FUNDAMENTALS OF
HIGH FREQUENCY CMOS ANALOG INTEGRATED CIRCUITS

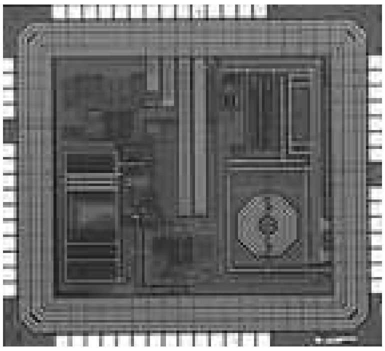
(Cambridge University Press, 2009)

ADDENDA

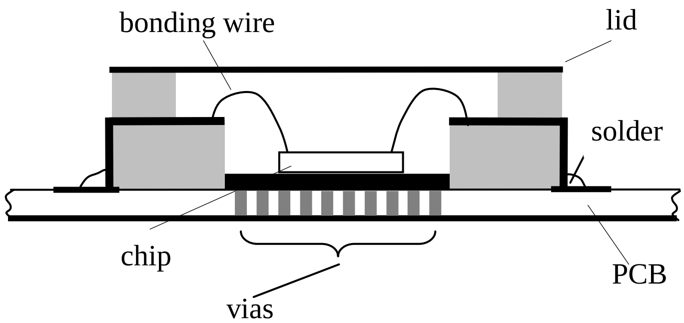
(Parts written after the publication of the book)

1.2.4. Pads and Packages

A chip must be protected from environmental (mechanical and chemical) hazards and must have features to make connections to the rest of the circuit. In Fig. 1.37 (a) and (b) the photo micrograph of an RF chip and the cross-section of a typical RF package are given, respectively. As shown in Fig. 1.37 (a) the bonding pads are placed along the edges of the chip that provide lands to make ultrasonic or thermo-compression bondigs for bonding wires to connect input, output, DC supply and ground nodes of the IC to the corresponding legs of the package[1]. The dimensions of the metal (Al or Cu) bonding pads are in the range of 80 to 100 μm and the diameter of the typical (Al or Au) bonding wires is 25 μm.



(a)



(b)

Figure 1.37. (a) The photo-micrograph of an RF chip. (b) The schematic cross-section of a typical RF package.

In addition to serve as bonding areas, the bonding pads must have features to protect the IC from over- voltages that may break-down the gate oxides or the p-n junctions of the MOS transistors connected to this pad. In Fig. 1.38 a simple electrostatic discharge (ESD) protection circuit for an input pad is shown. It can be easily seen that this circuit provides a lower limit equal to one diode voltage (approximately 0.7 V) smaller than the ground potential and an upper limit one diode voltage higher than V_{DD} for the voltage reaching to the circuit input node, during normal operation. In the event of an external ESD pulse with very high voltage levels, the protection diodes are designed to go into reverse breakdown mode in order to drain a very large curent to V_{DD} or to ground, instantaneously. The reverse breakdown voltage of the diodes must be smaller than the gate oxide breakdown voltage of the input transistor, to prevent permanent damage. Also note that a relatively small-valued current limiting series resistor (also called the ballast resistor) is needed to prevent overheating.

ESD protection circuits for digital and analog ICs are different in nature. The essential requirements for analog circuits are minimal parasitics that may affect the performance of the circuit, especially at the higher end of the operating frequency range, minimal additional noise and low nonlinearity. These requirements unfortunately conflict with the physical constraints that are dictated by the protection devices.

[1] Another possibility is not to use any package, to fix the chip directly on to the printed circuit board (PCB) and make the bondings from bonding pads on the chip to the corresponding pads on the PCB. In this case the chip and the bonding wires must be protected with an appropriate glue.

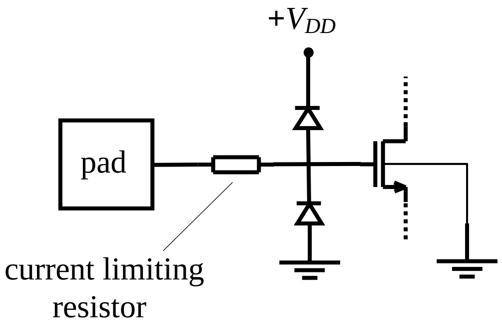


Figure 1.38. The circuit diagram of a typical input ESD protection circuit.

In particular, the ESD protection diodes must be designed with a relatively large effective junction area (using inter-digitated structures) in order to support the large current density that may flow in the event of an ESD pulse. The diode must also be surrounded by a single or double guard ring to prevent stray currents (Fig. 1.39). The large junction area results in large parasitic capacitances that appear in parallel to the input (or output) node, in addition to the pad metal to silicon capacitance. Similarly, the current limiting resistors must be realized as diffusion resistors in the substrate (to allow heat dispersion), which introduces relatively large parasitic capacitances with respect to the substrate. Hence, the RC parasitics that are due to ESD protection circuitry are virtually unavoidable in the signal path.

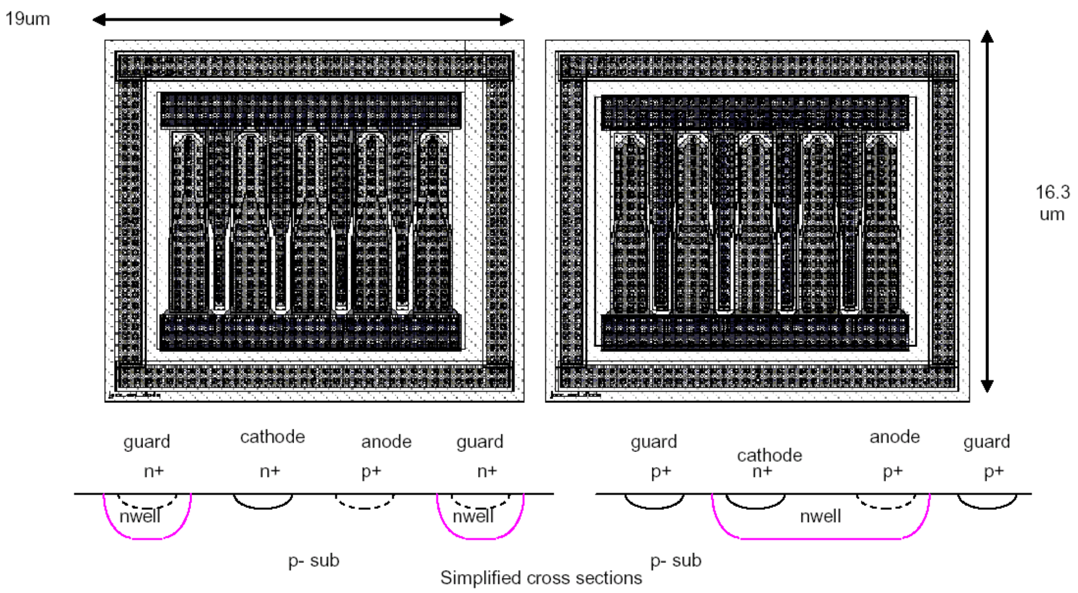


Figure 1.39. Typical large-area ESD protection diode structures.

In addition to input and output protection diodes, the power supply networks must also contain so-called power clamp devices between V_{DD} and ground that are capable of conducting large curents without sustaining permanent damage, if the supply network itself is subjected to an ESD event. In chips with multiple power supply domains, all independent supply networks must be interconnected using a bridge device (usually consisting of two cross-coupled junction diodes) that is inactive under normal conditions, but is capable of passing a large instantaneous current in case of ESD event.

Packages have several functions. One of them is to protect the chip and the bonding wires, as mentioned before. The second function is to connect the internal nodes of the chip to the rest of the circuit via the bonding pads, the bonding wires and the legs of the package that are soldered to the appropriate copper strips on the PCB. Finally the metal floor of the package on which the the chip eutectic bonded or adhered, provides a relatively low thermal resistance from the back side of the chip to the PCB. Forming large number of metal filled vias between upper copper area corresponding to the metal floor of the package and the lower copper plate of the PCB as shown in 1.37-a reduces the thermal resistance from the chip to the ambient and effectively improves the cooling efficiency.

The most important parasitics of the packages are the self inductances of the bonding wires that are connected in series to the inputs, outputs, DC supply ald ground lines. The self inductance of a 25 μm diameter bonding wire is approximately 1 nH. Depending on tha dimensions of the package an the distance from a bonding pad to the corresponding leg, the parasitic inductance of a bonding wire is usually in the range of several nanohenries. These parasitic inductances influences the frequency characteristics of the gain and the input and output impedances.

In Fig. 1.40 a chip is shown with its bonding wire inductances. For the sake of simplicity it is assumed that there are only two different internal blocks in the chip and they are marked with (1) and (2). These blocks may be different stages of an analog circuit or, if the chip is a mixed mode IC, one block may be an analog circuit and the other, digital.

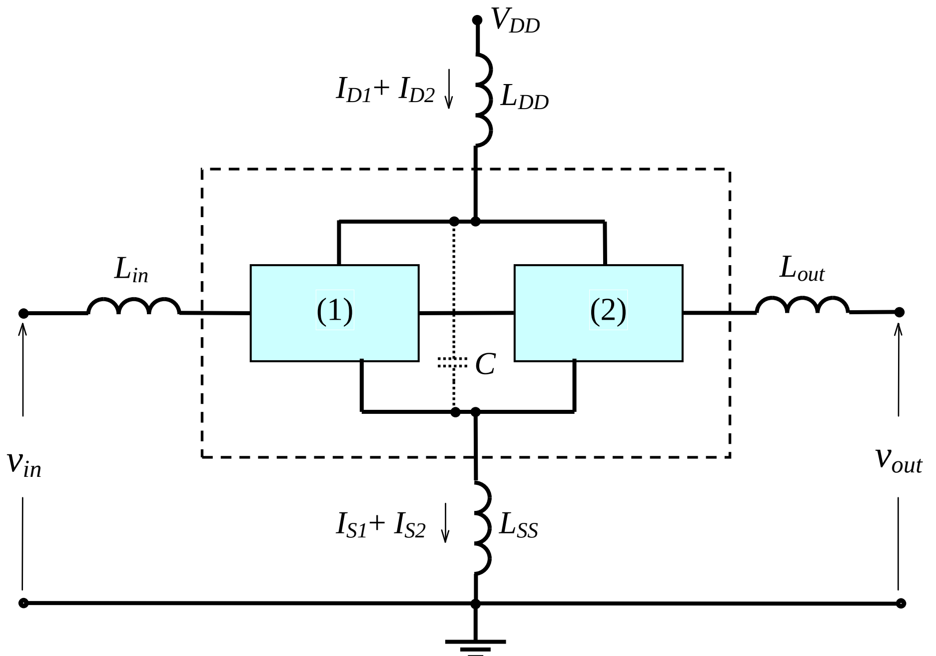


Figure 1.40. Schematic of an IC with the bonding wire parasitics. *C* is a high capacity on-chip capacitor to by-pass the high frequency components of the DC supply current and hence to maintain the internal V_{DD} to internal ground voltage constant.

The input bonding wire inductance is in series with the input capacitance of (1) and forms a series resonance circuit. This resonance effect, that will be investigated in Chapter 4, can be useful or harmful depending on the nature of the circuit. The effects of this inductance on the frequency characteristic of gain and on the input impedance must be simulated and tried to find sulutions to improve unwanted results. The effects of the output bonding wire is similar.

The efects of the ground connection bonding wire is important for fully analog as well as for mixed-mode chips. In case of an analog chip the ground return currents of (1) and (2) (and the currents of others if there are more than 2 stages) flow over L_{SS}. The signal voltage drop on this impedance is in the input loop of (1) and the resulting frequency dependent feedback can severely affect the performance of the circuit.

The voltage drop on L_{DD} due the signal components of the supply currents of (1) and (2) results in a non constant internal V_{DD} that causes also a –relatively weak- feedback. This effect can be minimized with an on-chip by-pass capacitor connected between the internal V_{DD} node and the internal ground, as shown in Fig. 1.40. This high capacitance capacitor is usually realized using all the possible area on the chip surface.

In case of a mixed- mode circuit; for example (1) is an analog block and (2) is digital, the pulsating supply and ground currens of (2) injects an unwanted noise to the analog block. To prevent (at least to minimize) this effect the supply and ground lines of the analog and digital blocks must be separated and must be connected to the legs of the package through different bonding wires.

These explanations show that the bonding wire inductances are severe problem sources and it is necessary to reduce them. One of the most frequently used solution is to use more than one bonding pads and bonding wires in parallel. Another solution is to use bonding ribbons (that exhibit smaller self inductance per micron) instead of round bonding wires. The more effective solution is to completely eliminate the bonding wires and use flip-chip (or similar) bonding techniques for the circuits operating in the microwave frequencies.

Another very effective method to minimize the adverse effects of the DC supply and ground bonding wires is to design the circuit on the chip with differential building blocks. Since the input signal of an analog (or digital) circuit controls the sharing ratio of currents flowing through the two halves of the circuit and the sum of these currents is constant, there is essentially no signal voltage drop on L_{SS} and L_{DD}, and consequently any feedback or cross noise injection risks are eliminated.

3.8.4 A Band Widening Technique: Inductive Peaking

All resistance-loaded amplifiers can be considered as "wide-band amplifiers" by their nature. The low end of the band is determined by the coupling capacitors (if there are any), and the high end by the capacitance parallel to the load resistor. A wide-band amplifier is characterized with its gain that must be flat and equal to the low frequency gain with an acceptable tolerance up to the high end of the band, and the upper cut-off frequency where the gain drops to 3dB below the low frequency gain.

For a resistance-loaded common source amplifier[2], the 3 dB frequency, the low frequency gain and the gain-bandwidth product (the figure of merit for wide band amplifiers) were given previously as

$$\omega_p = \frac{G_o}{C_L} \simeq \frac{1}{C_L R_L} \quad \text{for } r_{ds} \gg R_L \tag{3.4b}$$

$$A_v(0) = -\frac{g_m}{G_o} \simeq -g_m R_L \tag{3.5}$$

$$GBW = \frac{g_m}{2\pi C_L} \tag{3.6}$$

where R_L is the load resistance of the transistor and $C_L = C_o + C_L^{'}$ is the total capacitance of the output node, where C_o is the output capacitance of the transistor, mainly determined by the junction capacitance of the drain region, and $C_L^{'}$ the input capacitance of the succeeding stage (or the load). $C_L^{'}$ is usually given as one of the input parameters of the design problem. C_o can be initially neglected, or an estimated value depending on the technology can be given. (after the calculation of the dimensions of the transistor the correct value of C_o must be found and the design must be updated, if necessary).

To reach the targeted 3dB frequency that is equal to the pole frequency for a resistance loaded common source amplifier, the appropriate R_L value can be calculated from (3.4b) as

$$R_L = \frac{1}{2\pi f_p C_L} \tag{3.91}$$

To obtain the targeted gain, according to (3.5) the transconductance must be

$$g_m = \frac{|A_v(0)|}{R_L} \tag{3.92}$$

The transconductance of a transistor operating in the saturation region is determined by the DC drain current and the aspect ratio of the transistor.

$$g_m = \sqrt{2\mu C_{ox} \frac{W}{L} I_D} \tag{1.33}$$

[2] The following investigations can be applied to other amplifier configurations with minor modifications (whenever necessary).

According to (1.33), any combination of *(W/L)* and I_D can be used to obtain the desired g_m value. Choosing lower drain current values will correspond to wider transistors, and therefore, to higher parasitic capacitance values, for the same g_m value. For higher drain current values, the parasitic capacitance can be decreased but the power consumption increases. Another concern related to the drain DC current (the quiescent current) is its position on the output characteristic curves. To obtain maximum dynamic range for the output voltage and minimum nonlinear (harmonic and intermodulation) distortion for a certain amplitude, the operating point must be placed at the center of the saturated operating region (as shown in Fig. 2.1), corresponding to

$$I_D = \frac{(V_{DD} - V_{DS(sat)})}{2R_L} \tag{3.93}$$

which can be considered as the optimum current for the majority of applications. Using this I_D, the aspect ratio of the transistor can be calculated from (1.33):

$$\frac{W}{L} = \frac{g_m^2}{2\mu C_{ox} I_D} \tag{3.94}$$

From (3.6) it is obvious that the factor limiting the bandwidth of an amplifier is C_L, i.e., the sum of the input capacitance of the load and the output parasitic capacitance of the transistor. The only chance to reduce the value of C_L and to increase the gain-bandwidth product is to reduce the width of the transistor that helps to reduce the parasitic junction capacitance of the drain region, at the expense of increasing the drain current and reducing the output signal dynamic range.

Another possibility to decrease the effect of the output capacitance is to compensate it with an inductance in the frequency region around the pole frequency. This technique, called as "inductive peaking", has been investigated and used from the early days of the electronic engineering[3] to present day[4]. The two basic approaches to compensate the load capacitance with an inductor are shown in Fig. 3.47.

The circuit shown in Fig. 3.47(a) is called "parallel (or shunt) compensation" and its small signal equivalent circuit is given in Fig. 3.47(b). In this approach, the output capacitance is resonated with an inductance (that increases the load impedance and hence the gain) in the vicinity of the 3 dB frequency of the basic amplifier, where the gain starts to decrease.

The circuit shown in Fig. 3.47(c) is called "series compensation" and its small signal equivalent circuit is given in Fig. 3.47(d). In this approach, the increase of the capacitance voltage at resonance with respect to the input voltage of the series resonance circuit is used to compensate the decrease of the output voltage (See Chapter 4, Section 4.1.2 for the details of main concepts).

---

[3] See: A. V. Bedford, G. L. Fredendall, "Transient Response of Multistage Video-frequency Amplifiers, Proc. IRE, Vol. 27, 1939, pp. 277-284, and Samuel Seely, "Electronic Engineering", McGraw-Hill, 1956, pp.101-110.

[4] See: Peter Staric, Eric Margan, "Wideband Amplifiers", Springer, 2007, and S. S. Mohan, M. del Mar Hershenson, S.P. Boyd, T.H. Lee, "Bandwidth Extension in CMOS with Optimized On-Chip Inductors", JSSC, Vol. 35, 2000, pp.346-355.

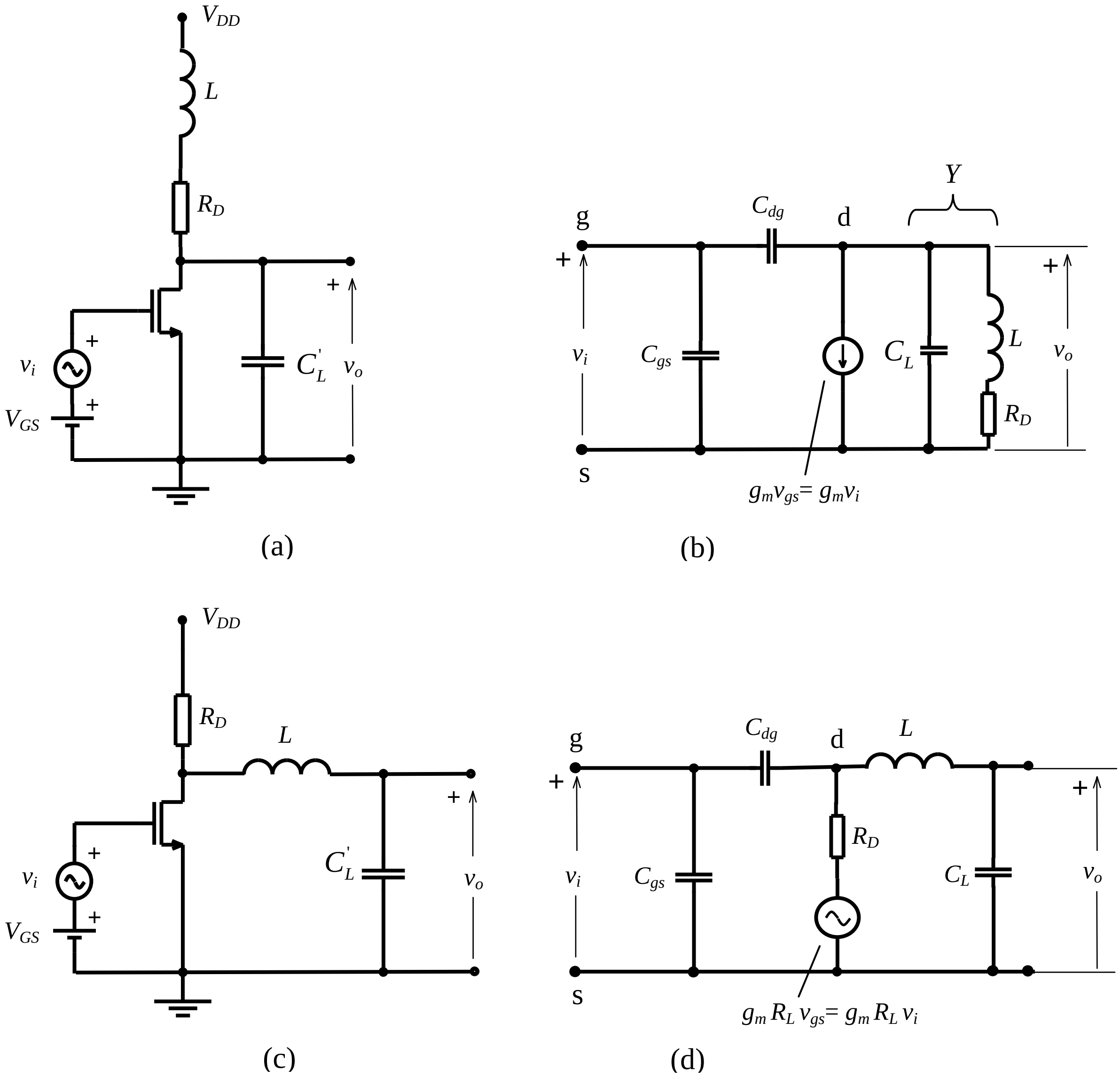


Figure 3.47 (a) Simplified circuit diagram of a resistive loaded common source amplifier with parallel inductive peaking (b) Small signal equivalent circuit. (c) Simplified circuit diagram of a resistive loaded common source amplifier with serial inductive peaking. (d) Small signal equivalent circuit.

Let us first investigate the parallel compensated amplifier. From the equivalent circuit shown in Fig. 3.47(b), the voltage gain can be found as

$$A_v = \frac{v_o}{v_i} = -g_m \frac{R_D + sL}{s^2 LC_L + sC_L R_D + 1} \tag{3.95}$$

The gain in frequency domain, normalized with respect to the low frequency gain:

$$\bar{A} = \frac{|A_v(\omega)|}{|A_v(0)|} = \frac{1 + j\omega \frac{L}{R_D}}{(1 - \omega^2 LC_L) + j\omega C_L R_D} \tag{3.96}$$

Using the pole frequency of the basic (non-compensated) amplifier ($\omega_p = 1/R_D C_L$), the resonance frequency of the LC circuit ($\omega_o = 1/\sqrt{LC_L}$) and a parameter α that is defined as $\alpha = \omega_p / \omega_0$, (3.92) can be arranged in normalized form as

$$\bar{A} = \frac{1 + j(\omega/\omega_P)\alpha^2}{\left[1 - (\omega/\omega_P)^2 \alpha^2\right] + j(\omega/\omega_P)} \qquad (3.97)$$

It must be noted that α can be written in terms of the circuit parameters and helps to calculate of the value of L, which is the most important design parameter:

$$\alpha = \frac{1}{R_D C_L}\sqrt{LC_L} \quad \rightarrow \quad L = \alpha^2 R_D^2 C_L \qquad (3.98)$$

The variation of the normalized gain as a function of the normalized frequency for different values of α is given in Fig. 3.48. It can be seen from this graph that;

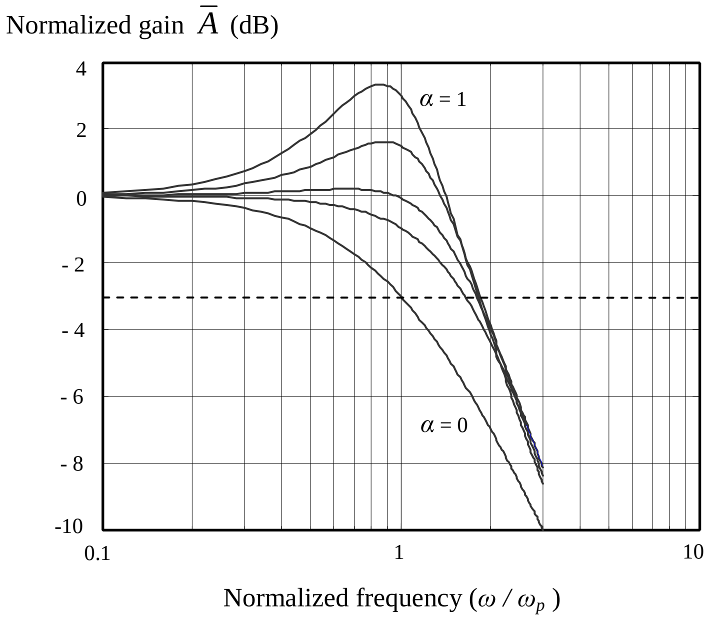


Figure 3.48 Variation of the normalized gain as a function of the normalized frequency for α = 0, 0.6, 0.7, 0.85 and 1.

