F_2 \approx F_{min2} = F_{min1}$). By performing noise matching, the interstage gain is fixed, which may lead to lower gain than conjugate matching, or maximizing A_{v1} by realizing a higher Z_{out1}. However, noise is minimized using this approach.

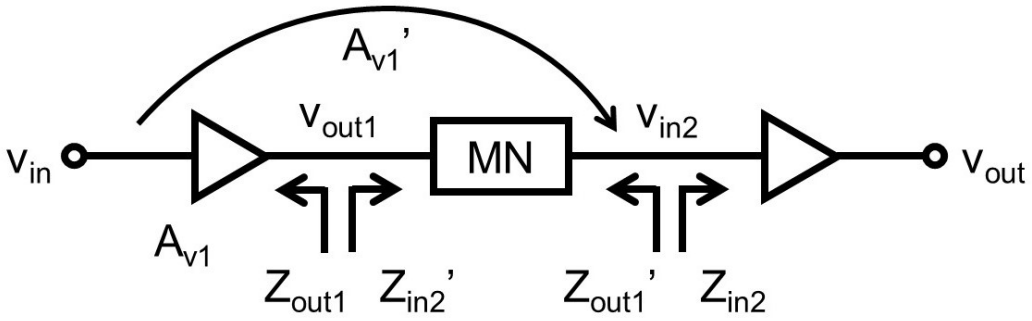


Figure 3.15: Interstage matching setup

The block diagram of Fig. 3.15 is used to compare the cascaded noise factor ($F_{cascade}$) of these two approaches.

Assuming that both gain stages in Fig. 3.14 consist of a single common-source[1] amplifier, the transconductance (g_{m1}), output impedance (Z_{out1}), and minimum noise factor (F_{min1}) of the first stage simulated at 28 GHz are 0.13 S, 17-$j24$ Ω and 0.6 dB, respectively.

For the first approach, the interstage transformer synthesizes Z_{out1} to its complex conjugate (Z'_{out1}). The second stage input impedance (Z_{in2}) and noise factor with respect to a source impedance of Z'_{out1} ($F_{2,wrtZ'_{out1}}$) are 12-$j32$ Ω and 2 dB, respectively. With a noise-matching interstage (the second approach), Z_{out1} is transformed to the optimum source impedance for minimum noise of the second stage ($Z_{n,opt2} = 49 + j80$ Ω) and the noise factor of the second stage is reduced to 0.6 dB (F_{min2}).

For simplicity, capacitive tuning (C_{tune}) is applied to tune out the magnetizing inductance at the operating frequency, such that Eq. 3.5 is simplified (i.e., $sk^2L_s||[Z_{out1}/(\frac{n}{k})^2] \approx Z_{out1}/(\frac{n}{k})^2$).

Applying Eq. 3.4 yields $NF_{cascade,approach1}$ of 0.81 dB and $NF_{cascade,approach2}$ of 0.68 dB (0.13 dB improvement). In real designs, the difference is expected to be higher as the transformer is not ideal, and capacitive tuning will only optimize the insertion loss of the transformer over a limited bandwidth (i.e., the term $\frac{sk^2L_s||[Z_{out1}/(\frac{n}{k})^2]}{s(1-k^2)L_s+sk^2L_s||[Z_{out1}/(\frac{n}{k})^2]}$ in Eq. 3.4 is less than 1, which leads to less suppression of F_2 by A_{v1}).

Assuming that reactances Z_{out1} and Z_{in2} are capacitive (i.e., FET drain output and gate input, respectively) and that the transformer is ideal, a noise match requires

$$Z'_{out1} = Z_{out1}/n^2 = Z_{n,opt2}. \tag{3.7}$$

Since the left side of Eq. 3.7 (Z'_{out1}) is capacitive and the right side ($Z_{n,opt2}$) is inductive (i.e., for FET gate input), Eq. 3.7 predicts that a noise match requires inductive tuning to transform Z'_{out1} to $Z_{n,opt2}$. The transformer turns ratio (n) may be selected to realize the transformation of the resistive part. An additional inductor may be tuned to realize the reactance needed for matching. Note that, by properly selecting $Z_{n,opt2}$ (e.g., via transistor sizing), the a voltage step-up transformer may be realized to boost A_{v1} in Eq. 3.6, which further improves the noise factor.

In the case of a linear transformer (Fig. 3.14b), the leakage inductance ($(1-k^2)L_s$) may be utilized as a tuning inductor for noise matching. In fact, a high-k transformer is desired for its wideband response, the magnitude of the leakage inductance is often insufficient for

[1]The values associated with the following calculations are taken from a single common-source amplifier (i.e., parameters such as noise factor, input/output impedance are subject to change if other topologies are used).

noise matching. An additional inductor placed in series with the leakage inductance may still be required.

In conclusion, utilizing a linear transformer for interstage matching requires: 1) the turns ratio to be selected for realizing transformation of the resistive component; 2) the magnetic coupling factor to be as high as possible for wideband frequency response; 3) the leakage inductance to be selected for realizing (or partially realizing) the reactance match.

3.1.5 Drawback of Interstage Noise Matching

Interstage noise matching is inherently narrowband. In this section, a two-port S-parameter simulation is performed with a post-layout, 20-nm long, 80-μm wide NFET. Reflection coefficient for the optimum noise factor is converted to impedance (i.e., Z_{opt}). The source resistance and reactance for lowest noise factor (i.e., R_{opt} and X_{opt}) versus frequency is depicted in Fig. 3.16.

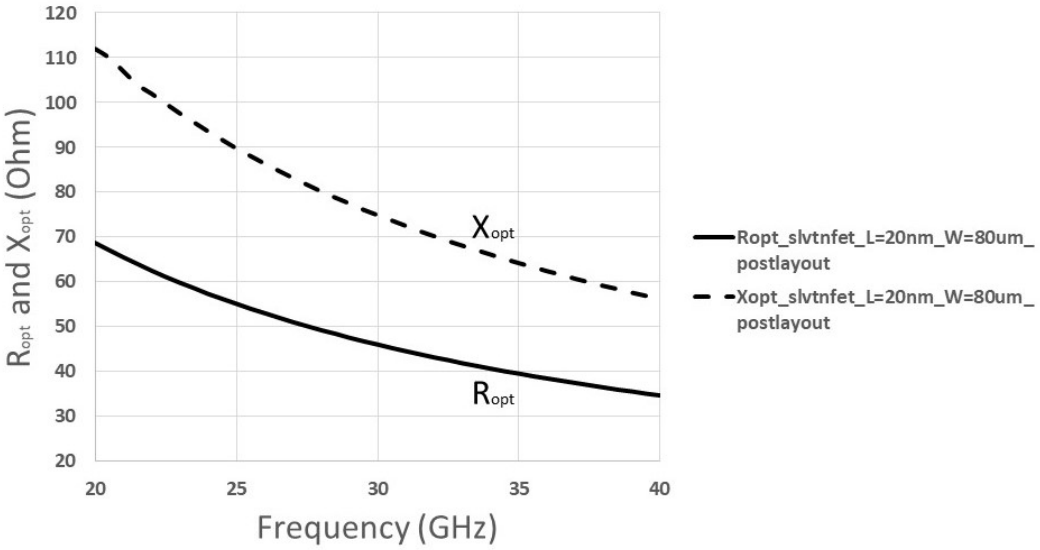


Figure 3.16: Source impedance for lowest noise across frequencies

The optimum resistance (R_{opt}) has a decreasing trend, as predicted by Eq. 2.8. The real part of the output impedance of the first stage (Z_{out1}) decreases as frequency increases, which matches with the trend of R_{opt} in the figure above. However, the optimum reactance (X_{opt}) is positive, and has a decreasing trend (i.e., a negative capacitance), which makes

it hard to match across a wide frequency range. Thus, the noise factor of the second stage is optimized only across a limited bandwidth.

3.1.6 Input Stage

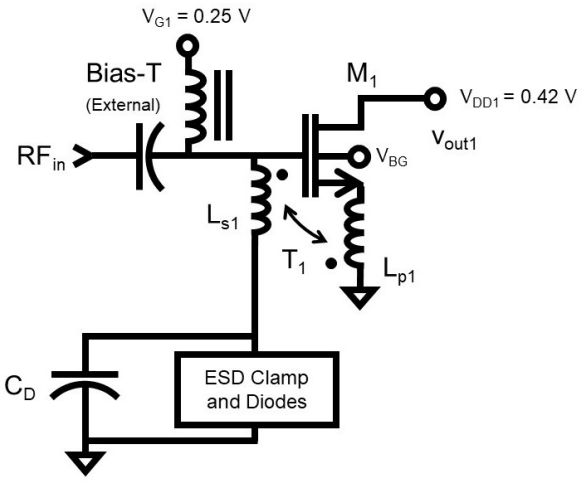


Figure 3.17: Input stage of the two-stage amplifier

The input stage consists of a common-source amplifier with a wideband input matching network, as shown in Fig. 3.17. Transformer T_1 couples the source and gate of M_1 such that the input impedance is dependent only on the transconductance of M_1 (g_{m1}) and the turns ratio (n) of T_1 (i.e., $Z_{in} \approx \frac{n^2}{(n+1)g_{m1}}$, as explained in Section 2.2.2). Transistor M_1 is scaled and biased for minimum noise, as explained in Section 3.1.2. As lowest noise factor is one of the major objectives, the cascode topology is not used due to the noise added by the common-gate stage. The supply voltage is set to 0.42 V, which almost halves the power consumption of the input stage (compared to the nominal supply voltage of 0.8 V), but at the cost of 0.06 dB increase in NF_{min} of M_1 (from 0.56 dB to 0.62 dB, as depicted in Fig. 3.8).

As explained in previous sections, isolation is poor for a single transistor stage. As a result, system stability must be considered, especially out-of-band, as the gain of the common-source amplifier is high at low frequencies. A common solution is to add a low-frequency damping network at the drain (for a low-noise amplifier) or at the gate (for a

power amplifier) of M_1, which consists of a parallel R-C network. At out-of-band frequencies, the added resistance in series with the load (when put at the drain), or in series with the gate reduces the gain, thereby improving stability. With respect to bandwidth, the capacitor shorts out the resistor, and the gain of the amplifier is not affected. However, additional noise is introduced by the resistor, which degrades the amplifier noise factor.

The transformer-coupled input matching network provides a low impedance path (via L_{s1} to ground, as depicted in Fig. 3.17) at low frequencies, which avoids the need for additional damping networks. Furthermore, an electrostatic discharge (ESD) clamp and diodes are attached between L_{s1} and ground, which provides ESD protection on the input port and extra damping. Simulation predicts a protection range of -3 kV to +600 V.

In summary, the total width of M_1 is set to 80 μm (with a unit finger width of 0.8 μm) to realize R_{opt} of 50 Ω. The gate is biased at 0.25 V, the supply voltage is set to 0.42 V, and the back-gate is biased at 2 V to ensure M_1 operates around $\frac{I_{DS}}{W}_{opt}$ for the lowest device noise. The output impedance of the first stage (Z_{out1}) is approximately $26 - j62$ Ω.

3.1.7 Output Stage

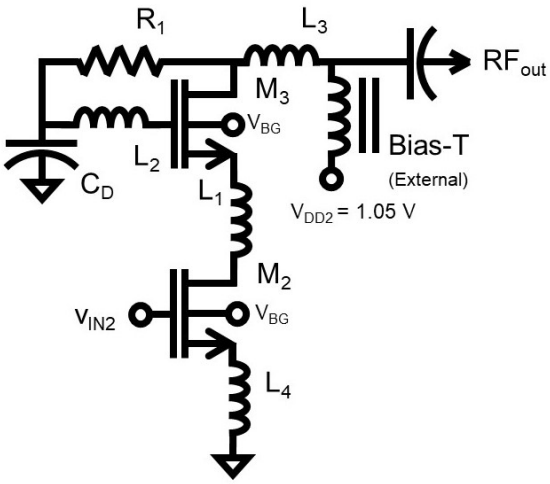


Figure 3.18: Output stage of the two-stage amplifier

The gate of M_2 (as depicted in Fig. 3.18) is resistively biased to the same gate voltage (0.25 V) as M_1 from the input stage to ensure the transistor operates at $\frac{I_{DS}}{W}_{opt}$ for the lowest

noise. Inductors L_1, L_2 and L_3 are used to peak the frequency response (as explained in Section 2.3.2). The gate of M_3 is resistively biased from the supply voltage, and the value R_1 is tuned such that the real part of the output impedance is 50 Ω. The reactive part is nulled by L_3 from 24 to 32 GHz.

The width of M_2 and M_3 are chosen based on the specification for gain, noise factor, linearity and power consumption. Assuming that transistor M_2 is biased close to $\frac{I_{DS}}{W}_{opt}$, the transconductance (g_{m2}) and power consumption increase as M_2 is scaled wider. The gain of the second stage also increases (i.e., the gain may be estimated by $g_{m2}Z_L$, where Z_L consists of two 50 Ω in parallel). Furthermore, increasing the width of M_2 lowers $R_{n,opt2}$, and lowers the turns ratio required for transforming the output impedance of the first stage (Z_{out1}) to $Z_{n,opt2}$, which leads to a lower gain from the first stage. In this design, the width of M_2 and M_3 are set to the same as M_1, and $Z_{n,opt2}$ is approximately $49 + j73$ Ω and the real part of output impedance is 50 Ω.

3.1.8 Interstage Transformer

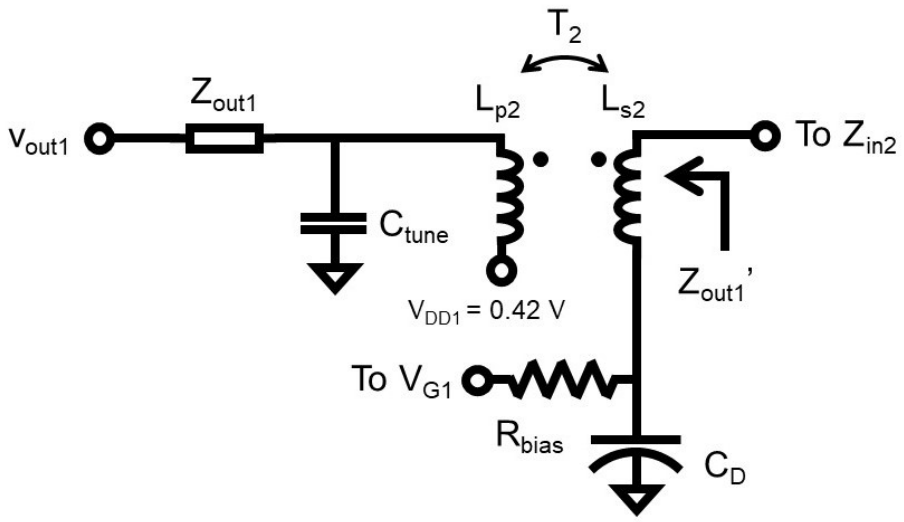


Figure 3.19: Interstage coupling for the two-stage amplifier

As depicted in Fig. 3.19, the interstage transformer is designed to transform Z_{out1} ($26 - j62$ Ω) to $Z_{n,opt2}$ ($49 + j73$ Ω). The synthesized impedance (Z'_{out1}) is given by (C_{tune} excluded)

$$Z'_{out1} = sL_{s2} + \frac{\omega^2 M^2 R_{11}}{R_{11}^2 + X_{11}^2} + \frac{-j\omega^2 M^2 X_{11}}{R_{11}^2 + X_{11}^2}, \tag{3.8}$$

where L_{s2} is the inductance of the secondary winding, ω is the operating frequency (rads/sec), and R_{11} and X_{11} represents the total resistance and reactance on the primary side, respectively (i.e., $R_{11} = R_{out1} + R_{p2}$, $X_{11} = \omega L_{p2} + X_{out1}$). The mutual inductance (M) is given by $k\sqrt{L_{p2}L_{s2}}$ (i.e., knL_{p2} assuming $L_{s2} = n^2 L_{p2}$).

