4 Gain Stages — Amplifiers I

Part 4.A—Frequency response and handling poles

CMOS Common-Source with single-pole

The common-source stage allows high-gain, in the order of 100's (dependent on a large number of factors). Assume that we sweep a design parameter, such as I_{bias},

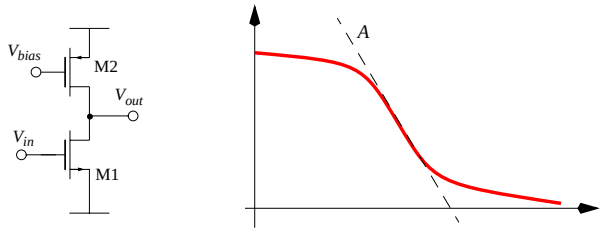


Figure 4.1: Common-source stage and its approximate transfer function.

W_1, W_2, V_{bias}, $V_{in,\,DC}$, we will achieve a characteristics similar to that in the figure above. By examining the slope of the curve (as function of any parameter) we find the gain.

$$A = \frac{\partial V_{out}}{\partial \text{ parameter}}$$

"Coarse" Large-Signal Design

We can also consider the large signal definitions (in the saturation regions)

$$I_P = \alpha_P \cdot (V_{DD} - V_{bias} - V_{TP})^2 \cdot (1 + \lambda_P \cdot (V_{DD} - V_{out})) \text{ and}$$
$$I_N = \alpha_N \cdot (V_{in} - V_{TN})^2 \cdot (1 + \lambda_N \cdot V_{out})$$

These two currents must be equal, hence we have

$$\alpha_N \cdot (V_{in} - V_{TN})^2 \cdot (1 + \lambda_N \cdot V_{out}) = \ldots$$
$$\ldots = \alpha_P \cdot (V_{DD} - V_{bias} - V_{TP})^2 \cdot (1 + \lambda_P \cdot (V_{DD} - V_{out}))$$

Assume that all values are fixed and you increase the DC input value, V_{in}. The LHS of the equation will increase. The RHS must also increase to achieve equilibrium. **The only possible way is that the output voltage decreases**. The same holds if we increase α_N, i.e., increase W_N. If we increase V_{bias} on the RHS, I_P will decrease, and to compensate the LHS, V_{out} must once again decrease.

In analog design it is important to be aware of properties such as the ones mentioned above. But we also have to care about the voltage swings, we want the transistors to be saturated and a maximum possible output swing. (Note that a maximum possible swing is not necessarily equivalent with maximum gain!). First, we have the obvious relations

$$V_{in} > V_{TN} \text{ and } V_{DD} - V_{bias} > V_{TP}$$

But we can also state that

$$V_{in} - V_{TN} < V_{out} \text{ and } V_{DD} - V_{bias} - V_{TP} < V_{DD} - V_{out} \text{ or } V_{out} < V_{bias} + V_{TP}$$

Hence $V_{out} \in [V_{in} - V_{TN}, V_{bias} + V_{TP}]$ to keep the transistors in their saturation region.

Using the current formulas, we also have that

$$V_{in} - V_{TN} \approx \sqrt{I_N / \alpha_N} \text{ and } V_{DD} - V_{bias} - V_{TP} \approx \sqrt{I_P / \alpha_P}$$

and we get

$$V_{out} \in [\sqrt{I_0 / \alpha_N}, V_{DD} - \sqrt{I_o / \alpha_P}]$$

We should also always aim for a design margin. Typically, this design margin is 0.1 - 0.15 V for CMOS.

Frequency-domain Properties

However, as we have found these large signal properties, we examine the circuit in the frequency domain we have the approximate transfer function shown in Fig. 4.2.

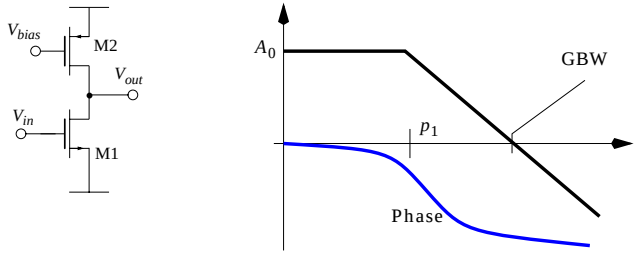


Figure 4.2: Common-source stage and its approximate transfer function.