- Increasing α (increasing L) effectively helps to increase of the bandwidth (and the gain-bandwidth product) of the amplifier. For α = 0.6 the bandwidth is 62% higher than that of a non-compensated amplifier. For higher values of α, the increase of the bandwidth approaches 87%.

- For α = 0.7, the frequency characteristic remains almost flat with a peaking of only 0.2 dB.

- For higher values of α, the bandwidth does not exhibit any further increase but the peaking becomes unacceptably high for many applications.

- Therefore, it can be concluded that the optimum value of α is 0.7.

But this conclusion has to be checked from the point of view of the delay characteristic of the amplifier.

Amplifiers are typically used to amplify complex waveforms, in most cases pulses and square waves. The flatness of the gain-versus-frequency characteristic guarantees the uniform amplification of all Fourier components associated with the waveform, up to the 3dB frequency. It must not be overlooked that to preserve the correct shape of the output waveform, the delay of all Fourier components must be same, in other words, the delay characteristic (not the phase characteristic) must also be flat.

The relation of the phase shift and the delay for a certain frequency is illustrated in Fig. 3.49. From this figure it can be seen that the delay (τ) corresponding to a certain phase shift φ is

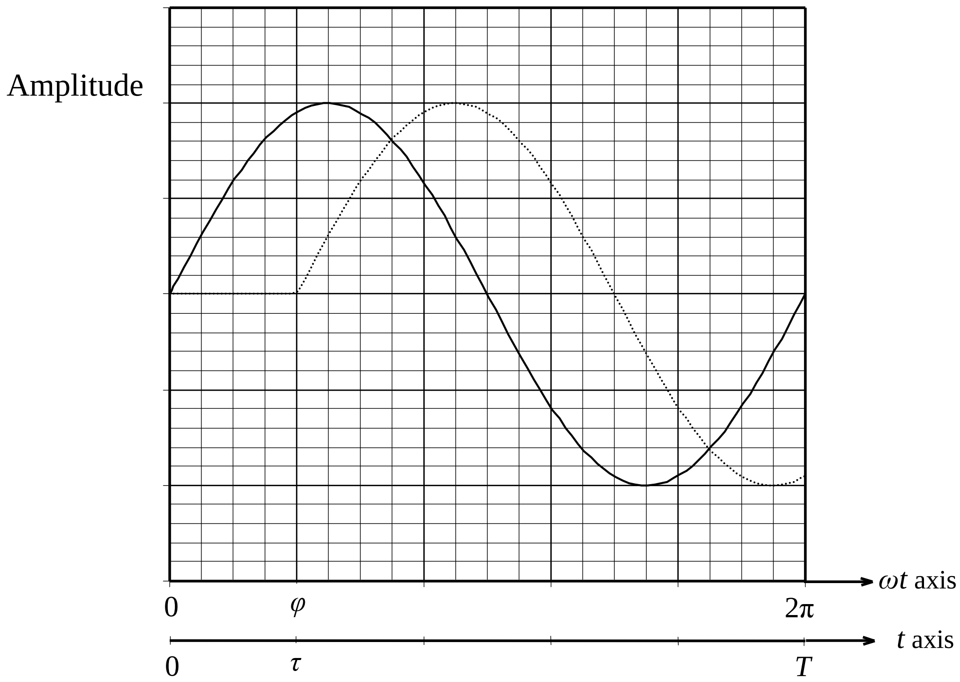


Figure 3.49 Relation of phase shift and time delay.

$$\tau = \varphi \frac{T}{2\pi} = \frac{\varphi}{2\pi f} = \frac{\varphi}{\omega} \qquad (3.99)$$

The normalized delay characteristic obtained from the normalized phase characteristic is given in Fig. 3.50. From these curves it can be seen that:

- The delay characteristic corresponding to the flat magnitude characteristic (e.g. for $\alpha = 0.7$) is not flat.
- The flat delay characteristic corresponds to $\alpha = 0.6$.
- Therefore, for amplification of complex waveforms, despite 13% smaller bandwidth the appropriate value is $\alpha = 0.6$.

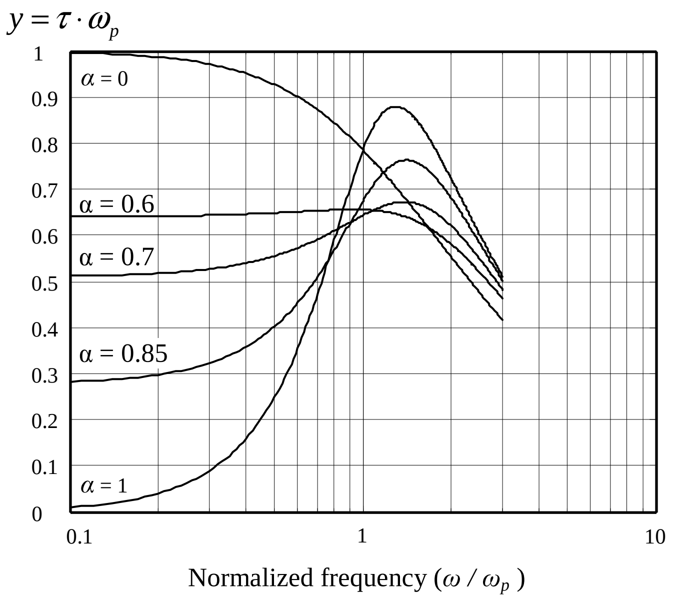


Figure 3.50 Variation of the normalized delay as a function of the normalized frequency for α = 0, 0.6, 0.7, 0.85 and 1.

To illustrate the effect of the phase delay, the PSpice step response simulation results for an amplifier for various values of α (corresponding to various values of L) are given in Fig. 3.51, underlining the importance of the delay characteristic for a wide band amplifier.

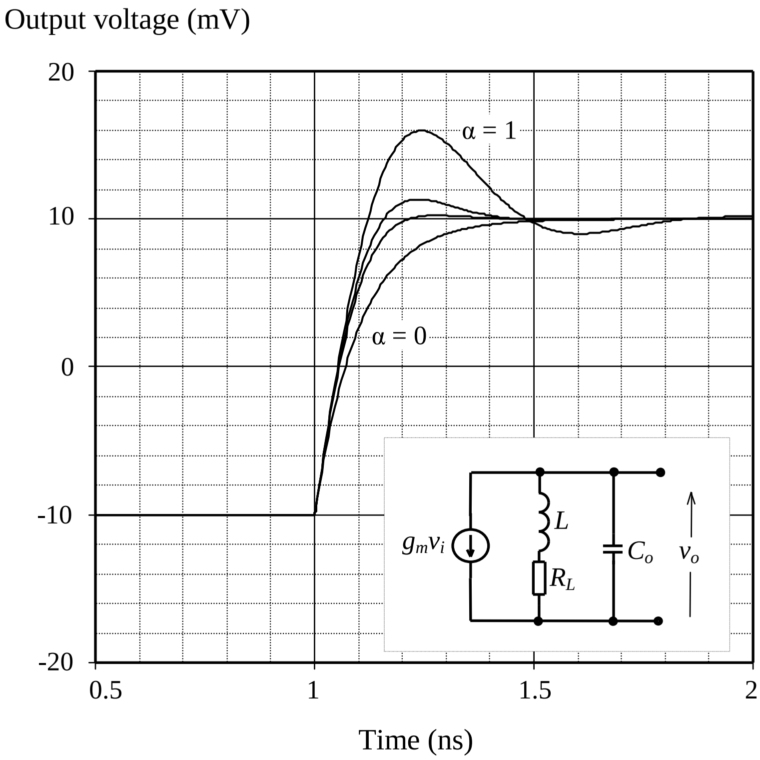


Figure 3.51 The step response of **a** wide-band amplifier model for +1 mV to -1 mV input step voltage. The parameters of the amplifier model are g_m = 10 mS, R_L = 1k ohm and C_o = 100 fF. The responses correspond to α = 0 (L = 0), α = 0.6 (L = 36 nH), α = 0.7 (L = 49 nH) and α = 1 (L = 100 nH).

Example 3.4 Design of a 3GHz bandwidth and 14 dB voltage gain, parallel inductive peaked wide band amplifier.

Technology: AMS035
Supply voltage: V_{DD} = + 3V
Load capacitance: $C_L^{'}$ = 75 fF
Frequency response: Minimum delay distortion
Signal source internal resistance: R_s = 50 ohm

Since the 3 dB frequency of an amplifier having minimum delay distortion (flat delay characteristic) is 62 % wider than that of a non peaked amplifier, the starting point must be a resistance loaded amplifier having 14 dB (A_v = 5) voltage gain and f_p = 3 GHz/1.62 = 1.85 GHz bandwidth.

The load capacitance is given as 75 fF. The other component of the total parallel output capacitance, the output capacitance of the transistor that depends on the dimensions has to be guessed in the beginning and must be checked later on, and updated if necessary. Let us assume the output capacitance of the transistor is also 75 fF, then the total parallel capacitance is C_L = 150 fF.

(3.91) gives the value of the collector load resistance:

$$R_L = \frac{1}{2\pi f_p C_L} = \frac{1}{2\pi \times (1.85\times 10^9)\times(150\times 10^{-12})} = 573.5 \text{ ohm}$$

The value of the transconductance to obtain the targeted gain with this load resistance can be calculated from (3.92):

$$g_m = \frac{5}{573.5} = 8.7 \text{ mS}$$

To obtain maximum output voltage dynamic range, the drain quiescent current, according to (3.93), must be

$$I_D \cong \frac{(3-0.5)}{2\times 573.5} = 2.18 \text{ mA}$$

The aspect ratio of the transistor can be calculated from (3.94):

$$\frac{W}{L} = \frac{(8.7\times 10^{-3})^2}{2\times 300\times(4.54\times 10^{-7})\times(2.18\times 10^{-3})} = 127.46 \quad \rightarrow \quad W \cong 45\ \mu\text{m}$$

Finally, the value of the peaking inductance for a flat phase delay characteristic from (3.98):

$$L = (0.6)^2 \times (573.5)^2 \times (150\times 10^{-15}) = 17.76 \text{ nH}$$

and for comparison purpose, the value of the peaking inductance for a flat gain characteristic;

$$L = (0.7)^2 \times (573.5)^2 \times (150\times 10^{-15}) = 24.17 \text{ nH}$$

The PSpice simulations:

With the calculated values the circuit diagram of the amplifier is given in Fig. 3.52. To obtain the calculated I_D value, a DC sweep must be applied to V_G, the gate bias voltage. The AC simulation with the calculated circuit parameters and the found gate bias voltage shows that the voltage gain is $|A_v(0)| \cong 3$, considerably smaller than the targeted value. This is due to the smaller value of the simulated transconductance as explained in Chapter 1 and in principle can be compensated by increasing the aspect ratio and/or the DC drain current. But the maximum output voltage dynamic range forces us to keep the calculated I_D value. Therefore, to fine tune the gain to 5, the aspect ratio and the bias voltage must be adjusted. The obtained results are $W = 84$ µm and $V_G = 0.95$V (which is in agreement with the pre-estimated $V_{DS(sat)}$ = 0.5 V).

The PSpice netlist is given below, covering the DC sweep, AC gain and delay characteristics, the transient response for a pulse input voltage to check and compare the delay distortion, and the transient response for a 1 GHz sinusoidal input voltage to check the output dynamic range (the *L* values correspond to the fine-tuned amplitude and delay characteristics):

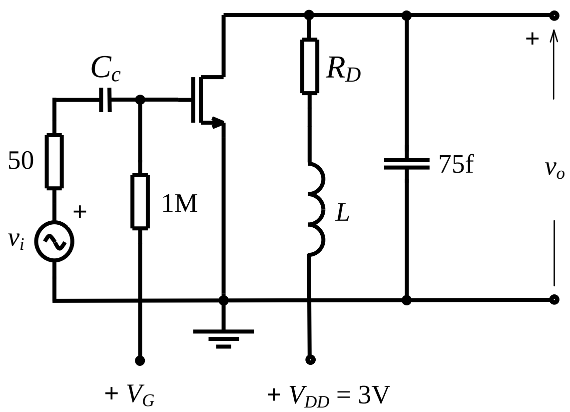
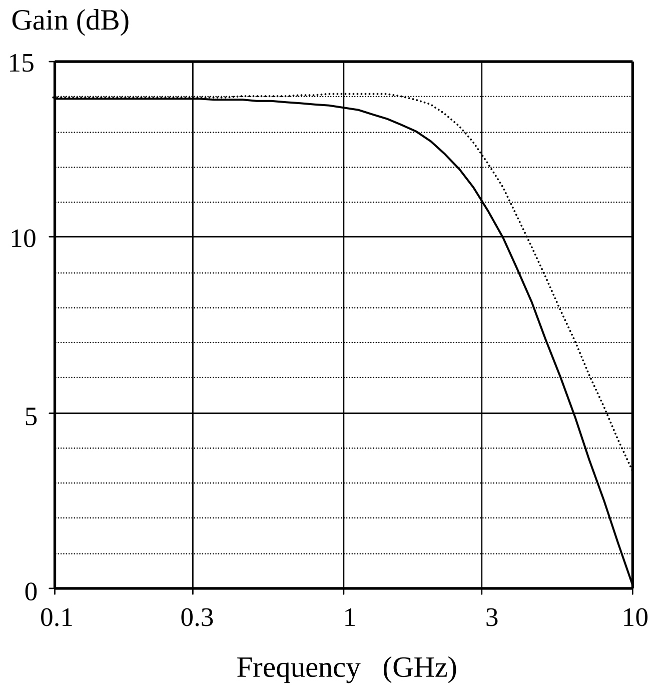


Figure 3.52 Schematic diagram of the amplifier.

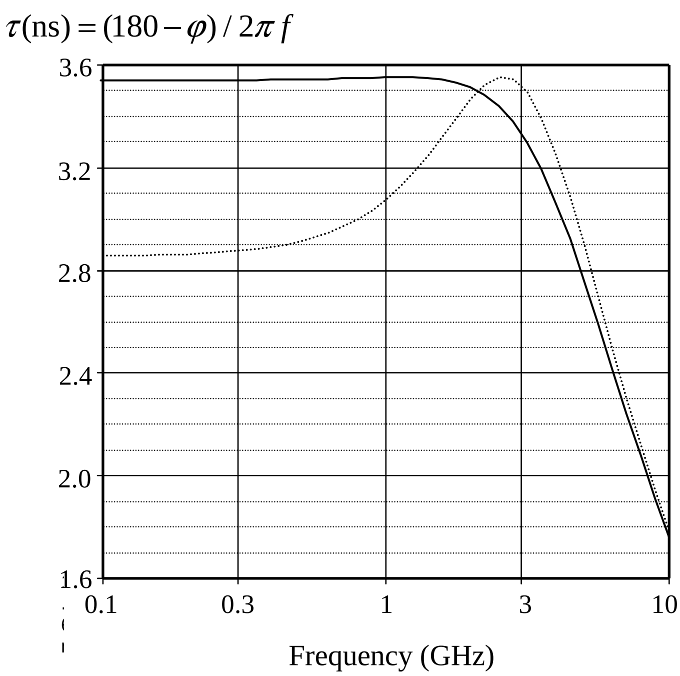
```
*CS INDUCTIVE SHUNT PEAKED AMPLIFIER*

.LIB "ams035.lib"
VDD  100  0  3
M1  1 2 0 0 modn L=.35U  W=84U ad=70e-12 as=70e-12 Pd=84u Ps=84u
CL' 1 0 75f
RL 11 1 573.5
L 100 11 16.6n
*L 100 11 24n
RG 2 200  100k
VG 200 0 .95
vin  23 0 ac 10m
*vin 23 0 pulse 10m -10m .5n 1p 1p 1n 2n
*vin 23 0 sin (0 200mV 1G)
Rs 23 22 50
cc 22 2 100p
.DC VGG .5  2  10M
.AC DEC 20 .1G 10G
*.TRAN .1n 2.5n 0 5p
.PROBE
.END
```

Simulation results for frequency characteristics of the gain and the signal delay are shown in Fig.3.53 (a) and (b). The pulse response corresponding to the flat gain characteristic and the flat delay characteristic are shown in Fig.3.54. The simulation performed to check the position of the operating point, is shown in Fig.3.55.



(a)



(b)

Figure 3.53 PSpice simulation results for (a) the gain characteristics, (b) the delay characteristics. The solid lines correspond to flat delay characteristic (*L* = 16.6 nH) and the dotted lines to the flat gain characteristic (*L* = 24 nH).

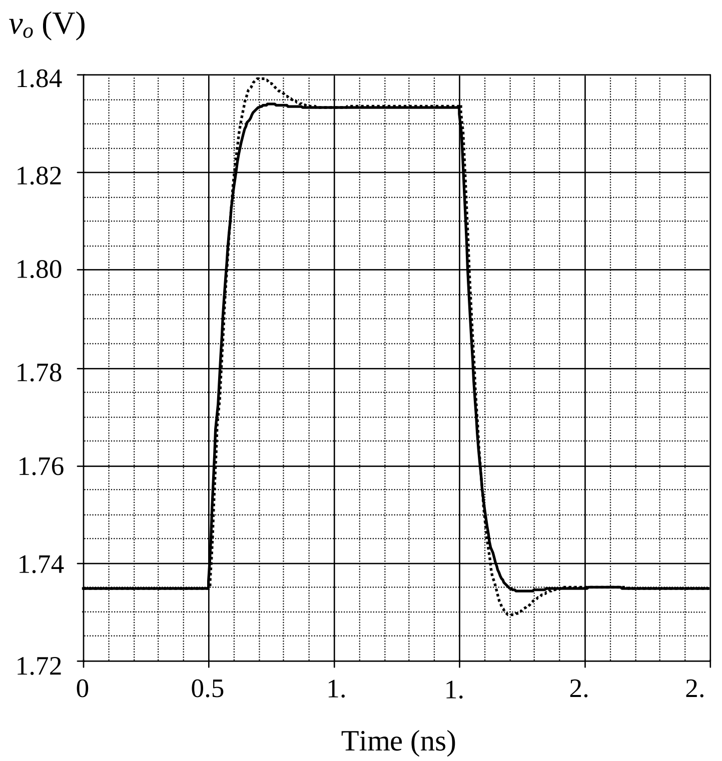


Figure 3.54 The pulse response of the amplifier. The solid line corresponds to the flat delay characteristic (L = 16.6 nH) and the dotted line to L = 24 nH.

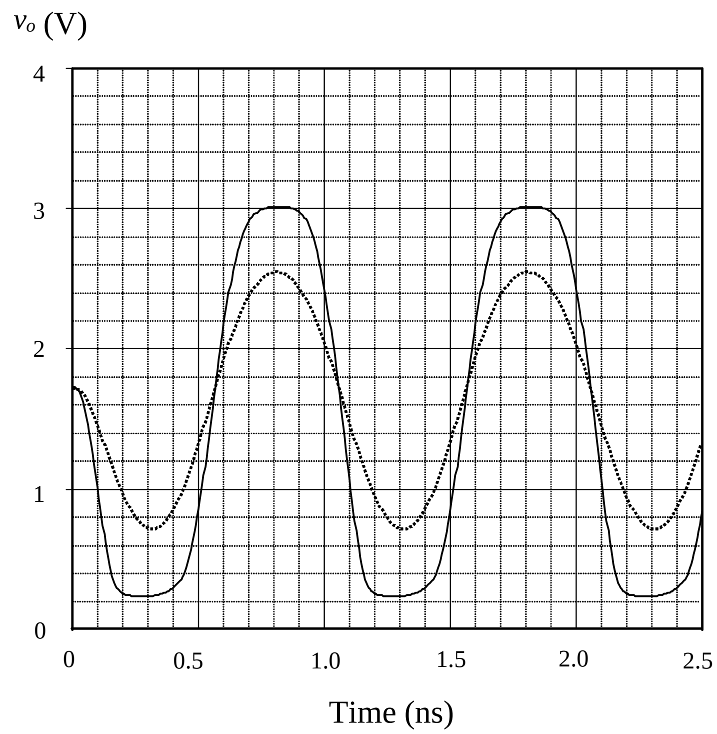


Figure 3.55 The output voltage of the amplifier at 1 GHz input frequency. The solid line corresponds to 500 mV input voltage amplitude and the dotted line to 200 mV amplitude.

These simulation results can be interpreted as follows:

- The delay characteristic corresponding to L = 16.6 nH shown in Fig.3.53(b) (the solid line) is flat up to 2GHz within 1.4% and up to 3 GHz within 5.8% [5]. The fine tuning of the inductance value (from the calculated 17.76 nH to 16.6 nH) indicates that the assumed value of the output parasitic capacitance of the transitor was somewhat higher than the assumed value. But the difference is small and therefore, iterative re-design is not considered to be necessary.

- The 3 dB frequency of the gain characteristic corresponding to minimum delay distortion (Fig.3.53(a), the solid line) fulfills the 3GHz bandwidth requirement.

- The 3 dB frequency corresponding to the flat gain characteristic is even higher than expected (3.38 GHz), at the expense of excessive signal delay at high frequencies, affecting the transient response.

- The transient responses corresponding to the flat delay characteristic and to the flat gain characteristic are given in Fig.3.54. The response corresponding to the flat delay characteristic settles to the final value without overshoot, with a rise time of 110.2 ps. The response corresponding to the flat gain characteristic has 5% overshoot, which can be tolerated for many applications. The rise time for this case is 94.8 ps, which may make it preferable.

- Since the inductance values are high, the quality factors for on-chip realization will be considerably low and series resistances high. For example, the series resistance of a L =16.6 nH, Q = 5 inductance at 1.85 GHz is 38.6 ohm (See Chapter 4, Section 4.1.1.1). Therefore, the value of R_L must be reduced to 535 ohm.

- In Fig. 3.55 the responses to 1 GHz, 500 mV amplitude and 200 mV amplitude sinusoidal input signals are shown. The symmetrical clipping of the output waveform corresponding to 500 mV input signal indicates that the position of the operating point on the output characteristic curves (the value of the drain quiescent current) is appropriate for small nonlinear distortion.

- If the power consumption has prime importance, a drain quiescent current smaller than the value given with (3.93) can be used, at the expense of reducing the output voltage dynamic range and increasing the output capacitance due to the increase of W.

- For higher signal source resistance values, it may be necessary to take into account the effect of the R_sC_{gs} low-pass section.

- The lower 3 dB frequency of the amplifier with C_c = 10 pF and R_G = 100 kohm is 160 kHz, which may be unnecessarily low for many applications. To increase the value of the lower 3 dB frequency a solution is reduce the value of C_c. But it must not be overlooked that C_c and C_{gs} form a capacitive, frequency independent voltage divider and reduce the signal reaching the gate of the transistor, and consequently, the voltage gain decreases in the whole band.

---

[5] Note that to eliminate the 180º basic phase shift of the amplifier, the signal delay is calculated as $(180 - \varphi) / (2\pi f)$.

A final remark: After the layout and the post lay-out simulations, it may be necessary to perform a second fine-tuning, due to the layout related parasitics.

Problem 3.4

Derive the expressions for the series-peaked wide band amplifier shown in Fig. 3.47(c)

Problem 3.5

(a) Design a series-peaked amplifier fulfilling the requirements given in Example 3.4.

(b) Compare the results corresponding to the parallel and the series compensated amplifiers, and discuss.

4.6.3. Noise in Amplifiers

At this point, it is worthwhile to examine the notion of noise, and modeling of noise in fundamental circuits. Amplifiers are being used to enhance the signal that is generated by a weak signal source; for example the signal generated by a microphone (that is proportional to the sound pressure), connected to the input of an audio amplifier, or the signal at the output of an antenna applied to the input of a receiver, that is proportional with the electromagnetic field strength at the tuning frequency of the receiver. In addition to the original signal to be amplified, there is usually a physical phenomenon that disturbs the amplified signal in a certain way: for example the high-frequency "hiss" that we hear from the loudspeaker connected to the output of the audio amplifier when the input signal
level is low, or the random speckles we see on the screen of a TV receiver when we receive a distant transmitter. This phenomenon is called "noise" that is generated in all kinds of resistors and devices like diodes and transistors, due to the random motions of electrons. In the scope of this book we will concentrate on the noise of LNAs. But it must be kept in mind that the physical mechanisms and the basic definitions are common for all kinds of amplifiers.

LNAs are being used to receive and amplify weak signals transmitted by distant transmitters, which may be fixed or mobile, or installed on board of a satellite. The total signal received at the input of the amplifier does not only consist of the signal sent by the transmitter[1], but in addition, it includes the unavoidable noise signal originating from the internal resistance of the antenna. To obtain a sufficiently high level of signal power with a reasonable signal-to-noise ratio ($S\ /\ N$) at the output of the LNA, the noise inherently generated in the amplifier must be kept as low as possible.

