THE DESIGN OF NARROWBAND CMOS RF LOW-NOISE AMPLIFIERS

Thomas H. Lee
Stanford University Center for Integrated Systems
Stanford, California, USA

ABSTRACT

General conditions for minimizing the noise figure of any linear two-port are reviewed before considering the specific case of a MOSFET low-noise amplifier (LNA). It is shown that the minimum noise figure cannot be obtained over an arbitrarily large bandwidth with networks of low order. For narrowband operation, however, one may construct simple amplifiers whose noise figure and power gain are close to the theoretical optima allowed within an explicit power constraint, and which simultaneously present a specified impedance to the driving source. The effects of overlap (drain-gate) capacitance, short-channel carrier heating, substrate resistance ("epi noise"), and gate interconnect resistance are also considered. Amplifier noise figures of 1.5dB or better at 10mW are achievable in the 1-2GHz range with 0.5μm technology, and improve with scaling.

1. INTRODUCTION

That the signals delivered by the antenna in modern wireless systems can be in the submicrovolt range underscores the acute need for low-noise amplification. Furthermore, the transfer characteristics of filters interposed between the antenna and LNA are frequently quite sensitive to the quality of the terminations. Thus, the design problem is often compounded by the additional requirement that the amplifier exhibit a specified (and usually real) input impedance. It should seem intuitively reasonable that it is difficult to satisfy this last condition with an inherently capacitive device such as a MOSFET, if broadband low-noise oper-

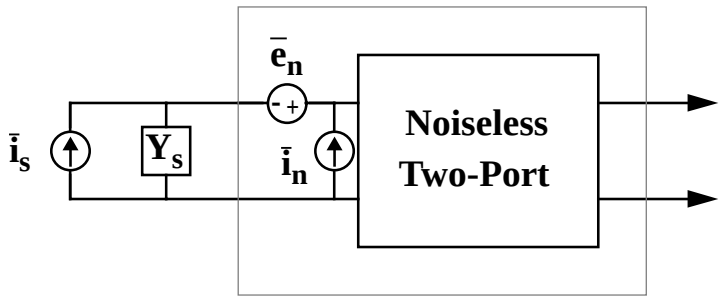
ation is required. This paper will first establish a theoretical foundation for that intuition, then explore a narrowband topology that provides near-optimum gain and noise figure while providing a real input resistance. Addition of an explicit power consumption design constraint to "classical" noise optimization will be seen to lead to a definite optimum device width. The effect of drain-gate overlap capacitance will be considered, as will the degradation in noise figure arising from short-channel effects and the thermal noise of both the substrate and gate interconnect material.

2. CLASSICAL NOISE OPTIMIZATION

What we term "classical" noise optimization begins with the assumption that one is given a linear two-port whose characteristics are fixed. The classical approach then yields the optimum source impedance (from a noise performance viewpoint).

A noisy two-port may be modeled as a noiseless two-port to which a noise voltage and noise current are connected:

FIGURE 1. Noisy two-port (source model shown outside dashed boundary)



The noise factor for this model is then:

$$F = \frac{\overline{i_s^2} + \overline{\left| i_n + Y_s e_n \right|^2}}{i_s^2} \tag{1}$$

Because e_n and i_n model the noise due to all of the sources within the original two-port, they may be correlated. Thus, express i_n as the sum of two currents: i_c, which is correlated with e_n, and i_u, which is not:

$$i_n = i_c + i_u \tag{2}$$

Because i_c is fully correlated with e_n, they are proportional:

$$i_c = Y_c e_n \tag{3}$$

The *correlation admittance* Y_c is not necessarily related to a measurable input admittance.