Assuming that the magnetic coupling factor (k) is 0.6 and a tuning capacitor is applied on the primary side (i.e., the magnetizing inductance is resonated out), the model shown in Fig. 3.14b (with the ideal transformer driving a pure resistive shunt component) is utilized to simplify the calculation. The transformer turns ratio may be calculated (n) from the ratio of the real part (i.e., $n/k = \sqrt{26/49} \approx 0.73$, and $n \approx 0.44$). The leakage inductance (or the leakage padded with additional inductance in series) matches the reactive component. Once the value of the leakage inductance is fixed, the secondary winding inductance (L_{s2}) may be found from $L_{s2}/(1-k^2)$. Finally, the primary winding inductance is calculated via the turns ratio. Note that the calculation above does not include parasitic capacitances of the windings or the interwinding capacitance, which both have a significant impact on the result. Therefore, the calculated values may only be used as initial estimates for simulation. The actual dimensions of the transformer physical layout are fixed after performing electromagnetic simulations.

3.1.9 Summary of Design Methodology

The design methodology of the two-stage, transformer-based, interstage noise matching amplifier is summarized as follows (may be extended to more than two stages):

1. Characterize available transistors in simulation to find the optimum biasing condition (i.e., $\frac{I_{DS}}{W}_{opt}$) and dimensions of the transistor connected to the input port (i.e., for input noise matching). All stages targeted for lowest noise should be biased close to the same $\frac{I_{DS}}{W}_{opt}$.

2. Design the input matching network to achieve a conjugate and noise match simultaneously. The output impedance of the first stage should be defined (Z_{out1}) during this step.

3. The dimensions of transistors in the second stage (or subsequent stages) should be selected based on gain, noise factor, linearity and power consumption specifications, etc.

4. Design the output matching network to realize an output conjugate match (may vary, depending on the block interfacing with the amplifier). The input impedance and the optimum noise source impedance of the second stage should be determined (Z_{in2} and $Z_{n,opt2}$).

5. Design the interstage transformer which transforms Z_{out1} to $Z_{n,opt2}$.

6. Sweep bias voltages at the gates of transistors for lowest noise factor, as the optimum bias point may vary after lossy matching networks are added to the circuit.

7. Due to finite isolation between stages (i.e., Z_{out1} affects the output matching and Z_{in2} affects the input matching), steps 2, 4, 5 and 6 should be iterated in a fixed order to converge on the final design.

3.1.10 Simulation Results

The post-layout simulation results are summarized in Table 3.1 below. Two modes (low-noise and low-power) are distinguished by different back-gate biasing (i.e., current density). In low-noise mode, transistors are biased close to $\frac{I_{DS}}{W}_{opt}$ and the lowest noise figure of 1.63 dB occurs at 28 GHz. In low-power mode, the total power consumption of the amplifier is 5.6 mW, down from 17.3 mW.

[1] P_{-1dB} is simulated in the time domain at peak-gain frequency (23.5 GHz)

[2] IP_3 is simulated in the time domain for two-tone signal (f_1 = 23.5 GHz, f_2 = 23.5005 GHz) with FFT

Table 3.1: Simulation results of the two-stage amplifier (Design A)

Parameter	Low-Noise Mode $V_{BG} = 2$ V	Low-Power Mode $V_{BG} = 0.62$ V	Target Spec.
Bandwidth (GHz)	23 - 32.5	23 - 32.5	As wide as possible
S_{21} (dB)	20.6 ± 1.5	18.3 ± 1.1	> 15
NF (dB)	1.79 ± 0.16	2.21 ± 0.13	< 2
S_{11} (dB)	< -10	< -9.9	< -10
S_{22} (dB)	< -15	< -8.5	< -10
$P_{-1dB,out}$ (dBm)	-3.5[1]	-8.3[1]	> -9.6
OIP_3 (dBm)	+6[2]	+0.8[2]	> 0
P_{DC} (mW)	17.3	5.6	< 40

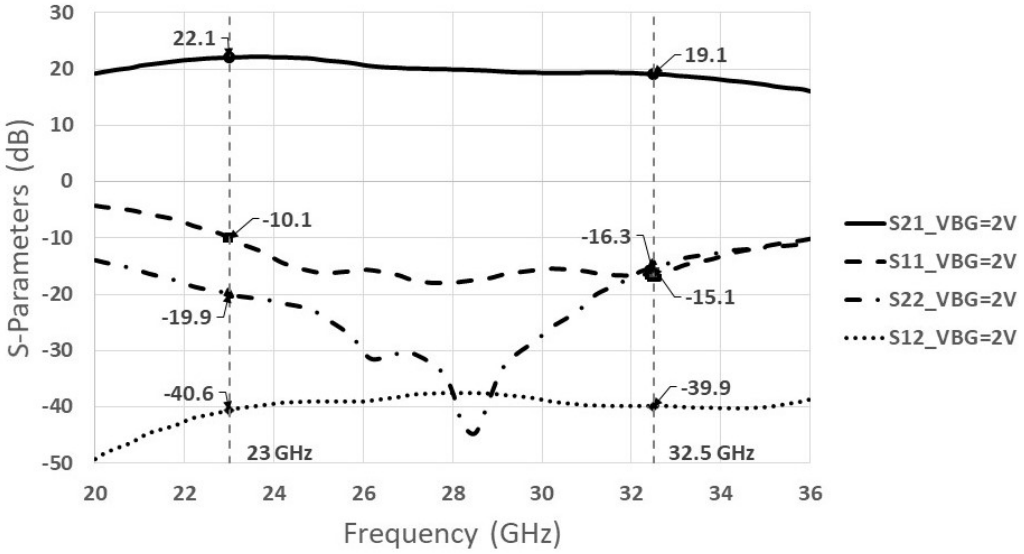


Figure 3.20: Simulated S-parameters of the two-stage amplifier in low-noise mode

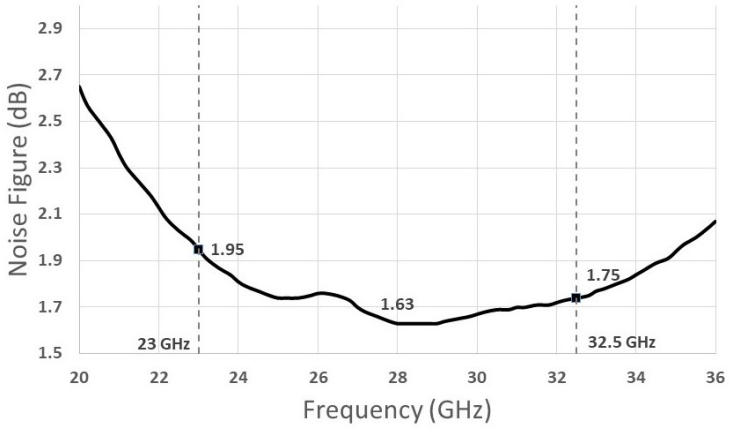


Figure 3.21: Simulated noise figure of the two-stage amplifier in low-noise mode

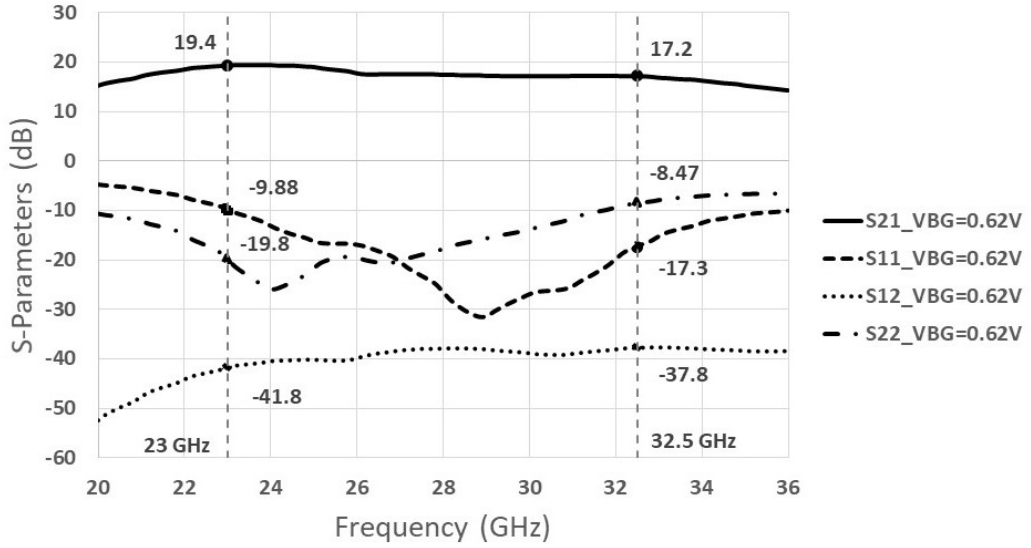


Figure 3.22: Simulated S-parameters of the two-stage amplifier in low-power mode

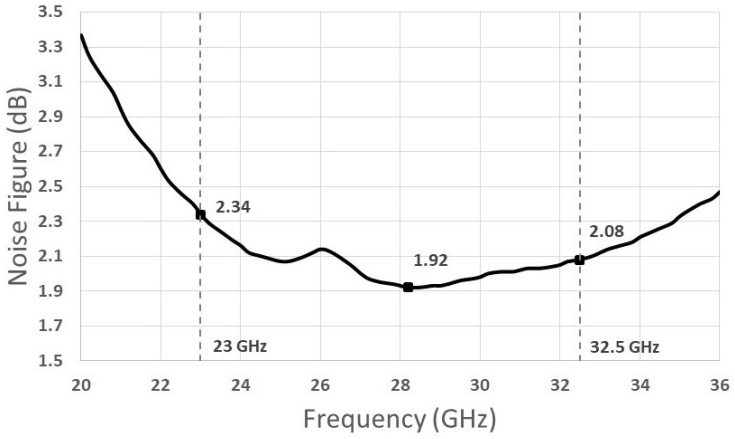


Figure 3.23: Simulated noise figure of the two-stage amplifier in low-power mode

3.2 Design B: a Two-Stage, High-Linearity Amplifier with Interstage Noise Matching

To improve the bandwidth and linearity of the amplifier, another two-stage amplifier is designed with optimizations to both stages. The transistor characterization is the same as covered in Section 3.1.3.

3.2.1 Proposed Circuit Topology

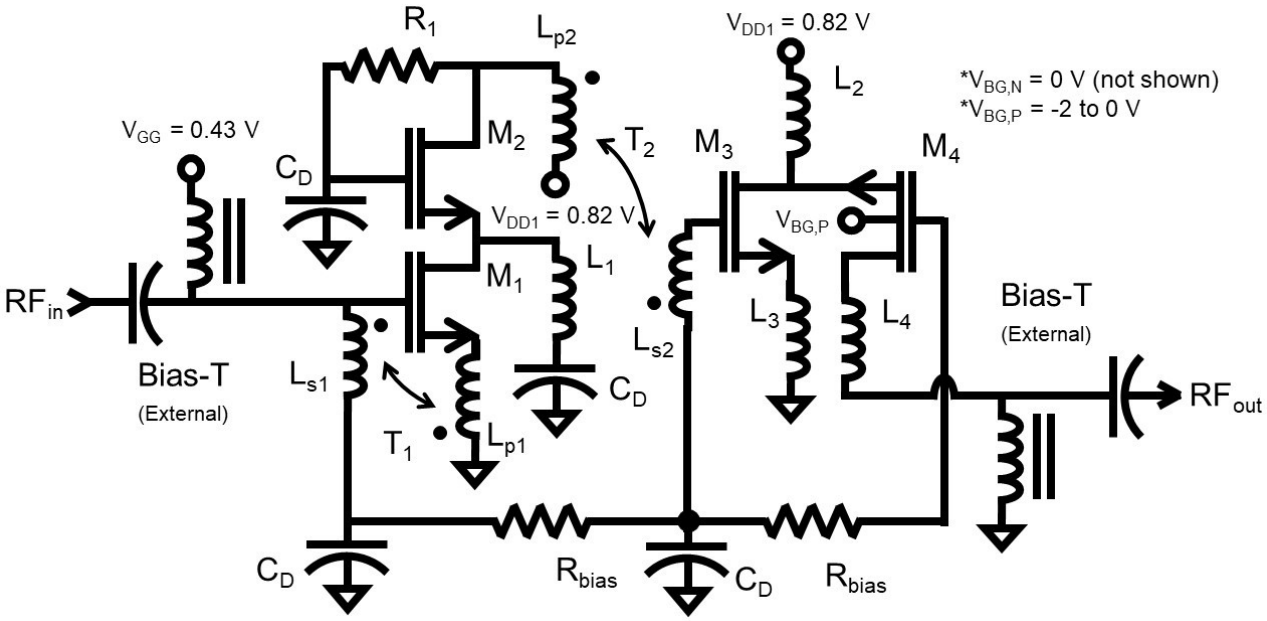


Figure 3.24: Topology of a two-stage, high-linearity amplifier with interstage noise matching

3.2.2 Input Stage Design

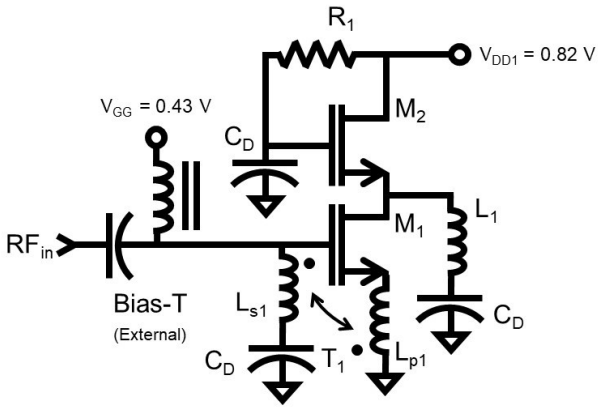


Figure 3.25: Input stage of the two-stage, high-linearity amplifier

The topology of the first stage changes from a common-source to a cascode which improves isolation, suppresses the Miller effect and increases the bandwidth by 6.5 GHz (24 - 40 GHz, up from 23 - 32.5 GHz). Noise contributed by the common-gate stage is suppressed by the shunt peaking inductor L_1, as explained in Section 2.3.2. Transistor M_2 is resistively biased via R_1, which is set to a large value to limit the noise contributed by the cascode.

One drawback is that the supply voltage is now set to 0.82 V (compared to 0.42 V used in the previous design) to ensure that M_1 and M_2 are in the saturation region, which almost doubles the power consumption. The benefit is that both stages are on the same supply, which eliminates the need for dual supply voltage generation and regulation.

The total width of M_1 and M_2 are set to 80 μm (with a unit finger width of 0.8 μm) to realize R_{opt} of 50 Ω. The gate of M_1 is biased at 0.43 V, and the back-gate is grounded to ensure that M_1 operates close to $\frac{I_{DS}}{W}_{opt}$ for the lowest device noise.

