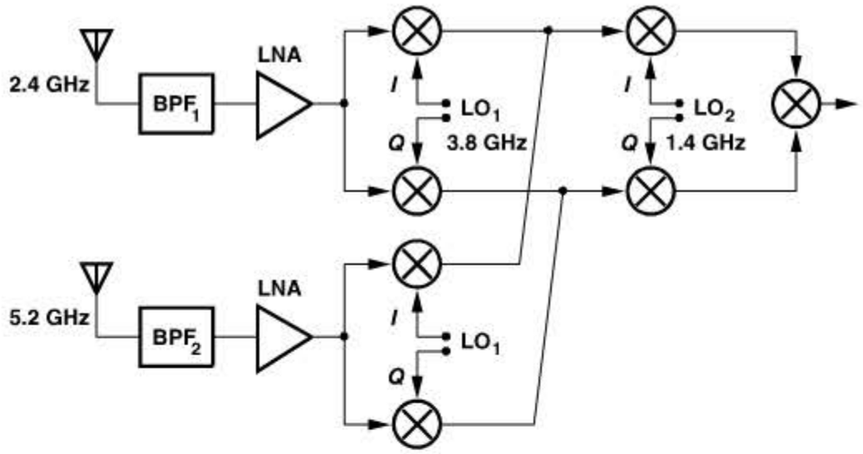


Figure 4.121 *Dual-band RX.*

an SNR of 20 dB is required for the signal to be detected properly. The Weaver architecture provides an image rejection ratio of 45 dB.

(a) Suppose the receiver must detect a −85-dBm signal in the 2.4-GHz mode while receiving at the same antenna a −10-dBm 5.2-GHz component as well. Determine the amount of rejection required of BPF_1 at 5.2 GHz.

(b) Suppose the receiver operates in the 5.2-GHz band but it also picks up a strong component at 7.2 GHz. It is possible for this component to be mixed with the third harmonics of LO_1 and LO_2 and appear in the baseband. Does the Weaver architecture prohibit this phenomenon? Explain in detail.

4.28. Consider the single-sideband mixer shown in Fig. 4.122. In the ideal case, the output has only one component at $\omega_1 + \omega_2$. Now suppose the ports sensing ω_2 suffer from third- and fifth-order nonlinearity. Plot the output spectrum if (a) $\omega_1 > 3\omega_2$ or (b) $\omega_1 < 3\omega_2$. Identify the frequency of each component.

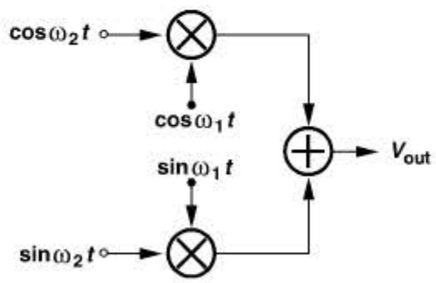


Figure 4.122 *SSB mixer.*

4.29. Explain why injection pulling is more serious in Fig. 4.114(b) than in Fig. 4.114(a).

CHAPTER

5

LOW-NOISE AMPLIFIERS

Following our system- and architecture-level studies in previous chapters, we move farther down to the circuit level in this and subsequent chapters. Beginning with the receive path, we describe the design of low-noise amplifiers. While our focus is on CMOS implementations, most of the concepts can be applied to other technologies as well. The outline of the chapter is shown below.

Basic LNA Topologies	Alternative LNA Topologies	Nonlinearity of LNAs
▪ CS Stage with Inductive Load	▪ Variants of CS LNA	▪ Nonlinearity Calculations
▪ CS Stage with Resistive Feedback	▪ Noise–Cancelling LNAs	▪ Differential and Quasi–Differential LNAs
▪ CG Stage	▪ Differential LNAs	
▪ CS Stage with Inductive Degeneration		

5.1 GENERAL CONSIDERATIONS

As the first active stage of receivers, LNAs play a critical role in the overall performance and their design is governed by the following parameters.

Noise Figure The noise figure of the LNA directly adds to that of the receiver. For a typical RX noise figure of 6 to 8 dB, it is expected that the antenna switch or duplexer contributes about 0.5 to 1.5 dB, the LNA about 2 to 3 dB, and the remainder of the chain about 2.5 to 3.5 dB. While these values provide a good starting point in the receiver design, the exact partitioning of the noise is flexible and depends on the performance of each stage in the chain. In modern RF electronics, we rarely design an LNA in isolation. Rather, we view and design the RF chain as one entity, performing many iterations among the stages.

To gain a better feel for a noise figure of 2 dB, consider the simple example in Fig. 5.1(a), where the noise of the LNA is represented by only a voltage source. Rearranging

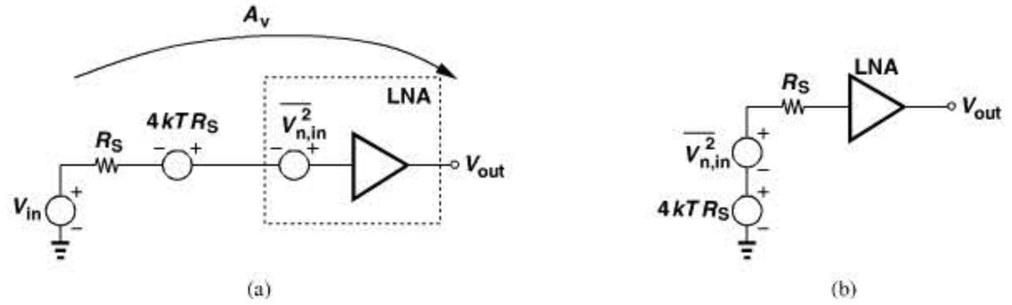


Figure 5.1 *(a) LNA with input-referred noise voltage, (b) simplified circuit.*

the input network as shown in Fig. 5.1(b), we have from Chapter 2

$$\text{NF} = \frac{\overline{V_{n,out}^2}}{A_v^2} \cdot \frac{1}{4kTR_S} \tag{5.1}$$

$$= 1 + \frac{\overline{V_{n,in}^2}}{4kTR_S}. \tag{5.2}$$

Thus, a noise figure of 2 dB with respect to a source impedance of 50 Ω translates to $\sqrt{\overline{V_{n,in}^2}} = 0.696\ \text{nV}/\sqrt{\text{Hz}}$, an extremely low value. For the gate-referred thermal noise voltage of a MOSFET, $4kT\gamma/g_m$, to reach this value, the g_m must be as high as $(29\,\Omega)^{-1}$ (if $\gamma = 1$). In this chapter, we assume $R_S = 50\,\Omega$.

Example 5.1

A student lays out an LNA and connects its input to a pad through a metal line 200 μm long. In order to minimize the input capacitance, the student chooses a width of 0.5 μm for the line. Assuming a noise figure of 2 dB for the LNA and a sheet resistance of 40 mΩ/□ for the metal line, determine the overall noise figure. Neglect the input-referred noise current of the LNA.

Solution:

We draw the equivalent circuit as shown in Fig. 5.2, pretending that the line resistance, R_L, is part of the LNA. The total input-referred noise voltage of the circuit inside the box is

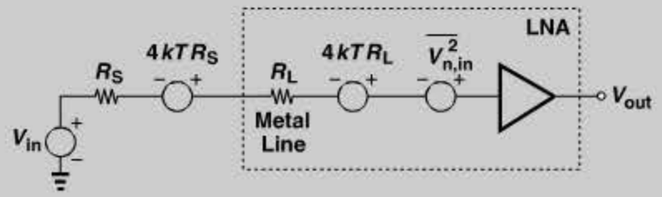


Figure 5.2 *LNA with metal resistance in series with its input.*

Example 5.1 *(Continued)*

therefore equal to $\overline{V_{n,in}^2} + 4kTR_L$. We thus write

$$\text{NF}_{\text{tot}} = 1 + \frac{\overline{V_{n,in}^2} + 4kTR_L}{4kTR_S} \tag{5.3}$$

$$= 1 + \frac{\overline{V_{n,in}^2}}{4kTR_S} + \frac{R_L}{R_S} \tag{5.4}$$

$$= \text{NF}_{\text{LNA}} + \frac{R_L}{R_S}, \tag{5.5}$$

where NF_{LNA} denotes the noise figure of the LNA without the line resistance. Since $\text{NF}_{\text{LNA}} = 2\ \text{dB} \equiv 1.58$ and $R_L = (200/0.5) \times 40\ \text{m}\Omega/\square = 16\,\Omega$, we have

$$\text{NF}_{\text{tot}} = 2.79\ \text{dB}. \tag{5.6}$$

The point here is that even small amounts of line or gate resistance can raise the noise figure of LNAs considerably.

The low noise required of LNAs limits the choice of the circuit topology. This often means that *only one transistor*—usually the input device—can be the dominant contributor to NF, thus ruling out configurations such as emitter or source followers.

Gain The gain of the LNA must be large enough to minimize the noise contribution of subsequent stages, specifically, the downconversion mixer(s). As described in Chapter 2, the choice of this gain leads to a compromise between the noise figure and the linearity of the receiver as a higher gain makes the nonlinearity of the subsequent stages more pronounced. In modern RF design, the LNA directly drives the downconversion mixer(s) with no impedance matching between the two. Thus, it is more meaningful and simpler to perform the chain calculations in terms of the voltage gain—rather than power gain—of the LNA.

It is important to note that the noise and IP_3 of the stage following the LNA are divided by *different* LNA gains. Consider the LNA/mixer cascade shown in Fig. 5.3(a), where the input-referred noise voltages are denoted by $\overline{V_{n,LNA}^2}$ and $\overline{V_{n,mixer}^2}$ and input noise currents

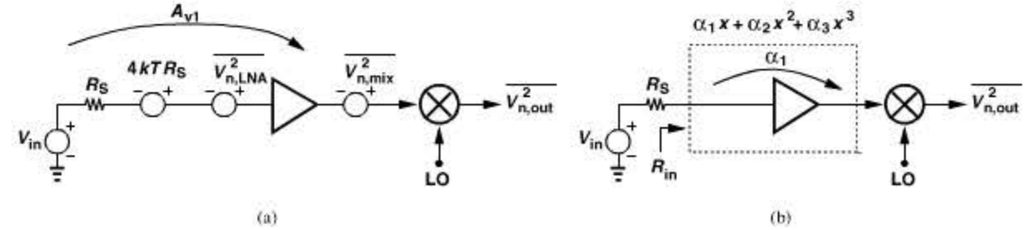


Figure 5.3 *Appropriate choice of gain for referring (a) noise and (b) IP_3 of a mixer to LNA input.*

are neglected. Assuming a unity voltage gain for the mixer for simplicity, we write the total output noise as $A_{v1}^2(\overline{V_{n,LNA}^2} + 4kTR_S) + \overline{V_{n,mix}^2}$. The overall noise figure is thus equal to

$$\mathrm{NF_{tot}} = \frac{A_{v1}^2(\overline{V_{n,LNA}^2} + 4kTR_S) + \overline{V_{n,mix}^2}}{A_{v1}^2}\frac{1}{4kTR_S} \tag{5.7}$$

$$= \mathrm{NF_{LNA}} + \frac{\overline{V_{n,mix}^2}}{A_{v1}^2}\cdot\frac{1}{4kTR_S}. \tag{5.8}$$

In other words, for NF calculations, the noise of the second stage is divided by the gain from the input voltage source to the LNA output.

Now consider the same cascade repeated in Fig. 5.3(b) with the nonlinearity of the LNA expressed as a third-order polynomial. From Chapter 2, we have[1]

$$\frac{1}{\mathrm{IP}_{3,\mathrm{tot}}^2} = \frac{1}{\mathrm{IP}_{3,\mathrm{LNA}}^2} + \frac{\alpha_1^2}{\mathrm{IP}_{3,\mathrm{mixer}}^2}. \tag{5.9}$$

In this case, α_1 denotes the voltage gain from the *input of the LNA* to its output. With input matching, we have $R_{in} = R_S$ and $\alpha_1 = 2A_{v1}$. That is, the mixer noise is divided by the *lower* gain and the mixer $\mathrm{IP_3}$ by the *higher* gain—both against the designer's wish.

Input Return Loss The interface between the antenna and the LNA entails an interesting issue that divides analog designers and microwave engineers. Considering the LNA as a *voltage* amplifier, we may expect that its input impedance must ideally be infinite. From the noise point of view, we may precede the LNA with a transformation network to obtain minimum NF. From the signal *power* point of view, we may realize *conjugate* matching between the antenna and the LNA. Which one of these choices is preferable?

We make the following observations. (1) the (off-chip) band-select filter interposed between the antenna and the LNA is typically designed and characterized as a high-frequency device and with a standard termination of 50 Ω. If the load impedance seen by the filter (i.e., the LNA input impedance) deviates from 50 Ω significantly, then the pass-band and stopband characteristics of the filter may exhibit loss and ripple. (2) Even in the absence of such a filter, the antenna itself is designed for a certain real load impedance, suffering from uncharacterized loss if its load deviates from the desired real value or contains an imaginary component. Antenna/LNA co-design could improve the overall performance by allowing even non-conjugate matching, but it must be borne in mind that, if the antenna is shared with the transmitter, then its impedance must contain a negligible imaginary part so that it *radiates* the PA signal. (3) In practice, the antenna signal must travel a considerable distance on a printed-circuit board before reaching the receiver. Thus, poor matching at the RX input leads to significant reflections, an uncharacterized loss, and possibly voltage attenuation. For these reasons, the LNA is designed for a 50-Ω resistive input impedance. Since none of the above concerns apply to the other interfaces within the RX (e.g., between the LNA and the mixer or between the LO and the mixer), they are typically designed to maximize *voltage* swings rather than power transfer.

1. The $\mathrm{IM_3}$ components arising from second-order terms are neglected.

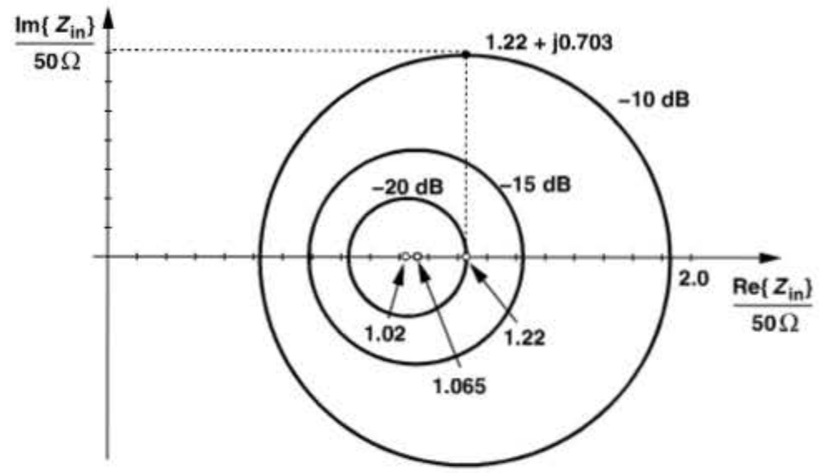


Figure 5.4 *Constant-Γ contours in the input impedance plane.*

The quality of the input match is expressed by the input "return loss," defined as the reflected power divided by the incident power. For a source impedance of R_S, the return loss is given by[2]

$$\Gamma = \left|\frac{Z_{in} - R_S}{Z_{in} + R_S}\right|^2, \tag{5.10}$$

where Z_{in} denotes the input impedance. An input return loss of −10 dB signifies that one-tenth of the power is reflected—a typically acceptable value. Figure 5.4 plots contours of constant Γ in the Z_{in} plane. Each contour is a circle with its center shown. For example, $Re\{Z_{in}\} = 1.22 \times 50\,\Omega = 61\,\Omega$ and $Im\{Z_{in}\} = 0.703 \times 50\,\Omega = 35.2\,\Omega$ yield $S_{11} = -10\,\mathrm{dB}$. In Problem 5.1, we derive the equations for these contours. We should remark that, in practice, a Γ of about −15 dB is targeted so as to allow margin for package parasitics, etc.

Stability Unlike the other circuits in a receiver, the LNA must interface with the "outside world," specifically, a poorly-controlled source impedance. For example, if the user of a cell phone wraps his/her hand around the antenna, the antenna impedance changes.[3] For this reason, the LNA must remain stable for all source impedances at *all frequencies*. One may think that the LNA must operate properly only in the frequency band of interest and not necessarily at other frequencies, but if the LNA begins to oscillate at any frequency, it becomes highly nonlinear and its gain is very heavily compressed.

A parameter often used to characterize the stability of circuits is the "Stern stability factor," defined as

$$K = \frac{1 + |\Delta|^2 - |S_{11}|^2 - |S_{22}|^2}{2|S_{21}||S_{12}|}, \tag{5.11}$$

2. Note that Γ is sometimes defined as $(Z_{in} - R_S)/(Z_{in} + R_S)$, in which case it is expressed in decibels by computing 20 log Γ (rather than 10 log Γ).

3. In the presence of a front-end band-select filter, the LNA sees smaller changes in the source impedance.

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

Example 5.2

A cascade stage exhibits a high reverse isolation, i.e., $S_{12} \approx 0$. If the output impedance is relatively high so that $S_{22} \approx 1$, determine the stability conditions.

Solution:

With $S_{12} \approx 0$ and $S_{22} \approx 1$,

$$K \approx \frac{1 - |S_{22}|^2}{2|S_{21}||S_{12}|} > 1 \tag{5.12}$$

and hence

$$|S_{21}| < \frac{1 - |S_{22}|^2}{2|S_{12}|}. \tag{5.13}$$

In other words, the forward gain must not exceed a certain value. For $\Delta < 1$, we have

$$S_{11} < 1, \tag{5.14}$$

concluding that the input *resistance* must remain positive.

The above example suggests that LNAs can be stabilized by maximizing their reverse isolation. As explained in Section 5.3, this point leads to two robust LNA topologies that are naturally stable and hence can be optimized for other aspects of their performance with no stability concerns. A high reverse isolation is also necessary for suppressing the LO leakage to the input of the LNA.

LNAs may become unstable due to ground and supply parasitic inductances resulting from the packaging (and, at frequencies of tens of gigahertz, the on-chip line inductances). For example, if the gate terminal of a common-gate transistor sees a large series inductance, the circuit may suffer from substantial feedback from the output to the input and become unstable at some frequency. For this reason, precautions in the design and layout as well as accurate package modeling are essential.

Linearity In most applications, the LNA does not limit the linearity of the receiver. Owing to the cumulative gain through the RX chain, the latter stages, e.g., the baseband amplifiers or filters tend to limit the overall input IP_3 or P_{1dB}. We therefore design and optimize LNAs with little concern for their linearity.

An exception to the above rule arises in "full-duplex" systems, i.e., applications that transmit and receive simultaneously (and hence incorporate FDD). Exemplified by the

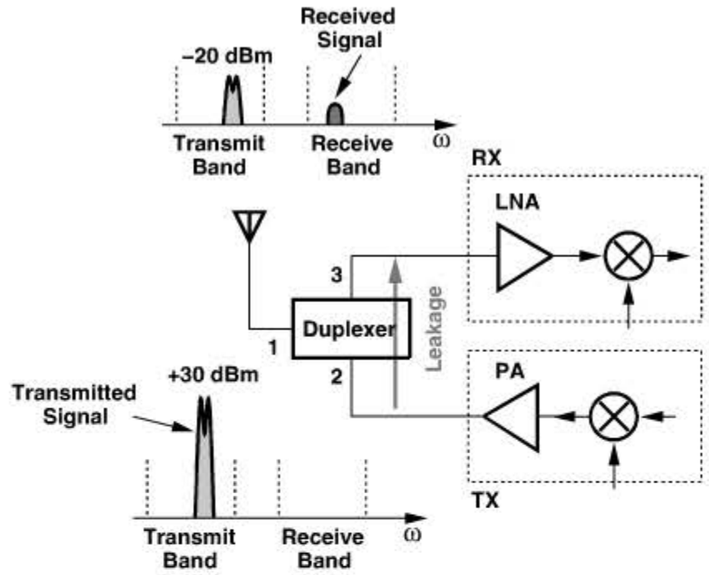


Figure 5.5 *TX leakage to RX in a full-duplex system.*

CDMA systems studied in Chapter 3, full-duplex operation must deal with the leakage of the strong transmitted signal to the receiver. To understand this issue, let us consider the front end shown in Fig. 5.5, where a duplexer separates the TX and RX bands. Modeling the duplexer as a three-port network, we note that S_{31} and S_{21} represent the losses in the RX and TX paths, respectively, and are about 1 to 2 dB. Unfortunately, leakages through the filter and the package yield a finite isolation between ports 2 and 3, as characterized by an S_{32} of about -50 dB. In other words, if the PA produces an average output power of $+30$ dBm (1 W), then the LNA experiences a signal level of -20 dBm in the TX band while sensing a much smaller received signal. Since the TX signal exhibits a variable envelope, its peak level may be about 2 dB higher. Thus, the receiver must remain uncompressed for an input level of -18 dBm. We must therefore choose a P_{1dB} of about -15 dBm to allow some margin.

Such a value for P_{1dB} may prove difficult to realize in a receiver. With an LNA gain of 15 to 20 dB, an input of -15 dBm yields an output of 0 to $+5$ dBm (632 to 1124 mV_{pp}), possibly compressing the LNA at *its output*. The LNA linearity is therefore critical. Similarly, the 1-dB compression point of the downconversion mixer(s) must reach 0 to $+5$ dBm. (The corresponding mixer IP_3 is roughly $+10$ to $+15$ dBm.) Thus, the mixer design also becomes challenging. For this reason, some CDMA receivers interpose an off-chip filter between the LNA and the mixer(s) so as to remove the TX leakage [1].

The linearity of the LNA also becomes critical in wideband receivers that may sense a large number of strong interferers. Examples include "ultra-wideband" (UBW), "software-defined," and "cognitive" radios.

Bandwidth The LNA must provide a relatively flat response for the frequency range of interest, preferably with less than 1 dB of gain variation. The LNA -3-dB bandwidth must therefore be substantially larger than the actual band so that the roll-off at the edges remains below 1 dB.

In order to quantify the difficulty in achieving the necessary bandwidth in a circuit, we often refer to its "fractional bandwidth," defined as the total −3-dB bandwidth divided by the center frequency of the band. For example, an 802.11g LNA requires a fractional bandwidth greater than 80 MHz/2.44 GHz = 0.0328.

Example 5.3

An 802.11a LNA must achieve a −3-dB bandwidth from 5 GHz to 6 GHz. If the LNA incorporates a second-order *LC* tank as its load, what is the maximum allowable tank Q?

Solution:

As illustrated in Fig. 5.6, the fractional bandwidth of an *LC* tank is equal to $\Delta\omega/\omega_0 = 1/Q$. Thus, the Q of the tank must remain less than 5.5 GHz/1 GHz = 5.5.

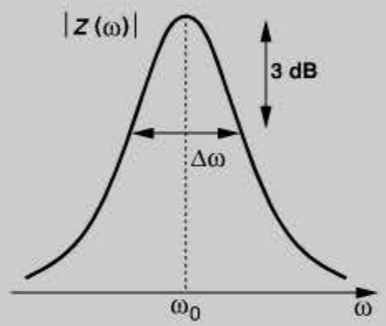


Figure 5.6 *Relationship between bandwidth and Q of a tank.*

LNA designs that must achieve a relatively large fractional bandwidth may employ a mechanism to *switch* the center frequency of operation. Depicted in Fig. 5.7(a) is an

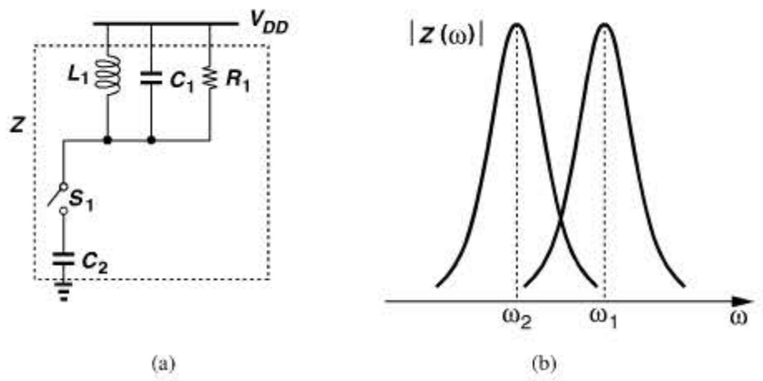


Figure 5.7 *(a) Band switching, (b) resulting frequency response.*

example, where an additional capacitor, C_2, can be switched into the tank, thereby changing the center frequency from $\omega_1 = 1/\sqrt{L_1 C_1}$ to $\omega_2 = 1/\sqrt{L_1(C_1 + C_2)}$ [Fig. 5.7(b)]. We return to this concept in Section 5.5.

Power Dissipation The LNA typically exhibits a direct trade-off among noise, linearity, and power dissipation. Nonetheless, in most receiver designs, the LNA consumes only a small fraction of the overall power. In other words, the circuit's noise figure generally proves much more critical than its power dissipation.

5.2 PROBLEM OF INPUT MATCHING

As explained in Section 5.1, LNAs are typically designed to provide a 50-Ω input resistance and negligible input reactance. This requirement limits the choice of LNA topologies. In other words, we cannot begin with an arbitrary configuration, design it for a certain noise figure and gain, and then decide how to create input matching.

Let us first consider the simple common-source stage shown in Fig. 5.8, where C_F represents the gate-drain overlap capacitance. At very low frequencies, R_D is much smaller than the impedances of C_F and C_L and the input impedance is roughly equal to $[(C_{GS} + C_F)s]^{-1}$. At very high frequencies, C_F shorts the gate and drain terminals of M_1, yielding an input *resistance* equal to $R_D||(1/g_m)$. More generally, the reader can prove that the real and imaginary parts of the input admittance are, respectively, equal to

$$Re\{Y_{in}\} = R_D C_F \omega^2 \frac{C_F + g_m R_D (C_L + C_F)}{R_D^2 (C_L + C_F)^2 \omega^2 + 1} \tag{5.15}$$

$$Im\{Y_{in}\} = C_F \omega \frac{R_D^2 C_L (C_L + C_F)\omega^2 + 1 + g_m R_D}{R_D^2 (C_L + C_F)^2 \omega^2 + 1}. \tag{5.16}$$

Is it possible to select the circuit parameters so as to obtain $Re\{Y_{in}\} = 1/(50\,\Omega)$? For example, if $C_F = 10$ fF, $C_L = 30$ fF, $g_m R_D = 4$, and $R_D = 100\,\Omega$, then $Re\{Y_{in}\} = (7.8\ \text{k}\Omega)^{-1}$ at 5 GHz, far from $(50\,\Omega)^{-1}$. This is because C_F introduces little feedback at this frequency.

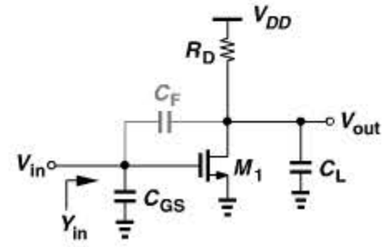


Figure 5.8 *Input admittance of a CS stage.*

Example 5.4

Why did we compute the input admittance rather than the input impedance for the circuit of Fig. 5.8.

Solution:

The choice of one over the other is somewhat arbitrary. In some circuits, it is simpler to compute Y_{in}. Also, if the input capacitance is cancelled by a *parallel* inductor, then $Im\{Y_{in}\}$ is more relevant. Similarly, a series inductor would cancel $Im\{Z_{in}\}$. We return to these concepts later in this chapter.

Can we employ simple resistive termination at the input? Illustrated in Fig. 5.9(a), such a topology is designed in three steps: (1) M_1 and R_D provide the required noise figure and gain, (2) R_P is placed in parallel with the input to provide $Re\{Z_{in}\} = 50\,\Omega$, and (3) an inductor is interposed between R_S and the input to cancel $Im\{Z_{in}\}$. Unfortunately, as explained in Chapter 2, the termination resistor itself yields a noise figure of $1 + R_S/R_P$. To calculate the noise figure at low frequencies, we can utilize Friis' equation[4] or simply treat the entire LNA as one circuit and, from Fig. 5.9(b), express the total output noise as

$$\overline{V_{n,out}^2} = 4kT(R_S||R_P)(g_m R_D)^2 + 4kT\gamma g_m R_D^2 + 4kTR_D, \tag{5.17}$$

where channel-length modulation is neglected. Since the voltage gain from V_{in} to V_{out} in Fig. 5.9(a) is equal to $-[R_P/(R_P + R_S)]g_m R_D$, the noise figure is given by

$$\text{NF} = 1 + \frac{R_S}{R_P} + \frac{\gamma R_S}{g_m(R_S||R_P)^2} + \frac{R_S}{g_m^2(R_S||R_P)^2 R_D}. \tag{5.18}$$

For $R_P \approx R_S$, the NF exceeds 3 dB—perhaps substantially.

