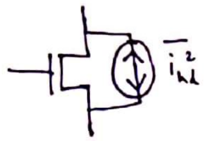
Noise in MOSFETs

Thermal noise

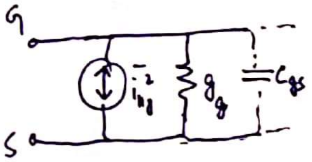
$$\overline{i_{nd}^2} = 4KT\gamma g_{d0}B$$

→ Theemal noise
→ From channel cz channel is essentially a voltage controlled resistor
→ g_{do} is drain-source conductance at $V_{DS}=0$
→ γ is process dependent factor
  → $\gamma = 1$ at $V_{DS}=0$ (means in triode)
  → $\gamma = \frac{2}{3}$ in saturation for long-channel
  → $\gamma = 2-3$ for short channel due to carrier heating by presence of large electric field. Therefore, you should try to makes V_{DS} smaller. Like just out of triode. Don't go deep in saturation.



Blue noise

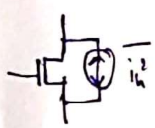
$$\overline{i_{ng}^2} = 4KT\delta g_g B$$



→ Blue noise because g_g increases with frequency
→ noise in channel gets capacitively coupled to gate
→ $g_g = \dfrac{\omega^2 C_{gs}^2}{5 g_{do}}$
→ Important to notice that it is a current source at gate
→ $\overline{i_{ng} i_{nd}^*} = c\sqrt{\overline{i_{g}^2}\cdot\overline{i_{nd}^2}}$ where c is correlation coff and usually imaginary

Flicker noise

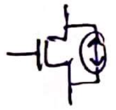
$$\overline{i_n^2} = K\cdot\frac{1}{f}\cdot\frac{g_m^2}{WLC_{ox}^2}\cdot B \quad \text{OR} \quad K\cdot\frac{1}{f}\cdot\frac{1}{WL}\cdot\omega_T^2\cdot B$$



→ Charge trap and release produce fluctuations in channel potential which is what we call noise
→ Bigger device & thin dielectric would increase gate capacitance which in turn would smooth out these fluctuations

epi noise

$$\overline{i_n^2} = 4KTR_{sub}\, g_{mb}^2\, B$$



→ Theemal noise of resistance between back gate and body
→ Produced as drain-source current like $\overline{i_{nd}^2}$
→ We compare this noise to the one coming from source R_s and we say this noise should be much smaller than R_s one

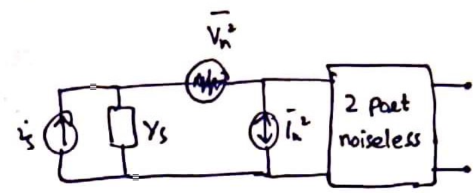
$$g_{mb}^2 R_{sub} << g_m^2 R_s$$

→ To reduce R_{sub}, we should put many taps. Break big transistors into small ones and put taps around em.

Noise parameters

- We can use two port noise parameters (R_N, G_u, G_c, B_c) to compare two technologies to find out how much NF would degrade if noise mismatch etc.

Derivation



($\overline{V_n^2}$ and $\overline{I_n^2}$ are sufficient and mandatory to capture two port noise of linear circuit)

$$F = \frac{\text{total noise power}}{\text{noise power due to source}}$$

$$= \frac{\overline{i_s^2} + \overline{|i_n + Y_s V_n|^2}}{\overline{i_s^2}}$$

Since i_n is usually partly correlated with en

$$\therefore \quad i_n = i_u + i_c \quad \& \quad \boxed{i_c = Y_c V_n} \rightarrow \text{so bascially } Y_c \text{ links } V_n \text{ to } I_{nc}$$

$$= 1 + \frac{\overline{|i_u + (Y_c+Y_s)V_n|^2}}{\overline{i_s^2}}$$

$$= 1 + \frac{\overline{i_u^2} + |Y_c+Y_s|^2\,\overline{V_n^2}}{\overline{i_s^2}}$$

we can represent
$\overline{V_n^2}$ by $R_N = \overline{V_n^2}/4KTB$
$\overline{i_u^2}$ by $G_u = \overline{i_u^2}/4KTB$
$\overline{i_s^2}$ by $G_s = \overline{i_s^2}/4KTB$

So we can write above expression in terms of $R_N, G_u, G_c, B_c, G_s, B_s$ and derivate the expression to find out minima for F