3.2.3 Output Stage Design

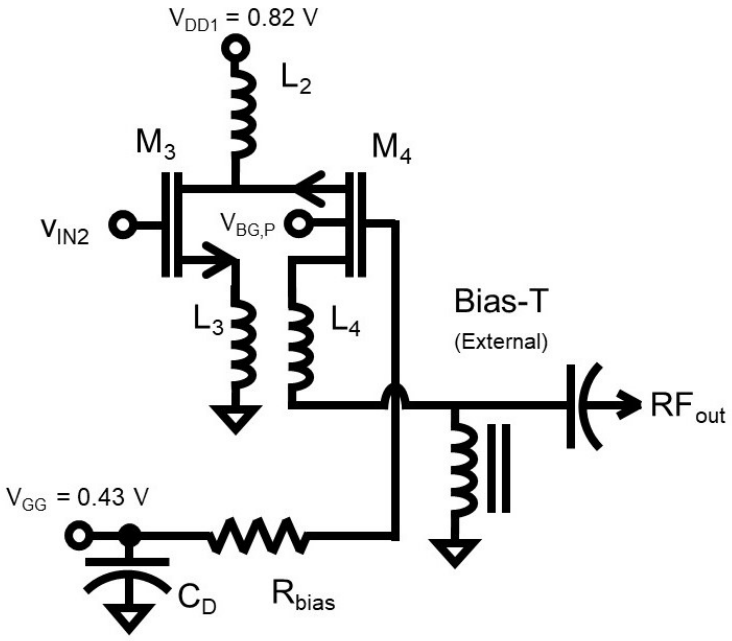


Figure 3.26: Output stage of the two-stage, high-linearity amplifier

Instead of a cascode, transistor M_4 is folded such that the voltage headroom at the drain of M_3 increases. As a result, P_{-1dB} at the output is improved from -5.8 dBm (with cascode output stage in Section 3.1.10) to +4.3 dBm. The gate of M_3 and M_4 are resistively biased at the same voltage as M_1 from the first stage, which ensures that M_3 operates close to $\frac{I_{DS}}{W}_{opt}$ to lower the noise contributed by the second stage. Peaking inductor L_2 serves as a DC-feed and suppresses noise contributed by M_4. Inductor L_3 provides degeneration to improve stability.

The drawback of utilizing a folded-cascode topology is that the DC current flowing through M_4 is not reused, which leads to higher power consumption than a regular cascode (i.e., from 17.3 mW in Section 3.1.10 to 35.5 mW). The widths of M_3 and M_4 are set to 80 μm (with a unit finger width of 0.8 μm) to ensure sufficient gain and linearity, as well as keeping the overall power consumption below 40 mW.

3.2.4 Simulation Results

The post-layout simulation results are summarized in Table 3.2. The bandwidth is improved from 9.5 to 16 GHz. The lowest noise figure in-band is 1.66 dB, which is close to 1.63 dB of the previous design. As explained in previous sections, noise matching has the disadvantage of being narrowband, which may be seen from the 0.38 dB in-band noise figure variation, up from 0.16 dB. The improvement in linearity comes at the cost of power consumption, which rises from 17.3 mW (maximum) to 35.5 mW. The 1-dB compression point at the output improves from -5.8 to 4.3 dBm and the third-order intercept point at the output increases to 15.4 dBm. Note that the reverse isolation (S_{12}) and output return loss (S_{22}) have an increasing trend beyond 45 GHz. These are due to the series resonance formed by L_2, C_{ds4} and L_4, which makes the output impedance deviate from 50 Ω rapidly.

Table 3.2: Simulation results of the two-stage, high-linearity amplifier (Design B)

Parameter	Design B	Design A Low-Noise mode	Delta
Bandwidth (GHz)	24 - 40	23 - 32.5	+6.5
S_{21} (dB)	20.1 ± 1.3	20.6 ± 1.5	-0.5
NF (dB)	2.04 ± 0.38	1.79 ± 0.16	-0.03
S_{11} (dB)	< -10	< -10	-
S_{22} (dB)	< -10	< -15	+5
$P_{-1dB,out}$ (dBm)	+4.3[1]	-5.8	+10.1
OIP_3 (dBm)	+15.4[2]	3.7	+11.7
P_{DC} (mW)	35.5	17.3	+18.2

[1] P_{-1dB} is simulated in the time domain at 28 GHz

[2] IP_3 is simuated in the time domain for two-tone signal ($f_1 = 28$ GHz, $f_2 = 28.01$ GHz) with FFT

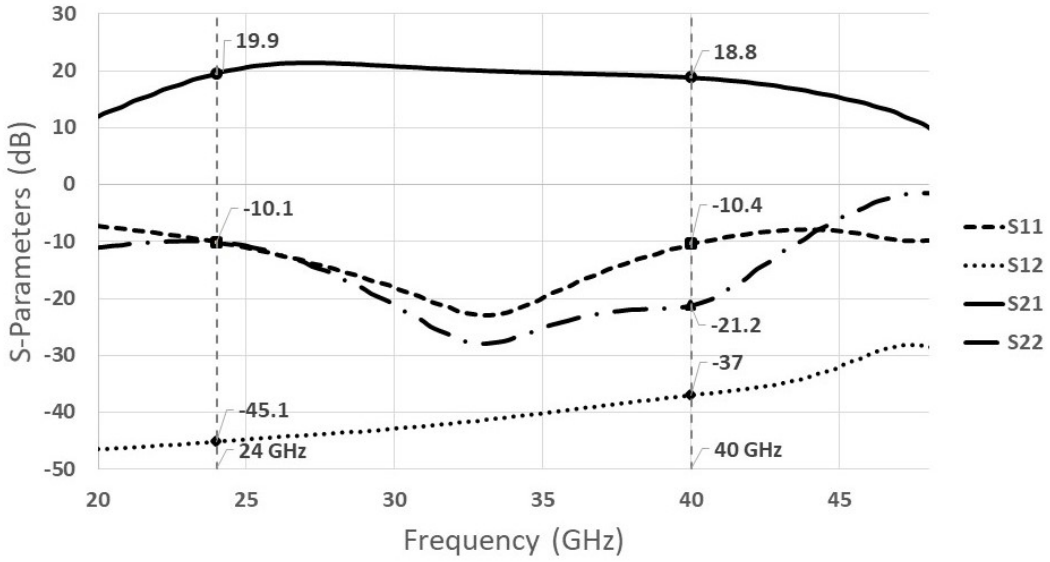


Figure 3.27: Simulated S-parameters of the two-stage, high-linearity amplifier

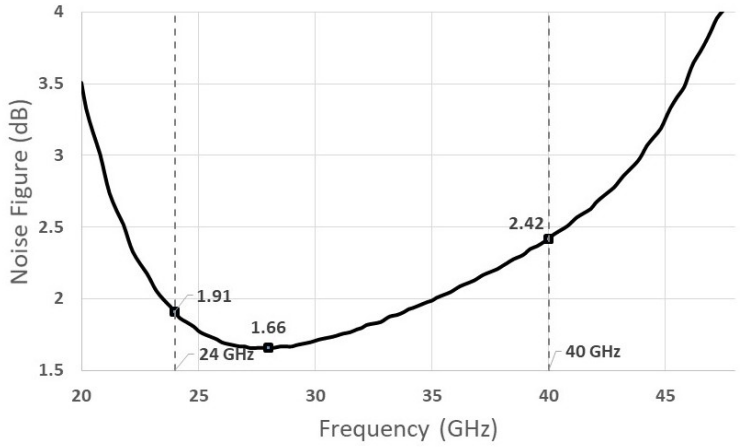


Figure 3.28: Simulated noise figure of the two-stage, high-linearity amplifier

Chapter 4

Layout and Implementation

As semiconductor processes scale down, the pitch and thickness of many metal interconnects also decrease. This leads to higher resistance and lower electromigration capability. The 22-nm CMOS-SOI process offers a 10 or 11-layer metal stack-up, with 7 thin layers (M_1 to M_7) and up to 4 thick metal layers (M_8 to M_{10} or M_{11}).

4.1 General Considerations of FET Layout for Low-Noise Designs

As mentioned in Section 2.1.1, the gate resistance (r_{gg}) and transition frequency (f_T) of the transistor are major contributors to its noise factor. Thus, the transistor has to be laid out with interconnect resistance and capacitance minimized, which is challenging due to an R-C trade-off (i.e., if interconnects are scaled wide, resistance is low but capacitance is high due to the increased metal area). For narrow interconnects, the capacitance is optimized but resistance increases. Another critical factor is electromigration, which defines the long-term current carrying capability (DC and AC currents) of the interconnect. For example, an 80-μm transistor biased at 0.2 mA/μm requires interconnections that are capable of flowing 16 mA of DC current.

To tackle these challenges, transistors laid out in a grid pattern, with double-gate contact and dual-lane, staircase source/drain connections (see Fig. 4.1) are used.

4.1.1 Gate, Drain and Source Connections

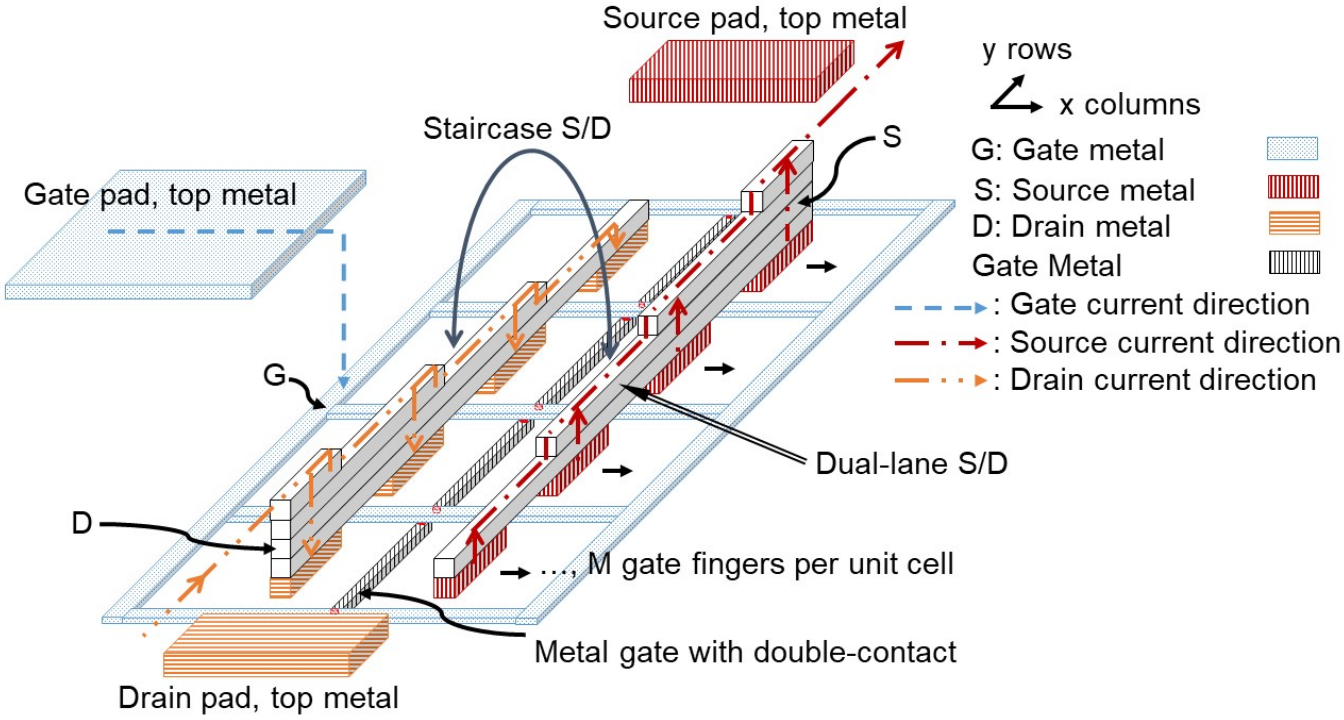


Figure 4.1: Simplified grid transistor layout (4-row x 1-column) utilizing 5 x 0.8 μm unit cells

Assuming that the gate signal is accessed at the top metal level, via stacks are utilized to ensure that the signal travels vertically with minimal added gate resistance, r_{gg}. The lowest two metal layers (M_1 and M_2) are shunted together and routed as a mesh to further reduce the interconnect resistance. However, this increases C_{gs} and C_{gd}, as it introduces sidewall capacitance between the gate and the other two terminals. Since M_1 and M_2 are relatively thin metals, the added capacitances are negligible. Moreover, a double-gate contact is implemented to lower r_{gg}, as explained in Section 2.1.1.

The drain connection (top metal) is placed adjacent to the gate such that the sidewall capacitance between the gate and drain top metals is reduced, as depicted in Fig. 4.1. The source connection is generally implemented by shunting lower metal layers to increase their current carrying capability. The source conducts current directly to the local ground plane,

which is also implemented in lower metals. Thus, source metals may be routed on both sides of the transistor, which relaxes the electromigration constraints by splitting the DC current into two paths. However, the transistor source may also connect to a transformer implemented in top metal (e.g., M_1 in Fig. 3.1 and Fig. 3.24) for the amplifier designed in this thesis, which makes the dual-source routing complicated. Thus, a single source connection is placed opposite from the drain in this case. As depicted in Fig. 4.1, the routing of the source connection is similar to the drain (but flipped along the y-axis), as both source and drain conduct the same amount of bias current.

4.1.2 Electromigration Considerations

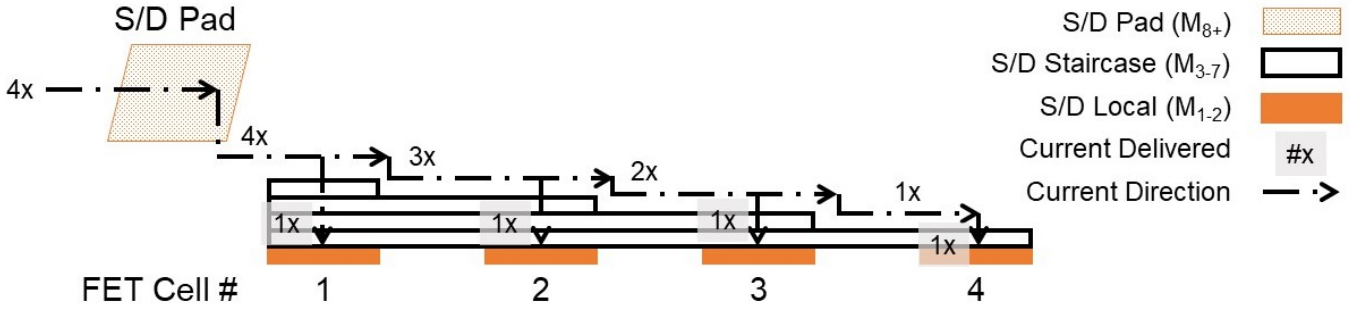
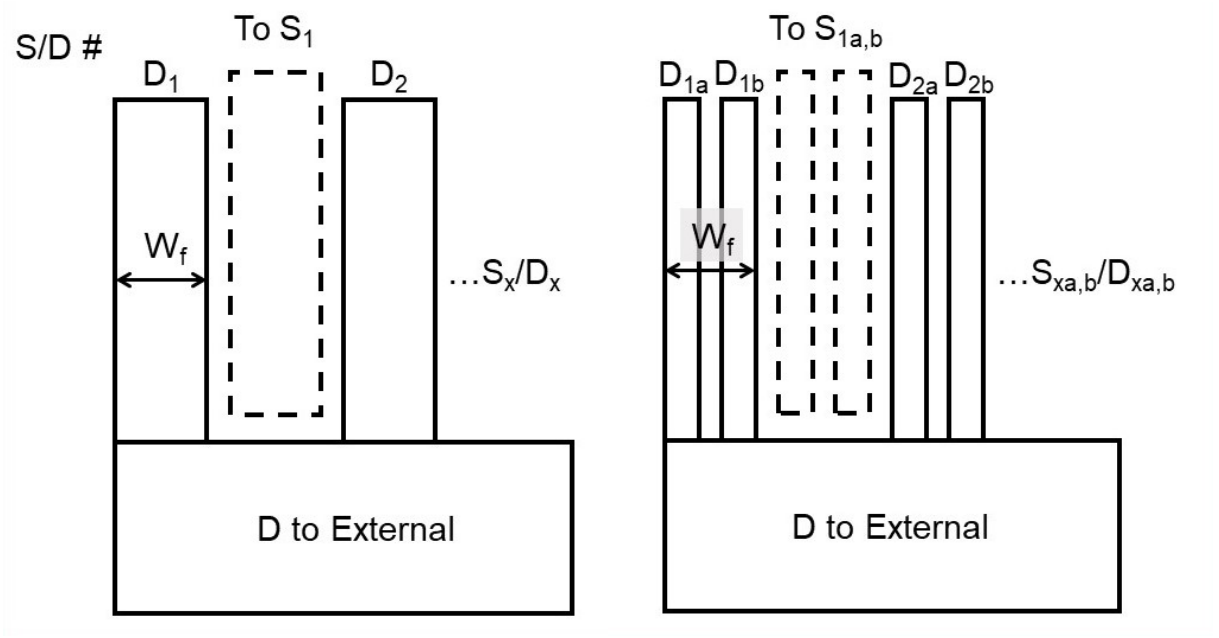


Figure 4.2: Simplified side view of the staircase layout of source/drain

To meet the electromigration requirements on the bias current, the source/drain connections are implemented using stack of metals rising from the active region to a specific layer. The total current carrying capacity is a multiple of the capacity for a single metal layer. However, this approach introduces sidewall capacitance (C_{ds}), which affects the output impedance as explained in Section 3.1.5. To minimize C_{ds}, a staircase-source/drain layout is utilized, as depicted in Fig. 4.1. As the current flow reduces because a portion of the current branches down to the active region (see Fig. 4.2), the metal required to carry the current decreases. Thus, implementing source and drain interconnects as a staircase of metals reduces C_{ds} and permits the electromigration rules to be satisfied. The downside of the staircase layout is that it increases the equivalent resistance of the source and drain interconnects, as fewer metals are shunted when the current flow requirement is smaller.