The key point in the foregoing study is that the LNA must provide a 50-Ω input resistance *without* the thermal noise of a physical 50-Ω resistor. This becomes possible with the aid of active devices.

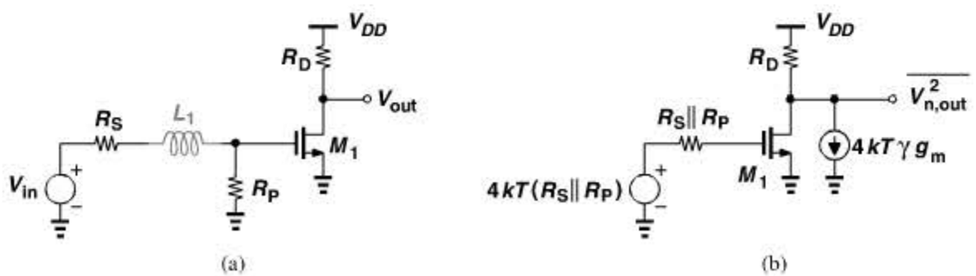


Figure 5.9 *(a) Use of resistive termination for matching, (b) simplified circuit.*

4. That is, consider R_P as one stage and the CS amplifier as another.

Example 5.5

A student decides to defy the above observation by choosing a *large* R_P and *transforming* its value down to R_S. The resulting circuit is shown in Fig. 5.10(a), where C_1 represents the input capacitance of M_1. (The input resistance of M_1 is neglected.) Can this topology achieve a noise figure less than 3 dB?

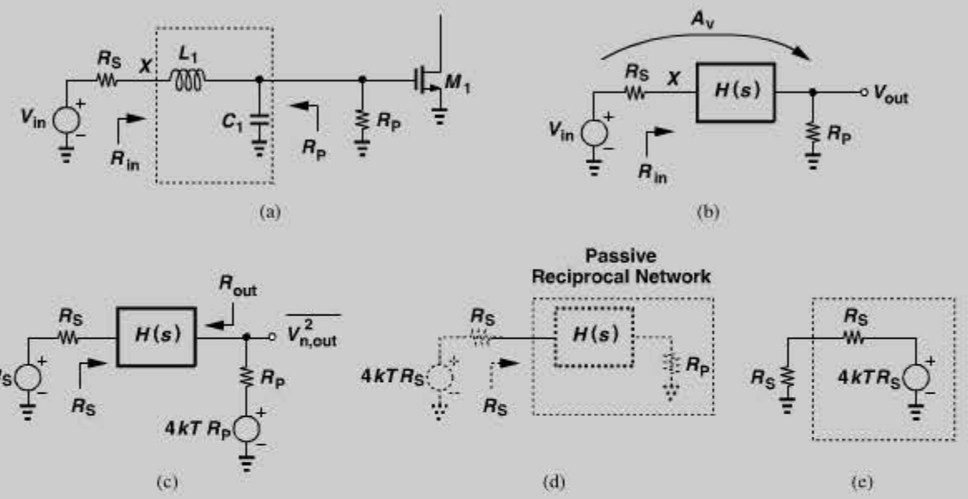


Figure 5.10 *(a) Use of matching circuit to transform the value of R_P, (b) general representation of (a), (c) inclusion of noise of R_P, (d) simplified circuit of (c), (e) simplified circuit of (d).*

Solution:

Consider the more general circuit in Fig. 5.10(b), where $H(s)$ represents a lossless network similar to L_1 and C_1 in Fig. 5.10(a). Since it is desired that $Z_{in} = R_S$, the power delivered by V_{in} to the input port of $H(s)$ is equal to $(V_{in,rms}/2)^2/R_S$. This power must also be delivered to R_P:

$$\frac{V_{in,rms}^2}{4R_S} = \frac{V_{out,rms}^2}{R_P}. \tag{5.19}$$

It follows that

$$|A_v|^2 = \frac{R_P}{4R_S}. \tag{5.20}$$

Let us now compute the output noise with the aid of Fig. 5.10(c). The output noise due to the noise of R_S is readily obtained from Eq. (5.19) by the substitution $\overline{V_{in,rms}^2} = 4kTR_S$:

$$\overline{V_{n,out}^2}|_{RS} = 4kTR_S \cdot \frac{R_P}{4R_S} \tag{5.21}$$

$$= kTR_P. \tag{5.22}$$

(*Continues*)

Example 5.5 *(Continued)*

But, how about the noise of R_P? We must first determine the value of R_{out}. To this end, we invoke the following thermodynamics principle: if R_S and R_P are in thermal equilibrium, then the noise power delivered by R_S to R_P must remain equal to the noise power delivered by R_P to R_S; otherwise, one heats up and the other cools down. How much is the noise delivered to R_S by R_P? We draw the circuit as depicted in Fig. 5.10(d) and recall from Chapter 2 that a passive reciprocal network exhibiting a real port impedance of R_S also produces a thermal noise of $4kTR_S$. From the equivalent circuit shown in Fig. 5.10(e), we note that the noise power delivered to the R_S on the left is equal to kT. Equating this value to the noise delivered by R_P to R_{out} in Fig. 5.10(c), we write

$$4kTR_P\left(\frac{R_{out}}{R_{out}+R_P}\right)^2\cdot\frac{1}{R_{out}}=kT \tag{5.23}$$

and hence

$$R_{out}=R_P. \tag{5.24}$$

That is, if $R_{in}=R_S$, then $R_{out}=R_P$. The output noise due to R_P is therefore given by

$$\overline{V_{n,out}^2}|_{RP}=kTR_P. \tag{5.25}$$

Summing (5.22) and (5.25) and dividing the result by (5.20) and $4kTR_S$, we arrive at the noise figure of the circuit (excluding M_1):

$$\text{NF}=2. \tag{5.26}$$

Unfortunately, the student has attempted to defy the laws of physics.

In summary, proper input (conjugate) matching of LNAs requires certain circuit techniques that yield a real part of 50 Ω in the input impedance without the noise of a 50-Ω resistor. We study such techniques in the next section.

5.3 LNA TOPOLOGIES

Our preliminary studies thus far suggest that the noise figure, input matching, and gain constitute the principal targets in LNA design. In this section, we present a number of LNA topologies and analyze their behavior with respect to these targets. Table 5.1 provides an overview of these topologies.

5.3.1 Common-Source Stage with Inductive Load

As noted in Section 5.1, a CS stage with resistive load (Fig. 5.8) proves inadequate because it does not provide proper matching. Furthermore, the output node time constant may prohibit operation at high frequencies. In general, the trade-off between the voltage gain and

Table 5.1 *Overview of LNA topologies.*

Common–Source Stage with	Common–Gate Stage with	Broadband Topologies
▪ Inductive Load	▪ Inductive Load	▪ Noise–Cancelling LNAs
▪ Resistive Feedback	▪ Feedback	▪ Reactance–Cancelling LNAs
▪ Cascode, Inductive Load, Inductive Degeneration	▪ Feedforward ▪ Cascode and Inductive Load	

the supply voltage in this circuit makes it less attractive as the latter scales down with technology. For example, at low frequencies,

$$|A_v|=g_mR_D \tag{5.27}$$

$$=\frac{2I_D}{V_{GS}-V_{TH}}\cdot\frac{V_{RD}}{I_D} \tag{5.28}$$

$$=\frac{2V_{RD}}{V_{GS}-V_{TH}}, \tag{5.29}$$

where V_{RD} denotes the dc voltage drop across R_D and is limited by V_{DD}. With channel-length modulation, the gain is even lower.

In order to circumvent the trade-off expressed by Eq. (5.29) and also operate at higher frequencies, the CS stage can incorporate an inductive load. Illustrated in Fig. 5.11(a), such a topology operates with very low supply voltages because the inductor sustains a smaller dc voltage drop than a resistor does. (For an ideal inductor, the dc drop is zero.) Moreover, L_1 resonates with the total capacitance at the output node, affording a much higher operation frequency than does the resistively-loaded counterpart of Fig. 5.8.

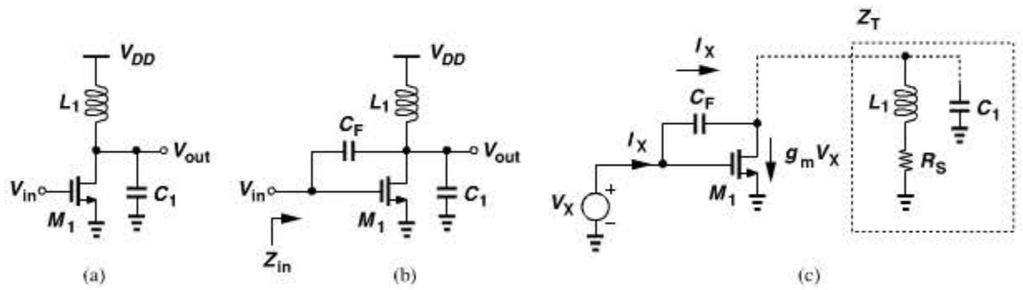


Figure 5.11 *(a) Inductively-loaded CS stage, (b) input impedance in the presence of C_F, (c) equivalent circuit.*

How about the input matching? We consider the more complete circuit shown in Fig. 5.11(b), where C_F denotes the gate-drain overlap capacitance. Ignoring the gate-source capacitance of M_1 for now, we wish to compute Z_{in}. We redraw the circuit as depicted in Fig. 5.11(c) and note that the current flowing through the output parallel tank is equal to

$I_X - g_m V_X$. In this case, the inductor loss is modeled by a *series* resistance, R_S, because this resistance varies much less with frequency than the equivalent parallel resistance does.[5] The tank impedance is given by

$$Z_T = \frac{L_1 s + R_S}{L_1 C_1 s^2 + R_S C_1 s + 1}, \tag{5.30}$$

and the tank voltage by $(I_X - g_m V_X) Z_T$. Adding the voltage drop across C_F to the tank voltage, we have

$$V_X = \frac{I_X}{C_F s} + (I_X - g_m V_X) Z_T. \tag{5.31}$$

Substitution of Z_T from (5.30) gives

$$Z_{in}(s) = \frac{V_X}{I_X} = \frac{L_1(C_1 + C_F)s^2 + R_S(C_1 + C_F)s + 1}{[L_1 C_1 s^2 + (R_S C_1 + g_m L_1)s + 1 + g_m R_S]C_F s}. \tag{5.32}$$

For $s = j\omega$,

$$Z_{in}(j\omega) = \frac{1 - L_1(C_1 + C_F)\omega^2 + jR_S(C_1 + C_F)\omega}{[-(R_S C_1 + g_m L_1)\omega + j(g_m R_S - L_1 C_1 \omega^2 + 1)]C_F \omega}. \tag{5.33}$$

Since the real part of a complex fraction $(a + jb)/(c + jd)$ is equal to $(ac + bd)/(c^2 + d^2)$, we have

$$Re\{Z_{in}\} = \frac{[1 - L_1(C_1 + C_F)\omega^2][-(R_S C_1 + g_m L_1)\omega] + R_S(C_1 + C_F)(g_m R_S - L_1 C_1 \omega^2 + 1)\omega^2}{D}, \tag{5.34}$$

where D is a positive quantity. It is thus possible to select the values so as to obtain $Re\{Z_{in}\} = 50\,\Omega$.

While providing the possibility of $Re\{Z_{in}\} = 50\,\Omega$ at the frequency of interest, the feedback capacitance in Fig. 5.11(b) gives rise to a *negative* input resistance at other frequencies, potentially causing instability. To investigate this point, let us rewrite Eq. (5.34) as

$$Re\{Z_{in}\} = \frac{g_m L_1^2(C_1 + C_F)\omega^2 + R_S(1 + g_m R_S)(C_1 + C_F) - (R_S C_1 + g_m L_1)}{D}\omega. \tag{5.35}$$

We note that the numerator falls to zero at a frequency given by

$$\omega_1^2 = \frac{R_S C_1 + g_m L_1 - (1 + g_m R_S)R_S(C_1 + C_F)}{g_m L_1^2(C_1 + C_F)}. \tag{5.36}$$

5. For example, if R_S simply represents the low-frequency resistance of the wire, its value remains constant and $Q = L\omega/R_S$ rises linearly with frequency. For a parallel resistance, R_P, to allow such a behavior for $Q = R_P/(L\omega)$, the resistance must rise in proprotion to ω^2 rather than remain constant.

Thus, at this frequency (if it exists), $Re\{Z_{in}\}$ changes sign. For example, if $C_F = 10\,\text{fF}$, $C_1 = 30\,\text{fF}$, $g_m = (20\,\Omega)^{-1}$, $L_1 = 5$ nH, and $R_S = 20\,\Omega$, then $g_m L_1$ dominates in the numerator, yielding $\omega_1^2 \approx [L_1(C_1 + C_F)]^{-1}$ and hence $\omega_1 \approx 2\pi \times (11.3\,\text{GHz})$.

It is possible to "neutralize" the effect of C_F in some frequency range through the use of parallel resonance (Fig. 5.12), but, since C_F is relatively small, L_F must assume a large value, thereby introducing significant parasitic capacitances at the input and output (and even *between* the input and output) and degrading the performance. For these reasons, this topology is rarely used in modern RF design.

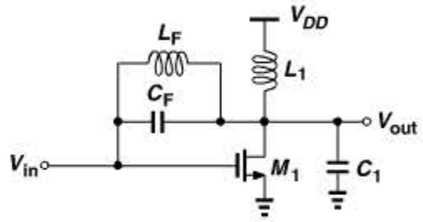


Figure 5.12 *Neutralization of C_F by L_F.*

5.3.2 Common-Source Stage with Resistive Feedback

If the frequency of operation remains an order of magnitude lower than the f_T of the transistor, the feedback CS stage depicted in Fig. 5.13(a) may be considered as a possible candidate. Here, M_2 operates as a current source and R_F senses the output voltage and returns a current to the input. We wish to design this stage for an input resistance equal to R_S and a relatively low noise figure.

If channel-length modulation is neglected, we have from Fig. 5.13(b),

$$R_{in} = \frac{1}{g_{m1}} \tag{5.37}$$

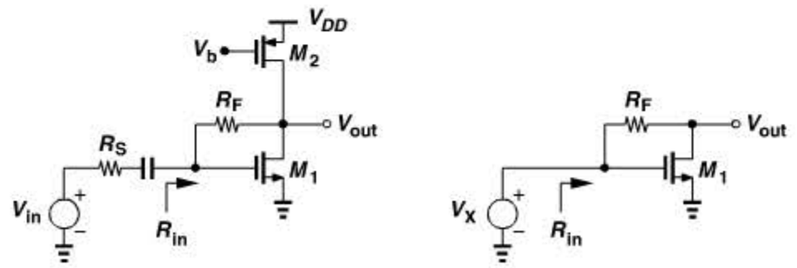


Figure 5.13 *(a) CS stage with resistive feedback, (b) simplified circuit.*

because R_F is simply in series with an ideal current source and M_1 appears as a diode-connected device. We must therefore choose

$$g_{m1} = \frac{1}{R_S}. \tag{5.38}$$

Figure 5.13(b) also implies that the small-signal drain current of M_1, $g_{m1}V_X$, entirely flows through R_F, generating a voltage drop of $g_{m1}V_X R_F$. It follows that

$$V_X - g_{m1}V_X R_F = V_{out} \tag{5.39}$$

and hence

$$\frac{V_{out}}{V_X} = 1 - g_{m1}R_F \tag{5.40}$$

$$= 1 - \frac{R_F}{R_S}. \tag{5.41}$$

In practice, $R_F \gg R_S$, and the voltage gain from V_{in} to V_{out} in Fig. 5.13(a) is equal to

$$A_v = \frac{1}{2}\left(1 - \frac{R_F}{R_S}\right) \tag{5.42}$$

$$\approx -\frac{R_F}{R_S}. \tag{5.43}$$

In contrast to the resistively-loaded CS stage of Fig. 5.8, this circuit does not suffer from a direct trade-off between gain and supply voltage because R_F carries no bias current.

Let us determine the noise figure of the circuit, assuming that $g_{m1} = 1/R_S$. We first compute the noise contributions of R_F, M_1, and M_2 at the output. From Fig. 5.14(a), the reader can show the noise of R_F appears at the output in its entirety:

$$\overline{V_{n,out}^2}|_{RF} = 4kTR_F. \tag{5.44}$$

The noise currents of M_1 and M_2 flow through the output impedance of the circuit, R_{out}, as shown in Fig. 5.14(b). The reader can prove that

$$R_{out} = \left[\frac{1}{g_{m1}}\left(1 + \frac{R_F}{R_S}\right)\right] || (R_F + R_S) \tag{5.45}$$

$$= \frac{1}{2}(R_F + R_S). \tag{5.46}$$

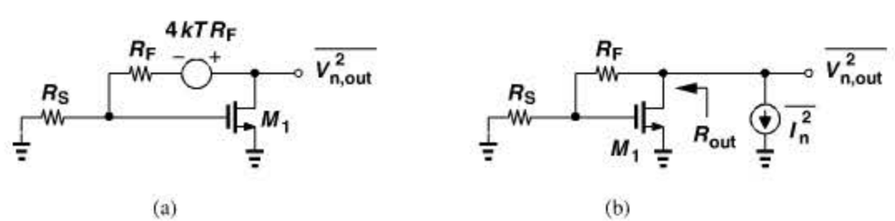


Figure 5.14 *Effect of noise of (a) R_F and (b) M_1 in CS stage.*

It follows that

$$\overline{V_{n,out}^2}|_{M1,M2} = 4kT\gamma(g_{m1} + g_{m2})\frac{(R_F + R_S)^2}{4}. \tag{5.47}$$

The noise of R_S is multiplied by the gain when referred to the output, and the result is divided by the gain when referred to the input. We thus have

$$\text{NF} = 1 + \frac{4R_F}{R_S\left(1 - \frac{R_F}{R_S}\right)^2} + \frac{\gamma(g_{m1} + g_{m2})(R_F + R_S)^2}{\left(1 - \frac{R_F}{R_S}\right)^2 R_S} \tag{5.48}$$

$$\approx 1 + \frac{4R_S}{R_F} + \gamma(g_{m1} + g_{m2})R_S \tag{5.49}$$

$$\approx 1 + \frac{4R_S}{R_F} + \gamma + \gamma g_{m2}R_S. \tag{5.50}$$

For $\gamma \approx 1$, the NF exceeds 3 dB even if $4R_S/R_F + \gamma g_{m2}R_S \ll 1$.

Example 5.6

Express the fourth term on the right-hand side of Eq. (5.50) in terms of transistor overdrive voltages.

Solution:

Since $g_m = 2I_D/(V_{GS} - V_{TH})$, we write $g_{m2}R_S = g_{m2}/g_{m1}$ and

$$\frac{g_{m2}}{g_{m1}} = \frac{(V_{GS} - V_{TH})_1}{|V_{GS} - V_{TH}|_2}. \tag{5.51}$$

That is, the fourth term becomes negligible only if the overdrive of the current source remains much higher than that of M_1—a difficult condition to meet at low supply voltages because $|V_{DS2}| = V_{DD} - V_{GS1}$. We should also remark that heavily velocity-saturated MOSFETs have a transconductance given by $g_m = I_D/(V_{GS} - V_{TH})$ and still satisfy (5.51).

Example 5.7

In the circuit of Fig. 5.15, the PMOS current source is converted to an "active load," amplifying the input signal. The idea is that, if M_2 amplifies the input in addition to injecting noise to the output, then the noise figure may be lower. Neglecting channel-length modulation, calculate the noise figure. (Current source I_1 defines the bias current, and C_1 establishes an ac ground at the source of M_2.)

(*Continues*)

Example 5.7 *(Continued)*

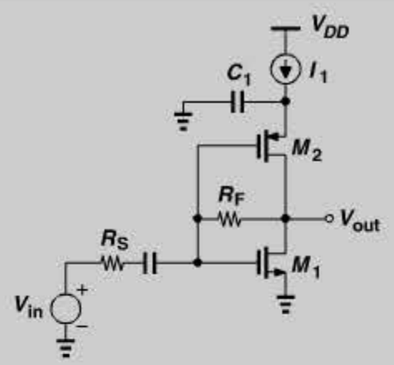


Figure 5.15 *CS stage with active load.*

Solution:

For small-signal operation, M_1 and M_2 appear *in parallel*, behaving as a single transistor with a transconductance of $g_{m1} + g_{m2}$. Thus, for input matching, $g_{m1} + g_{m2} = 1/R_S$. The noise figure is still given by Eq. (5.49), except that $\gamma(g_{m1} + g_{m2})R_S = \gamma$. That is,

$$\text{NF} \approx 1 + \frac{4R_S}{R_F} + \gamma. \tag{5.52}$$

This circuit is therefore superior, but it requires a supply voltage equal to $V_{GS1} + |V_{GS2}| + V_{I1}$, where V_{I1} denotes the voltage headroom necessary for I_1.

5.3.3 Common-Gate Stage

The low input impedance of the common-gate (CG) stage makes it attractive for LNA design. Since a resistively-loaded stage suffers from the same gain-headroom trade-off as its CS counterpart, we consider only a CG circuit with inductive loading [Fig. 5.16(a)]. Here, L_1 resonates with the total capacitance at the output node (including the input capacitance of the following stage), and R_1 represents the loss of L_1. If channel-length modulation and body effect are neglected, $R_{in} = 1/g_m$. Thus, the dimensions and bias current of M_1 are chosen so as to yield $g_m = 1/R_S = (50\,\Omega)^{-1}$. The voltage gain from X to the output node at the output resonance frequency is then equal to

$$\frac{V_{out}}{V_X} = g_m R_1 \tag{5.53}$$

$$= \frac{R_1}{R_S} \tag{5.54}$$

and hence $V_{out}/V_{in} = R_1/(2R_S)$.

Let us now determine the noise figure of the circuit under the condition $g_m = 1/R_S$ and at the resonance frequency. Modeling the thermal noise of M_1 as a voltage source in series

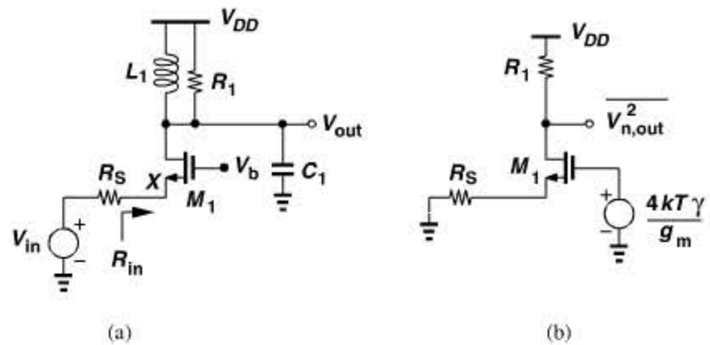


Figure 5.16 *(a) CG stage, (b) effect of noise of M_1.*

with its gate, $\overline{V_{n1}^2} = 4kT\gamma/g_m$ [Fig. 5.16(b)], and multiplying it by the gain from the *gate* of M_1 to the output, we have

$$\overline{V_{n,out}^2}|_{M1} = \frac{4kT\gamma}{g_m}\left(\frac{R_1}{R_S + \frac{1}{g_m}}\right)^2 \tag{5.55}$$

$$= kT\gamma\frac{R_1^2}{R_S}. \tag{5.56}$$

The output noise due to R_1 is simply equal to $4kTR_1$. To obtain the noise figure, we divide the output noise due to M_1 and R_1 by the gain and $4kTR_S$ and add unity to the result:

$$\text{NF} = 1 + \frac{\gamma}{g_m R_S} + \frac{R_S}{R_1}\left(1 + \frac{1}{g_m R_S}\right)^2 \tag{5.57}$$

$$= 1 + \gamma + 4\frac{R_S}{R_1}. \tag{5.58}$$

Even if $4R_S/R_1 \ll 1 + \gamma$, the NF still reaches 3 dB (with $\gamma \approx 1$), a price paid for the condition $g_m = 1/R_S$. In other words, a higher g_m yields a lower NF but also a lower input resistance. In Problem 5.8, we show that the NF can be lower if some impedance mismatch is permitted at the input.

Example 5.8

We wish to provide the bias current of the CG stage by a current source or a resistor (Fig. 5.17). Compare the additional noise in these two cases.

(Continues)

Example 5.8 *(Continued)*

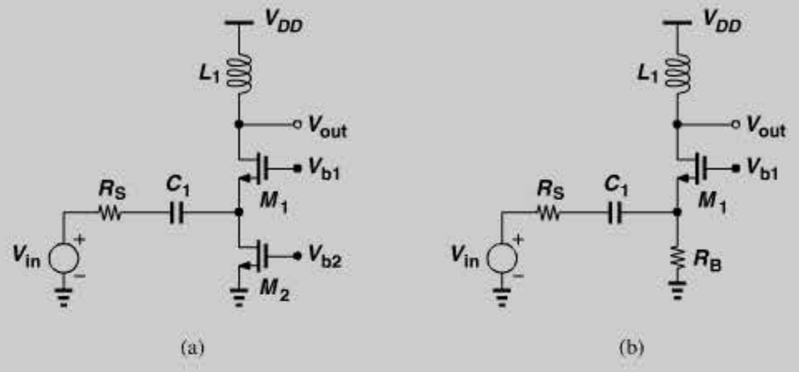


Figure 5.17 *CG stage biasing with (a) current source and (b) resistor.*

Solution:

For a given V_{b1} and V_{GS1}, the source voltages of M_1 in the two cases are equal and hence V_{DS2} is equal to the voltage drop across R_B ($=V_{RB}$). Operating in saturation, M_2 requires that $V_{DS2} \geq V_{GS2} - V_{TH2}$. We express the noise current of M_2 as

$$\overline{I_{n,M2}^2} = 4kT\gamma g_{m2} \tag{5.59}$$

$$= 4kT\gamma \frac{2I_D}{V_{GS2} - V_{TH2}}, \tag{5.60}$$

and that of R_B as

$$\overline{I_{n,RB}^2} = \frac{4kT}{R_B} \tag{5.61}$$

$$= 4kT\frac{I_D}{V_{RB}}. \tag{5.62}$$

Since $V_{GS2} - V_{TH2} \leq V_{RB}$, the noise contribution of M_2 is about twice that of R_B (for $\gamma \approx 1$). Additionally, M_2 may introduce significant capacitance at the input node.

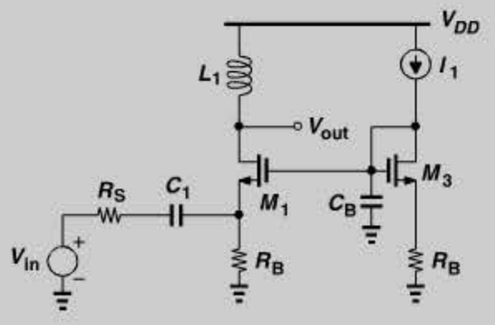


Figure 5.18 *Proper biasing of CG stage.*

Example 5.8 *(Continued)*

The use of a resistor is therefore preferable, so long as R_B is much greater than R_S so that it does not attenuate the input signal. Note that the input capacitance due to M_1 may still be significant. We will return to this issue later. Figure 5.18 shows an example of proper biasing in this case.

In deep-submicron CMOS technologies, channel-length modulation significantly impacts the behavior of the CG stage. As shown in Fig. 5.19, the positive feedback through r_O *raises* the input impedance. Since the drain-source current of M_1 (without r_O) is equal to $-g_m V_X$ (if body effect is neglected), the current flowing through r_O is given by $I_X - g_m V_X$, yielding a voltage drop of $r_O(I_X - g_m V_X)$ across it. Also, I_X flows through the output tank, producing a voltage of $I_X R_1$ at the resonance frequency. Adding this voltage to the drop across r_O and equating the result to V_X, we obtain

$$V_X = r_O(I_X - g_m V_X) + I_X R_1. \tag{5.63}$$

That is,

$$\frac{V_X}{I_X} = \frac{R_1 + r_O}{1 + g_m r_O}. \tag{5.64}$$

If the intrinsic gain, $g_m r_O$, is much greater than unity, then $V_X/I_X \approx 1/g_m + R_1/(g_m r_O)$. However, in today's technology, $g_m r_O$ hardly exceeds 10. Thus, the term $R_1/(g_m r_O)$ may become comparable with or even *exceed* the term $1/g_m$, yielding an input resistance substantially higher than 50 Ω.

