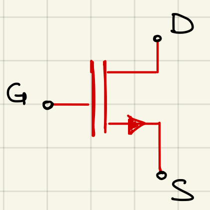
Analog Circuit Design

Andrea
Bertazzoni
A.A. 2020/21

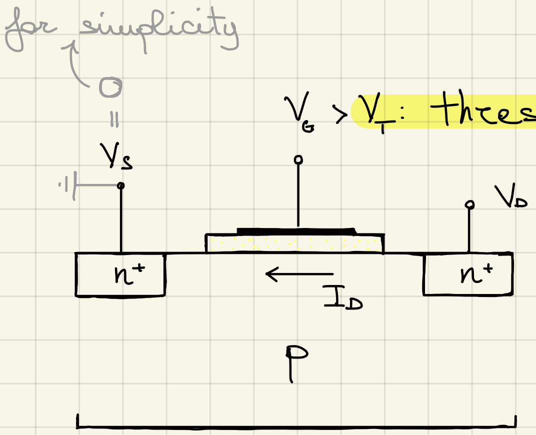


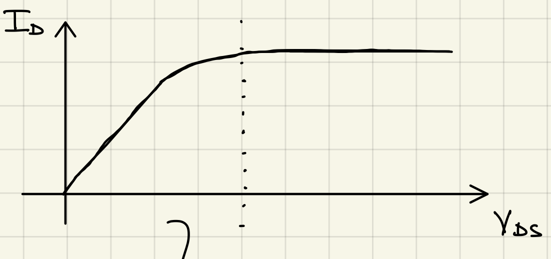
μ: maximum gain

g_m: transconductance

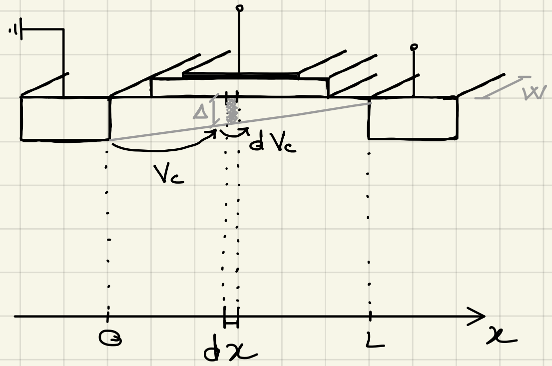
f_T: cut-off frequency

$V_G > V_T$: threshold voltage





how is it derived?



$$dV_c = I_D\, dR = I_D \frac{dx}{\underbrace{q}_{C}\,\mu_n \underbrace{n(x)}_{cm^{-3}}\, W\, \underbrace{\Delta(x)}_{cm}} =$$

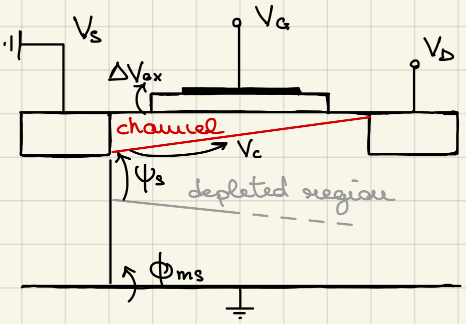
$$= I_D \frac{dx}{\mu_n \underbrace{Q'_n}_{\frac{C}{cm^2}} W}$$ charge surface density

What is Q'_n equivalent to?

$$V_G = \phi_{ms} + \psi_s + \Delta V_{ox}(0)$$ charge in the depleted region

$$V_G = \phi_{ms} + \psi_s + V_c + \frac{Q'_d + Q'_n}{C'_{ox}} \rightarrow Q'_n = C'_{ox}\left[V_G - \left(\phi_{ms} + \psi_s + V_c + \frac{Q'_d}{C'_{ox}}\right)\right]$$

charge in the channel



But $Q'_d = Q'_d(\psi_s + V_c)$ varies as the depleted region deepens (body effect).

$\Rightarrow$ First order approx.:

$$V_c + \underbrace{\phi_{ms} + \psi_s + \frac{Q'_d(\psi_s)}{C'_{ox}}}_{V_T} + \cancel{\underbrace{\frac{Q'_d(\psi_s + V_c) - Q'_d(\psi_s)}{C'_{ox}}}_{\Delta V_T}}$$

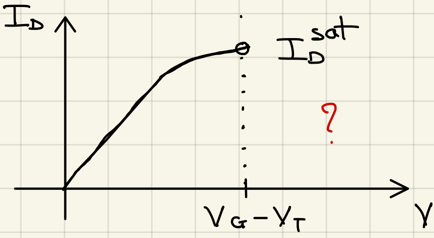
$$V_c + V_T + \cancel{\Delta V_T}$$

$\longrightarrow$ $Q'_n \simeq C'_{ox}[V_G - V_T - V_c]$ (charge-sheet model)

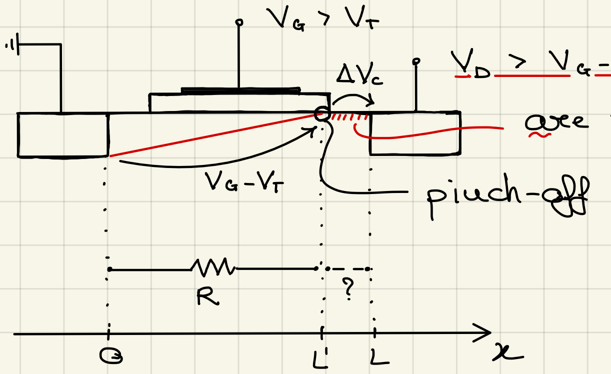
$$dV_c = \frac{dx}{\mu_n Q'_n W} I_D \simeq \frac{dx}{\mu_n C'_{ox}[V_G - V_T - V_c]\, W} \cdot I_D$$

$$\int_0^{V_{DS}} \mu_n C'_{ox} W \left[V_G - V_T - V_c\right] dV_c = \int_0^L I_D\, dx$$

$$\Rightarrow \quad \boxed{I_D = \mu_n C'_{ox}\frac{W}{L}\left[(V_G-V_T)V_{DS} - \frac{V_{DS}^2}{2}\right]}$$



$$\left[\; I_D^{sat} = \frac{\mu_n C'_{ox}}{2}\frac{W}{L}(V_G-V_T)^2 = k\,V_{ov}^2 \;\right]$$



are there any electrons?

$$\int_0^{V_G-V_T} dV_c = \int_0^{L'} \frac{I_D\,dx}{\mu_n W Q'_n}$$

$$\hookrightarrow I_D = \frac{\mu_n C'_{ox}}{2}\frac{W}{L'}(V_G-V_T)^2$$

$L' < L$ certainly, so knowing that if V_D grows bigger than V_{ov} ($= V_G - V_T$) then L' decreases, we can say for sure that as V_D grows I_D grows as well (being L' at the denominator)

Moving from source to drain, current stays the same but electron density goes down. Given the equation:

$$I \sim \downarrow n\, v\uparrow \quad \text{and} \quad \uparrow v = \mu_n F \uparrow$$

it means that electron speed and therefore electric field must be much bigger after the pinch-off.

$$\Delta V_c = V_D - \underbrace{(V_G - V_T)}_{V_D^{sat}} \qquad F \simeq \frac{\Delta V_c}{L - L'} \qquad \text{but} \qquad L' = L'(V_D)$$

First order expansion: $L'(V_D) = L + \left.\frac{dL'(V_D)}{dV_D}\right|_{V_D^{sat}} (V_D - V_D^{sat})$

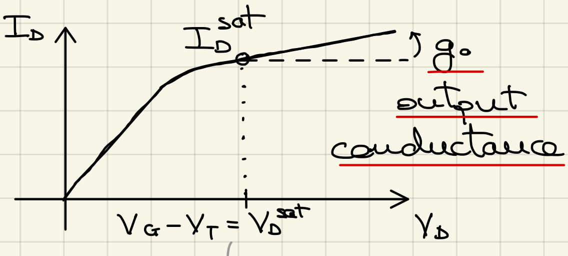
$$= L\left[1 + \frac{1}{L}\underbrace{\left.\frac{dL'}{dV_D}\right|_{V_D^{sat}}}_{<0}(V_D - V_D^{sat})\right]$$

$$= L\left[1 - \underset{\overset{\vee}{0}}{\lambda}\,(V_D - V_D^{sat})\right]$$

$$\Rightarrow I_D = \frac{1}{2}\mu_n C'_{ox}\frac{W}{L'}(V_G-V_T)^2$$

$$= \frac{1}{2}\mu_n C'_{ox}\frac{W\,(V_G-V_T)^2}{L\left[1 - \lambda(V_D - V_D^{sat})\right]}$$

$$\left\{ \frac{1}{1-x} \approx 1 + x \quad \text{for } x \to 0 \right\}$$



true only for non-short channel devices

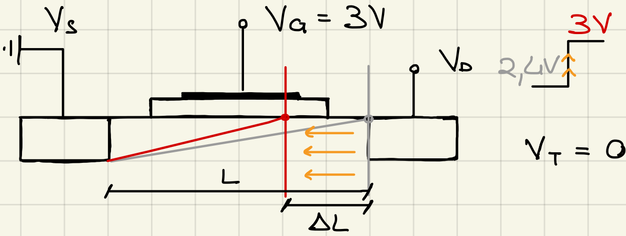
$$\left[\; I_D = \frac{1}{2}\mu_n C'_{ox}\frac{W}{L}(V_G-V_T)^2\left[1 + \lambda(V_D - V_D^{sat})\right] \;\right]$$

where $\lambda = \frac{1}{L}\left|\frac{dL'}{dV_D}\right|_{V_D = V_D^{sat}} = \frac{1}{V_A}$

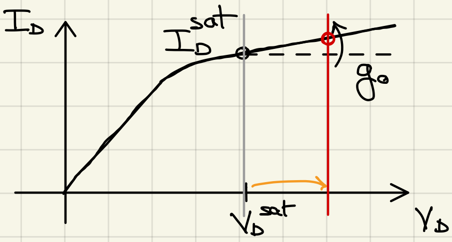
V_A: modulation voltage ("Early effect")

$V_A = \alpha L = \frac{V_A^\circ}{L_\circ} \cdot L$ tipically $L_\circ = 0{,}35\mu m$, $V_A^\circ = 7\,V$

longer channel → higher modulation voltage → lower output conductance



$V_T = 0{,}6V$



$g_o \sim \lambda = \frac{1}{L}\left|\frac{dL'}{dV_D}\right|_{V_D^{sat}}$

I_D should have a relative increase equal to $\frac{\Delta L}{L}$ ⇒ the longer the channel, the lower the increase

$I_D = I_D^{sat}\left[1 + \lambda\left(V_D - V_D^{sat}\right)\right]$ (source @ ground)

$\hookrightarrow I_D^{sat} = K_n\left[V_G - V_T\right]^2$, $K_n = 1/2\, \mu_n C'_{ox}\, W/L$

$V_D^{sat} = V_G - V_T = V_{ov}$ (non-short devices)

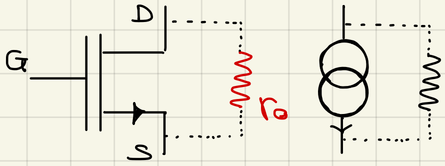
$g_o = \frac{dI_D}{dV_D} = K\left(V_G - V_T\right)^2 \cdot \lambda = \lambda I_D^{sat} = \frac{1}{L}\left|\frac{dI_D}{dV_D}\right|_{V_D^{sat}} \cdot I_D^{sat}$

$\Rightarrow g_o = \frac{I_D^{sat}}{V_A}$ $\boxed{r_o = \frac{V_A}{I_D^{sat}}} \simeq \frac{V_A}{I_D}$

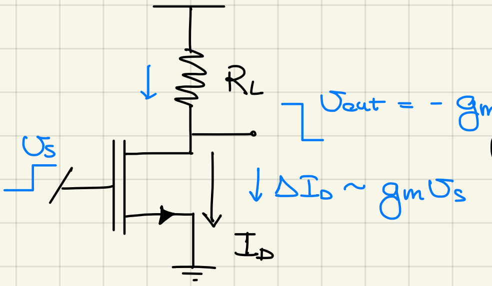
output resistance → increasing V_{ov} decreases the output resistance!

IMPLICATIONS:

The transistor is not an ideal current generator anymore

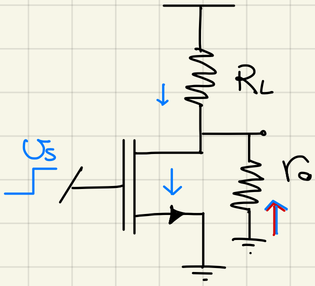


⇒ RESISTIVE COUPLING between source and drain



$v_{out} = -g_m v_s R_L$

$\hookrightarrow G = -g_m R_L$ voltage gain



$G = -g_m(R_L // r_o)$

capped at $g_m r_o$

$g_m = \frac{\partial I_D^{sat}}{\partial V_G} = \frac{\partial\left[K\left(V_G - V_T\right)^2\right]}{\partial V_G} = 2K\left(V_G - V_T\right)$ transconductance

$$\left[g_m = 2K(V_G - V_T) = 2K V_{ov} = \frac{2 I_D}{V_{ov}} = 2\sqrt{K I_D} \right]$$

Maximum gain $\mu = g_m r_o = \frac{2 I_D}{V_{ov}} \cdot \frac{V_A}{I_D} = \frac{2 V_A}{V_{ov}}]$ ↳ independent of bias current!

Ideally:

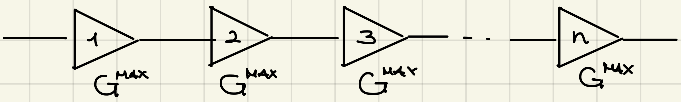
$$G = -g_m R_L \xrightarrow{R_L \to \infty} -\infty$$

Really:

$$G = -g_m (R_L // r_o) \xrightarrow{R_L \to \infty} -\overset{G^{MAX}}{\boxed{g_m r_o}} = -\mu$$

only transistor parameters → FIGURE OF MERIT

To build an OP. AMP. with a gain higher than the maximum gain we are then forced to use multiple transistors in cascade:



$$A_o = (G^{MAX})^n$$

Since $\mu = \frac{2V_A}{V_{ov}}$ you can use as little current as possible and still gain the most out of the transistor. This means you can amplify signals by a huge amount with almost no power consumption whatsoever.