The 22-nm CMOS-SOI process offers three metal interconnect pitches for the SLVT transistor (denoted by 1x, 2x and 3x FET, respectively). A higher pitch corresponds to

a wider distance between gate fingers, which leads to a larger physical size for a fixed transistor width. Therefore, the equivalent gate resistance and device noise factor increase with the increase in pitch (i.e., the gate signal has to travel further). To minimize the device noise factor, transistors with a lower pitch are preferred. However, due to electromigration limitations, even a 3x FET (i.e., maximum source/drain metal width) using the staircase layout is not capable of carrying the bias current required. To avoid utilizing manifold source/drain connections (which requires thick metal layers and increase C_{gs} and C_{gd}, e.g. transistor layout in [30]), a dual-lane source and drain are implemented.



(a) Single-lane layout

(b) Dual-lane layout

Figure 4.3: Dual-lane layout for source and drain to improve electromigration reliability

As depicted in Fig. 4.3, each pair of dual-lane source/drain wires (e.g. D_{1a} and D_{1b} in Fig. 4.3b) has less equivalent width compared to the single-lane alternative (i.e., D_1 in Fig. 4.3a). However, the dual-lane layout is capable of conducting more current if the width and length of the wire are lower than a certain threshold (defined according to electromigration limits). This may be explained by the short-length effect. The short-length effect represents the immunity of electromigration in a wire which is less than the Blech length [31]. The Blech length defines the threshold, where (partial) cancellation

occurs between electromigration (in the direction of current flow) and stress migration (in a direction opposite direction to the current flow). In conclusion, 2x is selected as the transistor pitch, which has a noise factor between 1x and 3x pitches, and its width is sufficient to implement the dual-lane layout for source and drain connections. The length of the source and drain wires (i.e., the transistor size in the y-direction in Fig. 4.1) are selected carefully to ensure that the short-length electromigration effect applies.

4.2 Transformer Layout

High magnetic coupling factor (k) is a main goal for the transformer realization, as it leads to wider bandwidth and lower losses. Minimum metal spacing and narrow metal width are utilized to achieve high k [29]. Table 4.1 records the performance of different transformer configurations from simulations. Two 1:1 transformers (planar and overlay) are laid out and simulated. The primary (or secondary) inductances are kept equal to permit a fair comparison. Note that the 11-layer metal stack-up is used to ensure the fairness of comparison for k, as it provides more than one metal layer with the same physical parameters (thickness, conductivity, etc.). The distance between the top of the substrate and the bottom of the transformer is kept the same for both configurations to realize similar substrate losses (i.e., M_9 is used to implement the planar transformer, and M_9 & M_{10} are used for the overlay candidate). No explicit shielding is used, as the transformer will be used in a single-ended amplifier.

Table 4.1: Simulated parameters for different transformer configurations (with metal width of 5 μm and spacing of 2 μm, using the 11-layer stack-up)

Winding Configuration	OD (μm)	L_p (pH)	R_p[1] (Ω)	R_s[1] (Ω)	k[1]	C_o (fF)	Minimum Attenuation (dB)	MAG (dB)
1:1 Planar Inverting	96	173	0.38	0.38	0.55	54	0.65 @ 77 GHz	-0.6 @ 65 GHz
1:1 Planar non-Inverting	96	173	0.38	0.38	0.55	54	4.67 @ 51 GHz	-0.84 @ 40 GHz
1:1 Overlay Inverting	90	173	0.34	0.36	0.65	68	0.53 @ 60 GHz	-0.5 @ 55 GHz
1:1 Overlay non-Inverting	90	173	0.34	0.36	0.65	68	3.26 @ 54 GHz	-0.66 @ 40 GHz

[1]Parameters simulated at 1 MHz (i.e., the skin effect of resistance is not included; self-inductance and magnetic coupling factor are taken at a frequency far from self-resonance frequency)

For insertion loss, the overlay transformer outperforms the planar layout due to its higher magnetic coupling factor (0.65 versus 0.55 for stacked and interleaved transformers, respectively). However, the planar transformer has a lower interwinding capacitance (C_o of 54 fF), which leads to better performance at higher frequencies. For winding resistances, the planar structure has better consistency between windings, whereas the overlay structure shows a small difference (i.e., 0.34 Ω in the primary winding (R_p) and 0.36 Ω in the secondary winding (R_s)). This may be explained by the fact that the primary winding is self-shielded by the secondary, thus losses due to the substrate create a difference between R_p and R_s that becomes more prominent at higher frequencies. For both configurations, the transformer has lower loss and wider bandwidth in inverting mode, which matches with the prediction in [29].

4.2.1 Interstage Transformer Layout

The interstage transformer is designed with the 10-layer metal stack-up, which has only three thick metal layers (i.e., M_8 to M_{10} and without M_{11}). The three metal layers have uneven physical parameters (thickness, conductivity, etc.). Metal M_9 is better than other two in term of ohmic loss), which increases the insertion loss of an overlay transformer (i.e., one of M_8 or M_{10} has to be used to implement one of the windings). Therefore, the transformer is implemented as a planar structure with only M_9. To improve the frequency response of the planar structure, the transformer is implemented using multi-turns instead of a single-turn. Table 4.2 shows the difference between single- and multi-turn layouts for a fixed winding inductance.

Table 4.2: Simulated parameters for single and multi-turn interleaved transformers (metal width of 6 μm and spacing of 2.6 μm for a 10-layer stack-up)

Winding Configuration	OD (μm)	L_p (pH)	R_p[1] (Ω)	R_s[1] (Ω)	k[1]	C_o (fF)	Minimum Attenuation (dB)	MAG (dB)
1:1 Planar Inverting	179	605	0.77	0.77	0.65	161	1.0 @ 25 GHz	-0.74 @ 10 GHz
1:1 Planar non-Inverting	179	605	0.77	0.77	0.65	161	3.6 @ 16 GHz	-0.88 @ 8 GHz
2:2 Planar Inverting	134	604	0.85	0.85	0.66	138	0.91 @ 29 GHz	-0.76 @ 14 GHz
2:2 Planar non-Inverting	134	604	0.85	0.85	0.66	138	3.3 @ 18 GHz	-0.89 @ 11 GHz

[1]Parameters simulated at 1 MHz (i.e., skin effect is not included; self-inductance and magnetic coupling factor are determined at a frequency far from self-resonant frequency)

As shown in Table 4.2, the multi-turn transformer has lower interwinding capacitance, and thus the bandwidth and frequency response of the transformer is improved. Another benefit is that the multi-turn transformer occupies less area than the single-turn alternative. The final layout of the interstage transformer is depicted in Fig. 4.4.

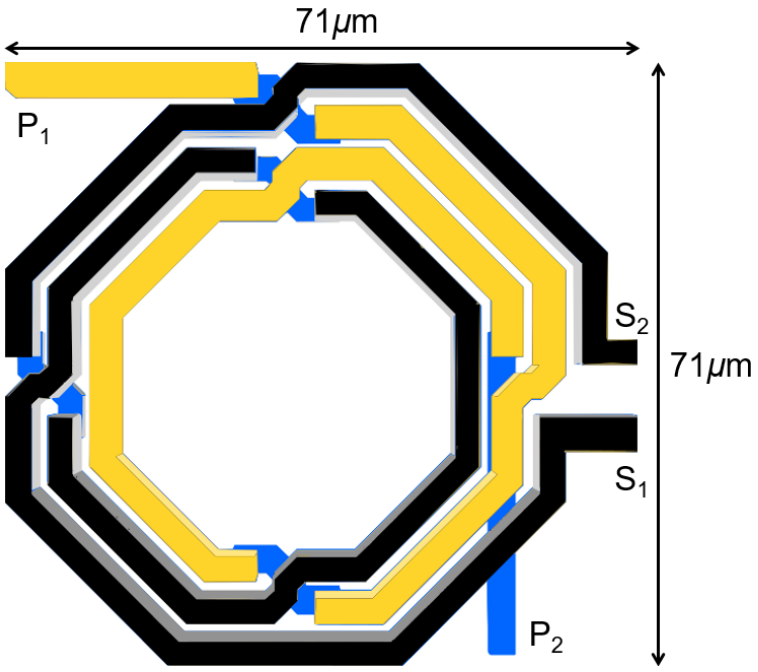


Figure 4.4: Layout of the interstage transformer

The primary and secondary windings are indicated in yellow and black traces, respectively. The connections (ports) to external circuitry are denoted by $P_1 - P_2$ for the primary and $S_1 - S_2$ for the secondary winding. The core of the transformer contains five octagonal edges in yellow (primary) and eight edges in grey (secondary), which indicates the turns ratio is 1:1.6 as drawn (Table 4.3 shows the electrical turns ratio is about 1:2.5). The interconnects provide additional inductance which improves the noise matching, as explained in Section 3.1.4.2. The interstage transformer parameters and performance are summarized in Table 4.3 and plotted in Fig. 4.5. Note that the main purpose of the transformer is to synthesize the optimum source impedance (Z_{opt2}). The input reflection coefficient is used as the indicator of how close the transformed impedance is to the optimum source impedance for minimum noise at the input of the second stage. It is depicted in Fig. 4.6.

Table 4.3: Simulated parameters for the interstage transformer

Winding Configuration	L_p (pH)	L_s (pH)	R_p[1] (Ω)	R_s[1] (Ω)	k[1]	C_o (fF)	Minimum Attenuation (dB)	MAG (dB)
Core Transformer non-Inverting	170	329	0.66	0.93	0.56	47	5.1 @ 43 GHz	-0.98 @ 36 GHz
Interconnect Parasitics[2]	139	10	0.42	0	-	15	-	-

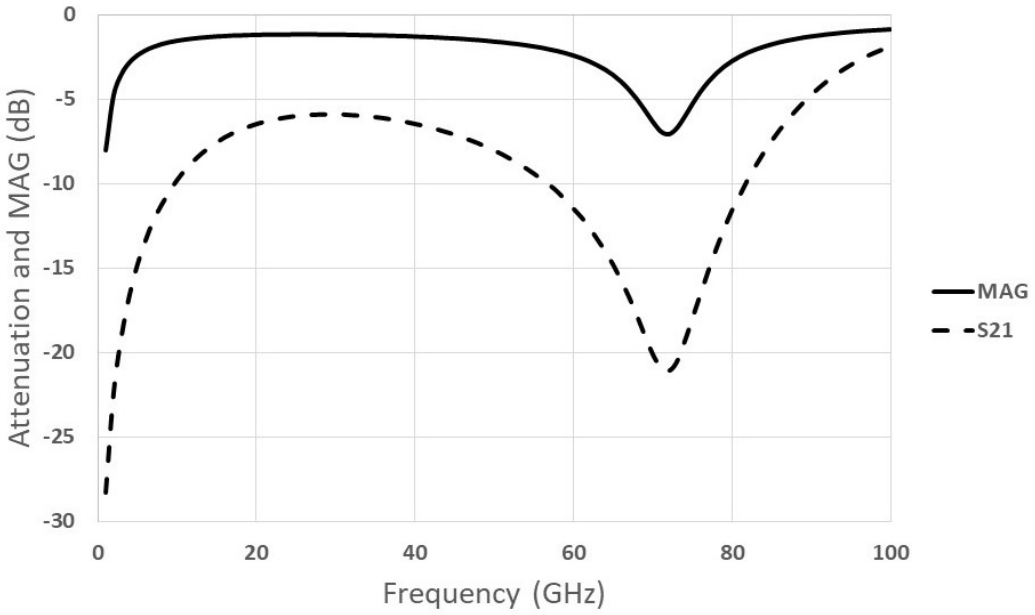


Figure 4.5: Interstage transformer insertion loss and maximum available gain

[1]Parameters simulated at 1 MHz (i.e., skin effect is not included; self-inductance and magnetic coupling factor are determined at a frequency far from self-resonant frequency)

[2]Interconnect resistance and inductance are added in series and capacitance in parallel on top of the core

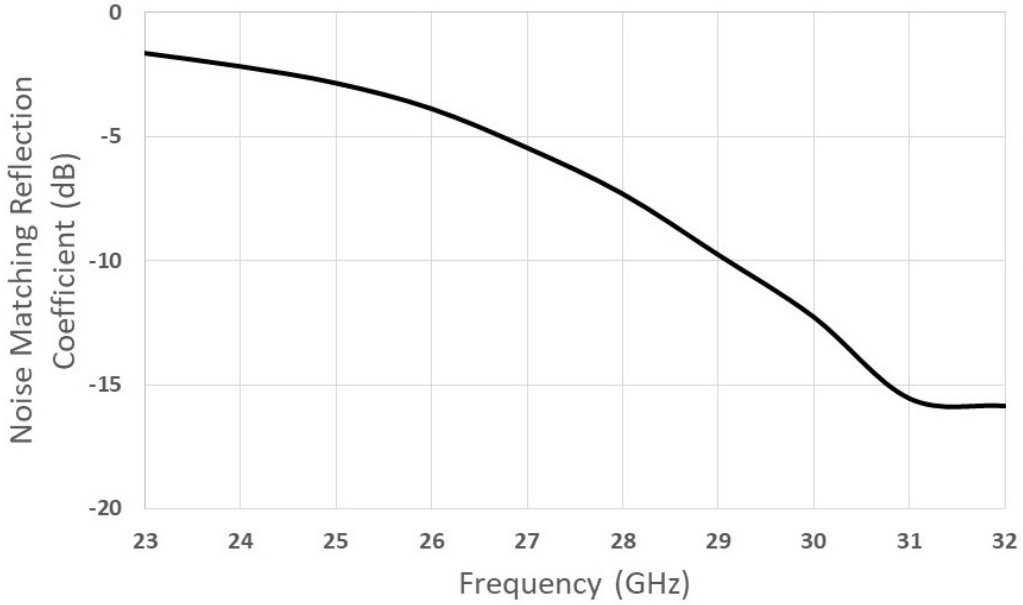
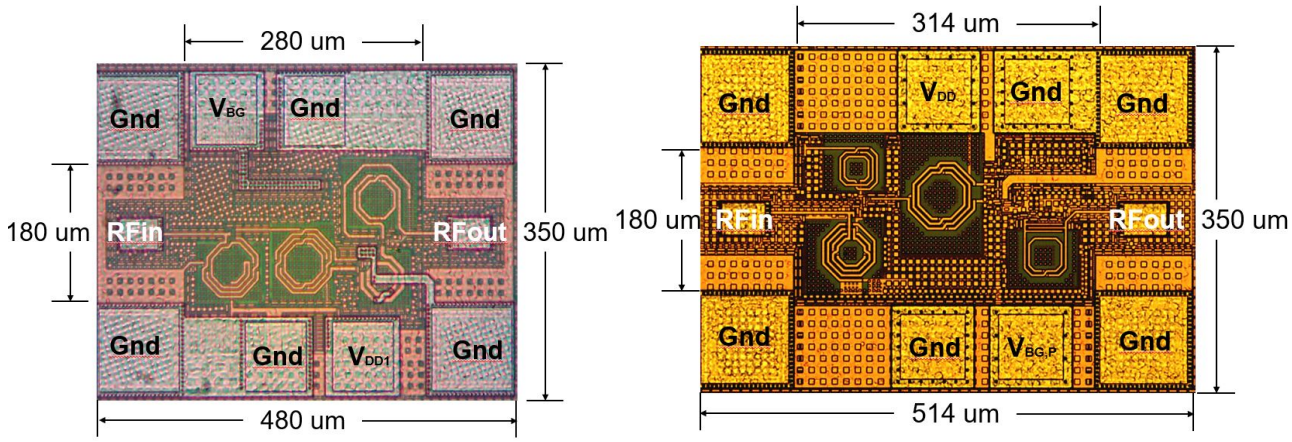