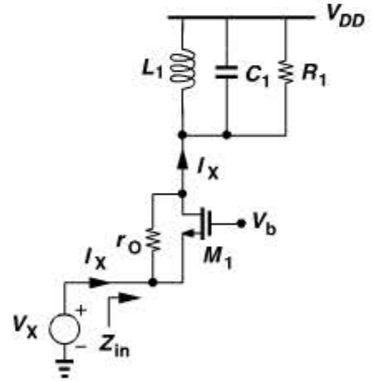


Figure 5.19 *Input impedance of CG stage in the presence of r_O.*

Example 5.9

Neglecting the capacitances of M_1 in Fig. 5.19, plot the input impedance as a function of frequency.

(Continues)

Example 5.9 *(Continued)*

Solution:

At very low or very high frequencies, the tank assumes a low impedance, yielding $R_{in} = 1/g_m$ [or $1/(g_m + g_{mb})$ if body effect is considered]. Figure 5.20 depicts the behavior.

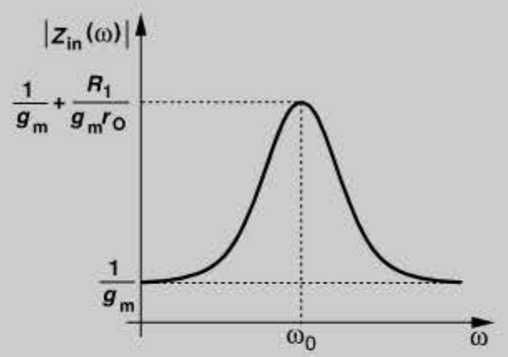


Figure 5.20 *Input impedance of CG stage with a resonant load.*

With the strong effect of R_1 on R_{in}, we must equate the actual input resistance to R_S to guarantee input matching:

$$R_S = \frac{R_1 + r_O}{1 + g_m r_O}. \tag{5.65}$$

The reader can prove that the voltage gain of the CG stage shown in Fig. 5.16(a) with a finite r_O is expressed as

$$\frac{V_{out}}{V_{in}} = \frac{g_m r_O + 1}{r_O + g_m r_O R_S + R_S + R_1} R_1, \tag{5.66}$$

which, from Eq. (5.65), reduces to

$$\frac{V_{out}}{V_{in}} = \frac{g_m r_O + 1}{2\left(1 + \dfrac{r_O}{R_1}\right)}. \tag{5.67}$$

This is a disturbing result! If r_O and R_1 are comparable, then the voltage gain is on the order of $g_m r_O/4$, a very low value.

In summary, the input impedance of the CG stage is too low if channel-length modulation is neglected and too high if it is not! A number of circuit techniques have been introduced to deal with the former case (Section 5.3.5), but in today's technology, we face the latter case.

In order to alleviate the above issue, the channel length of the transistor can be increased, thus reducing channel-length modulation and raising the achievable $g_m r_O$. Since the device width must also increase proportionally so as to retain the transconductance value, the gate-source capacitance of the transistor rises considerably, degrading the input return loss.

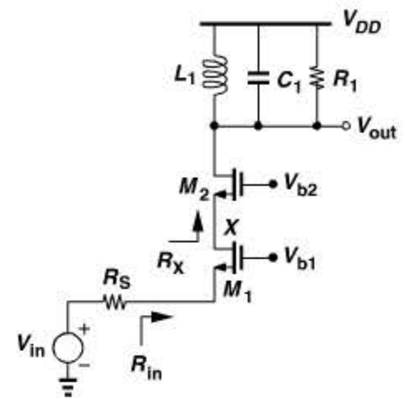


Figure 5.21 *Cascode CG stage.*

Cascode CG Stage An alternative approach to lowering the input impedance is to incorporate a cascode device as shown in Fig. 5.21. Here, the resistance seen looking into the source of M_2 is given by Eq. (5.64):

$$R_X = \frac{R_1 + r_{O2}}{1 + g_{m2} r_{O2}}. \tag{5.68}$$

This load resistance is now transformed to a lower value by M_1, again according to (5.64):

$$R_{in} = \left(\frac{R_1 + r_{O1}}{1 + g_{m2} r_{O2}} + r_{O1}\right) \div (1 + g_{m1} r_{O1}). \tag{5.69}$$

If $g_m r_O \gg 1$, then

$$R_{in} \approx \frac{1}{g_{m1}} + \frac{R_1}{g_{m1} r_{O1} g_{m2} r_{O2}} + \frac{1}{g_{m1} r_{O1} g_{m2}}. \tag{5.70}$$

Since R_1 is divided by the product of two intrinsic gains, its effect remains negligible. Similarly, the third term is much less than the first if g_{m1} and g_{m2} are roughly equal. Thus, $R_{in} \approx 1/g_{m1}$.

The addition of the cascode device entails two issues: the noise contribution of M_2 and the voltage headroom limitation due to stacking two transistors. To quantify the former, we consider the equivalent circuit shown in Fig. 5.22(a), where R_S ($= 1/g_{m1}$) and M_1 are replaced with an output resistance equal to $2r_{O1}$ (why?), and $C_X = C_{DB1} + C_{GD1} + C_{SB2}$. For simplicity, we have also replaced the tank with a resistor R_1, i.e., the output node has a broad bandwidth. Neglecting the gate-source capacitance, channel-length modulation, and body effect of M_2, we express the transfer function from V_{n2} to the output at the resonance frequency as

$$\frac{V_{n,out}}{V_{n2}}(s) = \frac{R_1}{\dfrac{1}{g_{m2}} + (2r_{O1}) \| \dfrac{1}{C_X s}} \tag{5.71}$$

$$= \frac{2r_{O1} C_X s + 1}{2r_{O1} C_X s + 2g_{m2} r_{O1} + 1} g_{m2} R_1. \tag{5.72}$$

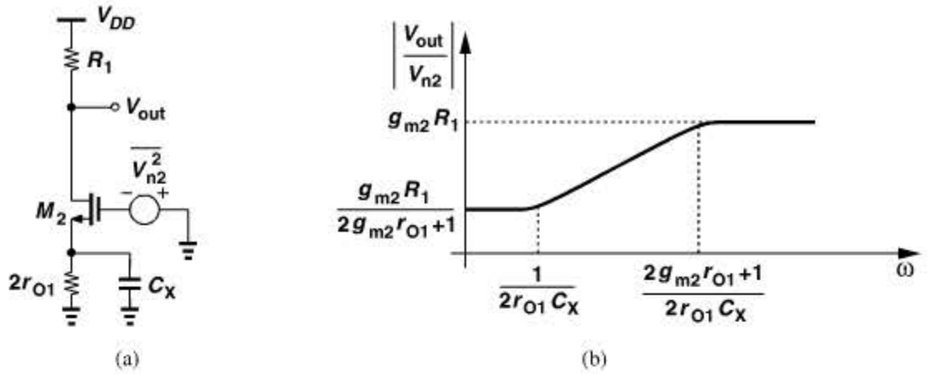


Figure 5.22 *(a) Cascode transistor noise, (b) output contribution as a function of frequency.*

Figure 5.22(b) plots the frequency response, implying that the noise contribution of M_2 is negligible for frequencies up to the zero frequency, $(2r_{O1}C_X)^{-1}$, but begins to manifest itself thereafter. Since C_X is comparable with C_{GS} and $2r_{O1} \gg 1/g_m$, we note that $(2r_{O1}C_X)^{-1} \ll g_m/C_{GS}$ ($\approx \omega_T$). That is, the zero frequency is much lower than the f_T of the transistors, making this effect potentially significant.

Example 5.10

Assuming $2r_{O1} \gg |C_X s|^{-1}$ at frequencies of interest so that the degeneration impedance in the source of M_2 reduces to C_X, recompute the above transfer function while taking C_{GS2} into account. Neglect the effect of r_{O2}.

Solution:

From the equivalent circuit shown in Fig. 5.23, we have $g_{m2}V_1 = -V_{out}/R_1$ and hence $V_1 = -V_{out}/(g_{m2}R_1)$. The current flowing through C_{GS2} is therefore equal to $-V_{out}C_{GS2}s/(g_{m2}R_1)$. The sum of this current and $-V_{out}/R_1$ flows through C_X, producing a voltage of $[-V_{out}C_{GS2}s/(g_{m2}R_1) - V_{out}/R_1]/(C_X s)$. Writing a KVL in the input loop gives

$$\left(-\frac{V_{out}C_{GS2}s}{g_{m2}R_1} - \frac{V_{out}}{R_1}\right)\frac{1}{C_X s} - \frac{V_{out}}{g_{m2}R_1} = V_{in} \tag{5.73}$$

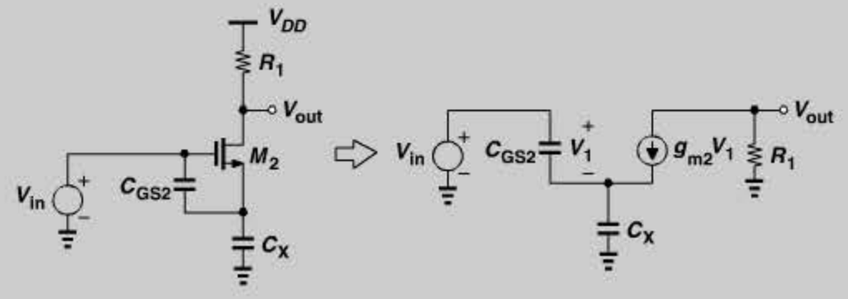


Figure 5.23 *Computation of gain from the gate of cascode device to output.*

Example 5.10 *(Continued)*

and hence

$$\frac{V_{out}}{V_{in}} = \frac{-g_{m2}R_1C_X s}{(C_{GS2}+C_X)s + g_{m2}}. \tag{5.74}$$

At frequencies well below the f_T of the transistor, $|(C_{GS2}+C_X)s| \ll g_{m2}$ and

$$\frac{V_{out}}{V_{in}} \approx -R_1C_X s. \tag{5.75}$$

That is, the noise of M_2 reaches the output unattenuated if ω is much greater than $(2r_{O1}C_X)^{-1}$ [but much less than $g_{m2}/(C_{GS2}+C_X)$].

The second issue stemming from the cascode device relates to the limited voltage headroom. To quantify this limitation, let us determine the required or allowable values of V_{b1} and V_{b2} in Fig. 5.21. Since the drain voltage of M_2 begins at V_{DD} and can swing below its gate voltage by as much as V_{TH2} while keeping M_2 in saturation, we can simply choose $V_{b2} = V_{DD}$. Now, $V_X = V_{DD} - V_{GS2}$, allowing a maximum value of $V_{DD} - V_{GS2} + V_{TH1}$ for V_{b1} if M_1 must remain saturated. Consequently, the source voltage of M_1 cannot exceed $V_{DD} - V_{GS2} - (V_{GS1} - V_{TH1})$. We say the two transistors consume a voltage headroom of one V_{GS} plus one overdrive $(V_{GS1} - V_{TH1})$.

It may appear that, so long as $V_{DD} > V_{GS2} + (V_{GS1} - V_{TH1})$, the circuit can be properly biased, but how about the path from the source of M_1 to ground? In comparison with the CG stages in Fig. 5.17, the cascode topology consumes an additional voltage headroom of $V_{GS1} - V_{TH1}$, leaving less for the biasing transistor or resistor and hence raising their noise contribution. For example, suppose $I_{D1} = I_{D2} = 2\,\text{mA}$. Since $g_{m1} = (50\,\Omega)^{-1} = 2I_D/(V_{GS1} - V_{TH1})$, we have $V_{GS1} - V_{TH1} = 200\,\text{mV}$. Also assume $V_{GS2} \approx 500\,\text{mV}$. Thus, with $V_{DD} = 1$ V, the voltage available for a bias resistor, R_B, tied between the source of M_1 and ground cannot exceed $300\,\text{mV}/2\,\text{mA} = 150\,\Omega$. This value is comparable with $R_S = 50\,\Omega$ and degrades the gain and noise behavior of the circuit considerably.

In order to avoid the noise-headroom trade-off imposed by R_B, and also cancel the input capacitance of the circuit, CG stages often employ an inductor for the bias path. Illustrated in Fig. 5.24 with proper biasing for the input transistor, this technique minimizes the additional noise due to the biasing element (L_B) and significantly improves the input matching. In modern RF design, both L_B and L_1 are integrated on the chip.

Design Procedure With so many devices present in the circuit of Fig. 5.24, how do we begin the design? We describe a systematic procedure that provides a "first-order" design, which can then be refined and optimized.

The design procedure begins with two knowns: the frequency of operation and the supply voltage. In the first step, the dimensions and bias current of M_1 must be chosen such that a transconductance of $(50\,\Omega)^{-1}$ is obtained. The length of the transistor is set to the minimum allowable by the technology, but how should the width and the drain current be determined?

Using circuit simulations, we plot the transconductance and f_T of an NMOS transistor with a given width, W_0, as a function of the drain current. For long-channel devices,

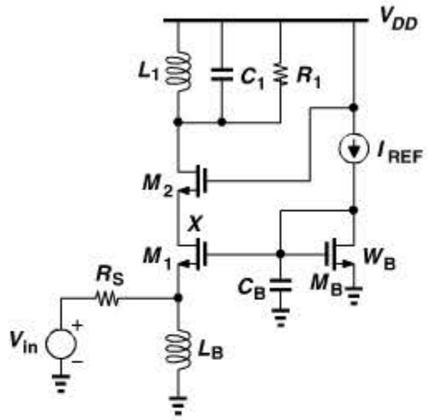


Figure 5.24 *Biasing of cascode CG stage.*

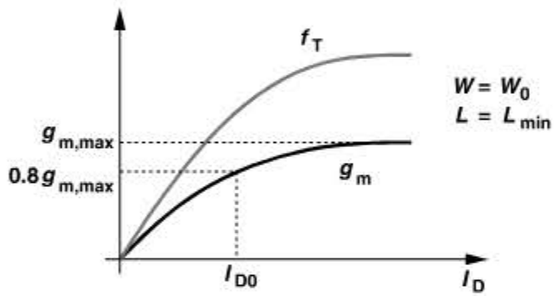


Figure 5.25 *Behavior of g_m and f_T as a function of drain current.*

$g_m \propto \sqrt{I_D}$, but submicron transistors suffer from degradation of the mobility with the vertical field in the channel, exhibiting the saturation behavior shown in Fig. 5.25. To avoid excessive power consumption, we select a bias current, I_{D0}, that provides 80 to 90% of the saturated g_m. That is, the combination of W_0 and I_{D0} (the "current density") is nearly optimum in terms of speed (transistor capacitances) and power consumption.

With W_0 and I_{D0} known, any other value of transconductance can be obtained by simply *scaling* the two proportionally. The reader can prove that if W_0 and I_{D0} scale by a factor of α, then so does g_m, regardless of the type and behavior of the transistor. We thus arrive at the required dimensions and bias current of M_1 (for $1/g_{m1} = 50\,\Omega$), which in turn yield its overdrive voltage.

In the second step, we compute the necessary value of L_B in Fig. 5.24. As shown in Fig. 5.26, the input of the circuit sees a pad capacitance to the substrate.[6] Thus, L_B must resonate with $C_{pad} + C_{SB1} + C_{GS1}$ and *its own capacitance* at the frequency of interest. (Here, R_p models the loss of L_B.) Since the parasitic capacitance of L_B is not known a priori, some iteration is required. (The design and modeling of spiral inductors are described in Chapter 7.)

6. The input may also see additional capacitance due to electrostatic discharge (ESD) protection devices that are tied to V_{DD} and ground.

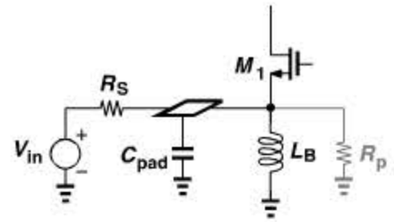


Figure 5.26 *Effect of pad capacitance on CG stage.*

Does L_B affect the performance of the circuit at resonance? Accompanying L_B is the parallel equivalent resistance $R_p = QL_B\omega$, which contributes noise and possibly attenuates the input signal. Thus, R_p must be at least ten times higher than $R_S = 50\,\Omega$. In other words, if the total capacitance at the input is so large as to dictate an excessively small inductor and R_p, then the noise figure is quite high. This situation may arise only at frequencies approaching the f_T of the technology.

In the third step, the bias of M_1 is defined by means of M_B and I_{REF} in Fig. 5.24. For example, $W_B = 0.2W_1$ and $I_{REF} = 0.2I_{D1}$ so that the bias branch draws only one-fifth of the current of the main branch.[7] Capacitor C_B provides a sufficiently low impedance (much less than $50\,\Omega$) from the gate of M_1 to ground and also bypasses the noise of M_B and I_B to ground. The choice of a solid, low-inductance ground is critical here because the high-frequency performance of the CG stage degrades drastically if the impedance seen in series with the gate becomes comparable with R_S.

Next, the width of M_2 in Fig. 5.24 must be chosen (the length is the minimum allowable value). With the bias current known ($I_{D2} = I_{D1}$), if the width is excessively small, then V_{GS2} may be so large as to drive M_1 into the triode region. On the other hand, as W_2 increases, M_2 contributes an increasingly larger capacitance to node X while its g_m reaches a nearly constant value (why?). Thus, the optimum width of M_2 is likely to be near that of M_1, and that is the initial choice. Simulations can be used to refine this choice, but in practice, even a twofold change from this value negligibly affects the performance.

In order to minimize the capacitance at node X in Fig. 5.24, transistors M_1 and M_2 can be laid out such that the drain area of the former is shared with the source area of the latter. Furthermore, since no other connection is made to this node, the shared area need not accommodate contacts and can therefore be minimized. Depicted in Fig. 5.27 and feasible only if $W_1 = W_2$, such a structure can be expanded to one with multiple gate fingers.

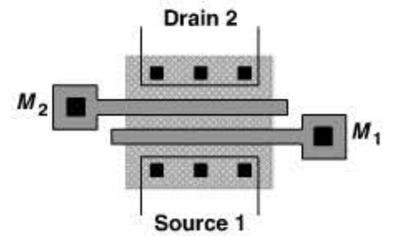


Figure 5.27 *Layout of cascode devices.*

7. For proper matching between the two transistors, M_1 incorporates five unit transistors (e.g., gate fingers) and M_B one unit transistor.

In the last step, the value of the load inductor, L_1, must be determined (Fig. 5.24). In a manner similar to the choice of L_B, we compute L_1 such that it resonates with $C_{GD2} + C_{DB2}$, the input capacitance of the next stage, and its own capacitance. Since the voltage gain of the LNA is proportional to $R_1 = QL_1\omega$, R_1 must be sufficiently large, e.g., 500 to 1000 Ω. This condition is met in most designs without much difficulty.

The design procedure outlined above leads to a noise figure around 3 dB [Eq. (5.58)] and a voltage gain, $V_{out}/V_{in} = R_1/(2R_S)$, of typically 15 to 20 dB. If the gain is *too high*, i.e., if it dictates an unreasonably high mixer IP_3, then an explicit resistor can be placed in parallel with R_1 to obtain the required gain. As studied in Section 5.7, this LNA topology displays a high IIP_3, e.g., +5 to +10 dBm.

Example 5.11

Design the LNA of Fig. 5.24 for a center frequency of 5.5 GHz in 65-nm CMOS technology. Assume the circuit is designed for an 11a receiver.

Solution:

Figure 5.28 plots the transconductance of an NMOS transistor with $W = 10\ \mu$m and $L = 60$ nm as a function of the drain current. We select a bias current of 2 mA to achieve a g_m of about 10 mS $= 1/(100\ \Omega)$. Thus, to obtain an input resistance of 50 Ω, we must double the width and drain current.[8] The capacitance introduced by a 20-μm transistor at the input is about 30 fF. To this we add a pad capacitance of 50 fF and choose $L_B = 10$ nH for resonance at 5.5 GHz. Such an inductor exhibits a parasitic capacitance of roughly 30 fF, requiring that a smaller inductance be chosen, but we proceed without this refinement.

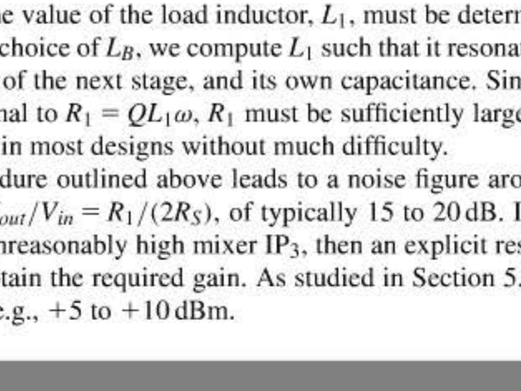


Figure 5.28 *Transcoductance of a 10 μm/60 nm NMOS device as a function of drain current.*

Next, we choose the width of the cascode device equal to 20 μm and assume a load capacitance of 30 fF (e.g., the input capacitance of subsequent mixers). This allows the use of a 10-nH inductor for the load, too, because the total capacitance at the output node amounts to about 75 fF. However, with a Q of about 10 for such an inductor, the LNA

8. The body effect lowers the input resistance, but the feedback from the drain to the gate raises it. We therefore neglect both.

Example 5.11 (*Continued*)

gain is exessively high and its bandwidth excessively low (failing to cover the 11a band). For this reason, we place a resistor of 1 kΩ in parallel with the tank. Figure 5.29 shows the design details and Fig. 5.30 the simulated characteristics. Note that the inductor loss

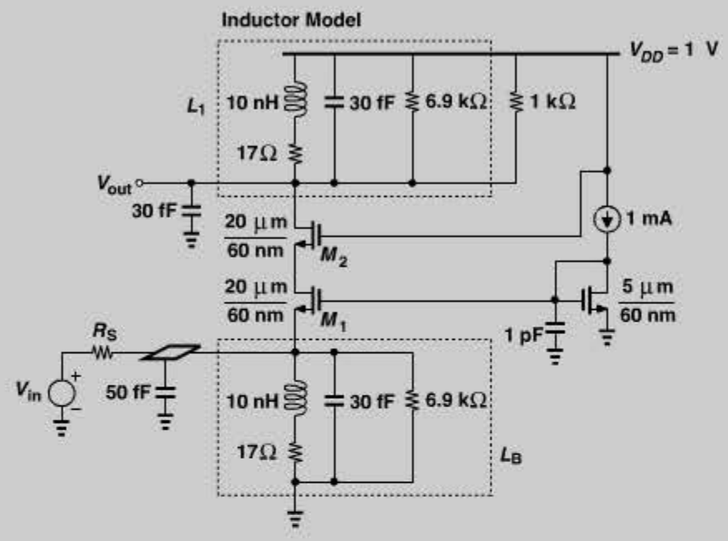


Figure 5.29 *CG LNA example.*

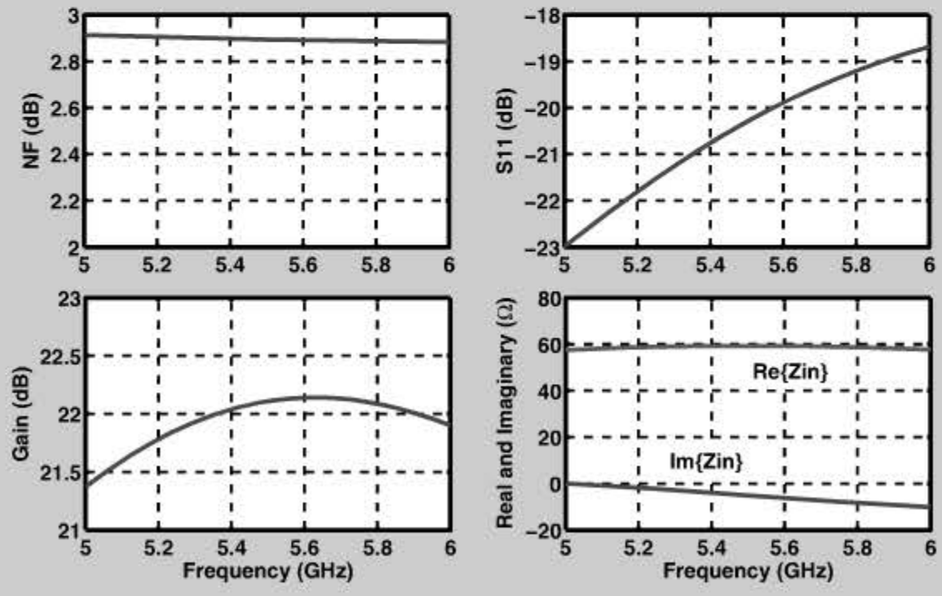


Figure 5.30 *Simulated characersitics of CG LNA example.*

(*Continues*)

Example 5.11 *(Continued)*

is modeled by series and parallel resistances so as to obtain a broadband representation (Chapter 7).

The simulation results reveal a relatively flat noise figure and gain from 5 to 6 GHz. The input return loss remains below −18 dB for this range even though we did not refine the choice of L_B.

5.3.4 Cascode CS Stage with Inductive Degeneration

Our study of the CS stage of Fig. 5.11(a) indicates that the feedback through the gate-drain capacitance many be exploited to produce the required real part, but it also leads to a negative resistance at lower frequencies. We must therefore seek a topology in which the input is "isolated" from the inductive load *and* the input resistance is established by means other than C_{GD}.

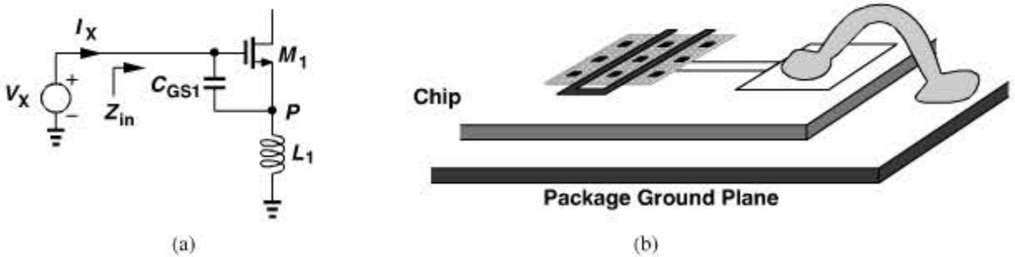
Let us first develop the latter concept. As mentioned in Section 5.2, we must employ active devices to provide a 50-Ω input resistance without the noise of a 50-Ω resistor. One such method employs a CS stage with inductive degeneration, as shown in Fig. 5.31(a). We first compute the input impedance of the circuit while neglecting C_{GD} and C_{SB}.[9] Flowing entirely through C_{GS1}, I_X generates a gate-source voltage of $I_X/(C_{GS1}s)$ and hence a drain current of $g_m I_X/(C_{GS1}s)$. These two currents flow through L_1, producing a voltage

$$V_P = \left(I_X + \frac{g_m I_X}{C_{GS1}s}\right) L_1 s. \tag{5.76}$$

Since $V_X = V_{GS1} + V_P$, we have

$$\frac{V_X}{I_X} = \frac{1}{C_{GS1}s} + L_1 s + \frac{g_m L_1}{C_{GS1}}. \tag{5.77}$$

Interestingly, the input impedance contains a frequency-independent real part given by $g_m L_1/C_{GS1}$. Thus, the third term can be chosen equal to 50 Ω.



(a) (b)

Figure 5.31 *(a) Input impedance of inductively-degenerated CS stage, (b) use of bond wire for degeneration.*

9. We also neglect channel-length modulation and body effect.

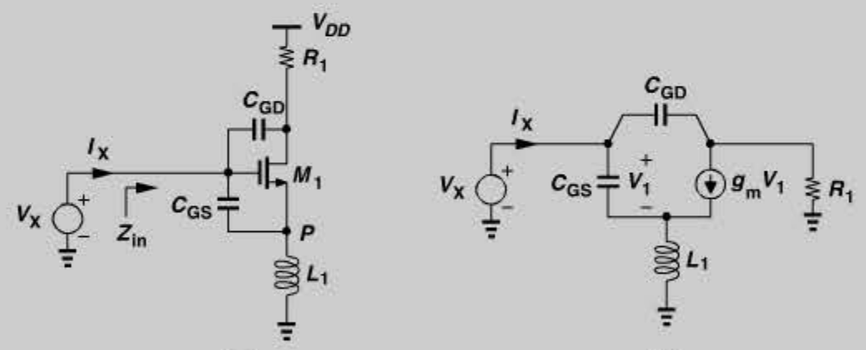
The third term in Eq. (5.77) carries a profound meaning: since $g_m/C_{GS1} \approx \omega_T\ (= 2\pi f_T)$, the input resistance is approximately equal to $L_1\omega_T$ and directly related to the f_T of the transistor. For example, in 65-nm technology, $\omega_T \approx 2\pi \times (160\,\text{GHz})$, dictating $L_1 \approx 50\,\text{pH}$ (!) for a real part of 50 Ω.

