Chapter 2

Low-Noise Amplifiers in CMOS Wireless Receivers

2.1 Introduction

This chapter aims to welcome the reader to the world of low-noise amplification in wireless receivers. The most important RF concepts are introduced in Section 2.2. These concepts include the quality factor of reactive elements, different types of matching, power gain and distortion. Topics that will return and gain significance in various discussions further on.

Section 2.3 introduces the device models that will be used in hand calculations. Hand calculations will be applied throughout the text to give intuitive insight in the behavior of different circuits and circuit aspects. The models have been fit in advance to numerical simulation results. Also an extended MOS model is introduced that has been used in conjunction with the numerical simulators. The most common noise sources in CMOS IC's are discussed together with their physical origin in Section 2.4. In Section 2.5, the LNA is described in its function and functionality within the receiver chain. The coherence and mutual dependence of the LNA with the other receiver blocks is investigated. Based on that, the main design criteria and performance requirements are derived.

To conclude this chapter, the most common LNA topologies in CMOS are classified and introduced with a simple —but not irrelevant— performance model in Section 2.6. Already a swift comparison can be made. Some specific and interesting designs, published in open literature but falling beyond the above classification are clarified briefly.

2.2 Some Important RF Concepts

2.2.1 Quality Factor of Reactive Elements and Series-Parallel Transformation

A few concepts that will reoccur often are the resonance, quality factor and series-parallel transformation of reactive elements. For a purely reactive element the current through the element is

90 degrees out of phase with the voltage over it. Hence no power is consumed in the element. Naturally we are talking about inductors and capacitors where the currents are respectively lagging and leading the voltage by 90 degrees. In real life however a purely reactive element does not exist and some power dissipation is always present. Moreover if there is power dissipation, there is a resistor and resistors give rise to thermal noise whereas reactive elements are completely noiseless. Consequently a means is needed to describe the '*purity*' of a reactive element.

This means is known as the quality factor Q of the reactive element. It is defined by:

$$Q \triangleq \frac{\text{average reactive power}}{\text{average power dissipated}}. \tag{2.1}$$

For a simple inductor or capacitor with a series resistor R_s this expression becomes

$$Q_L = \frac{\omega L}{R_s} \qquad \text{and} \qquad Q_C = \frac{1}{\omega C R_s} \tag{2.2}$$

respectively. This can be rewritten in one formula:

$$Q = \frac{X_s}{R_s}, \tag{2.3}$$

where X_s is the reactance of either inductor or capacitor at the given frequency. For an inductor or capacitor with a parallel resistor R_p, the quality factor is found as

$$Q = \frac{R_p}{B_p}, \tag{2.4}$$

where B_p is the susceptance of the inductor or capacitor at the given frequency.

A quality factor can also be constructed for a resonant RLC network. Consider a series RLC tank. The tank is characterized by its resonance frequency

$$\omega_r = \frac{1}{\sqrt{LC}}, \tag{2.5}$$

and by its quality factor

$$Q = \frac{\omega_r L}{R} = \frac{1}{\omega_r C R}. \tag{2.6}$$

This means that the Q defined by (2.2) is equivalent to the Q of a series RLC tank with resonant frequency $\omega = \omega_r$. Due to the series-parallel duality this equivalence applies also for a parallel RLC tank and equation (2.4).

For a simple RLC tank, the Q-factor has yet another meaning. Consider the impedance of a parallel tank. The quality factor of a RLC tank is related to the sharpness of the impedance peak, or mathematically:

$$Q = \frac{2\pi \mathrm{BW}}{\omega_r}, \tag{2.7}$$

where BW is the total (left and right) -3 dB bandwidth of the impedance magnitude centered around ω_r. Again, by duality (2.7) is also valid for a series tank but one needs to take the admittance bandwidth.

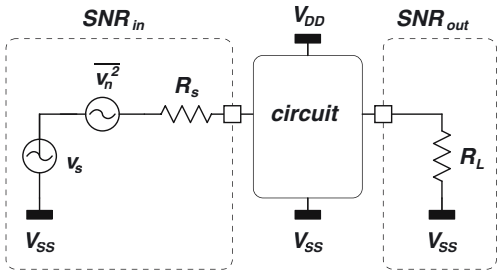


Figure 2.1: *Input and output Signal-to-Noise Ratio.*

2.2.2 SNR and Noise Figure

The SNR or Signal to Noise Ratio gives a measure for the purity of a signal. The definition is quite simple:

$$\mathrm{SNR} = \frac{\text{Available Signal Power}}{\text{Available Noise Power in Signal Bandwidth}}. \tag{2.8}$$

For instance the SNR_{in} of the signal source represented in Fig. 2.1 is

$$\mathrm{SNR}_{in} = \frac{P_{av,s}}{P_{av,n}} = \frac{\frac{v_s^2}{4R_s}}{kT\Delta f}, \tag{2.9}$$

where Δf is the signal bandwidth. Rewriting equation (2.9) as

$$\mathrm{SNR}_{in} = \frac{v_s^2}{4kTR_s\Delta f} \tag{2.10}$$

shows that it doesn't matter whether the ratio of squared voltages power or squared current is taken since both noise and signal have the same conversion factor, determined by the respective node impedance.

An ideal amplifying block operating on the signal will amplify both signal and noise equally and will not alter the SNR. However, any real-life —non-ideal— block will decrease the SNR since the block will add some noise to the signal. Mathematically this is expressed by the noise factor of the block:

$$F = \frac{\mathrm{SNR}_{in}}{\mathrm{SNR}_{out}} = \frac{\frac{P_{av,s}}{P_{av,n}}}{\frac{G\cdot P_{av,s}}{G\cdot P_{av,n}+G\cdot P_{n,eq}}}, \tag{2.11}$$

where $G \cdot P_{n,eq}$ is the excess noise power at the output and $P_{n,eq}$ is this power referred to the input. This can be simplified to

$$F = \frac{P_{av,n} + P_{n,eq}}{P_{av,n}}. \tag{2.12}$$

This means the noise factor is the total equivalent input noise power divided by the noise power of the source. Or equivalently, the noise factor is the total output noise divided by the output

noise resulting solely from the noise power of the input source. The noise figure is used much more often than the noise factor. It is related to the noise factor according to

$$\mathrm{NF} = 10\log(F). \tag{2.13}$$

Since F can be any number between 1 and ∞, NF is bounded by 0 and ∞. Noise figures lower than 0 should arouse serious suspicion since any sort of selective noise absorber has yet to be invented!

Now consider the specific case of the low noise amplifier. The LNA is usually driven by a 50 Ω source which can be either the impedance of the receive antenna or the output impedance of a band selecting SAW-filter. Consider the first case. The input SNR is given by

$$\mathrm{SNR}_{in} = \frac{v_s^2}{4kT_{\mathrm{eff}}R_s\Delta f} \tag{2.14}$$

(2.10) where $R_s = 50\,\Omega$ en T_{eff} is the effective noise temperature of the antenna. For the common case where the radiation resistance far exceeds the resistive losses in the antenna leads, T_{eff} is the average noise temperature seen by the antenna. It can be described by

$$T_{\mathrm{eff}} = \int_0^{4\pi} T(\Psi)G_A(\Psi)d\Psi, \tag{2.15}$$

where Ψ is the solid angle expressed in steradians, $T(\Psi)$ is the temperature at solid angle Ψ and $G_A(\Psi)$ is the antenna gain for solid angle Ψ. This temperature is largely dependent on the *view* of the antenna. For a GPS receiver for instance, T_{eff} will be very low at night looking into the sky with a temperature of only a few tens of Kelvin depending on the quality of the antenna. However in daylight, looking at the sun, the effective noise temperature will be much higher. In order to have a fixed noise factor for the LNA independent of the noise temperature of the antenna, F_{LNA} is defined with a fixed source noise temperature equal to the physical room temperature:

$$F_{LNA} = \frac{kT_r\Delta f + P_{n,eq}}{kT_r\Delta f} = 1 + \frac{P_{n,eq}}{kT_r\Delta f}. \tag{2.16}$$

Since most noise sources are proportional to the physical temperature, equation (2.16) shows that F_{LNA} should be independent of the actual room temperature during the measurements.

2.2.3 Impedance Matching, Power Matching, Noise Matching

Impedance matching is a term which is used frequently in the area of transmission lines. A transmission line is characterized by a characteristic impedance Z_c. Suppose the line is terminated with an impedance Z. A voltage wave V^+ travelling along the line will be partially reflected at the end of the line depending on the termination impedance. The reflected voltage V^- is given by

$$V^- = \Gamma V^+ \tag{2.17}$$

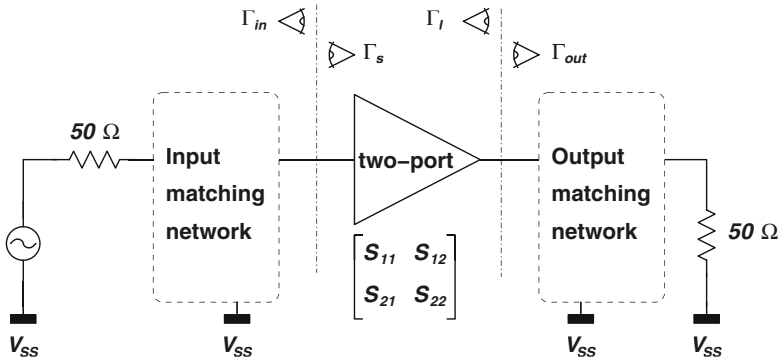


Figure 2.2: *An arbitrary two-port with lossless input and output matching networks.*

where

$$\Gamma = \frac{Z - Z_c}{Z + Z_c}. \tag{2.18}$$

Note that Γ is a complex number comprising both the amplitude ratio and the phase turn. If $Z = Z_c$ then $\Gamma = 0$ and no reflection occurs.

Power matching is in essence not related to impedance matching. The origin of power matching lies in the fundamental quest for energy efficiency. Suppose a voltage source (voltage V_S) with a source impedance Z_S drives a load impedance Z_L. The question is what value of Z_L maximizes the power dissipation in the load. It can easily be shown that this is achieved when

$$Z_L = Z_S^*, \tag{2.19}$$

with a maximum dissipated power in the load calculated as

$$P_{max} = \frac{V_S}{4\Re(Z_S)} \triangleq P_{av}. \tag{2.20}$$

This is also called the available source power.

Noise matching is completely unrelated to both previous types of matching. The origin here is the quest for good SNR and hence low noise figure. For a given two-port a noise match is obtained when the impedance of the source driving the two-port minimizes the noise figure of the resulting system. Referring to Appendix A this is achieved when $Z_S = Z_{opt}$.

In what follows the word '*matching*' must always be interpreted as '*impedance matching*' unless specifically stated otherwise.

2.2.4 Transducer Power Gain, Operating Power Gain and Available Power Gain

The concept of power gain of a two-port is not unambiguous. Several kinds of power gain are defined. Consider an arbitrary two-port as depicted in Fig. 2.2.

Transducer Power Gain or G_T is defined as follows:

$$G_T = \frac{\text{Power absorbed by the load}}{\text{Available power of the source}}. \tag{2.21}$$

Referring to Fig. 2.2, this can be rewritten as

$$G_T = \frac{1-|\Gamma_s|^2}{|1-\Gamma_{in}\Gamma_s|^2} \cdot |S_{21}|^2 \cdot \frac{1-|\Gamma_l|^2}{|1-S_{22}\Gamma_l|^2} \tag{2.22}$$

$$= \frac{1-|\Gamma_s|^2}{|1-S_{11}\Gamma_s|^2} \cdot |S_{21}|^2 \cdot \frac{1-|\Gamma_l|^2}{|1-\Gamma_{out}\Gamma_l|^2}, \tag{2.23}$$

where

$$\Gamma_{in} = S_{11} + \frac{S_{12}S_{21}\Gamma_l}{1-S_{22}\Gamma_l} \tag{2.24}$$

represents the reflection coefficient of the one-port constructed by the amplifier connected to the load Γ_l, and

$$\Gamma_{out} = S_{22} + \frac{S_{12}S_{21}\Gamma_s}{1-S_{11}\Gamma_s}. \tag{2.25}$$

is the equivalent representation of the output reflection coefficient.

The transducer power gain is most frequently used since the available source power is a given and the power in the load is what should be maximized.

Operating Power Gain or G_p is probably the most obvious definition. It is given by:

$$G_p = \frac{\text{Power absorbed by the load}}{\text{Power absorbed at the input}}. \tag{2.26}$$

Rewriting this in function of the reflection coefficients, yields

$$G_p = \frac{1}{1-|\Gamma_{in}|^2} \cdot |S_{21}|^2 \cdot \frac{1-|\Gamma_l|^2}{|1-S_{22}\Gamma_l|^2}. \tag{2.27}$$

Since this definition represents the output power normalized to the absorbed input power, it is independent of the actual equivalent source impedance represented by Γ_s.

Available Power Gain or G_{av} is defined as

$$G_{av} = \frac{\text{Available output power}}{\text{Available power of the source}}. \tag{2.28}$$

As a function of matching coefficients, this becomes

$$G_{av} = \frac{1-|\Gamma_s|^2}{|1-S_{11}\Gamma_s|^2} \cdot |S_{21}|^2 \cdot \frac{1}{1-|\Gamma_{out}|^2}. \tag{2.29}$$

Since the available power gain refers to the available output power it is independent of the actual equivalent load impedance represented by Γ_l.

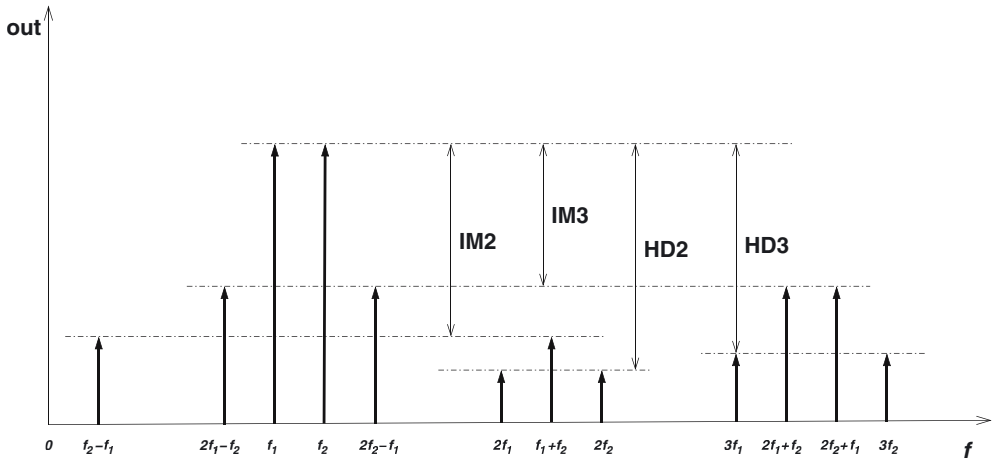


Figure 2.3: *Output tones in a two-tone test for a system with second and third order distortion.*

The available output power is always larger than the absorbed power by the load, therefore $G_T \leq G_{av}$. Similarly, the power absorbed at the input is always smaller than the available source power, hence $G_T \leq G_p$.