Figure 4.6: Noise matching reflection coefficient at the second-stage input (including the interstage transformer) from simulation

Chapter 5

Experimental Results

The micrographs of two amplifier prototypes fabricated in the 22-nm CMOS-SOI technology are shown in Fig. 5.1. The total chip area is 0.48×0.35 mm^2, and the active chip area is 0.28×0.18 mm^2 for Design A. For Design B, the total chip area is 0.51×0.35 mm^2, and the active chip area is 0.31×0.18 mm^2. The RF performance of the two amplifier designs described in Chapter 3 are measured and compared to representation samples from the recent literature. S-parameters, noise figure and the third-order intercept point (IP_3) were measured on-wafer for multiple samples. Note that, in this chapter, Design A corresponds to the design of Section 3.1, and Design B corresponds to the design of Section 3.2.

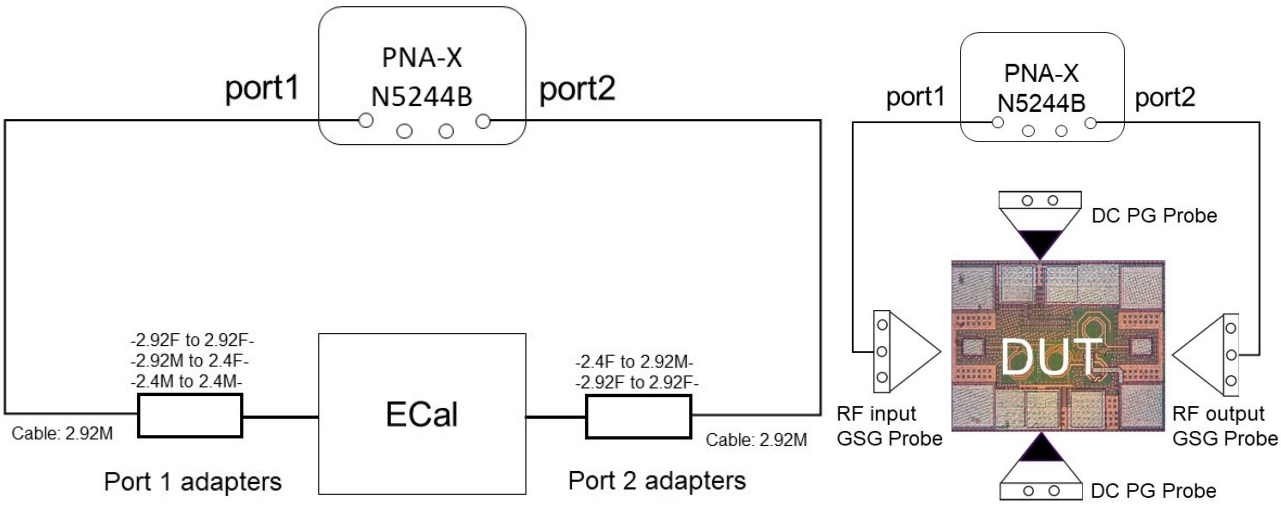


(a) Design A (b) Design B

Figure 5.1: Micrographs of two amplifier prototypes

5.1 Calibration and Measurement Setup

Two-port S-parameters and noise figure were measured using a vector network analyzer (Keysight™N5244B) as depicted in Fig. 5.2. Coaxial calibration is performed with an electronic calibration (ECal) module to bring the measurement reference plane to the end of cable and adapters (adapter is required since the cable has a 2.92-mm male connector and the Ecal has 2.4-mm a male/female connector). The measurement is performed on-wafer, and the probe and adapter losses are accounted for in the gain and noise figure measurements (details summarized in Appendix B)



(a) Calibration (b) Measurement

Figure 5.2: Two-port measurement setup

The measured gain is corrected by the difference between insertion loss of the probe and adapter at both input and output ports (port 1 and 2, respectively). For noise figure measurements, only the difference in loss (between the probe and adapter) at port 1 is included for correction, as the effect of loss at the output is insignificant due to the gain of the amplifier. The maximum measurement uncertainty is $\pm$ 0.24 dB for gain and $\pm$ 0.13 dB for noise figure.

5.2 Design A - Two-Stage Amplifier with Interstage Noise-Matching

In this section, the measurement results of the amplifier designed in Section 3.1 (Design A) are summarized. Gain, return losses, noise figure and IP_3 are detailed. Note that bandwidth may be defined by -3 dB points with respect to the peak gain, or when the input or output return losses are less than 10 dB.

5.2.1 Low-Noise Mode (Biased at 17.3 mW)

For the low-noise mode, the bias at the back-gate is set to 2 V. The measured DC power consumption is 17.3 mW. Fig. 5.3 shows the measured and simulated small-signal gains, which reveal a peak gain of approximately 20.7 dB at 22 GHz and a -3 dB bandwidth of 17.2 GHz (i.e., from 19 to 36.2 GHz). The measured bandwidth matches well with simulation. The peak gain is 1 dB below simulation, which may be due to processing variation and probe contact uncertainty. Note that, the result has a maximum uncertainty of 0.24 dB beyond 25 GHz due to adapter losses (as explained in Appendix B).

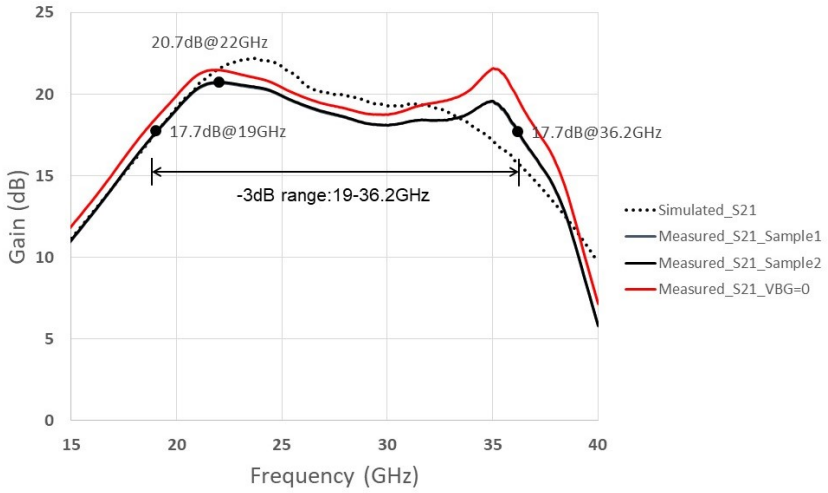


Figure 5.3: Measured vs. simulated small-signal gain of Design A biased at 17.3 mW in the low-noise mode

The effect of back-gate biasing on the gain is plotted on the red curve. When the bias at the back-gate is set to 0 V, the front-gate is biased at 0.43 V (instead of 0.25 V with

2 V back-gate biasing) to maintain the same bias currents, and a 1-dB increase is seen in the gain. This effect was not observed during the design phase, as simulations show that these two biasing conditions should have the same result in gain.

Fig. 5.4 shows the input and output return losses. Better than 10-dB return loss is observed from 22.5 to 32.2 GHz, which indicates a good 50-Ω match at the input and output. The frequency range where return losses are higher than 10 dB matches well with simulation. Note that the measured output return loss is not as good as the simulated result. This may be due to the inaccuracy within the FET model and extraction of the output capacitance (the design kit is still at its early stage), which leads to the discrepancy in output impedance at mm-wave frequencies (i.e., degradation in S_{22}). Note that, S_{22} is not considered as critical as S_{11} for monolithic systems as the amplifier will be interfacing with the next stage (i.e., frequency converter) on-chip.

Overall, the effective bandwidth is taken as 22.5 to 32.2 GHz (9.7 GHz), where the gain is within -3 dB of the peak gain and return losses are better than 10 dB.

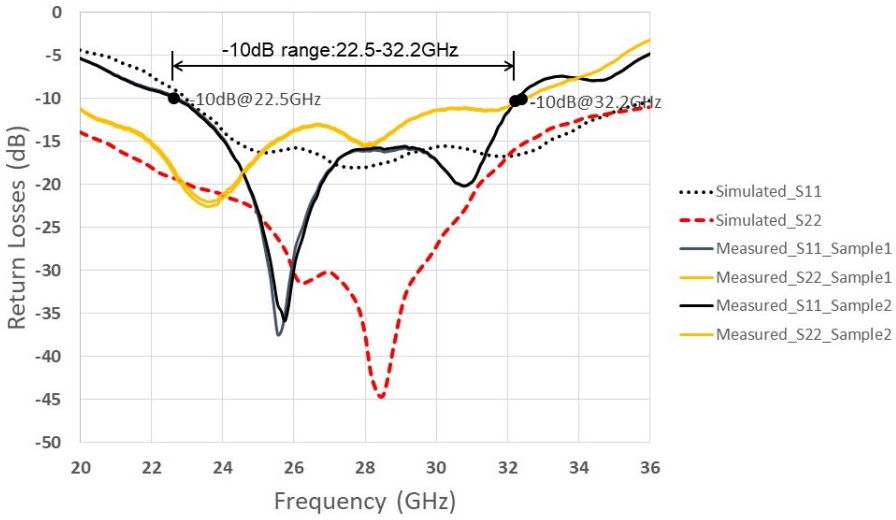


Figure 5.4: Measured vs. simulated input and output return losses of Design A biased at 17.3 mW in the low-noise mode

The noise figure measurement is depicted in Fig. 5.5. The noise figure reaches its minimum of 1.81 dB at 27.9 GHz. Simulations predict the lowest noise figure as 1.63 dB at 28 GHz. Within the defined bandwidth (22.5 to 32.2 GHz), the mean, measured noise figure is 2.1 dB with a variation of ±0.29 dB. When the bias at the back-gate is set to

0 V (red curve), the lowest noise figure is improved to 1.74 dB due to an increase in gain (as shown in Fig. 5.3). The increasing discrepancy between simulation and measurement beyond 28 GHz may be explained by two factors. Firstly, the noise matching condition at the interstage node may be affected by inaccurate modeling of the output capacitance of NFETs (which is also seen in the measurement of S_{22} above). A second reason is measurement uncertainty as frequency increases due to loss in cable adapters used in the measurement. Note that, the measured results have a maximum uncertainty of 0.13 dB beyond 25 GHz due to the adapters, as explained in Appendix B.

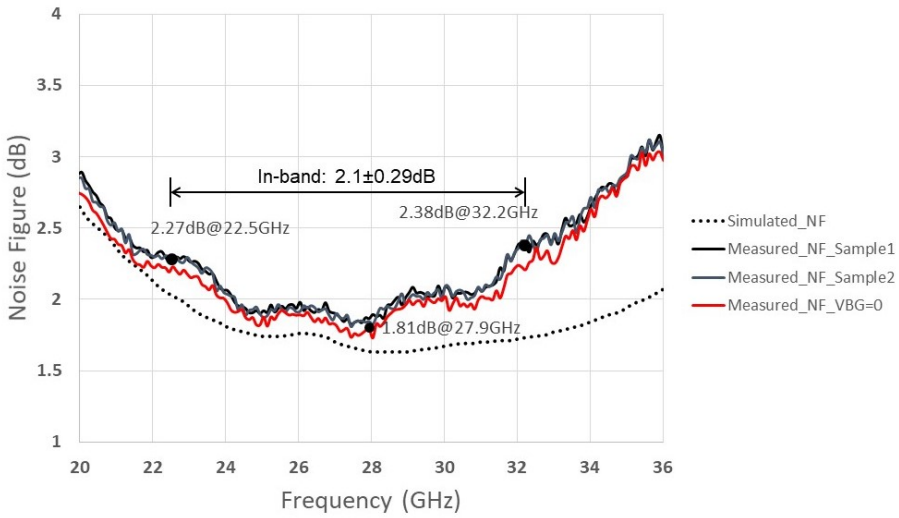


Figure 5.5: Measured vs. simulated noise figure for Design A biased at 17.3 mW in the low-noise mode

A two-tone test was conducted to measure the third-order intercept point (IP_3). The frequencies of two input tones were selected where the peak gain occurs (i.e., at 22.5 GHz) with a tone spacing of 500 kHz for maximizing resolution of the VNA. The signal power is set to -30 dBm (near small signal) to ensure the intermodulation is contributed mainly by the third order. The power of the main tone (P_{in}) and the third-order intermodulation (IM_3) products are obtained from the measured frequency spectrum. In this measurement, output powers of -53.8, -13.1, -13.3, -54.5 dBm are observed at the lower IM_3, lower tone, upper tone and upper IM_3 frequency, respectively. The IP_3 referred to the output (OIP_3) is calculated by Eq. 1.5 (i.e., OIP_3 = -13.1+$\frac{-13.1+53.8}{2}$), which yields an OIP_3 of 7.3 dBm (or IIP_3 of -13.4 dBm). Simulation predict an OIP_3 of 6 dBm at the peak-gain frequency (i.e., 22.2 dB at 23.5 GHz). The gain is 1.5 dB higher in simulation than measurement,

thus 1.5 dB has to be added to the simulated IP_3. After accounting the gain difference, the discrepancy in IP_3 is around 0.2 dB, which indicates good consistency between simulation and measurement.

5.2.2 Low-Power Mode (Biased at 5.6 mW)

In the low-power mode, the bias at the back-gate is decreased from 2 V to 0.62 V, and the DC power consumption is reduced from 17.3 mW to 5.6 mW. Fig. 5.6 shows the measured and simulated small-signal gains, which reveal a peak gain of 17.9 dB at 22.5 GHz, and a -3 dB bandwidth of 16 GHz from 19.6 to 35.6 GHz. The measured frequency range matches well with simulations. The measured gain is 1.4-dB lower, which may be due to variation in process corners and probe contacts. Note that, the result has a maximum uncertainty of 0.24 dB beyond 25 GHz due to adapter losses.