Derivate wrt B_s gives $\longrightarrow$ $B_{opt} = -B_c$

Derivate wrt G_s gives $\longrightarrow$ $G_{opt} = \sqrt{\frac{G_u}{R_N} + G_c^2}$

Put these value in F to get F_{min}

$$F_{min} = 1 + 2R_N\left(\sqrt{\frac{G_u}{R_N} + G_c^2} + G_c\right)$$

We can then express F as:

$$F = F_{min} + \frac{R_N}{G_s}\left[Y_s - Y_{OPT}\right]^2$$

Derivation of MOSFET Noise parameters

$$\overline{V_n^2} = \frac{\overline{i_{nd}^2}}{g_m^2} = \frac{4KT\gamma g_{do}B}{g_m^2} \rightarrow \boxed{R_N = \frac{\gamma g_{do}}{g_m^2}}$$

$$\overline{i_{n1}^2} = \frac{\overline{i_{nd}^2}}{g_m^2}\cdot(j\omega C_{gs})^2 = \overline{V_n^2}\cdot(j\omega C_{gs})^2 \rightarrow Y_c = j\omega C_{gs} \quad \text{(if there is no correlated noise)}$$

But we have gate noise partly correlated

$$\overline{i_{n2}^2} = \overline{V_n^2}\cdot Y_c^2$$

$$\overline{i_{n2}^2} = \frac{\overline{i_{nd}^2}}{g_m^2}\cdot Y_c^2 \rightarrow Y_c = g_m\cdot\sqrt{\frac{\overline{i_{n2}^2}}{\overline{i_{nd}^2}}} = g_m\cdot\sqrt{\frac{|c|^2\,\overline{i_{ng}^2}}{\overline{i_{nd}^2}}} = g_m\cdot\sqrt{\frac{4KTB\delta\cdot\omega^2C_{gs}^2/5g_{do}}{4KTB\gamma g_{do}}}$$

$$= \frac{g_m\cdot c}{g_{do}}\sqrt{\frac{\delta}{5\gamma}}\,\omega C_{gs}$$

We used $\overline{i_{n2}^2} = |c|^2 \overline{i_{ng}^2}$ → this is only true for magnitude. Otherwise there is imaginary component. Refer Thomas Lee. He derives Y_c from gate noise little differently. However, knowing that c is imaginary we arrive at same result.

$$\rightarrow \quad Y_c = j\omega C_{gs} + \frac{g_m}{g_{do}} \cdot jc \sqrt{\frac{\delta}{5\gamma}} \cdot \omega C_{gs}$$

$$\boxed{Y_c = j\omega C_{gs}\left(1 + \alpha |c| \sqrt{\frac{\delta}{5\gamma}}\right)} \quad \text{where } \alpha = \frac{g_m}{g_{do}}$$

→ This shows correlation impedance is different than input impedance ($Z_{in} = j\omega C_{gs}$)
So this also indicates that conjugate and noise match are different. Remember we need $-B_c$ for noise matching.

→ We can also see $-B_c$ is negative cap. ∴ noise match is inherently narrowband if we use inductor for source impedance because $-C$ and $+L$ have different frequency behavior.

→ Uncorrelated noise from gate does not have imaginary component cz it is only given by G_u. There is no B_u. Correlated has imaginary component cz of imaginary c.

$$\therefore \quad \overline{i_{ng}^2} = \overline{(i_{ngc} + i_{ngu})^2} = 4KTBR\delta g_g |c|^2 + \underbrace{4KTB\delta g_g (1-|c|^2)}_{i_{ngu}}$$

Putting g_g in there, we get

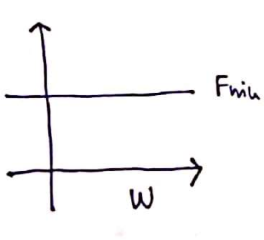
$$\boxed{G_u = \frac{\delta \omega^2 C_{gs}^2 (1-|c|^2)}{5 g_{do}}}$$

Also with shorter channel F_{min} goes down because increase in ω_T is more than increase in γ, δ factors.