We will assume that we have a dominating pole, e.g., the load capacitance (the total amount of capacitance associated with the output) is much larger than other capacitances. For example, we have

$$C_L = C_{db1} + C_{db2} + C_{gd2} + C_L'$$

where C_L' is the capacitance associated with following stages, wires, gates, etc.

Consider the small signal schematics in Fig. 4.3. Note the directions on the dependent current sources. We have defined the PMOS dependent source as

$$\frac{\partial I_d}{\partial V_{sg}} \cdot v_{sg} = g_{mp} \cdot v_{sg}$$

We now want to derive the output as function of the input, and we have that

$$\frac{V_{out}}{V_{in}} = -\frac{g_{mn}}{g_{dsn} + g_{dsp} + sC_L}$$

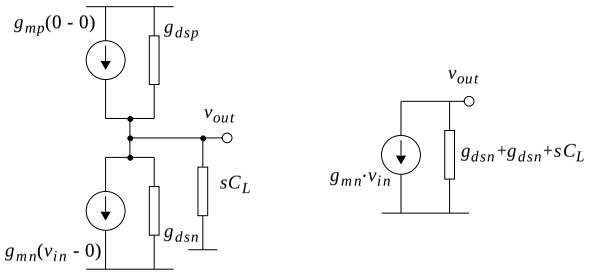


Figure 4.3: Small signal schematics of a common-source stage.

Note the negative sign and the very convenient way of using conductance instead of resistance. We are now about to compare this expression with a common notation:

$$A(s) = \frac{A_0}{1 + s/p_1}$$

We have that

$$\frac{V_{out}}{V_{in}} = -\frac{\frac{g_{mn}}{g_{dsn} + g_{dsp}}}{1 + \frac{s}{(g_{dsn} + g_{dsp})/C_L}}$$

Hence, we may identify the **DC gain** and the **dominant (-3 dB) pole**

$$A_0 = -\frac{g_{mn}}{g_{dsn} + g_{dsp}} = -\frac{g_{mn}}{g_{out}} \text{ and}$$

$$p_1 = \frac{g_{dsn} + g_{dsp}}{C_L} = \frac{g_{out}}{C_L}$$

Notice the definition of A_0, the negative sign is incorporated as well. We may also find the **unity-gain frequency** given by the expression

$$|A(j\omega_u)| = 1 \text{ (or where is the gain equal to 1?)}$$

Using the expressions above, we have that

$$|A(j\omega_u)|^2 = \frac{A_0^2}{1 + \omega_u^2/p_1^2} = 1 \text{ and } \omega_u^2 = (A_0^2 - 1) \cdot p_1^2$$

Further we assume that A_0 is large and we have that

$$\omega_u \approx A_0 \cdot p_1$$

$A_0 \cdot p_1$ is sometimes also referred to as the gain-bandwidth product (GBW). However, the GBW is not always equal to the ω_u.

Using our knowledge about the circuit it-self, we have that

$$\omega_u = \frac{g_{mn}}{g_{out}} \cdot \frac{g_{out}}{C_L} = \frac{g_{mn}}{C_L}$$

We also define the **phase margin**. It is the difference between -180 (or -360 if the amplifier is inverting–as in our case) degrees and the phase of the system at the unity-gain frequency. This is related to feedback and stability issues that we will come back to later on. For a single-pole system with a high-frequency ω_u, we have that it is 90 degrees, since

$$\arg\{A(j\omega_u)\} \approx \arg\left\{\frac{A_0}{j\omega_u/p_1}\right\} = \arg\left\{\frac{-\omega_u}{j\omega_u}\right\} = -270°$$

Comparing this result with -360, we have a phase margin of 90 degrees which is stable. We want to guarantee as high phase margin as possible and preferably higher than 45-60 degrees.