In practice, the degeneration inductor is often realized as a bond wire with the reasoning that the latter is inevitable in packaging and must be incorporated in the design. To minimize the inductance, a "downbond" can directly connect the source pad to a ground plane in the package [Fig. 5.31(b)], but even this geometry yields a value in the range of 0.5 to 1 nH—far from the 50-pH amount calculated above! That is, the input resistance provided by modern MOSFETs tends to be substantially higher than 50 Ω if a bond wire inductance is used.[10]

How do we obtain a 50-Ω resistance with $L_1 \approx 0.5$ nH? At operation frequencies far below f_T of the transistor, we can *reduce* the f_T. This is accomplished by increasing the channel length or simply placing an explicit capacitor in parallel with C_{GS}. For example, if $L_1 = 0.5$ nH, then f_T must be lowered to about 16 GHz.

Example 5.12

Determine the input impedance of the circuit shown in Fig. 5.32(a) if C_{GD} is not neglected and the drain is tied to a load resistance R_1. Assume $R_1 \approx 1/g_m$ (as in a cascode).



(a) (b)

Figure 5.32 *(a) Input impedance of CS stage in the presence of C_{GD}, (b) equivalent circuit.*

Solution:

From the equivalent circuit depicted in Fig. 5.32(b), we note the current flowing through L_1 is equal to $V_1 C_{GS}s + g_m V_1$ and hence

$$V_X = V_1 + (V_1 C_{GS}s + g_m V_1) L_1 s. \tag{5.78}$$

(Continues)

10. This is a rare case in which the transistor is too fast!

Example 5.12 *(Continued)*

Also, the current flowing through R_1 is equal to $I_X - V_1 C_{GS} s - g_m V_1$, leading to

$$V_X = (I_X - V_1 C_{GS} s - g_m V_1) R_1 + (I_X - V_1 C_{GS} s) \frac{1}{C_{GD} s}. \tag{5.79}$$

Substituting for V_1 from (5.78), we have

$$\frac{V_X}{I_X} = \frac{\left(R_1 + \dfrac{1}{C_{GD} s}\right)(L_1 C_{GS} s^2 + g_m L_1 s + 1)}{L_1 C_{GS} s^2 + (R_1 C_{GS} + g_m L_1) s + g_m R_1 + C_{GS}/C_{GD} + 1}. \tag{5.80}$$

If $R_1 \approx 1/g_m \ll |C_{GD} s|^{-1}$ and C_{GS}/C_{GD} dominates in the denominator, (5.80) reduces to

$$\frac{V_X}{I_X} \approx \left(\frac{1}{C_{GS} s} + L_1 s + \frac{g_m L_1}{C_{GS}}\right)\left[1 - \frac{2C_{GD}}{C_{GS}} - L_1 C_{GD} s^2 - \left(R_1 C_{GD} + g_m L_1 \frac{C_{GD}}{C_{GS}}\right) s\right] \tag{5.81}$$

Assuming that the first two terms in the square brackets are dominant, we conclude that the input resistance falls by a factor of $1 - 2C_{GD}/C_{GS}$.

Effect of Pad Capacitance In addition to C_{GD}, the input pad capacitance of the circuit also lowers the input resistance. To formulate this effect, we construct the equivalent circuit shown in Fig. 5.33(a), where C_{GS1}, L_1, and R_1 represent the three terms in Eq. (5.77), respectively. Denoting the series combination $jL_1\omega - j/(C_{GS1}\omega)$ by jX_1 and $-j/(C_{pad}\omega)$ by jX_2, we first transform $jX_1 + R_1$ to a parallel combination [Fig. 5.33(b)]. From Chapter 2,

$$R_P = \frac{X_1^2}{R_1}. \tag{5.82}$$

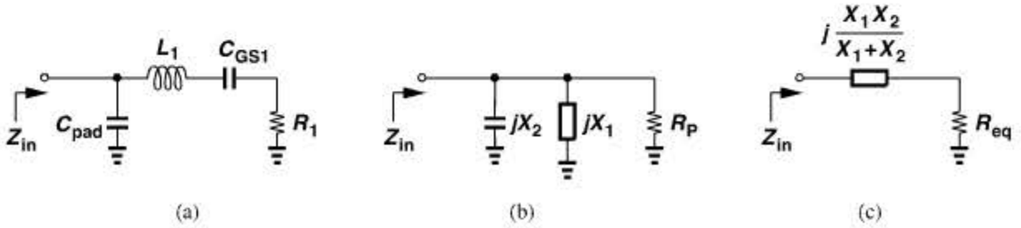


Figure 5.33 *(a) Equivalent circuit for inclusion of pad capacitance, (b) simplified circuit of (a), (c) simplified circuit of (b).*

We now merge the two parallel reactances into $jX_1X_2/(X_1 + X_2)$ and transform the resulting circuit to a series combination [Fig. 5.33(c)], where

$$R_{eq} = \left(\frac{X_1 X_2}{X_1 + X_2}\right)^2 \cdot \frac{1}{R_P} \tag{5.83}$$

$$= \left(\frac{X_2}{X_1 + X_2}\right)^2 R_1. \tag{5.84}$$

In most cases, we can assume $L_1\omega \ll 1/(C_{GS1}\omega) + 1/(C_{pad}\omega)$ at the frequency of interest, obtaining

$$R_{eq} \approx \left(\frac{C_{GS1}}{C_{GS1} + C_{pad}}\right)^2 R_1. \tag{5.85}$$

For example, if $C_{GS1} \approx C_{pad}$, then the input resistance falls by a factor of four.

We can now make two observations. First, the effect of the gate-drain and pad capacitance suggests that the transistor f_T need not be reduced so much as to create $R_1 = 50\,\Omega$. Second, since the degeneration inductance necessary for $Re\{Z_{in}\} = 50\,\Omega$ is insufficient to resonate with $C_{GS1} + C_{pad}$, another inductor must be placed in series with the gate as shown in Fig. 5.34, where it is assumed L_G is off-chip.

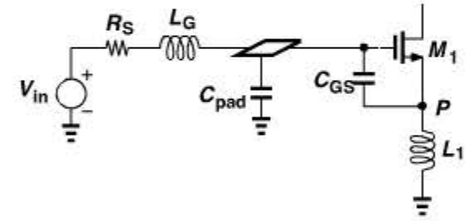


Figure 5.34 *Addition of L_G for input matching.*

Example 5.13

A 5-GHz LNA requires a value of 2 nH for L_G. Discuss what happens if L_G is integrated on the chip and its Q does not exceed 5.

Solution:

With $Q = 5$, L_G suffers from a series resistance equal to $L_G\omega/Q = 12.6\,\Omega$. This value is not much less than $50\,\Omega$, degrading the noise figure considerably. For this reason, L_G is typically placed off-chip.

NF Calculation Let us now compute the noise figure of the CS circuit, excluding the effect of channel-length modulation, body effect, C_{GD}, and C_{pad} for simplicity (Fig. 5.35). The noise of M_1 is represented by I_{n1}. For now, we assume the output of interest is the

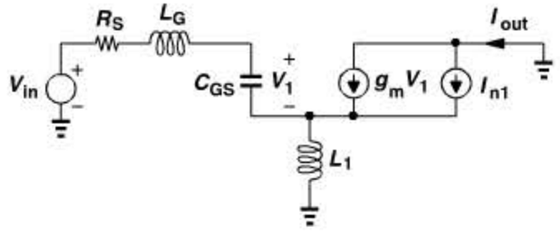


Figure 5.35 *Equivalent circuit for computation of NF.*

current I_{out}. We have

$$I_{out} = g_m V_1 + I_{n1}. \tag{5.86}$$

Also, since L_1 sustains a voltage of $L_1 s(I_{out} + V_1 C_{GS1} s)$, a KVL around the input loop yields

$$V_{in} = (R_S + L_G s)V_1 C_{GS1} s + V_1 + L_1 s(I_{out} + V_1 C_{GS1} s). \tag{5.87}$$

Substituting for V_1 from (5.86) gives

$$V_{in} = I_{out} L_1 s + \frac{(L_1 + L_G)C_{GS1} s^2 + 1 + R_S C_{GS1} s}{g_m}(I_{out} - I_{n1}). \tag{5.88}$$

The input network is designed to resonate at the frequency of interest, ω_0. That is, $(L_1 + L_G)$ $C_{GS1} = \omega_0^{-2}$ and hence, $(L_1 + L_G)C_{GS1} s^2 + 1 = 0$ at $s = j\omega_0$. We therefore obtain

$$V_{in} = I_{out}\left(jL_1\omega_0 + \frac{jR_S C_{GS1}\omega_0}{g_m}\right) - I_{n1}\frac{jR_S C_{GS1}\omega_0}{g_m}. \tag{5.89}$$

The coefficient of I_{out} represents the transconductance gain of the circuit (including R_S):

$$\left|\frac{I_{out}}{V_{in}}\right| = \frac{1}{\omega_0\left(L_1 + \dfrac{R_S C_{GS1}}{g_m}\right)}. \tag{5.90}$$

Now, recall from Eq. (5.77) that, for input matching, $g_m L_1/C_{GS1} = R_S$. Since $g_m/C_{GS1} \approx \omega_T$,

$$\left|\frac{I_{out}}{V_{in}}\right| = \frac{\omega_T}{2\omega_0}\cdot\frac{1}{R_S}. \tag{5.91}$$

Interestingly, the transconductance of the circuit remains independent of L_1, L_G, and g_m so long as the input is matched.

Setting V_{in} to zero in Eq. (5.89), we compute the output noise due to M_1:

$$|I_{n,out}|_{M1} = |I_{n1}|\frac{R_S C_{GS1}}{g_m L_1 + R_S C_{GS1}}, \tag{5.92}$$

which, for $g_m L_1/C_{GS1} = R_S$, reduces to

$$|I_{n,out}|_{M1} = \frac{|I_{n1}|}{2}, \tag{5.93}$$

and hence

$$\overline{I^2_{n,out}}|_{M1} = kT\gamma g_m. \tag{5.94}$$

Dividing the output noise current by the transconductance of the circuit and by $4kTR_S$ and adding unity to the result, we arrive at the noise figure of the circuit [2]:

$$\text{NF} = 1 + g_m R_S \gamma\left(\frac{\omega_0}{\omega_T}\right)^2. \tag{5.95}$$

It is important to bear in mind that this result holds only at the input resonance frequency and if the input is matched.

Example 5.14

A student notes from Eq. (5.95) above that, if the transistor width and bias current are scaled *down* proportionally, then g_m and C_{GS1} decrease while $g_m/C_{GS1} = \omega_T$ remains constant. That is, the noise figure decreases while the power dissipation of the circuit also decreases! Does this mean we can obtain NF = 1 with zero power dissipation?

Solution:

As C_{GS1} decreases, $L_G + L_1$ must increase proportionally to maintain a constant ω_0. Suppose L_1 is fixed and we simply increase L_G. As C_{GS1} approaches zero and L_G infinity, the Q of the input network ($\approx L_G\omega_0/R_S$) also goes to infinity, providing an *infinite voltage gain at the input*. Thus, the noise of R_S overwhelms that of M_1, leading to NF = 1. This result is not surprising; after all, in a circuit such as the network of Fig. 5.36, $|V_{out}/V_{in}| = (R_S C_a \omega_0)^{-1}$ at resonance, implying that the voltage gain approaches infinity if C_a goes to zero (and L_a goes to infinity so that ω_0 is constant). In practice, of course, the inductor suffers from a finite Q (and parasitic capacitances), limiting the performance.

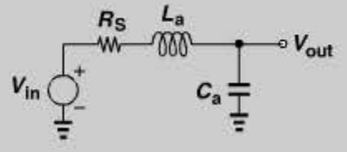


Figure 5.36 *Equivalent circuit of CS input network.*

What if we keep L_G constant and increase the degeneration inductance, L_1? The NF still approaches 1 but the transconductance of the circuit, Eq. (5.90), falls to zero if C_{GS1}/g_m remains fixed.[11] That is, the circuit provides a zero-dB noise figure but with zero gain.

11. If C_{GS1}/g_m is constant and L_1 increases, the input cannot remain matched and Eq. (5.95) is invalid.

The above example suggests that maximizing L_G can minimize the noise figure by providing voltage *gain* from V_{in} to the gate of M_1. The reader can prove that this gain is given by

$$\frac{V_G}{V_{in}} = \frac{1}{2}\left(1 + \frac{L_G\omega_0}{R_S}\right). \tag{5.96}$$

Note that $L_G\omega_0/R_S$ represents the Q of the series combination of L_G and R_S. Indeed, as explained below, the design procedure begins with the maximum available value of L_G (typically an off-chip inductor) whose parasitic capacitances are negligible. The voltage gain in the input network (typically as high as 6 dB) does lower the IP_3 and P_{1dB} of the LNA, but the resulting values still prove adequate in most applications.

We now turn our attention to the output node of the circuit. As explained in Section 5.3.1, an inductive load attached to a common-source stage introduces a negative resistance due to the feedback through C_{GD}. We therefore add a cascode transistor in the output branch to suppress this effect. Figure 5.37 shows the resulting circuit, where R_1 models the loss of L_D. The voltage gain is equal to the product of the circuit's transconductance [Eq. (5.91)] and the load resistance, R_1:[12]

$$\frac{V_{out}}{V_{in}} = \frac{\omega_T}{2\omega_0}\frac{R_1}{R_S} \tag{5.97}$$

$$= \frac{R_1}{2L_1\omega_0}. \tag{5.98}$$

The effect of C_{GD1} on the input impedance may still require attention because the impedance seen at the source of M_2, R_X, rises sharply at the output resonance frequency. From Eq. (5.64),

$$R_X = \frac{R_1 + r_{O2}}{1 + g_m r_{O2}}. \tag{5.99}$$

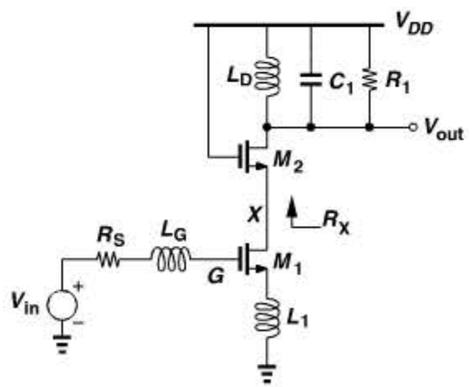


Figure 5.37 *Inductively-degenerated cascode CS LNA.*

12. The output impedance of the cascode is assumed much higher than R_1.

Using the transconductance expression in (5.90) and V_G/V_{in} in (5.96), we compute the voltage gain from the gate to the drain of M_1:

$$\frac{V_X}{V_G} = \frac{R_S}{L_1\omega_0} \cdot \frac{R_1 + r_{O2}}{(1 + g_{m2}r_{O2})(R_S + L_G\omega_0)}. \tag{5.100}$$

Since $R_S \gg L_1\omega_0$ (why?) and the second fraction is typically near or higher than unity, C_{GD1} may suffer from substantial Miller multiplication at the output resonance frequency.

In the foregoing noise figure calculation, we have not included the noise contribution of M_2. As formulated for the cascode CG stage in Section 5.3.3, the noise of the cascode device begins to manifest itself if the frequency of operation exceeds roughly $(2r_{O1}C_X)^{-1}$.

Example 5.15

Determine the noise figure of the cascode CS stage of Fig. 5.37, including the noise contributed by R_1 but neglecting the noise of M_2.

Solution:

Dividing the noise of R_1 by the gain given by (5.98) and the noise of R_S and adding the result to the noise figure in (5.95), we have

$$\text{NF} = 1 + g_m R_S \gamma \left(\frac{\omega_0}{\omega_T}\right)^2 + \frac{4R_S}{R_1}\left(\frac{\omega_0}{\omega_T}\right)^2. \tag{5.101}$$

Design Procedure Having developed a good understanding of the cascode CS LNA of Fig. 5.37, we now describe a procedure for designing the circuit. The reader is encouraged to review the CG design procedure. The procedure begins with four knowns: the frequency of operation, ω_0, the value of the degeneration inductance, L_1, the input pad capacitance, C_{pad}, and the value of the input series inductance, L_G. Each of the last three knowns is somewhat flexible, but it is helpful to select some values, complete the design, and iterate if necessary.

Governing the design are the following equations:

$$\frac{1}{(L_G + L_1)(C_{GS1} + C_{pad})} = \omega_0^2 \tag{5.102}$$

$$\left(\frac{C_{GS1}}{C_{GS1} + C_{pad}}\right)^2 L_1\omega_T = R_S. \tag{5.103}$$

With ω_0 known, C_{GS1} is calculated from (5.102), and ω_T and g_m ($= \omega_T C_{GS1}$) from (5.103). We then return to the plots of g_m and f_T in Fig. 5.25 and determine whether a transistor width can yield the necessary g_m and f_T simultaneously. In deep-submicron technologies and for operation frequencies up to a few tens of gigahertz, the f_T is likely to be "too high," but the pad capacitance alleviates the issue by transforming the input resistance to a lower value. If the requisite f_T is quite low, a capacitance can be added to C_{pad}. On the other hand, if the pad capacitance is so large as to demand a very high f_T, the degeneration inductance can be increased.

In the next step, the dimensions of the cascode device are chosen equal to those of the input transistor. As mentioned in Section 5.3.3 for the cascode CG stage, the width of the cascode device only weakly affects the performance. Also, the layout of M_1 and M_2 can follow the structure shown in Fig. 5.27 to minimize the capacitance at node X.

The design procedure now continues with selecting a value for L_D such that it resonates at ω_0 with the drain-bulk and drain-gate capacitances of M_2, the input capacitance of the next stage, and the inductors's own parasitic capacitance. If the parallel equivalent resistance of L_D results in a gain, $R_1/(2L_1\omega_0)$, greater than required, then an explicit resistor can be placed in parallel with L_D to lower the gain and widen the bandwidth.

In the last step of the design, we must examine the input match. Due to the Miller multiplication of C_{GD1} (Example 5.12), it is possible that the real and imaginary parts depart from their ideal values, necessitating some adjustment in L_G.

The foregoing procedure typically leads to a design with a relatively low noise figure, around 1.5 to 2 dB—depending on how large L_G can be without displaying excessive parasitic capacitances. Alternatively, the design procedure can begin with known values for NF and L_1 and the following two equations:

$$\text{NF} = 1 + g_{m1}R_S\gamma\left(\frac{\omega_0}{\omega_T}\right)^2 \tag{5.104}$$

$$R_S = \left(\frac{C_{GS1}}{C_{GS1} + C_{pad}}\right)^2 L_1\omega_T, \tag{5.105}$$

where the noise of the cascode transistor and the load is neglected. The necessary values of ω_T and g_{m1} can thus be computed ($g_{m1}/C_{GS1} \approx \omega_T$). If the plots in Fig. 5.25 indicate that the device f_T is too high, then additional capacitance can be placed in parallel with C_{GS1}. Finally, L_G is obtained from Eq. (5.102). (If advanced packaging minimizes inductances, then L_1 can be integrated on the chip and assume a small value.)

The overall LNA appears as shown in Fig. 5.38, where the antenna is capacitively tied to the receiver to isolate the LNA bias from external connections. The bias current of M_1 is

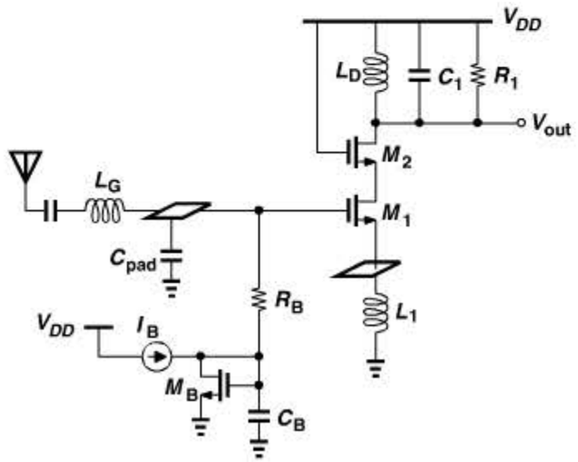


Figure 5.38 *Inductively-degenerated CS stage with pads and bias network.*

established by M_B and I_B, and resistor R_B and capacitor C_B isolate the signal path from the noise of I_B and M_B. The source-bulk capacitance of M_1 and the capacitance of the pad at the source of M_1 may slightly alter the input impedance and must be included in simulations.

Example 5.16

How is the value of R_B chosen in Fig. 5.38?

Solution:

Since R_B appears *in parallel* with the signal path, its value must be maximized. Is $R_B = 10R_S$ sufficiently high? As illustrated in Fig. 5.39, the series combination of R_S and L_G can be transformed to a parallel combination with $R_P \approx Q^2R_S \approx (L_G\omega_0/R_S)^2R_S$. From Eq. (5.96), we note that a voltage gain of, say, 2 at the input requires $Q = 3$, yielding $R_P \approx 450\,\Omega$. Thus, $R_B = 10R_S$ becomes *comparable* with R_P, raising the noise figure and lowering the voltage gain. In other words, R_B must remain much greater than R_P.

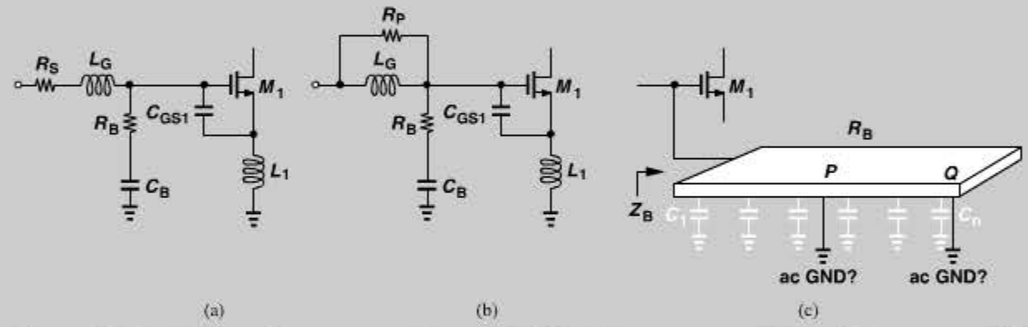


Figure 5.39 *(a) Effect of bias resistor R_B on CS LNA, (b) conversion of R_S and L_G to a parallel network, (c) effect of distributed capacitance of R_B.*

Large resistors may suffer from significant parasitic capacitance. However, increasing the *length* of a resistor does not load the signal path anymore even though it leads to a larger overall parasitic capacitance. To understand this point, consider the arrangement shown in Fig. 5.39(b), where the parasitic capacitance of R_B is represented as distributed elements C_1-C_n. Which node should be bypassed to ground, P or Q? We recognize that Z_B is *higher* if Q is bypassed even though the longer resistor has a higher capacitance. Thus, longer bias resistors are better. Alternatively, a small MOSFET acting as a resistor can be used here.

The choice between the CG and CS LNA topologies is determined by the trade-off between the robustness of the input match and the lower bound on the noise figure. The former provides an accurate input resistance that is relatively independent of package parasitics, whereas the latter exhibits a lower noise figure. We therefore select the CG stage if the required LNA noise figure can be around 4 dB, and the CS stage for lower values.

An interesting point of contrast between the CG and CS LNAs relates to the contribution of the load resistor, R_1, to the noise figure. Equation (5.58) indicates that in a CG stage, this contribution, $4R_S/R_1$, is equal to 4 divided by the voltage gain from the input

source to the output. Thus, for a typical gain of 10, this contribution reaches 0.4, a significant amount. For the inductively-degenerated CS stage, on the other hand, Eq. (5.101) reveals that the contribution is equal to $4R_S/R_1$ *multiplied* by $(\omega_0/\omega_T)^2$. Thus, for operation frequencies well below the f_T of the transistor, the noise contribution of R_1 becomes negligible.

Example 5.17

It is believed that input matching holds across a wider bandwidth for the CG stage than for the inductively-degenerated CS stage. Is this statement correct?

Solution:

Consider the equivalent circuits shown in Fig. 5.40 for the two LNA configurations, where $R_1 = 50\,\Omega$, C_1 and C_2 are roughly equal, and the inductors represent (inevitable) bond

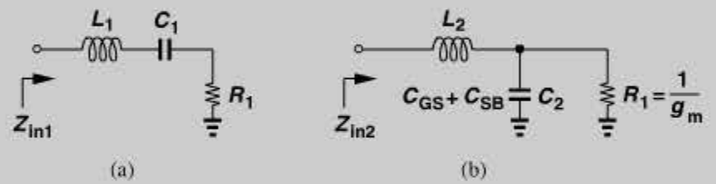


Figure 5.40 *Input networks of (a) CS and (b) CG LNAs.*

wires. For the CS stage [Fig. 5.40(a)], we have

$$Re\{Z_{in1}\} = R_1 \tag{5.106}$$

$$Im\{Z_{in1}\} = \frac{L_1 C_1 \omega^2 - 1}{C_1 \omega}. \tag{5.107}$$

If the center frequency of interest is ω_0 ($= 1/\sqrt{L_1 C_1}$) and $\omega = \omega_0 + \Delta\omega$, then

$$Im\{Z_{in1}\} \approx 2L_1 \Delta\omega \frac{L_1 \Delta\omega}{\omega_0}. \tag{5.108}$$

That is, the imaginary part varies in proportion to deviation from the center frequency, limiting the bandwidth across which $|S_{11}|$ remains acceptably low.

In the network of Fig. 5.40(b), on the other hand,

$$Re\{Z_{in2}\} = \frac{R_1}{1 + R_1^2 C_2^2 \omega^2} \tag{5.109}$$

$$Im\{Z_{in2}\} = L_2\omega - \frac{R_1^2 C_2 \omega}{1 + R_1^2 C_2^2 \omega^2}. \tag{5.110}$$

Example 5.17 *(Continued)*

In practice, $1/(R_1C_2)$ is comparable with the ω_T of the transistor [e.g., if $R_1 = 1/g_m$ and $C_2 = C_{GS}$, then $1/(R_1C_2) \approx \omega_T$]. Thus, for $\omega \ll \omega_T$,

$$Re\{Z_{in2}\} \approx R_1 \tag{5.111}$$

$$Im\{Z_{in2}\} \approx (L_2 - R_1^2 C_2)\omega. \tag{5.112}$$

Interestingly, if $L_2 = R_1^2 C_2$, then $Im\{Z_{in2}\}$ falls to zero and becomes *independent* of frequency. Thus the CG stage indeed provides a much broader band at its input, another advantage of this topology.

Example 5.18

Design a cascode CS LNA for a center frequency of 5.5 GHz in 65-nm CMOS technology.

Solution:

We begin with a degeneration inductance of 1 nH and the same input transistor as that in the CG stage of Example 5.11. Interestingly, with a pad capacitance of 50 fF, the input resistance happens to be around 60 Ω. (Without the pad capacitance, $Re\{Z_{in}\}$ is in the vicinity of 600 Ω.) We thus simply add enough inductance in series with the gate ($L_G = 12$ nH) to null the reactive component at 5.5 GHz. The design of the cascode device and the output network is identical to that of the CG example.

Figure 5.41 shows the details of the design and Fig. 5.42 the simulated characteristics. We observe that the CS stage has a higher gain, a lower noise figure, and a narrower bandwidth than the CG stage in Example 5.11.