For example: $V_A^0 = 7V$ $L_0 = 0{,}35\mu m$ $L = 1\mu m$ $V_{ov} = 0{,}1 V$

$$\Rightarrow V_A = \frac{V_A^0}{L_0} \cdot L = 20V \Rightarrow \mu = \frac{2V_A}{V_{ov}} = 400$$

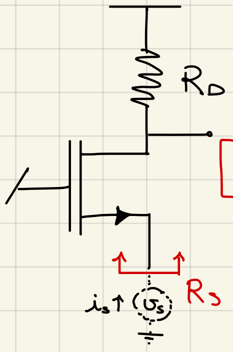
$$2 \text{ stages} \Rightarrow A_o = \mu^2 > 10^5 \text{ !!!}$$

"The maximum gain increases with longer channels" bad news

BUT

"The cut-off frequency increases with shorter channels" good news

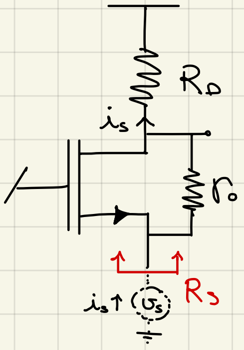
Ideally:



$$\left[R_S = \frac{v_s}{i_s} = \frac{v_s}{g_m v_s} = \frac{1}{g_m} \right]$$

independent of R_D!

Really:



$$\frac{v_s - R_D i_s}{r_o} + g_m v_s = i_s$$

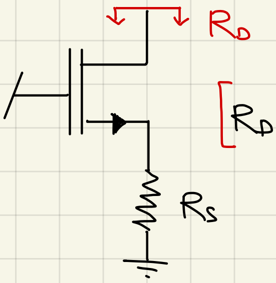
$$v_s \left(\frac{1}{r_o} + g_m \right) = i_s \left(1 + \frac{R_D}{r_o} \right)$$

$$\frac{v_s}{i_s} = \left[R_S = \frac{r_o + R_D}{1 + g_m r_o} \right]$$

$R_D \gg r_o$ ↙ $\frac{R_D}{\mu}$

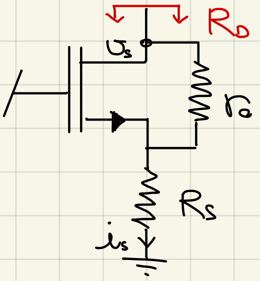
$r_o = \infty$ ↘ $\frac{1}{g_m}$

Ideally:



$\boxed{R_D \simeq \infty}$ → independent of R_s!

Really:



$$i_s = \frac{v_s - R_s i_s}{r_o} - g_m R_s i_s$$

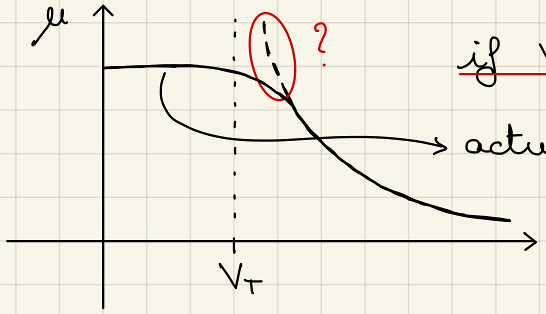
$$i_s\left(1 + \frac{R_s}{r_o} - g_m R_s\right) = \frac{v_s}{r_o}$$

$$\frac{v_s}{i_s} = \boxed{R_D = r_o + R_s(1+g_m r_o)}$$

$R_s > r_o$ → μR_s

$r_o = \infty$ → ∞

Note that $\mu = g_m r_o = \frac{2V_A}{V_{ov}}$ means not only that we can get a greater maximum gain by having a larger channel (and therefore a bigger V_A) – that is, increase the numerator – but also that we can get a very big maximum gain by having an almost-zero V_{ov} – that is, decrease the denominator.



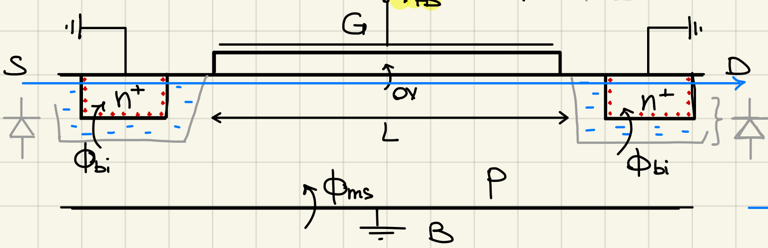
if $V_G \to V_T$ ($V_{ov} \to 0$) then $\mu \to +\infty$? NO

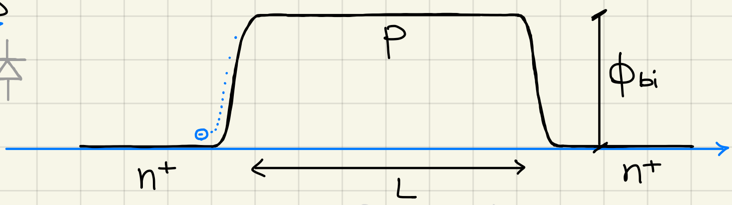
→ actual slope – how is it derived?

↓

Let's study the SUBTHRESHOLD OPERATION of the transistor:

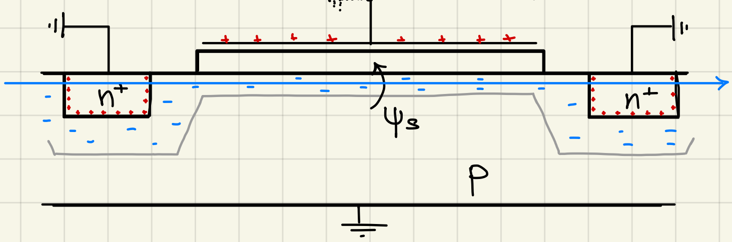
$V_{FB} \simeq -1V$ → Flat Band voltage

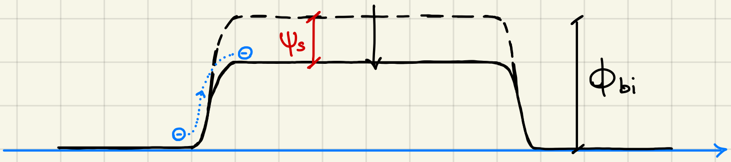




remember that higher potential = lower voltage for electrons

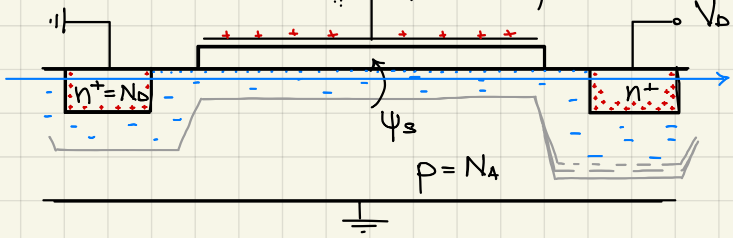
$V_{FB} < V_G < V_T = 0,6V$

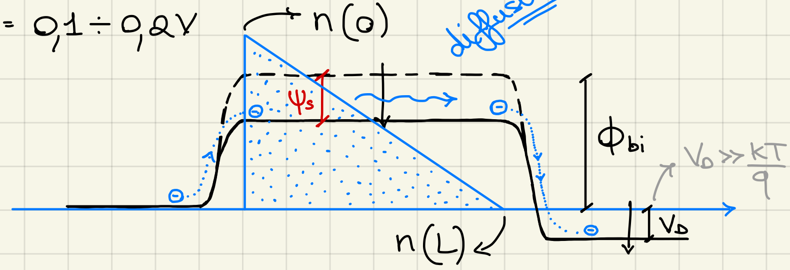




$V_G < V_T = 0,6V$

$V_D = 0,1 \div 0,2V$





$n(0) \simeq N_D e^{-\frac{q(\phi_{bi} - \psi_s)}{KT}}$ $\qquad\qquad$ $n(L) \simeq 0$

A few electrons at the source side have enough thermal energy to cross the potential barrier, which is now lowered by ψ_s. The number of electrons is given by Boltzmann's law.

Almost all electrons at the drain side will get over the potential drop to reach a lower potential level, which is deeper than the source side thanks to V_D.

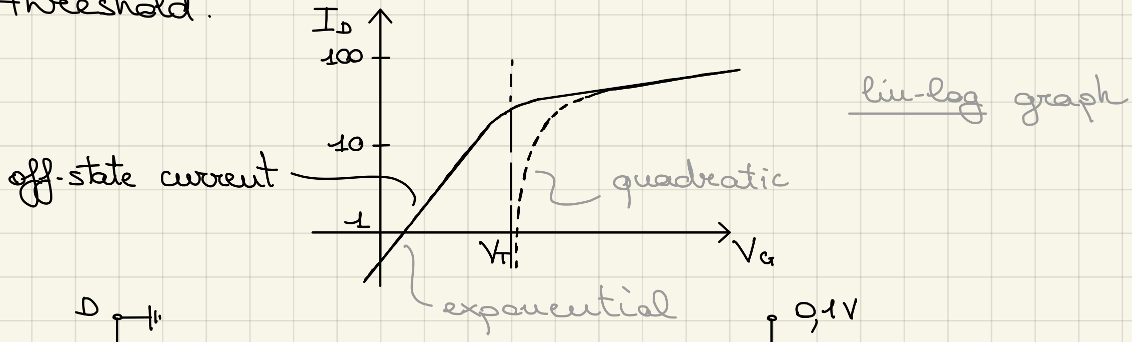
$$I_D = \underbrace{q D_n \frac{dn(x)}{dx}}_{J_{diff}} A \simeq q D_n \overset{\text{constant if } L < L_{diff}}{\underbrace{\frac{n(0) - n(L)}{L}}} A = q \frac{D_n A}{L} N_D e^{-\frac{q\phi_{bi}}{KT}} e^{\frac{q\psi_s}{KT}} = \underbrace{q \frac{D_n A}{L} \frac{n_i^2}{N_A}}_{I_0} e^{\frac{q\psi_s}{KT}}$$

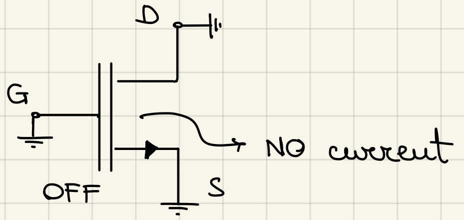
$$\phi_{bi} = \frac{KT}{q} \ln \frac{N_A N_D}{n_i^2}$$

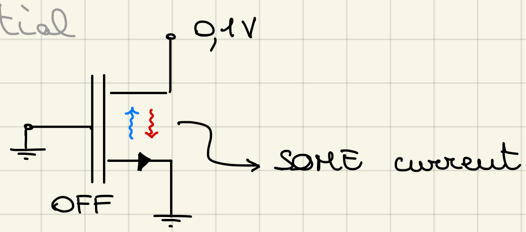
$$= I_0 e^{\frac{q\psi_s}{KT}}$$ diffusion current

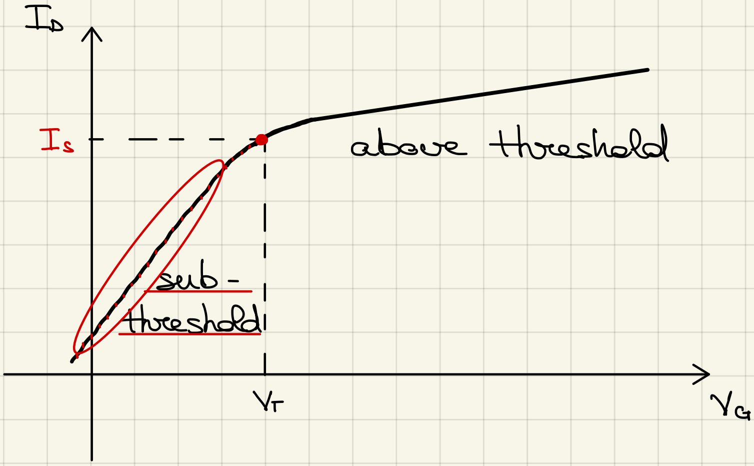
⟹ even if the transistor is off there is still some leakage current given by a diffusion term

Since $\psi_s \sim V_G$ then I_D grows exponentially with V_G while below threshold.



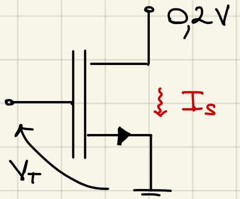


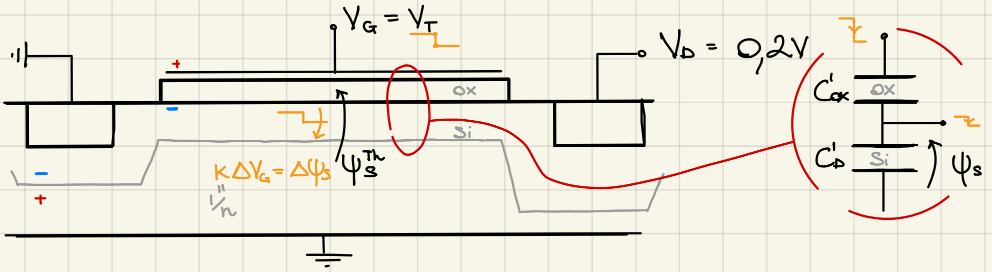




$I_D(V_G = V_T) = I_D(V_{ov} = 0) := I_s \neq 0$!!

I_s is the "threshold current".