We may also identify some other issues concerning the gain, bandwidth, and unity-gain. Consider Table 1 and Fig. 4.4. Notice that the load capacitance is not a design variable. Mostly the channel length is also set to a specific value or at least equally large for the whole circuit With relaxed design parameter, we only have the widths and the bias current. (The DC voltages of the input and the bias should also be taken in account of course, but they are not as visible in the small-signal schematics).

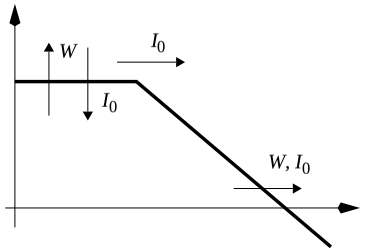


Figure 4.4: Design issues for a common-source stage.

Parameter	Small-signal parameters	Design parameters	Relaxed design parameters
A_0	$\frac{g_{mn}}{g_{dsn} + g_{dsp}}$	$\frac{\sqrt{\frac{W}{L}} \cdot \sqrt{I_0}}{I_0} \sim \frac{\sqrt{\frac{W}{L}}}{\sqrt{I_0}}$	$\frac{\sqrt{W}}{\sqrt{I_0}}$
p_1	$\frac{g_{dsn} + g_{dsp}}{C_L}$	I_0	I_0
ω_u	$\frac{g_{mn}}{C_L}$	$\sqrt{\frac{W}{L}} \cdot \sqrt{I_0}$	$\sqrt{W} \cdot \sqrt{I_0}$

Table 1: Design parameters.

Two-pole system

As we can see, the phase margin is always large for a single-pole system, but this is not the case for a dual-pole system. In analog design, we always have to also consider the small additional capacitances that are hidden somewhere in the mess.

Assume that we have the system

$$A(s) = \frac{A_o}{(1 + s/p_1) \cdot (1 + s/p_2)}$$

The phase margin will now become lower than 90 degrees and if we use a feedback the output will swing more when we apply an input step signal.

However, mostly we consider the case of a dominant pole at the first stages of the design. Then we use the simulator (and some hand calculations) to find the phase margin. We will get back to this when we discuss operational amplifier.

Settling

If we have a single-pole system (dominant pole) and apply a step signal at the input, $V_{in} = V_{IN} + v_{step}$, the output will describe an exponential waveform:

$$v_{out}(t) = V_{OUT} - v_{step} \cdot A_0 \cdot (1 - e^{-t/\tau})$$

where V_{OUT} is the DC output voltage and v_{step} is the input voltage difference, τ is the time constant, hence $\tau = 1/p_1$, and A_0 is the DC gain. The dominant pole is given by the output conductance and the load capacitance.

This phenomenon is referred to as linear settling and we should always design for linear settling. However, it may not be possible to achieve linear settling **if the input voltage step is to large**.

Any change of the output voltage is limited by the time it takes to charge/discharge the capacitance at the output. However, this "theory" also depends on that we have an infinite current available. This is naturally not the case. The maximum current that can be used for charging/discharging the load capacitance is limited by the bias current through the stage. The "speed" to charge the output is given by the time derivative of the output voltage

$$\frac{\partial v_{out}}{\partial t} = \frac{i_{out}}{C_L} \text{ (i.e., definition of the voltage over a capacitor)}$$

We want to consider the maximum possible derivative, and it must be given by

$$\max \left| \frac{\partial v_{out}}{\partial t} \right| = \max \left| \frac{i_{out}}{C_L} \right| = \frac{\max |i_{out}|}{C_L} = \frac{I_0}{C_L} = \text{SR}$$

which is the slew rate (SR). Basically, these two cases can be sketched as in Fig. 4.5.