Using the foregoing definitions, the noise factor may be expressed as follows:

$$F = 1 + \frac{G_u + \left|Y_c + Y_s\right|^2 R_n}{G_s} = 1 + \frac{G_u + \left[\left(G_c + G_s\right)^2 + \left(B_c + B_s\right)^2\right] R_n}{G_s} \tag{4}$$

where we have explicitly decomposed each admittance into a sum of a conductance G and a susceptance B, and where the following traditional substitutions have been made:

$$R_n \equiv \frac{\overline{e_n^2}}{4kT\Delta f}, G_u \equiv \frac{\overline{i_u^2}}{4kT\Delta f}, G_s \equiv \frac{\overline{i_s^2}}{4kT\Delta f} \tag{5}$$

The noise factor is a minimum when the following conditions hold:

$$B_s = -B_c = B_{opt} \tag{6}$$

$$G_s = \sqrt{\frac{G_u}{R_n} + G_c^2} = G_{opt} \tag{7}$$

Hence, when the source and correlation susceptances are algebraic inverses, and the source conductance is equal to the value in Eqn. 7, the following minimum noise factor is achieved:

$$F_{min} = 1 + 2R_n\left[G_{opt} + G_c\right] = 1 + 2R_n\left[\sqrt{\frac{G_u}{R_n} + G_c^2} + G_c\right] \tag{8}$$

In general, the noise factor may be expressed as:

$$F = F_{min} + \frac{R_n}{G_s}\left[\left(G_s - G_{opt}\right)^2 + \left(B_s - B_{opt}\right)^2\right] \tag{9}$$

Contours of constant noise factor are circles centered about (G_{opt}, B_{opt}) in the admittance plane (and also on a Smith chart, because the bilinear transformation that maps the two preserves circles).

Although minimizing the noise factor is qualitatively similar to maximizing power transfer, the source admittances leading to these two conditions are almost never the same, as the correlation and input admittances, for example, are rarely equal (except by coincidence). Therefore, one generally cannot enjoy maximum power gain and minimum noise figure simultaneously [1].

3. MOSFET TWO-PORT NOISE PARAMETERS

The MOSFET noise model has two sources arising from thermal fluctuations of channel charge. First, there is a drain noise current source:

$$\overline{i_{nd}^2} = 4kT\gamma g_{d0}\Delta f \tag{10}$$

where g_{d0} is g_{ds} evaluated at zero V_{DS}, and γ is unity in triode and 2/3 in saturation, at least in the long-channel limit. Some older references use g_m, but g_{d0} is better related to channel charge in short-channel devices.

A gate noise current (unmodeled by SPICE) also flows, from channel potential fluctuations coupling capacitively into the gate terminal:

$$\overline{i_{ng}^2} = 4kT\delta g_g \Delta f \tag{11}$$

where δ is twice γ in long devices, and

$$g_g = \frac{\omega^2 C_{gs}^2}{5g_{d0}} \tag{12}$$

The gate and drain noise currents have a correlation coefficient c of $j0.395$ for long-channel devices. The value of c in the short-channel regime is presently unknown.

In what follows, we neglect C_{gd} to simplify the derivation. The primary effect of C_{gd} (in the cascoded designs we will consider) is on the input impedance. To derive the four noise parameters, first short-circuit the input port. Reflect the drain current noise back to the input as a noise voltage and recognize that the ratio of these quantities is simply g_m. Thus,

$$\overline{e_n^2} = \frac{\overline{i_{nd}^2}}{g_m^2} = \frac{4kT\gamma g_{d0}\Delta f}{g_m^2} \tag{13}$$

from which it is apparent that the equivalent input noise voltage is completely correlated, and in phase, with the drain current noise. Thus, we can immediately determine that

$$R_n \equiv \frac{\overline{e_n^2}}{4kT\Delta f} = \frac{\gamma g_{d0}}{g_m^2} \tag{14}$$

The equivalent input noise voltage by itself does not fully account for the drain noise, however, because a noisy drain current also flows even when the input is *open*-circuited (and even if we additionally ignore induced gate current). Under this open-circuit condition, dividing the drain current noise by the transconductance yields an equivalent input voltage which, when multiplied in turn by the input admittance, gives us the value of an equivalent input current noise that completes the modeling of i_{nd}:

$$\overline{i_{n1}^2} = \frac{\overline{i_{nd}^2}\left(j\omega C_{gs}\right)^2}{g_m^2} = \frac{4kT\gamma g_{d0}\Delta f\left(j\omega C_{gs}\right)^2}{g_m^2} = \overline{e_n^2}\left(j\omega C_{gs}\right)^2 \tag{15}$$

where we have assumed that the input admittance of a MOSFET is purely capacitive. This assumption is satisfied well below ω_T and if appropriate high-frequency layout practice is observed to minimize gate resistance. Given this assumption, Eqn. 15 shows that the input noise current i_{n1} is

in quadrature, and therefore completely correlated, with the equivalent input noise voltage e_n.