$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} \cdot A \qquad C_D = \frac{\varepsilon_{si}}{W_D} \cdot A$

$$\Delta\psi_s = \Delta V_G \frac{C'_{ox}}{C'_{ox} + C'_D} = \frac{\Delta V_G}{n}$$

$$n = \frac{C'_{ox} + C'_D}{C'_{ox}} = 1 + \frac{C'_D}{C'_{ox}} \geqslant 1$$

(typically $n \simeq 1{,}5$)

$$I_D = I_S \, e^{\frac{q(\psi_s - \psi_s^{Th})}{KT}} \quad \text{in fact} \quad I_D(\psi_s = \psi_s^{Th}) = I_S = I_0 \, e^{q\frac{\psi_s^{Th}}{KT}}$$

$$= I_S \, e^{\frac{q\Delta\psi_s}{KT}} \quad \text{where} \quad \Delta\psi_s = \psi_s - \psi_s^{Th} \quad \psi_s^{Th} := \psi_s(V_G = V_T) = \psi_s(V_{ov} = 0)$$

$$= I_S \, e^{\frac{q\Delta V_G}{nKT}} \quad \text{since} \quad \frac{\Delta V_G}{n} = \Delta\psi_s$$

$$\Rightarrow \boxed{I_D = I_S \, e^{\frac{q(V_G - V_T)}{nKT}}}$$

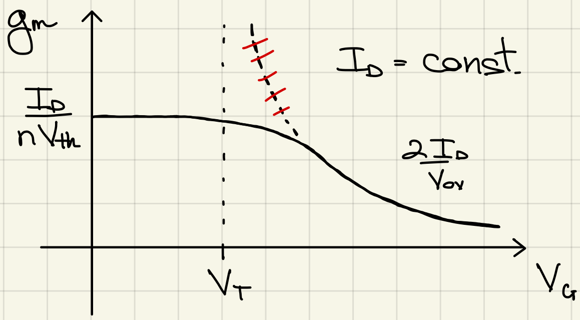
$$g_m = \frac{dI_D}{dV_G} = I_S \, e^{\frac{q(V_G - V_T)}{nKT}} \cdot \frac{q}{nKT} = \frac{I_D \, q}{nKT}$$

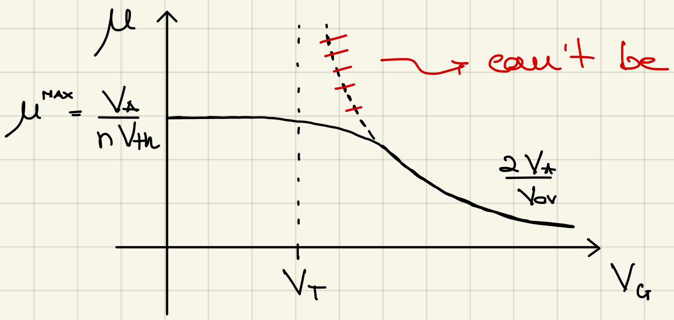
$$\Rightarrow \boxed{g_m = \frac{I_D}{n V_{th}} \quad \text{sub-threshold transconductance}}$$

$V_{th} \simeq 25\text{mV}$ @ 300K

Now g_m DEPENDS on the BIAS CURRENT and is INDEPENDENT of the OVERDRIVE VOLTAGE

→ there is a fixed limit to the transconductance and the maximum gain, which grows with the bias current and therefore with power consumption





$$g_m = \frac{2K(V_G - V_T)^2}{V_G - V_T} = \frac{2I_D}{V_{ov}} \xrightarrow{V_{ov}\to 0} \cancel{\infty}$$

$$\mu = g_m r_o = \frac{2I_D}{V_{ov}} \cdot \frac{V_A}{I_D} = \frac{2V_A}{V_{ov}} \xrightarrow{V_{ov}\to 0} \cancel{\infty}$$

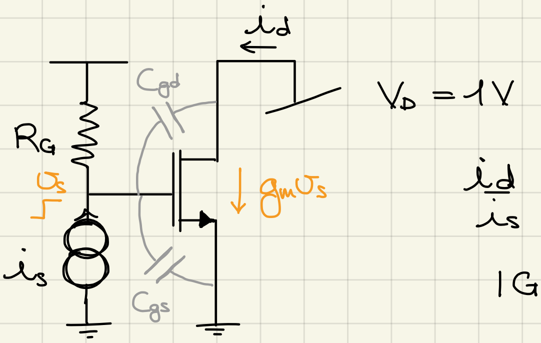
$\Longrightarrow$ around-threshold and sub-threshold values:

$$\left[g_m = \frac{I_D}{nV_{th}} \right] \qquad \left[\mu = g_m r_o = \frac{I_D}{nV_{th}} \cdot \frac{V_A}{I_D} = \frac{V_A}{\underbrace{nV_{th}}_{\text{fixed}}} \right]$$

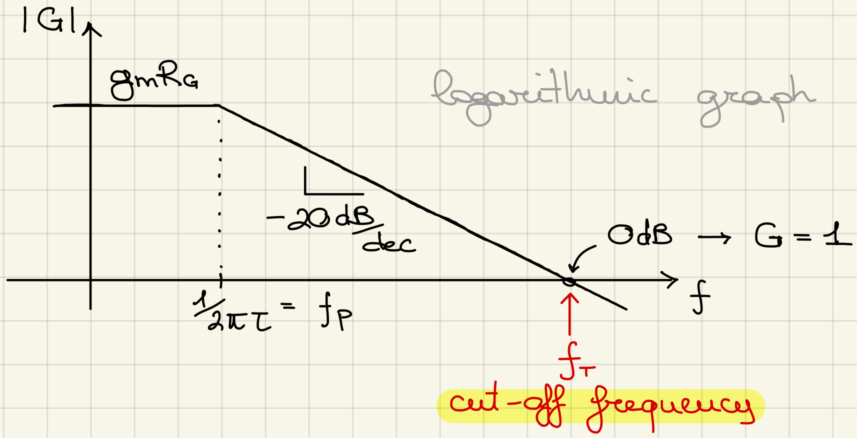
For example: $L = 1\mu m \quad V_A \simeq 20V \quad n = 3$

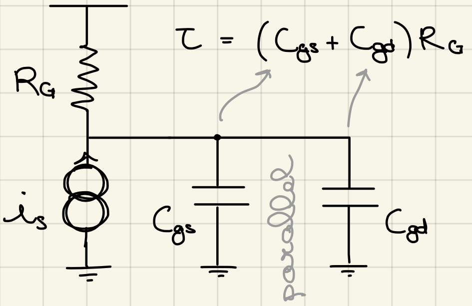
$$\Rightarrow \mu^{MAX} = \frac{20}{3 \cdot 25} \cdot 10^3 = 600$$

We necessarily need more than one transistor to obtain really high gains ($A_0 = 10^6 \div 10^7$).



$$\frac{i_d}{i_s} = G_0 = \frac{i_s R_G g_m}{i_s} = g_m R_G$$





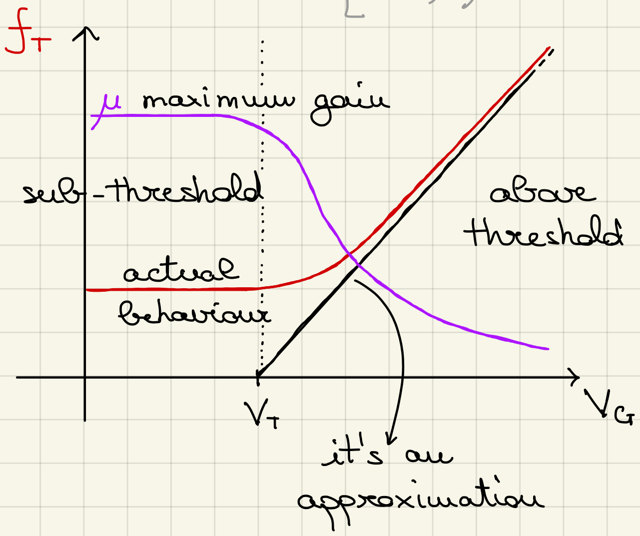
$$G(f_T)\cdot f_T = G(f_P)\cdot f_P$$

$$1\cdot \boxed{f_T} = g_m \cancel{R_G}\cdot\frac{1}{2\pi(C_{gs}+C_{gd})\cancel{R_G}} \simeq \boxed{g_m\frac{1}{2\pi C_{ox}}}$$

$$f_T = \frac{g_m}{2\pi(C_{gs}+C_{gd})} \simeq \frac{g_m}{2\pi C_{ox}} = \frac{2K V_{ov}}{2\pi C'_{ox}WL} = \frac{\cancel{2}\frac{1}{\cancel{2}}\mu_n \cancel{C'_{ox}}\frac{\cancel{W}}{L}V_{ov}}{2\pi\cancel{C'_{ox}}\cancel{W}L} =$$

$$= \frac{\mu_n V_{ov}}{2\pi L^2} = \frac{\bar{v}}{2\pi L} = \frac{1}{2\pi t_{TR}}$$ → transit time

$\frac{V_{ov}}{L} = F$, $\mu_n F = \bar{v}$

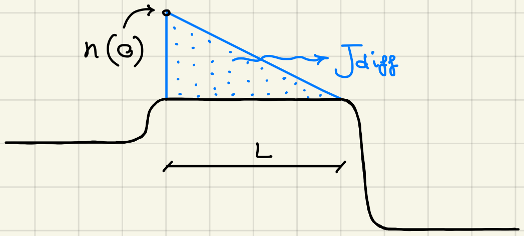


$$\left[f_T = \frac{1}{2\pi t_{TR}} = \frac{\mu_n V_{ov}}{2\pi L^2}\right]$$

shorter channel
higher bandwidth

higher overdrive
higher bandwidth
↓
lower gain
higher bandwidth!
↑
higher overdrive
lower gain

TRADE-OFF!

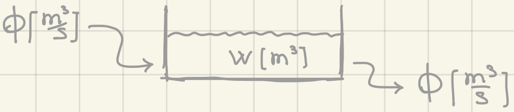


Sub-threshold value:

$$J_{diff} = qD_n\frac{dn}{dx} \simeq qD_n\frac{n(0)}{L}$$

$$Q' = \frac{q}{2}n(0)L$$

$$t_{diff} = \frac{Q'}{J_{diff}} = \frac{\cancel{q}/2\; n(0) L}{\cancel{q}D_n\frac{\cancel{n(0)}}{L}} = \frac{L^2}{2D_n} \leftarrow t_{TR}$$

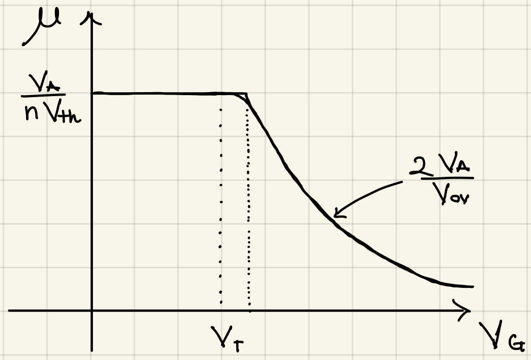


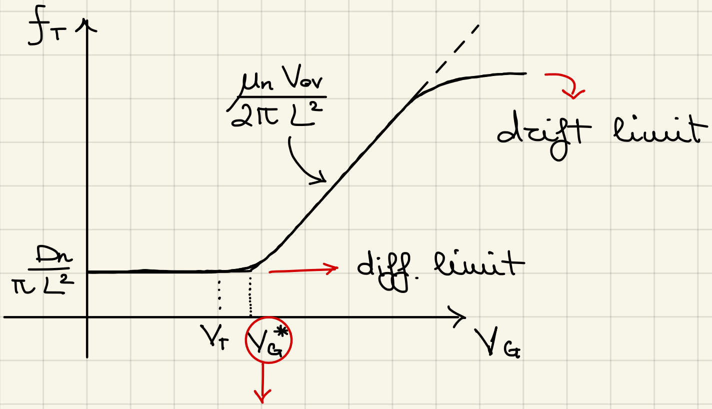
just like a water tank, the ratio W/Φ should give us the average transit time

$$\Rightarrow \left[f_T = \frac{1}{2\pi t_{TR}} = \frac{D_n}{\pi L^2}\right]$$

Now f_T is INDEPENDENT of the OVERDRIVE VOLTAGE.

GAIN/BW TRADE OFF



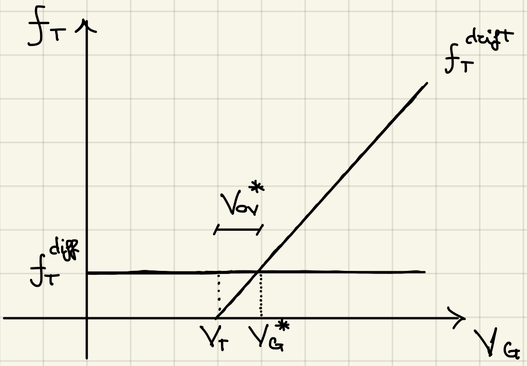


where is this transition?

We need to look for the point where $t_{TR}^{drift} = t_{TR}^{diff}$:

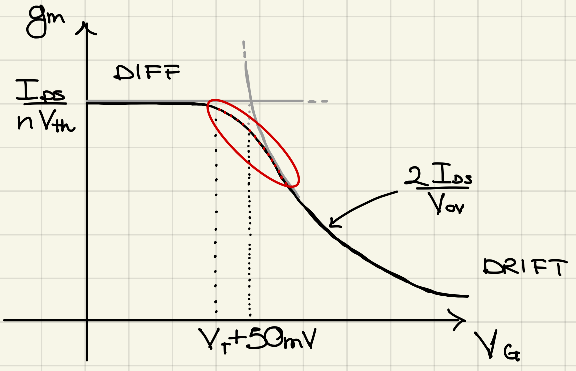
$$t_{TR}^{drift} = \frac{\cancel{L^2}}{\mu_n V_{ov}} = \frac{\cancel{L^2}}{2 D_n} = t_{TR}^{diff}$$

$$\frac{2 D_n}{\mu} = V_{ov} = V_{ov}^*$$



Einstein's equation: $D_n = \frac{KT}{q} \mu_n$

$$\Rightarrow \boxed{V_{ov}^* = \frac{2KT}{q}} \simeq \underline{50 mV}$$



EKV model — to study the exact behaviour of gm at very low overdrive (around threshold)

Enz
Krummenacher
Vittoz

$$\left[g_m = \frac{I_D}{n V_{th}} \cdot \frac{2}{1 + \sqrt{1 + 4\, IC}} \right]$$

IC → Inversion Coefficient

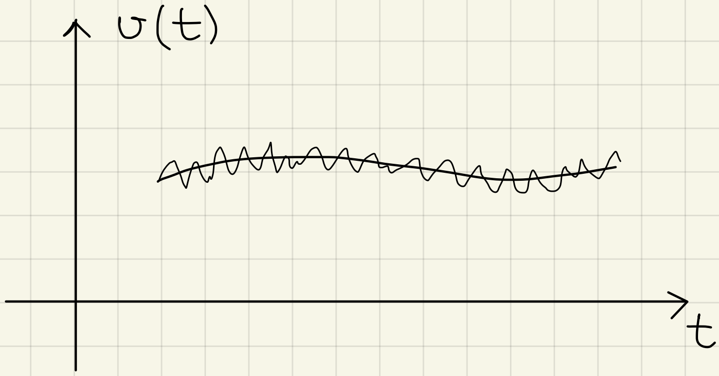
$$IC = \frac{I_D}{n \left[K (2 V_{th})^2 \right]} = \frac{I_D}{n \left[\frac{1}{2} \mu_n C'_{ox} \frac{W}{L} 4 V_{th}^2 \right]} = \frac{I_D}{2 n \mu_n C'_{ox} V_{th}^2 \frac{W}{L}}$$