It is interesting to find when there is slew-rate limiting. Compare the derivative of the linear settling with the SR.

$$\left| \frac{\partial v_{out}}{\partial t} \right| = \left| \frac{\partial v_{step} \cdot A_0 \cdot (1 - e^{-t/\tau})}{\partial t} \right| = v_{step} \cdot A_0 \cdot \frac{1}{\tau} \cdot e^{-\frac{t}{\tau}} = v_{step} \cdot A_0 \cdot p_1 \cdot e^{-t/\tau}$$

which is equal to

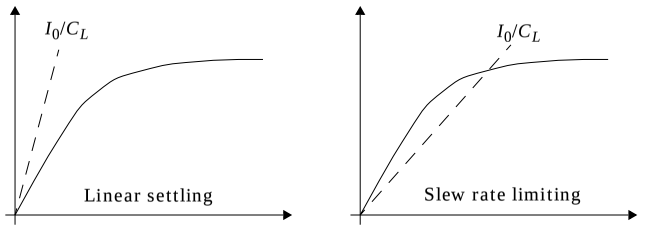


Figure 4.5: Illustration of linear and non-linear settling.

$$\left| \frac{\partial v_{out}}{\partial t} \right| = v_{step} \cdot \omega_u \cdot e^{-t/\tau} = \frac{v_{step} \cdot g_m}{C_L} \cdot e^{-t/\tau}$$

Notice, that the maximum value of this derivative (see Fig. 4.5) is found for $t = 0$. So, if we do not want slew-rate limiting, we must guarantee that the SR is much higher than the derivative of the linear settling, hence consider for $t = 0$

$$\text{SR} = \frac{I_0}{C_L} > \frac{v_{step} \cdot g_m}{C_L} \Rightarrow I_0 > v_{step} \cdot g_m$$

or we can use any of the other ways to express g_m, ω_u, etc. The equation above states that; the higher g_m we design for, the higher must the current be or the lower must the input step voltage be.

Another interesting value is the time-point at where the non-linear settling is changing into linear settling. This is easily found and we have that

$$\frac{I_0}{C_L} = \frac{v_{step} \cdot g_m}{C_L} \cdot e^{-\frac{T_{NL}}{\tau}} \Rightarrow T_{NL} = \frac{1}{p_1} \cdot \ln \frac{v_{step} \cdot g_m}{I_0}$$

If we have a switched system (discrete-time, e.g., switched-capacitor) these issues are very important. We have to design our circuits so that the non-linear settling can be eliminated during one clock cycle, or similar design requirement.

Part 4.B—Feedback

The phase margin is very important when using the circuits in feedback. This is actually also how we should use them. Unless you have a comparator, high gain is never attractive unless you use the high-gain amplifier in a feedback configuration to achieve a circuit with a lower gain, but with a higher bandwidth.

Basically, we want to avoid positive feedback since then a change at the input will amplify it self and the system will saturate. With negative feedback this is not really the case. However, we can allow positive feedback as long as the forward gain is less than zero (almost correct anyway). Therefore we use the parameters; phase margin and unity-gain frequency, e.g., how close are we to positive feedback?

Consider the feedback in Fig. 4.6. Assume that A is a single-pole system and that β does not have any pole.

Then we have that

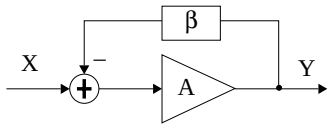


Figure 4.6: Feedback configuration.

$$\frac{Y}{X} = \frac{A(s)}{1+\beta\cdot A(s)} = \frac{\dfrac{A_0}{1+s/p_1}}{1+\beta\cdot\dfrac{A_0}{1+s/p_1}} = \frac{A_0}{1+\beta\cdot A_0+s/p_1} =$$

$$= \frac{\dfrac{A_0}{1+\beta\cdot A_0}}{1+\dfrac{s}{p_1\cdot(1+\beta\cdot A_0)}} = \frac{1}{\beta}\cdot\frac{\dfrac{1}{1+1/(\beta\cdot A_0)}}{1+\dfrac{s}{\beta\cdot A_0\cdot p_1\cdot 1+1/(\beta\cdot A_0)}} =$$

$$= \frac{1}{\beta}\cdot\frac{\dfrac{1}{1+1/(\beta\cdot A_0)}}{1+\dfrac{s}{\beta\cdot\omega_u}\cdot\dfrac{1}{1+1/(\beta\cdot A_0)}} \approx \frac{1}{\beta}\cdot\frac{1}{1+\dfrac{s}{\beta\cdot\omega_u}}$$

Hence, once we feedback, the gain is determined by the feedback factor, $1/\beta$, and the bandwidth is significantly increased to $\beta\cdot\omega_u$.