The total equivalent input current noise is the sum of the reflected drain noise contribution of Eqn. 15 and the induced gate current noise. The induced gate noise current itself consists of two terms, i_{ngc}, and i_{ngu}, which are fully correlated and uncorrelated, respectively, with the drain current noise. Hence, the correlation admittance is:

$$Y_c = j\omega C_{gs} + \frac{i_{ngc}}{e_n} = j\omega C_{gs} + \frac{g_m}{i_{nd}} \cdot i_{ngc} = j\omega C_{gs} + g_m \cdot \frac{i_{ngc}}{i_{nd}} \tag{16}$$

which may ultimately be expressed as:

$$Y_c = j\omega C_{gs} + g_m \cdot c\sqrt{\frac{\delta\omega^2 C_{gs}^2}{5\gamma g_{d0}^2}} = j\omega C_{gs} + \frac{g_m}{g_{d0}} \cdot c\sqrt{\frac{\delta}{5\gamma}} \cdot \omega C_{gs} \tag{17}$$

If c continues to be purely imaginary, even in the short-channel regime, we finally have:

$$Y_c = j\omega C_{gs} + j\omega C_{gs}\frac{g_m}{g_{d0}} \cdot |c|\sqrt{\frac{\delta}{5\gamma}} = j\omega C_{gs}\left(1 + \alpha|c|\sqrt{\frac{\delta}{5\gamma}}\right) \tag{18}$$

where

$$\alpha = \frac{g_m}{g_{d0}} \tag{19}$$

Since α is unity for long-channel devices, and progressively decreases as channel lengths shrink, it is one measure of the departure from the long-channel regime.

Eqn. 18 shows that the correlation admittance is purely imaginary, so that $G_c = 0$. More significant, however, is the fact that Y_c does not equal the admittance of C_{gs}, although it is proportional it. Hence, the MOSFET is typical in that one cannot optimize power transfer and noise figure simultaneously. To explore further the important implications of this observation, we derive the last noise parameter, G_u.

Using the definition of the correlation coefficient, we may express the induced gate noise as follows:

$$\overline{i_{ng}^2} = \overline{\left(i_{ngc} + i_{ngu}\right)^2} = 4kT\Delta f\delta g_g|c|^2 + 4kT\Delta f\delta g_g\left(1-|c|^2\right) \tag{20}$$

The very last term in Eqn. 20 is the uncorrelated portion of the gate noise current, so that, finally:

$$G_u \equiv \frac{\overline{i_u^2}}{4kT\Delta f} = \frac{4kT\Delta f\delta g_g\left(1-|c|^2\right)}{4kT\Delta f} = \frac{\delta\omega^2 C_{gs}^2\left(1-|c|^2\right)}{5g_{d0}} \tag{21}$$

The four noise parameters are summarized in the following table:

TABLE 1. Equivalent two-port noise parameters for MOSFET

Parameter	Expression
G_c	≈ 0
B_c	$\omega C_{gs}\left(1+\alpha\lvert c\rvert\sqrt{\frac{\delta}{5\gamma}}\right)$
R_n	$\frac{\gamma g_{d0}}{g_m^2} = \frac{\gamma}{\alpha}\cdot\frac{1}{g_m}$
G_u	$\frac{\delta\omega^2 C_{gs}^2\left(1-\lvert c\rvert^2\right)}{5g_{d0}}$

With these parameters, we can determine the source impedance that minimizes the noise figure:

$$B_{opt} = -B_c = -\omega C_{gs}\left(1+\alpha|c|\sqrt{\frac{\delta}{5\gamma}}\right) \tag{22}$$

$$G_{opt} = \sqrt{\frac{G_u}{R_n} + G_c^2} = \alpha\omega C_{gs}\sqrt{\frac{\delta}{5\gamma}\left(1 - |c|^2\right)} \tag{23}$$

The optimum source susceptance is thus inductive in character, but has a capacitive frequency variation. Furthermore, the optimum source conductance varies linearly with frequency. Synthesizing a network to provide these characteristics over a large frequency range is challenging to say the least, so that achieving a broadband noise match to a MOSFET is fundamentally difficult.