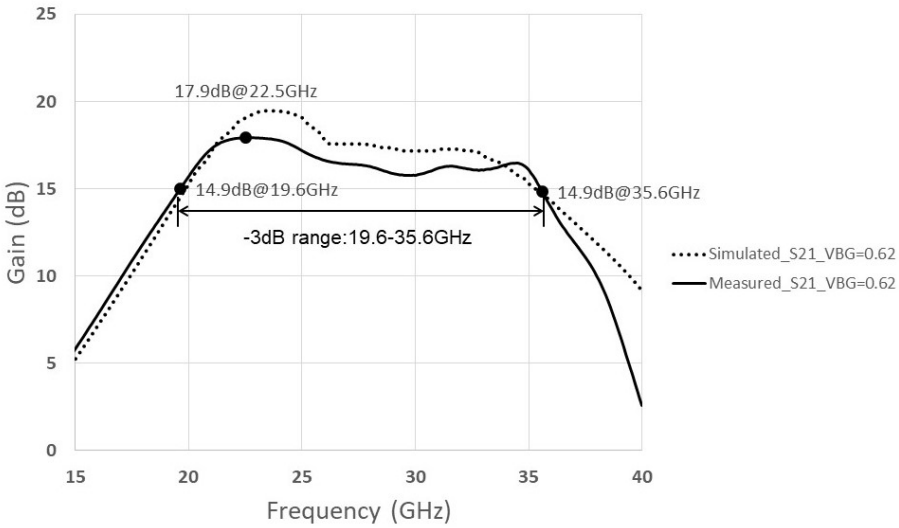


Figure 5.6: Measured vs. simulated small-signal gain of Design A biased at 5.6 mW in the low-power mode

Fig. 5.7 shows the input and output return losses. The input return loss is not affected by changes in bias at the back-gate. From 22.4 to 32 GHz, the input return loss is better than 10 dB (22.5 to 32.2 GHz when biased in the low-noise mode). The range of less than -10-dB output return loss is reduced to 21.2 to 28.5 GHz (from 22.5 to 32.2 GHz in the

low-noise mode). This may be explained by changes in the output impedance of FETs due to variation in the drain current.

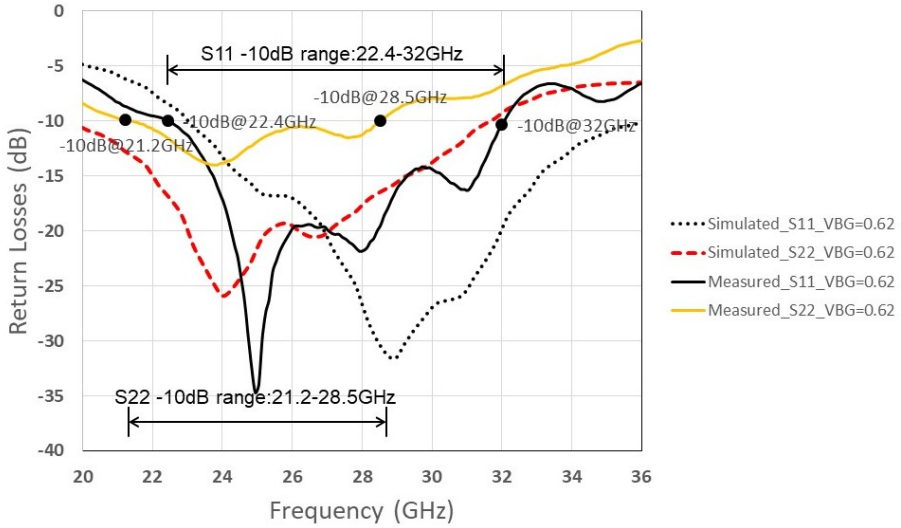


Figure 5.7: Measured vs. simulated input and output return losses of Design A biased at 5.6 mW in the low-power mode

Overall, the effective bandwidth is taken as 22.5 to 28.5 GHz (6 GHz), where the gain is within -3 dB of the peak gain and both return losses are better than 10 dB.

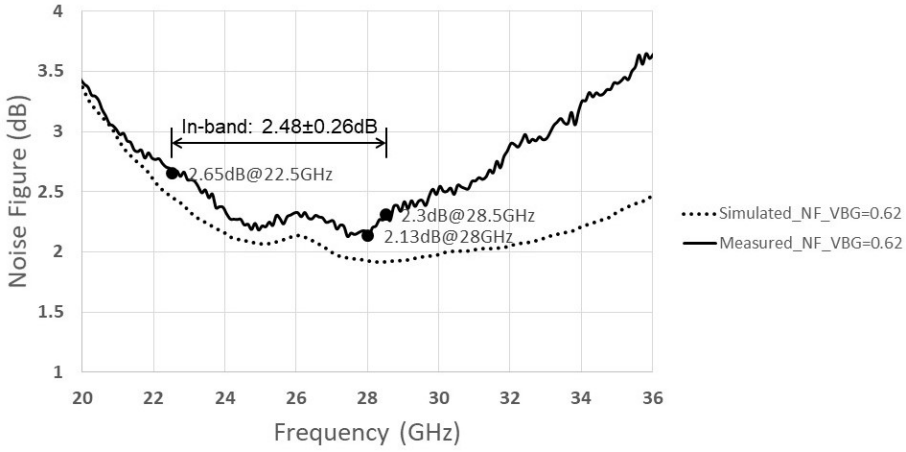


Figure 5.8: Measured vs. simulated noise figure of the two-stage amplifier biased at 5.6 mW in the low-power mode

The noise figure measurement is depicted in Fig. 5.8. The noise figure reaches its minimum of 2.13 dB at 28 GHz. Simulations show the lowest noise figure as 1.92 dB at 28 GHz. The average noise figure is 2.48 dB with ±0.26 dB fluctuation from 22.5 to 28.5 GHz. Similar to the low-noise mode, the noise figure curve begins deviating from simulations above 28 GHz, as mismatches may happen at the interstage. Note that, the result has a maximum uncertainty of 0.13 dB beyond 25 GHz due to adapter losses.

The same two-tone setup was used from measuring IP_3 in the low-power mode, as its peak gain occurs at the same frequency as in the low-noise mode (i.e., tones at 22.5 GHz with 500 kHz spacing and -30 dBm input power). Output powers of -55.4, -16.1, -16.3, -55.9 dBm are observed and OIP_3 = -55.4+$\frac{-16.1+55.4}{2}$ = 3.5 (i.e., IIP_3 of -14.5 dBm). The predicted OIP_3 is 0.8 dBm with a peak-gain of 19.5 dB at 23.5 GHz, where the gain is 1.6 dB higher in simulation than measurement. The discrepancy in IP_3 is around 1.1 dBm after accounting the gain difference. Compared to the low-noise mode, lowering the bias at the back-gate decreases IP_3 by 2.5 dB, which is expected since the power consumption decreases from 17.3 mW to 5.6 mW.

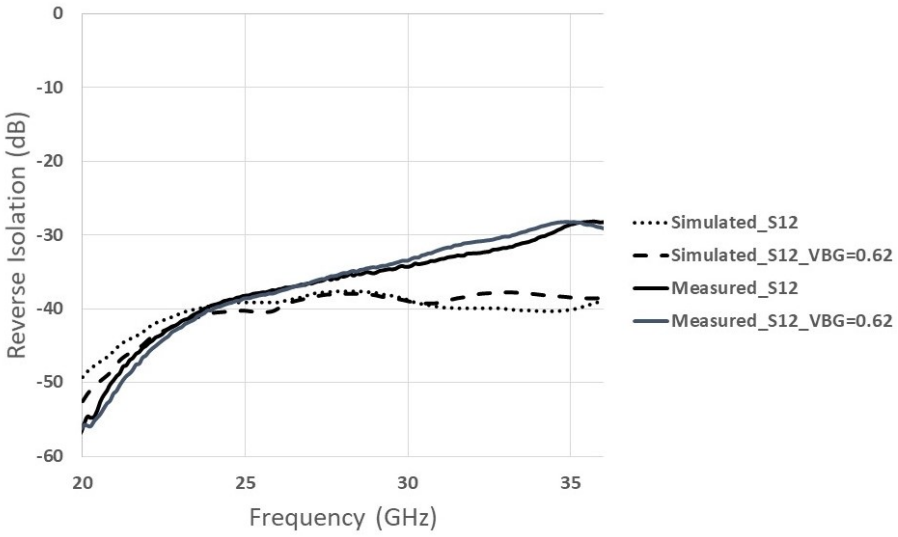


Figure 5.9: Measured vs. simulated isolation of Design A in the low-noise and low-power mode

The isolations of Design A in low-noise and low-power are plotted in Fig. 5.9. In both biasing conditions, the discrepancy between simulation and measurement in isolation becomes larger as frequency increases, which again may be due to the presence of unexpected parasitics.

5.3 Design B - Two-Stage, High-Linearity Amplifier with Interstage Noise-Matching

Design B refers to the work in Section 3.2, which contains a folded output stage to improve the linearity. During the design phase, the S-parameter (SP) and the linear simulation (AC) predict substantial differences in results for the magnitude of the gain, as shown in Fig. 5.10. The AC result has a wider bandwidth than SP, and the measurement is closer to the SP simulation. This may be due to an error either in the simulation tool or the models used in simulation.

The measured and simulated small-signal gains are shown in Fig. 5.10. The peak gain is 20.1 dB at 25 GHz and the -3 dB bandwidth for $|S_{21}|$ is 10.8 GHz (i.e., from 21.6 to 32.4 GHz). About 1.4 dB difference is seen between the measured and simulated results. The -3 dB frequency range experiences a detuning of around 10% (24-40 down to 21.6-32.4 GHz) and the shape is changed from a wideband response to a tuned response. Note that, the result has a maximum uncertainty of 0.24 dB beyond 25 GHz due to adapter losses.

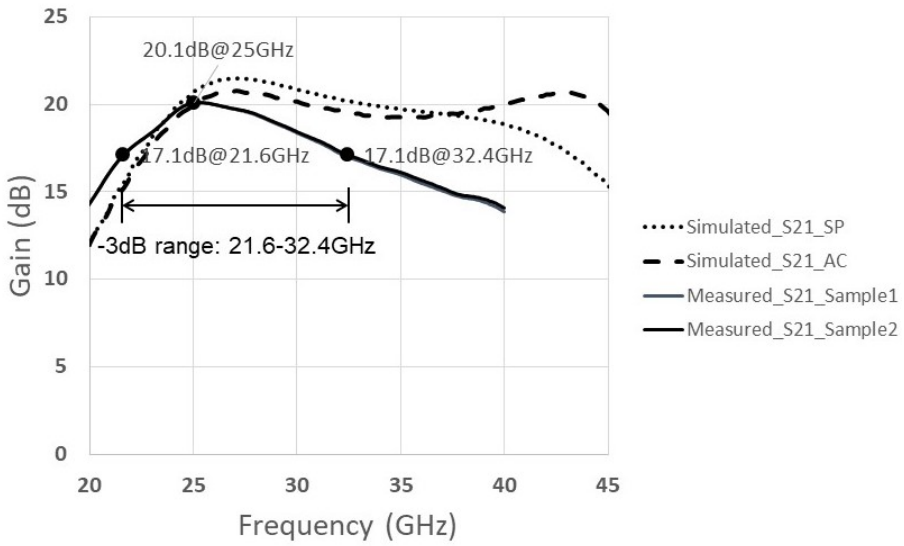


Figure 5.10: Measured vs. simulated small-signal gain of Design B

The 10% detuning may be caused by unexpected parasitics. As seen in measurements of the previous designs, good correlation between simulation and measurement are observed when the circuit contains NFETs only. Thus, the analysis of such discrepancy is focused

on the PFET. In addition to inaccurate modeling, one potential factor is the layout of inductors in the output stage of the amplifier (i.e., L_2 and L_3 in Fig. 3.26). As inductance of both inductors is low (L_2 = 40 pH and L_3 = 80 pH), they are implemented as simple metal traces without applying reduced metal filling (for spiral inductors, the reduced metal filling is required to boost the performance at mm-wave frequencies). As a result, the parasitic capacitance will affect the frequency response.

Except detuning, another possible cause is process-voltage-temperature (PVT) variations. To estimate the effect of PVT, 200 iterations of Monte-Carlo simulation are performed for each PVT condition (full corners, up to 10% voltage variation, 25, 55 and 85 °C). Fig. 5.11 shows spread of the gain, and up to 1.5-dB variation in gain is expected due to PVT variation.

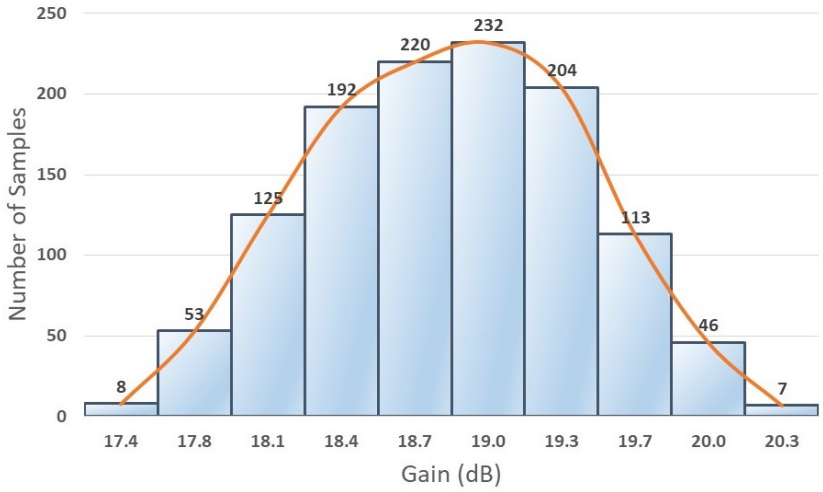


Figure 5.11: Gain distribution among various PVT conditions

The improvement on IP_3 via the folded output stage is substantial. OIP_3 of 13.4 dBm is measured (i.e., -64 and -12.35 dBm at IM_3 and fundamental tone, respectively), which indicates an increase of 6.1 dB compared to Design A (up from 7.3 dBm) in the low-noise mode, at the cost of 35 mW DC power consumption (up from 17.7 mW). Note that the measured is performed at peak-gain frequency (i.e., 25 GHz, 20.1-dB gain) with 500 kHz tone spacing. Simulations predict an OIP_3 of 15.4 dBm at the peak-gain frequency (i.e., 21.4 dB at 27 GHz), which is 3.3 dB higher than the measurement after accounting the gain difference. One potential cause is that the post-layout simplification option is used to to speed up time-domain and harmonic-balance simulations, which reduces the simulation accuracy.

Better than 10-dB return loss is observed from 23.8 to 40 GHz, which indicates a good 50-Ω match at the input and output, as seen in Fig. 5.12. The frequency range of better than 10-dB return losses matches well with prediction. Note that the correlation between measured and simulated results is better than Design A, as Design B is implemented with a later version of the design kit.

Overall, the effective bandwidth is taken as 23.8 to 32.4 GHz (8.6 GHz), where the gain is within -3 dB of the peak gain and return losses are better than 10 dB.