However, we see that if A_0 is not large enough there is a small deviation from the wanted $1/\beta$. This is crucial if we want to make a high-performance filter or similar where the requirements on the acuracy are very high.

Part 4.C—Improved gain in amplifiers

To improve the gain of the common-source stage we shall consider some different approaches.

Brute force

Consider the expression for the gain

$$A = \frac{g_m}{g_{out}} \sim \frac{\sqrt{\dfrac{W}{L}}\cdot\sqrt{I_d}}{I_d} \sim \frac{\sqrt{\dfrac{W}{L}}}{\sqrt{I_d}}$$

1) Decrease the current through the transistors. This will increase the gain, but it will reduce the unity-gain frequency and it will make the circuit more sensitive to noise.

2) Increase the size of the transistor. This will increase the gain and it will increase the unity gain frequency. However, the capacitance is increasing by WL at the input of the amplifier and it may be hard to achieve a good phase margin in the overall feedback, since this capacitance will reflect as a load capacitance in a feedback configuration.

Cascodes

Instead, consider the gain function again

$$A = \frac{g_m}{g_{out}} \text{ and } \omega_u = \frac{g_m}{C}$$

We see that if we decrease the g_{out} hence increase r_{out} we have that the unity-gain frequency is not effected. So, the first approach to increase the gain is to use cascodes. Consider Fig. 4.7 (a) .

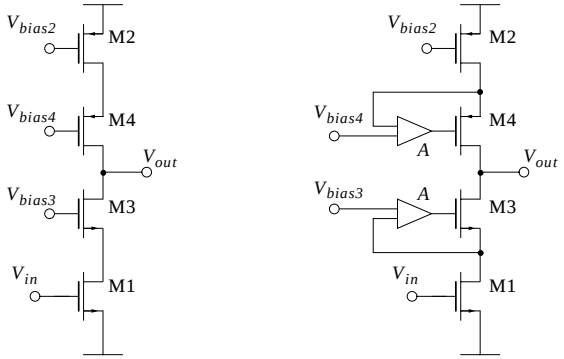


Figure 4.7: (a) Cascodes M3 and M4 are used to increase the output impedance and the DC gain. The gain (b) is increased further with gain-boosting (or regulated-cascode).

The output resistance is increased by a factor corresponding to the gain of the cascode transistor. This can be realized by considering the small signal schematics of the NMOS transistors only.

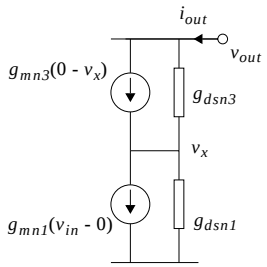


Figure 4.8: Small-signal schematics of cascoded NMOS transistor.

The output impedance is found as

$$r_{out} = \left.\frac{v_{out}}{i_{out}}\right|_{v_{in}=0}$$

Using Kirchhoff's laws we find that

$$i_{out} = g_{dsn1} \cdot v_x \text{ and}$$
$$i_{out} = g_{mn3} \cdot (-v_x) + g_{dsn3} \cdot (v_{out} - v_x)$$

This gives that

$$r_{out} = \frac{g_{mn3} + g_{dsn3} + g_{dsn1}}{g_{dsn3} \cdot g_{dsn1}} \approx \frac{g_{mn3}}{g_{dsn3}} \cdot \frac{1}{g_{dsn1}} = A_3 \cdot r_{out1}$$

However, since we are using more transistors, the voltage swing will be reduced (in order to keep all transistors in their saturation region). We have that

$$V_{out} \in \left[\sqrt{\frac{I_0}{\alpha_{N1}}} + \sqrt{\frac{I_0}{\alpha_{N3}}}, V_{DD} - \sqrt{\frac{I_0}{\alpha_{P2}}} - \sqrt{\frac{I_0}{\alpha_{P4}}} \right]$$

By adding even more cascodes the output resistance can be increased further. However, the voltage swing limits the concept.