Whenever a noise match is achieved, the corresponding minimum noise factor is:

$$F_{min} = 1 + 2R_n\left[G_{opt} + G_c\right] \approx 1 + \frac{2}{\sqrt{5}}\frac{\omega}{\omega_T}\sqrt{\gamma\delta\left(1 - |c|^2\right)} \tag{24}$$

Note that if there were no gate current noise, or *if the gate and drain current noise sources were fully correlated*, the minimum noise figure would be 0dB. Contrary to widely held beliefs, the mere presence of gate or drain noise by itself does not necessarily impose a fundamental noise figure penalty. MOS amplifiers exhibit nonzero minimum noise figures because of the existence of gate and drain noise currents that are uncorrelated with each other.

Improvements in ω_T that accompany scaling improve the noise figure at any given frequency. To illustrate this point, let us assign numerical values to the parameters in Eqn. 24. Because the behavior of some of these parameters in the short channel regime is unknown, we will make some pessimistic guesses to arrive at conservative estimates of F_{min}. Measurements of γ reveal that it can be 2-3 times larger in short devices than predicted by long-channel theory. Values for δ in the short channel regime have never been reported, unfortunately, so we will assume that it is also augmented by a factor of two to three. Because one mechanism, thermally-driven channel charge fluctuations, gives rise to both gate and drain noise, this last assumption appears reasonable. Finally, assume that $|c|$ remains equal to 0.395. The following table shows F_{min} as a function

of frequency (normalized to g_m/C_{gs}) if short channel effects cause a pessimistic tripling of γ and δ:

TABLE 2. Estimated F_{min} (γ = 2, δ = 4)

$g_m/\omega C_{gs}$	F_{min} (dB)
20	0.5
15	0.6
10	0.9
5	1.6

Clearly, excellent noise figures are possible for MOSFETs, even with increased γ and δ.

4. SECOND-ORDER EFFECTS

There are two additional noise sources that ought to be considered in any detailed LNA design. These are the thermal noise of the substrate (epi noise) and of the gate interconnect. Epi noise can be treated as primarily equivalent to increasing the effective value of γ [2]:

$$\frac{\overline{i_{nd}^2}}{\Delta f} = 4kT\left(\gamma g_{d0} + g_{mb}^2 R_{epi}\right) = 4kTg_{d0}\left(\gamma + \frac{g_{mb}^2 R_{epi}}{g_{d0}}\right) \quad (25)$$

The quantity in the last set of parentheses is the effective value of γ. Using typical values, one finds that the increase in γ is usually below the range of 0.1 to 0.2 (i.e., about 10%), and is thus generally negligible by itself. Furthermore, although epi noise also gives rise to a noisy gate current (whose magnitude can sometimes exceed that of the fundamental induced gate noise), this term is *fully correlated* with the epi-induced drain noise, so substrate resistance has only a minor effect on the fundamental limits of noise performance. The liberal use of nearby substrate taps is always helpful in any case.