The noise performance of an amplifier is usually expressed by the "noise factor", F, that is defined as

$$F = \frac{(S/N)_{input}}{(S/N)_{output}} = \frac{S_{in}/N_{in}}{S_{out}/N_{out}} \tag{4.79}$$

The LNA amplifies equally the incoming signal and the input noise generated in the signal source. For a noiseless amplifier the output signal power and the output noise power are equal to gain multiplied by the input signal and the input noise power, respectively. Therefore the noise factor of a noiseless amplifier is equal to unity. In a noisy amplifier, on the other hand, the output noise is the sum of the gain times the input noise and the output noise component representing the noise generated in the amplifier. From this consideration and (4.79), the noise factor can be written as

$$F = \frac{S_{in}}{S_{out}} \frac{N_{out}}{N_{in}} = \frac{S_{in}}{A_p \times S_{in}} \frac{(A_p \times N_{in}) + N_{amp}}{N_{in}} \tag{4.80}$$

where;

- A_p is the power gain of the amplifier,

---

[1] In addition to the signal of interest sent by the transmitter, certain natural (atmospheric, cosmic, etc.) and man-made (originating from switching of power lines, corona discharges, etc.) noises can reach to the input of the amplifier. Since these "external" noises are all sporadic and in some cases avoidable to some extent, they will be kept out of this discussion.

• N_{amp} is the noise power at the output, originating only from the amplifier itself, (i.e. the noise of the resistance of the signal source excluded).

Now (4.80) can be simplified as

$$F = 1 + \frac{N_{amp}}{A_p \times N_{in}} \tag{4.81}$$

The "noise figure", *NF* which is also used to express the noise performance of an amplifier, is the noise factor expressed in logarithmic scale:

$$NF(dB) = 10 \times \log F \tag{4.82}$$

The noise generated in an amplifier mainly originates from the random movements of charge carriers in resistors and devices, due to their thermal energy. Apart from this "thermal noise", there are several other types of noise; the partition noise, the multiplication noise, the flicker (or *1/f*) noise, etc. But since these other noise components are usually effective at lower frequencies and/or dominated by the thermal noise, they can be ignored for LNAs.

The thermal noise was first observed and measured by J.B. Johnson in 1928 and interpreted by H. Nyquist who derived an expression giving the value of the noise power in a conductor due to the random movements of electrons;

$$P_n = 4kTB \tag{4.83}$$

where k is the Boltzman constant (1.36×10^{-23} joules/K), T is the temperature of the conductor in K, and B is the bandwidth of interest. B can be placed anywhere on the frequency axis, indicating the "white" (frequency independent) character of the noise. Expression (4.83) shows that the noise power is same for all conductors for a certain temperature and for a bandwidth of interest, *regardless of the material and shape of the conductor*.

It must not be overlooked that the noise voltages and currents of the resistors and transistors in a circuit are processed (amplified, fed-back, mixed, etc) as other signals in the circuit, until they reach the output of the circuit. Hence, their contributions to the total noise power at the output must be calculated in terms of the mean square values and combined additively to find the total noise power.

4.6.3.1. Thermal Noise of a Resistor

If the electrical resistance of a conductor is R, the noise power can be expressed in terms of the mean square noise voltage between the terminals of the resistor, or the mean square noise current flowing through the resistor as

$$P_n = \overline{i_n^2} \times R \qquad \text{or} \qquad P_n = \overline{v_n^2} \times \frac{1}{R}$$

Consequently, the root mean square (effective) values of the noise current and the noise voltage can be written as

$$\overline{i}_n = \sqrt{4kTB\frac{1}{R}} \quad \text{and} \quad \overline{v}_n = \sqrt{4kTB \times R} \qquad (4.84)$$

and the noise equivalent circuit of a resistor can be drawn as seen in Fig.4.40.

To ease the comparison of noise behavior of different devices (or circuits) and to conform with the existing noise measurement systems that usually measure the noise in a narrow band, it is common practice to express the noise for 1Hz bandwidth. Hence, the mean square noise voltage and the mean square noise current for 1Hz bandwidth can be written as

$$\overline{v}_n^2\Big|_{B=1Hz} = 4kT \times R = S_v \quad \text{and} \quad \overline{i}_n^2\Big|_{B=1Hz} = 4kT / R = S_i \qquad (4.85)$$

called the "spectral density" of the mean square noise voltage and noise current, respectively.

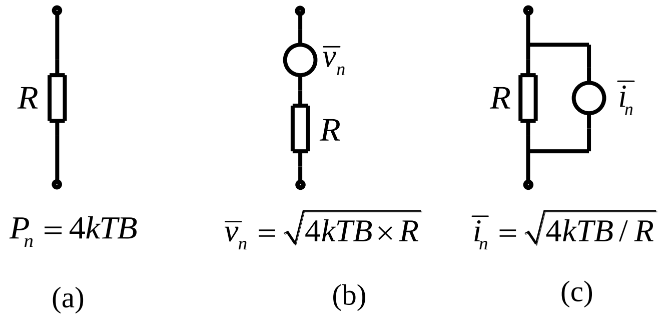


Figure 4.40. (a) A resistor and the associated noise power. The noise equivalent circuit with the noise voltage source (b) , and with the noise current source (c).

Problem 4.9

a) *Calculate the thermal noise voltage and thermal noise current of a 50 ohm resistance (a) for 30 ˚C, (b) for 100 ˚C (the bandwidth of interest is 10 MHz).*

b) *Calculate the thermal noise voltage and thermal noise current of a 1000 ohm resistance for the same temperature and the same bandwidth.*

c) *Compare and discuss the results.*

4.6.3.2. Thermal Noise of a MOS Transistor

The thermal noise of a MOS transistor was first investigated and modeled by A. van der Ziel in 1986 [4.10]. In the physical structure of a MOS transistor there are several "resistances" as shown in Fig.4.41, all of which generate noise according to (4.84).

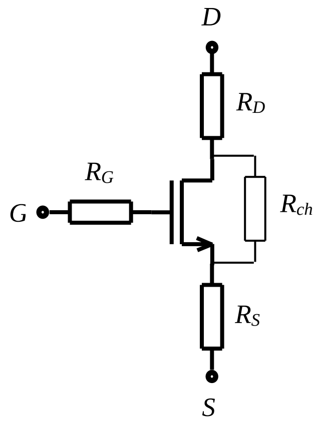


Figure 4.41. The MOS transistor with its noise generating resistances.

The total resistance between the external source node (*S* in Fig. 4.41) and the external drain node (*D*) of the transistor is the sum of the source series resistance (R_S), the channel resistance (R_{ch}) and the drain series resistance (R_D). The source series resistance (and similarly the drain resistance) is the sum of the intrinsic and extrinsic components and the equivalent contact resistance, as explained in Chapter 1. The source and drain series resistances are obviously technology- and geometry-dependent and can be more than one hundred ohms for small transistors and several ohms for large transistors. The gate series resistance, another noise source, is also technology- and geometry-dependent and can be minimized with appropriate finger structures [4.11].

It is known that the strongly dominant part of the MOS transistor noise is the thermal noise associated with the channel resistance [4.12]. In several publications this noise is investigated as the sum of the noises of the pre-pinch-off region of the channel and that of the pinched-off region [4.13], [4.14], [4.15]. On the other hand, it has been shown that the effect of V_{DS}, and consequently, the contribution of the pinched-off region on the drain current noise is negligible [4.16], [4.17].

In this section, the noise associated with the channel resistance of a MOS transistor operating in the saturation region will be derived with a similar but more straightforward approach based on the expression derived in Chapter 1 for the inversion charge.

The cross section of a non-velocity-saturated NMOS transistor in the saturation region is shown in Fig.4.42. According to (1.7), the effective gate voltage (and the inversion charge that is proportional to the effective gate voltage) varies with the square root of *y*, until the pinch-off point, *L'*, that corresponds to a channel voltage of $(V_{GS} - V_T)$. Along the saturation region (from *L'* to *L*) electrons travel with the saturation velocity, v_{sat}, and the inversion charge density is constant (see Fig. 1.6). It is obvious that these two sections of the channel resistance shown with r_1 and r_2 are different in nature and as mentioned above, from the point of view of noise, the strongly dominant part of the channel resistance is r_1. The value of r_1 corresponding to R_{ch} can be calculated based on the modified gradual channel approach developed in Chapter 1. The resistance of a channel element *dy* before the pinch off point can be written as

$$dr(y) = \frac{dV_c}{I_D} \qquad (4.86)$$

and the drain current, that is constant along the channel is

$$I_D = \frac{d\bar{Q}_i(y)}{dt} \tag{4.87}$$

where $d\bar{Q}_i(y)$ is the inversion charge in *dy* channel element that was calculated in (1.37) as

$$d\bar{Q}_i(y) = -C_{ox}W(V_{GS} - V_T)\sqrt{1 - \frac{y}{L'}} \ . \ dy$$

dt in (4.87) can be written as

$$dt = \frac{dy}{v} = \frac{dy}{\mu E(y)} = \frac{dy}{-\mu(dV_c/dy)} \tag{4.88}$$

From (4.86), (4.87) and (1.37) the resistance of a channel element *dy* can be calculated:

$$dr(y) = \frac{dy}{\mu C_{ox}W(V_{GS} - V_T)\sqrt{1 - \frac{y}{L'}}} \tag{4.89}$$

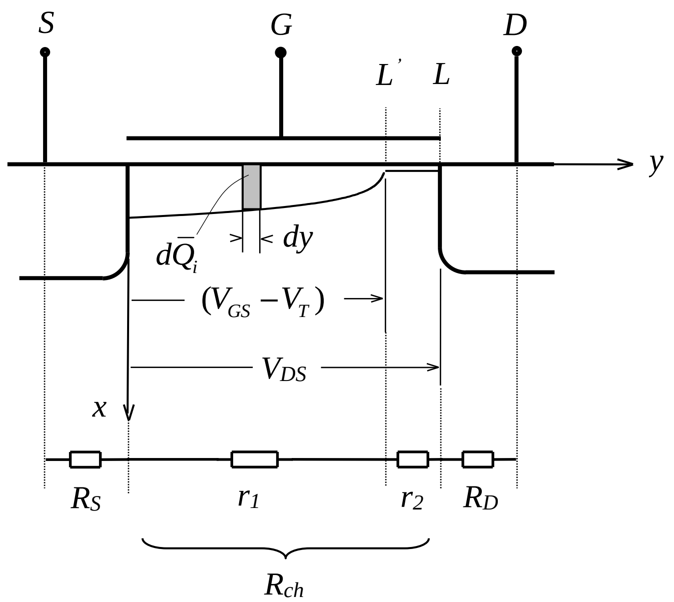


Figure 4.42. The cross section of a MOS transistor in saturation and components of the series resistances of the drain current path.

The integral of *dr(y)* from *y* = 0 to *y* = *L'* gives the value of the first (pre-pinch-off) section of the channel resistance[2]:

$$R_{ch} = \frac{2}{\mu C_{ox}\frac{W}{L'}(V_{GS} - V_T)} \tag{4.90}$$

where *L'* must be considered as

---

[2] Note that this is equal to $2/g_{do}$,where g_{do} is the output conductance corresponding to $V_{DS} \to 0$, in the original noise expressions of van der Ziel [4.10].

$$L' = L\frac{1+\lambda(V_{GS}-V_T)}{1+\lambda V_{DS}} \tag{4.91}$$

according to (1.14-a). It must be also noted that μ in (4.90) is a function of V_{GS} as mentioned in 1.1.2.1. and Appendix-A.

(4.90) can be arranged as

$$R_{ch} = \frac{(V_{GS}-V_T)}{I_D} \tag{4.92}$$

that is – not surprisingly – the DC resistance of the inversion region.

If we insert (V_{GS} – V_T) in terms of I_D, (4.92) can be rewritten as

$$R_{ch} = \sqrt{\frac{2}{I_D \mu\, C_{ox}\frac{W}{L}}} \tag{4.93}$$

or in terms of g_m ,

$$R_{ch} = \frac{2}{g_m} \tag{4.94}$$

These expressions can be interpreted as follows:

- R_{ch} decreases with the square root of the drain current, I_D.
- R_{ch} decreases with the square root of the aspect ratio.
- R_{ch} decreases with the square root of the mobility. The bias dependence of the mobility must not be ignored for small geometry devices.
- Since the mobility of holes is considerably smaller than the mobility of electrons, the channel inversion resistance of a PMOS transistor is higher than that of a NMOS transistor having the same geometry and the same drain current.
- R_{ch} decreases with the square root of C_{ox}. It means that the channel inversion resistance is inferior for a smaller geometry transistor having the same aspect ratio.

Now the noise current corresponding to the channel resistance of a MOS transistor can be written as

$$\overline{i}^2_{n-ch} = \frac{4kTB}{R_{ch}} = 4kTB\times\frac{1}{2}g_m \quad [\mathrm{A}^2] \tag{4.95}$$

It has been experimentally found and widely accepted that it is more realistic to modify this expression as:

$$\overline{i}^2_{n-ch} = 4kTB\times\gamma\times g_m \tag{4.95-a}$$

where γ ranges from 0.5 to 1.5, depending on the depth of saturation and the channel geometry, being 0.5 for large (W/L) ratios and higher for smaller geometries.

The interpretation of (4.96) for a non-velocity saturated transistor with

$$g_m = \sqrt{2\mu C_{ox}\frac{W}{L}I_D}$$

$$(1.33)$$

shows us that the channel current noise:

- increases with the drain DC current,
- increases with (W/L) ratio,
- increases for larger geometry devices due to higher C_{ox} values,
- related to the carrier mobility: lower for a PMOS transistor than that of a NMOS transistor having the same geometry and the same drain current.

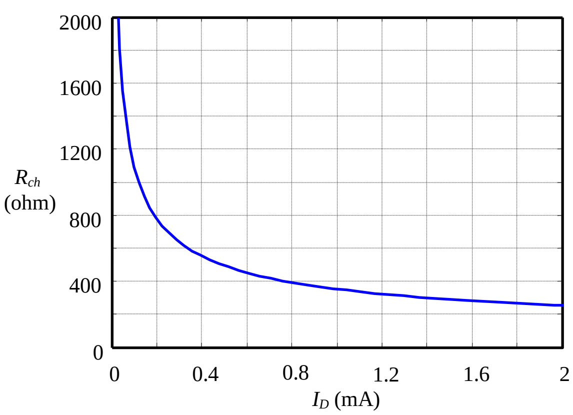
Similarly, for a velocity saturated transistor with

$$g_m = kWC_{ox}v_{sat} \tag{1.34}$$

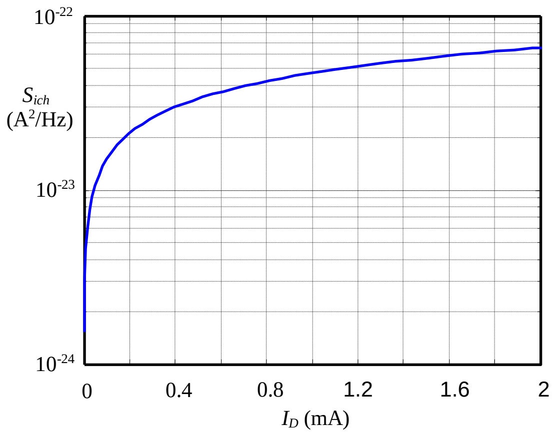
the channel current noise:

- does not depend on the drain DC current,
- increases with W,
- increases for smaller geometry devices due to higher C_{ox} values,
- almost equal for NMOS and PMOS transistors, since the velocity saturation values of holes and electrons are approximately same.

As an example, the variation of R_{ch} and S_{ich} for a AMS 035 (W = 35 µm, L = 0.35 µm) transistor as a function of the drain current are plotted in Fig. 4.43. Figure 4.43(a) shows that the inversion channel resistance acquires considerably low values for moderate to high drain currents. Therefore if the series source and drain resistances are not sufficiently small compared to R_{ch}, they must not be ignored.



(a)



(b)

Figure 4.43. (a) The inversion channel resistance and (b) the drain noise current spectral density of a 35µm / 0.35µm AMS transistor. (calculated with $\gamma = 0.5$, and the V_{GS} dependence of μ is taken into account.)

Since the noise contributions of the parasitic internal resistances (R_S, R_D and R_G) depend not only on the transistor but also on the circuit, they must be investigated for different circuit configurations separately. R_D can be considered as part of the the effective load resistance. To gain further insight on the contribution mechanisms of R_S and R_G, we will calculate their effects for the most basic common source amplifier.

The source series resistance has two effects on the noise behavior of the transistor:

(a) The noise voltage, $\overline{v}_{nRS1}$, generated by the source series resistance of the transistor (the thermal noise of R_S). This voltage, that is shown in Fig. 4.44(a) adds a component to the drain noise current equal to $\overline{i}_{ndS1} = y_{mSD}.\overline{v}_{nRS1}$, where y_{mSD} is the transadmittance from $\overline{v}_{nRS1}$ to the drain current:

$$y_{mSD} = -\frac{g_m}{R_S(g_m + j\omega C_{gs}) + (1 + j\omega C_{gs} R_G)}$$

$$= -\frac{g_m}{g_m R_S(1 + j\dfrac{\omega}{\omega_T}) + (1 + j\omega C_{gs} R_G)}$$

that is real and equal to

$$g_{mSD} \cong -\frac{g_m}{g_m R_S + 1} = -g_{m(eff)} \tag{4.96}$$

provided that $\omega \ll \omega_T$ and $\omega \ll 1/C_{gs}R_G$. The drain noise current component originating from the noise of R_S becomes

$$\overline{i}_{ndS1} = -\overline{v}_{nRS1} g_{m(eff)}$$

This noise component is totally uncorelated to the noise originating from the channel resistance and therefore must be added to the mean square channel noise as

$$\overline{i}^2_{ndS1} = \overline{v}^2_{nRS1} \cdot g^2_{m(eff)} = 4kTBR_S \cdot g^2_{m(eff)} \tag{4.97}$$

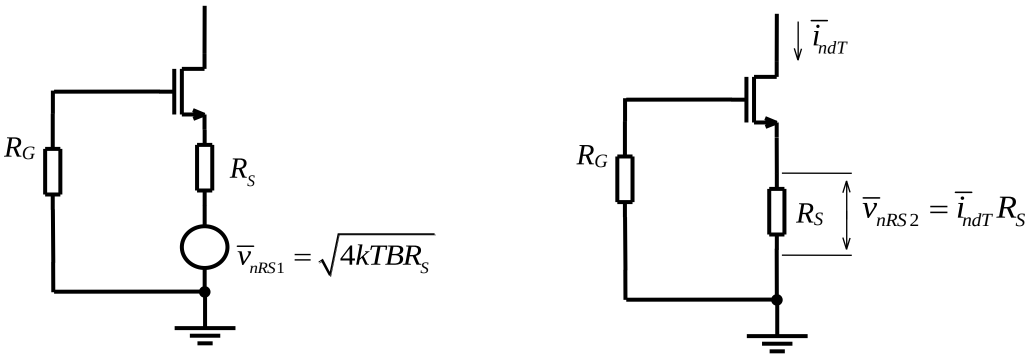
and the total mean square noise current (with the thermal noise of R_S included) becomes

$$\overline{i}^2_{nd} = \overline{i}^2_{n-ch} + \overline{i}^2_{ndS1} = 4kTB\left[\gamma g_m + R_S g^2_{m(eff)}\right] \tag{4.98}$$

(b) The additional effect of the noise voltage drop on R_S due to the noise current flowing through R_S, which is $\overline{v}_{nRS2} = \overline{i}_{nd} R_S$. This noise voltage provokes a drain noise current equal to

$$\overline{i}_{ndS2} = y'_{mSD}.\overline{i}_{nd} R_S \tag{4.99}$$

where y'_{mSD} is the transadmittance from $\overline{v}_{nRS2}$ to the drain current.



(a) (b)

Figure 4.44. (a) The noise voltage of the source series resistance. (b) The noise voltage due to the total drain noise current flowing through R_S .

$$y'_{mSD} = -\frac{g_m}{(1 + j\omega C_{gs} R_G)} \tag{4.100}$$

that is

$$y'_{mSD} \cong -g_m \tag{4.100-a}$$

provided that $\omega \ll 1/C_{gs}R_G$. From (4.99) and (4.100-a) the noise current component originating from the noise voltage drop on R_S becomes

$$\overline{i}_{ndS2} = -g_m . \overline{i}_{ndT} R_S \tag{4.101}$$

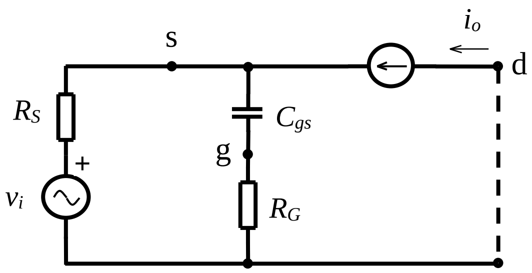


Figure 4.45. The small signal equivalent circuit used to calculate y_{mSD} .

that is apparantly related to the noise current originating from the channel resistance and is out of phase with it. Therefore, it reduces the channel noise current. This fact leads to the “noise cancelling feedback” concept that will be investigated in Section 4.4.8 with more detail. The channel noise current (with this feedback effect included) can be calculated from Fig. 4.46 as

$$\overline{i}_{ndT} = \overline{i}_{nd} \frac{1}{(1 + g_m R_S) + \dfrac{R_D + R_S}{r_{ds}}} \cong \overline{i}_{nd} \frac{1}{(1 + g_m R_S)} \tag{4.102}$$

and its mean square value is:

$$\overline{i}^2_{ndT} \cong \overline{i}^2_{nd} \frac{1}{(1 + g_m R_S)^2} \tag{4.102-a}$$

It must be noted that, if there is any other noise components not corelated with the channel noise, it must be added to $\overline{i}_{nd}^2$ before the inclusion of the feedback effect.

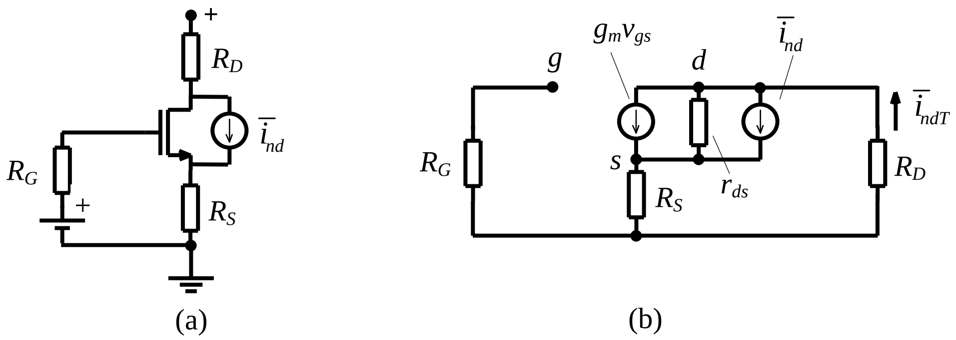


Figure 4.46. (a) The circuit and the channel noise current of the transistor, the thermal noise of R_S included. (b) The equivalent circuit to calculate the feedback modified value of the noise current.

Another important parasitic resistance of the MOS transistor is the gate resistance, whose value is strongly dependent on layout and process parameters.

The gate resistance generates a thermal noise voltage equal to $\overline{v}_{nRG1} = \sqrt{4kTBR_G}$ that is in the input loop of the transistor and provokes a drain noise current component equal to

$$\overline{i}_{ndG1} = \left| g_{m(eff)} \right| \overline{v}_{nR_G} \qquad (4.103)$$

where $g_{m(eff)}$ is the effective transconductance defined with (2.13-b). If there is an impedance (Z_S) connected in series to R_S, (2.13-b) must be modified as $g_{m(eff)} = 1/\left[(1+g_m(R_S+Z_S)\right]$.

This is a thermal (white) noise and is not correlated to the channel noise, therefore must be added with the mean square value to the channel noise. Together with this component, (4.98) becomes;

$$\overline{i}_{nd}^2 = \overline{i}_{n-ch}^2 + \overline{i}_{ndS1}^2 + \overline{i}_{ndG1}^2 = 4kTB\left[\gamma g_m + g_{m(eff)}^2 (R_S + R_G)\right] \qquad (4.98\text{-a})$$

It must be noted that if there is an external resistance connected in series to the gate, it must be considered together with the inherent gate resistance of the transistor.