Gain-boosting

Instead of using multiple cascodes we can use the concept of gain-boosting or regulated-cascode as shown in Fig. 4.7(b). It can easily be shown that the output impedance of the NMOS branch is approximately

$$r_{out} \approx A \cdot A_3 \cdot r_{out1}$$

Folded-cascode

We loose some in voltage level due to the cascodes, and it is hard to design the circuit to have the same output as input DC voltage levels. The folded-cascode common-source stage solves this problem. However, the disadvantage is obvious, there is now

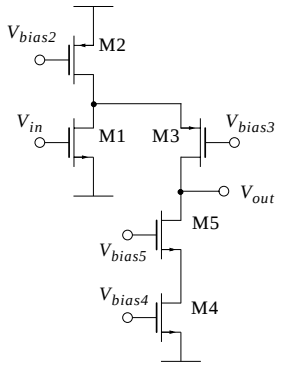


Figure 4.9: Folded-cascode gain stage.

a PMOS transistor in the signal path and the increased parasitic capacitance will influence the speed.

Part 4.D—Noise

The influence of noise is important. It is a fundamental limit on performance. We will in this course only consider the thermal noise and the simplest transistor models associated with this.

First, we have the definition of the signal-to-noise ratio (SNR) as the ratio between the signal and noise power.

$$\text{SNR} = 10 \cdot \log \frac{P_{sig}}{P_{noise}}$$

For our applications we aim for a 14-bit SNR, hence an SNR higher than 86 dB.

As model for our transistors, we only consider uncorrelated thermal noise and we have two cases as shown in Table 2.

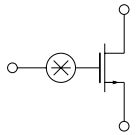
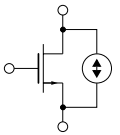
Input-referred noise	Output-referred noise
	
$v_i^2(f) = S_i(f) = \frac{8}{3}kT \cdot g_m$	$i_o^2(f) = S_o(f) = \frac{8}{3}kT \cdot \frac{1}{g_m}$

Table 2: Thermal noise sources on the CMOS transistor.

Spectral density and super function

Recapture from the statistics. The spectral density function describes the power density of a signal throughout the frequency range.

$$S_n(f) = v_n^2(f)$$

If the noise spectral density is constant, the noise is white.

Noise power

The total noise power is given by the integral of the spectral density over a specifed frequency band.

$$P_{noise} = \int_{f_1}^{f_2} S_n(f) df$$

For a sampled system, these frequencies are typically $f_1 = 0$ and $f_2 = f_s/2$ where f_s is the sample frequency.

Several noise sources

The contribution from several noise sources will sum at the output node. We may use the additional theorem and we get

$$S_{out}(f) = |H_{10}(f)|^2 \cdot S_1(f) + |H_{20}(f)|^2 \cdot S_2(f) + \ldots$$

where H_{10} is the transfer function from the noise source described by S_1 to the output, etc.

The superfunction describes the relation between the output and input signals in a linear(ized) system.

The total output noise power is given by

$$P_{out} = \int_0^\infty S_{out}(f)df = \int_0^\infty |H_{10}(f)|^2 \cdot S_1(f)df + \ldots$$

Noise bandwidth

Assume that the filtering function H_{ij} is a single-pole system and that only thermal noise is considered. Then we have

$$\int_0^\infty \frac{A_0^2}{|1 + j2\pi f/p_1|^2} \cdot S_{10}df = S_{10} \cdot A_0^2 \cdot \int_0^\infty \frac{1}{1 + (2\pi f)^2/p_1^2} \cdot df =$$

$$= S_{10} \cdot A_0^2 \cdot \frac{p_1}{2\pi} \cdot \int_0^\infty \frac{1}{1 + x^2} \cdot df = S_{10} \cdot A_0^2 \cdot \frac{p_1}{2\pi} \cdot \frac{\pi}{2} = S_{10} \cdot A_0^2 \cdot \frac{p_1}{4}$$

A noisy amplifier can be modelled in two different ways, either as a noiseless amplifier with a noise source at the input or with a noise source at the output.