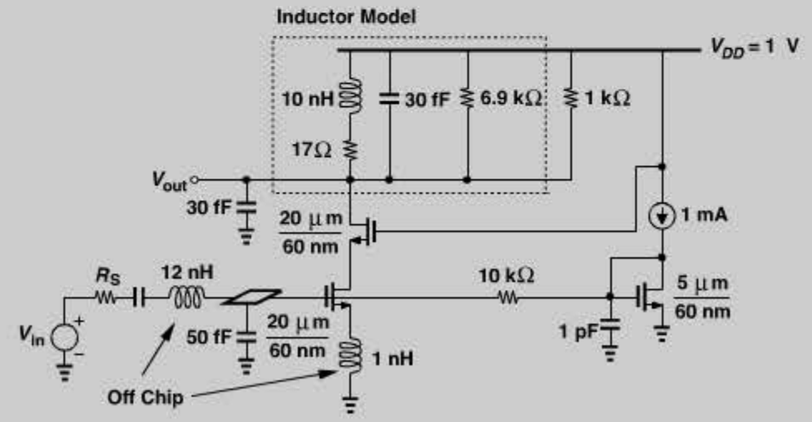


Figure 5.41 *CS LNA example.*

(Continues)

Example 5.18 *(Continued)*

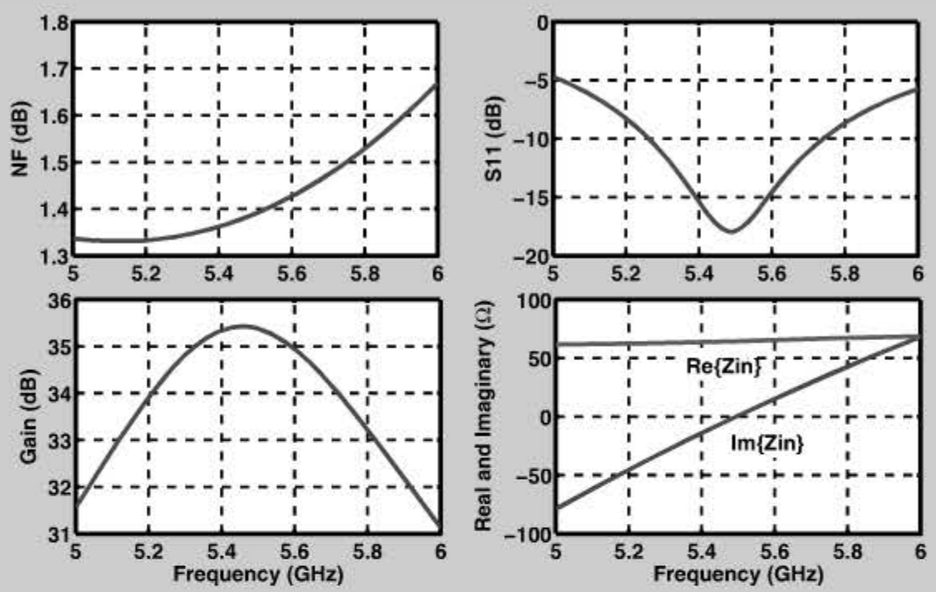


Figure 5.42 *Simulated characteristics of CS LNA example.*

5.3.5 Variants of Common-Gate LNA

As revealed by Eq. (5.57), the noise figure and input matching of the CG stage are inextricably related if channel-length modulation is negligible, a common situation in older CMOS technologies. For this reason, a number of efforts have been made to add another degree of freedom to the design so as to avoid this relationship. In this section, we describe two such examples.

Figure 5.43 shows a topology incorporating voltage-voltage feedback [3].[13] The block having a gain (or attenuation factor) of α senses the output voltage and subtracts a fraction thereof from the input. (Note that M_1 operates as a subtractor because $I_{D1} \propto V_F - V_{in}$.) The loop transmission can be obtained by breaking the loop at the gate of M_1 and is equal to $g_m Z_L \cdot \alpha$.[14] If channel-length modulation and body effect are neglected, the closed-loop input impedance is equal to the open-loop input impedance, $1/g_m$, multiplied by $1 + g_m Z_L \alpha$:

$$Z_{in} = \frac{1}{g_m} + \alpha Z_L. \tag{5.113}$$

At resonance,

$$Z_{in} = \frac{1}{g_m} + \alpha R_1. \tag{5.114}$$

13. This technique was originally devised for bipolar stages.
14. The input impedance of the feedback circuit is absorbed in Z_L.

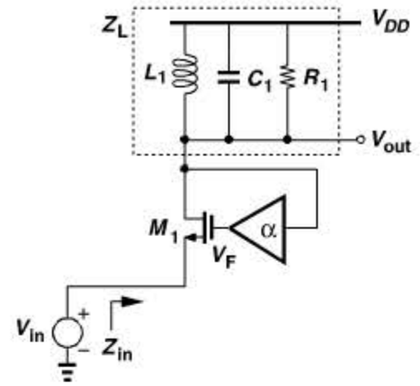


Figure 5.43 *CG LNA with feedback.*

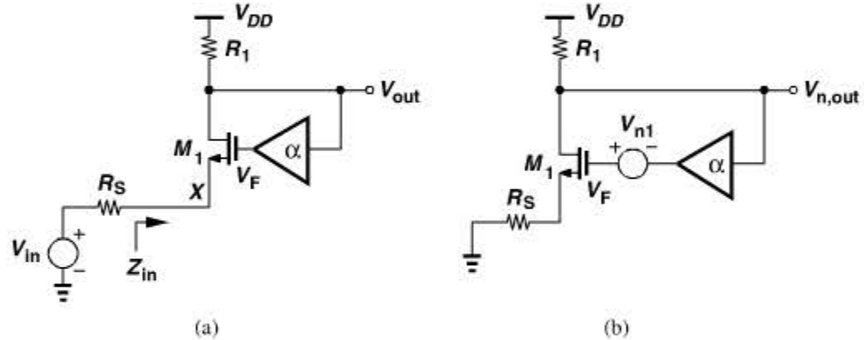


Figure 5.44 *(a) Input impedance and (b) noise behavior of CG stage with feedback.*

The input resistance can therefore be substantially *higher* than $1/g_m$, but how about the noise figure? We first calculate the gain with the aid of the circuit depicted in Fig. 5.44(a). The voltage gain from X to the output is equal to the open-loop gain, $g_m R_1$, divided by $1 + \alpha g_m R_1$ (at the resonance frequency). Thus,

$$\frac{V_{out}}{V_{in}} = \frac{Z_{in}}{Z_{in} + R_S} \cdot \frac{g_m R_1}{1 + \alpha g_m R_1} \tag{5.115}$$

$$= \frac{R_1}{\dfrac{1}{g_m} + \alpha R_1 + R_S}, \tag{5.116}$$

which reduces to $R_1/(2R_S)$ if the input is matched.

For output noise calculation, we construct the circuit of Fig. 5.44(b), where V_{n1} represents the noise voltage of M_1 and noise of the feedback circuit is neglected. Since R_S carries a current equal to $-V_{n,out}/R_1$ (why?), we recognize that $V_{GS1} = \alpha V_{n,out} + V_{n1} + V_{n,out} R_S / R_1$. Equating $g_m V_{GS1}$ to $-V_{n,out}/R_1$ yields

$$g_m \left(\alpha V_{n,out} + V_{n1} + \frac{R_S}{R_1} V_{n,out} \right) = -\frac{V_{n,out}}{R_1}, \tag{5.117}$$

and hence

$$V_{n,out}|_{M1} = \frac{-g_m V_{n1}}{g_m(\alpha + \frac{R_S}{R_1}) + \frac{1}{R_1}}. \tag{5.118}$$

The noise current of R_1 is multiplied by the output impedance of the circuit, R_{out}. The reader can show that R_{out} is equal to R_1 in parallel with $(1 + g_m R_S)/(\alpha g_m)$. Summing this noise and that of M_1, dividing the result by the square of (5.116) and $4kTR_S$, and assuming the input is matched, we have

$$\text{NF} = 1 + \frac{\gamma}{g_m R_S} + \frac{R_S}{R_1}\left(1 + \frac{1}{g_m R_S}\right)^2. \tag{5.119}$$

That is, the NF can be lowered by raising g_m. Note that this result is identical to that expressed by Eq. (5.57) for the simple CG stage, except that $g_m R_S$ need not be equal to unity here. For example, if $g_m R_S = 4$ and $\gamma = 1$, then the first two terms yield a noise figure of 0.97 dB. In Problem 5.15 we reexamine these results if channel-length modulation is not neglected.

Example 5.19

How is the feedback factor, α, chosen in the above circuit?

Solution:

The design begins with the choice of $g_m R_S$ and $R_1/(2R_S)$ to obtain the required noise figure and voltage gain, A_v. For input matching, $g_m R_S - 1 = \alpha g_m R_1 = \alpha g_m (2A_v R_S)$. It follows that

$$\alpha = \frac{g_m R_S - 1}{2 g_m R_S A_v}. \tag{5.120}$$

For example, if $g_m R_S = 4$ and $A_v = 6$ (= 15.6 dB), then $R_1 = 600\,\Omega$ and $\alpha = 1/16$.

Another variant of the CG LNA employs *feedforward* to avoid the tight relationship between the input resistance and the noise figure [4]. Illustrated in Fig. 5.45(a), the idea is to amplify the input by a factor of $-A$ and apply the result to the gate of M_1. For an input voltage change of ΔV, the gate-source voltage changes by $-(1 + A)\Delta V$ and the drain current by $-(1 + A)g_m \Delta V$. Thus, the g_m is "boosted" by a factor of $1 + A$ [4], *lowering* the input impedance to $R_{in} = [g_m(1 + A)]^{-1}$ and raising the voltage gain from the source to the drain to $(1 + A)g_m R_1$ (at resonance).

We now compute the noise figure with the aid of the equivalent circuit shown in Fig. 5.45(b). Since the current flowing through R_S is equal to $-V_{n,out}/R_1$, the source voltage is given by $-V_{n,out}R_S/R_1$ and the gate voltage by $(-V_{n,out}R_S/R_1)(-A) + V_{n1}$. Multiplying

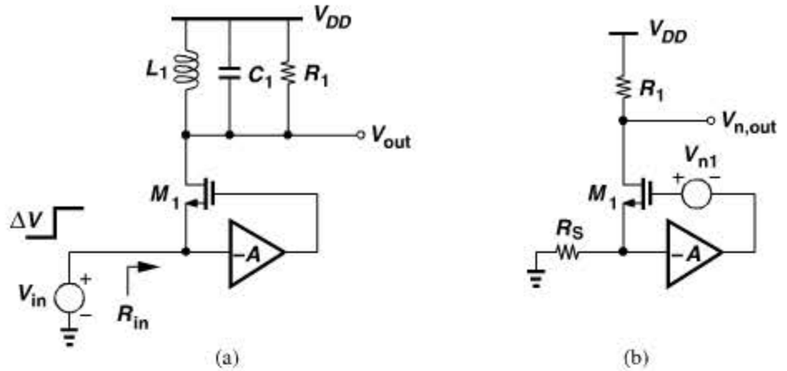


Figure 5.45 *(a) CG stage with feedforward, (b) calculation of NF.*

the gate-source voltage by g_m and equating the result to $-V_{n,out}/R_1$, we have

$$g_m\left(A\frac{R_S}{R_1}V_{n,out} + V_{n1} + \frac{R_S}{R_1}V_{n,out}\right) = -\frac{V_{n,out}}{R_1}, \tag{5.121}$$

and hence

$$V_{n,out}|_{M1} = \frac{-g_m R_1 V_{n1}}{(1 + A)g_m R_S + 1}. \tag{5.122}$$

This expression reduces to $-g_m R_1 V_{n1}/2$ if the input is matched, indicating that half of the noise current of M_1 flows through R_1.[15] With input matching, the voltage gain from the left terminal of R_S in Fig. 5.45(b) to the output is equal to $(1 + A)g_m R_1/2$. We therefore sum the output noise contribution of M_1 and R_1, divide the result by the square of this gain and the noise of R_S, and add unity:

$$\text{NF} = 1 + \frac{\gamma}{1 + A} + \frac{4R_S}{R_1}. \tag{5.123}$$

This equation reveals that the NF can be lowered by raising A with the constraint $g_m(1 + A) = R_S^{-1}$ (for input matching).

The above analysis has neglected the noise of the gain stage A in Fig. 5.45(a). We show in Problem 5.17 that the input-referred noise of this stage, V_{nA}^2, is multiplied by A and added to V_{n1} in Eq. (5.122), leading to an overall noise figure equal to

$$\text{NF} = 1 + \frac{\gamma}{1 + A} + \frac{4R_S}{R_1} + \frac{A^2}{(1 + A)^2}\frac{\overline{V_{nA}^2}}{4kTR_S}. \tag{5.124}$$

In other words, $\overline{V_{nA}^2}$ is referred to the input by a factor of $A^2/(1 + A)^2$, which is not much less than unity. For this reason, it is difficult to realize A by an active circuit.

It is possible to obtain the voltage gain through the use of an on-chip transformer. As shown in Fig. 5.46 [4], for a coupling factor of k between the primary and the secondary and

15. Where does the other half go?

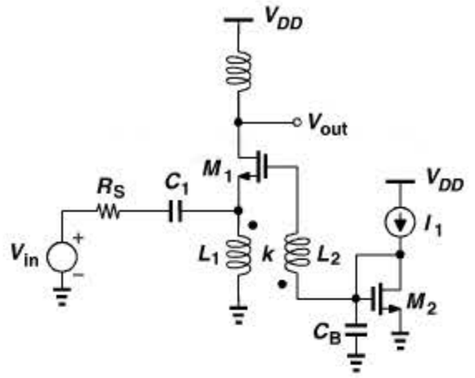


Figure 5.46 *CG stage with transformer feedforward.*

a turns ratio of n ($= \sqrt{L_2/L_1}$), the transformer provides a voltage gain of kn. The direction of the currents is chosen so as to yield a negative sign. However, on-chip transformer geometries make it difficult to achieve a voltage gain higher than roughly 3, even with stacked spirals [5]. Also, the loss in the primary and secondary contributes noise.

5.3.6 Noise-Cancelling LNAs

In our previous derivations of the noise figure of LNAs, we have observed three terms: a value of unity arising from the noise of R_S itself, a term representing the contribution of the input transistor, and another related to the noise of the load resistor. "Noise-cancelling LNAs" aim to cancel the second term [6]. The underlying principle is to identify *two* nodes in the circuit at which the signal appears with opposite polarities but the noise of the input transistor appears with the same polarity. As shown in Fig. 5.47, if nodes X and Y satisfy this condition, then their voltages can be properly scaled and summed such that the signal components add and the noise components cancel.

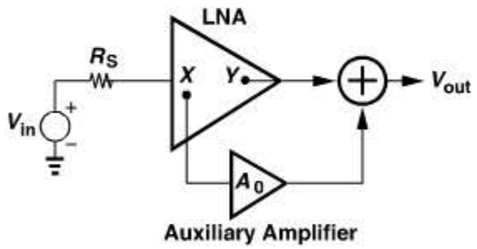


Figure 5.47 *Conceptual illustration of noise-cancelling LNAs.*

The CS stage with resistive feedback studied in Section 5.3.2 serves as a good candidate for noise cancellation because, as shown in Fig. 5.48(a), the noise current of M_1 flows through R_F and R_S, producing voltages at the gate and drain of the transistor with the same polarity. The signal, on the other hand, experiences inversion. Thus, as conceptually shown in Fig. 5.48(b), if V_X is amplified by $-A_1$ and added to V_Y, the noise of M_1 can be removed [6]. Since the noise voltages at nodes Y and X bear a ratio of $1 + R_F/R_S$ (why?), we choose $A_1 = 1 + R_F/R_S$. The signal experiences two additive gains: the original

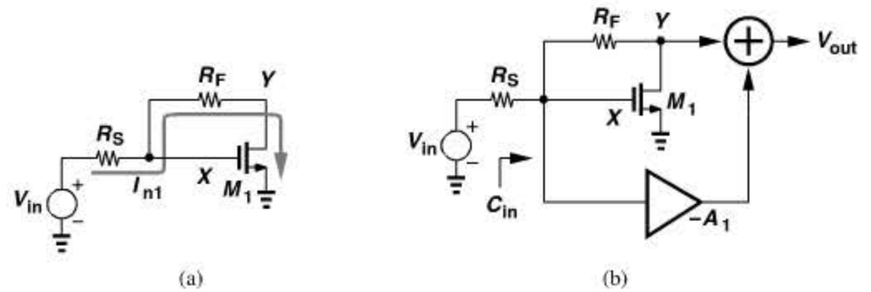


Figure 5.48 *(a) Noise of input transistor in a feedback CS stage, (b) cancellation of noise of M_1.*

gain, $V_Y/V_X = 1 - g_m R_F = 1 - R_F/R_S$ (if the input is matched), and the additional gain, $-(1 + R_F/R_S)$. It follows that

$$\frac{V_{out}}{V_X} = 1 - \frac{R_F}{R_S} - \left(1 + \frac{R_F}{R_S}\right) \tag{5.125}$$

$$= -\frac{2R_F}{R_S}, \tag{5.126}$$

if the input is matched. The gain V_{out}/V_{in} is half of this value.

Let us now compute the noise figure of the circuit, assuming that the auxiliary amplifier exhibits an input-referred noise voltage V_{nA1} and a high input impedance. Recall from Section 5.3.2 that the noise voltage of R_F appears directly at the output as $4kTR_F$. Adding this noise to $A_1^2\overline{V_{nA1}^2}$, dividing the result by $(R_F/R_S)^2$ and $4kTR_S$, and adding unity, we obtain the noise figure as

$$\text{NF} = 1 + \frac{R_S}{R_F} + A_1^2\overline{V_{nA1}^2}\frac{R_S}{4kTR_F^2}. \tag{5.127}$$

Since $A_1 = 1 + R_F/R_S$,

$$\text{NF} = 1 + \frac{R_S}{R_F} + \frac{\overline{V_{nA1}^2}}{4kTR_S}\left(1 + \frac{R_S}{R_F}\right)^2. \tag{5.128}$$

The NF can therefore be minimized by maximizing R_F and minimizing $\overline{V_{nA1}^2}$. Note that R_S/R_F is the inverse of the gain and hence substantially less than unity, making the third term approximately equal to $\overline{V_{nA1}^2}/(4kTR_S)$. That is, the noise of the auxiliary amplifier is directly referred to the input and must therefore be much less than that of R_S.

The input capacitance, C_{in}, arising from M_1 and the auxiliary amplifier degrades both S_{11} and the noise cancellation, thereby requiring a series (or parallel) inductor at the input for operation at very high frequencies. It can be proved [6] that the frequency-dependent noise figure is expressed as

$$\text{NF}(f) = \text{NF}(0) + [\text{NF}(0) - 1 + \gamma]\left(\frac{f}{f_0}\right)^2, \tag{5.129}$$

where NF(0) is given by (5.128) and $f_0 = 1/(\pi R_S C_{in})$.

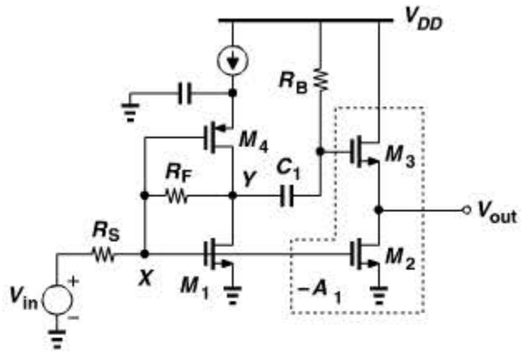


Figure 5.49 *Example of noise-cancelling LNA.*

Figure 5.49 depicts an implementation of the circuit [6]. Here, M_2 and M_3 serve as a CS amplifier, providing a voltage gain of $g_{m2}/(g_{m3} + g_{mb3})$, and also as the summing circuit. Transistor M_3 operates as a source follower, sensing the signal and noise at the drain of M_1. The first stage is similar to that studied in Example 5.7.

Example 5.20

Figure 5.50 shows an alternative implementation of a noise-cancelling LNA that also performs single-ended to differential conversion. Neglecting channel-length modulation, determine the condition for noise cancellation and derive the noise figure.

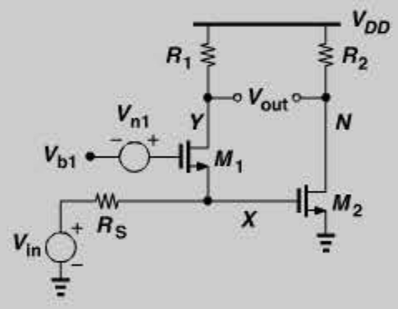


Figure 5.50 *CG/CS stage as a noise-cancelling LNA.*

Solution:

The circuit follows the noise cancellation principle because (a) the noise of M_1, V_{n1}, sees a source follower path to node X and a common-source path to node Y, exhibiting opposite polarities at these two nodes, and (b) the signal sees a common-gate path through X and Y, exhibiting the same polarity. Transistor M_1 produces half of its noise voltage at X if the input is matched (why?). Transistor M_2 senses this noise and amplifies it by a factor of $-g_{m2}R_2$. The reader can prove that the output noise of the CG stage due to M_1 (at Y) is equal to $(V_{n1}/2)g_{m1}R_1$. For noise cancellation, we must have

$$g_{m1}R_1\frac{V_{n1}}{2} = g_{m2}R_2\frac{V_{n1}}{2}, \tag{5.130}$$

Example 5.20 *(Continued)*

and, since $g_{m1} = 1/R_S$,

$$R_1 = g_{m2}R_2R_S. \tag{5.131}$$

If the noise of M_1 is cancelled, the noise figure arises from the contributions of M_2, R_1, and R_2. The noise at Y is equal to $4kTR_1$ and at N equal to $4kT\gamma g_{m2}R_2^2 + 4kTR_2$. Since the total voltage gain, V_{out}/V_{in}, is given by $(g_{m1}R_1 + g_{m2}R_2)/2 = g_{m1}R_1 = R_1/R_S$, we have

$$\text{NF} = 1 + \left(\frac{R_S}{R_1}\right)^2(4kTR_1 + 4kT\gamma g_{m2}R_2^2 + 4kTR_2)\frac{1}{4kTR_S} \tag{5.132}$$

$$= 1 + \frac{R_S}{R_1} + \gamma\frac{R_2}{R_1} + \frac{R_SR_2}{R_1^2}. \tag{5.133}$$

The principal advantage of the above noise cancellation technique is that it affords the broadband characteristics of feedback or CG stages but with a lower noise figure. It is therefore suited to systems operating in different frequency bands or across a wide frequency range, e.g., 900 MHz to 5 GHz.

5.3.7 Reactance-Cancelling LNAs

It is possible to devise an LNA topology that inherently cancels the effect of its own input capacitance. Illustrated in Fig. 5.51(a) [7], the idea is to exploit the inductive input impedance of a negative-feedback amplifier so as to cancel the input capacitance, C_{in}. If the open-loop transfer function of the core amplifier is modeled by a one-pole response, $A_0/(1 + s/\omega_0)$, then the input admittance is given by

$$Y_1(s) = \frac{s + (A_0 + 1)\omega_0}{R_F(s + \omega_0)}. \tag{5.134}$$

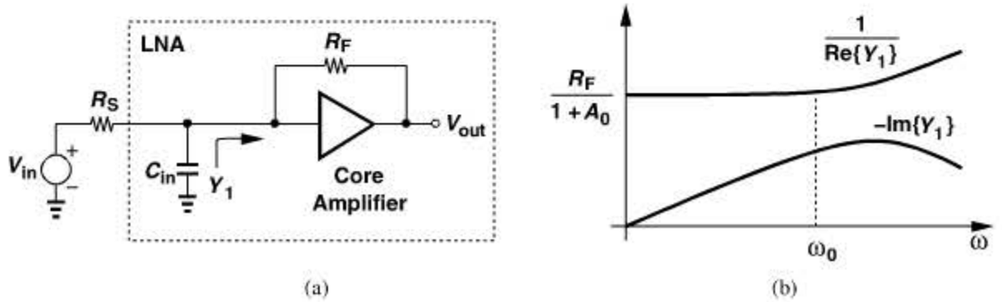


Figure 5.51 *(a) Reactance-cancelling LNA topology, (b) behavior of components of Y_1 with frequency.*

It follows that

$$\frac{1}{Re\{Y_1\}} = \frac{R_F(\omega^2 + \omega_0^2)}{(1 + A_0)\omega_0^2} \tag{5.135}$$

$$Im\{Y_1\} = \frac{-A_0\omega\omega_0}{R_F(\omega^2 + \omega_0^2)}. \tag{5.136}$$

At frequencies well below ω_0, $1/Re\{Y_1\}$ reduces to $R_F/(1 + A_0)$, which can be set equal to R_S, and $Im\{Y_1\}$ is roughly $-A_0\omega/(R_F\omega_0)$, which can be chosen to cancel $C_{in}\omega$. Figure 5.51(b) illustrates the behavior of $1/Re\{Y_1\}$ and $-Im\{Y_1\}$.

The input matching afforded by the above technique holds for frequencies up to about ω_0, dictating that the open-loop bandwidth of the core amplifier reach the maximum frequency of interest. The intrinsic speed of deep-submicron devices provides the gain and bandwidth required here.

The reader may wonder if our modeling of the core amplifier by a one-pole response applies to multistage implementations as well. We return to this point below.

Figure 5.52 shows a circuit realization of the amplifier concept for the frequency range of 50 MHz to 10 GHz [7]. Three common-source stages provide gain and allow negative feedback. Cascodes and source followers are avoided to save voltage headroom. The input transistor, M_1, has a large width commensurate with flicker noise requirements at 50 MHz, thus operating with a V_{GS} of about 200 mV. If this voltage also appears at node Y, it leaves no headroom for output swings, limiting the linearity of the circuit. To resolve this issue, current I_1 is drawn from R_F so as to shift up the quiescent voltage at Y by approximately 250 mV. Since $R_F = 1$ kΩ, I_1 need be only 200 μA, contributing negligible noise at the LNA input.[16]

With three gain stages, the LNA can potentially suffer from a small phase margin and exhibit substantial peaking in its frequency response. In this design, the open-loop poles at nodes A, B, X, and Y lie at 10 GHz, 24.5 GHz, 22 GHz, and 75 GHz, respectively, creating a great deal of phase shift. Nonetheless, due to the small feedback factor,

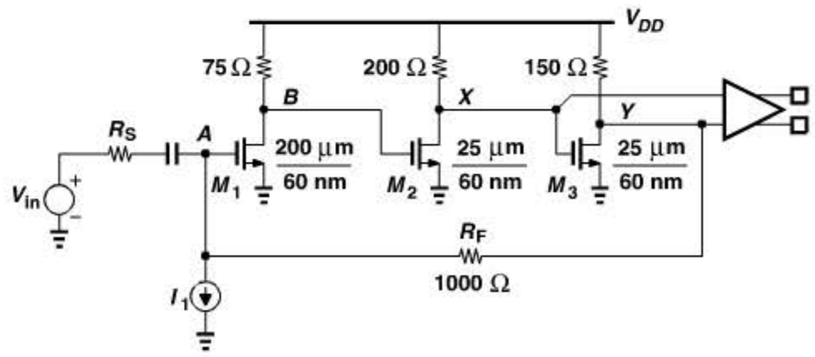


Figure 5.52 *Implementation of reactance-cancelling LNA.*

16. Alternatively, capacitive coupling can be used in the feedback path. But the large value necessary for the capacitor would introduce additional parasitics.

$R_S/(R_S + R_F) = 0.048$, simulations indicate that the circuit provides a phase margin of about 50° and a peaking of 1 dB in its closed-loop frequency response.

The multi-pole LNA of Fig. 5.52 contains an inductive component in its input impedance but with a behavior more complex than the above analysis suggests. Fortunately, behavioral simulations confirm that, if the poles at B, X, and Y are "lumped" (i.e., their time constants are added), then the one-pole approximation still predicts the input admittance accurately. The pole frequencies mentioned above collapse to an equivalent value of $\omega_0 = 2\pi(9.9\ \text{GHz})$, suggesting that the real and imaginary parts of Y_1 retain the desired behavior up to the edge of the cognitive radio band.

The LNA output is sensed between nodes X and Y. Even though these nodes provide somewhat unequal swings and a phase difference slightly greater than 180°, the pseudo-differential sensing still raises both the gain and the IP_2, the latter because second-order distortion at X also appears at Y and is thus partially cancelled in $V_Y - V_X$.[17]

5.4 GAIN SWITCHING

The dynamic range of the signal sensed by a receiver may approach 100 dB. For example, a cell phone may receive a signal level as high as −10 dBm if it is close to a base station or as low as −110 dBm if it is in an underground garage. While designed for the highest sensitivity, the receiver chain must still detect the signal correctly as the input level continues to increase. This requires that the gain of each stage be reduced so that the subsequent stages remain sufficiently linear with the large input signal. Of course, as the gain of the receiver is reduced, its noise figure rises. The gain must therefore be lowered such that the degradation in the sensitivity is less than the increase in the received signal level, i.e., the SNR does not fall. Figure 5.53 shows a typical scenario.