There is another noise component related to the gate resistance. At high frequencies a noise current due to the noise voltage on the inversion channel flows over the gate capacitance. This noise current ($\overline{i}_{ng}$) is not "white" since it increases with frequency, and is correlated to the inversion channel noise. The noise voltage drop on R_G due to this current, $\overline{v}_{nRG2} = \overline{i}_{ng} R_G$ is another noise voltage source in the input loop and adds another component on to the drain noise current (see Fig. 4.47-b).:

$$\overline{i}_{ndG2} = g_{m(eff)} \overline{v}_{nRG2} = g_{m(eff)} \overline{i}_{ng} R_G \qquad (4.104)$$

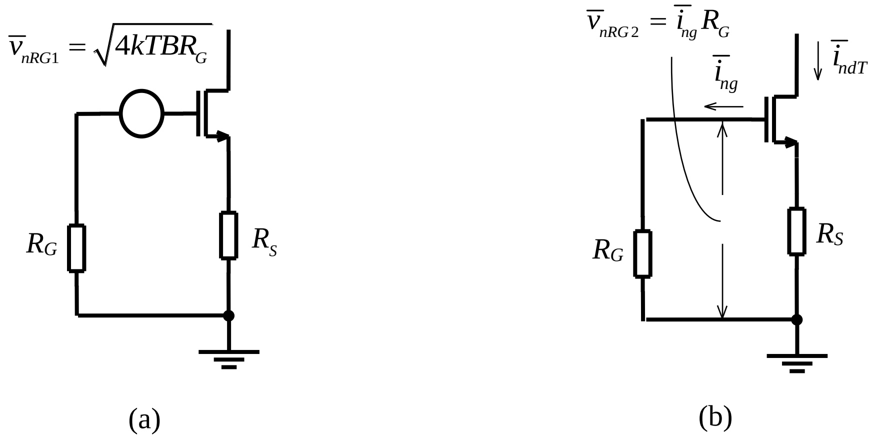


Figure 4.47. (a) The noise voltage source due to the gate series resistance.
(b) The input noise voltage source to the gate noise current.

The approach used to calculate the gate noise current ($\bar{i}_{ng}$) is shown in Fig.4.48. The noise voltage of a channel element *dy* at position *y* is

$$\bar{v}_n(y) = \bar{i}_{ndT} r(y)$$

where $\bar{i}_{ndT}$ is the channel noise current and *r*(*y*) is the resistance of the channel segment from the source end of the channel (*y* = 0) to *y*, that is equal to the integral of *dr* given with (4.89), from zero to *y*:

$$r(y) = \frac{2}{\mu C_{ox} \frac{W}{L'}(V_{GS} - V_T)} \left(1 - \sqrt{1 - \frac{y}{L'}} \right)$$

$\bar{v}_n(y)$ induces an incremental noise current over the incremental capacitance $dC_g = C_{ox} W dy$, that is

$$d\bar{i}_{ng} = j\omega(dC_g) \times \bar{v}_n(y) = j\omega(C_{ox} W dy) \times r(y) \bar{i}_{ndT}$$

and hence

$$\bar{i}_{ng} = j\omega C_{ox} W \frac{2}{\mu C_{ox} \frac{W}{L'}(V_{GS} - V_T)} \int_0^{L'} \left(1 - \sqrt{1 - \frac{y}{L'}} \right) dy \times \bar{i}_{ndT} \qquad (4.105)$$

It must be noted that $\bar{i}_{ng}$ is related to the channel noise current, but there is a 90° phase difference. Therefore, it does not contribute to any noise cancelling feedback.

For $L \cong L'$, and from (1.33) and (1.42) this expression can be arranged as[3]

$$\left| \bar{i}_{ng} \right| \cong \omega \frac{C_{gs}}{g_m} \times \bar{i}_{ndT} \cong \frac{\omega}{\omega_T} \times \bar{i}_{ndT} \qquad (4.106)$$

[3] It must be noted that this derivation is valid provided that $(1/\omega C_{gs}) >> r_d$, in other words the noise current deviated to the gate is sufficiently smaller than the drain noise current.

Hence $\overline{i}_{ndG2}$ becomes from (4.104)

$$\overline{i}_{ndG2} = g_{m(eff)}\overline{v}_{nRG2} = g_{m(eff)} \times \left(\omega \frac{C_{gs}}{g_m} \times \overline{i}_{ndT} \right) R_G \qquad (4.107)$$

and its mean square value, that must be added to the channel noise;

$$\overline{i}^2_{ndG2} = \left(\frac{\omega C_{gs} R_G}{1 + g_m R_S} \right)^2 \times \overline{i}^2_{ndT} \qquad (4.108)$$

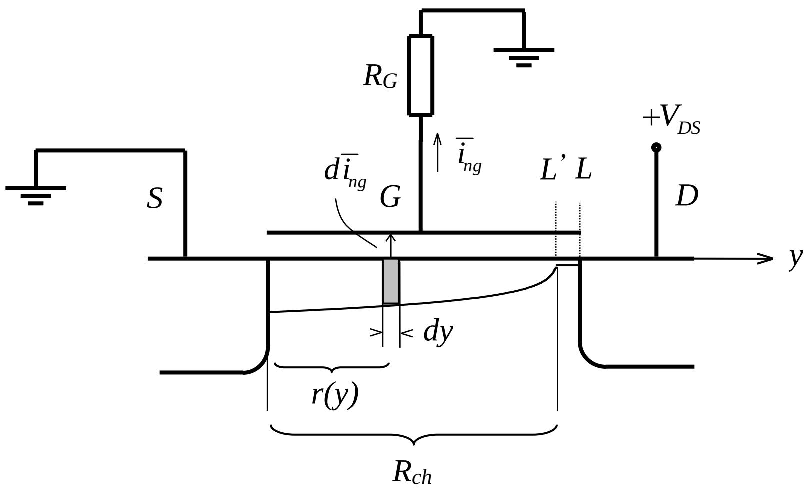


Figure 4.48. The gate noise current as the sum of the incremental components.

(4.108) can be interpreted as follows:

The drain current noise component associated with the gate noise current;
- increases with frequency,
- increases with C_{gs},
- is proportional to the total channel noise,
- is a capacitive current and has a 90º phase difference with the channel noise current, and therefore does not contribute to any noise cancelling feedback.
- must be added to the channel noise with its mean square value.

It can be seen that this noise component is usually very small and negligible except for very small oxide thicknesses and very high frequencies. Therefore, the actual output noise on the drain current (including noise feedback over R_S) can be calculated from (4.98-a) and (4.102-a) as

$$\overline{i}^2_{ndT} \cong \overline{i}^2_{nd} \frac{1}{(1 + g_m R_S)^2}$$

and then $\overline{i}^2_{ndG2}$ can be added if it is not negligibly small.

Example 4.9

In the following, the contribution of different components of the drain noise current will be calculated for a 35µm (7×5µm) / 0.35µm AMS NMOS transistor at f = 3 GHz for 10MHz bandwidth. The DC operating point of the transistor is I_D = 2 mA, V_{DS} = 3V. The temperature of the transistor will be assumed as 300 K.

The parasitic source and drain series resistances of a 5µm / 0.35µm transistor were calculated as 134 ohm each, in Example 1.3. The 35 µm width transistor is composed of seven parallel 5 µm channel regions in the form of a multi-finger transistor as shown in Fig. 1.23. Since source and drain regions are shared by the neighboring channel region and since they are connected in parallel, the source resistance of the 35 µm transistor is approximately 134/14≈10 ohm.

Since the poly gate sheet resistance for this technology is given as R_{sh} = 7 ohm/□, the resistance of one of the gate stripes is 7×(5 / 0.35) = 100 ohm. If the gate stripes are connected in parallel as shown in Fig. 1.23, the equivalent resistance is 100/7 = 14.3 ohm. Poly to metal contact resistance is given as 2 ohm / (0.4×0.4) micron contact. Assuming 10 contacts along the collecting stripe, the equivalent contact resistance is 0.2 ohm. Together with the resistance of the collecting stripe, the total resistance series to the gate can be taken as 15 ohm.

The gate-to-source voltage for 2 mA drain current and the mobility corresponding to this voltage can be found as 1V and 325 cm^2/ V.s, respectively. The transconductance of the transistor for 2 mA drain current is

$$g_m = \sqrt{2\times 325\times(4.54\cdot 10^{-7})\times(2\cdot 10^{-3})} = 7.53\cdot 10^{-3}\ \square\ 7.5\ \text{mS}$$

and the effective transconductance with R_S =10 ohm,

$$g_{m(eff)} = \frac{7.5\cdot 10^{-3}}{1+7.5\cdot 10^{-3}\times 10} = 6.98\cdot 10^{-3} \cong 7\ \text{mS}$$

With

$$4kTB = 4\times(1.38\times 10^{-23})\times 300\times(10\times 10^{6}) = 1.656\times 10^{-13}\quad [\text{W}]$$

the inversion channel noise of the transistor from (4.95-a):

$$\overline{i}^{2}_{n-ch} = 4kTB\gamma g_m = (1.656\times 10^{-13})\times 0.5\times(7.5\times 10^{-3}) = 6.21\times 10^{-16}\quad [\text{A}^2]$$

The drain current noise component related to the thermal noise of R_S from (4.97):

$$\overline{i}^{2}_{ndS1} = 4kTBR_S g^{2}_{m(eff)} = (1.656\times 10^{-13})\times 10\times(7\times 10^{-3})^2 = 0.81\times 10^{-16}\quad [\text{A}^2]$$

The drain current noise component related to the thermal noise of R_G from (4.103):

$$\overline{i}^{2}_{ndG1} = g^{2}_{m(eff)}\overline{v}^{2}_{nR_G} = g^{2}_{m(eff)}\times 4kTBR_G = (7\times 10^{-3})^2\times(1.656\times 10^{-13})\times 15 = 1.217\times 10^{-16}\quad [\text{A}^2]$$

The sum of all these componets according to (4.98-a)

$$\overline{i}_{nd}^2 = \overline{i}_{n-ch}^2 + \overline{i}_{ndS1}^2 + \overline{i}_{ndG1}^2 = 6.21\times10^{-16} + 0.81\times10^{-16} + 1.217\times10^{-16} = 8.237\times10^{-16} \quad [\mathrm{A}]$$

The drain mean square noise current after noise feedback over R_S from (4.102-a):

$$\overline{i}_{ndT}^2 \cong \overline{i}_{nd}^2 \frac{1}{(1+g_m R_S)^2} = (8.237\times10^{-16})\frac{1}{\left[1+(7.5\times10^{-3})\times10\right]^2} = 7.125\times10^{-16} \quad \left[\mathrm{A}^2\right]$$

Now the noise component originating from the capacitive gate noise current must be calculated from (4.108). Using $A = 2.47\times10^{11}$ calculated for this transistor,

$$\overline{i}_{ndG2}^2 = \left(\frac{(2\pi\times3\times10^9)(39\times10^{-15})\times15}{1+(7.5\times10^{-3})\times10}\right)^2 \times \overline{i}_{ndT}^2 = 0.0001\times\overline{i}_{ndT}^2$$

it can be seen that is very small compared to $\overline{i}_{ndT}^2$.

This example shows that the series resistances of the gate and source electrodes considerably affect the total mean square drain noise. Therefore, utmost care is necessary on the layout of the device to keep the external series resistances as low as possible

Another aspect that must not to be overlooked is the “temperature”. The “*T”* in the noise expressions is the temperature of the resistance, in the case of the MOS transistor the temperature of the channel region. This can be considerably higher than average surface temperature of the die[4] and certainly higher than the ambient temperature. This fact imposes higher noise for higher power densities, consequently higher device temperature. For example, at 400 K the mean square noise currents increase approximately 33%.

Another aspect that is important for the calculation of the noise factor is the difference of the temperature of the chip and the external resistive components. An example for this could be the resistance of the signal source or the resistance of the antenna.

Problem 4.9.

Derive an expression to calculate approximately the gate noise current assuming that the noise voltage along the channel is constant and equal to the noise voltage at the mid-point of the channel. Compare the result with (4.100) and discuss.

Problem 4.10.

[4] From the publications related to the thermal simulation and mapping of ICs [7.18], [7.19] it can be seen that the temperatures of directly heated micro-regions (hot spots) can be considerably higher than the average temperature of the die.

The die area of an integrated circuit is 10 mm^2 and its thickness is 0.5 mm. The die is mounted into an Amkor MLF, 44 lead miniature package, whose thermal resistance from the ambient to the bottom of the die is 24˚C/W. The power consumption of the circuit is 1W and the ambient temperature is 30˚C. Calculate the average surface temperature of die (The specific thermal conductance of silicon is 1.5 W/cm.˚C).

Answer: 54.33˚C! (It is obvious that the temperature of the channel regions of the individual MOS transistors on the die are considerably higher than this value, depending on their power densities.)

4.6.4. The Source Degenerated Tuned LNA and its Noise

The circuit diagram of a typical tuned LNA is shown in Fig. 4.49. The main gain stage is a cascode circuit whose load is a parallel resonance circuit tuned to the operating frequency, ω_0. The inductor connected in series to the source (L_S) helps to obtain a low value resistive component for the input impedance as mentioned in Section 3.6. The gate series inductor is used to resonate all reactive components at ω_0 , and obtain a purely resistive input impedance.

The key properties to be fulfilled by this circuit are;

- Good input matching at the tuning frequency,
- Voltage gain equal to or higher then a target value,
- Noise figure as low as possible at the tuning frequency.
- Output swing as high as possible without excessive nonlinear distortion that determines the usable maximum input voltage (see Section 4.6.6).
- In addition, the inductors in the circuit must be realisable in sence terms of their value and the quality factor, for the specified fabrication technology.

Another important feature of a tuned amplifier is its bandwidth. However, since the bandwidth of the tuned amplifier is imposed by the quality factor values of the on-chip inductors, setting the bandwidth as a primary design spec usually leads to unrealizable constraints. The usual approach is to accept the bandwidth of the amplifier as imposed by the quality factors of the on-chip inductors (which is usually wider than anticipated), and then rely on additional input and/or output filtering to sharpen the frequency characteristic.

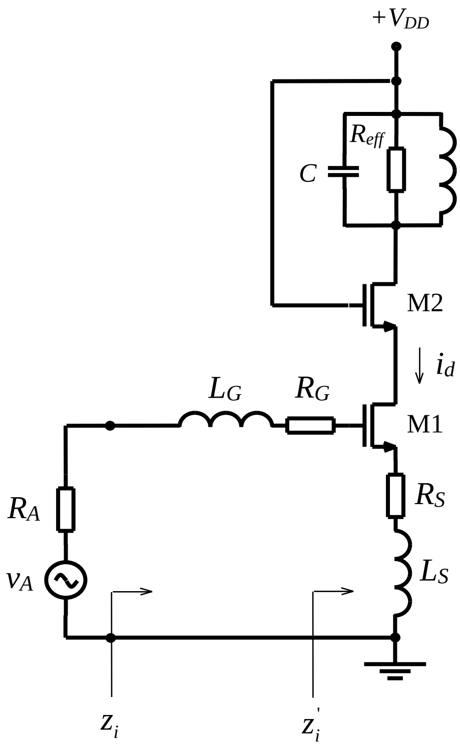


Figure 4.49. Most frequently used form of the source degenerated LNA.

Since each of these conditions mentioned above is related to the parameter values of several components in the circuit, all of these conditions influence each other. Therefore, an iterative design process becomes necessary. An analysis of the circuit leading to the calculation of the component values with some design hints is given below.

The most influential device parameter is the transadmittance, g_m of the input transistor, which determines the gain of the amplifier and the noise originating from this transistor. It must be noted that the voltage gain of the circuit has two factors; the gain from the v_{gs} of the input transistor to the output, i.e. the voltage gain of the cascode circuit, A_C, and the voltage transfer ratio from v_A to v_{gs} , that is a "series resonance gain" and equals to the value of the quality factor of the input loop (see 4.1.2.1.):

$$A_R = \frac{v_{gs}}{v_A} = Q_{in}$$

In case of the impedance matching at ω_0, since the real part of z_i is equal to R_A and ($L_G + L_S$) is in resonance with the input capacitance of M1, the quality factor of the input loop

is $$Q_{in} = \frac{(L_G + L_S)\omega_0}{2R_A}$$

that has a small value, usually slightly higher than unity.

The gain of the cascode circuit, A_C can be written as

$$|A_C| \cong g_{m1} \times R_{eff}$$

provided that $r_{i2} \ll r_{o1}$, where R_{eff} is the impedance of the load at resonance. The output resistance of the cascode circuit is very high and the quality factor of the total resonance capacitance is considerably higher than that of the on-chip inductor. Therefore

$$R_{eff} \cong L\omega_0 \times Q_L$$

Hence the total voltage gain becomes

$$|A_v| = A_R \times |A_C| \cong Q_{in} \times \left(g_{m1} \times R_{eff}\right)$$

Due to the high series resistance in the input loop, Q_{in} is very low and usually close to unity. Therefore the total gain can be assumed equal to A_C to start the design.

Since the realisable on-chip inductance values can not exceed a certain value and the quality factor depends on the technology, operating frequency and the value of the inductance, ,it is necessary to choose an appropriate L and Q_L pair to maximize A_C for a certain g_m . Using these L and Q_L values, g_m can be calculated corresponding to a target value for A_C as

$$g_{m1} = \frac{|A_C|}{L\omega_0 \times Q_L} \qquad (4.109)$$

This helps to calculate W and consequently C_{gs} parameters of the input.

For a non velocity saturated transistor from (1.33)

$$W = \frac{g_m^2 \times L}{2\mu C_{ox} I_D} \tag{4.110}$$

and for a velocity saturated transistor from (1.34)

$$W = \frac{g_m}{k C_{ox} v_{sat}} \tag{4.110-a}$$

Since g_m and corresponding W and C_{gs} are determined[1] (at least as a first approximation) now we can deal with the input matching:

From (3.54), the input impedance seen from the gate terminal of an inductive source degenerated transistor that is shown with z_i', can be written as

$$z_i' \cong \frac{1}{y_{in}} = \frac{1+(g_m + sC_{gs})(sL_S + R_S)}{sC_{gs}}$$

$$= \left(\frac{g_m L_S}{C_{gs}} + R_S\right) + sL_S + \frac{1+g_m R_S}{sC_{gs}}$$

and in frequency domain,

$$z_i' \cong \left(\frac{g_m L_S}{C_{gs}} + R_S\right) + j\left(\omega L_S - \frac{1+g_m R_S}{\omega C_{gs}}\right) \tag{4.111}$$

Note that $C_i' = C_{gs}/(1+g_m R_S)$ corresponds to the input capacitance of M1, together with R_S series its source. For $g_m R_S \ll 1$ -that is usually valid- C_i' is approximately equal to C_{gs}.

Now the input loop of M1 together with the signal source and the gate inductance can be drawn as shown in Fig. 4.50. In order to establish impedance matching at the operating frequency, ω_0 between R_A and its load; namely z_i , it is necessary

$$\left(\frac{g_m L_S}{C_{gs}} + R_S + R_G\right) = R_A \tag{4.112}$$

$$\left(\omega_0 (L_S + L_G) - \frac{1}{\omega_0 C_i'}\right) = 0 \tag{4.112-a}$$

[1] The calculated W values are usually very high. Therefore, multi-finger channel structure is necessary to achieve small parasitic resistances and small source and drain parasitic capacitances.

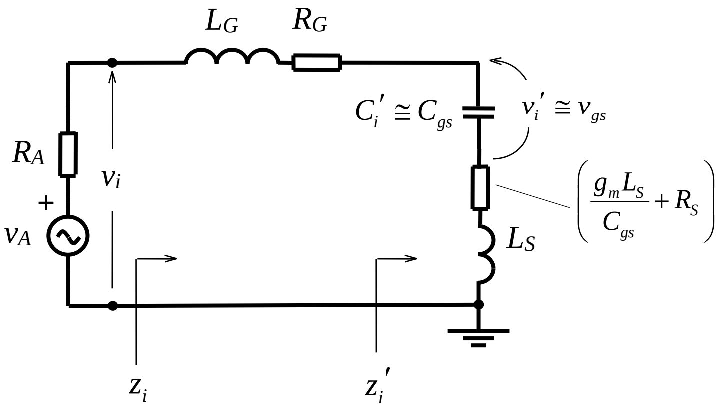


Figure 4.50. The input loop of the amplifier. R_S and R_G represent the equivalent series resistances of L_S and L_G corresponding the total losses at the operating frequency; in terms of the quality factors, $R_S = L_S\omega_0 / Q_{L_S}(\omega_0)$, $R_G = L_G\omega_0 / Q_{L_G}(\omega_0)$.

(4.112) gives the value of the of L_S as

$$L_S = \frac{C_{gs}}{g_m}\left[R_A - \left(R_S + R_G\right)\right] \tag{4.113}$$

Note that $R_G = R_{LG} + R_{Gi}$, $R_S = R_{LS} + R_{Si}$ where R_{Si} and R_{Gi} are the internal parasitic source and gate series resistances and $R_{LS} = L_S\omega_0 / Q$, $R_{LG} = L_G\omega_0 / Q$.

(4.112-a) indicates a series resonance between $C_i' \cong C_{gs}$ and $(L_S + L_G)$, that helps to calculate $(L_S + L_G)$ as

$$\left(L_G + L_S\right) = \frac{1}{\omega_0^2 C_i'} \cong \frac{1}{\omega_0^2 C_{gs}} \tag{4.114}$$

In many cases the calculated $(L_S + L_G)$ is very high, in the order of tens of nH, and L_S calculated from (4.113) is less than 1 nH. Therefore the dominant part of the resonance inductance, L_G is not realizable as an on-chip inductance. The conventional solution to reduce L_G to a realisable value, is to increase C_i' with an additional capacitor C_p connected between G and S terminals of the input transistor and turn back to (4.113), replace C_{gs} with $(C_{gs}+C_p)$ and continue. But related to this solution there is an interesting problem that affects the input matching.

This is related to the behavior of the C_{gs}, R_{Gi} , C_p combination at the input of M1. Note that C_p can not be connected directly parallel to the C_{gs} of the transistor, but before R_{Gi} as shown in Fig. 4.51 (a)[2]. The equivalent of this combination at ω_0 is shown in Fig. 4.51 (b) and its real and imaginary parts at ω_0 can be calculated in terms of C_{gs}, R_{Gi} and C_p as

[2] Since R_{Si} is very small compared to R_{Gi} its effect is ignored.

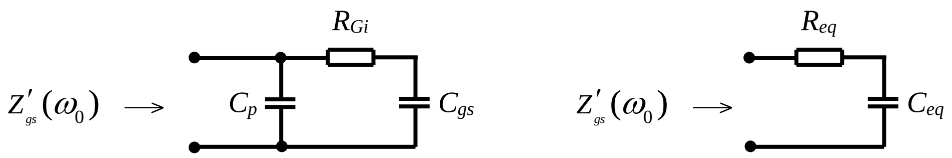
$$\mathrm{Re}\{Z'_{gs}\} = \frac{R_{Gi}C_{gs}^2}{\left(\omega_0 C_{gs}C_pR_{Gi}\right)^2 + \left(C_{gs}+C_p\right)^2} = R_{eq} \qquad (4.115)$$

$$\mathrm{Im}\{Z'_{gs}\} = -j\omega_0 \frac{\omega_0^2 C_{gs}^2 C_p R_{Gi}^2 + \left(C_{gs}+C_p\right)}{\omega_0^2\left[\left(\omega C_{gs}C_pR_{Gi}\right)^2 + \left(C_{gs}+C_p\right)^2\right]} \qquad (4.116)$$

The corresponding equivalent capacitance is

$$C_{eq} = \frac{\left[\left(\omega_0 C_{gs}C_pR_{Gi}\right)^2 + \left(C_{gs}+C_p\right)^2\right]}{\omega_0^2 C_{gs}^2 C_p R_{Gi}^2 + \left(C_{gs}+C_p\right)} \qquad (4.117)$$

that acts as the resonance capacitance of the input loop. From (4.117) C_p , and insering it in (4.115) R_{eq} can be calculated.



(a) (b)

Fig. 4.51 (a) The input of M1 together with C_p. (b) Its equivalent at ω_0.

It is obvious that C_{gs} and R_{Gi} must be replaced with C_{eq} and R_{eq} in (4.113):

$$\begin{aligned} L_S &= \frac{C_{eq}}{g_m}\left[R_A - \left(R_{LS}+R_{Si}+R_{eq}+R_{LG}\right)\right] \\ &\cong \frac{C_{eq}}{g_m}\left[R_A - \left(R_{Si}+R_{eq}+R_{LG}\right)\right] \end{aligned} \qquad (4.113\text{-a})$$

provided $R_{LS} \ll R_{Si}+R_{eq}+R_{LG}$, that is usually valid.

M2 is a unity gain current amplifier insulating M1 from the load to minimize the feedback over C_{dg1}. The dimension of M2 must be determined with care, considering noise and linearity. Smaller g_{m2} (therefore smaller gate width) reduces the contribution of M2 on the total output noise. But it must be kept in mind that the supply voltage is divided between M1 and M2, and the D.C. voltages of the drain nodes must be suitable for the voltage swings of these nodes without excessive nonlinear distortion. Otherwise in addition to the self nonlinearities of M1 and M2, the intrusion of the negative peaks of the output signal into the drain voltage of M1 severely limits the usable output swing, as will be shown Example 4.10.