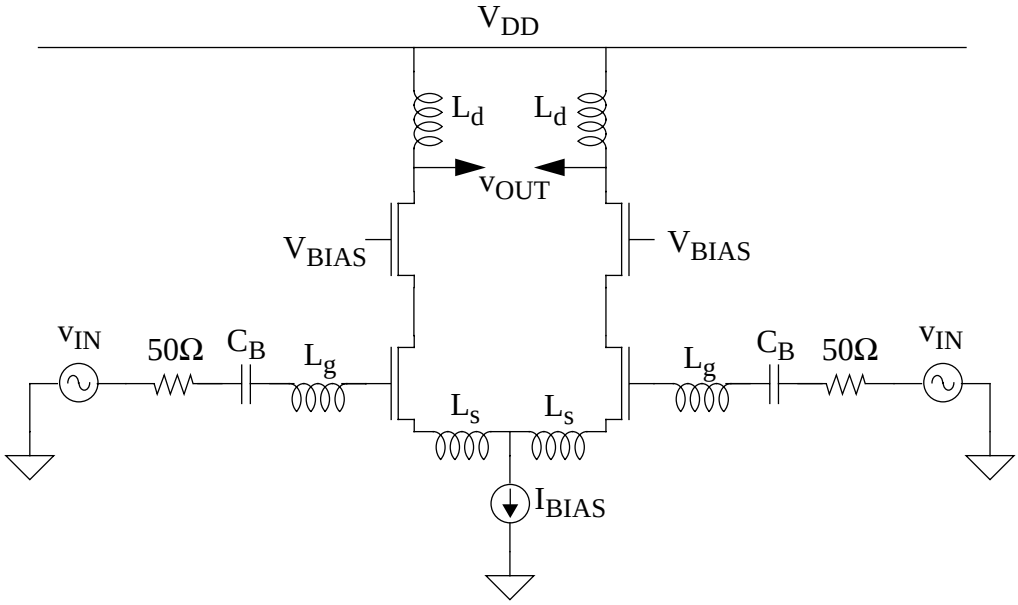
The thermal noise of the resistive gate electrode material is easily mitigated with careful layout that is already consistent with good high-fre-

quency practice. If a gate finger is contacted at only one end, it may be shown that this distributed *RC* structure has an equivalent noise resistance of $R_{SH}W/3L$, where R_{SH} is the sheet resistivity of the gate material. Thus, one must subdivide a device into a sufficient number of segments to make the total noise resistance as small as desired. Excessively short fingers are to be avoided because of the added parasitic capacitance that accompanies the wiring for each finger. Typical finger widths tend to be in the range of 5-20μm, although this is hardly a fixed rule.

5. CIRCUIT IMPLICATIONS AND IMPLEMENTATIONS

Having accepted the difficulty of achieving broadband low-noise operation, we now consider a topology designed specifically for narrowband applications:

FIGURE 2. Differential narrowband LNA (simplified)



In this differential cascoded structure, the source-degenerating inductance L_S interacts with the gate-source capacitance and g_m to produce a real term in the input impedance without paying the noise penalty of an ordinary resistance. Inductance L_G provides the additional degree of freedom required in general to allow for a desired resonant frequency of the input loop.

As discussed previously, the optimum source susceptance has an inductive character; this circuit provides that inductance at one frequency. We have also noted that the correlation susceptance of a MOSFET differs somewhat from that of the gate-source capacitance. However, using typical parameter values, the difference is seen to be reasonably small (under 25%). Hence, the source susceptance that yields optimum noise performance differs little from the value that produces a resonance at the operating frequency. Thus, this topology permits the near-simultaneous achievement of optimum noise and maximum gain. At the same time, it presents a specified real impedance to the input source.

It may be shown that the input resistance is modified downward by the drain-gate capacitance in this cascoded circuit. Although an exact analysis is difficult, a useful approximation is given by:

$$Re\left[Z_{in}\right] \approx \frac{\omega_T L_S}{1 + 2C_{gd}/C_{gs}} \tag{26}$$

The foregoing assumes equal widths for the main and cascoding devices.

Although the cascoding device does contribute noise of its own, the isolation it provides between source and load is highly desirable. The noise figure penalty is generally 0.5dB or less. In situations where that degradation is unacceptable and the coupling between input and output circuits may be tolerated, the cascoding device may be eliminated to yield an improved noise figure.

The differential connection consumes twice the power for a given noise figure than its single-ended counterpart, but has the valuable attribute of insensitivity to common-mode parasitic reactance in series with the biasing current source. Hence, one need not control or model this reactance. Single-ended designs require infinitely more attention in order to function as expected.

What is missing from the foregoing is what is missing from classical noise optimization: Guidance on how to choose the width of the transistors. In the integrated circuit design problem, the device is not fixed, but classical optimization ignores this valuable degree of freedom. Additionally, power consumption is not considered at all in the classical approach.

The necessary modification of the design procedure has been worked out in [3], and it leads to the following approximate expression for the optimum device width:

$$W_{optP} = \frac{3}{2}\frac{1}{\omega L C_{ox} R_s Q_{sP}} \approx \frac{1}{3\omega L C_{ox} R_s} \tag{27}$$

where Q_{sP} is a parameter related to the ideal Q of the input network and has a value in the range of 3-5. For typical process parameters of today, the product of width and operating frequency work out to very roughly 500μm-GHz for a 50Ω system.