In the remainder of this work we will only use the transducer power gain which is shortened to power gain. In most cases the power gain of the LNA is simply equal to S_{21} unless specifically mentioned otherwise. The amplifiers are designed such that input and output impedance are sufficiently close to $50\ \Omega$ to justify this simplification. In other words Γ_s, Γ_{in}, Γ_l and Γ_{out} are sufficiently close to zero.

2.2.5 Intermodulation Distortion

Basically two kinds of non-linearities can be distinguished: weak non-linearities and hard non-linearities. The first kind can be described by a Taylor series and can be approximated with arbitrary accuracy by simply taking sufficient terms in the expansion. An example of a weak non-linearity is the $i_{ds} - v_{gs}$ relation for a MOS transistor in saturation. Hard non-linearities, for instance clipping, can not be described with a finite Taylor expansion. Typical for hard non-linearities is that almost no non-linearity is present for very small input amplitudes but all of a sudden the system behaves extremely non-linear (for instance when clipping starts). In the further analysis all non-linearities are considered to be weak non-linearities.

The linearity of circuits is usually investigated by means of harmonic distortion analysis or intermodulation distortion analysis. The first one assumes a sine wave is applied to the input. The fundamental and harmonics at the output are studied. For intermodulation, two tones are applied at the input and the intermodulation terms together with the fundamental tones are investigated.

Fig. 2.3 shows the output tones for a system with second and third order distortion. Suppose two tones are applied at frequencies f_1 and f_2. Besides the fundamental tones, the output also

shows second and third order harmonics and second and third order intermodulation products. The second order intermodulation gives rise to tones at $\pm(f_1-f_2)$ and at $\pm(f_1+f_2)$ as illustrated in Fig. 2.3. If f_1 and f_2 are located around the carrier f_c, then $f_1 - f_2 \approx 0$ and $f_1 + f_2 \approx 2f_c$. The first will be rejected by the DC-offset compensation and the second will usually be filtered out. Moreover, second order intermodulation terms are often very low due to differential implementations such that second order terms appear as common mode. Still, sufficient care needs to be taken in the receiver design since out of band signals may have a second order intermodulation term falling in the band of interest. The CMRR should be good enough to avoid signal degradation as a result of these signals.

Third order intermodulation will cause tones at frequencies $\pm(2f_1 - f_2)$, $\pm(2f_2 - f_1)$ and $\pm(2f_1 + f_2)$. When the applied tones are close to the carrier, the last intermodulation tone will be close to $3f_c$ and be filtered out but the first two will be within the band of interest. Since they are not linearly correlated with the input signal they can be considered as noise disturbing the signal. This is why the definition of SNR discussed in Section 2.2.2 is extended to SNDR, the signal to noise and distortion ratio:

$$\text{SNDR} = \frac{P_{av,s}}{P_{av,n} + P_{im}}, \tag{2.30}$$

where P_{im} is the combined power of the in band intermodulation signals.

The ratio of the amplitude of the third order intermodulation signals and the amplitude of the fundamental signal is called IM3. Consequently IM3 increases with the square of the input signal amplitude. Consider a system described by

$$y(t) = f(x(t)). \tag{2.31}$$

Performing a Taylor expansion of $y(t)$ yields

$$y(t) = a_0 + a_1x(t) + a_2x^2(t) + a_3x^3(t) + O(x^4(t). \tag{2.32}$$

where

$$a_k = \frac{1}{k!}\frac{d^k y}{dx^k}. \tag{2.33}$$

Two tones are applied at the input:

$$x(t) = U\sin(\omega_1 t) + U\sin(\omega_2 t). \tag{2.34}$$

IM3 is found as

$$\text{IM3} = \frac{3}{4}\left|\frac{a_3}{a_1}\right| \cdot U^2, \tag{2.35}$$

where U is the input signal amplitude. Note that (2.35) can be used for any weakly non-linear circuit. The input amplitude for which IM3 $= 1$ is called the input referred third order intermodulation intercept point or IIP3:

$$\text{IIP3} = \sqrt{\frac{4}{3}\left|\frac{a_1}{a_3}\right|}. \tag{2.36}$$

If a_1 is very large compared to a_3 and neglecting a_2 then IIP3 will be large and the output will be a linearly scaled version of the input signal for a wide input range.

2.3 The Deep Sub-Micron MOS Transistor at Radio Frequencies

2.3.1 MOS Model for Hand Calculations

The models and equations discussed in this section will be used throughout this work for hand calculations. The model is quite similar to [HSp01] MOS level 3 . The drain-source current of a NMOS in saturation is described by

$$I_{DS} = \frac{\mu C_{ox}}{2} \frac{W}{(L - \Delta L)} \frac{V_{GST}^2}{(1 + \Theta V_{GST})}, \tag{2.37}$$

where μ is the mobility, $C_{ox} = \epsilon_{ox}/t_{ox}$, Θ models the mobility degradation due to both longitudinal electric field (velocity saturation) and transverse electric field and V_{GST} symbolizes $V_{GS} - V_T$ in order to reduce the complexity of the expressions. The factor $L - \Delta L$ takes into account the channel length modulation. $\Delta L/L$ is — within a limited range — proportional to V_{DS} and therefore, (2.37) can be rewritten as

$$I_{DS} = K \frac{W}{L} \frac{V_{GST}^2}{(1 + \Theta V_{GST})} \frac{1}{(1 - \Lambda V_{DS})}, \tag{2.38}$$

where K, Θ and Λ are extracted from simulations (HSpice or Eldo) by means of the MOSCAL tool [Van02a]. Equation (2.38) describes the behavior of the transistor well within a selected region of operation. Naturally as this region is increased the model becomes increasingly inaccurate. Hence for very fine calculations — for instance during design optimization —, the design space needs to be split up into several smaller regions with its own set of describing parameters.

The small signal parameters used in the hand calculations can be derived from (2.38). The transconductance is found by differentiating I_{DS}:

$$g_m \triangleq \frac{\partial I_{DS}}{\partial V_{GS}} = 2K \frac{W}{L} V_{GST} \frac{1}{(1 + \Theta V_{GST})} \left(\frac{1 + \frac{\Theta}{2} V_{GST}}{1 + \Theta V_{GST}} \right) \frac{1}{(1 - \Lambda V_{DS})} \tag{2.39}$$

The cut-off pulsation neglecting C_{gd} is now found as

$$\omega_T = \frac{g_m}{C_{gs}} = \frac{\mu V_{GST}}{L^2} \frac{1}{(1 + \Theta V_{GST})} \left(\frac{1 + \frac{\Theta}{2} V_{GST}}{1 + \Theta V_{GST}} \right) \frac{1}{(1 - \Lambda V_{DS})} \tag{2.40}$$

The finite output resistance due to the channel length modulation is approximated by

$$r_{ds} = \frac{1}{\Lambda I_{DS}}. \tag{2.41}$$

This can easily be understood from (2.38) if the factor $1/(1 - \Lambda V_{DS})$ is replaced with $(1 + \Lambda V_{DS})$ which is justified if $\Lambda V_{DS} \ll 1$. Note that even though Λ is inversely proportional to the effective channel length, this dependence can be ignored since all transistors feature the minimal length.

NMOS

K_n	V_{Tn}	Θ_n	Λ_n	α	α_{gd}	α_{gb}	$\alpha_{db} = \alpha_{sb}$ $V_{DB} = 0.5$ V	$\alpha_{db} = \alpha_{sb}$ $V_{DB} = 1.5$ V	γ	δ
[μA/V^2]	[V]	[V^{-1}]	[V^{-1}]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
192	0.52	4.15	0.07	0.83	0.23	0.16	0.68	0.53	2	4

PMOS

K_p	$\lvert V_{Tp}\rvert$	Θ_p	Λ_p	α	α_{gd}	α_{gb}	$\alpha_{db} = \alpha_{sb}$ $V_{DB} = 0.5$ V	$\alpha_{db} = \alpha_{sb}$ $V_{DB} = 1.5$ V	γ	δ
[μm]	[μA/V^2]	[V]	[V^{-1}]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
55	0.50	3.87	0.07	0.83	0.23	0.16	0.68	0.53	1	2

COMMON

L_{eff} [μm]	t_{ox} [nm]
0.2	5.5

Table 2.1: *Hand calculation parameters for the NMOS and PMOS in the 0.25 μm CMOS technology of Kawasaki Microelectronics (extracted for $V_{GS} - V_T$ values between 0.1 and 0.3 V).*

Unless specifically stated otherwise, all illustrated calculations have been done based on the hand calculation parameters in Table 2.1. Hereby, α is defined as

$$\begin{aligned} \alpha &= \frac{g_m}{g_{d0}} \\ &\approx \frac{1}{n} \triangleq \frac{g_m}{g_m + g_{mb}}, \end{aligned} \tag{2.42}$$

where g_{d0} is the drain-source conductance at zero V_{DS}. Parameters α_{xy} are defined as

$$\alpha_{xy} = \frac{C_{xy}}{C_{gs}}, \tag{2.43}$$

and γ and δ represent the excess noise factors discussed in Section 2.4.2.

2.3.2 Linearity of the short-channel MOS transistor

Since for an LNA, the main non-linearity problem is the 3rd order intermodulation, this subsection will evaluate the intermodulation performance of a MOS transistor by means of the IV3. This is the gate-source voltage amplitude for which the intermodulation drain current intercepts the fundamental drain current. In principle it is identical to the IIP3 —as it was introduced in Section 2.2.5— in as far as the gate-source voltage amplitude is the actual input and no conversion for input reference is required. The symbol IV3 is used here for the more general case where the input signal is different which will be the case in the amplifiers discussed further on. This

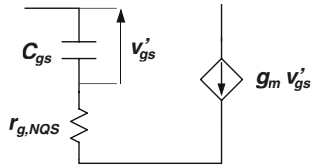


Figure 2.4: *Non-Quasi Static model for the delay in the channel charge buildup.*

will avoid confusion with the actual input referred intercept point or IIP3. The small signal input is $v_{gs}(t)$ and the output of the transistor is the current $i_{ds}(t)$. The total output current I_{DS} of a NMOS is described by (2.38). This reduces to

$$I_{DS} = K\frac{W}{L}\frac{V_{GST}^2}{(1+\Theta V_{GST})}. \tag{2.44}$$

where V_{DS} is assumed constant. This assumption is usually justified as shown mathematically in [Jan01]. Now the small signal current i_{ds} can be written as

$$i_{ds} = K\frac{W}{L}\left(\frac{(V_{GST}+v_{gs})^2}{(1+\Theta\,(V_{GST}+v_{gs}))} - \frac{V_{GST}^2}{(1+\Theta V_{GST})}\right). \tag{2.45}$$

This function can be expanded in a Taylor series which after some calculation, similar to the general derivation in Section 2.2.5, yields the following expression for IV3:

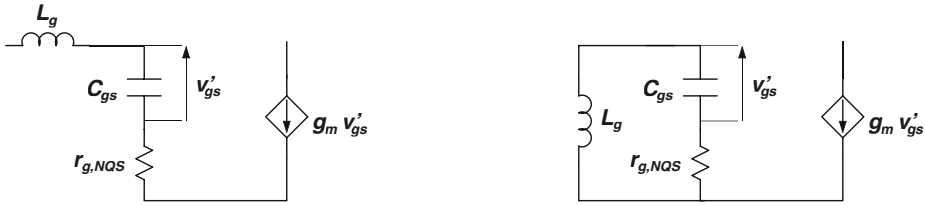
$$\text{IV3} = \sqrt{\frac{4}{3}\left|\frac{a_1}{a_3}\right|} \tag{2.46}$$

$$= \sqrt{\frac{4}{3}\frac{V_{GST}\,(2+\Theta V_{GST})\,(1+\Theta V_{GST})^2}{\Theta}}, \tag{2.47}$$

where IV3 is expressed in Volt amplitude. For any NMOS in saturation, the gate-source intercept voltage is given by (2.47).

2.3.3 Non-Quasi Static Model

The classical quasi static model of the MOS transistor behavior assumes that any change in charge at the gate is instantly reflected with an equal but opposite amount of charge in the channel. However, in reality there will always be a delay in the channel charge buildup. The physics of the MOS transistor tells us that the channel is built by means of inversion. Remember the behavior of the NMOS capacitor where the channel depletion starts when the gate voltage is increased above 0 V. Above V_T, electrons will be drawn from the bulk material creating an excess of inversion carriers in the channel. Considering this, it is intuitively clear that the process of adding an extra electron to the channel has a finite time constant.



(a) Series resonance. (b) Parallel resonance.

Figure 2.5: *The capacitive input of the MOS transistor tuned out with an inductor.*

This effect has been described and modelled by Y. Tsividis in [Tsi87, Jan99a]. A simplified model valid in strong inversion and within the long-channel approximation yields the following time constant associated with C_{gs}:

$$\tau_{gs} = \frac{C_{gs}}{5g_m} = \frac{1}{5\omega_T}. \tag{2.48}$$

Consequently, this delay effect can be modelled by adding a resistor $r_{g,NQS}$ in series with C_{gs}:

$$r_{g,NQS} = \frac{1}{5g_m}. \tag{2.49}$$

This model is illustrated in Fig. 2.4.

The frequency corresponding to this time constant is $5 \times f_T$ so one would think that this effect is not important at realistic operating frequencies much smaller than f_T. However in bandpass applications, the input capacitance may be tuned out with a series inductor. This means that for a given input current the voltage over C_{gs} is cancelled by the equal but opposite voltage over the inductor. The input impedance of the transistor is now purely resistive as shown in Fig. 2.5(a):

$$Z_{in,s} = r_{g,NQS} = \frac{1}{5g_m}. \tag{2.50}$$

Similarly, when the input capacitance is tuned out with a parallel inductor, the input is again purely resistive as demonstrated in Fig. 2.5(b):

$$Z_{in,p} \approx \frac{1}{\omega_r^2 C_{gs} r_{g,NQS}} = \frac{5f_T^2}{g_m f_r^2} \tag{2.51}$$

where f_r is the resonance frequency.

In short-channel MOS transistors the value of the non-quasi static resistor is still under discussion. It is generally assumed that the proportionality with g_m remains but the constant might be changed. Therefore (2.49) is rewritten as

$$r_{g,NQS} = \frac{1}{\kappa g_m}, \tag{2.52}$$

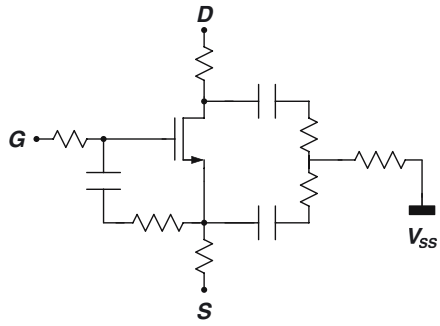


Figure 2.6: *NMOS transistor with 10 additional components for more accurate simulations.*

where κ represents the Elmore constant. The Elmore constant is usually represented by ϵ but κ is used here to avoid confusion with the permittivity. This NQS model is also implemented in MOS model BSIM3V3. All numerical examples in this text will assume $\kappa = 5$ unless specifically stated otherwise [Enz02].