Weak inversion: $IC \leqslant 0{,}1$

Moderate inversion: $0{,}1 \leqslant IC \leqslant 10$

Strong inversion: $IC \geqslant 10$

Signal, Noise and Disturbs



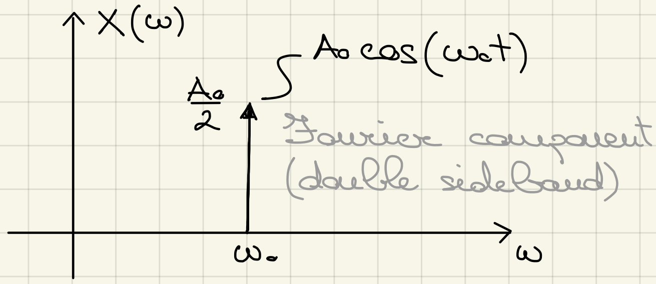
need to handle noise

$$v(t) = \underline{s(t)} + \cancel{d(t)} + \underline{n(t)}$$

can be reduced to negligible values with proper screening

$s(t) \rightarrow X(\omega)$ Spectrum

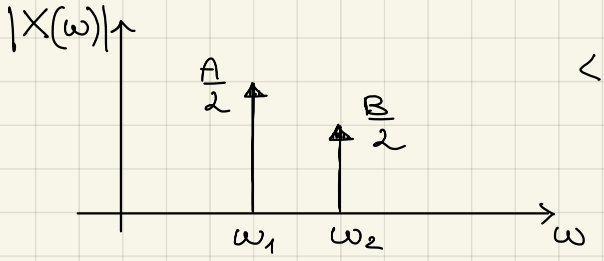
$n(t) \rightarrow S_n(\omega)$ Power Spectral Density (PSD)



Fourier component (double sideband)

$\langle n(t) \rangle = 0$! noise fluctuations have a null average

therefore we consider $\underline{\langle n^2(t) \rangle \neq 0}$ (unless $n(t) \equiv 0$)



$$\langle n^2(t) \rangle = \langle (x_1(t) + x_2(t))^2 \rangle$$

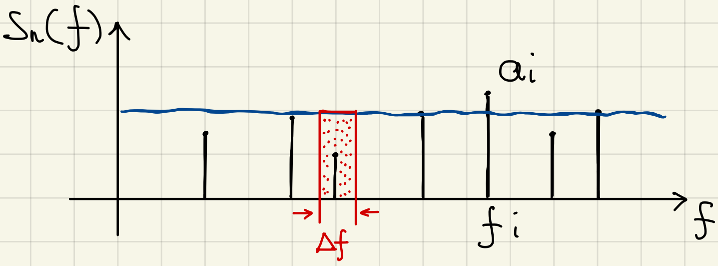
$$= \langle (A \sin(\omega_1 t + \varphi_1) + B \sin(\omega_2 t + \varphi_2))^2 \rangle$$

$$= \langle A^2 \sin^2(\omega_1 t + \varphi_1) + B^2 \sin^2(\omega_2 t + \varphi_2) + + 2AB \sin(\omega_1 t + \varphi_1) \sin(\omega_2 t + \varphi_2) \rangle$$

$$= \lim_{T \to +\infty} \left\{ \frac{1}{T} \int_0^T A^2 \sin^2(\omega_1 t + \varphi_1) dt + \frac{1}{T} \int_0^T B^2 \sin^2(\omega_2 t + \varphi_2) dt + + 2\frac{AB}{T} \cancel{\int_0^T \sin(\omega_1 t + \varphi_1) \sin(\omega_2 t + \varphi_2) dt} \right\} =$$

$$\frac{1}{2} \{ \cos[(\omega_1 - \omega_2)t + \varphi_{1-2}] - \cos[(\omega_1 + \omega_2)t + \varphi_{2-1}] \}$$

$$\Longrightarrow \langle n^2(t) \rangle = \frac{A^2}{2} + \frac{B^2}{2}$$

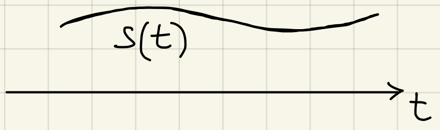


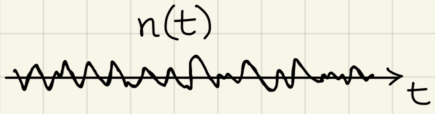
Noise made by many different harmonics throughout ideally the whole spectrum:

$$n(t) = \sum_i a_i \sin(\omega_i t + \varphi_i)$$

$$\langle n^2(t) \rangle = \sum_i \frac{a_i^2}{2} = \int_0^{+\infty} S_n(f) df$$

$$\left[S_n(f) := \frac{\langle n^2(t) \rangle |_{\Delta f}}{\Delta f} \right]$$





$$[n(t)] = V \rightarrow [\langle n^2(t)\rangle] = V^2 \rightarrow [S_n(f)] = \frac{V^2}{Hz}$$

contains ~ 68% of the noise fluctuations

$$p(x) = K e^{-x^2/2\sigma_n^2} \rightarrow$$ the lower σ_n, the more concentrated is the gaussian curve around 0V

The average square value or variance of the noise is a measure of the amplitude of the fluctuation

$$\sigma_n = \sqrt{(\sigma_n^2)} = \sqrt{\langle n^2(t)\rangle}$$

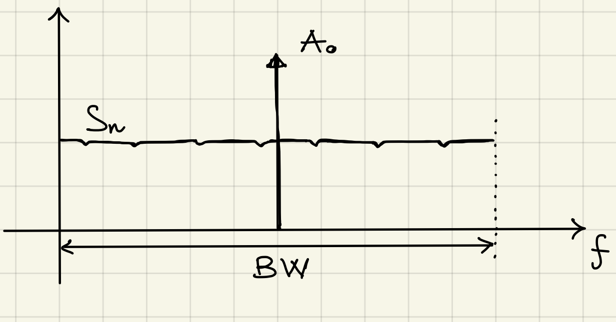
σ_n ↙ root mean square (RMS)

σ_n^2 ↘ variance

The RMS value gives a more precise measure of the average amplitude of the noise:

- within $\pm\sigma_n$: around 68,3% of all fluctuations
- " $\pm 2\sigma_n$: " 95,5% " " "
- " $\pm 3\sigma_n$: " 99,7% " " "

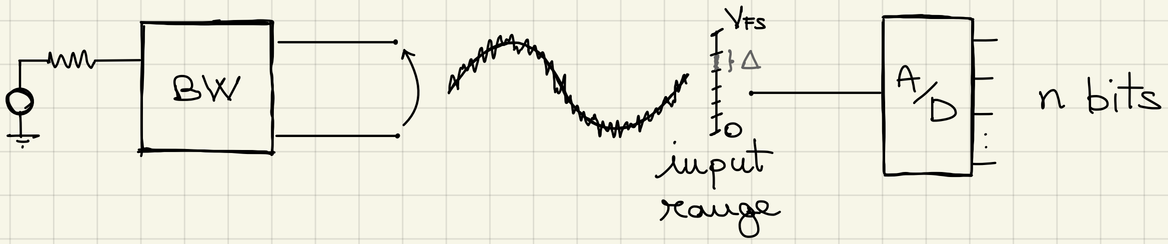
Signal-to-Noise Ratio



SNR: how much larger the signal is compared to the noise

$$\left(\frac{S}{N}\right)^2 = \frac{\langle s^2(t)\rangle}{\langle n^2(t)\rangle} \qquad \longrightarrow \int_0^{BW} S_n(f)\,df$$

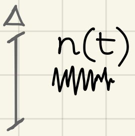
The SNR is related to the information content and quality of our signal



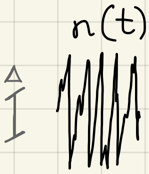
V_{FS} = full scale range

Δ = quantization interval

$\rightarrow \Delta = \frac{V_{FS}}{2^n}$



→ too little precision



→ too much precision

$\Longrightarrow$ Δ should be in the order of $3\sigma_n \div 4\sigma_n$ to have an optimized system

Consider $\Delta = \alpha \cdot \sigma_n$.

$$\left(\frac{S}{N}\right)^2_{MAX} = \frac{\langle (S_{MAX}(t))^2 \rangle}{\langle n^2(t) \rangle} = \frac{\left(\frac{V_{FS}}{2}\right)^2/2}{\sigma_n^2} = \frac{V_{FS}^2\,\alpha^2}{8\Delta^2}$$

$$= \frac{V_{FS}^2\,\alpha^2\,(2^n)^2}{8\,V_{FS}^2} = \underbrace{\boxed{\frac{\alpha^2}{8}}}_{\sim 1}\,2^{2n}$$

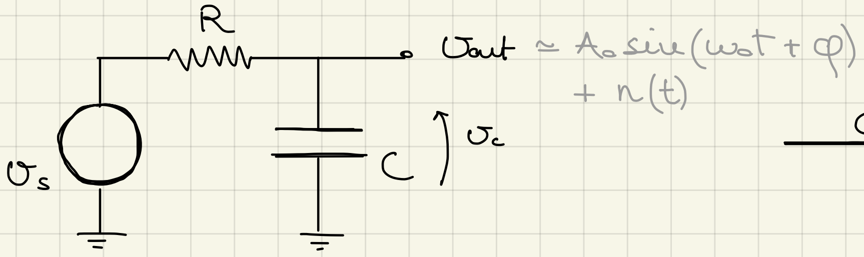
$$\Longrightarrow \left(\frac{S}{N}\right)^2_{MAX} \simeq 2^{2n}$$

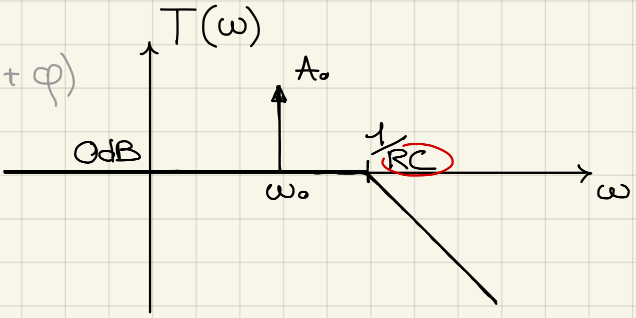
$$10\log\left(\frac{S}{N}\right)^2 = 20\log\left(\frac{S}{N}\right) = 10\log 2^{2n} = 20\log 2^n =$$

$$= n \cdot \underbrace{20\log 2}_{6{,}02\,dB}$$

$\Longrightarrow$ $\boxed{\left(\frac{S}{N}\right)_{dB} = n \cdot 6{,}02\,dB}$ → the number of bits (information content) is defined by the signal-to-noise ratio

Example: RC network





$v_s = A_0 \sin(\omega_0 t)$

$$\left(\frac{S}{N}\right)^2 = \frac{A_0^2/2}{\langle v_c^2 \rangle}$$

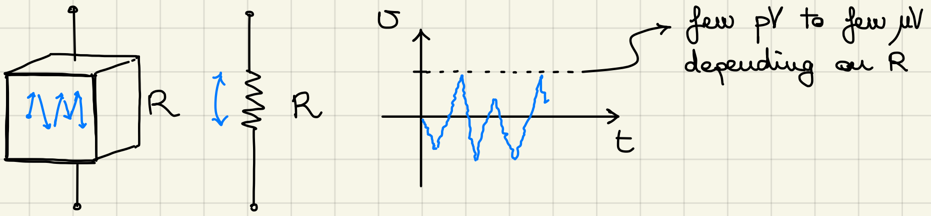
↓

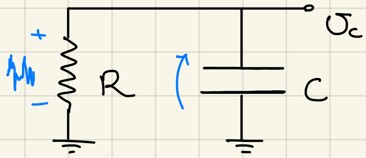
1) where is the noise coming from ?

2) $S_n(f) = ?$

3) $\int_0^{+\infty} S_n(f)\,df = \sigma_n^2 = ?$

1) The RESISTOR is a noise source due to thermal agitation of electrons



3) 

We are looking for $\langle v_c^2 \rangle = \int_0^{+\infty} S_n(f)\,df$ with the signal turned off.

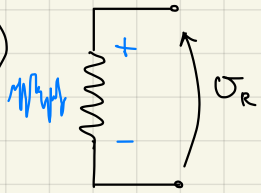
$$\mathcal{E} = \frac{1}{2} C v_c^2(t) \qquad \langle \underline{\underline{\mathcal{E}}} \rangle = \frac{1}{2} C \langle \underline{\underline{v_c^2(t)}} \rangle \overset{\downarrow}{=} \boxed{\frac{1}{2} KT} \longrightarrow$$

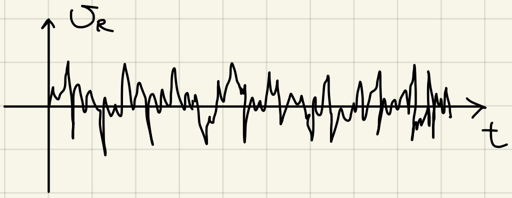
Boltzmann's law for 1 d.o.f. systems

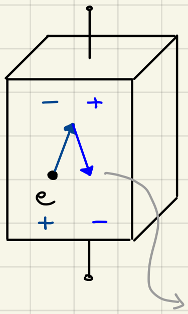
The energy of the system can only be stored in the capacitance. The only source of energy of the system is thermal energy. Therefore the two must be equal

$$\Longrightarrow \uparrow\left(\frac{S}{N}\right)^2 = \frac{A_0^2/2}{\langle v_c^2 \rangle} = \frac{A_0^2/2}{KT/C} = \frac{C}{KT}\uparrow \frac{A_0^2}{2}$$

To have a high SNR while keeping the circuit pole at the desired frequency, we must increase C and decrease R accordingly.

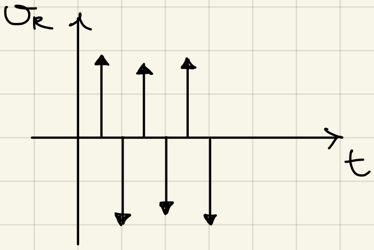
2) 



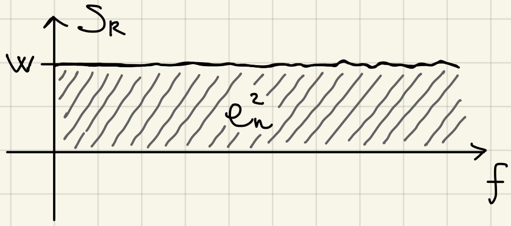


τ is in the order of [fs] which is much smaller than the typical observation time.