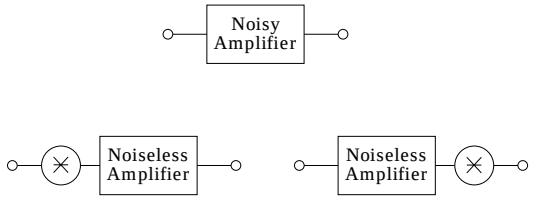


Figure 4.10: Models of input-referred and output-referred noise sources.

Noisy common-source stage

Consider the common-source stage in Fig. 4.1. Assume a dominant load capaitance and that both transistors are noisy. Derive the total noise power at the output and the input-referred noise voltage source.

We have already derived the transfer function from the gate of transitor M1 to the output;

$$H_{10}(s) = \frac{g_{mn}/g_{out}}{1 + s/(g_{out}/C_L)}$$

It is obvious that the transfer function from the gate of M2 must be equal to this expression but with a different g_m, hence

$$H_{20}(s) = \frac{g_{mp}/g_{out}}{1 + s/(g_{out}/C_L)}$$

The spectral densities of the noise at M1 and M2 are given by

$$S_1(f) = \frac{8}{3}kT \cdot \frac{1}{g_{mn}} \text{ and } S_2(f) = \frac{8}{3}kT \cdot \frac{1}{g_{mp}}$$

Then we can use the superfunction

$$S_{out}(f) = |H_{10}(f)|^2 \cdot S_1(f) + |H_{20}(f)|^2 \cdot S_2(f) =$$

$$= \left|\frac{g_{mn}/g_{out}}{1 + j2\pi f/(g_{out}/C_L)}\right|^2 \cdot \frac{8}{3}kT \cdot \frac{1}{g_{mn}} + \left|\frac{g_{mp}/g_{out}}{1 + j2\pi f/(g_{out}/C_L)}\right|^2 \cdot \frac{8}{3}kT \cdot \frac{1}{g_{mp}}$$

The total noise power at the output (over all frequencies) is given by

$$P_{out} = \int_0^\infty S_{out}(f)df$$

and since both H_{10} and H_{20} are described by single-pole systems, we can use the concept of noise bandwidth:

$$P_{out} = \left(\frac{g_{mn}}{g_{out}}\right)^2 \cdot \frac{g_{out}/C_L}{4} \cdot \frac{8}{3}kT \cdot \frac{1}{g_{mn}} + \left(\frac{g_{mp}}{g_{out}}\right)^2 \cdot \frac{g_{out}/C_L}{4} \cdot \frac{8}{3}kT \cdot \frac{1}{g_{mp}} =$$

$$= \frac{g_{mn}}{g_{out}} \cdot \frac{1}{C_L} \cdot \frac{2}{3}kT + \frac{g_{mp}}{g_{out}} \cdot \frac{1}{C_L} \cdot \frac{2}{3}kT = \frac{g_{mn} + g_{mp}}{g_{out}} \cdot \frac{1}{C_L} \cdot \frac{2}{3}kT$$

Input-referred noise is easily derived by comparing the total output noise spectral density and we have

$$S_{in}(f) = \frac{S_{out}(f)}{|H_{10}(f)|^2} = S_1(f) + \frac{|H_{20}(f)|^2}{|H_{10}(f)|^2} \cdot S_2(f) =$$

$$= \frac{8}{3}kT \cdot \frac{1}{g_{mn}} + \frac{g_{mp}^2}{g_{mn}^2} \cdot \frac{8}{3}kT \cdot \frac{1}{g_{mp}} = \frac{8}{3}kT \cdot \frac{1}{g_{mn}} \cdot \left(1 + \frac{g_{mp}}{g_{mn}}\right)$$