2.3.4 Extended MOS Model for Simulation

Most simulations are performed using HSpice or Eldo in combination with Berkeley MOS model BSIM3V3 (level 49 in [HSp01] and level 53 in [Eld01]). Even though this model is far more complex than the one used for hand calculations (cf. Section 2.3.1), it is still lacking intrinsic accuracy for RF simulations. The designer however has a lot of options to solve this. An NMOS transistor is shown in Fig. 2.6 where 10 extra components have been added to better describe the behavior at high frequencies and more accurately predict the noise behavior.

If the NQS effect is not included in the MOS model that is used it can be modelled by placing the equivalent parallel resistor in parallel with C_{gs} given by

$$R_{p,NQS} \approx \frac{1}{\omega_0^2 C_{gs} r_{g,NQS}} = \frac{5 f_T^2}{g_m f_0^2}. \tag{2.53}$$

This model is only valid at frequency f_0 and should be used with care. An extra coupling capacitance is added in series with $R_{p,NQS}$ in order not to disturb the operating point of the circuit.

Fig. 2.6 also shows the series resistors for the gate, source and drain region. The gate resistor represents the resistance of the poly gate. Taking it into account is important for accurate noise simulations. The same goes for the source resistor which models the resistance of the n+ source region. The drain resistor can also be important for instance in switched power amplifiers where it will increase the on-resistance of the switch.

Both at drain and source a capacitor is added to represent their respective junction capacitances. They are resistively coupled to the bulk node. Finally a resistor is added representing the resistance from the bulk node to the actual bulk contact. This resistor is relatively large since it is formed in a high ohmic p-well or n-well region.

Figure 2.7: *Noise voltage and noise current of an arbitrary resistor.*

2.4 The Origin of Noise

2.4.1 Resistor Thermal Noise

Probably the most well known noise source is the thermal noise of a resistor (also called Johnson noise). It is white noise since the PSD of the noise signal is flat throughout the frequency band. The noise is also called gaussian which means the amplitude of the noise signal has a gaussian distribution. The noise power is proportional to absolute temperature. The available noise power which is the same for every resistor is given by

$$P_{av,n} = kT\Delta f. \tag{2.54}$$

where k is Boltzmann's constant ($\sim 1.38 \times 10^{-23}$ J/K), T is the absolute temperature in K and Δf is the noise bandwidth in Hz. For ease of calculation, this available noise power is usually converted to a noise voltage source in series or a current source in parallel with the resistor as shown in Fig. 2.7. They are respectively given by

$$\overline{v_{nr}^2} = 4kTR\Delta f \qquad \text{and} \qquad \overline{i_{nr}^2} = \frac{4kT}{R}\Delta f. \tag{2.55}$$

Even though the available noise power is independent of the resistance, these voltage and current sources are not. Consequently the choice of a specific resistor is very important also from a noise point of view. For instance for a high impedance node, a minimum amount of injected noise current is desired. A high resistance is then preferred. However in series with an input voltage source a low resistance is preferred to keep the noise voltage low. The fact that they have the same available noise power is not relevant here.

2.4.2 Thermal Noise in MOS transistors

2.4.2.1 Classical MOS Channel Noise

It is quite clear that MOS transistors in the linear region need to display some sort of thermal noise. After all, the linear MOS transistor is essentially a controlled resistor. The drain noise current (Fig. 2.8) was calculated by [vdZ62]:

$$\overline{i_{nd}^2} = 4kT\gamma g_{d0}\Delta f, \tag{2.56}$$

where g_{d0} is the drain-source conductance at zero V_{DS}. Parameter γ is one at zero V_{DS} and — for long devices— decreases to a value of 2/3 in saturation. However, in short-channel NMOS

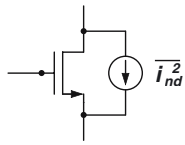


Figure 2.8: *Classical drain noise current for an NMOS transistor.*

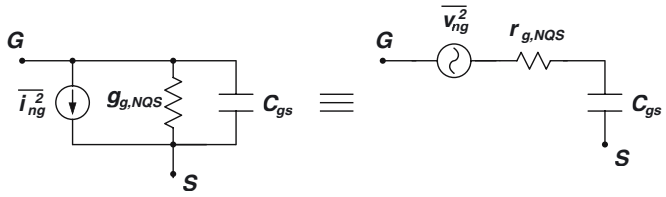


Figure 2.9: *Schematic of induced gate noise current and the equivalent voltage.*

devices the effective temperature of the carriers is significantly larger due to the high electric field in the channel. γ values of 2, 3 and more have been reported [Lee98]. Since the electric field for a fixed device is proportional to the V_{DS} it is important to keep this voltage as low as possible. Sometimes it may even be advised to use a non-minimum length transistor if this does not inhibit the required frequency performance. The PMOS transistor usually exhibits lower γ values than its NMOS counter part. Therefore, it could also be interesting to use a PMOS transistor for as far as other criteria allow this. In order to simplify calculations with MOS transistors in the saturation region, (2.56) is rewritten as

$$\overline{i_{nd}^2} = 4kT\frac{\gamma}{\alpha}g_m\Delta f, \tag{2.57}$$

where α is given by (2.42).

2.4.2.2 Induced Gate Noise

Since the gate is capacitively coupled with the channel, the drain noise also leads to a noisy gate voltage as shown by [vdZ86, Enz02].

$$\overline{i_{ng}^2} = 4kT\delta g_{g,NQS}\Delta f, \tag{2.58}$$

where

$$g_{g,NQS} = \omega^2 C_{gs}^2 r_{g,NQS} \tag{2.59}$$

and δ is 4/3 for long-channel transistors as shown in [vdZ86]. This means that $\delta = 2 \times \gamma$. Since exceedingly hot carriers that increase the drain noise are also expected to increase the induced gate noise, it can be justified to state that also for short channels this equation remains valid. This was postulated in [Lee98].

The induced gate noise is clearly linked to the non-quasistatic gate resistance. In fact it can be considered as the thermal noise of this resistor. Consequently, the noise voltage $\overline{v_{ng}^2}$, (as shown in Fig. 2.9) may be expressed as

$$\overline{v_{ng}^2} = 4kT\delta r_{g,NQS}\Delta f, \tag{2.60}$$

Even though this expression is correct, care should be taken. Since the induced gate noise behaves partly as a capacitive reflection of the channel noise, both noise sources are not uncorrelated. The correlation coefficient for both noise currents, defined as

$$c \triangleq \frac{\overline{i_{ng} \cdot i_{nd}^*}}{\sqrt{\overline{i_{ng}^2} \cdot \overline{i_{nd}^2}}} \tag{2.61}$$

is equal to $j0.395$ for long-channel MOS transistors as shown by [vdZ86]. For ease of calculation it is assumed throughout the rest of this work that

$$c = j0.4, \tag{2.62}$$

for all regimes.

2.4.3 1/f Noise

Even though 1/f noise, pink noise or Flicker noise is very important in CMOS RF receivers, it will not be discussed in this text. As the name already stipulates, this type of noise has a PSD increasing towards low frequencies. Hence it will not be significant in low noise amplifiers operating in the GHz range. It will however have a prominent impact on the behavior and design of VCO's and down-conversion mixers. For the sake of completeness the most common expression for the PSD of the 1/f noise current in a MOS transistor is given below:

$$\overline{i_{nf}^2} = \frac{K_f}{f} \cdot \omega_T^2 \cdot WL \cdot \Delta f, \tag{2.63}$$

where K_f is a constant, f is the frequency, ω_T is the cut-off pulsation and WL is the transistor area. The location of the noise source is identical to that of the classical drain noise represented in Fig. 2.8.

2.4.4 Shot Noise

This noise mechanism was first introduced by Schottky and is based on the discrete nature of electrical charge. It occurs when a current flow crosses a potential barrier where the discreteness of the arrival times of the individual charges give rise to the noise current. Equivalent to thermal noise the PSD is flat and hence it is also a sort of white noise. The shot noise current is given by:

$$\overline{i_{nsh}^2} = 2qI_{DC}\Delta f, \tag{2.64}$$

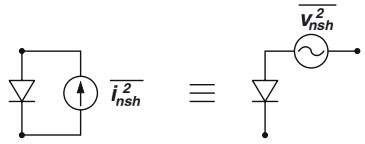


Figure 2.10: *Shot noise in silicon diodes.*

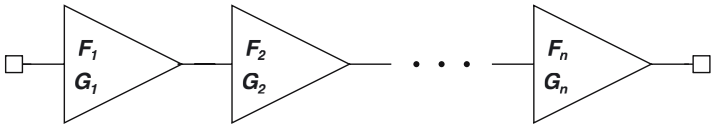


Figure 2.11: *Noise figure of a cascade of linear noisy gain blocks.*

where q is the elementary charge of an electron ($\sim 1.6 \times 10^{-19}\ C$) and I_{DC} is the DC current through the barrier.

In silicon, shot noise is most commonly found in pn-junctions (see Fig. 2.10). Consequently it is the most dominant noise source in bipolar transistors where large currents cross the base-collector junction.

2.5 The LNA in the Receiver Chain

2.5.1 Cascading Non-Ideal Building Blocks

2.5.1.1 Noise in a Cascade

Understanding what happens with the noise figure and distortion components when several non-ideal blocks are cascaded is crucial in the design of any receiver. Consider a cascade of linear gain blocks with power gain G_i and noise factor F_i as depicted in Fig. 2.11. Each block is assumed to be matched to 50 Ω at both input and output. The equivalent input noise of the final block —characterized by its noise factor F_n— can be referred to the input of the preceding block by dividing it by the gain of this block. This yields an equivalent noise factor for the cascade configuration of block n and block $n-1$ given by:

$$F_{n-1,n} = F_{n-1} + \frac{F_n - 1}{G_{n-1}}. \tag{2.65}$$

Continuing this technique all the way to the input of block 1 yields

$$F = F_1 + \sum_{i=2}^{n} \frac{F_i - 1}{\prod_{k=2}^{i} G_{i-1}}. \tag{2.66}$$

This means the noise added in each stage is suppressed by all the gains of the preceding stages. Hence the noise of subsequent stages becomes progressively less important. This can also be

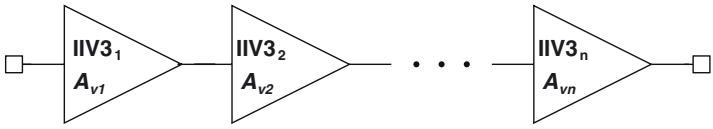


Figure 2.12: *IIP3 of a cascade of non-linear gain blocks.*

understood well from the fact that signal levels become higher proceeding through the cascade and additive noise becomes more and more negligible.

2.5.1.2 IIV3 of a Cascade

However as signal levels grow larger, the amount of distortion increases. Similar to the noise figure, also the IIV3 (the input referred intercept voltage) can be referred back to the input. Consider the cascade in Fig. 2.12. The blocks are now considered noiseless and non-linear and are characterized by their voltage gain A_{vi} and $\mathrm{IIV3}_i$. For simplicity the input impedance of each block is considered infinite while the output impedance is zero. Consider the output voltage of block 1. The third order intermodulation terms are

$$V_{im3,1} = \frac{A_{v1}V^3}{\mathrm{IIV3}_1^2} \tag{2.67}$$

where V is the input voltage amplitude. It is clearly seen that $V_{im3,1} = A_{v1}V$ for $V = \mathrm{IIV3}_1$ which follows directly from the definition of IIV3. The intermodulation terms at the output of block 2 consist of the amplified intermodulation terms at the output of block 1 and the intermodulation terms generated by block 2:

$$\begin{aligned} V_{im3,tot2} &= A_{v2}V_{im3,1} + \frac{A_{v2}\left(A_{v1}V\right)^3}{\mathrm{IIV3}_2^2} && (2.68)\\ &= \frac{A_{v2}A_{v1}V^3}{\mathrm{IIV3}_1^2} + \frac{A_{v2}\left(A_{v1}V\right)^3}{\mathrm{IIV3}_2^2}. && (2.69) \end{aligned}$$

The total IIV3 of the cascaded blocks can now be calculated.

$$\begin{aligned} \frac{1}{\mathrm{IIV3}_{tot}^2} &= \frac{V_{im3,tot2}}{A_{v2}A_{v1}V^3} && (2.70)\\ &= \frac{1}{\mathrm{IIV3}_1^2} + \frac{A_{v1}^2}{\mathrm{IIV3}_2^2}. && (2.71) \end{aligned}$$

This method can be extended for an arbitrary amount of blocks yielding the following formula for the total IIV3:

$$\frac{1}{\mathrm{IIV3}_{tot}^2} = \frac{1}{\mathrm{IIV3}_1^2} + \sum_{i=2}^{n} \frac{\prod_{k=1}^{i-1} A_{vk}^2}{\mathrm{IIV3}_i^2}. \tag{2.72}$$

This shows that the IIV3 of a block becomes more important as the amount of gain preceding the block increases, which means that the signals fed to this block grow larger. Since in a normal

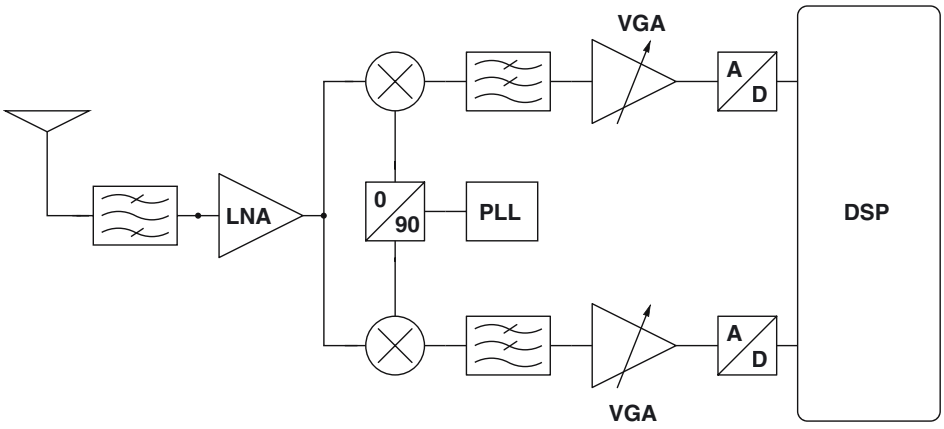


Figure 2.13: *Architecture of a direct conversion receiver.*

receiver the signals increase the further you proceed in the receiver chain, the IIV3 of the later blocks should be progressively higher than that of the first block.

2.5.2 Wireless Receiver Architectures

Clearly an ideal cascade of amplifying stages is no sufficient model for a wireless receiver. In general a receiver consists of an antenna, a band-select filter, a low-noise amplifier, one or more downconversion stages with a phase-locked loop and with or without interstage filtering, and an A/D converter with or without variable gain amplification. In classical superheterodyne receivers the downconversion was done in several steps. This required intermediate high-Q filtering for image cancellation after all stages. This topology is not interesting for a full CMOS wireless receiver since these high quality mixers have to be implemented externally and they would excessively increase the overall system cost. Therefore, two other topologies have been devised which don't need these high-Q filters: the zero-IF receiver [Abi95, Raz97] and the low-IF receiver [Cro98]. Both are direct conversion receivers since they don't require a second downconversion. The general schematic of this architecture is shown in Fig. 2.13.