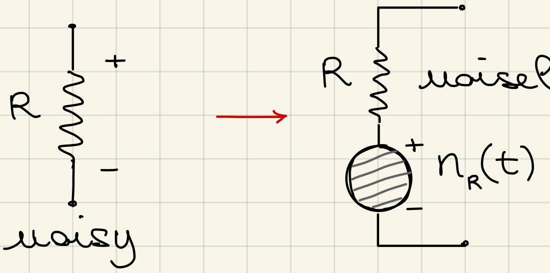
it takes a time equal to τ for the electron to bounce back to its original position, where τ is the scattering time due to the resistor's particles reticulum.



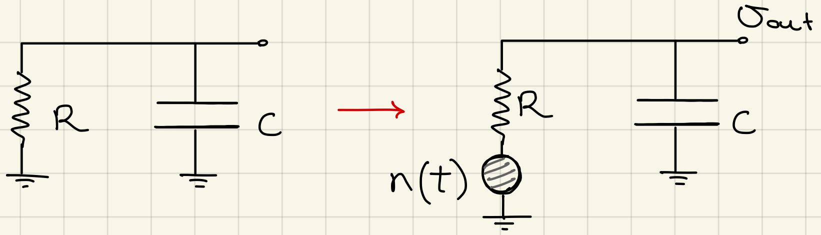
v_R can then be approximated as a series of delta-like pulses, whose correlation time tends to zero.



Therefore the PSD of the thermal noise can be represented by a white noise as it is the superposition of the Fourier transforms of delta-like pulses.



$$\rightarrow \quad S_R(f) = W = \text{const.}$$



$$T(s) = \frac{1/sC}{1/sC + R} = \frac{1}{1+sCR} = \frac{1}{1+\frac{s}{\omega_0}}$$

$$RC = \frac{1}{\omega_0}$$

We don't care about the phase shift of the output noise, just about its amplitude:

$$|T(j\omega)| = \left|\frac{1}{1+j\omega/\omega_0}\right| = \frac{1}{\sqrt{1+\omega^2/\omega_0^2}}$$

$$\langle v_{out}^2(t)\rangle = \langle (n(t)\,|T(j\omega)|)^2\rangle = \langle n^2(t)\rangle\,|T(j\omega)|^2 = e_n^2\,|T(j\omega)|^2$$

$$= \frac{e_n^2}{1+\omega^2/\omega_0^2} = \frac{S_R(f)\,\Delta f}{1+\omega^2/\omega_0^2}$$

def. of $S_n(f)$

$$\Longrightarrow \quad \langle v_{out}^2(t)\rangle = \int_0^{+\infty} \frac{S_R(f)\,df}{1+\omega^2/\omega_0^2}$$

$$\langle v_{out}^2(t)\rangle = \langle v_C^2(t)\rangle = \int_0^{+\infty} \frac{W\,df}{1+\omega^2/\omega_0^2} = \frac{\omega_0 W}{2\pi}\int_0^{+\infty} \frac{df\,2\pi/\omega_0}{1+\omega^2/\omega_0^2} =$$

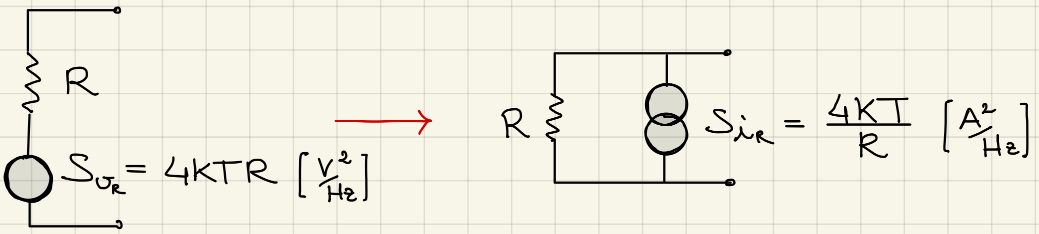
$$x = f\,\frac{2\pi}{\omega_0} = \frac{\omega}{\omega_0}$$

$$= \frac{\omega_0 W}{2\pi}\int_0^{+\infty}\frac{dx}{1+x^2} = \frac{\omega_0 W}{2\pi}\,\frac{\pi}{2} = \frac{\omega_0 W}{4}$$

$$\frac{W}{4RC} = \langle v_c(t)\rangle = \frac{kT}{C}$$

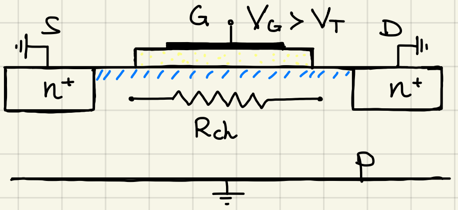
$$\Longrightarrow \quad W = \boxed{S_{v_R}(f) = 4kTR}$$

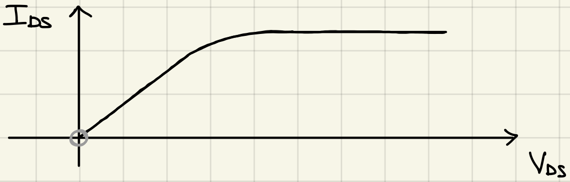
Thevenin / Norton transformations

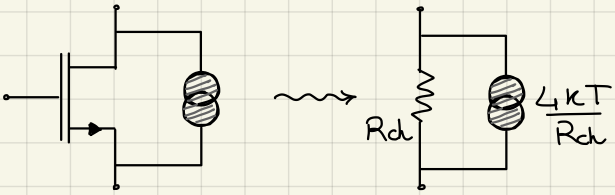


For a resistor R = 1kΩ the associated power spectral density is $S_{v_R} = (4\,nV)^2/Hz$ (@ room temperature T = 300K)

What about thermal noise in transistors?







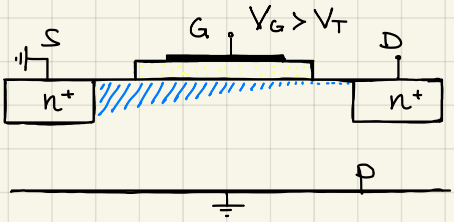
$$I_{DS} = K\left[2(V_{GS}-V_T)V_{DS} - V_{DS}^2\right]$$

$[V_{DS} \to 0]$ then $I_{DS} \to 2K(V_{GS}-V_T)V_{DS} = G_{ch} V_{DS}$

$$\Longrightarrow \boxed{R_{ch} \simeq \frac{1}{2K(V_{GS}-V_T)} = \frac{1}{g_m} \quad \text{when } V_{DS} \simeq 0}$$

$$\longrightarrow \boxed{S_{i_T} = \frac{4kT}{R_{ch}} = 4kTg_m}$$

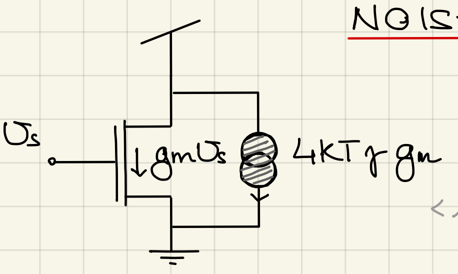
What about other operating points, such as $[V_{DS} > V_{ov}]$?



$$\boxed{S_{i_T} = 4kT\gamma g_m} \begin{cases} \gamma \simeq 1 & \text{ohmic } \boxed{V_{DS} \simeq 0} \\ \gamma \simeq \frac{2}{3} & \text{saturation } \boxed{V_{DS} > 0} \end{cases}$$

for long channel devices; $\gamma \simeq 2$ for short channel devices

NOISE vs. POWER TRADE-OFF



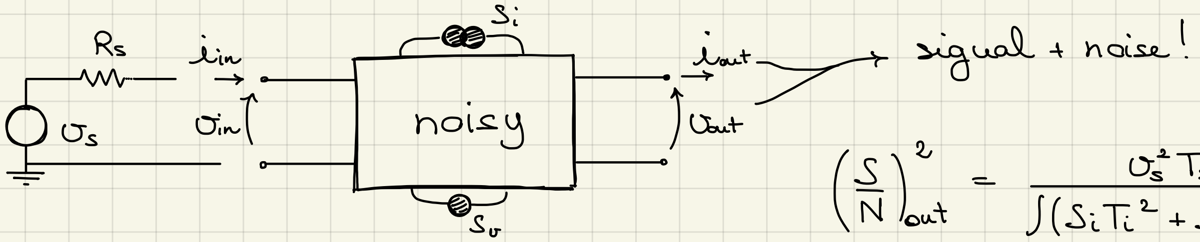
$v_s = A_0 \sin(\omega t)$

$$\left(\frac{S}{N}\right)^2 = \frac{\frac{(g_m v_s)^2}{2}}{4kT\gamma g_m \cdot BW} = \frac{\frac{v_s^2}{2} \cdot g_m}{4kT\gamma \, BW} \qquad g_m \to \frac{2I_D}{V_{ov}} \qquad N: \langle i_n^2 \rangle$$

$\Longrightarrow$ $I_D \downdownarrows$ then $\left(\frac{S}{N}\right)\downarrow$

⟹ By reducing the current (power consumption) we are impairing the signal-to-noise ratio (information content).

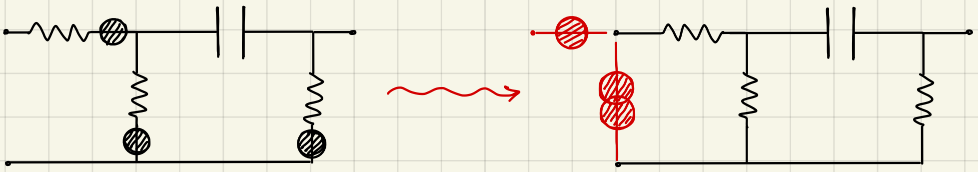
Input-referred noise sources



$$\left(\frac{S}{N}\right)^2_{out} = \frac{v_s^2 T_s^2}{\int (S_i T_i^2 + S_v T_v^2) df}$$

↳ superposition principle

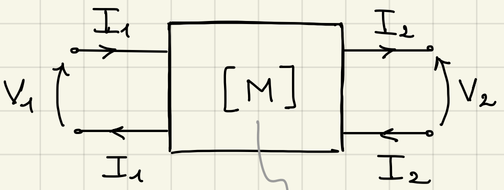
We can simplify the system by "moving" all noise sources at the input and consider everything else noiseless.



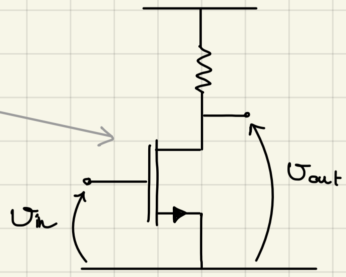
① It is always possible to represent a noisy network as a noiseless network with voltage and current input-referred noise sources, as long as the network is a <u>two-port</u> network

② The PSD of the input-referred noise sources is <u>independent</u> of the input and output (source and load) resistances

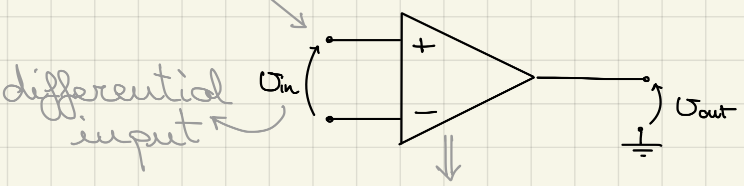
TWO-PORT NETWORKS



$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix}$$



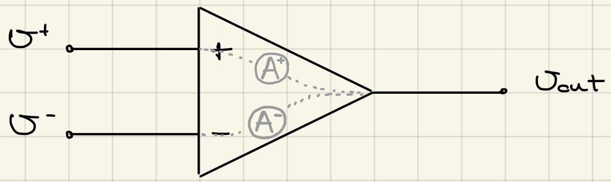
?



⇓

it's a two-port network <u>only if its common mode gain is $G_{cm} = 0$</u>, that is, CMRR = ∞

We can use the input-referred noise representation for a normal amplifier only if its CMRR is very high



$$\begin{cases} v_d = v^+ - v^- \\ v_{cm} = \frac{v^+ + v^-}{2} \end{cases} \rightarrow \begin{cases} v_d = v^+ - v^- \\ 2v_{cm} = v^+ + v^- \end{cases}$$

$$v_d + 2v_{cm} = 2v^+$$

$$v_d - 2v_{cm} = -2v^-$$

$$\Leftarrow \boxed{\begin{aligned} v^+ &= v_{cm} + \frac{v_d}{2} \\ v^- &= v_{cm} - \frac{v_d}{2} \end{aligned}}$$

$$v_{out} = A^+v^+ - A^-v^-$$

$$v_{out} = A^+\left(v_{cm} + \frac{v_d}{2}\right) - A^-\left(v_{cm} - \frac{v_d}{2}\right)$$

$$= \underbrace{(A^+ - A^-)}_{A_{cm}} v_{cm} + \underbrace{\left(\frac{A^+ + A^-}{2}\right)}_{A_d} v_d$$

The amplifier is a good <u>differential</u> amplifier only if

$$A^+ = A^- \Rightarrow A_{cm} = 0, \quad CMRR = \infty$$

and therefore can be considered a two-port network.

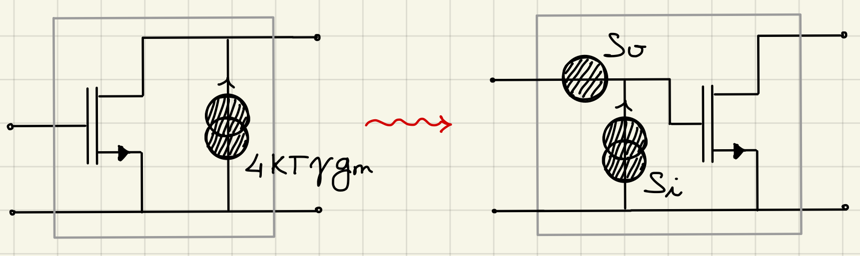
Indeed, if $A_{cm} \neq 0$ then the output would vary by just increasing v^+ and v^- by the same amount, however their difference would still be the same. The amplifier could not be considered a two-port network, since the output would change without any change on the (differential) input.

A good differential amplifier can be represented as a noiseless amplifier through the use of input-referred noise sources.