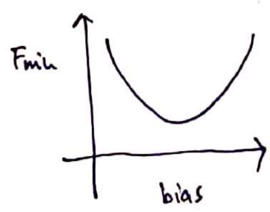
We can get F_{min} now as:

$$\boxed{F_{min} = 1 + \frac{2}{\sqrt{5}} \frac{\omega}{\omega_T} \sqrt{\gamma\delta(1-|c|^2)}}$$

→ Interesting to note that with bigger device size F_{min} does not change but F goes down because R_N goes down.
∴ Use bigger device $\left[F = F_{min} + \frac{R_N}{G_s}(Y_s - Y_{opt})^2\right]$



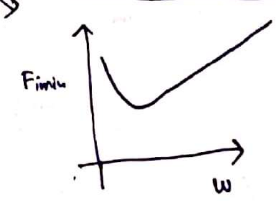
F_{min} has no dependence on device width because γ, δ and c are width independent factors (more process related)
Also ω_T does not change because g_m & C_{gs} scales proportionally.



→ At low bias γ and δ are very high. e.g. $\gamma = 1$ in triode. ω_T is also low.

→ At high bias ω_T is high but γ and δ are high again because of carrier heating.

→ ∴ there is optimal bias

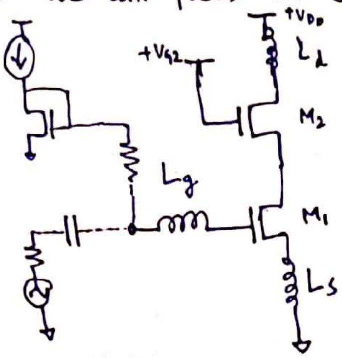


→ At low freq, flicker noise is dominant

→ Other than that noise increases with high freq because
→ ω/ω_T decreases
→ δ "blue noise" happens (gate induced noise)

LNA Design

* We will focus on cascode as it is pretty much the standard.



My way:

① As F_{min} has optima for bias, choose bias for optima point.

② Increase device size to get required Q of network.

$$Q = \frac{1}{\omega C_{gs} \cdot 2R_s} \quad \text{(when input is designed to be matched)}$$

↑ Device ↓Q ↑ BW

③ Add L_s to introduce $\omega_T L_s$ real part in Z_{in}.

④ Add ~~C_{gs}~~ L_g so that $L_g + L_s$ resonate C_{gs} out. Now you have pure real input resistance.

2a Use smallest channel length as it has lower F_{min} because ω_T increases. Note you can also increase ω_T by bias but γ and δ also increases which may counter the effect.

2b Bigger width is also better because it reduces R_N. Hence, sensitivity to noise mismatch is reduced.

⑤ For M_2, you want to increase its size to increase its g_m and thus reduce M_1 gain. This will reduce miller effect. Otherwise C_{gd} will reduce real part of M_1 as:

$$Re(Z_{in}) = \frac{\omega_T L_s}{1 + 2C_{gd}/C_{gs}}$$

However, this big size will add too much cap at ~~paras~~ intermediate node increasing M_2 noise contribution which starts to act after this frequency $(2r_o \cdot C_x)^{-1}$ where C_x is cap at intermediate noise.

Considering this you may want to decrease M_2 size but you will run into real part and larger overdrive problems.

∴ M_2 best size is same as M_1. This way you can also easily merge M_1 drain and M_2 source, reducing cap further. You can also pull out more current from M_2 source just to decrease M_1 gain by enhancing M_2 g_m.

⑥ Choose L_d to resonate out parasitic cap at drain. You may choose input and output resonance freq to be little different to enhance BW.

⑦ V_{g2} bias is normally chosen to be V_{DD} but it has two problems:

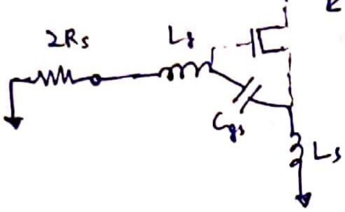
(i) Max voltage swing at output is reduced. The minimum it can go now is $V_{DD} - V_{th}$ whereas if you chose V_{g2} to be $2V_{GS} - V_{th}$ then it can go upto $2V_{ov}$.

(ii) Also $V_{g2} = V_{DD}$ will make V_{DS1} large which may increase γ by carrier heating.