In the zero-IF architecture, the LO has the same frequency as the RF-carrier. Consequently the RF-signal is its own image and hence cannot be filtered out as such. This is solved by mixing the RF signal with both sine and cosine of the LO. This is called quadrature mixing since a differential representation of both sine and cosine of the LO yields four 90 degrees shifted LO signals. If both signal paths are perfectly matched, the signals at the output of the mixers can be recombined (usually in the digital domain) to completely cancel the image, leaving only the wanted signal. The same technique is used for low-IF receivers. Since these topologies no longer require high-Q image filters, they are ideally suited for CMOS implementation. Moreover the image cancellation can be carried out in the digital back-end which is after all the core competence of CMOS.

The design of such direct-conversion receivers still poses some problems.

- Even though the actual image cancellation is done in the digital domain, it is the matching of the two quadrature paths in the analog front-end which determines the IMRR. Values of 30 to 40 dB have been achieved.
- Since there is no filtering in the RF front-end (the moderate band-select filtering aside), the signals reaching the ADC have a high dynamic range due to the possible presence of large blocking signals.
- Especially for zero-IF receivers, the presence of $1/f$ noise and DC-offset can severely limit the performance.

2.5.3 LNA Requirements

2.5.3.1 Matching

Referring to Section 2.2.3 the importance of the different types of matching for the input of the LNA will be discussed. Suppose the LNA is fed through a 50 Ω transmission line coming from the antenna or an off-chip band-select filter. First consider the power matching requirement. Since the MOS transistor is basically a voltage driven current source, an input power match is not required for a large output power. Moreover, it may be interesting to have an open circuit at the input since this would give the largest input voltage and hence the largest output current. Conclusion: power matching is not required.

However, for a 50 Ω source, the power match is identical to the impedance match. The reason for input impedance matching in the LNA is twofold. First, it avoids reflections over the transmission line feeding the LNA. And second it supplies a correct termination for the possible SAW-filter preceding the amplifier. This termination resistance is required in order to guarantee the frequency characteristic of the filter, both in the pass-band and in the stop-band. In pass-band, an incorrect termination resistance may lead to extra attenuation. The filter is the first block in the receiver. It attenuates the signal but due to its 50 Ω output impedance it has the same output noise power as the antenna (neglecting temperature differences). Hence an attenuation of 3 dB lowers the SNR with 3 dB which is equivalent to a noise figure of 3 dB. Any extra dB attenuation increases the noise figure with one dB. This must be avoided at all cost.

Also in the stop-band a correct termination is desired. Without it the attenuation in the stop-band could be reduced or the behavior of the stop-band ripple might be altered. Even though this has no direct influence on the wanted signal, it could lead to insufficiently suppressed blocker signals which can yield large in-band intermodulation products. Conclusion: a close to 50 Ω input impedance is very important in the signal band. It is also desired outside the signal band.

Noise matching aims at providing this equivalent source impedance to a given circuit which minimizes the noise figure of the circuit (cf. Appendix A). Often, the noise figure has quite a flat behavior around its optimum. such that an impedance match yields a sufficiently low noise figure. In classic microwave design, the amplifier (or transistor) is fixed once it has been chosen. The design is then done by choosing the equivalent source and load impedance that yield a

Specification	Receiver 1	Receiver 2
NF_{LNA}	3 dB	3 dB
$\mathrm{IIP3}_{LNA}$	0 dBm	0 dBm
$A_{v,LNA}$	15 dB	30 dB
NF_{mix}	15 dB	15 dB
$\mathrm{IIP3}_{mix}$	15 dBm	15 dBm

Table 2.2: *LNA and mixer specifications in Fig. 2.14.*

stable amplifier with a sufficient impedance match and a good noise figure. However in our case there are many more degrees of freedom. Consequently, the noise optimization can be done on transistor level while taking the impedance match as a constraint.

2.5.3.2 Noise Figure

Neglecting the channel-select filter, the LNA is the first building block in the receiver. As such it sets a lower bound on the attainable noise figure for the entire receiver. A very low noise figure is crucial. This becomes even more important for high sensitivity receivers like for the GPS system where the signal levels that need to be detected are extremely small.

2.5.3.3 Voltage Gain or Power Gain

The gain of the LNA should be large for more or less the same reason. It was learned from Section 2.5.1.1 that the noise of the stages following the LNA is suppressed by the gain of the LNA. Consequently for a receiver, the gain should be very large to minimize the noise figure contribution from the down-conversion mixer. Since the mixer is usually driven by a voltage, it is the voltage gain that should be optimized. Only if the LNA drives an external 50 Ω source (stand-alone LNA's), the power gain is considered in the optimization. For lab realizations and prototypes, the LNA output is often designed to drive 50 Ω as well, in order to ease the measurements.

In a real life environment however the LNA output stage is determined by the attached load, namely the input of the mixer. For a linear mixer this can in principal be either the capacitive load of the gate of a linear MOS transistor or the resistive load of its drain-source conductance. If the RF signal drives the gate, the voltage over the gate should be maximized and clearly the voltage gain is the main criterium. If the LNA drives the source of the mixing transistor, then the current through this transistor should be maximized. Since the current through the mixing transistor is proportional to the voltage over it, again the voltage gain should be maximized. For Gilbert type implementations, the load is always capacitive.

During the design of the LNA it might be tempting to just maximize the gain regardless of other building blocks or architectural considerations. However this is not advised. Increasing the gain of the LNA increases the signal levels in the mixer and this could give linearity problems. Equation (2.72) shows indeed that the IIP3 contribution of the mixer increases linearly with the gain of the LNA. For a Gilbert mixer which is not very linear, the gain of the LNA is usually kept

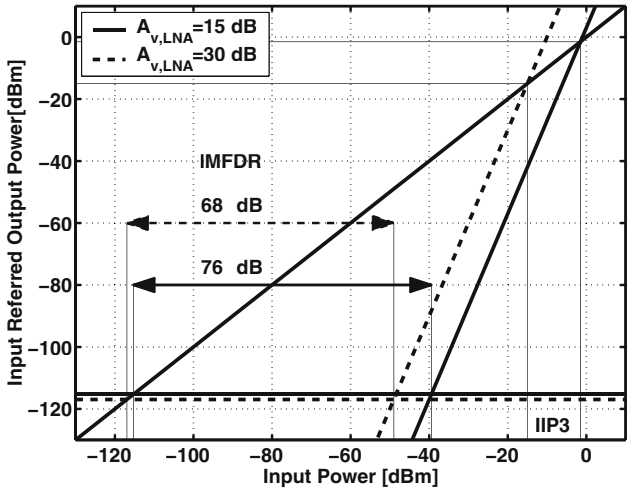


Figure 2.14: *Influence of the LNA gain on the IMFDR of a receiver.*

lower than when driving a linear mixer. To put it in another way, the voltage gain of the LNA should be set to maximize the dynamic range of the total receiver. If the next blocks are very linear but noisy, the gain is increased and vice versa. Fig. 2.14 shows the intermodulation free dynamic range (IMFDR) for a receiver consisting of an LNA and a mixer. IMFDR is defined by

$$\text{IMFDR [dB]} \triangleq \frac{2}{3}\text{IIP3 [dBm]} - \frac{2}{3}(P_{av,n} + P_{n,eq}) \text{ [dBm]}, \tag{2.73}$$

where $P_{av,n} + P_{n,eq} = 10\log(kT) + \text{NF}$ is the total input referred noise power of the receiver. It is often called the noise floor of the receiver since any signal below it is not visible on a spectrum analyzer. The IMFDR is the signal range between the level where the fundamental tone becomes visible and the level where the third order intermodulation terms become visible. The specifications of LNA and mixer used in Fig. 2.14 are listed in Table 2.2. The only difference between the two receivers is the LNA voltage gain of 15 dB and 30 dB respectively. Fig. 2.14 shows that the dynamic range of the receiver decreases with 8 dB when increasing the voltage gain of the LNA leaving the other specs unaltered. The reason is that the linearity of the mixer in this example is very (unrealistically) poor and the noise performance is rather good. Consequently the mixer is better off with lower signal powers.

It may be interesting to compare the high-level design of the LNA gain with the exposure control in a digital camera. The image sensor of the camera is representative for the fixed receiver front-end excluding the LNA. In high-end digital SLR cameras, a real-time histogram of the viewfinder image is used to set the correct exposure in order to fully exploit the dynamic range of the image sensor. If some region is too dark, the exposure needs to be increased, if it is too bright, the exposure can be reduced. A similar investigation of the possible signal levels allows

to choose the best gain setting considering the dynamic range[1] of the subsequent building blocks The main difference is that the gain of the LNA is usually fixed after design. That is why, for the LNA, not one picture, but all possible pictures (read: signal levels) need to be considered in this gain optimization. In a complete receiver front-end design, of course, not only the gain of the LNA but all the specifications of the building blocks need to be considered in a global high-level optimization, also taking into account the power of blocking signals. This is done by investigating the minimum SNDR as the signal proceeds through the receiver.

2.5.3.4 Intermodulation Distortion

Similar to the receiver noise figure, which is lower bounded by the noise figure of the LNA, the IIP3 of the receiver is upper bounded by the IIP3 of the LNA. In many applications, the linearity specifications for the LNA don't pose many difficulties in the design of the receiver. Some applications have stronger linearity requirements, for instance because they need to be able to receive very large signals when the distance to the transmitter becomes small. Another reason can be the presence of large blocking signals in a neighboring band.

Even though the linearity requirements become more stringent further in the receiver path, the relative power of blocking signals usually decreases since each block has some intrinsic filtering. Consequently, the dynamic range of the signals is reduced further in the receiver, when the blocking signals are dominant in the linearity specification. This means the LNA requires the highest dynamic range of all receiver building blocks.

2.5.3.5 Reverse Isolation

The reverse isolation is defined as $-S_{12}$ where S_{12} is the reverse gain of the LNA. Basically three driving forces exist for increasing the reverse isolation. The first one is based on the spurious emission specification of the receiver. The signal coming from the local oscillator may couple back through the mixer to the output of the LNA. This signal can reach the antenna through the reverse gain of the LNA. The higher the reverse isolation, the smaller the spurious LO tone at the antenna. The amount of reverse isolation required depends on the LO signal amplitude, the coupling through the mixer and the allowed spurious signal level at the antenna. Usually a value of 25 dB to 30 dB is sufficient.

Another reason for increasing the reverse isolation is that the input matching becomes considerably more reliable when the reverse isolation is high. Often a low reverse isolation goes hand in hand with a reduced gain since the inherent feedback of the reverse gain reduces the signal efficiency.

A final driving force is the intrinsic stability of the amplifier which is discussed next.

[1] The dynamic range implies in fact two numbers here, the minimum and maximum signal level, not just the difference between them. Otherwise no conclusion about the required gain could be drawn.

2.5.3.6 Stability

Several techniques exist to describe and design stable amplifiers. For RF and microwave amplifiers, it is interesting to define the concept of unconditional stability. A circuit is unconditionally stable if for any combination of source and load impedance, the circuit is stable. A single parameter μ_s was defined in [Edw92][2] which can describe the unconditional stability of an amplifier as a function of its S-parameters:

$$\mu_s = \frac{1 - |S_{11}|^2}{|S_{22} - S_{11}^* \Delta| + |S_{21} S_{12}|} > 1 \tag{2.74}$$

where Δ is the determinant of the S-parameter matrix:

$$\Delta = S_{11} S_{22} - S_{12} S_{21}. \tag{2.75}$$

A completely equivalent but not equal parameter can be found by interchanging indexes 1 and 2. Parameter μ can be calculated by the simulator in order to investigate the stability of the amplifier. Generally μ increases for decreasing $|S_{12}|$. Hence, increasing the reverse isolation will ease the design for stability. Moreover, if $S_{12} = 0$ then (2.74) reduces to

$$\mu_s = \frac{1}{|S_{22}|} > 1 \qquad \text{or} \qquad |S_{22}| < 1. \tag{2.76}$$

Again indexes 1 and 2 can be interchanged.

For integrated CMOS LNA's, simulating or designing for unconditional stability is quite difficult, especially at high frequency. This is due to the lack of RF-models for the devices as they are laid out by the designer. Some companies provide S-parameter models for a limited set of devices which can be implemented with parameterized cells. Simulating the stability with those cells is possible and yields good results when the S-parameters have been extracted correctly from test chip measurements. Drawback of the use of these devices is that it takes away a lot of the design freedom.

2.5.3.7 Single-ended vs. Differential

The main reason for using a differential topology is the common mode rejection. This has some major advantages. The noise on the power supply lines appears as common-mode and is not seen at the output. This largely reduces possible problems with digital switching noise when integrating both RF, baseband analog and digital on the same die. Moreover, the even harmonics also appear as common mode and are similarly suppressed. Some blocking signals that have a second harmonic in the signal band are this way largely disarmed.

The main drawback of differential topologies is that they require more or less twice the power for the same performance. In handheld or wearable applications, where power consumption is the bottleneck, this is often unacceptable. Naturally this trade-off should be considered for every application separately.

[2]The original symbol is μ but μ_s is used here to avoid confusion with the mobility μ.

2.6 Topologies for Low-Noise Amplifiers

2.6.1 The Inductively Degenerated Common Source LNA

2.6.1.1 From Basic Common-Source Amplifier to Inductively Degenerated Common-Source LNA

This section will discuss the origin and the basic schematic of one of the most popular LNA topologies, known as the inductively degenerated common source LNA. Based on Fig. 2.15, the gradual evolution from basic common source amplifier to inductively degenerated CS LNA is explained. The circuit, depicted in Fig. 2.15(a) shows a simple baseband one-transistor CS amplifier. The first LNA criterium is already fulfilled, the positive gain requirement. One of the problems with this circuit is that it has a purely capacitive input impedance (at least according to the classical quasi-static MOS model).

In order to create a resistive input, it suffices to place a termination resistor parallel to the LNA input (Fig. 2.15(b)). In this figure, the termination resistor is connected to ground. In reality this would upset the DC biasing of the amplifier. Consequently, resistor R_t should be connected to the DC biasing node. This node can be decoupled from the ground with a large decoupling capacitor $C_{dc,in}$. Therefore, R_t is connected to AC ground and the AC performance ($\omega \gg 1/(R_t C_{dc,in})$) remains unaltered. The input impedance for $\omega \gg 1/(R_t C_{dc,in})$ is given by

$$Z_{in} = \frac{R_S}{1 + \frac{j\omega}{\omega_p}}, \tag{2.77}$$

where

$$\omega_p = \frac{1}{R_S(C_{gs} + MC_{gd})}, \tag{2.78}$$

M is the Miller factor and $R_S = 50\ \Omega$, the source resistance. The Miller effect for this circuit is very pronounced,

$$M = 1 + g_{m1}R_L. \tag{2.79}$$

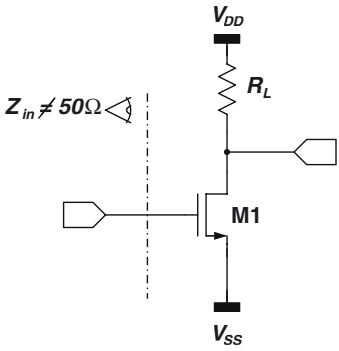
This effect strongly limits the frequency performance and gives rise to a very poor reverse isolation.