Gain switching in an LNA must deal with several issues: (1) it must negligibly affect the input matching; (2) it must provide sufficiently small "gain steps"; (3) the additional devices performing the gain switching must not degrade the speed of the original LNA;

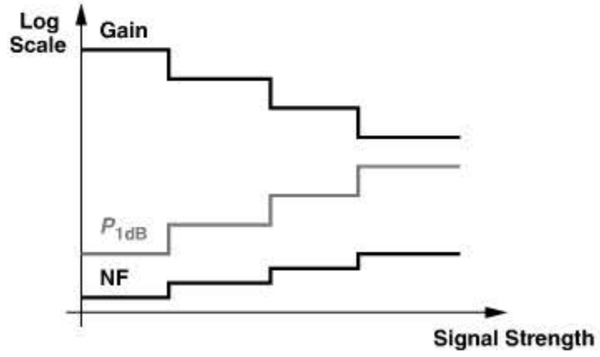


Figure 5.53 *Effect of gain switching on NF and P_{1dB}.*

17. To ensure stability in the presence of package parasitics, a capacitor of 10-20 pF must be placed between V_{DD} and GND.

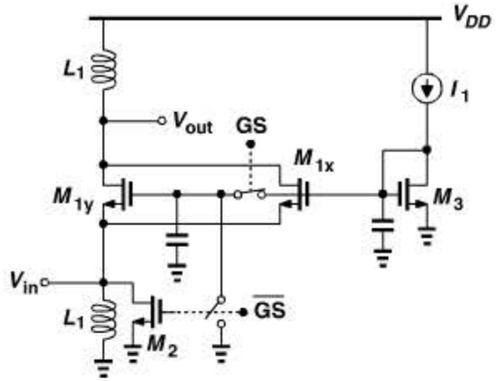


Figure 5.54 *Example of gain switching in CG stage.*

(4) for high input signal levels, gain switching must also make the LNA *more linear* so that this stage does not limit the receiver linearity. As seen below, some LNA topologies lend themselves more easily to gain switching than others do.

Let us first consider a common-gate stage. Can we reduce the transconductance of the input transistor to reduce the gain? To switch the gain while maintaining input matching, we can insert a physical resistance in parallel with the input as g_m is lowered. Figure 5.54 shows an example [8], where the input transistor is decomposed into two, M_{1x} and M_{1y}, and transistor M_2 introduces a parallel resistance if it is on. In the "high-gain mode," the gain select line, GS, is high, placing M_{1x} and M_{1y} in parallel, and M_2 is off. In the "low-gain mode," M_{1y} turns off, reducing the gain, and M_2 turns on, ensuring that $R_{on2}||(g_{m1x} + g_{mb1x})^{-1} = R_S$. For example, to reduce the gain by 6 dB, we choose equal dimensions for M_{1x} and M_{1y} and $R_{on2} = (g_{m1x} + g_{mb1x})^{-1} = 2R_S$ (why?). Also, the gate of M_{1y} is secured to ground by a capacitor to avoid the on-resistance of the switch at high frequencies.

Example 5.21

Choose the devices in the above circuit for a gain step of 3 dB.

Solution:

To reduce the voltage gain by $\sqrt{2}$, we have

$$\frac{W_{1x}}{W_{1x} + W_{1y}} = \frac{1}{\sqrt{2}}, \tag{5.137}$$

and hence $W_{1y}/W_{1x} = \sqrt{2} - 1$. We also note that, with M_{1y} off, the input resistance rises to $\sqrt{2}R_S$. Thus, $R_{on2}||(\sqrt{2}R_S) = R_S$ and hence

$$R_{on2} = \frac{\sqrt{2}}{\sqrt{2} - 1} R_S. \tag{5.138}$$

In Problem 5.21, we calculate the noise figure after the 3-dB gain reduction.

In the above calculation, we have neglected the effect of channel-length modulation. If the upper bound expressed by Eq. (5.67) restricts the design, then the cascode CG stage of Fig. 5.24 can be used.

Another approach to switching the gain of a CG stage is illustrated in Fig. 5.55, where the on-resistance of M_2 appears in parallel with R_1. With input matching and in the absence of channel-length modulation, the gain is given by

$$\frac{V_{out}}{V_{in}} = \frac{R_1||R_{on2}}{2R_S}. \tag{5.139}$$

For multiple gain steps, a number of PMOS switches can be placed in parallel with R_1. The following example elaborates on this point.

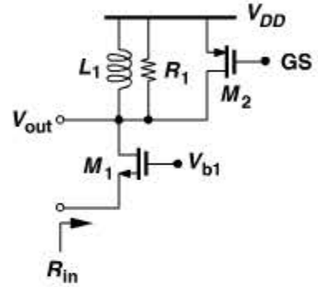


Figure 5.55 *Effect of load switching on input impedance.*

Example 5.22

Design the load switching network of Fig. 5.55 for two 3-dB gain steps.

Solution:

As shown in Fig. 5.56, M_{2a} and M_{2b} switch the gain. For the first 3-dB reduction in gain, M_{2a} is turned on and

$$R_1||R_{on,a} = \frac{R_1}{\sqrt{2}}, \tag{5.140}$$

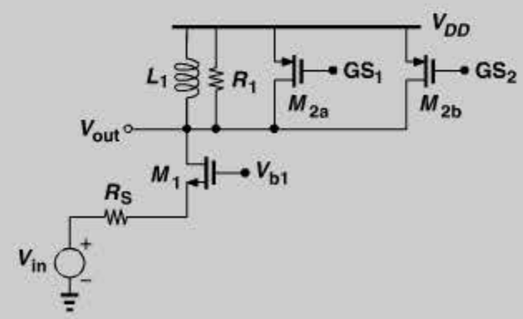


Figure 5.56 *Load switching for 3-dB gain steps.*

(Continues)

Example 5.22 *(Continued)*

i.e., $R_{on,a} = R_1/(\sqrt{2} - 1)$. For the second 3-dB reduction, *both* M_{2a} and M_{2b} are turned on and

$$R_1||R_{on,a}||R_{on,b} = \frac{R_1}{2}, \tag{5.141}$$

i.e., $R_{on,b} = R_1/(2 - \sqrt{2})$. Note that if only M_{2b} were on in this case, then it would need to be wider, thus contributing a greater capacitance to the output node.

The principal difficulty with switching the load resistance in a CG stage is that it alters the input resistance, as expressed by $R_{in} = (R_1 + r_O)/(1 + g_m r_O)$. This effect can be minimized by adding a cascode transistor as in Fig. 5.24. The use of a cascode transistor also permits a third method of gain switching. Illustrated in Fig. 5.57, the idea is to route part of the drain current of the input device to V_{DD}—rather than to the load—by means of another cascode transistor, M_3. For example, if M_2 and M_3 are identical, then turning M_3 on yields $\alpha = 0.5$, dropping the voltage gain by 6 dB.

The advantage of the above technique over the previous two is that the gain step depends only on W_3/W_2 (if M_2 and M_3 have equal lengths) and not the absolute value of the on-resistance of a MOS switch. The bias and signal currents produced by M_1 split between M_3 and M_2 in proportion to W_3/W_2, yielding a gain change by a factor of $1 + W_3/W_2$. As a result, gain steps in the circuit of Fig. 5.57 are more accurate than those in Figs. 5.54 and 5.55. However, the capacitance introduced by M_3 at node Y degrades the performance at high frequencies. For a single gain step of 6 dB, we have $W_3 = W_2$, nearly doubling the capacitance at this node. For a gain reduction by a factor of N, $W_3 = (N - 1)W_2$, possibly degrading the performance considerably.

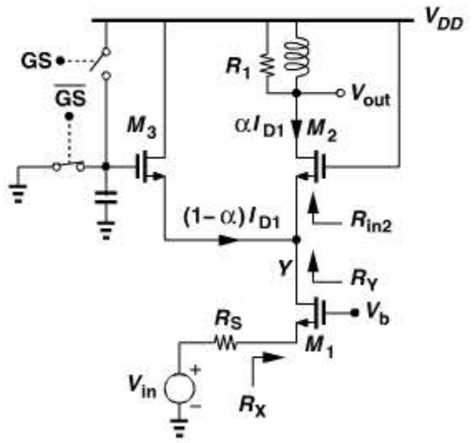


Figure 5.57 *Gain switching by cascode device.*

Example 5.23

If $W_3 = W_2$ in Fig. 5.57, how does the input impedance of the circuit change from the high-gain mode to the low-gain mode? Neglect body effect.

Solution:

In the high-gain mode, the input impedance is given by Eq. (5.70). In the low-gain mode, the impedance seen looking into the source of M_2 changes because both g_{m2} and r_{O2} change. For a square-law device, a twofold reduction in the bias current (while the dimensions remain unchanged) translates to a twofold increase in r_O and a $\sqrt{2}$ reduction in g_m. Thus, in Fig. 5.57,

$$R_{in2} = \frac{R_1 + 2r_{O2}}{1 + \sqrt{2}g_{m2}r_{O2}}, \tag{5.142}$$

where g_{m2} and r_{O2} correspond to the values while M_3 is off. Transistor M_3 presents an impedance of $(1/g_{m3})||r_{O3}$ at Y, yielding

$$R_Y = \frac{1}{g_{m3}}||r_{O3}||\frac{R_1 + 2r_{O2}}{1 + \sqrt{2}g_{m2}r_{O2}}. \tag{5.143}$$

Transistor M_1 transforms this impedance to

$$R_X = \frac{R_Y + r_{O1}}{1 + g_{m1}r_{O1}}. \tag{5.144}$$

This impedance is relatively independent of the gain setting because R_Y is on the order of $1/g_m$.

In order to reduce the capacitance contributed by the gain switching transistor, we can *turn off* part of the main cascode transistor so as to create a greater imbalance between the two. Shown in Fig. 5.58 (on page 310) is an example where M_2 is decomposed into two devices so that, when M_3 is turned on, M_{2a} is turned off. Consequently, the gain drops by a factor of $1 + W_3/W_{2b}$ rather than $1 + W_3/(W_{2b} + W_{2a})$.

Example 5.24

Design the gain switching network of Fig. 5.58 for two 3-dB steps. Assume equal lengths for the cascode devices.

Solution:

To reduce the gain by 3 dB, we turn on M_3 while M_{2a} and M_{2b} remain on. Thus,

$$1 + \frac{W_3}{W_{2a} + W_{2b}} = \sqrt{2}. \tag{5.145}$$

(Continues)

Example 5.24 *(Continued)*

For another 3-dB reduction, we turn off M_{2b}:

$$1 + \frac{W_3}{W_{2a}} = 2. \tag{5.146}$$

It follows from Eqs. (5.145) and (5.146) that

$$W_3 = W_{2a} = \frac{W_{2b}}{\sqrt{2}}. \tag{5.147}$$

In a more aggressive design, M_2 would be decomposed into *three* devices, such that one is turned off for the first 3-dB step, allowing M_3 to be narrower. The calculations are left as an exercise for the reader.

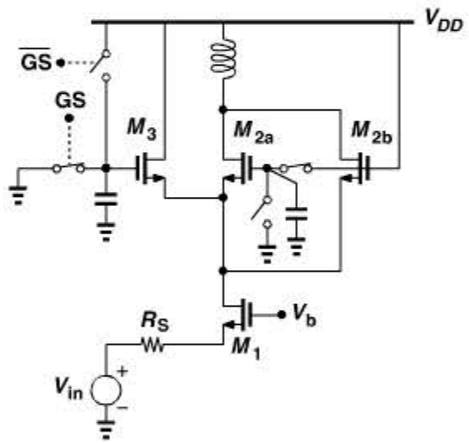


Figure 5.58 *Gain switching by programmable cascode devices.*

We now turn our attention to gain switching in an inductively-degenerated cascode LNA. Can we switch part of the input transistor to switch the gain (Fig. 5.59)? Turning M_{1b} off does not alter ω_T because the current density remains constant. Thus, $Re\{Z_{in}\} = L_1\omega_T$ is relatively constant, but $Im\{Z_{in}\}$ changes, degrading the input match. If the input match is somehow restored, then the voltage gain, $R_1/(2L_1\omega)$, does not change! Furthermore, the thermal noise of S_1 degrades the noise figure in the high gain mode. For these reasons, gain switching must be realized in other parts of the circuit.

As with the CG LNA of Fig. 5.55, the gain can be reduced by placing one or more PMOS switches in parallel with the load [Fig. 5.60(a)]. Alternatively, the cascode switching scheme of Fig. 5.57 can be applied here as well [Fig. 5.60(b)]. The latter follows the calculations outlined in Example 5.24, providing well-defined gain steps with a moderate

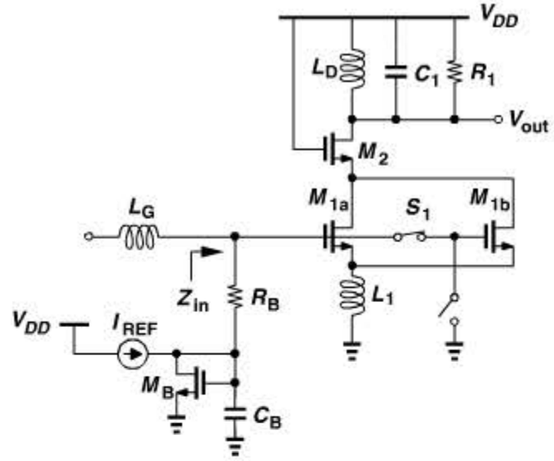


Figure 5.59 *Gain switching in CS stage.*

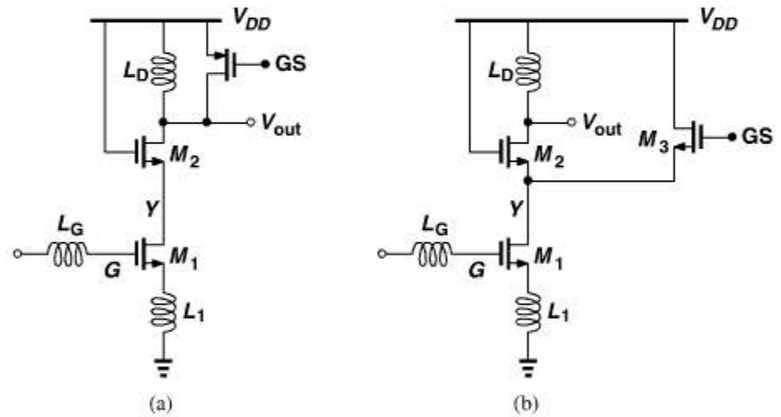


Figure 5.60 *Gain switching in cascode CS stage by (a) load switching, (b) additional cascode device.*

additional capacitance at node Y. It is important to bear in mind that cascode switching is attractive because it reduces the current flowing through the load by a well-defined ratio *and* it negligibly alters the input impedance of the LNA.

For the two variants of the CG stage studied in Section 5.3.3, gain switching can be realized by cascode devices as illustrated in Fig. 5.57. The use of feedback or feedforward in these topologies makes it difficult to change the gain through the input transistor without affecting the input match.

Lastly, let us consider gain switching in the noise-cancelling LNA of Fig. 5.48(b). Since $V_Y/V_X = 1 - R_F/R_S$ and R_{in} is approximately equal to $1/g_{m1}$ and independent of R_F, the gain can be reduced simply by lowering the value of R_F. Though not essential in the low-gain mode, noise cancellation can be preserved by adjusting A_1 so that it remains equal to $1 + R_F/R_S$.

Which one of the foregoing gain reduction techniques also makes the LNA *more linear*? None, except for the last one! Since the CG and CS stages retain the gate-source voltage swing (equal to half of the input voltage swing), their linearity improves negligibly. In the feedback LNA of Fig. 5.48(b), on the other hand, a lower R_F strengthens the negative feedback, raising the linearity to some extent.

Receiver designs in which the LNA nonlinearity becomes problematic at high input levels can "bypass" the LNA in very-low-gain modes. Illustrated conceptually in Fig. 5.61, the idea is to omit the LNA from the signal path so that the mixer (presumably more linear) directly senses the received signal. The implementation is not straightforward if input matching must be maintained. Figure 5.62 depicts a common-gate example, where M_1 is turned off, M_2 is turned on to produce a 50-Ω resistance, and M_3 is turned on to route the signal to the mixer.

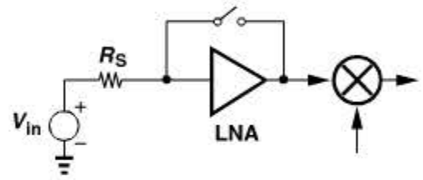


Figure 5.61 *LNA bypass.*

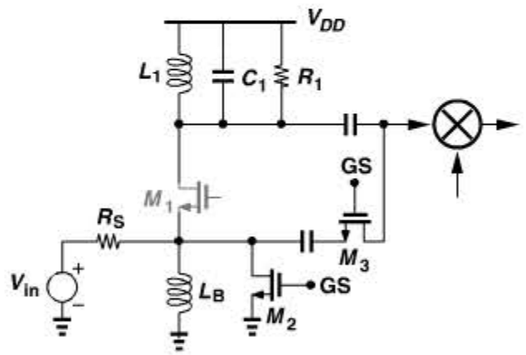


Figure 5.62 *Realization of LNA bypass.*

5.5 BAND SWITCHING

As mentioned in Section 5.1, LNAs that must operate across a wide bandwidth or in different bands can incorporate band switching. Figure 5.63(a) repeats the structure of Fig. 5.7(a), with the switch realized by a MOS transistor. Since the bias voltage at the output node is near V_{DD}, the switch must be a PMOS device, thus contributing a larger capacitance for a given on-resistance than an NMOS transistor. This capacitance lowers the tank resonance frequency when S_1 is *off*, reducing the maximum tolerable value of C_1 and hence limiting the size of the input transistor of the following stage. (If L_1 is reduced to compensate for the higher capacitance, then so are R_1 and the gain.) For this reason, we prefer the implementation in Fig. 5.63(b), where S_1 is formed as an NMOS device tied to ground.

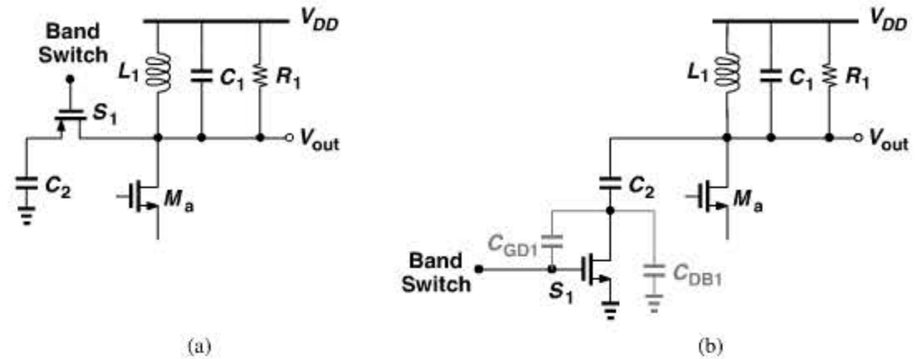


Figure 5.63 *(a) Band switching, (b) effect of switch parasitics.*

The choice of the width of S_1 in Fig. 5.63(b) proves critical. For a very narrow transistor, the on-resistance, R_{on1}, remains so high that the tank does not "feel" the presence of C_2 when S_1 is on. For a moderate device width, R_{on1} limits the Q of C_2, thereby lowering the Q of the overall tank and hence the voltage gain of the LNA. This can be readily seen by transforming the series combination of C_2 and R_{on1} to a parallel network consisting of C_2 and $R_{P1} \approx Q^2 R_{on1}$, where $Q = (C_2 \omega R_{on1})^{-1}$. That is, R_1 is now shunted by a resistance $R_{P1} = (C_2^2 \omega^2 R_{on1})^{-1}$.

The foregoing observation implies that R_{on1} must be minimized such that $R_{P1} \gg R_1$. However, as the width of S_1 in Fig. 5.63(b) increases, so does the capacitance that it introduces in the *off state*. The equivalent capacitance seen by the tank when S_1 is off is equal to the series combination of C_2 and $C_{GD1} + C_{DB1}$, which means C_1 must be less than its original value by this amount. We therefore conclude that the width of S_1 poses a trade-off between the tolerable value of C_1 when S_1 is off and the reduction of the gain when S_1 is on. (Recall that C_1 arises from M_a, the input capacitance of the next stage, and the parasitic capacitance of L_1.)

An alternative method of band switching incorporates two or more tanks as shown in Fig. 5.64 [8]. To select one band, the corresponding cascode transistor is turned on while the other remains off. This scheme requires that each tank drive a copy of the following stage, e.g., a mixer. Thus, when M_1 and band 1 are activated, so is mixer MX_1. The principal drawback of this approach is the capacitance contributed by the additional cascode device(s) to node Y. Also, the spiral inductors have large footprints, making the layout and routing more difficult.

5.6 HIGH-IP$_2$ LNAS

As explained in Chapter 4, even-order distortion can significantly degrade the performance of direct-conversion receivers. Since the circuits following the downconversion mixers are typically realized in differential form,[18] they exhibit a high IP$_2$, leaving the LNA and the

18. And since they employ large devices and hence have small mismatches.

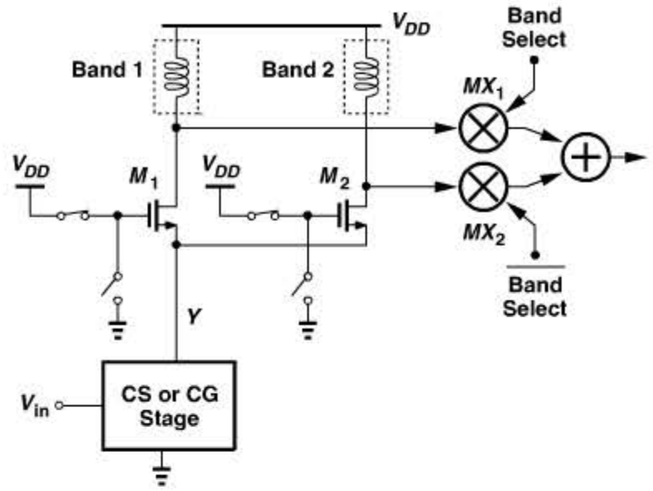


Figure 5.64 *Band switching by programmable cascode branches.*

mixers as the IP_2 bottleneck of the receivers. In this section, we study techniques of raising the IP_2 of LNAs, and in Chapter 6, we do the same for mixers.

5.6.1 Differential LNAs

Differential LNAs can achieve high IP_2's because, as explained in Chapter 2, symmetric circuits produce no even-order distortion. Of course, some (random) asymmetry plagues actual circuits, resulting in a finite, but still high, IP_2.

In principle, any of the single-ended LNAs studied thus far can be converted to differential form. Figure 5.65 depicts two examples. Not shown here, the bias network for the input transistors is similar to those described in Sections 5.3.3 and 5.3.4.

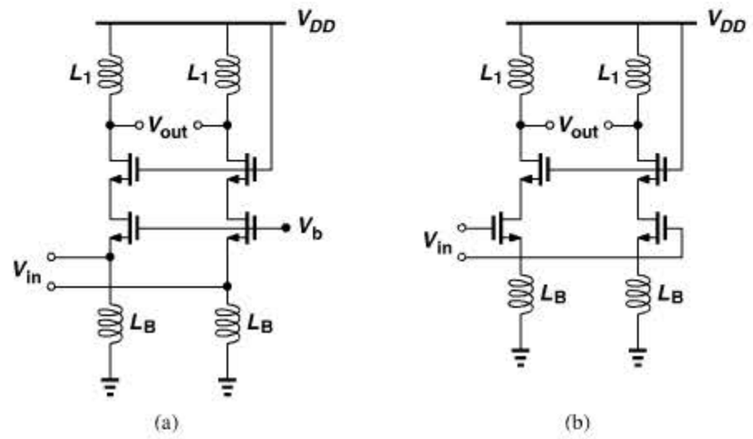


Figure 5.65 *Differential (a) CG and (b) CS stages.*

But what happens to the noise figure of the circuit if it is converted to differential form? Before answering this question, we must determine the source impedance driving the LNA. Since the antenna and the preselect filter are typically single-ended, a transformer must precede the LNA to perform single-ended to differential conversion. Illustrated in Fig. 5.66(a), such a cascade processes the signal differentially from the input port of the LNA to the end of the baseband section. The transformer is called a "balun," an acronym for "balanced-to-unbalanced" conversion because it can also perform differential to single-ended conversion if its two ports are swapped.

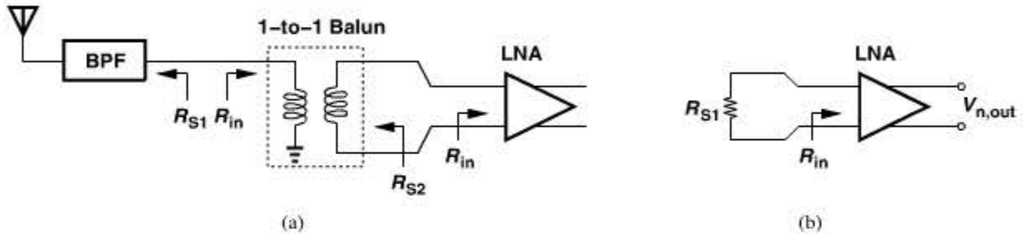


Figure 5.66 *(a) Use of balun at RX input, (b) simplified circuit.*

If the source impedance provided by the antenna and the band-pass filter in Fig. 5.66(a) is R_{S1} (e.g., 50 Ω), what is the differential source impedance seen by the LNA, R_{S2}? For a lossless 1-to-1 balun, i.e., for a lossless transformer with an equal number of turns in its primary and secondary, we have $R_{S2} = R_{S1}$. We must thus obtain the noise figure of the differential LNA with respect to a differential source impedance of R_{S1}. Figure 5.66(b) shows the setup for output noise calculation.

Note that the differential input impedance of the LNA, R_{in}, must be equal to R_{S1} for proper input matching. Thus, in the LNAs of Figs. 5.66(a) and (b), the *single-ended* input impedance of each half circuit must be equal to $R_{S1}/2$, e.g., 25 Ω.

Differential CG LNA We now calculate the noise figure of the differential CG LNA of Fig. 5.65(a), assuming it is designed such that the impedance seen between each input node and ground is equal to $R_{S1}/2$. In other words, each CG transistor must provide an input resistance of 25 Ω. Figure 5.67(a) shows the simplified environment, emphasizing that the noise figure is calculated with respect to a source impedance of R_{S1}. Redrawing Fig. 5.67(a) as shown in Fig. 5.67(b), we recognize from the symmetry of the circuit that

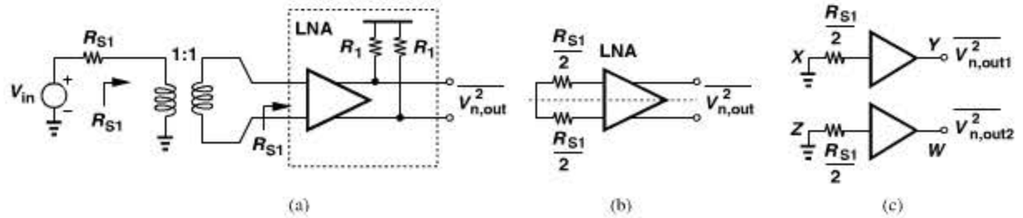


Figure 5.67 *(a) Cascade of balun and LNA, (b) simplified circuit of (a), and (c) simplified circuit of (b).*

we can compute the output noise of each half circuit as in Fig. 5.67(c) and add the output powers:

$$\overline{V_{n,out}^2} = \overline{V_{n,out1}^2} + \overline{V_{n,out2}^2}. \tag{5.148}$$

Since each half circuit provides matching at the input, the CG results of Section 5.3.3 apply here as well with the substitution $R_S = R_{S1}/2$. Specifically, the voltage gain from X to Y is equal to $R_1/(2R_{S1}/2)$, where R_1 denotes the load resistance of the CG half circuit. The output noise consists of (1) the input transistor contribution, given by Eq. (5.56), (2) the load resistor contribution, $4kTR_1$, and (3) the source impedance contribution, $(4kTR_{S1}/2)[R_1/(2R_1/2)]$:

$$\overline{V_{n,out1}^2} = kT\gamma\frac{R_1^2}{R_{S1}/2} + 4kTR_1 + 4kT\frac{R_{S1}}{2}\left(\frac{R_1}{\frac{2R_{S1}}{2}}\right)^2. \tag{5.149}$$

From Eq. (5.148), the total output noise power is twice this amount. Noting that the total voltage gain $A_v = (V_Y - V_W)/(V_X - V_Z)$ is equal to that of half of the circuit, V_Y/V_X $(= R_1/R_{S1})$, we compute the noise figure with respect to a source impedance of R_{S1} as

$$\text{NF} = \frac{\overline{V_{n,out}^2}}{A_v^2}\cdot\frac{1}{4kTR_{S1}} \tag{5.150}$$

$$= 1 + \gamma + \frac{2R_{S1}}{R_1}. \tag{5.151}$$

Interestingly, this value is lower than that of the single-ended counterpart [Eq. (5.58)]. But why? Since in Fig. 5.67(c), $V_Y/V_X = R_1/(2R_{S1}/2) = R_1/R_{S1}$, we observe that the voltage gain is twice that of the single-ended CG LNA. (After all, the transconductance of the input transistor is doubled to lower the input impedance to $R_{S1}/2$.) On the other hand, the overall differential circuit contains *two* R_1's at its output, each contributing a noise power of $4kTR_1$. The total, $8kTR_1$, divided by $(R_1/R_{S1})^2$ and $4kTR_{S1}$ yields $2R_{S1}/R_1$. Of course, the value stipulated by Eq. (5.151) can be readily obtained in a single-ended CG LNA by simply doubling the load resistance. Figure 5.68 summarizes the behavior of the two circuits, highlighting the greater voltage gain in the differential topology. If identical gains are desired, the value of the load resistors in the differential circuit must be halved, thereby yielding identical noise figures.