Example 4.10

A tuned LNA will be designed for a 0.18 micron technology The tuning frequency is f_0 = 2 GHz. Internal impedance of the signal source is 50 ohm resistive and must be matched with the input impedance at 2 GHz. The voltage gain must not be lower then 20. The DC supply voltage is 1.8 V and the DC power consumption must not exceed 10 mW. The circuit diagram with all necessary components for PSpice simulation is given in Fig.4.51

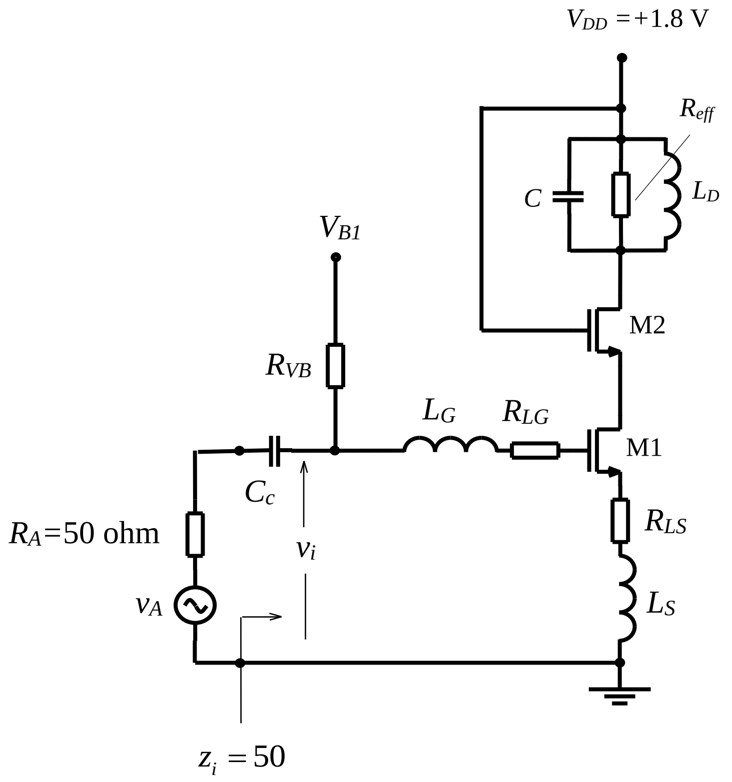


Fig. 4.51. The complete circuit diagram of the circuit with all parameters to be determined.

The important basic parameters and restrictions for this technology are given as follows:

The maximum value of the realisable on-chip inductors is 10 nH
The quality factors of on chip inductors are 10.
TOX = 4.2 nm that corresponds to C_{ox}= 8.2 e-7 F/cm^2
UO = 314 cm^2/V.s, that corresponds to μ_n = 220 cm^2/V.s for $V_{GS} \approx 0.7$ V
VTHO = 0.3 V
CGSO = CGDO =2.35 e-10 F/m
The sheet resistance of gate poly = 6 ohm/□

The overall voltage gain A_v was defined as the multiplication of the resonance gain of the input loop, A_R and the valtage gain of the cascode stage, A_C. Since A_R is usually slightly higher then unity, $|A_v| \approx |A_C|$ is a good approximation to attack the problem, which simplifes the calculation of g_{m1} and provides a margin for the final value of the gain. Since the possible maximum value of L_D is given as 10 nH and Q =10, g_{m1} of the input transistor can be calculated from (4.109) as

$$g_{m1} = \frac{20}{(10\times10^{-9})(2\pi\times2\times10^{9})\times10} \cong 16 \text{ mS}$$

and assuming V_B = 0.7, the width of the transistor from (4.110)

$$W = \frac{0.016\times(0.18\times10^{-4})}{220\times(8.2\times10^{-7})\times(0.7-0.3)} = 40\times10^{-4} \text{ cm} = 40\ \mu\text{m}$$

To assure the value of the transconductance, it is useful to fine-tune the width and/or the bias voltage with Spice simulation. In this example we short circuited L_S and L_G, connected the drain of M1 to ground with a high value capacitance to prevent any feedback, applied a 10 mV input voltage sweeped around 2GHz and adjusted the signal current of M1 with $W1$ and V_{B1} to 160 µA (that corresponds to g_{m1} = 16mS). The appropriate values found as $W1$ = 46 µm and V_{B1} = 0.77 V.

From (1.50), the maximum channel width to prevent excessive attenuation on the gate electrode is calculated as 35 µm. This leads to -at least- a 2-finger transistor with 23 µm finger length. But for this case the calculated internal parasitic gate resistance is unacceptably high. Therefore to have a reasonable internal resistances the number of finger increased to 11 with 4.5 µm finger length that provides 14 ohm internal gate series resistance and approximately 0.5 ohm internal source resistance. Another advantage of the finger structure is the decrease of the source and drain parasitic junction capacitances. The input capacitance of this W_1 = 11×4.5 µm = 49.5 µm transitor is calculated from (1.40) as

$$C_{gs} = \frac{2}{3}C_{ox}WL + \text{CGSO} \times W$$
$$= \frac{2}{3}(8.2\times10^{-7})(49.5\times10^{-4})(0.18\times10^{-4}) + (2.35\times10^{-12})(49.5\times10^{-4}) = 60.3 \text{ fF}$$

Since it is not effective on the input matching and the gain, let us start with W_2 = (0.5×W_1), and fine-tune later on for noise and linearity.

The sum of the source and gate inductances to resonate at 2 GHz with this capacitance can be calculated as 67 nH, that is impossible to realise on-chip. The solution is to use the possible maximum value of inductance for L_G, to estimate a small value for L_S and to connect a parallel capacitance (C_p) between G and S nodes for resonance. For (L_G + L_S)=11 nH the equivalent capacitance, C_{eq} defined with Fig. 4.51 becomes

$$C_{eq} = \frac{1}{\omega_0^2(L_G + L_S)} = \frac{1}{(1.26\times10^{10})^2(11\times10^{-9})} = 573 \text{ fF}$$

With C_{gs}=60.3 fF, R_{Gi}=14 ohm C_p from (4.117) and then R_{eq} from (4.115) can be calculated as 512 fF and 0.13 ohm, respectively.

Now L_S can be calculated from (4.113-a) with $R_{Si} \cong 0.5$ ohm , $R_{Gi} \cong 14$ ohm and

$$R_{LG} \approx \frac{L_G \omega_0}{Q} \approx \frac{(10\times10^{-9})(1.26\times10^{10})}{10} = 12.6 \text{ ohm}$$

as

$$L_S \cong \frac{(573\times10^{-15})}{(16\times10^{-3})}\left[50 - \left(0.5 + 0.13 + 12.6\right)\right] = 1.3 \text{ nH}$$

that is close enough to the previously estimated 1 nH.

With the resonance gain of the input loop, that is

$$A_R = \frac{v_{gs}}{v_A} = \frac{(L_G + L_S)\omega_0}{2R_A} = \frac{(10+1.3)\times10^{-9}\times(1.26\times10^{10})}{2\times50} = 1.42$$

and the gain of the cascode circuit,

$$\left|A_C\right| = \frac{v_o}{v_{gs}} \cong g_m \times R_{eff} = (16\times10^{-3})\times1260 = 20.16$$

the overall voltage gain becomes 28.6, that is higher then the target value with a considerable margin. Since the selectivity of the input resonance is very low, the bandwidth of the circuit is determined by the output resonance and can be calculated from (4.25) as B = 200 MHz

The component values calculated above and used for the PSpice simulation are as follows:

$W1$ = 11×4.5µm = 49.5 µm
$W2$ = 5×5µ = 25 µm
L_G = 10 nH, R_{LG} =12,6 ohm
C_p = 512 fF, R_{eq} =0.13 ohm
L_S = 1.3 nH, R_{LS} =1.64 ohm
C_c = 100 pF
R_{VB} = 10 k ohm

Usually a discrepency occurs between the hand calculated values and the simulation results. This results from the differences of models and parameters used. Especially for the simulation of tuned circuit a "fine tuning" is always necessary. For this case the fine tuning strategy was as follows:

- Apply a small AC input voltage, (say, 10 mV in amplitude) to sweep around 2 GHz.

- To adjust the g_{m1} to 16 mS;
  1. Connect a high value capacitance from the drain of M1 to the ground to prevent any feedback.
  2. Short circuit the gate and source inductors. Adjust the drain signal current to 160 µA at 2 GHz with V_B (it was 0.77V).

- To adjust the input resonance circuit to 2 GHz; disconnect the shorts on L_G and L_S , tune i_{LG} (the current of the input series resonance circuit) to maximum at 2GHz with C_p.

- Disconnect C_{d1} , tune the output voltage to maximum at 2GHz with C.

- To check the input matching, plot v_i and measure its value at 2 GHz (it can be fine-tuned to 5 mV with C_p for a perfect matching).

- Plot the output frequency characteristic. Determine the gain at 2 GHz and the bandwidth (They were A_v= 24.5 and B =200 MHz, in a good aggreement with the targetted values).

To finalize the design, we must fine-tune the width of M2 for good linearity at the output of the circuit, as well as at the drain node of M1. The procedure we applied was as follows:

- Apply a 2 GHz sinusoidal signal to the input with a low amplitude (for example 10 mV) and observe the voltage waveforms at the output and d1 node. Measure the magnitude of the output signal and calculate the voltage gain (found v_o =242 mV and A_v =24.2 for our circuit).

- Calculate the value of the gain corresponding to the "-1 dB point" that is a metric for the linearity of the amplifier. (See Part 4.6.7). (For our circuit it was $A_{v(-1dB)} = A_v /1.122 = 21.7$.

- Increase the amplitude step by step until the voltage gain drops to 21.7. In our case the input amplitude corresponding to $A_{v(-1dB)}$ was V_A = 60 mV. This is -by definition- the input voltage corresponding to the -1 dB point of the amplifier, in other words the maximum input voltage swing with an acceptable nonlinearity.

The waveforms of the output voltage and the voltage of the drain node of M1 are shown in Fig. 4.52. The output voltage has 1.30 V amplitude with a reasonable waveform, but the signal at the drain of M1 (that is the input voltage of M2) is excessively distorted. Since the input signal is pure sinusoidal, the distortion terms of this signal (the harmonics of the input frequency) are filtered-out by the output resonator and do not affect the output waveform. But in case of a modulated input signal containing side-frequencies, the intermodulations of these components may take place in the band. Therefore the maximum input amplitude must be lowered to a level that do not exhibite a considerable distortion at the input of M2. For our circuit it was approximately 30 mV, that is the considerably lower than the -1dB level[3].

To reduce W_2 increases the swing area of M1 and the distortion resulting from the intrusion of negative peaks of the output swing into the drain voltage of M1 appears at higher input levels. This also reduces the contribution of M2 on the total output noise. But excessive reduction of W_2 results in several disadvantages:

[3] This example shows that the -1dB level definition does not reflect the nonlinarities of the previous stages but only the output transistor.

- The increase of the input impedance of M2 increases the signal voltage on the drain of M1 that increases the unwanted feedback over C_{dg1}.
- Since $r_{i2} \ll r_{o1}$ does not hold any more, i_{d2} becomes smaller than i_{d1}, that reduces the gain.
- Decrease of the DC voltage share of M1 leads to nonlinear distortion at the output of this transistor.

Therefore the optimisation of W2 must be done with care. To optimize W2 in our example, after controlling the performance for $W_2 = 20\mu$m, 15µm and 10µm we decided to $W_2 = 3\times3.5\mu$m = 10.5 µm that provides a voltage gain of $A_v = 20.3$, -1 dB point input signal amplitude of $V_{A(-1dB)} = 65$ mV and 1.185 V output signal amplitude without resonable distortion on the drain of M1, as seen in Fig. 4.53.

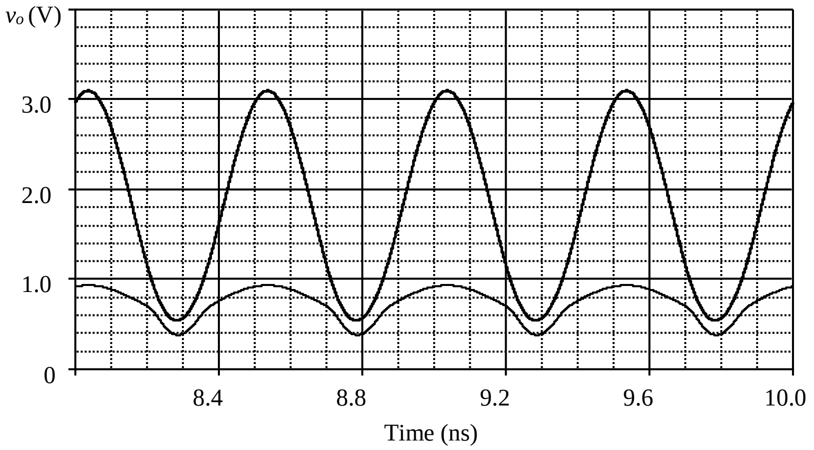


Fig. 4.52. The waveforms of the output voltage (upper trace) and the drain voltage of M1 for 60 mV input voltage amplitude ($W_1 = 49.5\mu$ and $W_2 = 25\ \mu$). Note the distortion of the signal at the drain of M1.

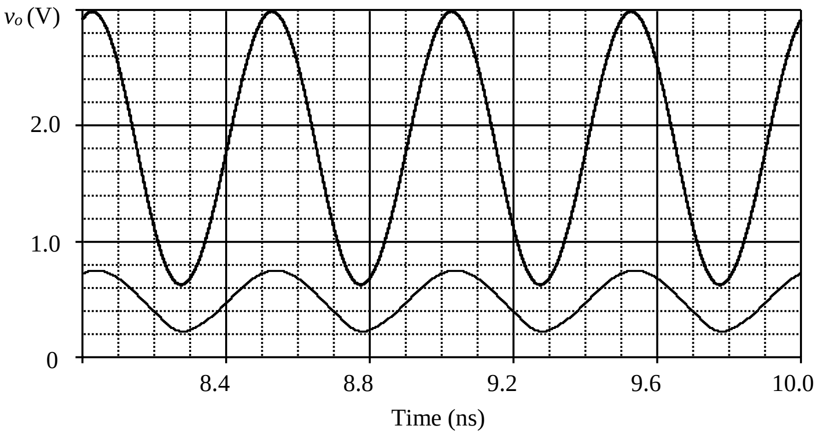


Fig. 4.53. The waveforms of the output voltage (upper trace) and the drain voltage of M1 for 65 mV input signal amplitude ($W_1 = 55\ \mu$ and $W_2 = 10.5\ \mu$). Note the clean waveforms of both signals.

4.6.4.1. Noise of a tuned LNA

The circuit diagram of a typical source degenerated LNA and the noise sources related to its noisy components are shown in Fig. 4.54. The approach used to analyse this circuit in 4.6.4 helps to calculate the noise contributions of the resistors placed in the input loop. The noise voltage sources situated in series in the input loop behave same as the signal source, v_A, i.e. the output noise components corresponding to each of these sources is equal to $g_m \times \overline{v}_n$.

The mean square noise components of the load current for 1 Hz bandwidth originating from each the noisy components are as follows:

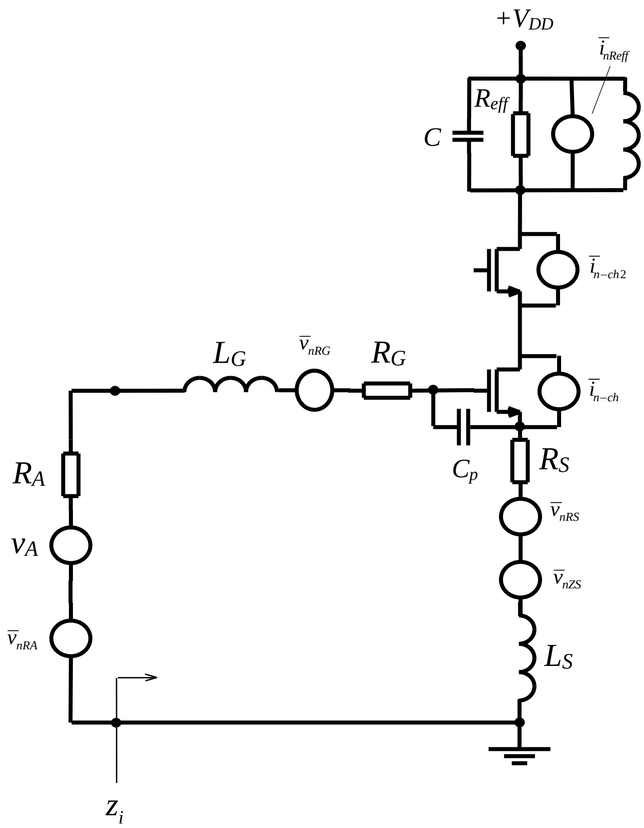


Fig. 4.54. Inductive source degenerated cascode LNA with its noise sources

Output noise originating from M1: $\overline{i}^2_{n-ch} = 4kT_C \times \gamma \times g_{m1}$

Output noise originating from M2: $\overline{i}^2_{n-ch2} = 4kT_C \times \gamma \times g_{m2}$

Output noise originating from R_A: $\overline{i}_{ndRA} = \overline{v}_{nRA} \times g_m \rightarrow \overline{i}^2_{ndRA} = 4kT_A R_A \times g^2_{m1}$ (4.118)

Output noise originating from R_G: $\overline{i}_{ndRG} = \overline{v}_{nRG} \times g_m \rightarrow \overline{i}^2_{ndRG} = 4kT_C R_G \times g^2_{m1}$

Output noise originating from R_S: $\overline{i}_{ndRS} = \overline{v}_{nRS} \times g_m \rightarrow \overline{i}^2_{ndRS} = 4kT_C R_S \times g^2_{m1}$

where T_A is the temperature of the antenna (the signal source) and T_C is the average temperature of the chip[4]. R_G is the sum of the R_{LG} and R_{eq} given with (4.115). The total mean square drain noise current of M1 is

$$\overline{i}_{ndT}^2 = \overline{i}_{n-ch}^2 + \overline{i}_{nRS}^2 + \overline{i}_{nRG}^2$$

This noise is transferred to the load via M2, together with its own channel noise current, $\overline{i}_{n-ch2}^2$. Hence the total noise current flowing through R_{eff}, as well as through M1 and M2, becomes

$$\overline{i}_{ndT}^2 = \overline{i}_{n-ch}^2 + \overline{i}_{nRS}^2 + \overline{i}_{nRG}^2 + \overline{i}_{n-ch2}^2 + \overline{i}_{nR(eff)}^2 \qquad (4.119)$$

There are two additional noise components related to the total drain noise current of M1. One of them originates from the noise voltage drop on the impedance series to the source (Z_S) at ω_0, that can be written as

$$\overline{v}_{nZS} = \overline{i}_{ndT} \times Z_S(\omega_0) \cong \overline{i}_{ndT} \times (L_S \omega_0)$$

and the corresponding mean square noise component on the output,

$$\overline{i}_{ndZS}^2 = (L_S \omega_0)^2 \times g_{m1}^2 \times \overline{i}_{ndT}^2 \qquad (4.120)$$

The other noise component related to the total drain noise current of M1 originates from the gate noise current given with (4.106). This current flows over ($R_G + R_A$) to the ground and produces a noise voltage equal to $\overline{v}_{nG} = (R_G + R_A) \times \overline{i}_{ng}$. The resulting drain noise component is

$$\overline{i}_{nG} = \left[(R_G + R_A) \times \left(\omega_0 \frac{C_{gs}}{g_m} \right) \overline{i}_{ndT} \right] \frac{g_{m1}}{(1 + g_{m1} Z_S)}$$

and its mean square value

$$\overline{i}_{nG}^2 = \left[(R_G + R_A) \frac{g_{m1}}{(1 + g_{m1} Z_S)} \right]^2 \left(\omega_0 \frac{C_{gs}}{g_{m1}} \right)^2 \times \overline{i}_{ndT}^2 \qquad (4.121)$$

Together with these additional components, the total mean square drain current, as well as the total output noise current flowing through the load becomes

$$\overline{i}_{ndT}^2 = \overline{i}_{n-ch}^2 + \overline{i}_{nRS}^2 + \overline{i}_{nRG}^2 + \overline{i}_{n-ch2}^2 + \overline{i}_{nR(eff)}^2 + A \times \overline{i}_{ndT}^2 + B \times \overline{i}_{ndT}^2 \qquad (4.122)$$

where $$A = (L_S \omega_0 \times g_{m1})^2 \qquad (4.123)$$

and $$B = \left[(R_G + R_A) \frac{g_{m1}}{(1 + g_{m1} Z_S)} \left(\omega_0 \frac{C_{gs}}{g_{m1}} \right) \right]^2 \qquad (4.123\text{-a})$$

[4] In reality the temperature of the channel may be considerably higher than the average temperature of the chip.

From (4.122), $\overline{i}_{ndT}^2$ can be solved as

$$\overline{i}_{ndT}^2 = \frac{\overline{i}_{n-ch}^2 + \overline{i}_{nRS}^2 + \overline{i}_{nRG}^2 + \overline{i}_{n-ch2}^2 + \overline{i}_{nR}^2}{1-(A+B)} \qquad (4.124)$$

and the total output noise power dissipated on R_{eff},

$$P_{no(tot)} = 4kT_C R_{eff} \left\{ \frac{(\gamma \times g_{m1}) + (R_G \times g_{m1}^2) + (R_S \times g_{m1}^2) + (\gamma \times g_{m2}) + (1/R_{eff})}{1-(A+B)} \right\} \qquad (4.125)$$

Since the noise power at the output originating from R_A is

$$P_{noRA} = \overline{i}_{ndRA}^2 \times R_{eff} = 4kT_A R_A g_{m1}^2 R_{eff} \qquad (4.126)$$

the noise factor of the amplifier becomes

$$F = 1 + \frac{P_{no(tot)}}{P_{noRA}} = 1 + \frac{T_C}{T_A} \frac{\left\{ \frac{(\gamma \times g_{m1}) + (R_G \times g_{m1}^2) + (R_S \times g_{m1}^2) + (\gamma \times g_{m2}) + (1/R_{eff})}{1-(A+B)} \right\}}{R_A g_{m1}^2} \qquad (4.127)$$

and can be re-arranged as

$$F = 1 + \frac{P_{no(tot)}}{P_{noRA}} = 1 + \frac{T_C}{T_A} \frac{\gamma\left(1 + \frac{g_{m2}}{g_{m1}}\right) + g_{m1}(R_G + R_S) + \left(\frac{1}{g_{m1} R_{eff}}\right)}{[1-(A+B)] R_A g_{m1}} \qquad (4.127\text{-a})$$

(4.127-a) provides some hints to improve the design for better noise performance:

- It is useful to reduce the resistances series to the gate and the source M1. To reduce the inductor related components of these resistances, it is necessary to increase the *Q* values that necessitates better inductor design. The parasitic resistances of the transistor are strongly geometry dependent. The finger-structure and the connecting lines must be optimized for minimum resistances.
- The increase of g_{m1} improves the noise performance to some extent. It means the increase of the the width of the transistor, that results in higher parasitic capacitanses. For non-velocity-saturated transistors, increasing the current is another way but also increases the power consumption.
- g_{m2} can be reduced with care, to prevent inconvenient operating conditions for M1 and M2 that affects the linearity as discussed in Example 4.10.
- The overall power consumption of the chip and the thermal resistance to the ambient must be as small as possible for a small (T_C/T_A) ratio. In addition M1 and M2 must not be close to the high power consuming parts on the chip, to prevent the increase of the -already high- local temperatures of the channel regions.

Example – 4.11.