→ Is it true that any 4-terminal network is a two-port network? NO

Let's now see how to compute the input-referred noises S_v and S_i

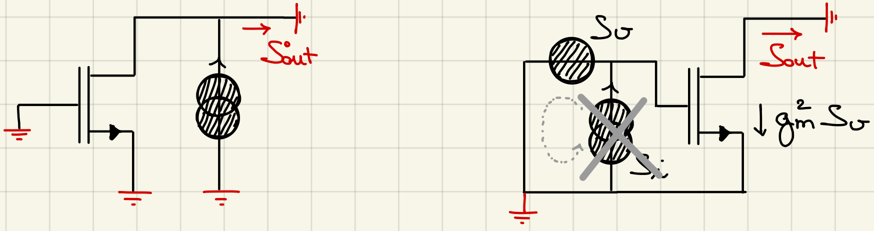
Example:



REAL

MODEL

- To calculate S_v, short input and output terminals and compare the real current output PSD with the model one.



$$\overset{\circ}{S}_{out} = 4kT\gamma g_m = S_{out} = g_m^2 S_v \Rightarrow \boxed{S_v = \frac{4kT\gamma}{g_m}}$$

- To calculate S_i, short output and open input terminal and compare the current output PSD

this current can only flow through the transistor's parasitic capacitances

$$C_{ox} \simeq C_{gs} + C_{gd}$$

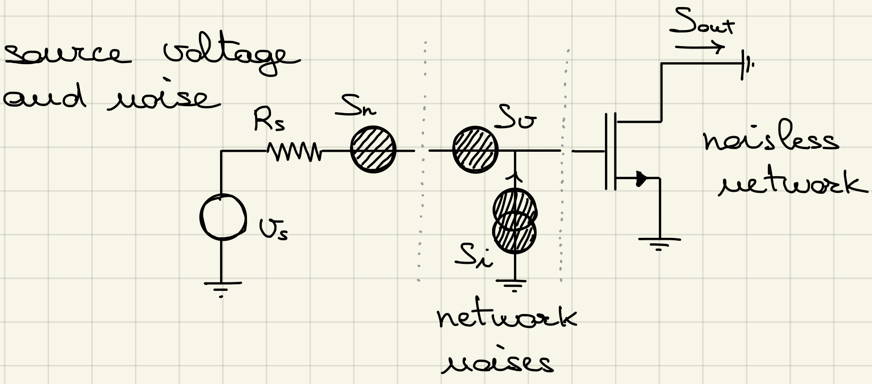
$$S_{out} \approx g_m^2 S_{v_g} = g_m^2 S_i \left| \frac{1}{\omega C_{ox}} \right|^2$$

$$= S_i \frac{g_m^2}{\omega^2 C_{ox}^2} = \overset{\circ}{S}_{out} = 4kT\gamma g_m$$

$$\Longrightarrow S_i \approx 4kT\gamma g_m \cdot \frac{\omega^2 C_{ox}^2}{g_m^2} \quad \omega_T = 2\pi f_T$$

$$= \boxed{4kT\gamma g_m \left(\frac{\omega}{\omega_T}\right)^2}$$

Note that, because of rule ②, these results are still valid even if there was a resistance load attached to the transistor drain



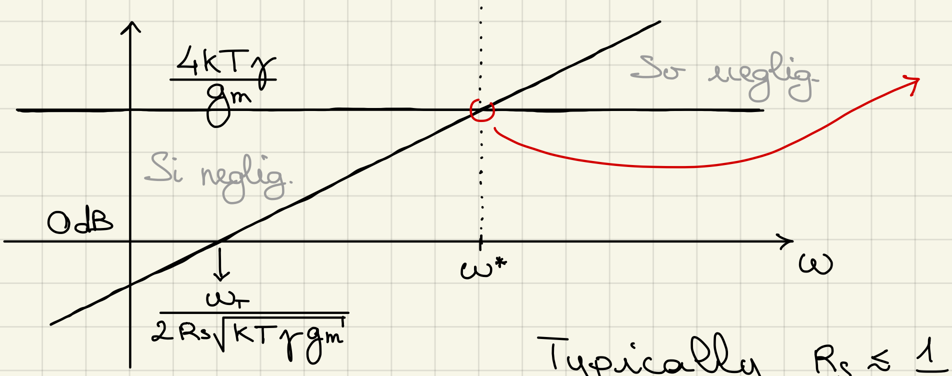
Instead of computing each noise contribution on Sout we can calculate their effects on the network input (the transistor gate) and then utilize the network transfer function to obtain S_{out}. In this way we can avoid using many different transfer functions for each noise source.

$$S_{v_g} = S_R + S_v + S_i \cdot R_S^2 \quad \rightarrow \quad S_{out} = S_{v_g} \cdot |T(s)|^2$$

Note that it would be otherwise hard to compare different types of noise sources (voltage or current) on the output PSD.

In this particular case, we can see that the most relevant intrinsic noise source depends on the value of R_s

$$S_{out} \propto \underset{\swarrow \atop \frac{4kT\gamma}{g_m}}{S_v} + \underset{\searrow \atop 4kT\gamma g_m \left(\frac{\omega}{\omega_T}\right)^2 R_S^2}{R_S^2 S_i}$$



$$\frac{\cancel{4kT\gamma}}{g_m} = \cancel{4kT\gamma}\, g_m \left(\frac{\omega}{\omega_T}\right)^2 R_S^2$$

$$\omega^2 = \frac{\omega_T^2}{g_m^2 R_S^2}$$

$$\boxed{\omega^* = \frac{\omega_T}{g_m R_S}}$$

Typically $R_S \lesssim \frac{1}{g_m} \rightarrow \omega^* \gtrsim \omega_T \simeq 100\,\text{GHz}!$

⟹ <u>In standard conditions, S_i is negligible.</u> In case of very big R_s or very low ω_T then it should be considered.

An off-topic note: <u>Norton theorem</u>

Will be used to quickly compute transfer function of a network.

1) Compute output current (i_{cc}) as a function of input signal with output shorted to ground

2) Compute output impendance (R_{eq})

3) $T = \frac{V_{out}}{S_{in}} = \frac{i_{cc}(S_{in})}{S_{in}} \cdot R_{eq}$

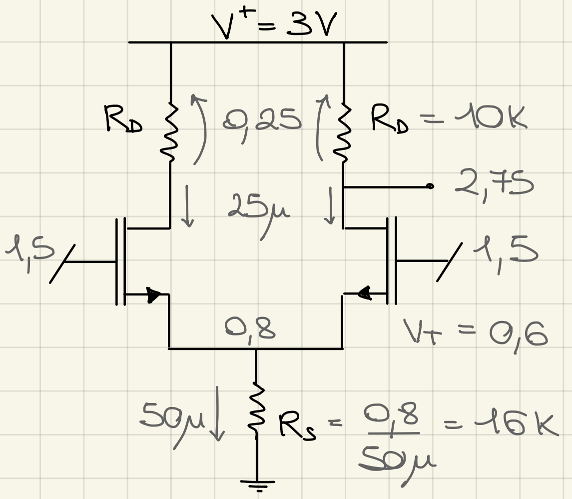
↳ can be either voltage or current signal

CIRCUIT DESIGN

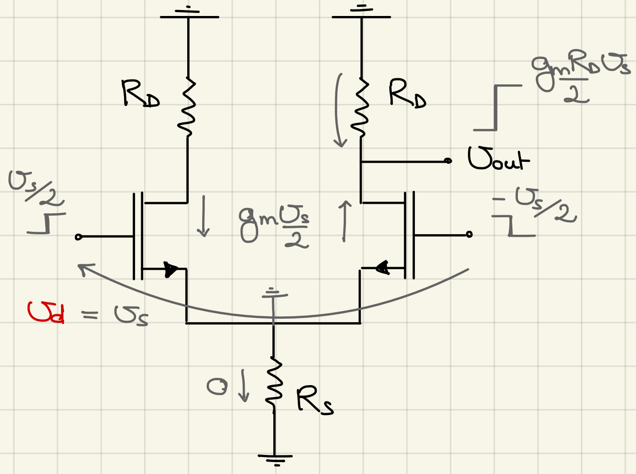
–

Prototypical differential stage

We need: very high $A_d \simeq 10^5$; very low $A_{cm} \simeq 0$ $\}\longrightarrow$ CMRR $\gtrsim$ 100dB



BIAS



SIGNAL

$$\Longrightarrow \boxed{G_d = \frac{v_{out}}{v_d} = \frac{g_m R_D}{2}} =$$

$$= \frac{2 I_D}{V_{ov}} \cdot \frac{R_D}{2} = \frac{I_D R_D \uparrow}{V_{ov} \downarrow} = 2$$

! POOR !

The differential gain is quite low!

To increase the gain, we can either

- decrease the overdrive to increase g_m

  ↓

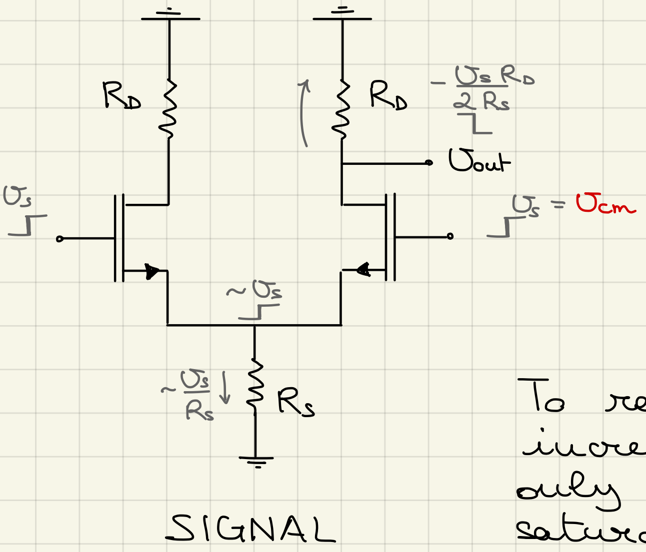
  this can be done only down to the point where g_m saturates to the thermal value $\frac{I_D}{nV_{th}}$ which would increase G_d only up to 5,33

- increase the load R_D

  ↓

  this can be done only up to when the transistors exit saturation and enter ohmic region, that is when the bias point of their drain goes below 0,9V, which represents a voltage drop over the load equal to 2,1V and a resistance R_D equal to 84kΩ, returning a maximum differential gain of 21

The gain can then be increased, but not by much and only through greater power consumptions.



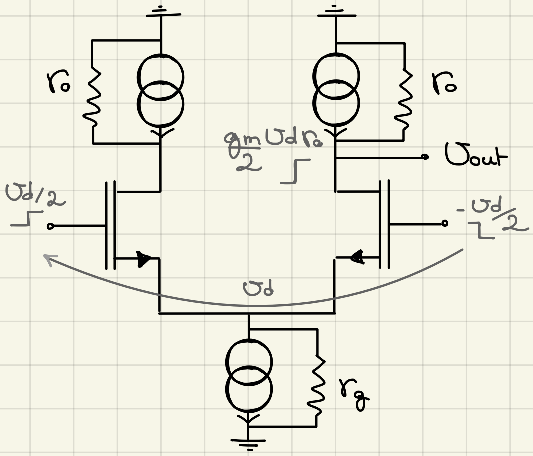
SIGNAL

$$\boxed{G_{cm}} \simeq -\frac{\overset{84K}{R_D}}{\underset{16K}{2R_s}} = -2{,}6 \text{ ! large !}$$

$$CMRR = \left|\frac{\overset{20}{G_d}}{\underset{2,6}{G_{cm}}}\right| \simeq 7{,}7 = 17\,dB \text{ ! POOR !}$$

To reduce the CMRR we can only increase the source resistance, but only up to when the transistors exit saturation.

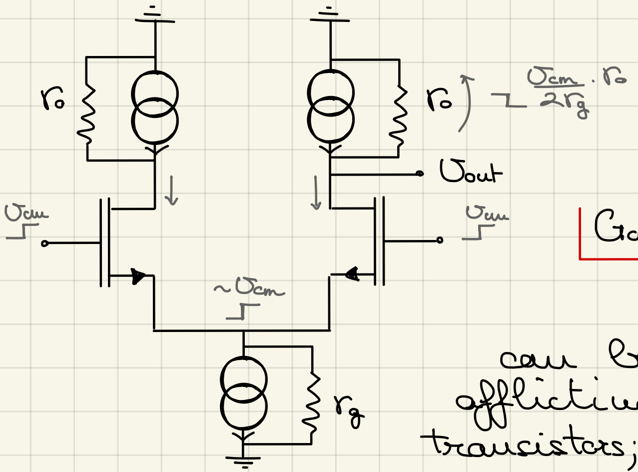
The CMRR can hardly be increased and only through greater power consumptions (higher power supply voltage to improve the stage dynamic).

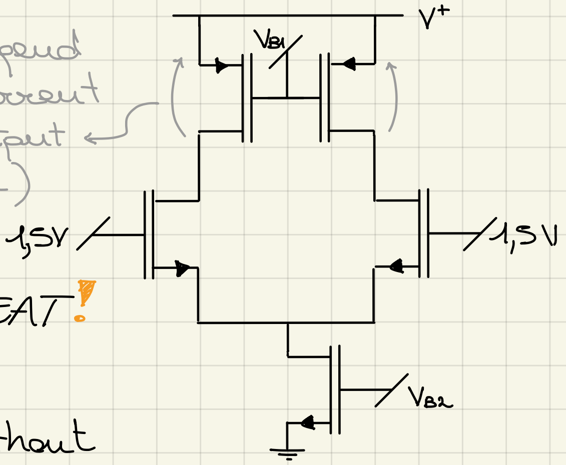


Possible solution:
use current generators (transistors) instead of the resistors

$$\boxed{G_d} = \frac{g_m r_0}{2} = \frac{\mu}{2} \gg 2$$

The stage gain has the same expression as before, but it is now higher and the voltage drop across the load is independent of the current.