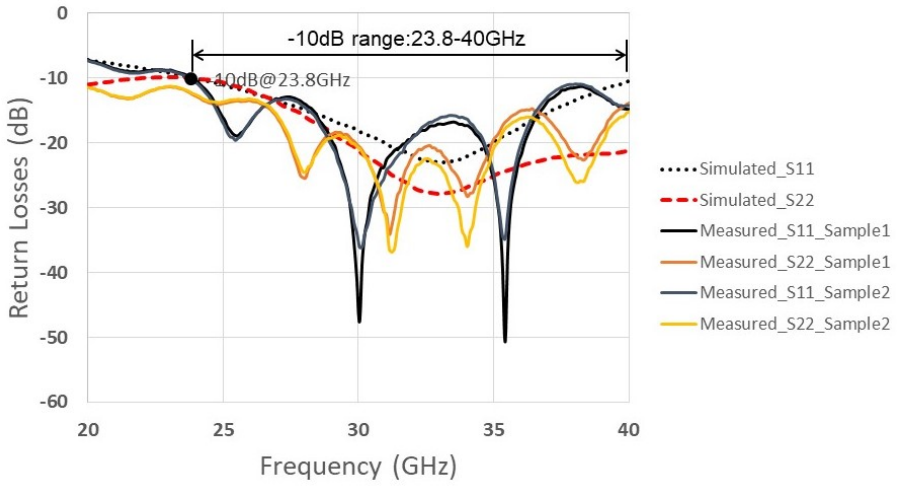


Figure 5.12: Measured vs. simulated input and output return losses of Design B

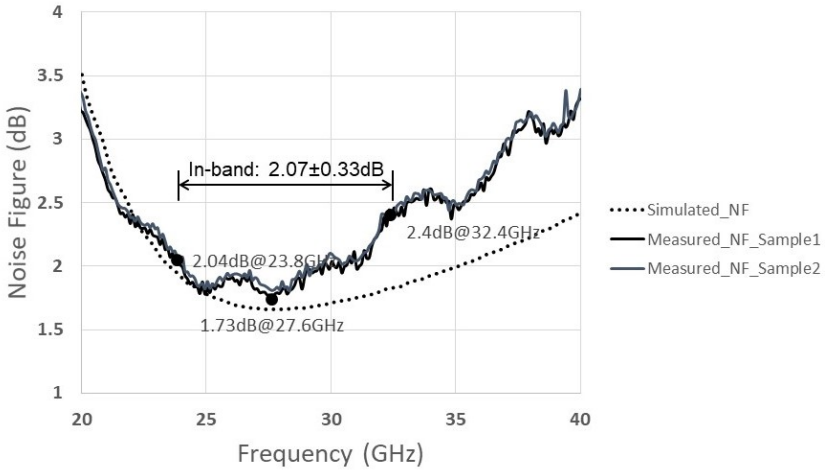


Figure 5.13: Measured vs. simulated noise figure of Design B

Simulations predict that Design A and B have similar noise figure. As depicted in Fig. 5.13, the lowest in-band noise figure is 1.73 dB at 27.6 GHz, which is improved by 0.08 dB from Design A. Within the defined bandwidth (23.8 to 32.4 GHz), the mean, measured noise figure is 2.07 dB with a variation of ±0.33 dB. The effect of narrowband interstage noise-matching is again observed at beyond 28 GHz, which leads to increasing discrepancy between the measurement and simulation. Note that, the result has a maximum uncertainty of 0.13 dB beyond 25 GHz due to adapter losses.

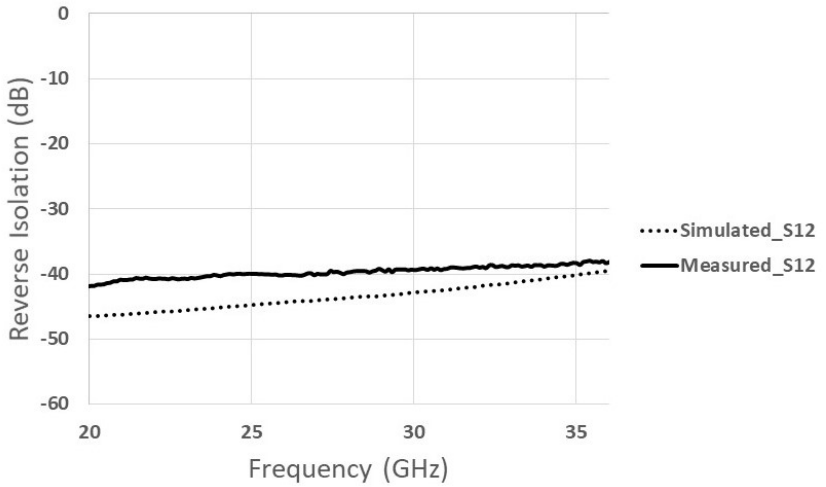


Figure 5.14: Measured vs. simulated isolation of Design B

Fig. 5.9 shows the reverse isolation of Design B. The measured result is a bit higher than the simulated result, which may be due to unexpected parasitics.

5.4 Comparison and Summary

The performance of both amplifier designs is summarized in Table 5.1. Other work reported in the recent literature at similar frequencies is added for comparison. As the designs in this thesis are fabricated in a 22-nm CMOS-SOI process, other designs fabricated in bulk CMOS and CMOS-SOI are included to compare different technologies.

The two designs in this thesis output 20-dB peak gain and have better than 10-dB return losses across an effective bandwidth of 9.7 GHz and 8.6 GHz, respectively. The lowest noise figures measured for the two amplifiers are 1.81 and 1.73 dB, respectively. With 17.3 mW power consumption, Design A achieves an IIP_3 of -13.4 dBm. With folded output stage, the IIP_3 is improved to -6.7 dBm in Design B at the cost of 35 mW of DC power. In addition, the bias at the back-gate offers some tunability between performance (gain, noise) and power consumption. Both designs occupy around 0.05 mm^2 active area, which is at least 33% lower than other candidates. Multiple samples are measured and excellent correlation is found between samples.

Compared to other work, the cascode amplifier in [32] achieves a minimum noise figure of 1.3 dB and an IIP_3 of +3.8 dBm which seems to be better than Design B. However, it has several drawbacks. To achieve the low noise figure, the design consists a single stage, which results in a peak gain of 14 dB only (thus higher IIP_3). Furthermore, its input and output are not properly matched to 50 Ω across the -3 dB $|S_{21}|$ bandwidth.

The LNA in [11] shows similar performance to the two amplifier designs in this thesis. However, it operates near 20 GHz (rather than 28 GHz) across an effective bandwidth of 4 GHz, which is about half compared to this work. The wideband amplifier in [33] shows the highest bandwidth (20.5 GHz) in Table 5.1. However, its mean noise figure of 4.2 dB is not comparable to others, and it consumes the highest amount of power (58 mW).

The LNA in [34] has digitally controlled gain. In the high-gain mode, a peak gain of 27.1 dB is achieved (i.e., higher than Design A and B) over 7.4 GHz bandwidth. However, the noise figures of Design A and B are 1.5 dB lower than [34]. In the low-gain mode, [34] has similar gain compared to Design A operating in low-power mode, but Design A only consumes 5.6 mW of DC power rather than 21.5 mW. Lastly, Design B has similar gain and IIP_3 compared to [35]. However, the bandwidth is higher (i.e., 8.6 over 2.6 GHz) and noise figure is lower (i.e., 1.73 over 2.9 dB).

Table 5.1: Summary and comparison of amplifier performance metrics

Parameter	DesignA Low-Noise	DesignA Low-Power	DesignB	[32]	[11]	[33]	[34] High-Gain	[34] Low-Gain	[35]
Eff. BW.[1](GHz)	22.5 - 32.2 9.7 GHz	22.5 - 28.5 6 GHz	23.8 - 32.4 8.6 GHz	-	20 - 24[1] 4 GHz	27 - 47.5 20.5 GHz	26 - 33.4[1] 7.4 GHz	26 - 35.3[1] 9.3 GHz	27.6 - 30.2 2.6 GHz
S_{21} (dB)	19.6 ± 1.1	16.9 ± 1	18.6 ± 1.5	12.5 ± 1.5	18 ± 1.5	18.5 ± 1.5	25.6 ± 1.5	17.2 ± 1.2[2]	18.5 ± 1.5
Peak S_{21} (dB) @ Freq.	20.7 @ 22.5 GHz	17.9 @ 22.5 GHz	20.1 @ 25 GHz	14 @ 24 GHz	19.5 @ 20 GHz	20 @ 28 GHz	27.1 @ 27.1 GHz	18.4 @ 27.8 GHz	20 @ 28.5 GHz
-3 dB BW. (GHz)	19 - 36.2 17.2 GHz	19.6 - 35.6 16 GHz	21.6 - 32.4 10.8 GHz	24 - 31 7 GHz	16 - 24 8 GHz	24 - 47.5 23.5 GHz	26 - 33.4[1] 7.4 GHz	26 - 35.3[1] 9.3 GHz	27.6 - 30.2 2.6 GHz
NF (dB)	2.1 ± 0.29	2.48 ± 0.26	2.07 ± 0.33	1.45 ± 0.15	2.3 ± 0.2[2]	4.85 ± 0.65	3.8 ± 0.5	3.9 ± 0.5	3.55 ± 0.65
Min. NF (dB) @ Freq.	1.81 @ 27.9 GHz	2.13 @ 28 GHz	1.73 @ 27.6 GHz	1.3 27.5 GHz[1]	2 19 GHz	4.2 30 GHz	3.3 31 GHz	3.4 31 GHz	2.9 28 GHz
-10 dB S_{11} range (GHz)	22.5 - 32.2	22.4 - 32	23.8 - 40	-	19 - 30[1]	27 - 48	26 - 40	26 - 40	28 - 34[1]
-10 dB S_{22} range (GHz)	20 - 32.2	21.2 - 28.5	20 - 40	-	20 - 24[1]	20 - 67	26 - 40	26 - 40	27 - 34[1]
$P_{-1dB,out}$ (dBm)[3] @ Freq.	-2.3 @ 22.5 GHz	-6.1 @ 22.5 GHz	+3.8 @ 25 GHz	+8.2 @ 28 GHz	+1 @ 20 GHz	+1 @ 28 GHz	+5.5 @ 27.1 GHz	+5 @ 27.8 GHz	+2 @ N/A
IIP_3 (dBm) @ Freq.; Δf	-13.4 @ 22.5G (500 kHz)	-14.4 @ 22.5G (500 kHz)	-6.7 @ 25G (500 kHz)	+3.8 @ 28G (100 MHz)	-8 @ 20G (N/A)	-9.4[3]@ 28G N/A	-12.6 @ 27.1G (N/A)	-4.9 @ 27.8G (N/A)	-7.5 @ N/A (N/A)
P_{DC} (mW) @ V_{dd}[4](V)	17.3 @ 1.05	5.6 @ 1.05	35 @ 0.8	15 @ 1.5	32.5 @ 1.5	58 @ 1.1	31.4 @ 1.1	21.5 @ 1.1	16.3 @ 1
Act. Area (mm^2)	0.05	0.05	0.056	0.3	0.15	0.2	0.26	0.26	0.67
Process	22-nm CMOS-SOI	22-nm CMOS-SOI	22-nm CMOS-SOI	45-nm CMOS-SOI High-ρ sub.	45-nm CMOS-SOI	45-nm CMOS-SOI High-ρ sub.	40-nm Bulk CMOS	40-nm Bulk CMOS	90-nm Bulk CMOS

[1] Effective bandwidth is defined as the frequency range where gain is within -3 dB of the peak gain, and both input and output return losses are better than 10 dB

[2] Estimated from plots

[3] P_{-1dB} is mathematically related to IP_3 (lower by 9.6 dB) if not provided

[4] The highest voltage is picked if multiple voltage levels occurs

Chapter 6

Conclusions and Future Work

The chapter concludes this thesis and is followed by thoughts raised during the design and implementation phases of the amplifier.

6.1 Conclusions

The demand for high-speed wireless communications drives the research & development of radio systems moving towards mm-wave frequencies for higher bandwidth. The increase in operating frequency introduces challenges such as insufficient gain for single-stage amplifiers and excessive noise factor for multistage amplifiers. The 22-nm CMOS-SOI technology demonstrates its excellent potential to be utilized as the building block for monolithic solutions. Besides the benefits such as an increase in transition frequency and a decrease in device noise factor from CMOS scaling, features such as flip-well structure and SOI substrate enable a dynamic control of circuits being high-performance low-power, which makes the process suitable for various applications.

This thesis has explored the use of the 22-nm CMOS-SOI technology for designing a multistage amplifier which targets on high-bandwidth, low-noise and high-linearity around 28 GHz. Emphasis is placed on minimizing the noise factor of a cascaded topology by realizing noise matching at the second-stage input via a transformer-coupled interstage. Chapter two covers the fundamentals of amplifier design for minimum noise and examines a few design examples at radio and mm-wave frequencies. Various inductive peaking techniques for extending the bandwidth are also discussed in this chapter. Chapter three begins with a noise analysis of cascaded blocks with transformer-coupled interstages. Then, an analysis

is accomplished to illustrate the gain of interstage noise match over conjugate match. A design methodology is summarized and two proposed amplifier topologies are presented with simulation results. Chapter four explained the layout philosophy for minimizing noise and maintaining sufficient electromigration reliability. An analysis of transformer configurations is performed with the metal stack-up provided by the process, and a transformer layout is depicted.

Overall, Design A achieves a peak gain of 20.7 dB and better than 10-dB input and output return losses within a frequency range of 22.5 to 32.2 GHz. The lowest noise figure of 1.81 dB is achieved within this frequency range. Output IP_3 of +7.3 dBm is achieved for 17.3-mW of DC power consumption. When the bias at the back-gate is lowered from 2 V to 0.62 V, the power is decreased to 5.6 mW and the peak gain drops down to 17.9 dB. Noise figure increases from 1.81 to 2.13 dB and output IP_3 drops to +3.5 dBm.

Design B implements a folded-output stage, which improves the output IP_3 to +13.4 dBm at the cost of 35 mW total power consumption. Shunt peaking is applied to the input cascode amplifier, which helps lower the minimum noise figure to 1.73 dB. The peak gain of the second design is 20.1 dB within a frequency range of 23.8 to 32.4 GHz. Both designs occupy about 0.05 mm^2 active area.

6.2 Future Work

Several challenges and interesting ideas for future work raised during the design, implementation and measurement phases of this thesis are discussed in this section.

6.2.1 Device Modeling

The 22-nm CMOS-SOI PDK, the Calibre® xACT™parasitic extraction tools were updated several times during the design phase of this project, and the simulation results were affected by these changes. Furthermore, the time required for post-layout simulation is significantly slowed by the number of circuit elements in the parasitic network generated by the extraction process. To speed up the simulation, the option of layout simplification was used, which degrades the accuracy. To overcome this problem, one approach is to build a lumped element network which models parasitic effects by electromagnetic (EM) analysis. In [36], a scalable, lumped-element model is developed for interconnects around a 0.25-μm GaAs pHEMT with manifold gate and drain. The idea may also be applied to the 22-nm CMOS-SOI process for different layout styles (i.e., grid instead of manifold).

6.2.2 Ground Impedance

At mm-wave frequencies, the impedance in the current return path (e.g., ground impedance) becomes significant. At the nodes which excessive impedance may cause the frequency response and matching condition to be detuned (e.g. gate and source of M_1 in Fig. 2.3), a low impedance current loop should be ensured during the design phase. A common approach during layout is to shunt multiple layers of metal together to increase the effective thickness of the path. Moreover, the incident and return paths may be placed physically close to each other so that the mutual inductance of the interconnects is minimized. Both approaches were implemented during the design phase. To further improve the immunity of the circuit to such parasitic impedances, a fully-differential topology may be considered. A fully-differential topology provides a virtual ground along a line of symmetry in the layout, which significantly reduces the current loop are and parasitic impedances, but at the cost of extra power consumption.

6.2.3 The Omni-Amplifier

Friis' equation (Eq. 1.4) states that the noise factor of a cascaded amplifier is dominated by the first few stages. The third-order intercept point (TOI) of a multistage amplifier may also be expressed by an equation in a similar form, and the TOI is dominated by the last few stages [22]. As the TOI is proportional to the output stage power consumption [4], an amplifier may be designed which targets low noise, sufficient gain and high linearity/output power at the same time (i.e., a high-linearity, low-noise amplifier). Technically, an "omni-amplifier" may be implemented by two stages or even with a single stage. Assuming that the gain of a single-stage amplifier is insufficient and isolation is finite, Fig. 6.1 depicts the concept at block diagram level with three active stages.