Calculation of the noise figure designed in Example 4.10:

Parameters necessary to calculate the noise are as follows:

R_A = 50 ohm
$R_{(eff)}$ = 1260 ohm
R_{LS} = 1.64 ohm
R_{LG} = 12.6 ohm
g_{m1} = 16 mS
g_{m2}= 3.4 mS
R_{Si} =0.5 ohm
R_{eq} = 0.13 ohm
L_S = 1.3 nH
γ = 0.7 (assumed)
T_A= 300 K
T_C= 330 K (assumed)

Instead of the calculation of the noise factor directly from (4.127-a), let us calculate the individual noise powers at the output for 1 Hz bandwidth, originating from several components from (4.123), to evaluate and compare the contribution of each component on noise. Since it is common for all components, we start with calculation of *A* and *B* from (4.123) and (4.123-a):

$$A = \left[(1.3\times10^{-9})(1.26\times10^{10})(16\times10^{-3})\right]^2 = 68.7\times10^{-3}$$

with $Z_S \cong L_S\omega_0$

$$B \cong \left[(12.73+50)\frac{(16\times10^{-3})}{\left[1+(16\times10^{-3})(1.3\times10^{-9})(1.26\times10^{10})\right]}\left((1.26\times10^{10})\frac{(60.3\times10^{-15})}{(16\times10^{-3})}\right)\right]^2 = 1.42\times10^{-3}$$

Then, $\left[1-(A+B)\right] = 0.93$

Another factor common to all component;

$$4kT_C R_{eff} = 4\times(1.36\times10^{-23})\times330\times1260 = 2.26\times10^{-17}$$

The noise power at the output originating from M1:

$$P_{n(M1)} = \frac{4kT_C R_{eff}}{\left[1-(A+B)\right]}\gamma g_{m1} = \frac{2.26\times10^{-17}}{0.93}\times0.7\times(16\times10^{-3}) = 2.72\times10^{-19}\ \text{W}$$

The noise power at the output originating from R_G, where $R_G = R_{LG} + R_{eq}$

$$P_{n(RG)} = \frac{4kT_C R_{eff}}{\left[1-(A+B)\right]}\left(R_G\times g_{m1}^2\right) = \frac{2.26\times10^{-17}}{0.93}\times12.73\times(16\times10^{-3})^2 = 0.79\times10^{-19}\ \text{W}$$

The noise power at the output originating from R_S:

$$P_{n(RS)} = \frac{4kT_C R_{eff}}{[1-(A+B)]}\left(R_S \times g_{m1}^2\right) = \frac{2.26\times10^{-17}}{0.93}\times1.64\times(16\times10^{-3})^2 = 0.1\times10^{-19}\ \text{W}$$

The noise power at the output originating from M2:

$$P_{n(M2)} = \frac{4kT_C R_{eff}}{[1-(A+B)]}\gamma g_{m2} = \frac{2.26\times10^{-17}}{0.93}\times0.7\times(4.85\times10^{-3}) = 0.83\times10^{-19}\ \text{W}$$

The noise power at the output originating from R_{eff}:

$$P_{n(Reff)} = \frac{4kT_C R_{eff}}{[1-(A+B)]}\left(\frac{1}{R_{eff}}\right) = \frac{2.26\times10^{-17}}{0.93}\times\frac{1}{1260} = 0.19\times10^{-19}\ \text{W}$$

The sum of all these components giver the total noise power at the output originating from the amplifier:

$$\begin{aligned} P_{no} &= P_{n(M1)} + P_{n(RG)} + P_{n(RS)} + P_{n(M2)} + P_{n(Reff)} \\ &= (2.72+0.79+0.1+0.83+0.19)\times10^{-19} = 4.63\times10^{-19}\ \text{W} \end{aligned}$$

The output noise originating from the source resistance, R_A;

$$P_{n(RA)} = 4kT_A R_A \times g_m^2 \times R_{eff} = 4\times(1.36\times10^{-23})\times300\times50\times(16\times10^{-3})^2\times1260 = 2.63\times10^{-19}\ \text{W}$$

Hence the noise factor and corresponding noise figure can be calculated as

$$N = 1+\frac{P_{no}}{P_{n(RA)}} = 1+\frac{4.63\times10^{-19}}{2.63\times10^{-19}} = 2.76 \qquad NF = 10\log(N) = 4.4\ \text{dB}$$

Note that if it were assumed that the temperature of the chip is equal to the ambient temperature, the noise figure should be found as 4.15 dB.

The PSpice noise simulation gives for the total output mean square noise voltage $\overline{v}_{no}^2 = 5.81\times10^{-16}\ \text{V}^2/\text{Hz}$ that corresponds to $P_{no} = (5.81\times10^{-16})/1260 = 4.6\times10^{-19}\ \text{W}$ and for the mean square output noise voltage originating from R_A, $\overline{v}_{n(RA)}^2 = 3.4\times10^{-16}\ \text{V}^2/\text{Hz}$ that corresponds to $P_{no} = (3.4\times10^{-16})/1260 = 2.69\times10^{-19}\ \text{W}$. Hence the simulation results for the noise factor and the noise figure becomes N = 2.71 and NF = 4.32 dB that are in a good agreement with the calculated values.

4.6.4.2. The Differential LNA

For some applications, designing the LNA as a differential amplifier is more convenient:

a) If the antenna is balanced (differential), i.e. signal voltage of either of the output nodes is not at ground potential, but they have equal magnitude and opposite in phase with respect to the ground [5].

b) If the circuit following the LNA has a differential input.

Another advantage of a differential LNA is that its even harmonics are small (theoretically zero), due to the symmetrical structure of the circuit.

The schematic of a typical differential LNA is given in Fig. 4.55(a). The circuit is re-drawn in Fig. 4.55(b) that helps to consider the circuit as two single-ended LNAs and use the expressions derived in Section 4.6.4.

A center-tapped inductor is composed of two identical parts and its inductance depends on the coupling between two parts.. If the coupling of the two halves of the inductor is zero or negligibly small, the total inductance is equal to $L_T = 2L$. This case corresponds to two separate, physically uncoupled, and identical inductors. If these inductors are closely coupled, for example designed as a center tapped inductor as shown in Fig. 1.35(b), the total inductance becomes $L_T = 2L(1 \mp k)$, where k is the coupling coefficient[6]. If the two halves of the inductor are "wound" in the same direction the k is positive and the value of the inductance is highher then the zero coupling case. This is an advantage from the point of view of the area and the effective series resistance of the inductor.

But there is another effect that must not be overlooked. The output noise voltage of M21, namely $\overline{v}_{nd1}$ applied to the left half of the center tapped inductor induces a voltage on the wright half of the inductor as a oto-transformer equal to $k \cdot \overline{v}_{nd1}$, and visa-versa. Hence the total mean square noise voltage becomes $(1+k)^2(\overline{v}_{no1}^2 + \overline{v}_{no2}^2)$.

[5] Another solution is to connect a balanced antenna to the input of a single ended LNA via a BALUN, a simple structure that converts a balanced signal to an unbalanced signal, and visa-versa, to the expense of extra parasitics and noise.
[6] Typical values of k for center tapped ünductors are in the range of 0.6 – 0.8.

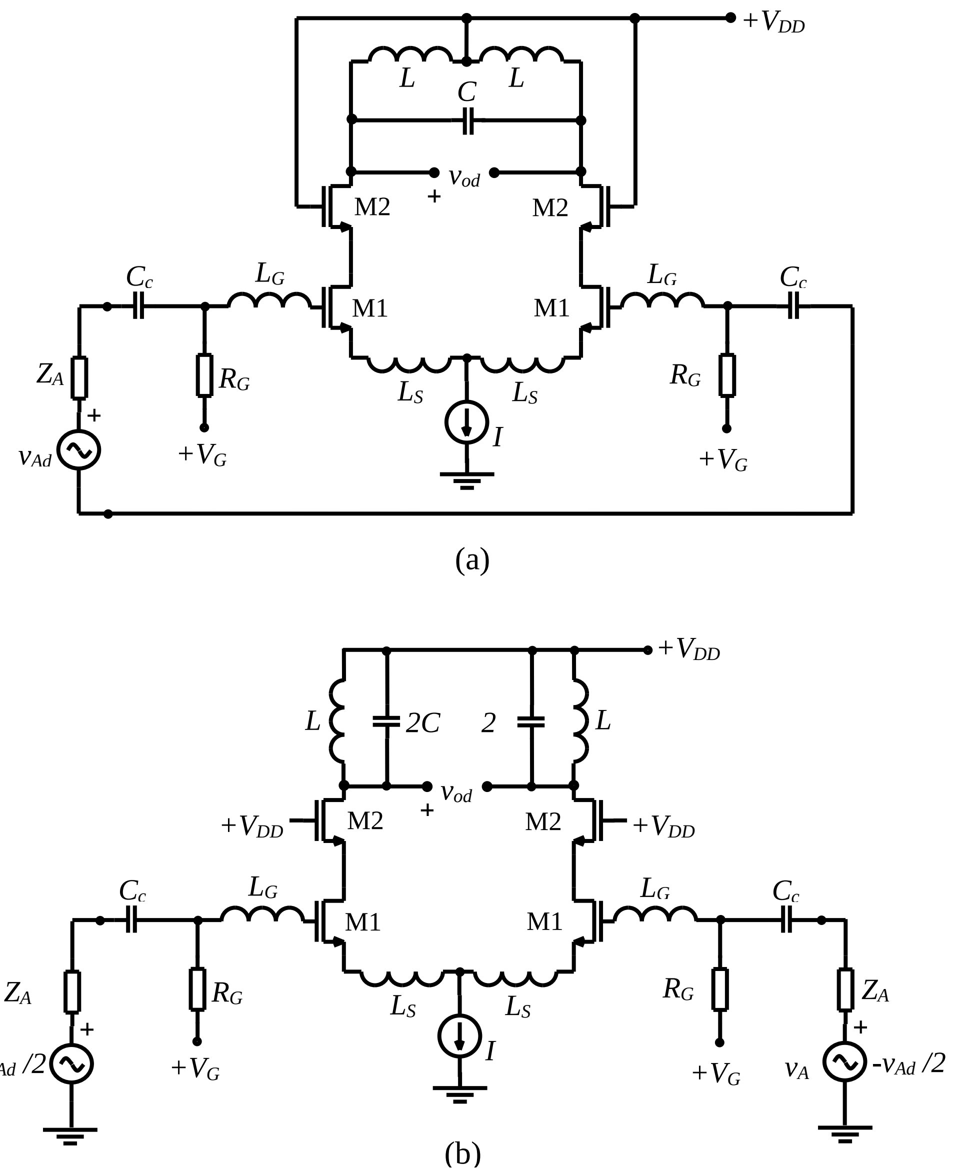


Figure 4.55 (a) Schematic of a typical differential LNA. (b) The re-arranged schematic that helps to use the expressions derived for the single-ended LNA

4.6.5. Wide-band LNAs, and "noise signal feedback"

Another approach to the LNAs different from the conventional amplifiers tuned to a certain frequency at the output, is to design a wide-band amplifier and to define the operating frequency at the input with a suitable passive band-pass filter (or directly using the tuned behavior of the antenna). With this approach it is possible to use a wide-band amplifier for any frequency, certainly in the band of the amplifier, only by changing the input filter or the antenna. Another advantage of this approach is reduced neighboring channel interferance.

It must not be overlooked that the noise voltages and currents of the resistors and transistors in a circuit are processed (amplified, fed-back, mixed, etc) as other signals in the circuit, until they reach to the output circuit. Then, their contributions to the total noise power at the output must be calculated and added on to find the total noise power. The r.m.s. values are necessary at this stage. To calculate the r.m.s values of the noises in advance masks some important effects; the "noise feedback" that is known since the early days of electronics[1] , and the "noise cancellation" technique appeared in several recent publications.

This fact will be examplified on several basic circuits below.

4.6.5.1. The common gate amplifier as a wide-band LNA

In a common gate amplifier given in Fig.4.56, there are three noisy components; the internal resistance of the input signal source, R_A, the MOS transistor and the load resistor, R_L .

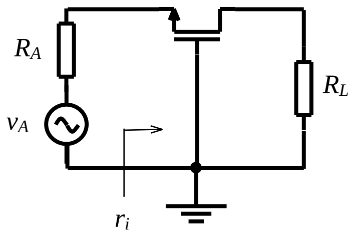


Figure 4.56 . The principal circuit diagram of a common gate amplifier

The contributions of these noise components to the output noise are as follows:

a) Contribution of R_A: The instantaneous value of the noise voltage of the source resistance (v_{nA}) is transferred to the input of the amplifier as

$$v_{ni} = \frac{r_i}{R_A + r_i} v_{nA}$$

and then amplified by the transistor with a voltage gain of $A_v \cong g_m R_L$. Provided that the amplifier can be considered as linear for this signal level, the amplified noise voltage at the output is a replica of the noise voltage at the input and the r.m.s value of this voltage is equal to

$$\overline{v}_{nAo} = A_v \times \overline{v}_{ni} = A_v \times \frac{r_i}{R_A + r_i} \overline{v}_{nA} \qquad (4.128)$$

---

[1] For example: H.J. Reich, " Theory and Applications of Electron Tubes", p. 201, McGraw-Hill, 1944.

and if the input is matched, i.e. $r_i = R_A$

$$\overline{v}_{nAo} = A_v \times \frac{1}{2}\overline{v}_{nA} = \frac{1}{2} g_m R_L \sqrt{4kTBR_A} \qquad (4.128\text{-a})$$

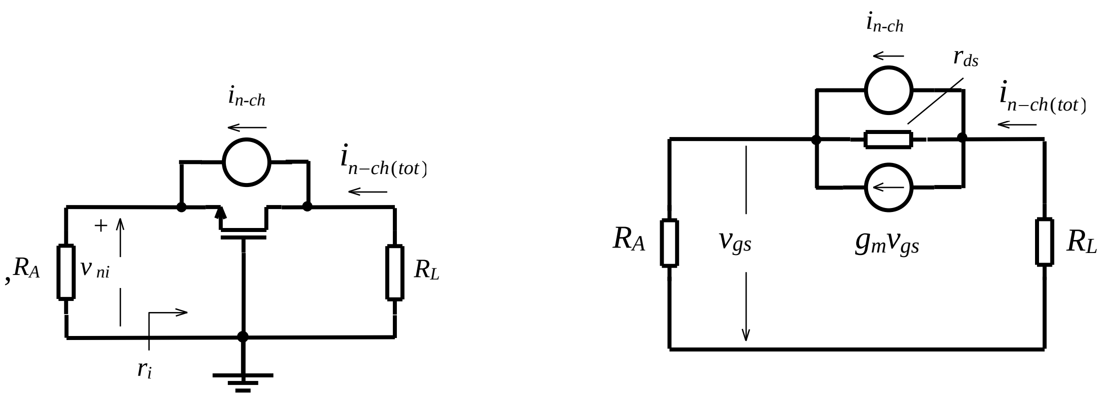
Since $R_A = r_i \cong 1/g_m$ for a common gate amplifier,

$$\overline{v}_{nAo} = \frac{1}{2} g_m R_L \sqrt{4kTB/g_m} \qquad (4.128\text{-b})$$

Hence the noise power dissipated on the load resistance owing to the noise of the source resistance becomes

$$P_{nAo} = \frac{\overline{v}^2_{nAo}}{R_L} = kTBg_m R_L \qquad (4.129)$$

b) Contribution of the channel noise: The channel noise current of the transistor $\overline{i}_{n-ch} = \sqrt{4kTB\gamma g_m}$ flows through the load resistor R_L as well as the internal resistance of the input signal source, R_A. The noise current flowing through R_A produces a voltage drop v_{ni} which provokes a channel noise current equal to $(v_{ni} \times g_m)$. This "feedback noise component" is in opposite direction with i_{n-ch} and reduces it. In other words, there is a "noise cancelling feedback". This qualitative explanation can be developed with the inclusion of the effect of the internal resistance of the transistor as shown in the equivalent circuit shown in Fig. 4.57(b) from which the resulting output noise current can be calculated as



(a) (b)

Figure 4.57. (a). The channel noise current, (b) the equivalent circuit to calculate the internal feedback effect of this current.

$$i_{n-ch(tot)} = i_{n-ch} \frac{r_{ds}}{2r_{ds} + R_A + R_L} = i_{n-ch} \times \delta \qquad (4.130)$$

where δ is the noise reduction parameter that approaches to ½ for

$$(R_A + R_L) \ll 2r_{ds}$$

The r.m.s. value of this noise current

$$\overline{i}_{n-ch(tot)} = \delta \overline{i}_{n-ch} = \delta\sqrt{4kTB\gamma g_m} \qquad (4.131)$$

The contribution of the channel noise on the output power now can be calculated as

$$P_{n-cho} = \overline{i}^2_{n-ch(tot)} \times R_L = \delta^2(4kTB\gamma g_m R_L) \qquad (4.132)$$

c) Contribution of the load resistance noise on the total output noise simply is

$$P_{n-RL} = 4kTB \qquad (4.133)$$

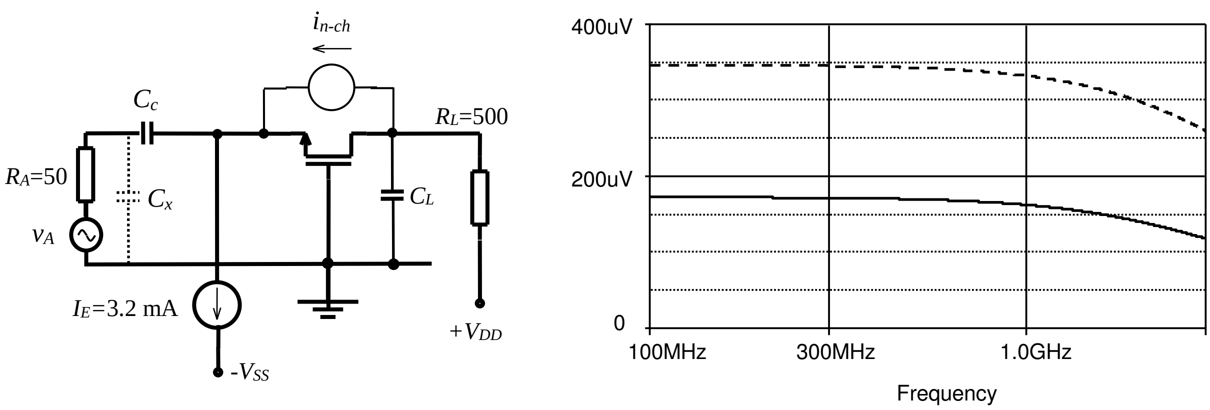
Now the noise factor of a common gate amplifier can be written as

$$F = 1 + \frac{P_{n-cho} + P_{n-RL}}{P_{n-Ao}} = 1 + \frac{\delta^2(4kTB\gamma g_m R_L) + 4kTB}{kTBg_m R_L}$$
$$F = 1 + 4\left(\delta^2\gamma + \frac{1}{g_m R_L}\right) = 1 + 4\left(\delta^2\gamma + \frac{1}{|A_v|}\right) \qquad (4.134)$$

As an example for R_A = 50 ohm, g_m = 20 mS, R_L = 500 ohm, r_{ds} = 1 kohm and γ = 0.7 the noise factor is F = 1.83 that corresponds to a noise figure of NF = 2.6 dB.

Example 4.12

The noise reduction parameter (δ) of the common gate amplifier given in Fig.4.58 (a) will be determined by pSpise simulation. Dimensions of the transistor are W =11×8μ, L =0.18μ.



(a) (b)

Figure 4.58. (a) The circuit diagram of the simulated amplifier. (b) The output noise voltage of the circuit originating from the emulated channel noise current (lover trace) and the noise voltage without noise reducing feedback (upper trace).

The size and the D.C current provides a 50 ohm input resistance and a voltage gain of 5. The channel noise current is emulated with a sinusoidal 1μV / 100 MHz current source. For v_A = 0, the “noise voltage “ at the output node measured as 172 μV as seen from Fig.4.58 (b). To eliminate the noise feedback R_A shunted with a high value capacitor. In this case the output voltage was 345 µV, showing us that the noise reduction parameter is 172/345 = 0.49 that is in a very good agreement with (4.139).

4.6.5.2. Parallel voltage feedback amplifier as wide-band LNA

A resistance loaded common source amplifier, applied a current feedback from the output voltage, that is a trans-impedance amplifier in nature, is shown in Fig.3-a. It is known that low frequency values of the trans-impedance, the voltage gain and the input and output impedances are[2]

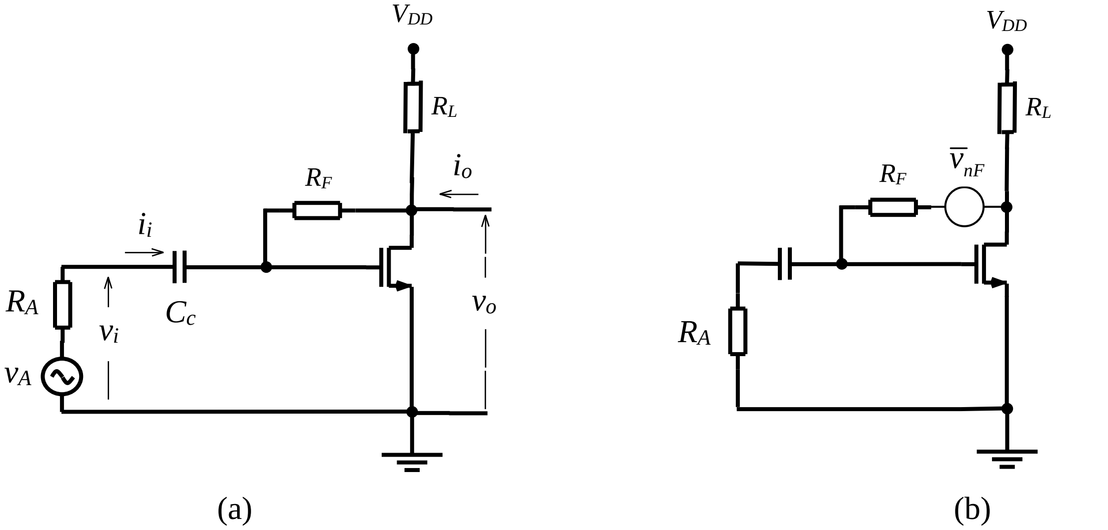


Figure 4.59. (a) The circuit diagram of a self-biased common source feedback amplifier. (b) Schematic to calculate the effect of the noise of R_F.

$$Z_m(0) = \frac{v_o}{i_i} = -R_L \frac{g_m R_F - 1}{g_m R_L + 1} \cong -R_F \qquad (4.135)$$

$$A_V(0) = \frac{v_o}{v_i} = \frac{Z_m(0)}{r_i} \cong -R_F g_m \qquad (4.136)$$

$$A_{VA}(0) = \frac{v_o}{v_A} = \frac{r_i}{(r_i + R_A)} A_v(0) \qquad (4.136\text{-a})$$

$$Z_i(0) = r_i = \frac{v_i}{i_i} = \frac{R_F + r_p}{1 + g_m r_p} \cong \frac{1}{g_m} \qquad (4.137)$$

$$Z_o(0) = r_o = \frac{v_o}{i_o} = \frac{r_p}{1 + g_m r_p} \cong \frac{1}{g_m} \qquad (4.138)$$

where r_p is the parallel equivalent of R_L and r_{ds} . These expressions can be approximated as mentioned, only provided that g_m , R_F and r_{ds} are sufficiently high.

---

[2] Effects of the reactive components and parasitics will be discussed later on.

The input matching condition can be written for $r_p \cong R_L$ from (4.137) as

$$R_A = \frac{R_F + R_L}{1 + g_m R_L} \quad (4.139)$$

and with (4.139) , (4.135) and (4.136) can be arranged as

$$\left|Z_m\right| = (R_F - R_A) \quad (4.140)$$

$$\left|A_v\right| = \frac{(R_F - R_A)}{R_A}, \qquad \left|A_{vA}\right| = \frac{1}{2}\frac{(R_F - R_A)}{R_A} \quad (4.141)$$

Problem 4.11.

Derive expression (4.140).

The noisy components in this circuit are the internal resistance of the input signal source, R_A, the MOS transistor, the load resistor, R_L. and the feedback resistor, R_F.

The contributions of these noise components to the output noise are as follows:

a) The equivalent noise voltage of the signal source internal resistance, v_{nA} is transferred to the input of the amplifier as

$$v_{nAi} = v_{nA}\frac{R_A}{R_A + r_i} \quad (4.142)$$

where r_i is the input resistance of the amplifier. Provided that the input is matched (i.e. $R_A = r_i$, the resulting output noise voltage signal

$$v_{nAo} = \frac{v_{nA}}{2} \times A_v$$

its RMS value

$$\bar{v}_{nAo} = \frac{\bar{v}_{nA}}{2}\left|A_v\right| \quad (4.143)$$

and the corresponding noise power at the output

$$P_{nAo} = \frac{\bar{v}_{nAo}^2}{R_L} = \bar{v}_{nA}^2 \frac{1}{4}\frac{\left|A_v\right|^2}{R_L}$$

Since $\bar{v}_{nA} = \sqrt{4kTR_A}$

$$P_{nAo} = kT\frac{R_A}{R_L}\left|A_v\right|^2 \quad (4.144)$$

b) The output noise power component related to the channel noise current of the transistor is

$$P_{n-cho} = \bar{i}_{n-ch}^2 R_L = 4kT\gamma g_m R_L \quad (4.145)$$

c) The noise power of the load resistor R_L itself is simply

$$P_{nLo} = 4kT \tag{4.146}$$

d) The contribution of R_F on the output noise is two-fold; owing to the noise of itself and the negative feedback effect over R_F.

Noise of R_F is represented with a noise voltage source, $\overline{v}_{nF}$ in Fig. 4.59(b). This source provokes a noise current along R_L, R_F and R_A:

$$\overline{i}_{nF} = \frac{\overline{v}_{nF}}{(R_L^{'} + R_F + R_A)} \tag{4.147}$$

where $R_L^{'} = R_L // r_{ds} \cong R_L$.