∴ Choose V_{g2} which will bias M_1 just outside triode (Not too deep saturation)

Interesting Facts

* G_m of LNA is independent of device size at resonance given that input is matched.



equivalent circuit at resonance

Note: we have $2R_s$ (R_s from source + R_s of Z_{in})

if $\frac{V}{2}$ is voltage across R_s (∵ if V is source voltage, it is divided by 2 between R_s)
→ $Q\frac{V}{2}$ is voltage across C_{gs}

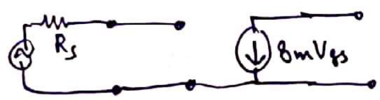
∴ voltage is Q times higher or $G_m = Q g_m$

As $Q = \frac{1}{\omega C_{gs} \cdot 2R_s}$ → $G_m = \frac{g_m}{C_{gs}} \cdot \frac{1}{\omega} \cdot \frac{1}{2R_s} = \frac{\omega_T}{2\omega_0 R_s}$

→ So you can only increase G_m by increasing device bias (thus, ω_T) (intuitively, ↑ device Q↓ ∴ G_m remains same)

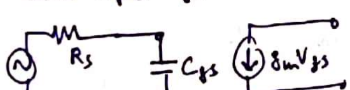
* A capacitive degenration can introduce -ive real part. ∴ be careful about source to substrate cap.

* Inductive degenration does not increase NF. ∵ Γ_{opt} stays unaffected.
  → To understand this consider,



$$NF = 1 + \left(\frac{\gamma}{\alpha}\right) \cdot \frac{1}{g_m R_s}$$

with input C_{gs}



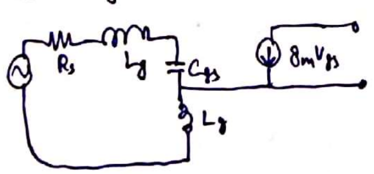
$$NF = 1 + \left(\frac{\gamma}{\alpha}\right) \cdot \frac{1}{g_m R_s} \cdot \omega^2 R_s^2 C_{gs}^2 \quad \text{OR} \quad 1 + \frac{\gamma}{\alpha} \cdot g_m R_s \cdot \left(\frac{\omega}{\omega_T}\right)^2$$

here $\omega^2 R_s^2 C_{gs}^2$ always >> 1

∴ NF increased.

Intuitively G_m decreased because of voltage division at input ∴ signal gain decreased but output is still producing with actual 'g_m'

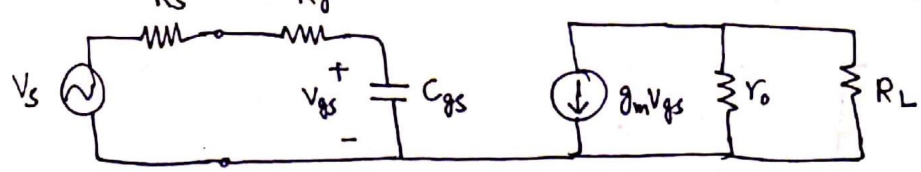
with degeration & resonance



Now, output noise current reduced by $\frac{1}{2}$ in amplitude or $\frac{1}{4}$ in power. Because part of it flows through input and activates it g_m. which produces correlated noise at output with opposite sign. We can see mathematically that it produces $-i_d/2$ noise at output. which cancells i_d and only $i_d/2$ flows at output. ∴ NF should decrease. However, G_m is decreased also (now $\frac{\omega_T}{\omega_0 \cdot 2R_s}$) which reduces signal gain again. ∴ Overall NF remains same.

$$NF = 1 + \frac{\gamma}{\alpha} \cdot g_m R_s \cdot \left(\frac{\omega}{\omega_T}\right)^2$$

CS LNA



$\overline{V_g^2} = 4KTR_g$

$\overline{V_s^2} = 4KTR_s$

$\overline{i_d^2} = 4KT\gamma g_m$ (long channel)   otherwise $g_{d0}\gamma = g_m \cdot \frac{\gamma}{\alpha}$

$\overline{i_L^2} = 4KTG_L$

$$\overline{i_T^2} = \overline{i_d^2} + \overline{i_L^2} + G_m^2\left(\overline{V_g^2} + \overline{V_s^2}\right)$$

where $G_m = \frac{1/sC_{gs}}{1/sC_{gs} + R_s + R_g} \cdot g_m = \frac{g_m}{1 + sC_{gs}(R_s + R_g)}$

Input-reffered noise voltage

$$\frac{\overline{i_d^2}}{G_m^2} + \frac{\overline{i_L^2}}{G_m^2} + \overline{V_g^2} + \overline{V_s^2} = \overline{V_{in}^2}$$

Divide by $\overline{V_s^2}$ to get NF

$$NF = 1 + \frac{\overline{V_g^2}}{\overline{V_s^2}} + \frac{\overline{i_d^2} + \overline{i_L^2}}{G_m^2 \overline{V_s^2}}$$

$$= 1 + \frac{4KTR_g}{4KTR_s} + \frac{4KT\gamma g_m + 4KTG_L}{4KTR_s} \cdot \frac{1}{G_m^2}$$

$$= 1 + \frac{R_g}{R_s} + \left(\frac{\gamma g_m}{R_s} + \frac{1}{R_L R_s}\right) \cdot \frac{1}{G_m^2}$$

You may be tempted to increase R_s seeing from this expression. But G_m is more strongly dependent on R_s. So you would need to decrease it.