It is shown in [3] that deviations from the optimum width do not cause dramatic degradation of noise figure except at low bias currents, where the absolute noise figure is likely to be high anyway. At higher power levels, the optimum conditions are relatively flat, so uncertainties in device models may be comfortably accommodated.

The existence of an optimum width may be understood qualitatively as follows. For a fixed power budget (actually, bias current), a very narrow device has high ω_T, which tends to improve noise figure. However, a narrow device also requires a high-impedance input matching network, and this increases the prominence of gate noise, which tends to degrade noise figure. A very wide device, on the other hand, has fewer problems with gate noise, but the low current density degrades ω_T, increasing the prominence of drain noise. The optimum width balances these two effects to yield the minimum noise figure for a given power budget.

6. EXPERIMENTAL RESULTS

Measurements of noise figure on single-ended, single-device LNAs are in excellent accord with the foregoing expectations. At 1GHz, minimum noise figures slightly under 1dB have been achieved in a 0.5μm technology on bias currents of 2-5mA, with power gains ranging from 13.5dB to 18.5dB. The noise figures remain below 2dB at 500μA, while the gain drops to 8dB [4]. The performance of these LNAs is not necessarily representative of what may be achieved in practice because these LNAs were not designed to provide a specified input resistance. Rather, low-

loss tuners were adjusted to yield these minima. Nevertheless, they provide a valuable check on the entries of Table 2.

As an example of a somewhat more practical design, a differential LNA for use in an integrated GPS receiver has achieved a 2.4dB noise figure at 1.6GHz with a total bias current of under 5mA (2.5mA per side) [5]. To keep noise figure small, on-chip spiral inductors are not used in the gate circuit because their lossiness is too significant. Perhaps more important than the noise figure achieved (only about 1-1.5dB above F_{min}), however, is that the measured performance is close to the predicted noise figure of 2.5dB (assuming a doubling of γ over long-channel values) obtained with a variant of SPICE modified to accommodate gate noise. Since measurements at relatively low noise figure values are much more sensitive to errors than those made at high noise figures, this level of tracking between theory and measurement is reassuring.

7. CONCLUSION

It has been shown that the portion of gate noise current that is uncorrelated with the drain noise current is of fundamental importance in setting a lower bound on the noise figure. Thermal noise from the substrate and gate interconnect material degrades noise performance, but these effects are either of second order, or may be made so through proper layout.

Acknowledgment of the correct noise model, coupled with a modification of classical noise optimization to incorporate an explicit power constraint, leads to a narrowband LNA architecture which simultaneously achieves near optimum noise and power gain while providing a specific input resistance. The optimum performance is achieved when the product of device width and operating frequency is approximately 500μm-GHz in a 50Ω system. Excellent agreement between theoretical predictions and experiment is observed.

8. ACKNOWLEDGMENTS

The author acknowledges with great pleasure having learned much during the almost-completed thesis work of Mr. Derek Shaeffer in the area of LNAs and related topics, and also thanks Mr. Gabriel Brenna for characterizing the noise of a great number of CMOS LNAs as part of his

Diploma Thesis work, carried out at Stanford and Advanced Micro Devices, with the assistance of Dr. Natalino Camilleri.

9. REFERENCES

[1] T. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, Chapter 10, Cambridge University Press, 1998.

[2] R. Jindal, "Distributed Substrate Resistance Noise in Fine-Line NMOS Field-Effect Transistors," *IEEE Trans. on Electron Devices*, vol. ED-32, no. 11, Nov. 1985, pp. 2450-2453.

[3] D. Shaeffer and T. Lee, "A 1.5V, 1.5GHz CMOS Low Noise Amplifier," *IEEE J. Solid-State Circuits*, vol. 32, no. 5, May, 1997, pp. 745-759.

[4] G. Brenna, "LNA Research," Diploma Thesis, École Polytechnique Fédérale de Lausanne, Feb. 1998.

[5] D. Shaeffer et al., "A 115mW CMOS GPS Receiver," *ISSCC Digest of Technical Papers*, Feb. 1998.