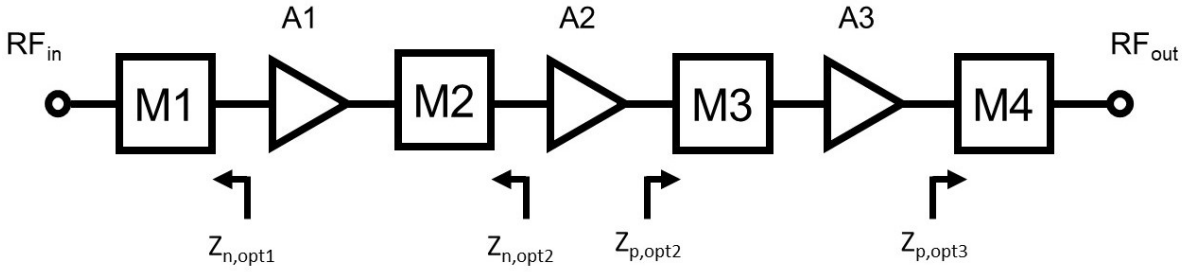


Figure 6.1: A three-stage omni-amplifier at block level

As shown in the figure above, $M_1 - M_4$ are matching networks and $A_1 - A_3$ represent the three active stages. The source and the output impedances of A_1 are transformed to the optimum source impedance for minimum noise factor ($Z_{n,opt1}$ and $Z_{n,opt2}$, respectively) via M_1 and M_2. This minimizes the noise factor of the amplifier. The load and input impedances of A_3 are transformed to the optimum load impedance for maximum power ($Z_{p,opt2}$ and $Z_{p,opt3}$, respectively) via M_3 and M_4. This maximizes the output power (i.e., linearity) of the amplifier. The sizes of transistors used in A_1 and A_3 may be optimized by source and load pull analyses, while the size of A_2 may be selected to achieve the desired gain specification. As a result, a low-noise, high-linearity amplifier may be realized. Techniques such as noise-cancelling [37] and linearization [38] are not required. These methods often require an auxiliary amplifier/path where the auxiliary circuits performance is sensitive to parasitics of the physical layout at mm-wave frequencies.

References

[1] "Communication monitoring report 2017." https://crtc.gc.ca/eng/publications/reports/PolicyMonitoring/2017/cmr2017.pdf. (Accessed on 08/08/2018).

[2] "22nm FD-SOI technology." https://www.globalfoundries.com/sites/default/files/product-briefs/22fdx-product-brief_0.pdf. (Accessed on 08/08/2018).

[3] S. B. Chiah and X. Zhou, "Floating-body effect in partially/dynamically/fully depleted DG/SOI MOSFETs based on unified regional modeling of surface and body potentials," *IEEE Transactions on Electron Devices*, vol. 61, pp. 333–341, Feb 2014.

[4] B. Razavi, *RF Microelectronics*. Prentice Hall, 2nd ed., 2011.

[5] J. Proakis and M. Salehi, *Fundamentals of Communication Systems*. Pearson, 2nd ed., 2013.

[6] B. Razavi, *Design of Analog CMOS Integrated Circuits*. McGraw-Hill Education, 1st ed., 2000.

[7] X. Guan, H. Hashemi, and A. Hajimiri, "A fully integrated 24-GHz eight-element phased-array receiver in silicon," *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 2311–2320, November 2004.

[8] M. Sayginer and G. M. Rebeiz, "A W-band LNA/phase shifter with 5-db NF and 24-mw power consumption in 32-nm CMOS SOI," *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, pp. 1973–1982, February 2018.

[9] J. Dunworth and V. A. et al., "A 28GHz bulk-CMOS dual-polarization phased-array transceiver with 24 channels for 5G user and basestation equipment," in *2018 IEEE International Solid-State Circuits Conference*, Feb 2018.

[10] H.-K. Lim and J. Fossum, "Threshold voltage of thin-film Silicon-on-insulator (SOI) MOSFET's," *IEEE Transactions on Electron Devices*, vol. 30, pp. 1244–1251, Oct 1983.

[11] T. Kanar and G. M. Rebeiz, "A 16-24 GHz CMOS SOI LNA with 2.2 dB mean noise figure," in *2013 IEEE Compound Semiconductor Integrated Circuit Symposium*, Oct 2013.

[12] O. Inac, M. Uzunkol, and G. M. Rebeiz, "45-nm CMOS SOI technology characterization for millimeter-wave applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, pp. 1301–1311, June 2014.

[13] B.-J. Huang, K.-Y. Lin, and H. Wang, "Millimeter-wave low power and miniature CMOS multicascode low-noise amplifiers with noise reduction topology," *IEEE Transactions on Microwave Theory and Techniques*, vol. 57, pp. 3049–3059, November 2009.

[14] H.-H. Hsieh, P.-Y. Wu, C.-P. Jou, F.-L. Hsueh, and G.-W. Huang, "60GHz high-gain low-noise amplifiers with a common-gate inductive feedback in 65nm cmos," in *2011 IEEE Radio Frequency Integrated Circuits Symposium*, June 2011.

[15] J. Cao, Z. Li, J. Tian, H. Liu, and Q. Li, "A 24-48-GHz low power low noise amplifier using gain peaking techniques," in *2016 IEEE International Conference on Microwave and Millimeter Wave Technology (ICMMT)*, June 2016.

[16] H. Fukui, "The noise performance of microwave transistors," *IEEE Transactions on Electron Devices*, vol. ED-13, pp. 329–341, March 1966.

[17] S. Voinigescu, *High-Frequency Integrated Circuits*. Cambridge University Press, 4th ed., 2013.

[18] A. S. Sedra and K. C. Smith, *Microelectronic Circuits*. Oxford University Press, 7th ed., 2014.

[19] D. Cassan and J. Long, "A 1-V transformer-feedback low-noise amplifier for 5-GHz wireless LAN in 0.18-um CMOS," *IEEE Journal of Solid-State Circuits*, vol. 38, pp. 427–435, Mar 2003.

[20] M. T. Reiha and J. R. Long, "A 1.2 V reactive-feedback 3.1-10.6 GHz low-noise amplifier in 0.13 um CMOS," *IEEE Journal of Solid-State Circuits*, vol. 42, pp. 1023–1033, May 2007.

[21] S. Guo, T. Xi, P. Gui, D. Huang, Y. Fan, and M. Morgan, "A transformer feedback gm-boosting technique for gain improvement and noise reduction in mm-wave cascode LNAs," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, pp. 2080–2090, July 2016.

[22] G. D. Vendelin, A. M. Pavio, and U. L. Rohde, *Microwave Circuit Design Using Linear and Nonlinear Techniques*. Wiley, 1st ed., 1990.

[23] H.-T. Dabag, B. Hanafi, F. Golcuk, A. Agah, J. F. Buckwalter, and P. M. Asbeck, "Analysis and design of stacked-FET millimeter-wave power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, pp. 1543–1556, April 2013.

[24] J. Kim and J. F. Buckwalter, "Bandwidth enhancement with low group-delay variation for a 40-Gb/s transimpedance amplifier," *IEEE Transactions on Circuits and Systems-I: Regular Papers*, vol. 57, pp. 1964–1972, August 2010.

[25] A. Hokazono, S. Balasubramanian, K. Ishimaru, H. Ishiuchi, C. Hu, and T.-J. K. Liu, "Forward body biasing as a bulk-Si CMOS technology scaling strategy," *IEEE Transactions on Electron Devices*, vol. 55, pp. 2657–2664, September 2008.

[26] H.-S. Wong, M. H. White, T. J. Krutsick, and R. V. Booth, "Modeling of transconductance degradation and extraction of threshold voltage in thin oxide MOSFET's," *Solid-State Electronics*, vol. 30, pp. 953–968, February 1987.

[27] A. Bazigos, M. Bucher, J. Assenmacher, S. Decker, W. Grabinski, and Y. Papananos, "An adjusted constant-current method to determine saturated and linear mode threshold voltage of mosfets," *IEEE Transactions on Electron Devices*, vol. 58, pp. 3751–3758, August 2011.

[28] V. Jain and M. J. W. Rodwell, "Transconductance degradation in near-THz InP double-heterojuncation bipolar transistors," *IEEE Electron Device Letters*, vol. 32, pp. 1068–1070, June 2011.

[29] J. R. Long, "Monolithic transformer for silicon RF IC design," *IEEE Journal of Solid-State Circuits*, vol. 35, pp. 1368–1381, September 2000.

[30] H. Shakoor, "Fully integrated 60 GHz power amplifiers in 45 nm SOI CMOS," Master's thesis, University of Waterloo, September 2016.

[31] J. Lienig, "Electromigration and its impact on physical design in future technologies," in *Proceedings of the 2013 ACM International Symposium on Physical Design*, March 2013.

[32] C. Li, O. El-Aassar, A. Kumar, M. Boenke, and G. M. Rebeiz, "LNA design with CMOS SOI process-1.4db NF K/Ka band LNA," in *2018 IEEE/MTT-S International Microwave Symposium*, June 2018.

[33] V. Chauhan and B. Floyd, "A 24–44 GHz UWB LNA for 5G cellular frequency bands," in *2018 11th Global Symposium on Millimeter Waves*, May 2018.

[34] M. Elkholy, S. Shakib, J. Dunworth, V. Aparin, and K. Entesari, "A wideband variable gain LNA with high OIP3 for 5G using 40-nm bulk CMOS," *IEEE Microwave and Wireless Components Letters*, vol. 28, pp. 64–66, December 2017.

[35] E. Adabi, B. Heydari, M. Bohsali, and A. M. Niknejad, "30 GHz CMOS low noise amplifier," in *2007 IEEE Radio Frequency Integrated Circuits Symposium*, June 2007.

[36] D. Resca, A. Raffo, A. Santarelli, G. Vannini, and F. Filicori, "Scalable equivalent circuit FET model for MMIC design identified through FW-EM analyses," *IEEE Transactions on Microwave Theory and Techniques*, vol. 57, pp. 245–253, February 2009.

[37] S. C. Blaakmeer, E. A. M. Klumperink, D. M. W. Leenaerts, and B. Nauta, "Wideband balun-LNA with simultaneous output balancing, noise-canceling and distortion-canceling," *IEEE Journal of Solid-State Circuits*, vol. 43, pp. 1341–1350, June 2008.

[38] H. Zhang and E. Sánchez-Sinencio, "Linearization techniques for CMOS low noise amplifiers: A tutorial," *IEEE Transactions on Circuits and Systems - I: Regular Papers*, vol. 58, pp. 22–36, January 2011.

[39] H. Friis, "Noise figures of radio receivers," in *Proceedings of the IRE*, vol. 32, pp. 419–422, July 1944.

[40] D. M. Pozar, *Microwave Engineering*. Wiley, 4th ed., 2011.

[41] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*. Cambridge University Press, 2nd ed., 2003.

[42] W. Zhuo, X. Li, S. Shekhar, S. H. K. Embabi, J. P. de Gyvez, D. J. Allstot, and E. Sanchez-Sinencio, "A capacitor cross-coupled common-gate low-noise amplifier," *IEEE Transactions on Circuits and Systems-II: Express Briefs*, vol. 52, pp. 875–879, December 2005.

[43] X. Fan, H. Zhang, and E. Sanchez-Sinencio, "A noise reduction and linearity improvement technique for a differential cascode LNA," *IEEE Journal of Solid-State Circuits*, vol. 43, pp. 588–599, March 2008.

[44] D. Boudinet, G. L. Blevennec, C. Serbutoviez, J.-M. Verilhac, H. Yan, and G. Horowitz, "Contact resistance and threshold voltage extraction in n-channel organic thin film transistors on plastic substrates," *Journal of Applied Physics*, vol. 105, April 2009.

APPENDICES

Appendix A

μ-Stability Factor

For a two-port network, the μ-parameter defined as

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

$$|\Delta| = |S_{11}S_{22} - S_{12}S_{21}|, \tag{A.2}$$

which is utilized as it not only performs the unconditional-stability test (i.e., $\mu > 0$ indicates the network is unconditionally stable), but also can be used to compare stability across multiple systems (e.g., higher μ corresponds to better stability) [40]. Assuming perfect return loss (i.e., $S_{11} = S_{22} = 0$), high isolation (i.e., low $|S_{12}|$) improves stability.

For multistage amplifiers, the system stability cannot be concluded from a simple two-port test. The μ-stability test has to be conducted at all possible combinations of consecutive stages and interstages to ensure that the system is stable (i.e., for a two-stage amplifier with a passive interstage, the μ-stability test has to be taken individually for "stage 1" and "stage 2", as well as for "stage 1+interstage", "interstage+stage 2" and "stage 1+interstage+stage 2").

Appendix B

Correction of Measurement Results through Calibration

Fig. B.1 shows the measured losses of probes and adapters used to test the amplifiers developed in this work, as well as the correction applied to the gain and noise figure data. The insertion loss of the probes is provided by the manufacturer. The losses of cable adapters are measured after performing a transmission response calibration.

After performing calibration with the Ecal module, the reference plane for the vector-network analyzer (VNA) measurement is moved to the end of adapters. When probing, the on-wafer measurement is done with probes in place of adapters. Therefore, a correction to the measured data requires subtracting the insertion loss of the probes and then adding the loss of adapters to the results.

For gain measurements, losses at both ports have to be included for correction. For noise figure measurements, only the loss at the input port (port 1) was accounted for, as the effect of loss at the output port on the amplifier noise factor is insignificant, as its noise is being suppressed by the amplifier gain. As shown in Fig. B.1, the adapter losses vary more with increasing frequency above 20 GHz. A linear fit is applied to model its loss (i.e., the amount of $|S_{21}|$ correction), which introduces a maximum measurement uncertainty of $\pm$ 0.24 dB for gain and $\pm$ 0.13 dB for the noise figure.

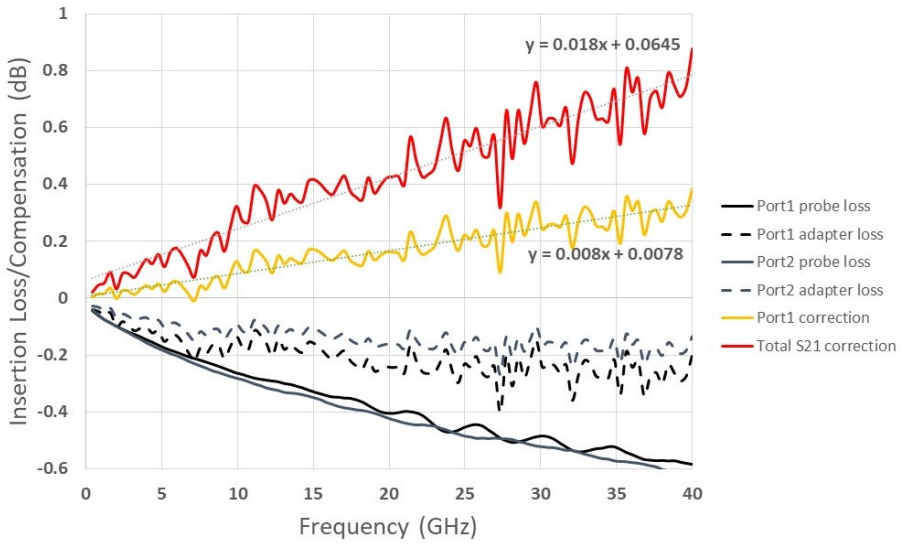


Figure B.1: Insertion loss of probes & adapters and the amount of correction