We need to take magnitude of $G_m \rightarrow$ so $|G_m|^2$ because NF is a real quantity

$$NF = 1 + \frac{R_g}{R_s} + \left(\frac{\gamma g_m}{R_s} + \frac{1}{R_L R_s}\right)\left(1 + \omega^2 C_{gs}^2 (R_s + R_g)^2\right) / g_m^2$$

Good layout $\rightarrow R_s >> R_g$ so $R_s + R_g \approx R_s$

Say high freq $\rightarrow \omega^2 C_{gs}^2 R_s^2 >> 1$

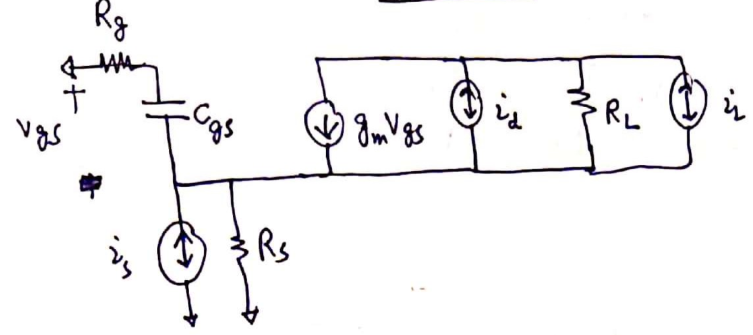
$$\therefore \quad NF = 1 + \cancelto{0}{\frac{R_g}{R_s}} + \frac{\gamma g_m}{R_s} \cdot \frac{\omega^2 C_{gs}^2 R_s^2}{g_m^2} + \frac{\omega^2 C_{gs}^2 R_s^2}{R_L R_s g_m^2}$$

$$= 1 + \gamma\left(\frac{\omega}{\omega_T}\right)^2 g_m R_s + \frac{\omega^2 C_{gs}^2 R_s}{R_L g_m^2}$$

$$= 1 + \gamma\left(\frac{\omega}{\omega_T}\right)^2 g_m R_s + \left(\frac{\omega}{\omega_T}\right)^2 \cdot \frac{R_s}{R_L}$$

To $\downarrow$ NF
- $\downarrow \omega$
- $\uparrow \omega_T$ by increasing ~~bias~~ g_m by increasing bias
- Reduce R_s because it affects G_m

CG LNA



at low-frequency ignoring C_{gs}

$$\overline{i_s^2} = \frac{4KT}{R} \rightarrow \overline{V_{gs}^2} = \overline{i_s^2} R_s^2 = 4KTR_s$$

$$\overline{i_T^2}\text{ (output)} = g_m^2 \overline{V_{gs}^2} + \overline{i_d^2} + \overline{i_L^2}$$

Dividing by output noise current from R_s only to get NF

$$NF = 1 + \frac{\overline{i_d^2} + \overline{i_L^2}}{g_m^2 \overline{V_{gs}^2}}$$

$$= 1 + \frac{4KT\gamma g_m + 4KTG_L}{g_m^2 \cdot 4KTR_s}$$

$$= 1 + \frac{\gamma}{g_m R_s} + \frac{1}{g_m^2 R_L R_s}$$

Since $g_m R_s = 1$, that's how you input match in CG by making $g_m = \frac{1}{R_s}$

$\rightarrow$ NF $= 1 + \gamma$ (ignoring R_L)

γ is usually $2/3$

so NF $\rightarrow$ 2-3 dB already

$\therefore$ CG have higher NF because you couldn't make g_m large enough

- In other words, you can also think that since in CG there is no current gain all of drain noise current flows through input without getting divided by gain $\therefore$ higher NF

$$NF = \frac{\overline{i_d^2} + \overline{i_s^2}}{\overline{i_s^2}} = 1 + \frac{\overline{i_d^2}}{\overline{i_s^2}} = 1 + \gamma g_m R_s = 1 + \gamma$$

so two reasons for high NF:

$\rightarrow$ g_m not high enough

OR

$\rightarrow$ no current gain