The contribution of this current on the output noise power can be calculated as

$$P_{nFo} = \overline{i}_{nF}^{\,2} R_L = \left[\frac{\overline{v}_{nF}}{(R_L^{'} + R_F + R_A)} \right]^2 R_L \cong 4kTR_F \frac{R_L}{(R_L + R_F + R_A)^2} \tag{4.148}$$

Hence the total noise power dissipating on the load resistor owing to the noisy components of the amplifier becomes

$$P_{no} = P_{n-cho} + P_{nFo} + P_{nL} = 4kT\left(\gamma g_m R_L + \frac{R_L R_F}{(R_L + R_F + R_A)^2} + 1 \right) \tag{4.149}$$

that corresponds to an RMS output noise voltage of

$$\overline{v}_{no} = \sqrt{P_{no} R_L} \tag{4.150}$$

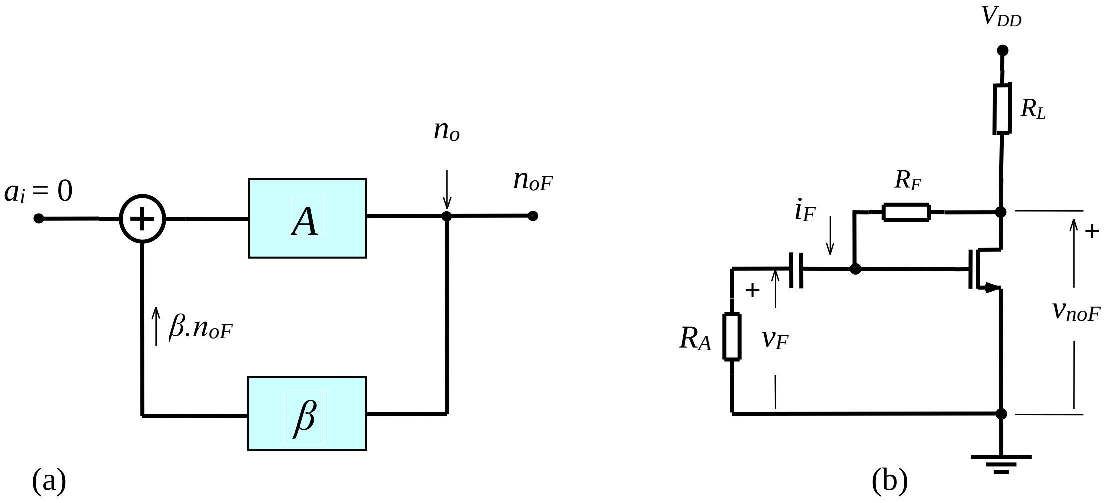


Figure 4.60. (a). The block diagram of a feedback amplifier with the output noise feedback. (b) Application to the parallel voltage feedback amplifier.

The noise feedback effect of a negative feedback amplifier can be modelled using the block diagram shown in Fig. 4.62(a) which is applicape to all single-loop feedback amplifiers. The input signal (a_i) is assumed zero. The noise signal at the output node of the amplifier originating from the components of the amplifier is shown as n_o. The total noise output signal n_{oF} is the sum of n_o and the fed-back noise signal:

$$n_{oF} = n_o + \beta A \times n_{oF}$$

that yields

$$n_{oF} = \frac{n_o}{(1-\beta A)}$$

This expression shows that if either β or A is negative (it means tha the feedback is negative) the noise signal at the output node becomes $(1-\beta A)$ times smaller then the original noise. It means that there is a noise cancelling feedback.

For the amplifier shown in Fig. 4.60(b) the voltage feedback factor and the voltage gain are

$$\beta = \frac{v_F}{v_{noF}} = \frac{R_A}{R_F + R_A} \quad , \qquad A_v = -\frac{|Z_m|}{r_i} = -\frac{|Z_m|}{R_A} \tag{4.151}$$

$$v_{noF} = v_{no} \frac{1}{1+\beta \frac{|Z_m|}{R_A}} = \frac{1}{1+\frac{(R_F - R_A)}{(R_F + R_A)}} \tag{4.152}$$

$$\delta = \frac{v_{noF}}{v_{no}} = \frac{1}{1+\frac{(R_F - R_A)}{(R_F + R_A)}} = \frac{(R_F + R_A)}{2R_F} \tag{4.153}$$

where δ is the output noise voltage reduction factor owing to the feedback and is always smaller than unity. The corresponding noise power reduction factor is δ^2 . Hence the output noise power and the noise factor with feedback become

$$P_{noF} = \delta^2 \times P_{no} \tag{4.154}$$

$$F \cong 1 + \frac{P_{noF}}{P_{nAo}} = 1 + \frac{4\delta^2 \left(\gamma g_m R_L + \frac{R_L R_F}{(R_L + R_F + R_A)^2} + 1 \right)}{\frac{R_A}{R_L} |A_v|^2} \tag{4.155}$$

Note that the noise factor increases with g_m, strongly increases with R_L and decreases with gain.

Now a design flow for an input matched self biasing parallel voltage feedback amplifier can be defined as follows:

The basic D.C. condition for the circuit is

$$V_{DS} = V_{GS} = V_{DD} - I_D R_L \tag{4.156}$$

The drain current for a non-velocity saturated transistor

$$I_D \cong \frac{1}{2}\mu C_{ox}\frac{W}{L}(V_{GS}-V_T)^2 = \frac{1}{2}g_m(V_{GS}-V_T) \qquad (4.157)$$

and for a velocity saturated transistor

$$I_D \cong kC_{ox}W(V_{GS}-V_T) = g_m(V_{GS}-V_T) \qquad (4.158)$$

Since the amplifier to be designed is a wide-band circuit, a velocity-saturated (small geometry) transistor must be preferred to extend the cut-off frequency. Therefore the design flow must be based on (4.140). These expressions lead to

$$R_F = \left(|A_v|+1\right)R_A \qquad (4.159)$$

$$g_m = \frac{1}{R_L}\left(\frac{R_F+R_L-R_A}{R_A}\right) \qquad (4.160)$$

Since $V_{DS}=V_{GS}$;

$$I_D = \frac{(V_{DD}-V_T)}{R_L+\dfrac{1}{g_m}} \qquad (4.161)$$

(4.159) directly gives the value of R_F corresponding to any source resistance and gain. (4.160) and (4.161) help to calculate the transconductance and the drain D.C. current corresponding to any anticipated load resistor that determines the -3dB frequency together with the load capacitance and the output node parasitics.

From (4.155) it can be seen that the noise factor increases linearly with g_m , quadratically with R_L and decreases quadratically with A_v . At the other hand the power consumption linearly decreases with the increase of the load resistor. The bandwidth depends on the load resistance and the transistor geometry which affects the corner frequencies of the output as well as the input node. Therefore there is a trade-off among noise, power consumption and bandwidth.

Example 4.13.

An input matched wide band LNA as shown in Fig. (4.61-a) will be designed. The technology is 0.18 micron UMC CMOS technology for which L =0.18 µm, V_T = 0.3V, C_{ox} = 8.2×10^{-7} F/cm^2 and V_{DD} = 2V. The internal resistance of the input signal source is R_A = 50 ohm and the load capacitance is 100 fF. The required voltage gain from the source is 10 dB ($|A_v| = 3.2$) that corresponds to a voltage gain from the input node of 16 dB ($|A_v| = 6.4$). The -3dB frequency and the noise figure at 1 GHz must be higher then 3 GHz and 3 dB, respectively. It will be assumed that the transistor is operating in the velocity saturated mode, that must be checked at the end of the design. The r_{ds} of the transistor will be assumed considerably higher than the drain load resistance.

(4.141) imposes the value of the feedback resistoras

$$R_F = (|A_V| + 1) R_A = 370 \text{ ohm}$$

The g_m and I_D values calculated from (4.160) and (4.161), and NF from (4.155) corresponding to different R_L values are given in the table below. The -3 dB frequencies obtained from the SPICE simulations are also given. From this table it can be seen that R_L = 200 ohm is the appropriate choice. The noise figure can be decreased to smaller values to the expense of a higher D.C. power consumption and higher transistor widths that impairs the band width.

R_L (ohm)	g_m (mS)	I_D (mA)	NF (dB)	BW (GHz) (simulation)
100	84	15	1.55	3.25
150	62.7	10.2	2.33	3.75
200	52	7.75	2.8	3.83
250	45.6	6.25	3.2	3.85
300	41.3	5.28	4.5	3.70

In Fig.4.61 the SPICE simulation results corresponding to R_L = 200 ohm is shown. The width of the transistor is optimized to obtain a good input matching, i.e. $r_i = R_A$ = 50 ohm. From these results it can be seen that;

- The voltage gain is 3.17, very close to the target value.
- The bandwidth is 3.83 GHz.
- The noise reduction factor (δ) and the noise factor calculated from (4.153) and (4.155) with these circuit parameters found as δ = 0.567 and $F \cong 1.91$ that corresponds to $NF = 2.8$ dB.

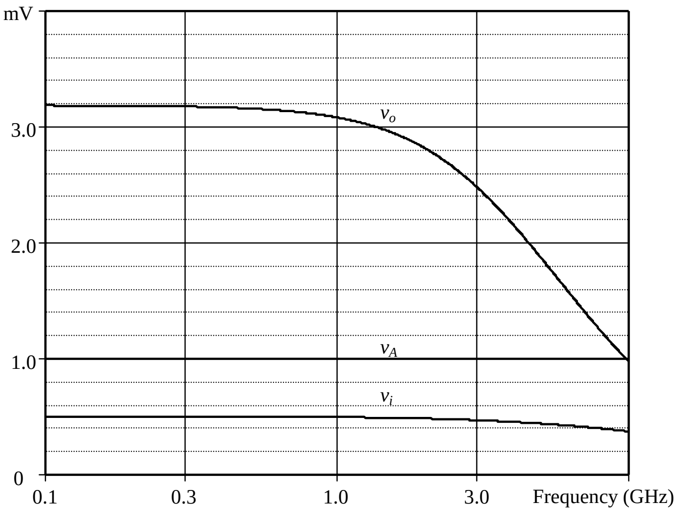


Figure 4.61, Simulation results for R_F = 370 ohm and R_L = 200 ohm.

- The input matching is maintained perfectly up to 1 GHz and reasonable up to 3 GHz. The slight decrease of v_{in} at higher frequencies indicate a phase shift due to the input capacitance that impairs the noise feedback and results in a frequency dependent increase of δ and consequent increase of the noise figure.

To proof the calculations about the noise reduction effect of the feedback a simulation was performed emulating the total noise current originating from the amplifier with a sinusoidal 1 μA current source in parallel to the load and the total output noise voltage simulated for $R_A = 0$ (that corresponds to no feedback) and $R_A = 50$ ohm (that corresponds to the case of feedback). The corresponding noise voltages at the output shown in Fig. 4.62 indicates a 0.57 noise voltage reduction owing to the feedback that is in a very good agreement with the calculated value.

There is an obvious drawback of the self biased amplifier shown in Fig. 4.59; the position of the D.C. operating point. Since $V_{DS} = V_{GS}$, the negative swing range of the output voltage is smaller then the positive swing range.This asymmetric and limited output dynamic range impairs the linearity.

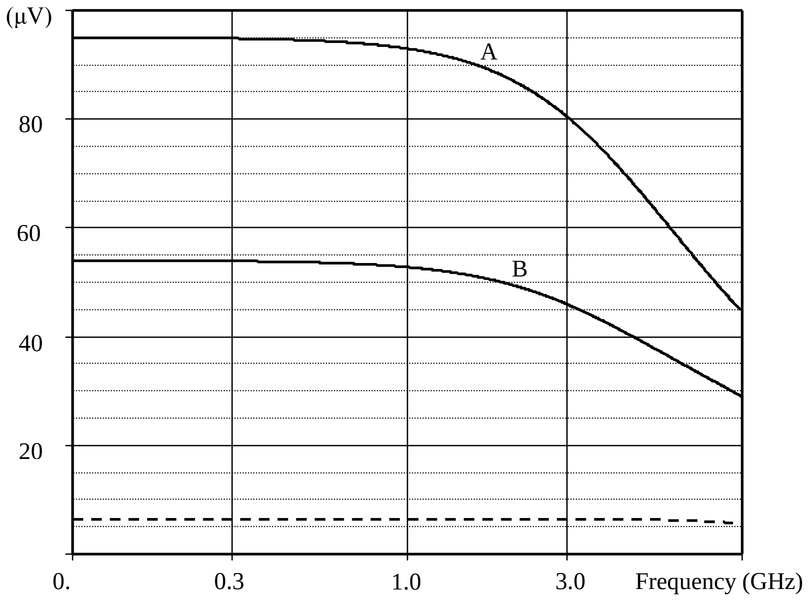


Figure 4.62. Simulation results showing the output noise voltage (A) feedback effect eliminated, (B) with feedback for a 1μA output noise current. The dotted curve shows the feedback voltage on R_A.

The symmetical version of this circuit shown in Fig. 4.63 is an effective solution to increase the symmetrical output dynamic range. This circuit is nothing but a CMOS inverter. The rail-to-rail symmetrical output swing capability of the circuit guarantees a good linearity around the operating point, situated in the middle of the output voltage swing range. The small signal equivalent of this circuit is exactly same as of the circuit given in Fig. 4.59, where g_m and r_{ds} are the sum of the transconductances and internal resistances of M1 and M2 and all expressions derived for the original circuit are valid for the symmetrical version (See Part 2.2).

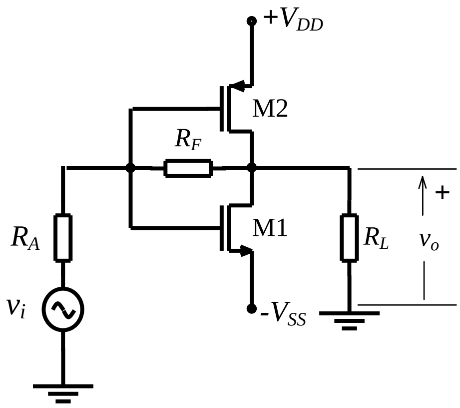


Figure 4.63. Symmetrical version of the parallel voltage feedback amplifier.

4.6.6. The noise cancelletion in wide-band amplifiers

The noise cancellation –in broader sense- means to obtain a negative replica of the noise signal at the output of a system and add it to the original output signal. Since the total signal at the output is the sum of the intentional output signal and the non-intentional output noise, the generated negative replica cancels out the noise at the output. In other words, a noise cancelling feedback is established.

This concept is applied to reduce the noise of a wide-band amplifier by F. Bruccoleri at.al.[1] The amplifier of which the output noise is intended to be cancelled is a self-biased common source feedback amplifier. The circuit diagram is drawn with the cancellation mechanism in Fig. 4.64. A3 is a negative gain amplifier. The gain of A2 is positive and small in magnitude, that can be smaller or higher than or equal to unity.

From 4.6.5.2 we know that the v_o output voltage is composed of two components; the normal output signal which is equal to ($v_i \times A_{v1}$) and the noise voltage signal at the output, v_{no}. A replica of this noise voltage, that is divided from v_{no} by R_F and R_A appears at the input node:

$$v_{noi} = v_{no} \frac{R_A}{R_A + R_F} \tag{4.162}$$

It can be seen that if

$$v_{noi} \times A_{v3} = -v_{no} \times A_{v2}$$

the noise voltages reaching to the summing node becomes equal in magnitude but opposite in sign, and consequently cancel out each other. The gain of A2 must fulfill

$$A_{v2} = |A_{v3}| \frac{R_A}{R_A + R_F} \qquad (4.163).$$

and be “fine tunable” for a perfect cancellation.

At the other hand, the signal voltages reaching to the summing point, namely $v_i(|A_{v1}| \times A_{v2})$ and $v_i \times |A_{v3}|$ are in-phase. Therefore the output voltage at the noise-compensated output becomes

$$v_{oNC} = v_i(|A_{v1}| \times A_{v2} + |A_{v3}|) \tag{4.164}$$

and from (4.141), (4.163) and (4.164)

$$v_{oNC} = v_i \left(A_{v2} \frac{2R_F}{R_A} \right) \tag{4.165}$$

that is almost twice higher that the voltage at the output of A1.

[1] F. Bruccoleri, E. A. M. Klumpering, B. Nauta, “ Wide-Band CMOS Low-Noise Amplifier Exploiting Thermal Noise Cancelling”, IEEE JSSC, Feb. 2004, pp. 275-282.

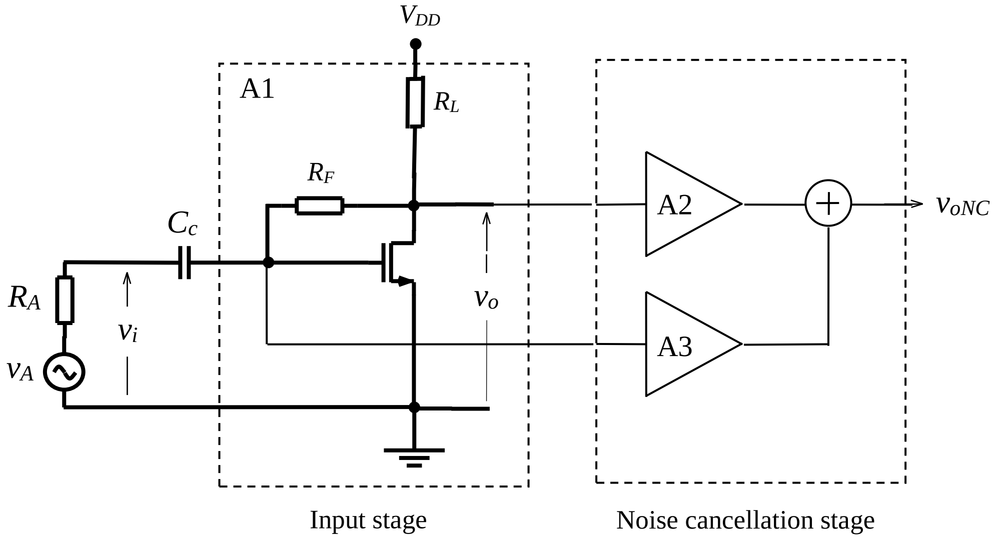


Figure 4.64. Principal circuit diagram of a wide-band LNA with noise cancellation feature.

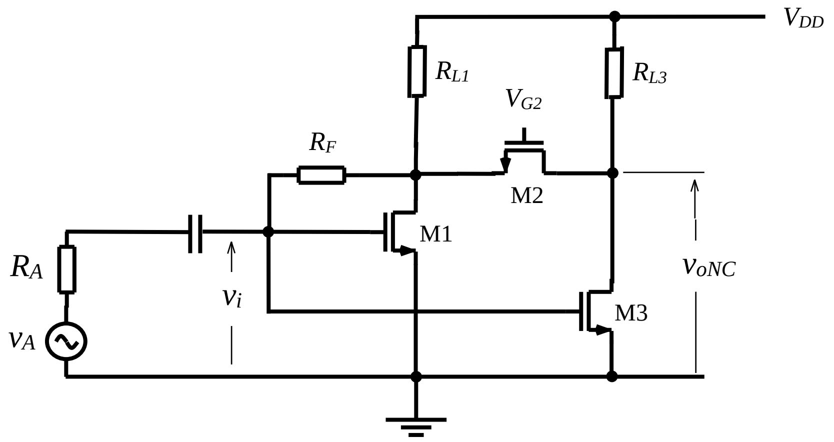
It must be noted that there are two realities affecting the noise cancellation advantage of the approach:

- For a perfect cancellation of the output noise of A1, the delay (or phase) characteristics of A2 and A3 must be identical along the bandwidth of the amplifier.
- The noises of A2 and A3 are out of the cancellation process and appear at the output of the summing circuit.

The noise at the output originating from A2 and A3 depend on the structure of these amplifiers and will be demonstrated in the following example.

Example 4.13.

The circuit diagram of a noise cancelled wide-band LNA is designed with UMC 018 technology is given in Fig. 4.67. M1 is the feedback amplifier investigated in Example 4.13 to provide a low input impedance for matching. M3 is a common source amplifier to provide the necessary negative voltage gain from the input node to the summing node. M2 is a grounded gate amplifier to provide a positive, low voltage gain. R_{L3} is shared by M2 and M3 and serves to sum the output voltages of M2 and M3. Since the voltage gain of M2 is low (around unity) it is a low transconductance, low current and relatively high input resistance circuit. The gate bias voltage of M2 helps to fine-tune the gain of this stage to obtain a perfect cancellation and minimum noise factor. The simulation results for the voltage gain of the amplifier from v_i is 10.66 (20.5 dB), the bandwidth 2.25 GHz and the power consumption is 25.2 mW.



V_{DD} = 2 V
V_{G3} = 1.43 V
M1: 500 /.18
M2: 20 /.18
M3: 200 /.18
R_A = 50 ohm
R_F = 370 ohm
R_{L1} = 200 ohm
R_{L3} = 200 ohm

Figure 4.65. The circuit diagram of the wide-band LNA with noise cancellation feature.

The results of a SPICE simulation to observe the noise cancellation is given in Fig. 4.66. For this purpose the total noise current generated in A1 is emulated with a 1μA, 100 MHz sinusoidal current, parallel to R_{L1}. (a) is the noise voltage at the output of M1 while (c) is the noise voltage at the output of the circuit that indicates a perfect noise cancellation.

But it can be seen that the cancellation process is very sensitive of the bias voltage (and gain) of M2 and permits a spread of the bias voltage in the order of only few tens of millivolts.

The simulation results given in Fig. 4.67 shows the variations of the noise voltage at the output of M1, at the input of M1 and at the output of the amplifier, corresponding to a 1 μA sinusoidal noise current parallel to R_{L1}. As seen from curve (c) the noise at the output is fully cancelled out along the frequency band of the amplifier. This indicates that the signal delays of M2 and M3 are matched. The slight increase of the noise at the high end of the band indicates increased discrepency of the delays with frequency.

As mentione before, the cancellation process cancels out the noise component at the output port originating from M1. The noises of M2 and M3 appears at the output and determine the noise figure of the amplifier.

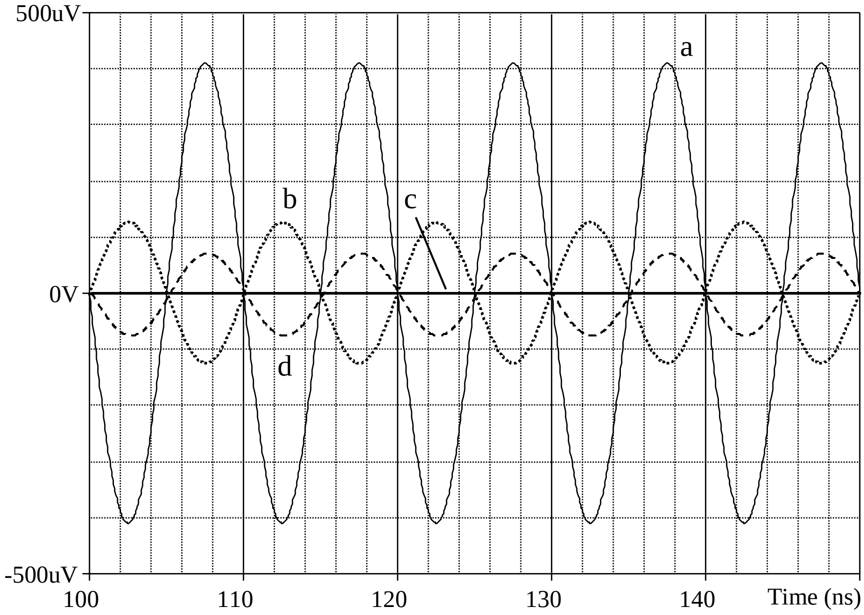


Figure 4.66. (a) The emulated noise voltage at the output of M1, corresponding to a 1 μA noise current on R_{L1}. (b) The noise voltage at the output of the amplifier for V_{GS2} = 1.335 V (the phase relation with (a) indicates an unsatisfactory gain of the A1-A2 chain. (c) The noise voltage at the output of the amplifier for V_{GS2} = 1.435 V that corresponds to a perfect noise cancellation. (d) Corresponds to V_{GS2} = 1.535 V. Note the phase reversal owing to the over- increased gain of M2.

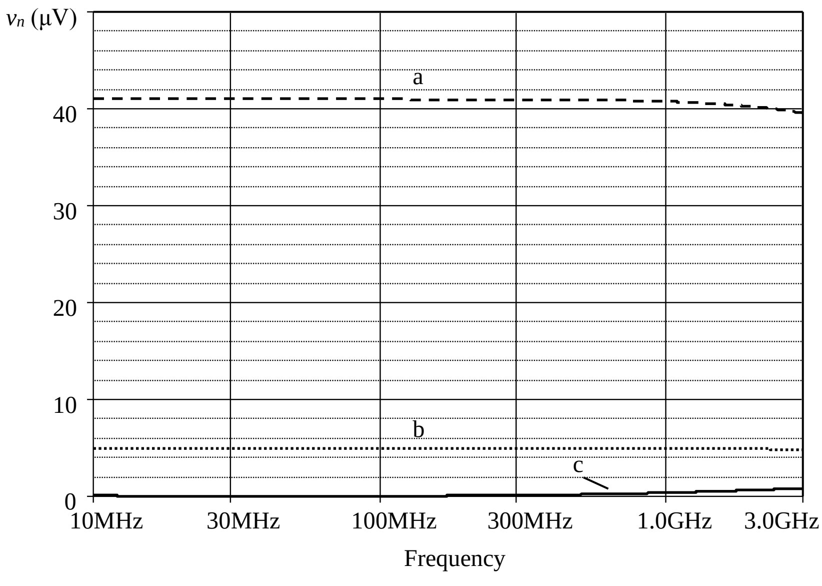


Figure 4.67. Variation of the emulated noise originaring from the first stage: (a) at the output of M1, (b) at the input of M1 fed-back over R_F , (c) the cancelled noise at the output of the amplifier.