In summary, a single-ended CG LNA can be converted to differential form according to one of three scenarios: (1) simply copy the circuit, in which case the differential input resistance reaches 100 Ω, failing to provide matching with a 1-to-1 balun; (2) copy the circuit but double the transconductance of the input transistors, in which case the input is matched but the overall voltage gain is doubled; (3) follow the second scenario but halve the load resistance to retain the same voltage gain. The second choice is generally preferable. Note that, for a given noise figure, a differential CG LNA consumes *four* times the power of a single-ended stage.[19]

19. To halve the input resistance, the transistor width and bias current must be doubled.

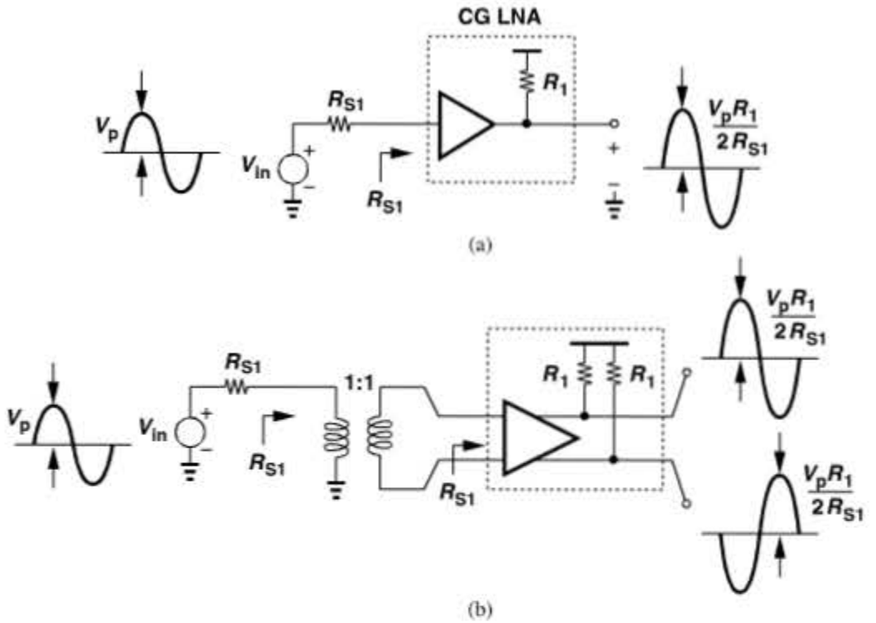


Figure 5.68 *Comparison of (a) single-ended and (b) differential CG LNAs.*

Our NF calculations have assumed an ideal balun. In reality, even external baluns have a loss as high as 0.5 dB, raising the NF by the same amount.

Example 5.25

An amplifier having a high input impedance employs a parallel resistor at the input to provide matching [Fig. 5.69(a)]. Determine the noise figure of the circuit and its differential version, shown in Fig. 5.69(b), where two replicas of the amplifier are used.

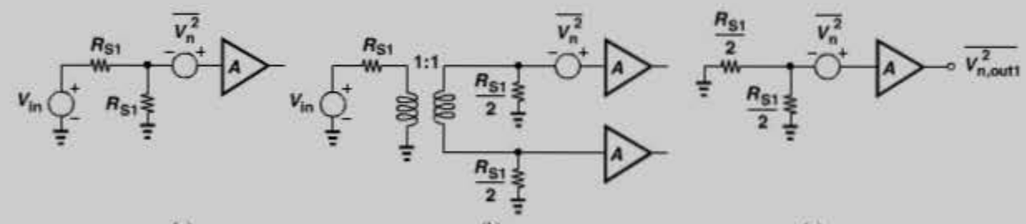


Figure 5.69 *(a) NF of an LNA with resistive termination, (b) differential version of (a), (c) simplified circuit of (b).*

Solution:

In the circuit of Fig. 5.69(a), the amplifier input-referred noise current is negligible and the total noise at the output is equal to $(4kTR_{S1}/2)A^2 + A^2\overline{V_n^2}$. The noise figure of the

(Continues)

Example 5.25 *(Continued)*

single-ended circuit is therefore given by

$$\mathrm{NF}_{sing} = \frac{4kT\frac{R_{S1}}{2}A^2 + A^2\overline{V_n^2}}{\frac{A^2}{4}} \cdot \frac{1}{4kTR_{S1}} \tag{5.152}$$

$$= 2 + \frac{\overline{V_n^2}}{kTR_{S1}}. \tag{5.153}$$

For the differential version, we write from the simplified half circuit shown in Fig. 5.69(c), $\overline{V_{n,out1}^2} = (4kTR_{S1}/4)A^2 + A^2\overline{V_n^2}$. The total output noise power of the differential circuit is twice this amount. The corresponding noise figure is then given by

$$\mathrm{NF}_{diff} = \frac{2\left(4kT\frac{R_{S1}}{4}A^2 + A^2\overline{V_n^2}\right)}{\frac{A^2}{4}} \cdot \frac{1}{4kTR_{S1}} \tag{5.154}$$

$$= 2 + \frac{2\overline{V_n^2}}{kTR_{S1}}. \tag{5.155}$$

In this case, the noise figure of the differential circuit is *higher*. We conclude that whether the differential version of an LNA exhibits a higher or lower NF depends on the circuit topology.

Differential CS LNA The differential CS LNA of Fig. 5.65(b) behaves differently from its CG counterpart. From Section 5.3.4, we recall that the input resistance of each half circuit is equal to $L_1\omega_T$ and must now be halved. This is accomplished by halving L_1. With input matching and a degeneration inductance of L_1, the voltage gain was found in Section 5.3.4 to be $R_1/(2L_1\omega_0)$, which is now doubled. Figure 5.70(a) illustrates the overall cascade of the balun and the differential LNA. We assume that the width and bias current of each input transistor are the same as those of the single-ended LNA.

To compute the noise figure, let us first determine the output noise of the half circuit depicted in Fig. 5.70(b). Neglecting the contribution of the cascode device, we note from Section 5.3.4 that, if the input is matched, half of the noise current of the input transistor flows from the output node. Thus,

$$\overline{V_{n,out1}^2} = kT\gamma g_{m1}R_1^2 + 4kTR_1 + 4kT\frac{R_{S1}}{2}\left(\frac{R_1}{L_1\omega_0}\right)^2. \tag{5.156}$$

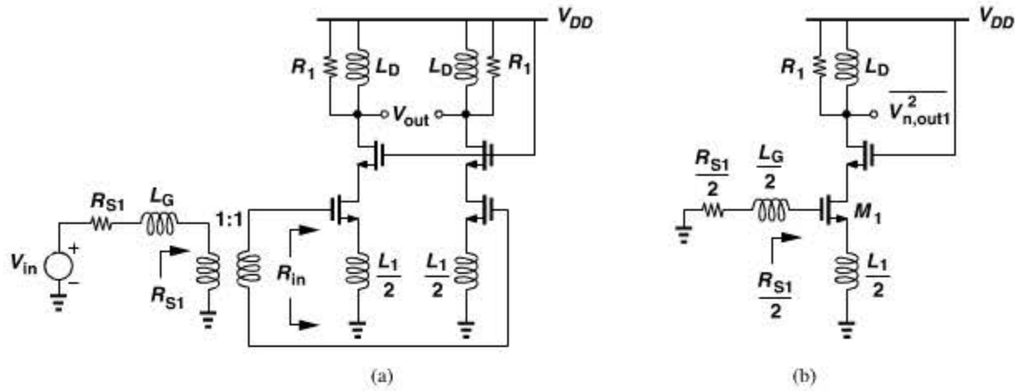


Figure 5.70 *(a) Differential CS LNA and (b) its half circuit.*

Multiplying this power by two, dividing it by $A_v^2 = R_1^2/(L_1\omega_0)^2$ and $4kTR_{S1}$, and noting that $L_1\omega_T/2 = R_{S1}/2$, we have

$$\mathrm{NF} = \frac{\gamma}{2}g_{m1}R_{S1}\left(\frac{\omega_0}{\omega_T}\right)^2 + \frac{2R_{S1}}{R_1}\left(\frac{\omega_0}{\omega_T}\right)^2 + 1. \tag{5.157}$$

How does this compare with the noise figure of the original single-ended LNA [Eq. (5.101)]? We observe that both the transistor contribution and the load contribution are halved. The transistor contribution is halved because g_{m1} and hence the transistor noise *current* remain unchanged while the overall *transconductance* of the circuit is doubled. To understand this point, recall from Section 5.3.4 that $G_m = \omega_T/(2\omega_0 R_S)$ for the original single-ended circuit. Now consider the equivalent circuit shown in Fig. 5.71, where the differential transconductance, $(I_1 - I_2)/V_{in}$, is equal to $\omega_T/(\omega_0 R_{S1})$ (why?). The differential output current contains the noise currents of both M_1 and M_2 and is equal to $2(kT\gamma g_{m1})$. If this power is divided by the square of the transconductance and $4kTR_{S1}$, the first term in Eq. (5.157) is obtained.

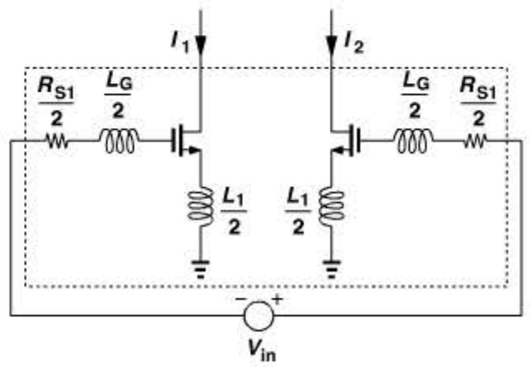


Figure 5.71 *Differential CS stage viewed as a transconductor.*

The reduction of the input transistor noise contribution in Eq. (5.157) is a remarkable property of differential operation, reinforcing the NF advantage of the degenerated CS stage over the CG LNA. However, this result holds only if the design can employ *two* degeneration inductors, each having *half* the value of that in the single-ended counterpart. This is difficult with bond wires as their physical length cannot be shortened arbitrarily. Alternatively, the design can incorporate on-chip degeneration inductors while converting the effect of the (inevitable) bond wire to a common-mode inductance. Figure 5.72 shows such a topology. With perfect symmetry, the bond wire inductance has no effect on the differential impedance seen between the gates. Nonetheless, as explained in Chapter 7, on-chip inductors suffer from a low quality factor (e.g., a high series resistance), possibly degrading the noise figure. We compare the power consumptions of the single-ended and differential implementations in Problem 5.22.

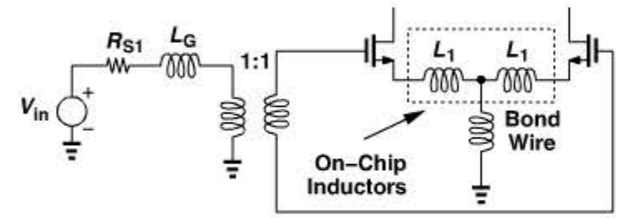


Figure 5.72 *Differential CS stage with on-chip degeneration inductors.*

The NF advantage implied by Eq. (5.157) may not materialize in reality because the loss of the balun is not negligible.

Is it possible to use a differential pair to convert the single-ended antenna signal to differential form? As shown in Fig. 5.73(a), the signal is applied to one input while the other is tied to a bias voltage. At low to moderate frequencies, V_X and V_Y are differential and the voltage gain is equal to $g_{m1,2}R_D$. At high frequencies, however, two effects degrade the balance of the phases: the parasitic capacitance at node P attenuates and delays the signal propagating from M_1 to M_2, and the gate-drain capacitance of M_1 provides a non-inverting feedforward path around M_1 (whereas M_2 does not contain such a path).

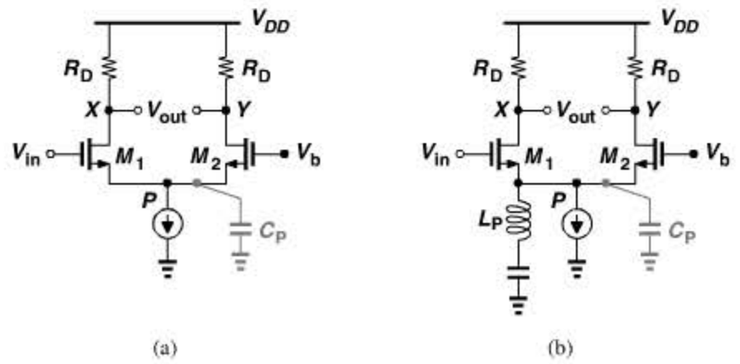


Figure 5.73 *Single-ended to differential conversion by (a) a simple differential pair, (b) a differential pair including tail resonance.*

The capacitance at P can be nulled through the use of a parallel inductor [Fig. 5.73(b)] [9], but the C_{GD1} feedforward persists. The tail inductor can be realized on-chip because its parallel equivalent resistance at resonance ($R_P = QL_P\omega_0$) is typically much greater than $1/g_{m1,2}$.

Example 5.26

A student computes C_P in Fig. 5.73(b) as $C_{SB1} + C_{SB2} + C_{GS2}$, and selects the value of L_P accordingly. Is this an appropriate choice?

Solution:

No, it is not. For L_P to null the phase shift at P, it must resonate with only $C_{SB1} + C_{SB2}$. This point can be seen by examining the voltage division at node P. As shown in Fig. 5.74, in the absence of $C_{SB1} + C_{SB2}$,

$$V_P = V_{in}\frac{Z_2}{Z_1 + Z_2}. \tag{5.158}$$

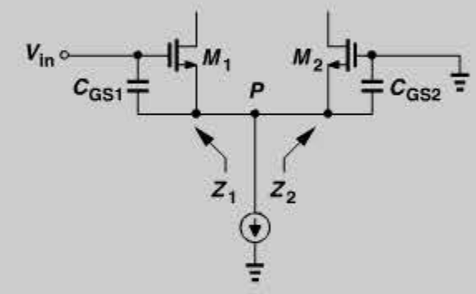


Figure 5.74 *Impedances seen at the common source of differential pair.*

For V_P to be exactly equal to half of V_{in} (with zero phase difference), we must have $Z_1 = Z_2$. Since each impedance is equal to $(g_m + g_{mb})^{-1}||(C_{GSS})^{-1}$, we conclude that C_{GS2} must *not* be nulled.[20]

The topology of Fig. 5.73(b) still does not provide input matching. We must therefore insert (on-chip) inductances in series with the sources of M_1 and M_2 (Fig. 5.75). Here, L_{P1} and L_{P2} resonate with C_{P1} and C_{P2}, respectively, and $L_{S1} + L_{S2}$ provides the necessary input resistance. Of course, $L_{S1} + L_{S2}$ is realized as one inductor. However, as explained in Section 5.7, this topology exhibits a lower IP_3 than that of Fig. 5.65(b).

Balun Issues The foregoing development of differential LNAs has assumed ideal 1-to-1 baluns. Indeed, external baluns with a low loss (e.g., 0.5 dB) in the gigahertz range are available from manufacturers, but they consume board space and raise the cost. Integrated baluns, on the other hand, suffer from a relatively high loss and large capacitances.

20. But the parasitic capacitance of I_{SS} must be nulled.

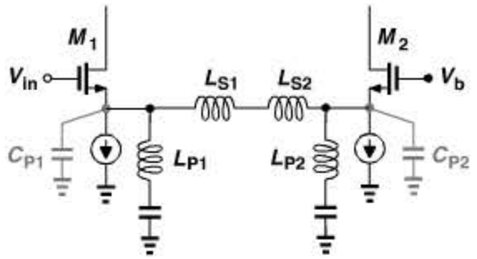


Figure 5.75 *Use of on-chip inductors for resonance and degeneration.*

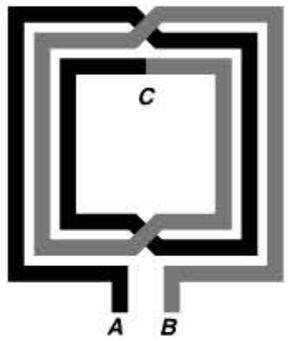


Figure 5.76 *Simple planar 1-to-1 balun.*

Shown in Fig. 5.76 is an example, where two spiral inductors L_{AC} and L_{CB} are intertwined to create a high mutual coupling. As explained in Chapter 7, the resistance and capacitance associated with the spirals and the sub-unity coupling factor make such baluns less attractive.

Example 5.27

A student attempts to use a 1-to-N balun with a differential CS stage so as to amplify the input voltage by a factor of N and potentially achieve a lower noise figure. Compute the noise figure in this case.

Solution:

Illustrated in Fig. 5.77, such an arrangement transforms the source impedance to a value of N^2R_S, requiring that each half circuit provide an input real part equal to $N^2R_S/2$. Thus, $L_1\omega_T = N^2R_S/2$, i.e., each degeneration inductance must be reduced by a factor of N^2. Since still half of the noise current of each input transistor flows to the output node, the noise power measured at each output is given by

$$\overline{V_{n,out1}^2} = \overline{V_{n,out2}^2} = 4kT\gamma g_{m1}\frac{R_1^2}{4} + 4kTR_1. \tag{5.159}$$

Example 5.27 *(Continued)*

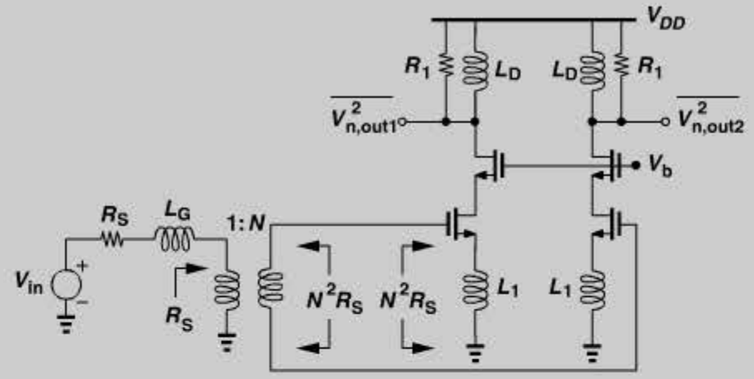


Figure 5.77 *Use of 1-to-N balun in an LNA.*

The gain from V_{in} to the differential output is now equal to $NR_1/(2L_1\omega_0)$. Doubling the above power, dividing by the square of the gain, and normalizing to $4kTR_S$, we have

$$\text{NF} = N^2\frac{\gamma}{2}g_{m1}R_S\left(\frac{\omega_0}{\omega_T}\right)^2 + 2N^2\frac{R_S}{R_1}\left(\frac{\omega_0}{\omega_T}\right)^2 + 1. \tag{5.160}$$

We note, with great distress, that the first two terms have *risen* by a factor of N^2![21] This is because the condition $L_1\omega_T = N^2R_S/2$ inevitably leads to an N^2-fold reduction in the transconductance of the circuit. Thus, even with the N-fold amplification of V_{in} by the balun, the overall voltage gain drops by a factor of N.

The reader may wonder if an N-to-1 (rather than 1-to-N) balun proves beneficial in the above example as it would multiply the first two terms of Eq. (5.160) by $1/N^2$ rather than N^2. Indeed, off-chip baluns may provide a lower noise figure if L_1 (a bond wire) can be reduced by a factor of N^2. On the other hand, on-chip baluns with a non-unity turns ratio are difficult to design and suffer from a higher loss and a lower coupling factor. Figure 5.78(a) shows an example [5], where one spiral forms the primary (secondary) of the balun and the series combination of two spirals constitutes the secondary (primary). Alternatively, as shown in Fig. 5.78(b), spirals having different numbers of turns can be embedded [10].

5.6.2 Other Methods of IP_2 Improvement

The difficulty with the use of off-chip or on-chip baluns at the input of differential LNAs makes single-ended topologies still an attractive choice. A possible approach to raising the IP_2 entails simply filtering the low-frequency second-order intermodulation product, called the beat component in Chapter 4. Illustrated in Fig. 5.79, the idea is to remove the beat by a simple high-pass filter (HPF) following the LNA. For example, suppose two interferers

21. Assuming that g_{m1} and ω_T remain unchanged.

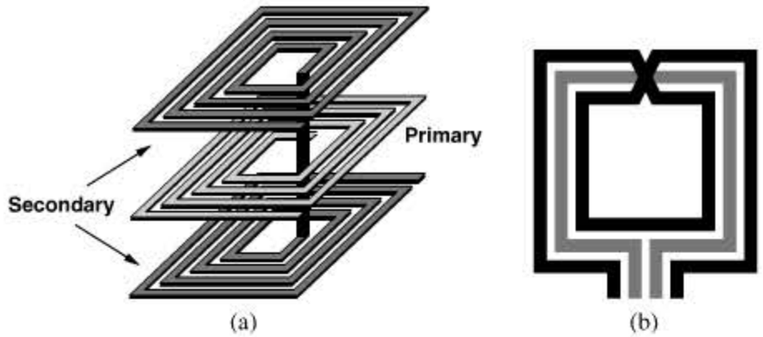


Figure 5.78 *Realization of 1-to-2 balun as (a) stacked spirals, (b) embedded spirals.*

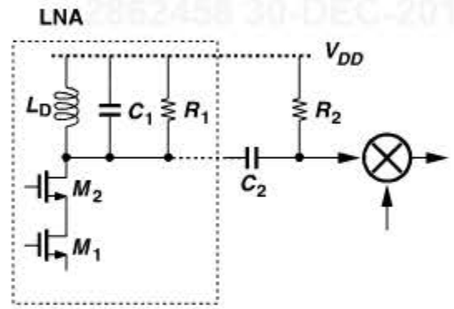


Figure 5.79 *Removal of low-frequency beat by first-order high-pass filter.*

are located at the edges of the 2.4-GHz band, $f_1 = 2.4\,\text{GHz}$ and $f_2 = 2.480\,\text{GHz}$. The beat therefore lies at 80 MHz and is attenuated by approximately a factor of $2400/80 = 30$ for a first-order HPF. With this substantial suppression, the IP_2 of the LNA is unlikely to limit the RX performance, calling for techniques that improve the IP_2 of *mixers* (Chapter 6).

Example 5.28

A student considers the above calculation pessimistic, reasoning that an 80-MHz beat leaking to the baseband of an 11b/g or Bluetooth receiver does not fall within the desired channel. Is the student correct?

Solution:

Yes, the student is correct. For a direct-conversion 11b/g receiver, the baseband signal spans $-10\,\text{MHz}$ to $+10\,\text{MHz}$. Thus, the worst-case beat occurs at 10 MHz, e.g., between two interferers at 2.400 GHz and 2.410 GHz. Such a beat is attenuated by a factor of $2400/10 = 240$ by the first-order HPF.

The filtration of the IM_2 product becomes less effective for wider communication bands. For example, if a receiver must accommodate frequencies from 1 GHz to 10 GHz, then two interferers can produce a beat *within* the band, prohibiting the use of filters to remove the beat. In this case, the LNA may become the receiver's IP_2 bottleneck.

5.7 NONLINEARITY CALCULATIONS

The general behavior of nonlinear systems was formulated in Chapter 2. In this section, we develop a methodology for computing the nonlinear characteristics of some circuits.

Recall from Chapter 2 that systems with weak static nonlinearity can be approximated by a polynomial such as $y = \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3$. Let us devise a method for computing α_1-α_3 for a given circuit. In many circuits, it is difficult to derive y as an explicit function of x. However, we recognize that

$$\alpha_1 = \frac{\partial y}{\partial x}|_{x=0} \tag{5.161}$$

$$\alpha_2 = \frac{1}{2}\frac{\partial^2 y}{\partial x^2}|_{x=0} \tag{5.162}$$

$$\alpha_3 = \frac{1}{6}\frac{\partial^3 y}{\partial x^3}|_{x=0}. \tag{5.163}$$

These expressions prove useful because we can obtain the derivatives by implicit differentiation. It is important to note that in most cases, $x = 0$ in fact corresponds to the *bias* point of the circuit with no input perturbation. In other words, the total y may not be zero for $x = 0$. For example, in the common-source stage of Fig. 5.80, M_1 is biased at a gate-source voltage of $V_{GS0} = V_b$ and V_{in} is superimposed on this voltage.

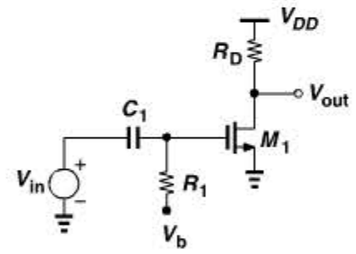


Figure 5.80 *CS stage with gate bias.*

5.7.1 Degenerated CS Stage

As an example, let us study the resistively-degenerated common-source stage shown in Fig. 5.81,

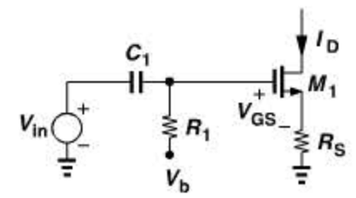


Figure 5.81 *CS stage for nonlinearity calculations.*

assuming the drain current is the output of interest. We wish to compute the IP_3 of the circuit. For a simple square-law device

$$I_D = K(V_{GS} - V_{TH})^2, \tag{5.164}$$

where $K = (1/2)\mu_n C_{ox}(W/L)$ and channel-length modulation and body effect are neglected. Since $V_{GS} = V_{in} - R_S I_D$,

$$I_D = K(V_{in} - R_S I_D - V_{TH})^2, \tag{5.165}$$

and hence

$$\frac{\partial I_D}{\partial V_{in}} = 2K(V_{in} - R_S I_D - V_{TH})\left(1 - R_S \frac{\partial I_D}{\partial V_{in}}\right). \tag{5.166}$$

We also note that

$$g_m = \frac{\partial I_D}{\partial V_{GS}} = 2K(V_{GS} - V_{TH}) \tag{5.167}$$

$$= 2K(V_{in0} - R_S I_{D0} - V_{TH}), \tag{5.168}$$

where V_{in0} ($= V_b$) and I_{D0} denote the bias values. Thus, in the absence of signals,

$$\frac{\partial I_D}{\partial V_{in}}|_{Vin0} = \alpha_1 = \frac{g_m}{1 + g_m R_S}, \tag{5.169}$$

an expected result.