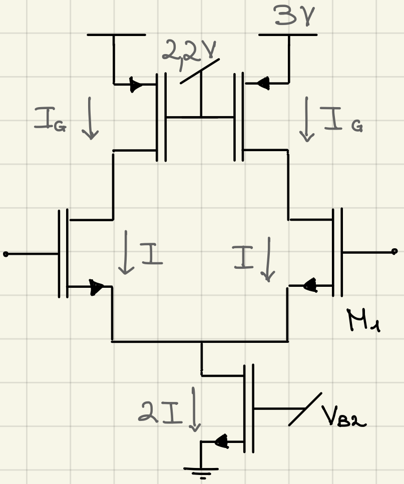


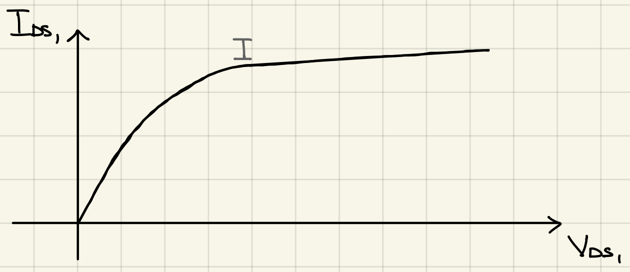
does not depend on the bias current (neglecting output resistance)

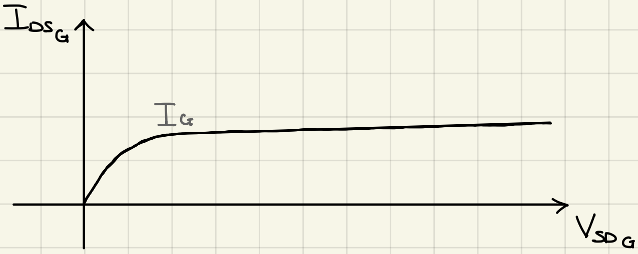
$$\boxed{G_{cm}} = -\frac{r_0}{2r_g} \text{ ! GREAT !}$$

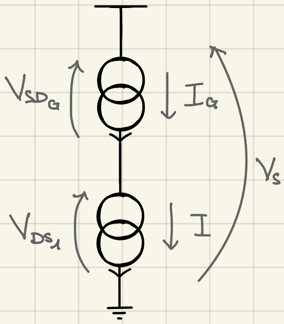
can be decreased without afflicting the operating region of the transistors; two transistors in cascode mode can give high resistance with low bias voltage drop.

Bias issue: current mismatch





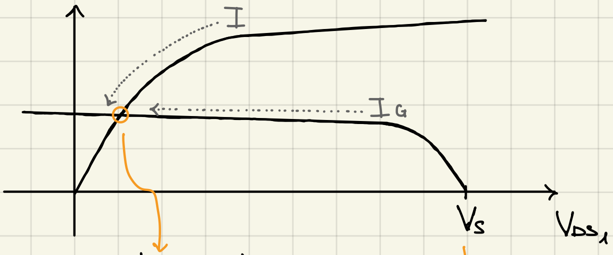




$$V_{SD_G} + V_{DS_1} = V_s$$

$$V_{SD_G} = V_s - V_{DS_1}$$

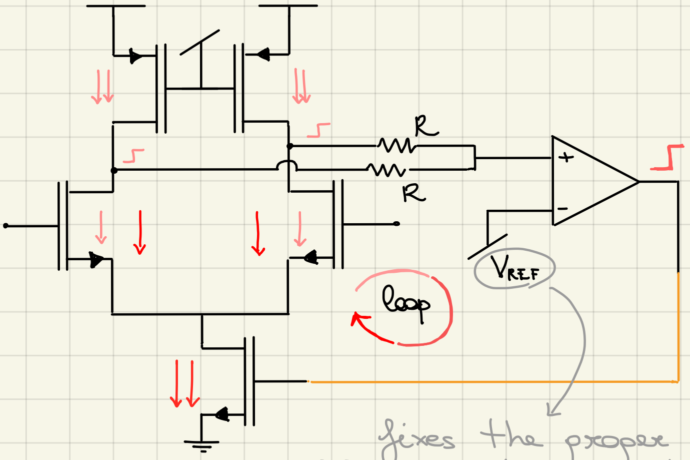
$$I_{SD_G}(V_{SD_G}) = I_{SD_G}(V_s - V_{DS_1})$$



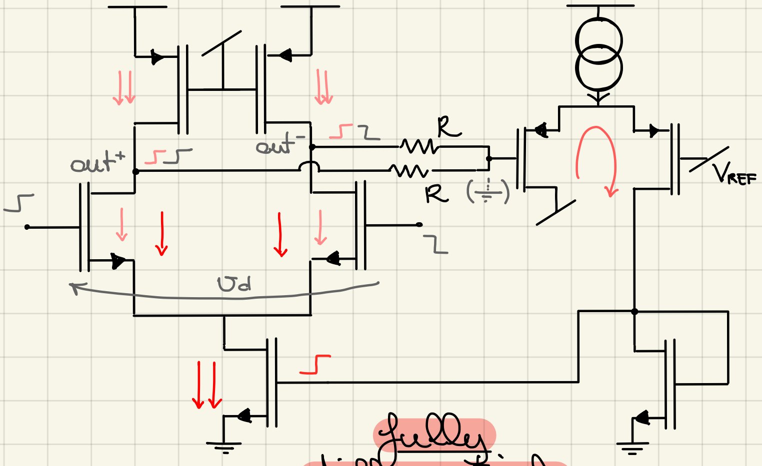
M_1 enters ohmic region!

A small difference in the transistors parameters or bias values will cause either of the two transistors (in each branch of the stage) to exit saturation region.

→ we need a common-mode feedback to properly fix the bias operating point of the stage, while allowing the signal to propagate.



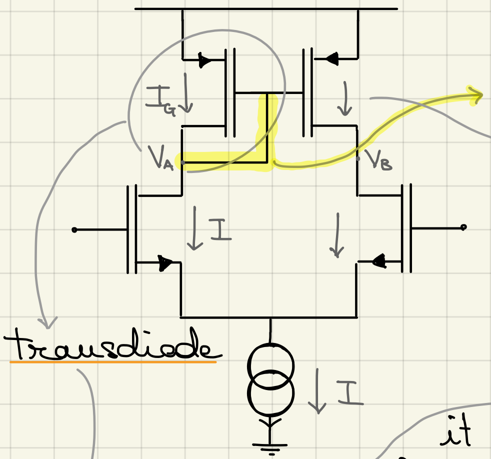
fixes the proper bias operating point



fully differential amplifier

It's a common-mode feedback because it does not affect the differential signal gain while controlling the effects of a common mode signal on the bias.

Better structure for the same amplifier:

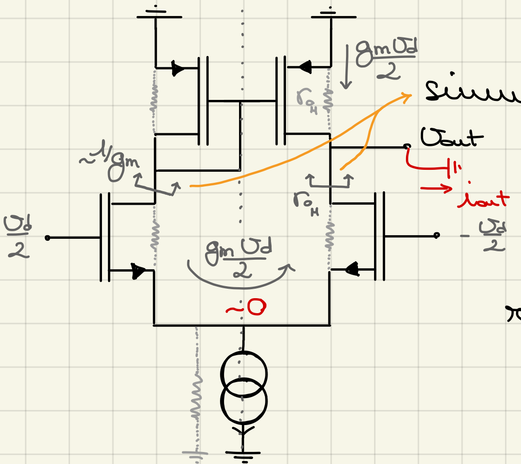


built-in feedback

the current through the right-hand side of the current mirror precisely matches the current on the left-hand side (provided that V_A and V_B are equal, considering the effects of the modulation voltage)

it detects the current flowing through the drain: if $I_G \neq I$ then there must be a voltage change at the drain, which in turn will adjust I_G through the connection to the gate in order to equate I.

Only issue with this structure is that it cannot provide a double-ended output (not fully differential)



simmetry is lost! → we can calculate the gain using Norton's theorem

↓

shorting the output to ground, we can approximate $\frac{1}{g_m} \simeq 0$ and consider the resistances seen from the two drains the same

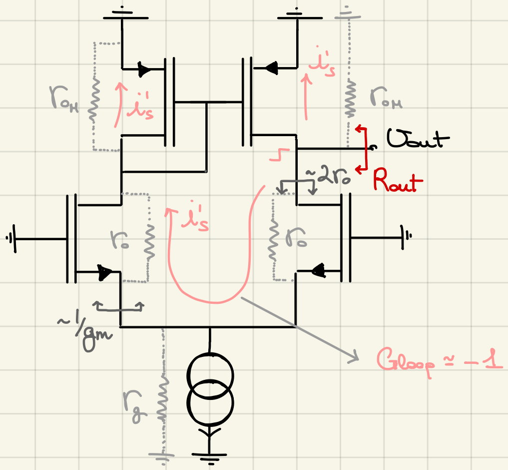
$$\boxed{G_d = \frac{i_{out}}{v_d} \cdot R_{out}}$$

→ $i_{out} = g_m v_d$ $\qquad R_{out} \simeq r_{o_H} // \frac{2r_0}{1-G_{loop}} = r_{o_H} // r_o \simeq \frac{r_o}{2}$

$$G_d = \frac{g_m r_0}{2} = \frac{2I}{V_{ov}} \cdot \frac{V_A}{I} \cdot \frac{1}{2} = \frac{V_A}{V_{ov}} \simeq 70$$

grows with L

does not depend on current

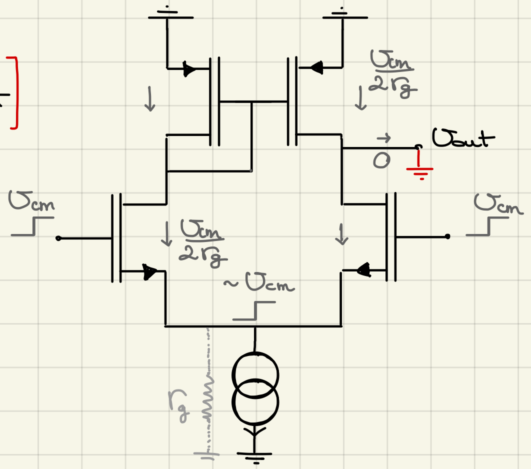


$$\boxed{G_{cm} = \frac{i_{out}}{v_{cm}} \cdot R_{out}}$$

↓

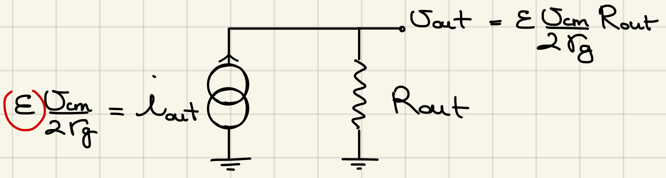
$i_{out} \simeq 0$

ideally $G_{cm} \simeq 0$



but in truth there might be a small current mismatch:

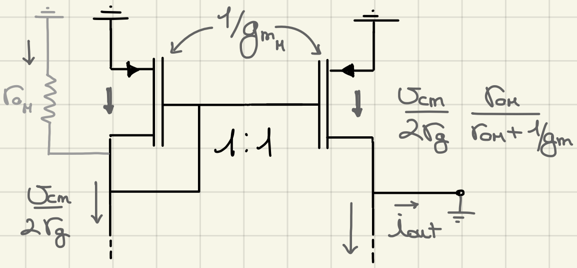
$$G_{cm} = \frac{\varepsilon}{2r_g} R_{out}$$



then the CMRR would be:

$$CMRR = \frac{G_d}{G_{cm}} = \frac{\frac{g_m r_o}{2}}{\varepsilon \frac{R_{out}}{2r_g}} = \frac{2 g_m r_g}{\varepsilon}$$

How to compute the current error ε?



$$i_{out} = \frac{v_{cm}}{2r_g} - \frac{v_{cm}}{2r_g}\left(\frac{r_{oM}}{r_{oM}+1/g_{m_M}}\right) =$$

$$= \frac{v_{cm}}{2r_g}\left(\frac{1/g_{m_M}}{r_{oM}+1/g_{m_M}}\right) = \frac{v_{cm}}{2r_g}\frac{1}{1+g_{m_M} r_{oM}} \simeq$$

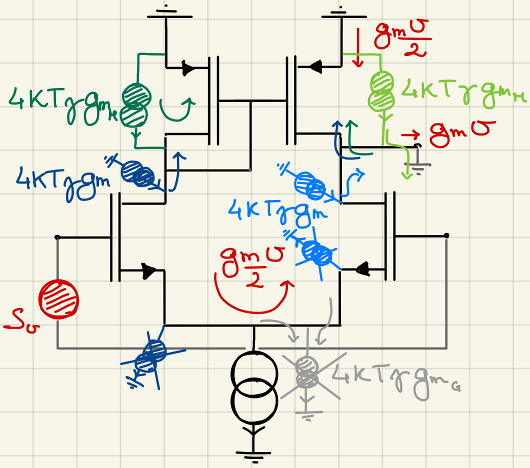
$$\simeq \frac{v_{cm}}{2r_g}\frac{1}{\mu}$$

$$\varepsilon \simeq \frac{1}{\mu} \simeq 10^{-2}$$

it's actually slightly larger because of other non-idealities

$$CMRR = \frac{2 g_m r_g}{\varepsilon} \simeq 2 \cdot 10^4 = 86 dB$$! GREAT !