Adding a cascode transistor as shown in Fig. 2.15(c) significantly decreases the Miller effect since it is now decoupled from the gain of the circuit. If $R_L \ll r_{ds2}$ where r_{ds2} is the output resistance of cascode transistor M2, then

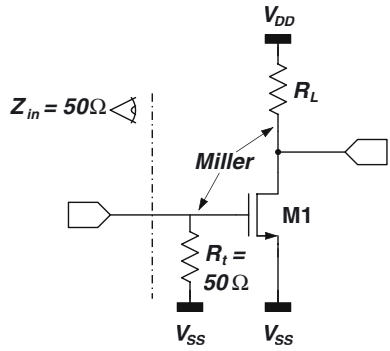
$$M = 1 + \frac{g_{m1}}{g_{m2}} \approx 2. \tag{2.80}$$

One of the problems with this circuit is that R_L needs to be large for a high gain. However this will cause a large DC voltage drop over R_L. In order for the circuit to operate within parameters, the voltages over M1 and M2 need to be larger then $V_{DS,sat} = V_{GS} - V_T$. Therefore,

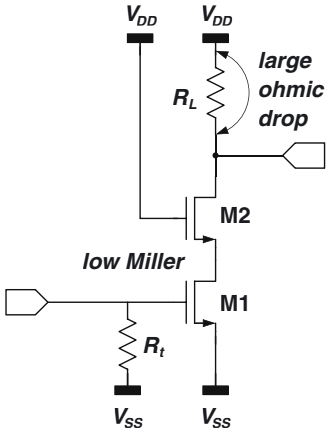
$$R_L < \frac{V_{DD} - V_{DS,sat,1} - V_{DS,sat,2}}{I_{DC}}, \tag{2.81}$$



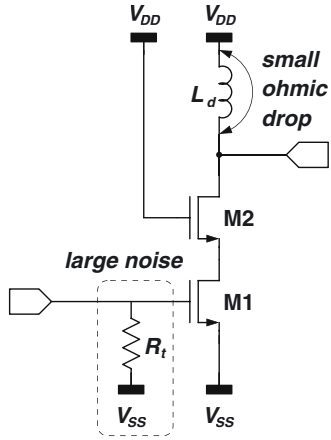
(a) Step 1.



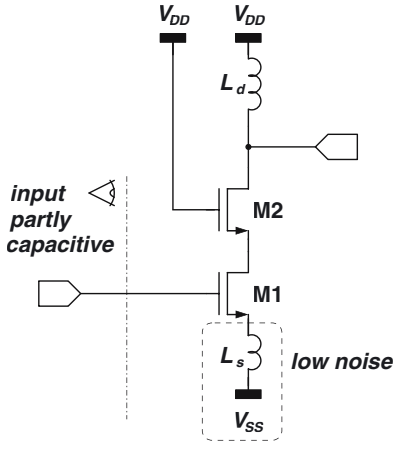
(b) Step 2.



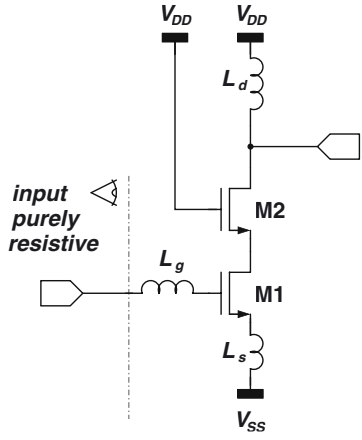
(c) Step 3.



(d) Step 4.



(e) Step 5.



(f) Step 6.

Figure 2.15: *Gradual evolution from basic CS amplifier to inductively degenerated CS LNA.*

which is usually limited to a few hundred Ohm. Suppose both $V_{DS,sat}$ are 0.25 V and V_{DD} is 1.5 V. The voltage gain of the circuit is then limited by

$$A_v < g_{m1}R_L = 2 \times \frac{V_{DD} - V_{DS,sat,1} - V_{DS,sat,2}}{V_{DS,sat,1}} = 8, \tag{2.82}$$

which corresponds to 18 dB.

One option to alleviate this is by using an active load element. If R_L is replaced by a PMOS transistor then the voltage drop over the load transistor can be limited by its own saturation voltage, which is a lot lower than the ohmic drop over a resistor with the same impedance. Disadvantages are the extra output capacitance, the need for an extra biasing node, the limited output swing, the fact that the gain is dependent on the r_{ds} of a transistor which is not accurately modelled and the non-linear behavior of the load transistor. The most important disadvantage is the noise introduced by the PMOS. The squared noise current of a load resistance is inversely proportional to its resistance,

$$\overline{i^2_{n,R_L}} = \frac{4kT\Delta f}{R_L}, \tag{2.83}$$

but the squared noise current of a load transistor M_p is proportional to its transconductance g_{mp},

$$\overline{i^2_{n,M_p}} = 4kT\gamma g_{mp}\Delta f, \tag{2.84}$$

which can be larger with a factor of 20 or more. This is unacceptable for a low noise amplifier.

Luckily another option is available for LNA's in wireless receivers. The LNA should only amplify in a relatively small frequency band around the carrier. Therefore, it is possible to replace R_L with an inductor L_d (see Fig. 2.15(d)). This inductor should go into parallel resonance with the parasitic capacitance on the output node precisely at the carrier frequency f_0. If the quality factor Q_L of the parallel tank was infinite, the equivalent load impedance and the gain at f_0 would be infinite. However due to the losses in the tank, the equivalent load resistance remains finite. Considering only the loss in the series resistance of the load inductor, the equivalent load resistance is given by

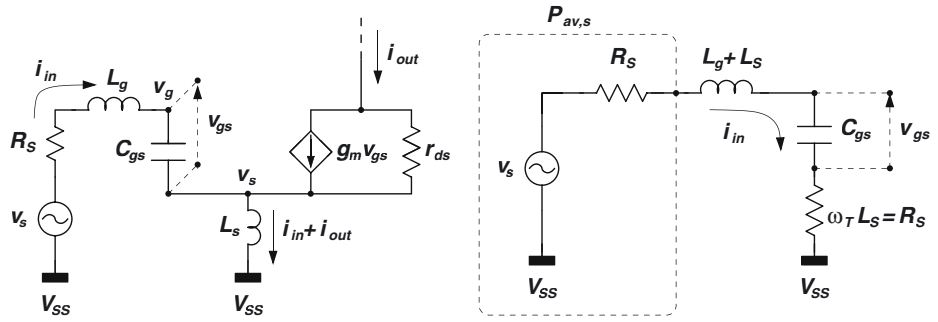
$$R_L = R_{L,s}(Q_L^2 + 1), \tag{2.85}$$

where $R_{L,s}$ is the series resistance of the load inductor and

$$Q_L = \frac{\omega_0 L_d}{R_{L,s}}. \tag{2.86}$$

Now, the DC voltage drop over the load is proportional to the DC series resistance of the inductor $R_{L,s}$ which is very small, in the order of a few Ohm. The gain is proportional to R_L which can be made large: a few hundred to over a thousand Ohm. The squared noise current from the load is inversely proportional to R_L. The noise factor of this circuit, neglecting the contribution of M2 is approximately

$$F \approx 2 + \frac{\gamma}{\alpha}\frac{4}{g_{m1}R_S} + \frac{4}{g_{m1}^2 R_S R_L}, \tag{2.87}$$



(a) Small-signal input schematic. (b) Equivalent input schematic.

Figure 2.16: *The origin of the resistive component in the input impedance of the CS LNA and the input quality factor Q_{in}.*

Real noise figure values for this type of amplifiers are in the order of 6 dB or more. This is too large for most applications. Another type of input matching is required.

Instead of adding a termination resistor, an inductor L_s is connected between the source of M1 and V_{SS} as shown in Fig. 2.15(e). Maybe somewhat unexpectedly, this creates a resistive part in the input impedance. Neglecting the gate-drain capacitance, the input impedance is given by

$$Z_{in} = \frac{1}{j\omega C_{gs}} + j\omega L_s + \omega_T L_s, \tag{2.88}$$

where the cut-off pulsation ω_T is defined by (2.40). At normal operating frequencies, the input still behaves capacitively due to C_{gs}. Therefore an extra gate resonance inductor L_g is connected in series with the input as indicated in Fig. 2.15(f). The input impedance is now

$$Z_{in} = \frac{1}{j\omega C_{gs}} + j\omega(L_g + L_s) + \omega_T L_s, \tag{2.89}$$

which reduces to

$$Z_{in} = \omega_T L_s, \tag{2.90}$$

at the operating frequency f_0 when

$$\omega_0(L_g + L_s) = \frac{1}{\omega_0 C_{gs}}. \tag{2.91}$$

In the design, L_s is chosen such that $\Re(Z_{in}) = \omega_T L_s = 50\ \Omega$. L_g is chosen according to (2.91) so that $\Im(Z_{in}) = 0$ at ω_0.

In what follows the subscript 1 for parameters relating to M1 will be left out for reasons of simplicity.

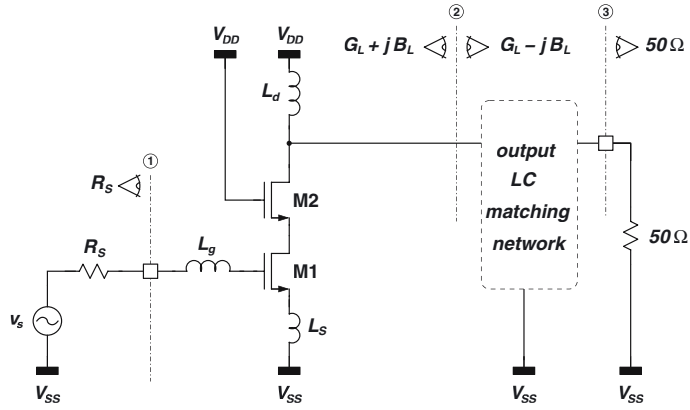


Figure 2.17: *Simplified schematic of the CS LNA matched at both input and output.*

2.6.1.2 Power Gain

Before calculating the power gain it is interesting to see where the resistive part of the input impedance comes from. Suppose a certain current i_{in} is flowing through the C_{gs} of M1 as depicted in Fig. 2.16(a). This will cause a v_{gs} which is lagging 90 degrees with respect to i_{in}. Hence, a current i_{out} will flow given by $g_m v_{gs}$ which is in phase with v_{gs} and also lagging 90 degrees with respect to i_{in}. This current (together with i_{in} itself) will flow through L_s giving rise to a voltage over L_s leading 90 degrees with respect to i_{out} and in phase with i_{in}:

$$\begin{aligned} v_s &= j\omega L_s(i_{in} + i_{out}) \\ &= i_{in}(j\omega L_s + \frac{g_m}{C_{gs}} L_s). \end{aligned} \tag{2.92}$$

Since v_s is part of v_{in} according to

$$v_{in} = j\omega L_g i_{in} + v_{gs} + v_s, \tag{2.93}$$

this in phase component is found also in Z_{in} given in (2.89).

For the derivation of the power gain it is assumed that both the input and output of the LNA are matched to 50 Ω. For the output this can be done with any lossless matching network as depicted in Fig. 2.17. Specific matching networks will be discussed in more detail in Section 4.8. The small signal equivalent of Fig. 2.17 is shown in Fig. 2.18. The available source power $P_{av,s}$ is given by

$$P_{av,s} = \frac{v_s^2}{4R_S}. \tag{2.94}$$

Since the input is matched, the input current i_{in} is

$$i_{in} = \frac{v_s}{2R_S} = \sqrt{\frac{P_{av,s}}{R_S}} \tag{2.95}$$

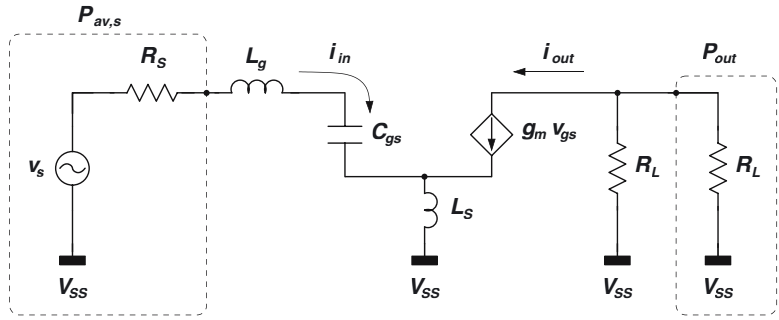


Figure 2.18: *Simplified small signal schematic of the LNA in Fig. 2.17.*

The input current gives rise to an output current according to

$$i_{out} = i_{in}\frac{\omega_T}{\omega_0}, \tag{2.96}$$

which is equivalent to the current gain β of a bipolar transistor. Since the output of the circuit is matched it is also matched at reference plane ② in Fig. 2.17. Both to the left and right of the reference plane, the current i_{out} sees a resistance R_L (also shown in Fig. 2.18). Thus only half of the output current is used to generate the output power. Consequently

$$P_{out} = \left(\frac{i_{out}}{2}\right)^2 R_L. \tag{2.97}$$

Equations (2.94) to (2.97) allow calculation of the power gain:

$$G_T = \frac{P_{out}}{P_{av,s}} = \frac{R_L}{4R_S}\left(\frac{\omega_T}{\omega_0}\right)^2. \tag{2.98}$$

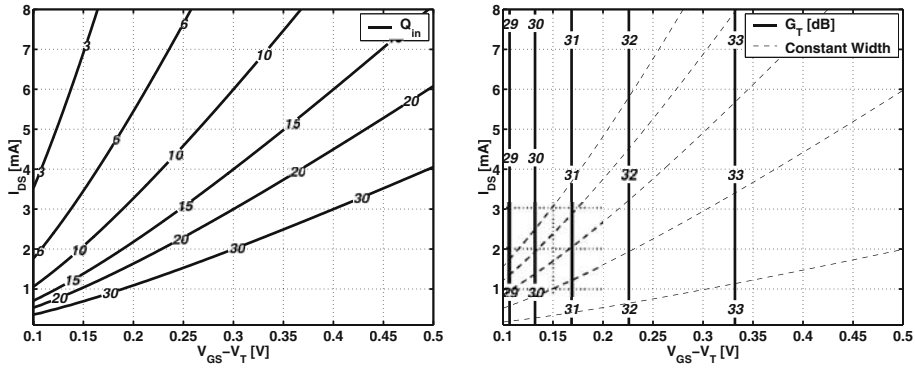
It is clear that the power gain increases with the load resistance. However practically the load resistance achievable with an on-chip inductor is limited as will be shown in Section 4.10.2. Notice also that the power gain goes up with increasing ω_T. As such, deeper submicron technologies automatically improve the gain of the LNA. Fig. 2.19(b) illustrates the behavior of the power gain in the M1 design space assuming a fixed load resistance of 500 Ω. This assumption is not entirely justified since a smaller M1 will usually result in a smaller output capacitance. This will allow a larger load inductor and hence a larger load resistance as will be shown in Section 4.8. For now, this is neglected. It is clearly seen in Fig. 2.19(b) that the gain is only dependent on $V_{GS} - V_T$ through ω_T.