We now compute the second derivative from Eq. (5.166):

$$\frac{\partial^2 I_D}{\partial V_{in}^2} = 2K\left(1 - R_S \frac{\partial I_D}{\partial V_{in}}\right)^2 + 2K(V_{in} - R_S I_D - V_{TH})\left(-R_S \frac{\partial^2 I_D}{\partial V_{in}^2}\right). \tag{5.170}$$

With no signals, (5.168) and (5.169) can be substituted in (5.170) to produce

$$\frac{\partial^2 I_D}{\partial V_{in}^2}|_{Vin0} = 2\alpha_2 = \frac{2K}{(1 + g_m R_S)^3}. \tag{5.171}$$

Lastly, we determine the third derivative from (5.170):

$$\frac{\partial^3 I_D}{\partial V_{in}^3} = 4K\left(1 - R_S \frac{\partial I_D}{\partial V_{in}}\right)\left(-R_S \frac{\partial^2 I_D}{\partial V_{in}^2}\right) + 2K\left(1 - R_S \frac{\partial I_D}{\partial V_{in}}\right)\left(-R_S \frac{\partial^2 I_D}{\partial V_{in}^2}\right)$$
$$- 2K(V_{in} - R_S I_D - V_{TH}) R_S \frac{\partial^3 I_D}{\partial V_{in}^3}, \tag{5.172}$$

which, from (5.169) and (5.171) reduces to

$$\frac{\partial^3 I_D}{\partial V_{in}^3}|_{Vin0} = 6\alpha_3 = \frac{-12K^2 R_S}{(1 + g_m R_S)^5}. \tag{5.173}$$

While lengthy, the foregoing calculations lead to interesting results. Equation (5.173) reveals that $\alpha_3 = 0$ if $R_S = 0$, an expected outcome owing to the square-law behavior assumed for the transistor. Additionally, α_1 and α_3 have *opposite* signs, implying a compressive characteristic—whereas the undegenerated transistor would exhibit an *expansive* behavior. In other words, resistive degeneration of a square-law device *creates* third-order distortion.

To compute the IP_3 of the stage, we write from Chapter 2,

$$A_{IIP3} = \sqrt{\frac{4}{3}\left|\frac{\alpha_1}{\alpha_3}\right|} \tag{5.174}$$

$$= \sqrt{\frac{2g_m}{3R_S}\frac{(1 + g_m R_S)^2}{K}}. \tag{5.175}$$

The 1-dB compression point follows the same expression but lowered by a factor of 3.03 (9.6 dB).

The reader may wonder if the above analysis of nonlinearity confuses large-signal and small-signal operations by expression α_1-α_3 in terms of the device transconductance. It is helpful to bear in mind that g_m in the above expressions is merely a short-hand notation for a *constant* value, $2K(V_{in0} - R_S I_{D0} - V_{TH})$, and independent of the input. It is, of course, plausible that α_1-α_3 must be independent of the input; otherwise, the polynomial's order exceeds 3.

Example 5.29

A student measures the IP_3 of the CS stage of Fig. 5.81 in the laboratory and obtains a value equal to *half* of that predicted by Eq. (5.175). Explain why.

Solution:

The test setup is shown in Fig. 5.82, where the signal generator produces the required input.[22] The discrepancy arises because the generator contains an internal output resistance $R_G = 50\,\Omega$, and it *assumes* that the circuit under test provides input matching, i.e., $Z_{in} = 50\,\Omega$. The generator's display therefore shows $A_0/2$ for the peak amplitude.

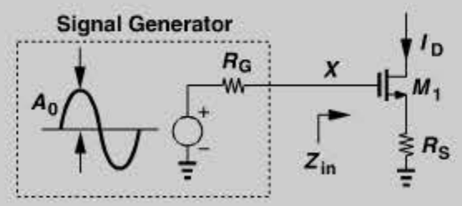


Figure 5.82 *CS stage driven by finite signal source impedance.*

(Continues)

22. In reality, the outputs of two generators are summed for a two-tone test.

Example 5.29 *(Continued)*

The simple CS stage, on the other hand, exhibits a high input impedance, sensing a peak amplitude of A_0 rather than $A_0/2$. Thus, the level that the student reads is half of that applied to the circuit. This confusion arises in IP_3 measurements because this quantity has been traditionally defined in terms of the *available* input power.

Example 5.30

Compute the IP_3 of a common-gate stage if the input is matched. Neglect channel-length modulation and body effect.

Solution:

The circuit is shown in Fig. 5.83, where we have

$$I_D = K(V_b - V_{in} - I_D R_S - V_{TH})^2, \tag{5.176}$$

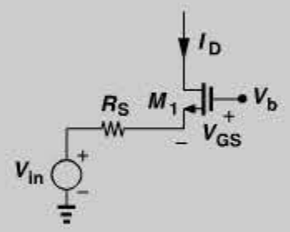


Figure 5.83 *CG stage for nonlinearity calculations.*

and $K = (1/2)\mu_n C_{ox}(W/L)$. Differentiating both sides with respect to V_{in} gives

$$\frac{\partial I_D}{\partial V_{in}} = 2K(V_b - V_{in} - I_D R_S - V_{TH})\left(-1 - R_S \frac{\partial I_D}{\partial V_{in}}\right). \tag{5.177}$$

In the absence of signals, $2K(V_b - V_{in0} - I_{D0}R_S - V_{TH})$ is equal to the transconductance of M_1, and hence

$$\frac{\partial I_D}{\partial V_{in}}|_{Vin0} = \frac{-g_m}{1 + g_m R_S}. \tag{5.178}$$

The second derivative is identical to that of the CS stage, Eq. (5.171):

$$\frac{\partial^2 I_D}{\partial V_{in}^2}|_{Vin0} = \frac{2K}{(1 + g_m R_S)^3}, \tag{5.179}$$

and the third derivative emerges as

$$\frac{\partial^3 I_D}{\partial V_{in}^3}|_{Vin0} = \frac{12K^2 R_S}{(1 + g_m R_S)^5}. \tag{5.180}$$

Example 5.30 *(Continued)*

Thus, the IP_3 expression in Eq. (5.175) applies here as well. For input matching, $R_S = 1/g_m$. However, as explained in Example 5.29, the definition of IP_3 is based on the *available* signal power, i.e., that which is delivered to a matched load. Thus, the peak value predicted by Eq. (5.175) must be divided by 2, yielding

$$A_{IIP3} = \frac{2}{K}\sqrt{\frac{2}{3}g_m} \tag{5.181}$$

$$= 4\sqrt{\frac{2}{3}}(V_{GS0} - V_{TH}), \tag{5.182}$$

where V_{GS0} denotes the bias value of the gate-source voltage.

5.7.2 Undegenerated CS Stage

Consider the CS stage shown in Fig. 5.80. Submicron transistors substantially depart from square-law characteristics. The effect of mobility degradation due to both vertical and lateral fields in the channel can be approximated as

$$I_D = \frac{1}{2}\mu_0 C_{ox}\frac{W}{L}\frac{(V_{GS} - V_{TH})^2}{1 + (\frac{\mu_0}{2v_{sat}L} + \theta)(V_{GS} - V_{TH})}, \tag{5.183}$$

where μ_0 denotes the zero-field mobility, v_{sat} the saturation velocity of the carriers, and θ the effect of the vertical field [11]. If the second term in the denominator remains much less than unity, we can write $(1 + \varepsilon)^{-1} \approx 1 - \varepsilon$ and hence

$$I_D \approx \frac{1}{2}\mu_0 C_{ox}\frac{W}{L}\left[(V_{GS} - V_{TH})^2 - \left(\frac{\mu_0}{2v_{sat}L} + \theta\right)(V_{GS} - V_{TH})^3\right]. \tag{5.184}$$

The input signal, V_{in}, is superimposed on a bias voltage, $V_{GS0} = V_b$. We therefore replace V_{GS} with $V_{in} + V_{GS0}$, obtaining

$$I_D \approx K[2 - 3a(V_{GS0} - V_{TH})](V_{GS0} - V_{TH})V_{in} + K[1 - 3a(V_{GS0} - V_{TH})]V_{in}^2 - KaV_{in}^3 + K(V_{GS0} - V_{TH})^2 - aK(V_{GS0} - V_{TH})^3, \tag{5.185}$$

where $K = (1/2)\mu_0 C_{ox}(W/L)$ and $a = \mu_0/(2v_{sat}L) + \theta$. We recognize the coefficient of V_{in} as the transconductance $(\partial I_D/\partial V_{in})$ and the last two terms as the bias current. It follows that

$$\alpha_1 = K[2 - 3a(V_{GS0} - V_{TH})](V_{GS0} - V_{TH}) \tag{5.186}$$

$$\alpha_3 = -Ka. \tag{5.187}$$

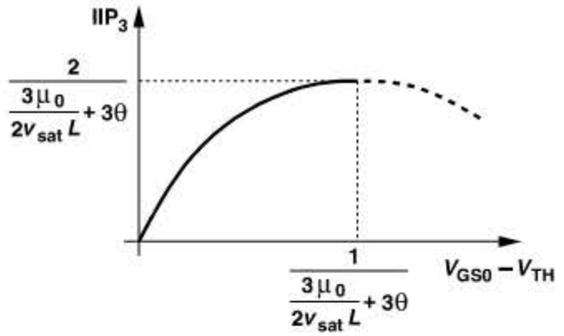


Figure 5.84 *Behavior of IP_3 as a function of overdrive.*

The IP_3 is given by

$$A_{IIP3} = \sqrt{\frac{4}{3} \times \frac{2 - 3a(V_{GS0} - V_{TH})}{a}(V_{GS0} - V_{TH})} \tag{5.188}$$

$$= \sqrt{\frac{\frac{8}{3}(V_{GS0} - V_{TH})}{\frac{\mu_0}{2v_{sat}L} + \theta} - 4(V_{GS0} - V_{TH})^2}. \tag{5.189}$$

We note that the IP_3 rises with the bias overdrive voltage, reaching a maximum of

$$A_{IIP3,max} = \frac{2}{3a} = \frac{2}{3}\frac{1}{\frac{\mu_0}{2v_{sat}L} + \theta} \tag{5.190}$$

at $V_{GS0} - V_{TH} = (3a)^{-1}$ (Fig. 5.84).

Example 5.31

If the second term in the denominator of Eq. (5.183) is only somewhat less than unity, a better approximation must be used, e.g., $(1 + \varepsilon)^{-1} \approx 1 - \varepsilon + \varepsilon^2$. Compute α_1 and α_3 with this approximation.

Solution:

The additional term $a^2(V_{GS} - V_{TH})^2$ is multiplied by $K(V_{GS} - V_{TH})^2$, yielding two terms of interest: $4Ka^2V_{in}(V_{GS} - V_{TH})^3$ and $4Ka^2V_{in}^3(V_{GS} - V_{TH})$. The former contributes to α_1 and the latter to α_3. It follows that

$$\alpha_1 = K[2 - 3a(V_{GS0} - V_{TH}) + 4a^2(V_{GS0} - V_{TH})^2](V_{GS0} - V_{TH}) \tag{5.191}$$

$$\alpha_3 = -Ka[1 - 4a(V_{GS0} - V_{TH})]. \tag{5.192}$$

5.7.3 Differential and Quasi-Differential Pairs

In RF systems, differential signals can be processed using the differential pair shown in Fig. 5.85(a) or the "quasi-differential" pair depicted in Fig. 5.85(b). The two topologies exhibit distinctly different nonlinear characteristics. We know from our above analysis that the dependence of the mobility upon vertical and lateral fields in the channel results in third-order nonlinearity in the quasi-differential pair and an IP_3 given by Eq. (5.189). To study the nonlinearity of the standard differential pair, we recall from basic analog circuits that

$$I_{D1} - I_{D2} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}V_{in}\sqrt{\frac{4I_{SS}}{\mu_n C_{ox}\frac{W}{L}} - V_{in}^2}, \tag{5.193}$$

where V_{in} denotes the input differential voltage. If $|V_{in}| \ll I_{SS}/(\mu_n C_{ox}W/L)$, then

$$I_{D1} - I_{D2} \approx \frac{1}{2}\mu_n C_{ox}\frac{W}{L}V_{in}\sqrt{\frac{4I_{SS}}{\mu_n C_{ox}\frac{W}{L}}}\left(1 - \frac{1}{2}\frac{V_{in}^2}{\frac{4I_{SS}}{\mu_n C_{ox}W/L}}\right). \tag{5.194}$$

That is,

$$\alpha_1 = \sqrt{\mu_n C_{ox}\frac{W}{L}I_{SS}} \tag{5.195}$$

$$\alpha_3 = -\left(\mu_n C_{ox}\frac{W}{L}\right)^{3/2}\frac{1}{8\sqrt{I_{SS}}}, \tag{5.196}$$

and hence

$$A_{IIP3} = \sqrt{\frac{6I_{SS}}{\mu_n C_{ox}W/L}} \tag{5.197}$$

$$= \sqrt{6}(V_{GS0} - V_{TH}), \tag{5.198}$$

where $(V_{GS0} - V_{TH})$ is the overdrive voltage of each transistor in equilibrium ($V_{in} = 0$).[23]

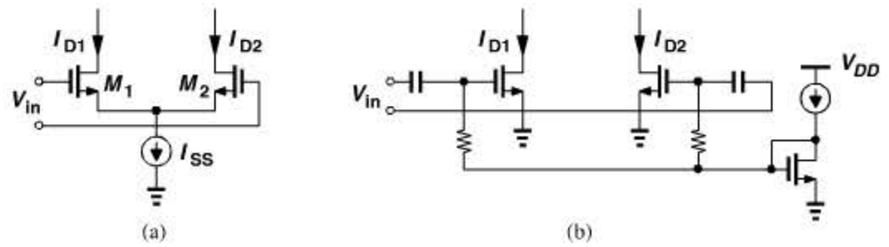


Figure 5.85 *(a) Differential and (b) quasi-differential pairs.*

23. Note that one transistor turns off if the differential input reaches $\sqrt{2}(V_{GS0} - V_{TH})$.

Interestingly, the standard differential pair suffers from third-order nonlinearity even in the absence of field-dependent mobility (i.e., with square-law devices). For this reason, the quasi-differential pair of Fig. 5.85(b) is preferred in cases where linearity is important. In fact, it is for this reason that the differential CS LNA of Fig. 5.65(b) does not employ a tail current source. The quasi-differential pair also saves the voltage headroom associated with the tail current source, proving more attractive as the supply voltage is scaled down.

5.7.4 Degenerated Differential Pair

Consider the degenerated pair shown in Fig. 5.86, where $I_{D1} - I_{D2}$ is the output of interest. Since $I_{D1} + I_{D2} = 2I_0$, we have $\partial(I_{D1} - I_{D2})/\partial V_{in} = 2\partial I_{D1}/\partial V_{in}$. Also, $V_{in1} - V_{GS1} - I_S R_S = V_{in2} - V_{GS2}$ and $I_S = I_{D1} - I_0$. It follows that

$$V_{in} - R_S I_{D1} + R_S I_0 = \frac{1}{\sqrt{K}}(\sqrt{I_{D1}} - \sqrt{I_{D2}}), \tag{5.199}$$

where $V_{in} = V_{in1} - V_{in2}$ and $K = (1/2)\mu_n C_{ox}(W/L)$. Differentiating both sides with respect to V_{in} yields

$$\frac{\partial I_{D1}}{\partial V_{in}}\left[R_S + \frac{1}{2\sqrt{K}}\left(\frac{1}{\sqrt{I_{D1}}} + \frac{1}{\sqrt{I_{D2}}}\right)\right] = 1. \tag{5.200}$$

At $V_{in} = 0$, $I_{D1} = I_{D2}$ and

$$\alpha_1 = \frac{1}{R_S + \dfrac{2}{g_m}}, \tag{5.201}$$

where $g_m = 2I_0/(V_{GS0} - V_{TH})$. Differentiating both sides of (5.200) with respect to V_{in} gives

$$\frac{\partial^2 I_{D1}}{\partial V_{in}^2}\left[R_S + \frac{1}{2\sqrt{K}}\left(\frac{1}{\sqrt{I_{D1}}} + \frac{1}{\sqrt{I_{D2}}}\right)\right] - \frac{\partial I_{D1}}{\partial V_{in}}\left[\frac{1}{4\sqrt{K}}\left(\frac{1}{I_{D1}^{3/2}}\frac{\partial I_{D1}}{\partial V_{in}} + \frac{1}{I_{D2}^{3/2}}\frac{\partial I_{D2}}{\partial V_{in}}\right)\right] = 0. \tag{5.202}$$

Note that for $V_{in} = 0$, we have $\partial^2 I_{D1}/\partial V_{in}^2 = 0$ because $\partial I_{D1}/\partial V_{in} = -\partial I_{D2}/\partial V_{in}$ and the term in the second set of square brackets vanishes. Differentiating once more and exploiting

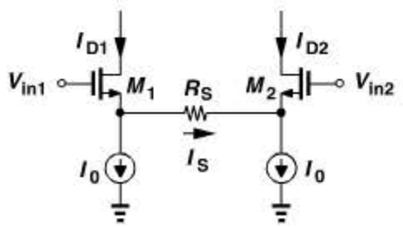


Figure 5.86 *Degenerated differential pair.*

this fact, we have

$$\frac{\partial^3 I_{D1}}{\partial V_{in}^3}|_{Vin=0} = \frac{-3}{(R_S + \dfrac{2}{g_m})^4 g_m I_0^2} = 6\alpha_3. \tag{5.203}$$

It follows that $6\alpha_3 = \partial^3(I_{D1} - I_{D2})/\partial V_{in}^3 = 2\partial^3 I_{D1}/\partial V_{in}^3$. We now have that

$$A_{IIP3} = \frac{2I_0}{3}\sqrt{g_m\left(R_S + \frac{2}{g_m}\right)^3}. \tag{5.204}$$

REFERENCES

[1] J. Rogin et al., "A 1.5-V 45-mW Direct-Conversion WCDMA Receiver IC in 0.13-m CMOS," *IEEE Journal of Solid-State Circuits,* vol. 38, pp. 2239–2248, Dec. 2003.

[2] D. K. Shaeffer and T. H. Lee, "A 1.5-V, 1.5-GHz CMOS Low Noise Amplifier," *IEEE J. Solid-State Circuits,* vol. 32, pp. 745–759, May 1997.

[3] P. Rossi et al., "A Variable-Gain RF Front End Based on a Voltage-Voltage Feedback LNA for Multistandard Applications," *IEEE J. Solid-State Circuits,* vol. 40, pp. 690–697, March 2005.

[4] X. Li, S. Shekar, and D. J. Allstot, "G_m-Boosted Common-Gate LNA and Differential Colpitts VCO/QVCO in 0.18-um CMOS," *IEEE J. Solid-State Circuits,* vol. 40, pp. 2609–2618, Dec. 2005.

[5] A. Zolfaghari, A. Y. Chan, and B. Razavi, "Stacked Inductors and 1-to-2 Transformers in CMOS Technology," *IEEE Journal of Solid-State Circuits,* vol. 36, pp. 620–628, April 2001.

[6] F. Bruccoleri, E. A. M. Klumperink, and B. Nauta, "Wideband CMOS Low-Noise Amplifier Exploiting Thermal Noise Canceling," *IEEE J. Solid-State Circuits,* vol. 39, pp. 275–281, Feb. 2004.

[7] B. Razavi, "Cognitive Radio Design Challenges and Techniques," *IEEE Journal of Solid-State Circuits,* vol. 45, pp. 1542–1553, Aug. 2010.

[8] B. Razavi et al., "A UWB CMOS Transceiver," *IEEE Journal of Solid-State Circuits,* vol. 40, pp. 2555–2562, Dec. 2005.

[9] M. Zargari et al., "A Single-Chip Dual-Band Tri-Mode CMOS Transceiver for IEEE 802.11a/b/g Wireless LAN," *IEEE Journal of Solid-State Circuits,* vol. 39, pp. 2239–2249, Dec. 2004.

[10] J. R. Long and M. A. Copeland, "The Modeling, Characterization, and Design of Monolithic Inductors for Silicon RF ICs," *IEEE J. Solid-State Circuits,* vol. 32, pp. 357–369, March 1997.

[11] B. Razavi, *Design of Analog CMOS Integrated Circuits,* Boston: McGraw-Hill, 2001.

PROBLEMS

5.1. Assuming $Z_{in} = x + jy$, derive an equation for constant-Γ contours in Fig. 5.4.

5.2. If $R_p = R_S$ and $g_m R_S \approx 1$, determine the NF in Eq. (5.18) by considering the first three terms. What value of g_m is necessary to achieve a noise figure of 3.5 dB?

5.3. Repeat Example 5.5 by solving the specific network shown in Fig. 5.10(a).

5.4. Determine the noise figure of the stages shown in Fig. 5.87 with respect to a source impedance of R_S. Neglect channel-length modulation and body effect.

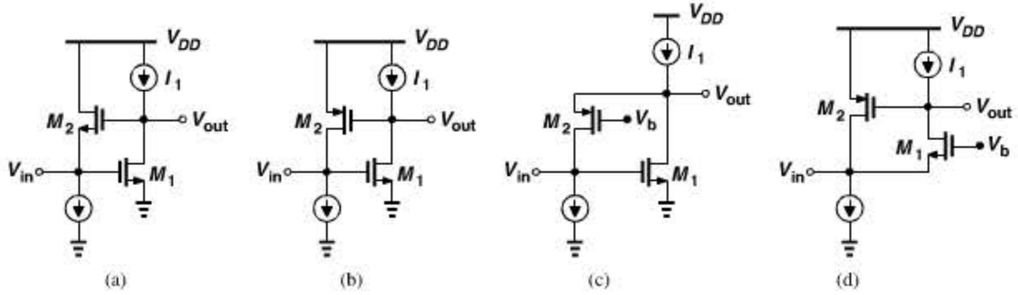


Figure 5.87 *Stages for NF calculation.*

5.5. For the inductively-loaded CS stage of Fig. 5.11(b), determine V_{out}/V_{in} and find the voltage gain at the resonance frequency, $\omega_0 = 1/\sqrt{L_1(C_1 + C_F)}$, if $|jC_1\omega_0| \ll g_m$.

5.6. For the CS stage of Fig. 5.13(a), determine the closed-loop gain and noise figure if channel-length modulation is not neglected. Assume matching at the input.

5.7. For the complementary stage shown in Fig. 5.15, determine the closed-loop gain and noise figure if channel-length modulation is not neglected. Assume matching at the input.

5.8. For the CG stage of Fig. 5.16(a), compute the noise figure at the output resonance frequency if $g_m \neq 1/R_S$. How can g_m be chosen to yield a noise figure lower than $1 + \gamma + 4R_S/R_a$?

5.9. A circuit exhibits a noise figure of 3 dB. What percentage of the output noise power is due to the source resistance, R_S? Repeat the problem for NF = 1 dB.

5.10. Determine the noise figure of the CG circuits shown in Fig. 5.17.

5.11. In Example, 5.10, we concluded that the noise of M_2 reaches the output unattenuated if ω is greater than $(R_1C_X)^{-1}$ but much less than $g_{m2}/(C_{GS2} + C_X)$. Does such a frequency range exist? In other words, under what conditions do we have $(R_1C_X)^{-1} < \omega \ll g_{m2}/(C_{GS2} + C_X) >$? Assume $g_{m2} \approx g_{m1}$ and recall that $g_{m1}R_1$ is the gain of the LNA and C_X is on the order of C_{SG2}.

5.12. If L_G in Fig. 5.34 suffers from a series resistance of R_t, determine the noise figure of the circuit.

5.13. The LNA shown in Fig. 5.88 is designed to operate with low supply voltages. Each inductor is chosen to resonate with the total capacitance at its corresponding node at the frequency of interest. Neglect channel-length modulation and body effect and the noise due to the loss in L_2. Determine the noise figure of the LNA with respect to a source resistance R_S assuming that L_1 can be viewed as a resistance equal to R_p at the resonance frequency. Make sure the result reduces to a familiar form if $R_p \to \infty$. (Hint: the equivalent transconductance of a degenerated common-source stage is given by $g_m/(1 + g_mR_1)$, where R_1 denotes the degeneration resistance.)

5.14. Determine S_{11} for both topologies in Fig. 5.40 and compute the maximum deviation of the center frequency for which S_{11} remains lower than -10 dB.

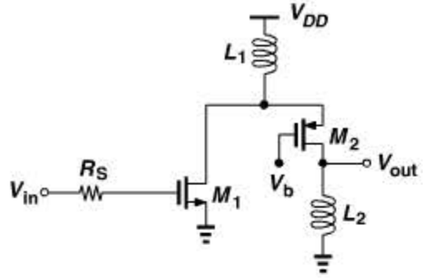


Figure 5.88 *Folded-cascode LNA.*

5.15. Repeat the analysis of the CG stage in Fig. 5.43 while including channel-length modulation.

5.16. Repeat the NF analysis of the CG stage in Fig. 5.43 while including the noise of the feedback network as a voltage, $\overline{V_{nF}^2}$, in series with its input.

5.17. Prove that the input-referred noise of the feedforward amplifier in Fig. 5.45(a) manifests itself as the fourth term in Eq. (5.124).

5.18. Repeat the analysis of the CG stage of Fig. 5.45(a) while including channel-length modulation.

5.19. Is the noise of R_F in Fig. 5.48(b) cancelled? Explain.

5.20. For the circuit shown in Fig. 5.89, we express the input-output characteristic as

$$I_{out} - I_0 = \alpha_1(V_{in} - V_0) + \alpha_2(V_{in} - V_0)^2 + \cdots, \tag{5.205}$$

where I_0 and V_0 denote the bias values, i.e., the values in the absence of signals. We note that $\partial I_{out}/\partial V_{in} = \alpha_1$ at $V_{in} = V_0$ (or $I_{out} = I_0$). Similarly, $\partial I_{out}^2/\partial^2 V_{in} = 2\alpha_2$ at $V_{in} = V_0$ (or $I_{out} = I_0$).

(a) Write a KVL around the input network in terms of V_{in} and I_{out} (with no V_{GS}). Differentiate both sides *implicitly* with respect to V_{in}. You will need this equation in part (b). Noting that $2\sqrt{KI_0} = g_m$, where $K = \mu_nC_{ox}W/L$, find $\partial I_{out}/\partial V_{in}$ and hence α_1.

(b) Differentiate the equation obtained in part (a) with respect to V_{in} once more and compute α_2 in terms of I_0 and g_m.

(c) Determine the IP_2 of the circuit.

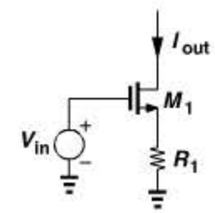


Figure 5.89 *Stage for IP_2 calculation.*

5.21. Determine the noise figure in Example 5.21 if the gain is reduced by 3 dB.

5.22. Compare the power consumptions of the single-ended and differential CS stages discussed in Section 5.6.1. Consider two cases: (a) the differential stage is derived by only halving L_1 (and hence has a lower noise figure), or (b) the differential stage is designed for the same NF as the single-ended circuit.

5.23. Repeat the analysis of the differential CG stage NF if a 1-to-2 balun is used. Such a balun provides a voltage gain of 2.

5.24. Consider a MOS transistor configured as a CS stage and operating in saturation. Determine the IP_3 and P_{1dB} if the device (a) follows the square-law behavior, $I_D \propto (V_{GS} - V_{TH})^2$, or (b) exhibits field-dependent mobility [Eq. (5.183)]. (Hint: IP_3 and P_{1dB} may not be related by a 9.6-dB difference in this case.)

CHAPTER

6

MIXERS

In this chapter, our study of building blocks focuses on downconversion and upconversion mixers, which appear in the receive path and the transmit path, respectively. While a decade ago, most mixers were realized as a Gilbert cell, many more variants have recently been introduced to satisfy the specific demands of different RX or TX architectures. In other words, a stand-alone mixer design is no longer meaningful because its ultimate performance heavily depends on the circuits surrounding it. The outline of the chapter is shown below.

- **General Considerations**
  - Mixer Noise Figures
  - Port–to–Port Feedthrough
  - Single–Balanced and Double–Balanced Mixers
  - Passive and Active Mixers
- **Passive Mixers**
  - Conversion Gain
  - Noise
  - Input Impedance
  - Current–Driven Mixers
- **Active Mixers**
  - Conversion Gain
  - Noise
  - Linearity
- **Improved Mixer Topologies**
  - Active Mixers with Current Source Helpers
  - Active Mixers with High IP_2
  - Active Mixers with Low Flicker Noise
- **Upconversion Mixers**
  - Passive Mixers
  - Active Mixers

6.1 GENERAL CONSIDERATIONS

Mixers perform frequency translation by multiplying two waveforms (and possibly their harmonics). As such, mixers have three distinctly different ports. Figure 6.1 shows a generic transceiver environment in which mixers are used. In the receive path, the downconversion mixer senses the RF signal at its "RF port" and the local oscillator waveform at its "LO port." The output is called the "IF port" in a heterodyne RX or the "baseband port" in a direct-conversion RX. Similarly, in the transmit path, the upconversion mixer input sensing the IF or the baseband signal is called the IF port or the baseband port, and the output port is called the RF port. The input driven by the LO is called the LO port.

How linear should each input port of a mixer be? A mixer can simply be realized as depicted in Fig. 6.2(a), where V_{LO} turns the switch on and off, yielding $V_{IF} = V_{RF}$ or