Now let us calculate the noise originating from M2 and M3 for this example and find the *NF* of the amplifier. The transconductances of M2 and M3 are 3.37 mS and 31.47 mS, respectively and they share the same load resistor, R_{L3}. The total noise power at the output originating from these components is

$$P_{no2,3} = 4kT\gamma(g_{m2} + g_{m3})R_L + 4kT$$

The noise voltage at the output owing to the signal source;

$$\begin{aligned} v_{nAo} &= \frac{1}{2} v_{nA} \left(|A_{v1}| A_{v2} + |A_{v3}| \right) \\ &= \frac{1}{2} v_{nA} |A_{v3}| \frac{R_F}{R_A + R_F} \end{aligned}$$

The noise power at the output owing to the signal source;

$$P_{nAo} = \frac{v_{nAo}^2}{R_L} = kTR_A |A_{v3}|^2 \left(\frac{2R_F}{R_A + R_F} \right)^2 \frac{1}{R_{L3}} = 4kT |A_{v3}|^2 \left(\frac{R_F}{R_A + R_F} \right)^2 \frac{R_A}{R_{L3}}$$

Hence the noise factor becomes

$$F = 1 + \frac{4kT \left(\gamma(g_{m2} + g_{m3})R_{L3} + 1 \right)}{4kT |A_{v3}|^2 \left(\frac{R_F}{R_A + R_F} \right)^2 \frac{R_A}{R_{L3}}} = 1.76$$

that corresponds to NF = 2.46 dB, that is only slightly better than the noise figure of the non-noise-compensated first stage.

These results show that;

- The noise noise of the input stage is effectively cancelled out at the output.
- The voltage gain is approximately doubled.

But it must be noted that;

- The cancelletion stage (for the proposed circuit M2 and M3) adds a noise comparable with the non-compensated noise of the input transistor. With a better design of the cancellation stage it is possible to reach to a smaller owerall noise.
- For the cancellation stage, the equality of the signal delays of the paths subject to be added have prime importance for wide-band noise cancellation.
- As a more important fact, the cancellation process is very sensitive to the gain of M2 (or M3) and needs a fine-tuning.
- Another disadvantage is the increase of the power consumption.

4.6.7 The Limits of Usable Input Levels for LNAs

The signal voltage level delivered to the input of an LNA from the antenna may vary in a very wide interval, from very weak signals comparable to the noise level, to high amplitude signals resulting in severe nonlinear (harmonic and intermodulation) distortion.

For weak signal levels, the signal-to-noise ratio at the output determines the lower limit of the input signal. The acceptable minimum value of the output S/N ratio depends on the area of application and the demodulation techniques applied to the incoming carrier. As mentioned in Section 4.6.3, the output S/N ratio is determined by the noise factor (F) of the amplifier that is a measure of the internally generated noise of the amplifier and the input S/N ratio. On the other hand, the input S/N ratio depends on the field strength of the incoming wave, the antenna and the coupling scheme connecting the output of the antenna to the input of the amplifier.

An antenna can be considered as a "transducer" delivering a voltage to its output port, proportional to the strength of the electromagnetic wave illuminating the antenna. The proportionality factor depends on the direction of the incoming wave, the structure of the antenna and the frequency. It must be kept in mind that an antenna is a resonating system (with very few exceptions) such that the output voltage reaches its maximum at a certain frequency, depending on the structure (dimensions, shape, etc.) of the antenna. It is obvious that to maximize the input signal voltage and correspondingly improve the input S/N ratio, it is necessary to operate the antenna at its resonance frequency.

The open-ended maximum output voltage of an antenna ($v_o^{antenna}$) can be expressed in terms of the magnitude of the electric field component of the incoming wave (E) and the maximum value of the proportionality factor ($\alpha^{antenna}$) that corresponds to the resonance frequency and zero angle between the direction of the wave and the axis of the main lobe of the radiation pattern of the antenna:

$$v_o^{antenna} = \alpha^{antenna} \times E \qquad (4.166)$$

Since the current of a MOS amplifier is controlled by the input signal voltage (and not signal power), $v_o^{antenna}$ must be transferred to the input of the amplifier with possible maximum efficiency to increase the input S/N ratio and correspondingly, the output S/N ratio.

The signal voltage reaching the input port of the LNA depends on the interface between the output of the antenna and the input of the amplifier. If we denote the efficiency of this interface with $\beta^{coupling}$, the input signal of the amplifier corresponding to a certain value of E becomes

$$v_{in} = \alpha^{antenna} \times \beta^{coupling} \times E \qquad (4.167)$$

For example; if the antenna is very close to the input of the amplifier as mentioned in Section 4.6.1 and if the input impedance of the amplifier is matched to the antenna at the resonance frequency, the input voltage of the amplifier is equal to one half of $v_o^{antenna}$, in other words $\beta^{coupling} = 0.5$. On the other hand, if the input impedance is very high compared to the internal impedance of the antenna, the input voltage of the amplifier is equal to $v_o^{antenna}$ and $\beta^{coupling} = 1$.

If a transmission line has to be used between the output of the antenna and the input of the amplifier, the conventional approach is to match the input impedance of the amplifier to the characteristic impedance of the line and to use the antenna at a point where its impedance is real and equal to the line impedance. The $\beta^{coupling}$ in this case is considerably smaller and obviously impairs the S/N ratio.

For high signal levels there is another problem – not only for LNAs but all kinds of amplifiers; the nonlinear distortion of the output signal which produces harmonics of the sinusoidal components of the input signal and their intermodulation products. The amount of nonlinearity depends on the supply voltage, the load, the position of the operating point on the output characteristic curves and obviously, on the input signal level. This basic behavior of amplifiers will be examined on the most frequently used common source amplifier, in the following.

The nonlinearity is primarily related to the relation between the drain current and the gate-source voltage. The output voltage is equal to the voltage drop on the load. In case of a wide-band amplifier, the load is resistive up to the vicinity of the -3dB frequency. Therefore, the nonlinearity of the drain current directly reflects on the output voltage. But in the case of a tuned amplifier, the tuned load filters out the harmonics and the intermodulation products falling outside of its pass band.

The drain current depends not only on the gate-source voltage but also on the drain-source voltage to some extent, and consequently, on the load. In Fig. 4.55(a) the “static” (DC) and the “dynamic” load lines of a common source amplifier loaded with a parallel resonance circuit are shown. The static and the dynamic load lines are determined by the DC resistance of the inductance (r_L) and the resonance impedance of the load (R_{eff}), respectively. The input voltage-to-output current characteristic of this circuit, called the “dynamic transfer characteristic”, can be derived from Fig. 4.68(a) and is shown in Fig. 4.68(b). The nonlinearity of the transfer curve is obvious; to obtain a large dynamic range for the drain current with an acceptable distortion, the operating point (Q) must be in the middle of this curve.

In case of a resistance loaded (wide-band) amplifier, the static and dynamic characteristics are same up to the vicinity of the upper cut-off frequency.

The nonlinearity of the drain current can be represented with a Taylor series referred to the operating point:

$$I_D = I_{DQ} + \left.\frac{dI_D}{dV_{GS}}\right|_Q \Delta V_{GS} + \frac{1}{2!}\left.\frac{d^2 I_D}{dV_{GS}^2}\right|_Q (\Delta V_{GS})^2 + \frac{1}{3!}\left.\frac{d^3 I_D}{dV_{GS}^3}\right|_Q (\Delta V_{GS})^3 + \cdots$$

This expression can be converted into

$$i_d = g_1 v_{gs} + g_2 v_{gs}^2 + g_3 v_{gs}^3 + g_4 v_{gs}^4 + \cdots \tag{4.168}$$

where v_{gs} and i_d are the small signal components of the gate-source voltage and the drain current, respectively. The output signal voltage is $v_o = i_d R_{eff}$ for a tuned amplifier with a bandwidth determined by the effective quality factor of the resonance circuit.

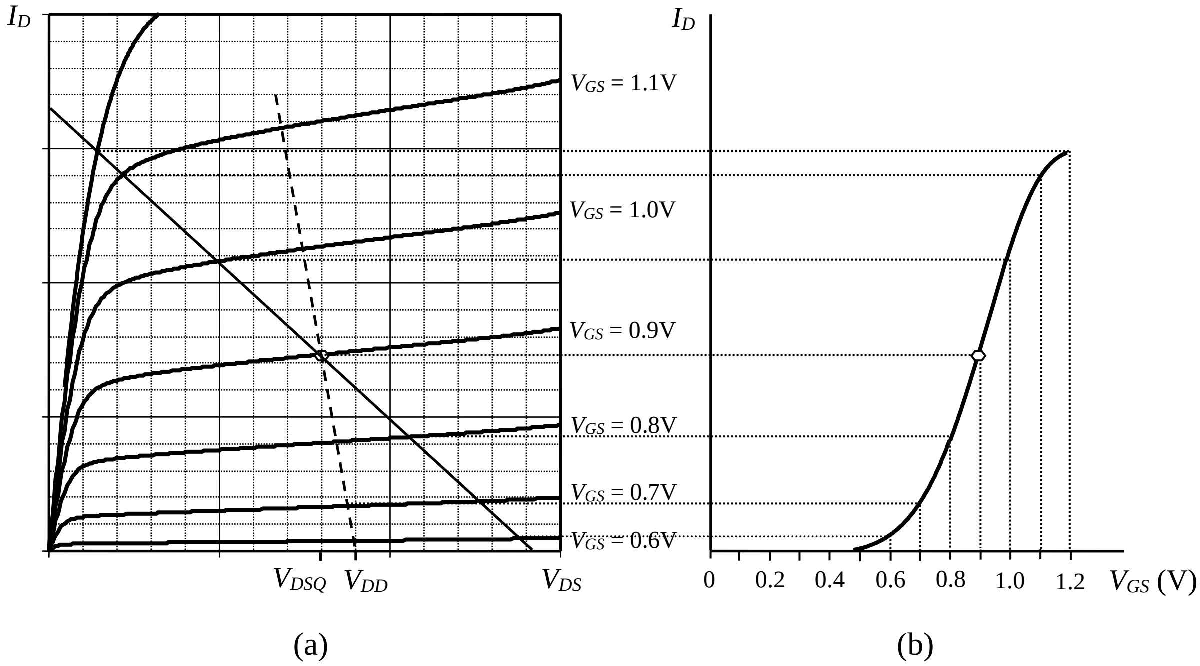


Figure 4.68 (a) The DC load line (dashed line) and the dynamic load line (solid line) on the output characteristic of a transistor. (b) The dynamic transfer characteristic corresponding to the AC load.

For a wide-band amplifier, the output voltage is $v_o = i_d R_D$ up to the 3 dB frequency of the circuit. The g_i coefficients of the series given in (4.168) depend on the shape of the dynamic transfer characteristic and each of them can be positive or negative. The first coefficient (g_1) in (4.168) is noting but the g_m of the transistor for this operating point, and others are the higher order derivatives of g_m [see references][1]. Since the magnitude of the coefficients usually decreases with increasing order, the fourth and higher order terms will be neglected to keep the expressions manageable.

To evaluate the harmonic distortion, a sinusoidal input signal voltage ($v_{gs} = v_i = V_i \cos \omega t$) must be applied to the input of the amplifier. The corresponding drain current can be solved as:

[1] See:
P. M. Jupp, D. R. Webster, "Application of Derivative Superposition to low IM3 Distortion IF Amplifiers", Roke Manor Research Ltd, 2002.
B. Razavi, "RF Microelectronics", Prentice Hall, 1998.

$$i_d = \left(\frac{1}{2}g_2V_i^2 + \frac{3}{8}g_4V_i^4 + \cdots\right)$$
$$+\left(g_1V_i + \frac{3}{4}g_3V_i^3 + \cdots\right)\cos\omega t$$
$$+\left(\frac{1}{2}g_2V_i^2 + \frac{1}{2}g_4V_i^4 + \cdots\right)\cos 2\omega t \qquad (4.169)$$
$$+\left(\frac{1}{4}g_3V_i^3 + \cdots\right)\cos 3\omega t$$
$$+\cdots$$

To evaluate the intermodulation distortion, we shall investigate the effects of two sinusoidal signals[2] with different frequencies arriving at the input of the amplifier:

$$v_{gs} = V_1\cos\omega_1 t + V_2\cos\omega_2 t\ .$$

For this case, the drain current becomes:

$$i_d = \frac{1}{2}g_2(V_1^2 + V_2^2)$$
$$+\left[g_1V_1 + \frac{3}{2}g_3(V_1V_2^2 + \frac{1}{2}V_1^3)\right]\cos\omega_1 t + \left[g_1V_2 + \frac{3}{2}g_3(V_1^2V_1 + \frac{1}{2}V_2^3)\right]\cos\omega_2 t$$
$$+\frac{1}{2}g_2V_1^2\cos 2\omega_1 t\frac{1}{2}g_2V_2^2\cos 2\omega_2 t$$
$$+\frac{1}{4}g_3V_1^3\cos 3\omega_1 t\frac{1}{2}g_3V_2^3\cos 3\omega_2 t \qquad (4.170)$$
$$+\frac{1}{2}g_2V_1V_2\cos(\omega_1+\omega_2)t + \frac{1}{2}g_2V_1V_2\cos(\omega_1-\omega_2)t$$
$$+\frac{3}{4}g_3V_1^2V_2\cos(2\omega_1+\omega_2)t + \frac{3}{4}g_3V_1V_2^2\cos(2\omega_2+\omega_1)t$$
$$+\frac{3}{4}g_3V_1^2V_2\cos(2\omega_1-\omega_2)t + \frac{3}{4}g_3V_1V_2^2\cos(2\omega_2-\omega_1)t$$

It is obvious that for $V_1 = 0$ or $V_2 = 0$ (4.170) reduces to (4.169)

From (4.169) it can be seen that:

- The drain current has acquired a DC term (the rectification term) due to the even harmonic coefficients. This term indicates a shift of the operating point.

[30] The assumption of two different signals represents the simplest case; if there are more than two signals the intermodulation occurs among all of these components.

- The coefficient of the fundamental frequency has two components; a term that increases linearly with the amplitude of the signal (Fig.4.69, line A), and a term originating from the third order nonlinearity and increases with the third power of the amplitude of the signal (Fig.4.69, curve B).

- For an S-shaped transfer curve it can be seen that the sign of the component originating from the third order nonlinearity is negative. Therefore, the actual variation of the fundamental term is equal to (A-B) on (Fig.4.69).

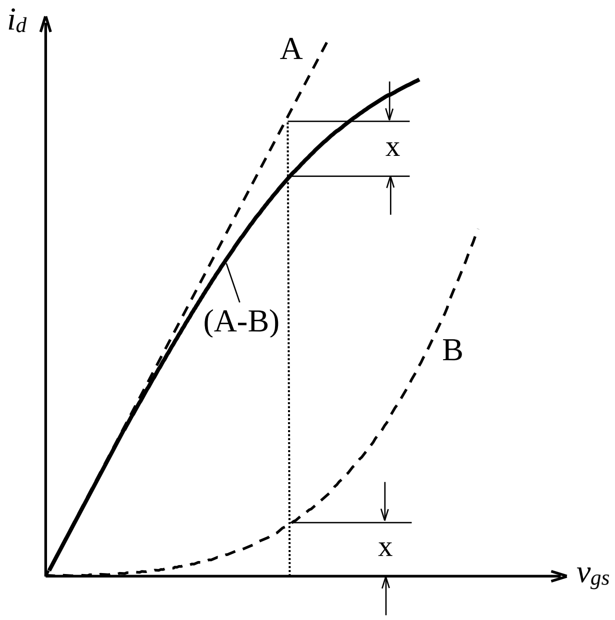


Figure 4.69 The linear term (A), and the cubic term (B) of the fundamental component of the drain current as a function of the input signal voltage. The solid line corresponds to the actual variation of the fundamental term of the drain current.

- This "saturation" of the drain current amplitude indicates that in the case of amplifying a modulated signal, the share of the term associated with the cubic term must be well below than that of the linear term in order to maintain the relative amplitudes of the carrier and side-frequencies originating from the modulation. The usually accepted criterion is not to exceed the input level corresponding to the -1dB fall of the output power which corresponds to a factor of 1.259 reduction of the output power, or factor of 1.122 reduction of the output signal current or voltage. The input voltage corresponding to this point can be calculated from

$$\frac{g_1 V_i}{g_1 V_i - \frac{3}{4}|g_3| V_i^3} = 1.122 \quad \rightarrow \quad V_i\big|_{-1\text{dB}} = 0.38\sqrt{\frac{g_1}{|g_3|}} \qquad (4.171)$$

- The coefficient of the second harmonic is composed only of the even order parameters, and the coefficient of the third harmonic is composed only of the odd order parameters. These terms are not important for tuned amplifiers, since these harmonics are normally filtered out by the output tuned circuit. But for wide band amplifiers they may be in the pass band of the amplifier.

- As already mentioned, symmetrical circuits eliminate the even harmonics. Another advantage of the elimination of the even harmonics is the elimination of the DC shift of the operating point due to the rectification term.

For (4.143), the above interpretations related to the rectification term as well as the second and the third harmonics are equally valid. The effects related to the intermodulation products can be interpreted as follows:

- The first order intermodulation products, $(\omega_1 + \omega_2)$ and $(\omega_1 - \omega_2)$ are far from ω_1 and ω_2 and will be filtered out by the output resonance circuit for tuned amplifiers.

- The second order intermodulation products, $(\omega_1 + 2\omega_2)$ and $(2\omega_1 + \omega_2)$ also are prone to be filtered out.

- The other set of the second order intermodulation products, $(2\omega_1 - \omega_2)$ and $(2\omega_2 - \omega_1)$, (we shall name them as the "close intermodulation products") are critical. If the difference between ω_1 and ω_2 is $\Delta\omega$, for example $\omega_2 = \omega_1 + \Delta\omega$, these second order intermodulation products become

$$(2\omega_1 - \omega_2) = \omega_1 - \Delta\omega$$
$$(2\omega_2 - \omega_1) = \omega_2 + \Delta\omega$$

In Fig.4.70(a) the relative positions of the fundamental components and these intermodulation products are shown for $V_1 = V_2 = V_i$. In reality ω_1 and ω_2 are modulated and have side-(modulation) bands having a certain width. Note that the side bands of $2\omega_1$ and $2\omega_2$ are twice larger than that of ω_1 and ω_2. In Fig. 4.70(b) these four frequencies are shown with their modulation side-bands. It can be easily seen that the interfering effects of the intermodulation products are more severe compared to the interfering effects of the two neighboring channels, ω_1 and ω_2.

It must be noted that in case of a wide-band LNA, all these components existing in the drain current produce corresponding output voltage components up to the 3 dB frequency of the amplifier. In this case an input filter is necessary to exclude all signals other than the carrier (and its side frequencies) intended to be received.

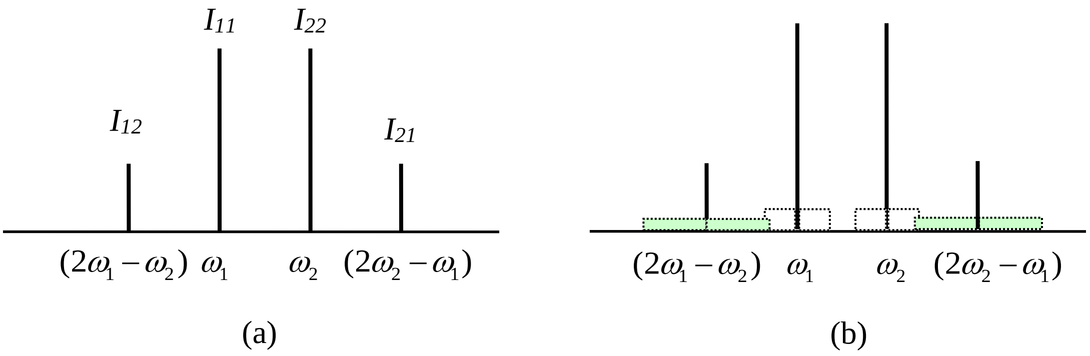


Figure 4.70 (a) The close intermodularion products of two sinusoidal signals having equal amplitude and different frequencies. (b) Illustration of the interfering effects of one of these intermodularion product on an LNA tuned to (or to the vicinity of) this frequency.

From (4.170) it can be seen that the magnitudes of the fundamentals and these second order intermodulation products are affected by g_3 and rapidly increase with the amplitudes of the ω_1 and ω_2 components, $V_1 = V_2 = V_i$, as described in the following

The magnitudes of the ω_1 and ω_2 components of the drain current:

$$I_{11} = I_{22} = g_1 V_i + \frac{9}{4} g_3 V_i^3 = g_1 V_i - \frac{9}{4} |g_3| V_i^3 \qquad (4.172)$$

The magnitudes of $(2\omega_1 - \omega_2)$ and $(2\omega_2 - \omega_1)$ components:

$$I_{12} = I_{21} = \frac{3}{4} |g_3| V_i^3 \qquad (4.173)$$

In case of a tuned amplifier, if all these components are close to each other and are in the pass band of the amplifier, the corresponding components of the output voltage become

$$V_{11} = V_{22} = R_{eff}\, g_1 V_i - \frac{9}{4} R_{eff}\, |g_3| V_i^3 \qquad (4.174)$$

and $$V_{12} = V_{21} = \frac{3}{4} R_{eff}\, |g_3| V_i^3 \qquad (4.175)$$

Therefore, the ratio of the magnitude of one of the fundamental components to the magnitude of one of the close intermodulation products can be written as

$$\frac{V_{11}}{V_{12}} = \frac{R_{eff}\, g_1 V_i - \frac{9}{4} R_{eff}\, |g_3| V_i^3}{\frac{3}{4} R_{eff}\, |g_3| V_i^3} = \frac{4}{3} \frac{g_1}{|g_3|} \frac{1}{V_i^2} - 3 \qquad (4.176)$$

The numerical value of this ratio for the input level corresponding to the -1 dB point can be calculated from (4.171) and (4.176):

$$\left.\frac{V_{11}}{V_{12}}\right|_{-1\text{dB}} = \frac{4}{3}\frac{g_1}{|g_3|}\frac{1}{V_i^2} - 3 = \frac{4}{3}\frac{V_i^2}{0.145}\frac{1}{V_i^2} - 3 = 6.195 \quad \rightarrow \quad 15.84 \text{ dB} \tag{4.177}$$

This relation provides a hint for a method to find the input level corresponding to the -1 dB point using a relatively easy intermodulation measurement.

Another metric to evaluate the nonlinearity of an amplifier is the input level corresponding to the so-called "third order intercept point", or "IP3" in short. IP3 is defined as the intercept point of the line corresponding to the linear term of one of the fundamental components of (4.170) and the line corresponding to the magnitude of one of the close intermodulation components, shown on a log-log axes or in dB scales (Fig. 4.71).

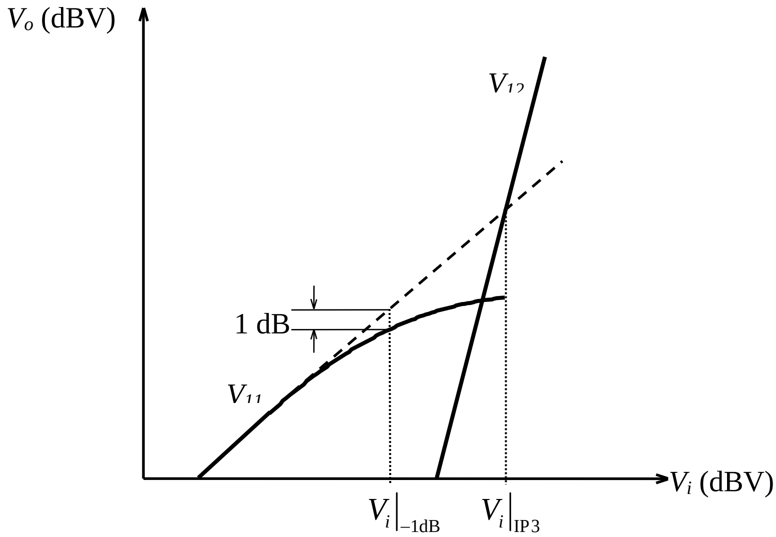


Figure 4.71 Definition of IP3, the third order intercept point.

The input levels corresponding to the IP3 can be calculated from (4.170):

$$g_1 V_i = \frac{3}{4}|g_3| V_i^3 \quad \rightarrow \quad V_i|_{\text{IP3}} = 1.154\sqrt{\frac{g_1}{|g_3|}} \tag{4.178}$$

The input voltage corresponding to the -1 dB point was calculated as:

$$V_i|_{-1\text{dB}} = 0.38\sqrt{\frac{g_1}{|g_3|}} \tag{4.179}$$

Therefore;

$$\frac{V_i|_{\text{IP3}}}{V_i|_{-1\text{dB}}} = 3.037 \quad \rightarrow \quad 9.65 \text{ dB} \tag{4.180}$$

This expression shows that the input level corresponding to the IP3 is well above the level corresponding to the -1 dB point, and practically not applicable due to the excessive nonlinearity[3].

[3] Although the intercept point lies outside the usable limits of the amplifiers, the reason behind the widespread acceptance of the IP3 as a metric of nonlinearity is the simplicity of its determination by a relatively easy measurement procedure.