We can rewrite (2.98) as

$$G_T = \frac{R_L}{4R_S}\left(\frac{\omega_T}{\omega_0}\right)^2 = Q_{in}^2 g_m^2 R_L R_S, \tag{2.99}$$

where

$$Q_{in} = \frac{v_{gs}}{v_s} = \frac{1}{2\omega_0 C_{gs} R_S}, \tag{2.100}$$



(a) $Q_{in}[\ ]$. (b) G_T [dB].

Figure 2.19: *Contours of the input quality factor and power gain of the CS LNA.*

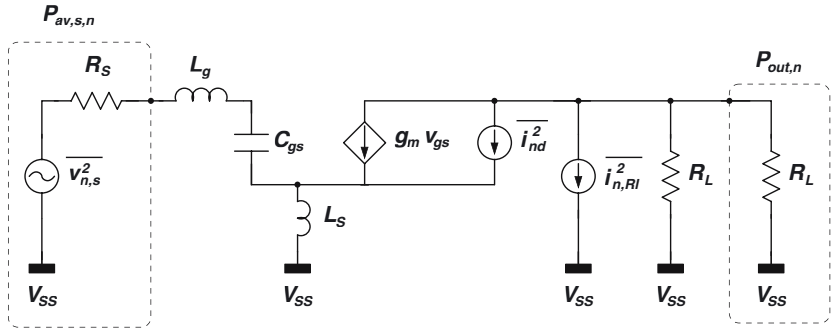
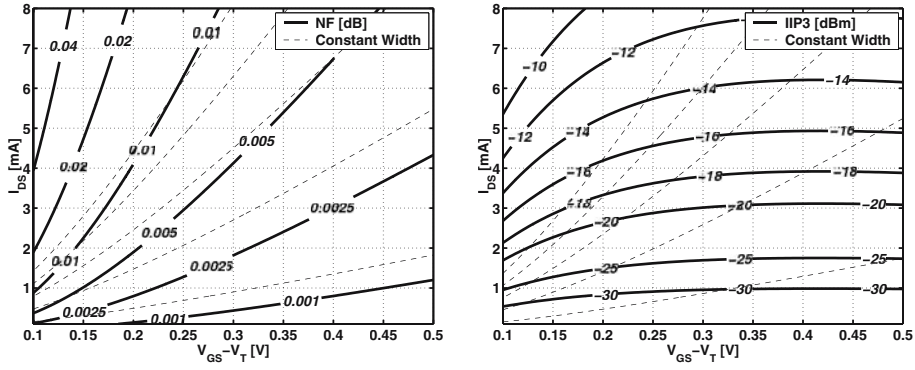


Figure 2.20: *Simplified small signal schematic of the CS LNA in Fig. 2.17 with noise sources.*

represents the quality factor of the series resonant input tank consisting of $2R_S$, $(L_g + L_s)$ and C_{gs} as indicated in Fig. 2.16(b). Contours of Q_{in} are shown in Fig. 2.19(a). This explains why G_T is only a function of $V_{GS} - V_T$. Equation (2.99) shows that the gain increases with g_m, which comes to no surprise. However, if g_m is increased for a constant $V_{GS} - V_T$ by increasing the width of M1, the gain remains constant. The reason is that the larger width and hence the larger C_{gs}, causes a proportional reduction in the quality factor of the input tank. This completely counters the increased g_m.

2.6.1.3 Noise Figure

For calculation of the noise factor of the LNA, only the classical channel noise of M1 and the thermal noise of R_L are considered as indicated in Fig. 2.20. The noise contribution of M2 and other parasitic noise sources are ignored. The derivation was based on the quasi static approximation and the influence of the cascode pole, the Miller-effect and other parasitics are neglected.



(a) NF [dB]. (b) IIP3 [dBm].

Figure 2.21: *Contours of the noise figure and IIP3 of the CS LNA.*

A more in depth discussion is given in Chapter 4 where also the influence of the main parasitics is covered. For now, the following expression approximates the noise factor of the amplifier:

$$\begin{aligned} F &\approx 1 + (F_d - 1) + (F_L - 1) \\ &\approx 1 + \frac{\gamma}{\alpha} g_m R_S \left(\frac{\omega_0}{\omega_T}\right)^2 + 4\left(\frac{\omega_0}{\omega_T}\right)^2 \frac{R_S}{R_L}, \end{aligned} \tag{2.101}$$

where $(F_d - 1)$ and $(F_L - 1)$ denote the contributions of transistor M1 and the equivalent load resistor R_L respectively. It might be surprising to find g_m in the numerator of the noise factor rather than in the denominator. However, $F_d - 1$ can be rewritten as

$$F_d - 1 = \frac{\gamma}{4\alpha g_m R_S} \cdot \frac{1}{Q_{in}^2}, \tag{2.102}$$

where Q_{in} is given by (2.100) and accounts for the input series resonance. In this equation g_m is found in the denominator as intuitively expected. This proportionality only holds for a fixed device width since then Q_{in} is a constant. The factor 4 in the denominator finds its origin in the fact that only half of the drain noise current finds its way to the output owing to the feedback inductor L_s. Note also that $(F_L - 1) = G_T^{-1}$ goes down with increasing R_L since an infinite R_L gives a zero noise current at the output. It also decreases with increasing $\frac{\omega_T}{\omega_0}$ since the gain of the circuit increases and the noise contributed by the load resistor becomes proportionally less important.

It is even more interesting to take a closer look at the behavior of $(F_d - 1)$ with regards to the dimensions of M1:

$$(F_d - 1) \approx \frac{\gamma}{\alpha} g_m R_S \left(\frac{\omega_0}{\omega_T}\right)^2 \propto \frac{g_m}{\omega_T^2}. \tag{2.103}$$

In order to illustrate the behavior of F_d a contour plot is depicted in Fig. 2.21(a). It is based on hand calculations with parameters extracted from the 0.25 μm Kawasaki technology in Table 2.1 and for an operating frequency of 1.57 GHz conform the primary GPS-band.

- For a fixed current through the device, (2.103) becomes

$$(F_d - 1) \propto \frac{1}{(V_{GS} - V_T)^3} \propto W^{3/2}. \tag{2.104}$$

  This shows that the noise figure decreases drastically with increasing $V_{GS} - V_T$, even if the current through the device does not change. This can be deduced also from Fig. 2.21(a) since a constant current is simply a horizontal line. Going to the right with increasing $V_{GS} - V_T$ yields a gradually lower noise figure.

- For a fixed device width,

$$(F_d - 1) \propto \frac{1}{(V_{GS} - V_T)} \propto \frac{1}{\sqrt{I}}. \tag{2.105}$$

  This means that for a given device if the bias voltage is increased, the noise figure improves due to the increase in ω_T^2 which is faster than the increase in g_m. This can easily be tested experimentally since after processing the device width tends to remain fixed. Contours for constant width are also represented in Fig. 2.21(a). Increasing either I_{DS} or $V_{GS} - V_T$ (following the contours from left to right) indeed shows a slowly decreasing noise figure.

- For a fixed overdrive voltage $V_{GS} - V_T$, the noise factor behaves as

$$(F_d - 1) \propto W \propto I. \tag{2.106}$$

  This is undoubtedly the strangest behavior; the noise figure increases with increasing current! It is illustrated graphically in Fig. 2.21(a) where the contours for constant $V_{GS} - V_T$ are just vertical lines. Increasing the current (following the vertical lines upward) indeed gives an increasing noise figure. This is very counterintuitive but can be understood as follows. A fixed $V_{GS} - V_T$ implies a fixed ω_T which means the current gain from (2.96) is also fixed and the squared output noise current due to the source is unchanged. However the squared output noise current from M1 is proportional to its g_m and hence proportional to the bias current for a fixed $V_{GS} - V_T$. Therefore the noise figure increases with increasing bias current.

As well $(F_L - 1)$ as $(F_d - 1)$ decrease with increasing ω_T/ω_0 which means that smaller technologies also improve the inherent classical noise contribution in the noise figure. However, the excess noise factor γ tends to increase for smaller gate lengths as mentioned in Section 2.4.2. The reason is that the electric field within the short channel increases causing an electron temperature notably larger than the lattice temperature. To decrease γ it is advised to lower the drain-source voltage in order to keep the electric field in the channel as low as possible. Increasing the gate length to reduce γ makes no sense since the increase in ω_T would far outweigh the benefits of a lower γ.

A final possibility for improvement appears to be given by the decrease of R_S (which would also increase the power gain as seen in (2.98). Of course, the actual source resistance is the output resistance of the antenna or the bandpass filter preceding the LNA and is consequently

fixed. However it is possible to use a lossless transformation network to synthesize an equivalent source resistance (analogous to the matching network at the output in Fig. 2.17) . Actually such a transformation network is inherently present albeit that it normally increases the equivalent source resistance. This will yield an even larger contribution of the classical channel noise in the noise figure (but it will reduce the contribution of the non-quasistatic induced gate noise). These issues will be clarified in Chapter 4.

2.6.1.4 Linearity

In Section 2.3.2 it was stated that for any NMOS in saturation, the third order intermodulated output current is equal to the fundamental output current ($\mathrm{IM3} = 1$) for an input voltage amplitude given by (2.47). This goes also for transistor M1 in Fig. 2.17. Consequently, referring to the small-signal equivalent in Fig. 2.18, the v_{gs} at which $\mathrm{IM3} = 1$ is given by (2.47):

$$\mathrm{IV3\ [V\ amp]} = \sqrt{\frac{4}{3}\frac{V_{GST}\left(2+\Theta V_{GST}\right)\left(1+\Theta V_{GST}\right)^2}{\Theta}}. \tag{2.107}$$

Equations (2.94) and (2.100) allow calculation of the available source power corresponding with this v_{gs}

$$\mathrm{IIP3\ [W]} = \frac{\mathrm{IV3}^2}{2\times 4Q_{in}^2 R_S} = \frac{\mathrm{IV3}^2\omega_0^2 C_{gs}^2 R_S}{2}. \tag{2.108}$$

Together with (2.107) this results in

$$\mathrm{IIP3\ [W]} = \frac{1}{8R_S}\frac{4}{3}\frac{V_{GST}\left(2+\Theta V_{GST}\right)\left(1+\Theta V_{GST}\right)^2}{\Theta}\frac{1}{Q_{in}^2}, \tag{2.109}$$

the available source power expressed in Watt. Converting this expression to [dBm] yields,

$$\mathrm{IIP3\ [dBm]} = 5.25 + 10\log\left(\frac{V_{GST}\left(2+\Theta V_{GST}\right)\left(1+\Theta V_{GST}\right)^2}{\Theta}\right) - 20\log\left(Q_{in}\right). \tag{2.110}$$

Substituting (2.100) in (2.110) gives

$$\mathrm{IIP3} = 5.25 + 10\log\left(\frac{V_{GST}\left(2+\Theta V_{GST}\right)\left(1+\Theta V_{GST}\right)^2}{\Theta}\right) + 20\log\left(2\omega_0 R_S C_{gs}\right), \tag{2.111}$$

The first term in this equation is fixed and consists of the product of the conversion factor for going from V to mW ($10\log(1000/(2\times 50))$ and the factor 4/3 in the IV3. The second term represents the IV3 of M1 itself while the last term is the quality factor of the input stage. This is logical indeed, if Q_{in} is higher then v_{gs} will be larger for the same input power, thus IIP3 decreases. However it should be noted that the output referred 3rd order intermodulation intercept point (OIP3) is independent of Q_{in}. It is given by

$$\mathrm{OIP3} = \mathrm{IIP3}\times G_T. \tag{2.112}$$

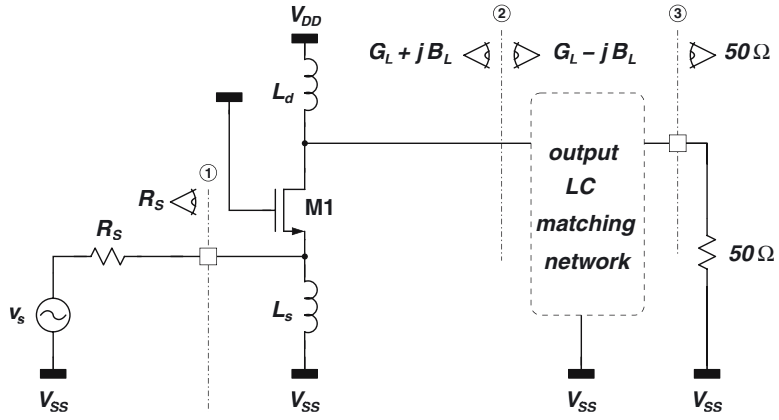


Figure 2.22: *Simplified schematic of the CG LNA matched at both input and output.*

substituting G_T by (2.99) results in

$$\mathrm{OIP3} = 5.25 + 10\log\left(\frac{V_{GST}\left(2+\Theta V_{GST}\right)\left(1+\Theta V_{GST}\right)^2}{\Theta}\right) + 10\log\left(g_m^2 R_L R_S\right), \quad (2.113)$$

independent of Q_{in}.

Fig. 2.21(b) plots the contours of IIP3 in the design space of M1. It is seen that the IIP3 is quite a weak function of $V_{GS}-V_T$ where for a simple NMOS it is only dependent on $V_{GS}-V_T$. The reason is the third term in (2.110) representing the input quality factor, defined by (2.100). Q_{in} is inversely proportional to the width of M1:

$$Q_{in} = \frac{1}{2\omega_0 C_{gs} R_S} \propto \frac{1}{W}. \quad (2.114)$$

Hence, when increasing $V_{GS}-V_T$, at a fixed current, the device width is inversely proportional to the square root of the overdrive voltage. Consequently the second term of (2.110) increases but the third term decreases. These effects almost cancel each other resulting in the relatively flat behavior of the contours. The only sure way to improve the linearity of the LNA is to increase the current consumption.

2.6.2 The Common-Gate LNA

The common-gate LNA, depicted schematically in Fig. 2.22 is the main competitor of the inductively degenerated common-source LNA. The reason is quite obviously the resistive input impedance. In the common-gate amplifier the signal is fed to the source of the input transistor rather than the gate. The input resistance is the inverse of the transconductance. Simply equating it to 50 Ω or 20 mS results in a correct impedance match. Actually, things are a bit more complicated but this basic principle is the reason for its popularity.

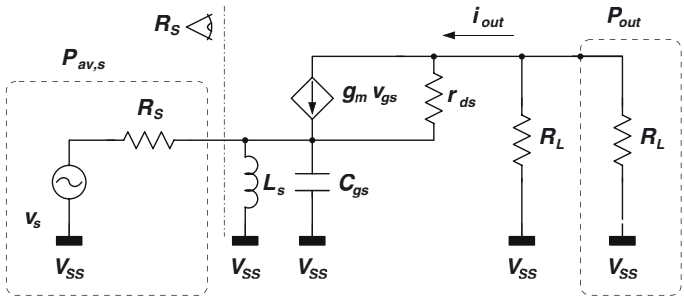


Figure 2.23: *Simplified small signal schematic of the CG LNA in Fig. 2.22.*

2.6.2.1 Input Matching

The small signal equivalent of the CG LNA is shown in Fig. 2.23. The following discussion is based on the classical MOS-model as described in Section 2.3.1. No non-quasistatic effects are taken into account at this point. The impedance seen at the input of the LNA (neglecting the input capacitance) is then

$$Z_{in} = \frac{1}{g_m + g_{mb}} \frac{2r_{ds} + R_L}{2r_{ds}}. \tag{2.115}$$

It can be set to equal the source resistance, R_S. For larger g_m, the required load resistance for acquiring a correct input impedance increases rapidly. Once M1 is fixed in its design space, the value of R_L for a 50 Ω input impedance can be found as:

$$R_L = 2\left(R_S\left(g_m + g_{mb}\right) - 1\right) r_{ds} = 2(n g_m R_S - 1) r_{ds}, \tag{2.116}$$

where

$$n = \frac{g_m + g_{mb}}{g_m}. \tag{2.117}$$

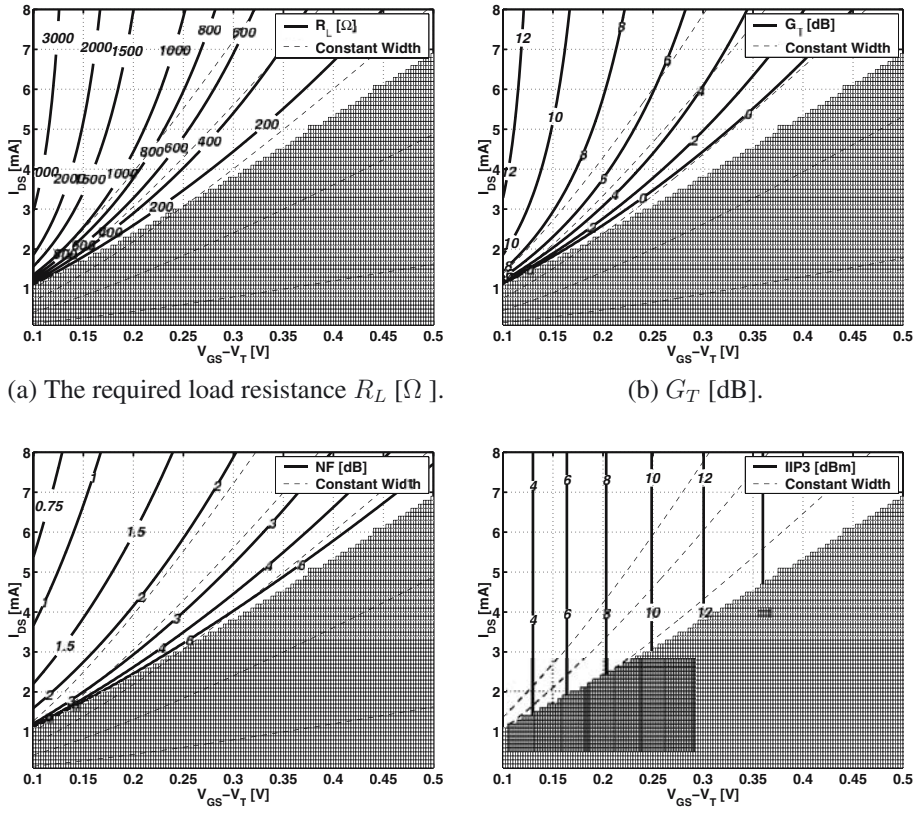
Substituting (2.41) in (2.116) yields

$$R_L = 2\frac{n g_m R_S - 1}{\Lambda I_{DS}} = \frac{4nR_S}{\Lambda\left(V_{gs} - V_T\right)} - \frac{2}{\Lambda I_{DS}}. \tag{2.118}$$

If the value of R_L, calculated from this expression becomes negative, an input match can no longer be realized without an extra matching network. Therefore it is required that

$$g_m > \frac{1}{nR_S} \approx 17\,\mathrm{mS}. \tag{2.119}$$

A serious disadvantage of the input impedance of the CG LNA is that it depends heavily on the value of r_{ds}. This resistance is not well known and not accurately modelled. An offset of 50% is quite possible. Consequently also the input impedance will show a large spread on process variations. This can be avoided by placing an extra resistor R_{dsx} between drain and



(a) The required load resistance R_L [Ω]. (b) G_T [dB].

(c) NF [dB]. (d) IIP3 [dBm].

Figure 2.24: *Contour plots illustrating the behavior of the CG LNA.*

source. An extra coupling capacitor can be placed in series with this resistor in order not to disturb the DC bias. This resistance will lower the spread of the input impedance on process variations. However, it will also reduce the gain and increase the noise figure and should not be done unless it is really necessary.

Equation (2.118) is illustrated graphically in Fig. 2.24(a). The patterned region indicates the part of the design space where no input match can be obtained. The required load resistance increases towards the upper left corner of the graph since this region features a low $V_{GS} - V_T$, increasing the positive term in (2.118), and a high current, reducing the negative term. Moreover, if g_m is significantly larger than proscribed by (2.119) the second term in (2.118) can be neglected making R_L a pure function of $V_{GS} - V_T$. This is seen in the steep behavior of the 1500 Ω contour. It should also be mentioned that for the common source LNA discussed previously, R_L was assumed constant and was set to 500 Ω. If higher values for R_L are possible then the gain of the CS LNA can be increased[3]. If not, the upper left region in Fig. 2.24 should be patterned since

[3]Care has to be taken since high gain values may compromise the stability of the amplifier.

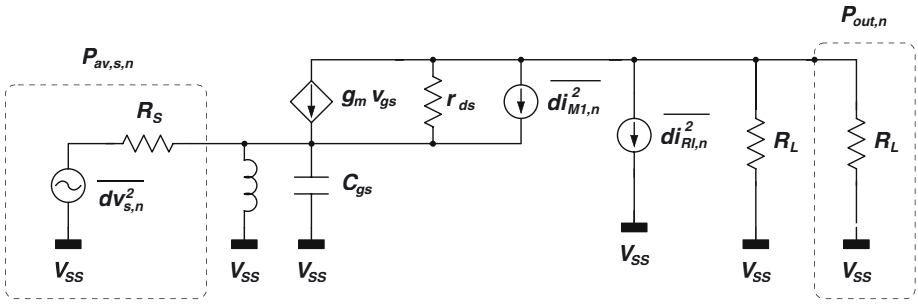


Figure 2.25: *Simplified small signal schematic of the CG LNA in Fig. 2.22 with noise sources.*

no input match can be obtained. $R_{L,max}$ is dependent on the technology. It depends on the size feature of the technology, but also on the amount of metal layers, their sheet resistance and the distance between them. It usually has an upper bound of 1 to 2 kilo Ohm. This will be discussed in more detail in Section 4.8.

2.6.2.2 Power Gain

Assuming the input of the LNA is matched, the output current, equal to the input current is given by

$$i_{out} = i_{in} = \sqrt{\frac{P_{av,s}}{R_S}} \tag{2.120}$$

A derivation similar to the one in Section 2.6.1.2 results in the following expression for the power gain:

$$G_T = \frac{R_L}{4R_S}. \tag{2.121}$$

Substituting (2.116) in (2.121) yields

$$G_T = \frac{1}{2}\left(ng_m - \frac{1}{R_S}\right) r_{ds}. \tag{2.122}$$

The result is plotted in Fig. 2.24(b). Comparing Fig. 2.24(a) and Fig. 2.24(b) clearly shows that the gain increase towards the upper left corner solely results from an increased load resistance. Also note that the power gain is much lower than for the CS LNA. Typical values are in the range of 5 to 10 dB compared to 25 dB for the CS LNA.

2.6.2.3 Noise Figure

Calculation of the noise factor of the CG LNA is based on the schematic in Fig. 2.25. Only the classical drain channel noise and the thermal noise of the load resistance are considered. The

noise factor is approximated by

$$\begin{aligned} F &\approx 1 + (F_d - 1) + (F_L - 1) \\ &\approx 1 + \frac{\gamma g_m}{\alpha} \frac{4 r_{ds}^2 R_S}{(2 r_{ds} + R_L)^2} + \frac{4 R_S}{R_L}, \end{aligned} \tag{2.123}$$

where $F_L - 1 = G_T^{-1}$. Substitution of (2.116) in (2.123) yields

$$F \approx 1 + \frac{\gamma}{n g_m R_S} + \frac{2 R_S}{(n g_m R_S - 1)\, r_{ds}}. \tag{2.124}$$

It is seen that $(F_d - 1)$ can theoretically be made arbitrarily low by simply increasing the transconductance. This is a big advantage of the CG amplifier. The performance of the CG LNA with respect to noise figure and gain is lower than that of the CS LNA at low frequencies. However for the CS LNA the excess noise is for the most part proportional to the square of the operation frequency and the power gain is inversely proportional to the square of the frequency. Consequently, the performance rapidly declines with increasing frequency. And at higher frequency the CG LNA performs better then its CS counterpart [Gua02]. Since this comparison requires a more detailed performance model taking into account non-quasistatic effects and several other parasitics and non-idealities it will be discussed in more detail in Section 4.11.

The noise figure of the CG LNA is presented graphically in Fig. 2.24(c). Similarly to the gain, also the noise figure improves towards the upper left, i.e. for larger g_m. This is easily understood from equation (2.124) since $(F_d - 1)$ is inversely proportional to g_m. It is quite interesting to see that the noise figure contours almost coincide with the contours for R_L. Intuitively this may be understood as follows. Since the input is matched, half of the source noise current is delivered to the output due to the nature of the matched CG stage. However, the channel noise current is not necessarily found at the output. Part of the current will simply run through r_{ds}. Suppose R_L were infinite then none of the channel noise current would find its way to the output; it would all flow through r_{ds} instead. Hence, a higher R_L will lead to a lower $(F_d - 1)$. Also $F_L - 1 = G_T^{-1}$ is only dependent on R_L.

2.6.2.4 Linearity

Analogously to the discussion in Section 2.6.1.4, the IIP3 of the CG LNA expressed in [dBm] is approximated by

$$\mathrm{IIP3} = 11.25 + 10 \log \left(\frac{V_{GST} \left(2 + \Theta V_{GST}\right) \left(1 + \Theta V_{GST}\right)^2}{\Theta} \right). \tag{2.125}$$

Compared to (2.110) the last term $(-20 \log(Q_{in}))$ is replaced with $+6$ dB since $v_{gs} = v_s/2$ instead of $v_{gs} = Q_{in} v_s$.

The result can be interpreted graphically by means of Fig. 2.24(d). Since there is no input quality factor, IIP3 is only function of $V_{GS} - V_T$ like for a simple MOS transistor. Indeed, the IIP3 contours in Fig. 2.24(d) are vertical lines. The absence of Q_{in} also allows much larger values for IIP3 compared to the CS LNA.

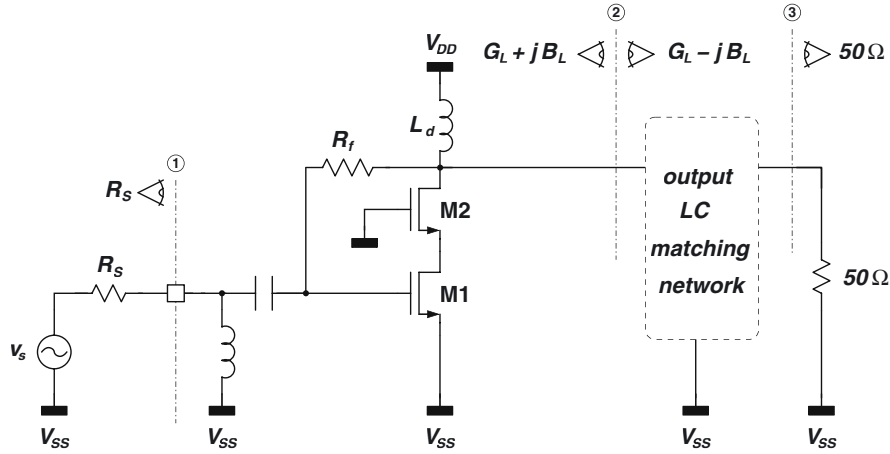


Figure 2.26: *Simplified schematic of the shunt feedback LNA.*

2.6.3 Shunt-Feedback Amplifier

A simplified schematic of a shunt-feedback LNA is shown in Fig. 2.26. The amplifier consists of a common source transistor with a cascode transistor on top to reduce the Miller effect. The resistor R_f is added to obtain a resistive part in the input impedance. The excess capacitance is tuned out with a parallel inductor, similar to the common-gate topology. If the feedforward through R_f can be neglected, then the input impedance is given by

$$Z_{in} = \frac{R_f}{1 + g_m \frac{R_{out}}{2}}, \tag{2.126}$$

where the output resistance R_{out} is found by

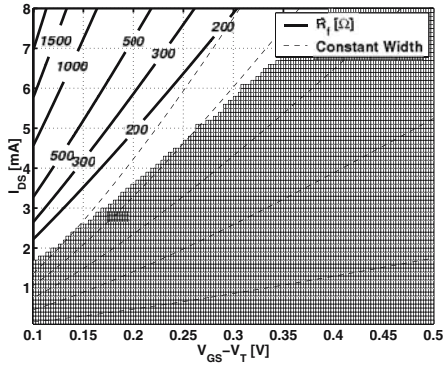
$$R_{out} = R_L \parallel R_f. \tag{2.127}$$

For a given g_m and R_L, R_f is chosen such that $Z_{in} = 50\ \Omega$. For $R_L = 1$ kΩ and a frequency of 1.5 GHz, the contours of R_f are plotted in the design space of the amplifying transistor in Fig. 2.27(a). The patterned area marks the region where the power gain is negative.

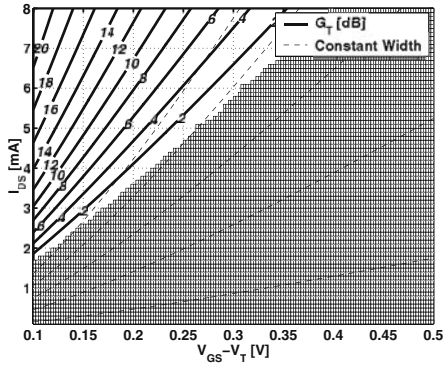
The power gain of the shunt-feedback LNA is calculated by

$$G_T = \frac{g_m^2 R_S R_L \parallel R_f}{4}. \tag{2.128}$$

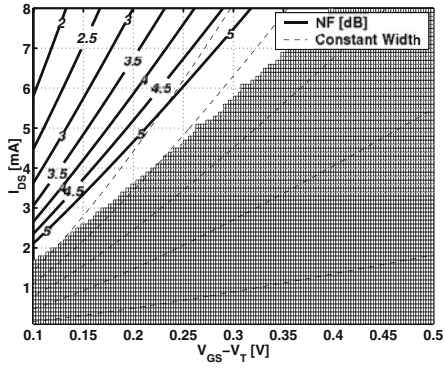
This is illustrated in Fig. 2.27(b). Indeed, the gain drops going to the lower right corner. In that region g_m is lowest and (2.126) implies that R_f needs to decrease in order to keep $Z_{in} = 50\Omega$. Thus the power gain drops even faster. A power gain of 15 to 20 dB is achievable depending on the power budget.



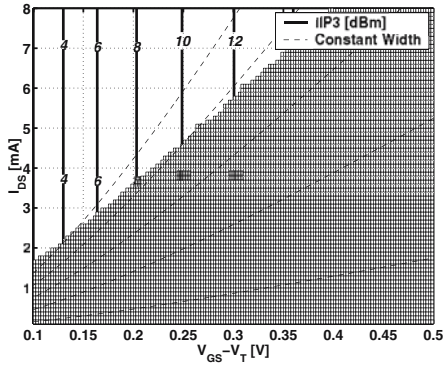
(a) The required feedback resistance R_f [Ω].



(b) G_T [dB].



(c) NF [dB].



(d) IIP3 [dBm].

Figure 2.27: *Contour plots illustrating the behavior of the shunt feedback LNA.*

The noise factor of the shunt-feedback LNA can be written as

$$F \approx 1 + \frac{R_S}{R_f} + \left(\frac{\gamma}{\alpha} g_m R_{out} + 1\right) G_T^{-1}. \tag{2.129}$$

The second term in this equation represents the contribution of the feedback resistor noise. It is usually the dominant contribution. Consequently, the noise figure can only be sufficiently low in the region where R_f is high according to (2.126). This is illustrated graphically in Fig. 2.27(c). Indeed, NF decreases towards the upper left. The noise performance is clearly inferior to both previous amplifiers.

The linearity of this type of amplifier is comparable to that of the common gate amplifier since their is no passive voltage amplification as with the tuned common-source amplifier. The

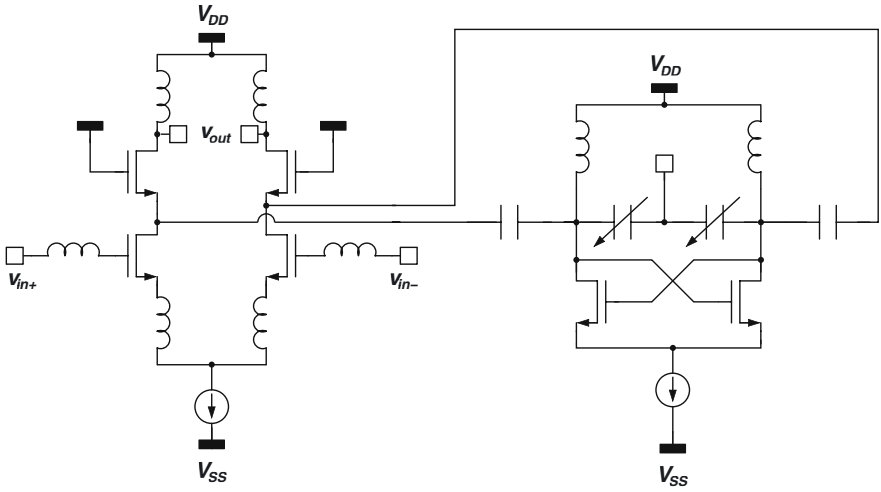


Figure 2.28: *Simplified schematic of an image reject LNA [Sam99].*

IIP3 is approximated by:

$$\mathrm{IIP3} = 11.25 + 10\log\left(\frac{V_{GST}\left(2+\Theta V_{GST}\right)\left(1+\Theta V_{GST}\right)^2}{\Theta}\right). \tag{2.130}$$

It is visualized in Fig. 2.27(d). This plot is identical to the common-gate amplifier. IIP3 is only function of $V_{GS} - V_T$.

One drawback of this circuit is the rather poor reverse isolation due to the direct connection of R_f to both input and output. It is in the order of 10 to 20 dB which is often too high for direct conversion receivers. Furthermore a small revers isolation will complicate the design for stability which becomes more problematic at high frequencies. Consequently this type of topology is only found at moderate frequencies.

An advantage of the shunt-feedback LNA is that, similar to the common-gate LNA, it can be used in baseband or wideband applications. The inductors used for tuning at the input and output are then left out. The output matching network is usually left out or replaced by an active buffer. Both amplifiers are for instance often used in optical receiver front-ends. These amplifiers are driven by a capacitive current source. They are intended to convert this current to a voltage and therefore in this context they are called transimpedance amplifiers.

In the remainder of this section, a few LNA designs or topologies will be discussed that don't completely fit within the foregoing classification.

2.6.4 Image Reject LNA's

Image reject LNA's have an extra functionality besides amplifying the wanted signal. They feature a sharp notch in their characteristic at the image frequency. The image frequency is

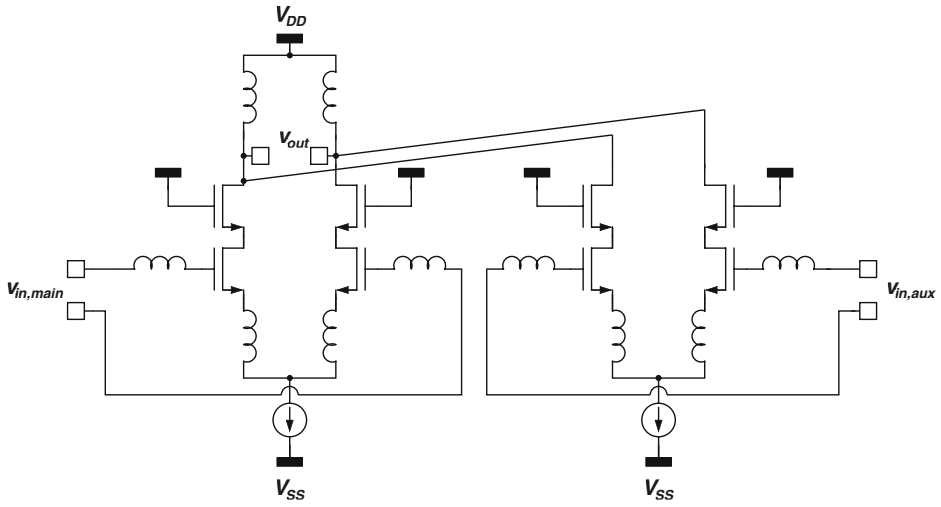


Figure 2.29: *LNA with feedforward cancellation of 3rd order distortion [Din01].*

located symmetrically to the wanted signal at the other side of the LO frequency. The notch is aimed at suppressing this image frequency.

One of the most interesting designs of an image reject LNA (and complete front-end) was done in [Sam99]. The basic schematic of the LNA + filter is shown in Fig. 2.28. The circuit was designed for a 5 GHz wireless LAN receiver and it was integrated within a complete RF front-end: LNA + image reject filter, PLL and mixer. The LNA is a differential, common-source amplifier with inductive degeneration. A series resonant tank is added at the cascode node, which shunts the signal to ground at the resonance frequency (the frequency of the image signal). The resonant tank that controls the notch is identical to the VCO in the PLL except for the larger supply current to sustain oscillation in the VCO. Further more, the center frequency of the notch is tuned by the same voltage that controls the VCO in the PLL. Consequently, locking the frequency of the VCO to the image frequency guarantees the suppression of the image frequency in the LNA.

2.6.5 Highly Linear Feedforward LNA

A highly linear LNA was presented at ISSCC in 2001 [Din01]. It achieves an IIP3 of 18 dBm through feedforward cancellation of the 3rd order terms. A schematic of the amplifier is shown in Fig. 2.29. It consists of a differential common-source amplifier with inductive source degeneration. A scaled copy of the amplifier is added in a feedforward configuration. This auxiliary amplifier is steered with a fraction $\beta > 1$ of the input signal. The output currents of both amplifiers are subtracted. The operation is as follows. Considering only 3rd order distortion, the

output signal of the main amplifier can be written as

$$y_{main}(x) = a_1 x \left(1 + \frac{a_3}{a_1} x^2\right), \tag{2.131}$$

where

$$a_3 = \frac{4}{3} \frac{a_1}{\mathrm{IIV3}^2_{main}}. \tag{2.132}$$

For the auxiliary amplifier the output is given by

$$y_{aux}(x) = a_{1,aux} \beta x \left(1 + \frac{a_{3,aux}}{a_{1,aux}} \beta^2 x^2\right), \tag{2.133}$$

where

$$\frac{a_{3,aux}}{a_{1,aux}} = \frac{a_3}{a_1}, \tag{2.134}$$

since it is a scaled copy. Now if

$$a_{1,aux} = \frac{a_1}{\beta^3}, \tag{2.135}$$

then

$$y(x) = y_{main}(x) - y_{aux}(x) = a_1 \left(1 - \frac{1}{\beta^2}\right) x, \tag{2.136}$$

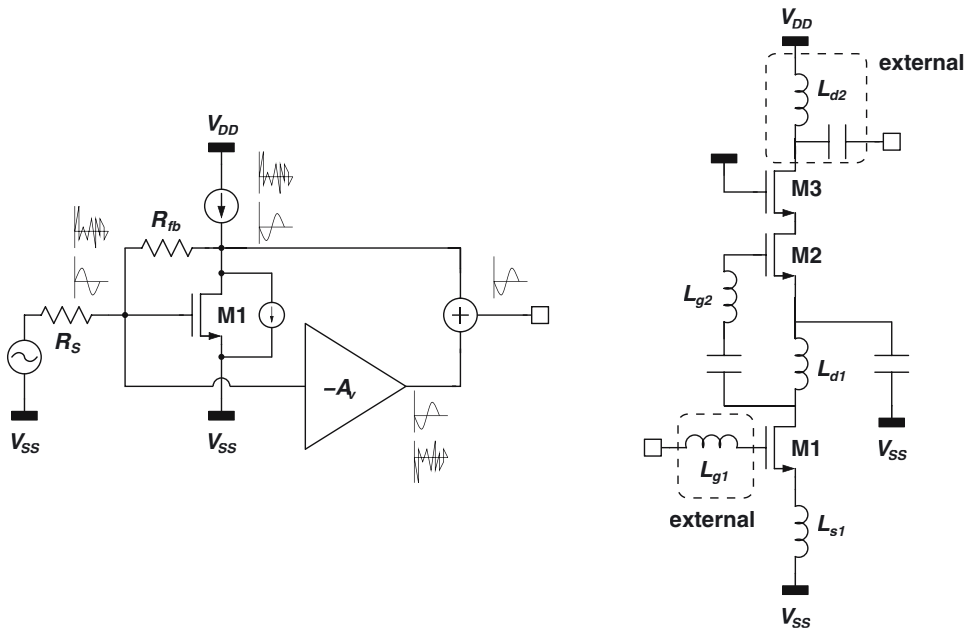
is a perfectly scaled version of x. Practically the linearity is limited by the mismatch between both amplifiers, higher order terms and non-linearity of the load impedance. A linearity improvement of 13 dB was achieved in [Din01]. However the comparison was done with an amplifier without feedforward but only using half the power. For the same power the improvement would be a few dB less.

2.6.6 The Noise-Cancelling Wide-band LNA

A very interesting LNA was introduced in [Bru02] at ISSCC 2002. The circuit is based on the shunt-feedback amplifier discussed in Section 2.6.3. The schematic is shown in Fig. 2.30(a). An extra inverting amplifier with voltage gain $-A_v$ is connected to the input. The output is added to the output of the shunt-feedback amplifier. The basic idea is that the input signal is 180 degrees out of phase with respect to the output signal while the input noise voltage is in phase with the output noise voltage. Consequently if the total input signal is fed forward through a (noiseless) inverting amplifier (with correct gain) and added to the output, then the noise of transistor M1 is completely cancelled while the signal is amplified.

2.6.7 Current Reuse LNA with Interstage Resonance

The current reuse topology implies that the LNA consists of two amplifier stages and only one current branch. A 5.2 GHz LNA employing this technique was published at ESSCIRC 2002 [Cha02, Cha03]. The basic circuit schematic is shown in Fig. 2.30(b). Operation is as follows.



(a) [Bru02] (b) [Cha03]

Figure 2.30: *(a) Schematic of the noise-cancelling wide-band LNA [Bru02] and (b) schematic of the current reuse LNA with interstage resonance [Cha03] (b).*

Transistor M1 is a common source amplifier with inductive degeneration and a tuned load impedance. The voltage at the drain of M1 is passively amplified by the series resonance network consisting of L_{g2} and C_{gs2}. This v_{gs2} is then converted into a current by M2 and via cascode M3 the current is dumped in a second tuned load impedance.

2.6.8 Transformer Feedback LNA

Another design which is more than interesting is a transformer based common source LNA [Cas03]. A schematic of the circuit is shown in Fig. 2.31. The LNA was designed for application in a 5 GHz wireless LAN receiver. The circuit is fully differential and uses magnetic coupling between the input and output to reduce the Miller effect. This is achieved by interwinding the source and drain inductors in both branches. The Circuit is matched at the input and achieves a power gain of 14.2 dB. The noise figure of 0.9 dB is the lowest value published for a CMOS amplifier at this frequency.

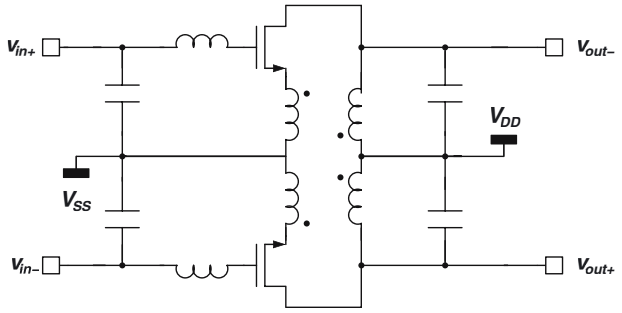


Figure 2.31: *Simplified schematic of the transformer feedback LNA [Cas03].*

2.7 Conclusion

In this chapter different concepts and topics, relevant to RF CMOS design have been introduced. A manageable MOS transistor model has been introduced that is able to reproduce the typical behavior of the device during design optimization. This model is closely related to Hspice MOS model level 3. The MOS model used in simulations has been extended with 10 extra resistors and capacitors to obtain more realistic simulation results with respect to noise, gain and linearity. Also the relevant non-quasistatic effect at the gate of the MOS device is modelled by this. All extra components are calculated taking into account the physical layout of the devices.

After a study of the different CMOS noise sources, the functionality of the LNA within the receiver chain has been studied. Based on the twoport cascading theory the conclusion was drawn that the LNA sets a minimum on the attainable noise figure of the complete receiver. Similarly, it places an upper bound on the attainable IIP3. The importance of the power gain and voltage gain has also been demonstrated. The value is determined by the dynamic range of the subsequent mixer stage. Also other LNA requirements have been investigated and placed in perspective: matching, reverse isolation, and stability.

The last section was devoted to the introduction of the most common LNA topologies. Even though no parasitic effects have yet been considered already a few interesting conclusions could be drawn. The common-source LNA with inductive degeneration seems to yield the best performance with respect to noise figure and gain. Unfortunately the noise figure increases rapidly with rising frequency. The common gate topology does not suffer from this effect. Moreover, the noise figure of the CG LNA can be improved continuously by increasing the current consumption. It may therefore become interesting to shift to this topology when the frequency and power budget are large. Both the CG amplifier and the shunt-feedback amplifier have the additional ability to handle baseband signals. This chapter concluded with a brief discussion of a few specific and interesting designs published recently.