Input referred noise



$CMRR > 10^4$ → stage can be seen as a two-port network

Compute S_v: $S_{out}|_v = S_v g_m^2$

$$S_{out} = 8KT\gamma g_{m_M} + 8KT\gamma g_m$$

→ $$S_v = \frac{8KT\gamma}{g_m}\left[1 + \frac{g_{m_M}}{g_m}\right]$$

$$= \frac{8KT\gamma}{g_m}\left[1 + \frac{\overset{0,1V}{V_{ov}}}{\underset{0,2V}{V_{ov_M}}}\right] \simeq \left(5 nV/\sqrt{Hz}\right)^2$$

higher transconductance

‖

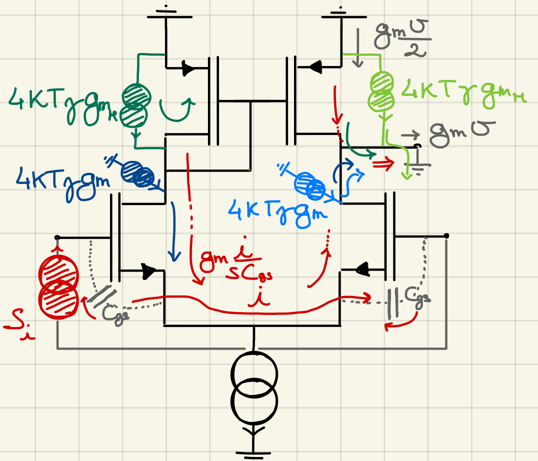
higher bias current

‖

greater power consumption

overdrive of the mirror higher than the input pair overdrive

lower noise



(neglecting C_{gd})

Compute S_i: $S_{out}\big|_i = 4S_i\left(\frac{\omega_T}{\omega}\right)^2 = S_i\left|2\frac{g_m}{j\omega C_{gs}}\right|^2$

$$S_{out} = 8KT\gamma g_m\left[1+\frac{g_{mM}}{g_m}\right]$$

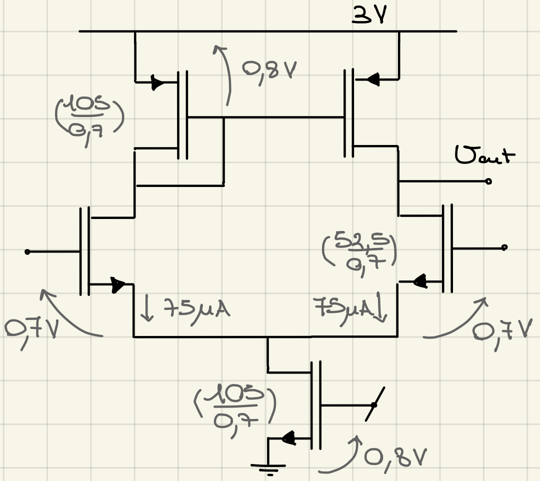
$$\longrightarrow S_i = 2KT\gamma g_m\left[1+\frac{V_{ov}}{V_{ovM}}\right]\left(\frac{\omega}{\omega_T}\right)^2$$

$$= S_v\frac{g_m^2}{4}\left(\frac{\omega}{\omega_T}\right)^2$$

at low frequencies $S_v \gg S_i$

(considering an input resistance in the order of $\frac{1}{g_m}$)

To summarize what we've got so far:



$K'_n = 50\,\mu A/V^2$ $\quad K'_p = 25\,\mu A/V^2$

$V_T = 0,6\,V$ $\quad V_A^\circ = 7\,V$ @ $L_{min} = 0,35\,\mu m$

→ $S_v = \frac{8KT\gamma}{g_{m\,in}}\left(1+\frac{V_{ov\,in}}{V_{ovM}}\right) \leq \left(5\frac{nV}{\sqrt{Hz}}\right)^2$

→ $V_{ov\,in} = 0,1V$ $\quad V_{ovM} = 0,2V$ $\quad g_{m\,in} \geq 1,2\,\frac{mA}{V}$

$\rightarrow g_{m\,in} = 1,5\,\frac{mA}{V}$ $\quad g_{mM} = 0,75\,\frac{mA}{2}$

→ $G_d = g_{m\,in}(r_{oM} /\!/ r_{o\,in}) = \frac{V_A}{V_{ov\,in}} \geq 100$ $\quad I = \frac{g_m V_{ov}}{2} = 75\mu A$

$V_A \geq 10V \rightarrow V_A = 2V_A^\circ = 14V$ $\quad \left(\frac{W}{L}\right)_{in} = 150$ $\quad \left(\frac{W}{L}\right)_M = 75$

$L = 2L_{min} = 0,7\,\mu m$ $\quad W_{in} = 105\,\mu m$ $\quad W_M = 52,5\,\mu m$

→ $V_{ovg} = 0,2V \rightarrow \left(\frac{W}{L}\right)_g = \frac{2I}{K'_n V_{ovg}} = 150 \longrightarrow W_g = 105\,\mu m$

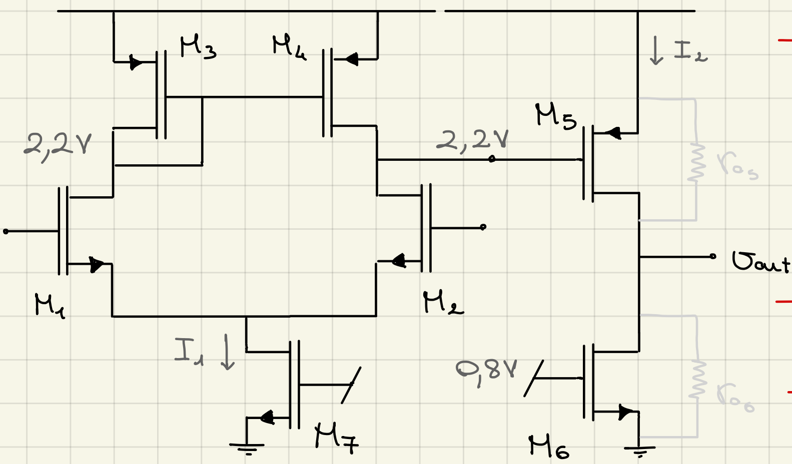
$G_{cm} \simeq \frac{r_{oM} /\!/ r_{o\,in}}{2\mu_M r_g}$ $\quad r_{o\,in} = r_{oM} = 186,6\,k\Omega$ $\quad r_g = 93,3\,k\Omega$

$$\simeq \frac{1}{2\mu_M} = \frac{V_{ovM}}{4V_A} = 0,00357$$

$G_d = 43dB$ $\quad CMRR = \frac{G_d}{G_{cm}} \simeq 2 g_{m\,in} r_g \mu_M = 92dB$

only parameter that should still be improved (up to 100 dB)

→ Add a second stage with high gain:



→ $G_2 = g_{m_5}(r_{o5}//r_{o6}) = \frac{2I_2}{V_{ov_5}} \cdot \frac{V_A}{2I_2} =$

$= \frac{V_A}{V_{ov_5}} = 140$ $\quad V_{ov_5} = V_{ov_{3,4}}$

$= 0,2V$

→ $V_{A_5} = 28V$ $\quad L_5 = L_6 = 1,4\mu m$

→ $I_2 = 150\mu A$ $\quad \left(\frac{W}{L}\right)_5 = 150$

→ $V_{ov_6} = 0,2V$ $\quad \left(\frac{W}{L}\right)_6 = 75$

$W_5 = 210\mu m$ $\quad W_6 = 105\mu m$

$r_{o_5} = r_{o_6} = 186,6K\Omega$ $\quad g_{m_5} = 1,5\frac{mA}{V}$

common source stage
(with active load)

$$G_d = g_{m_1}\underbrace{(r_{o_2}//r_{o_4})}_{140}\, g_{m_5}\underbrace{(r_{o_5}//r_{o_6})}_{140} \simeq 86dB$$! GREAT !

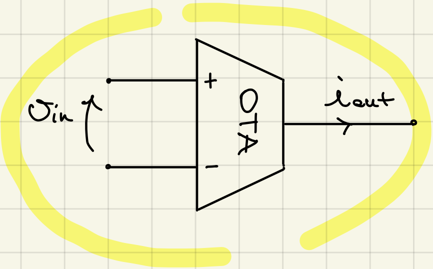
Note that the second stage adds a negligible contribution to the input-referred noise of the overall amplifier.

$$S_v = \frac{8KT\gamma}{g_{m_1}}\left(1 + \frac{V_{ov_1}}{V_{ov_3}}\right) + \frac{4KT\gamma(g_{m5} + g_{m6})}{\boxed{G_1^2} \cdot g_{m_5}}$$

referred to the input, the noise of the second stage is reduced by the gain of the first stage

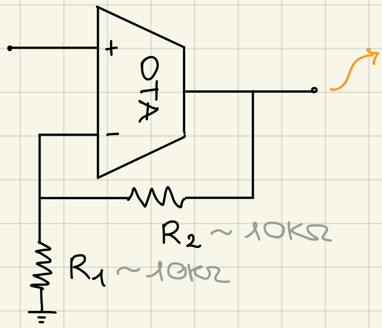
This prototypical differential stage is called Operational Transimpedence Amplifier (OTA)

its output impedance is very large (it amplifies voltage into current)



has a very high gain (transimpedence)

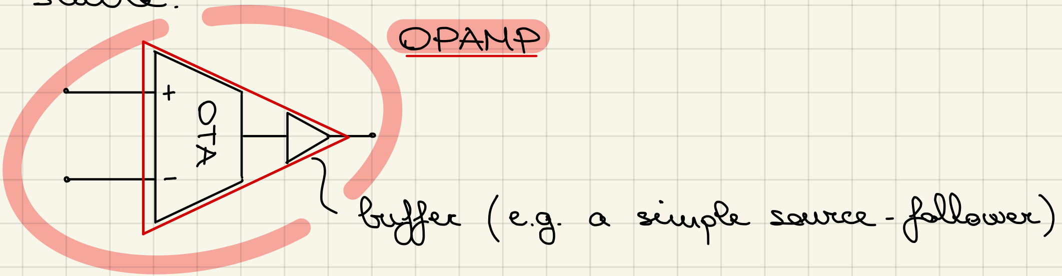
An OTA cannot be used with a low impedance load:



$R_{out} = (r_{o5}//r_{o6})//(R_1 + R_2)$
$\simeq R_1 + R_2$

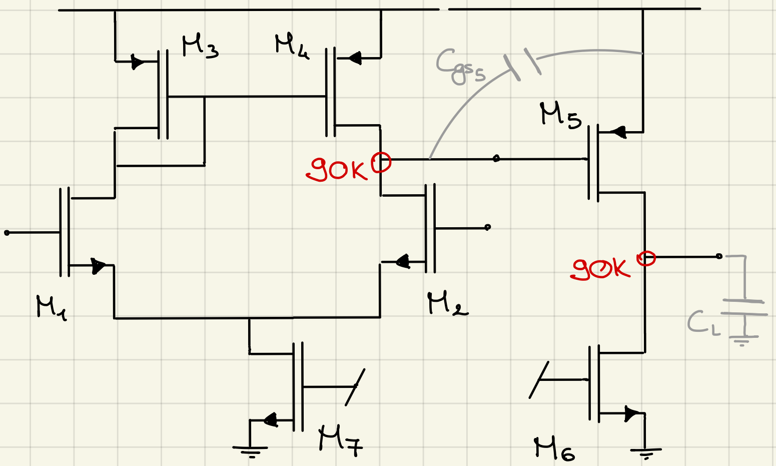
$\downarrow T_{OTA} \propto R_{out}\downarrow$ → the gain of the amplifier goes down with its output resistance

That is why a generic operational amplifier is made of an OTA connected to an output buffer so that its output impedance is not modified by the load and its gain remains stable.



An OPAMP can be connected to whatever load. An OTA must be connected to a high impedance load.

|| Frequency Response and Compensation ||



$$f_p = \frac{1}{2\pi C R_{eq}}$$

i.e. within the frequency range of interest

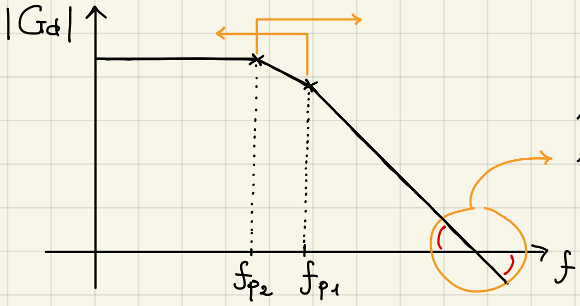
Since lower frequency poles are found at high resistance nodes, we are better off considering the capacitances seen at only those nodes.

Note: for each high-gain stage, there exists 1 high impedance node

Most relevant capacitances seen at the two high-imp. nodes:

$C_{gs_5} = C'_{ox} (WL)_5 \cdot 2/3 \simeq 1 pF$ $\qquad$ $C_L \simeq 2pF$

$f_{p1} = 17 MHz$ $\qquad$ $f_{p2} = 8,5 MHz$

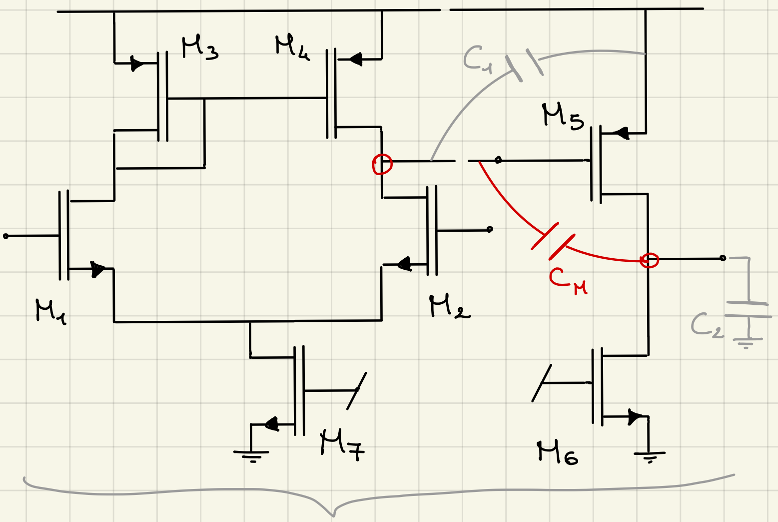


it can easily become unstable if used in any low-gain neg. feedback circuit (e.g. buffer)

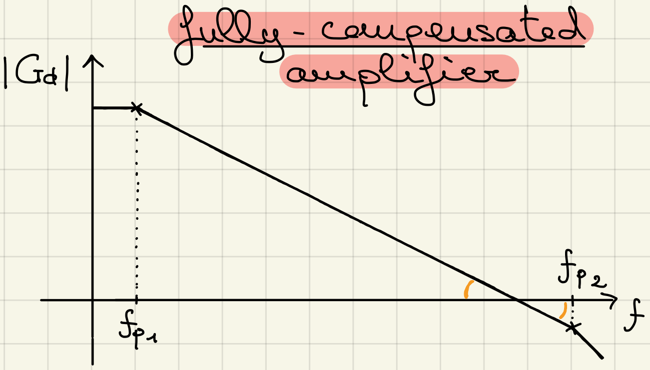
The two poles are too close and produce a bad closure angle of the transfer function.

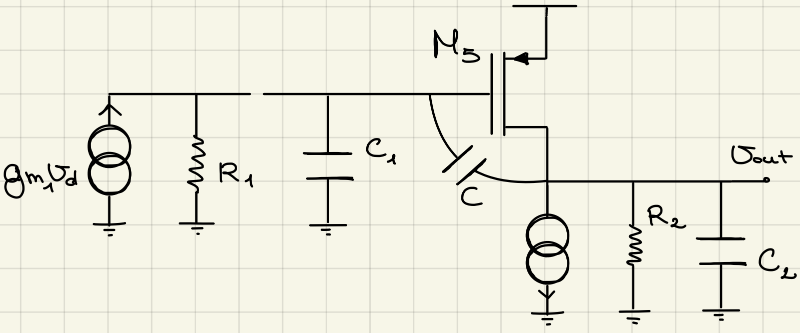
We have to split them apart in order to cut the 0dB axis with a $-20 \frac{dB}{dec}$ slope. How can it be done?

→ Insert a Miller capacitance that connects the two nodes



fully-compensated amplifier





$R_1 = r_{o2} // r_{o4}$ $R_2 = r_{o5} // r_{o6}$

$$T(s) = -[g_{m1}R_1 g_{m5}R_2] \frac{a_2s^2+a_1s+1}{b_2s^2+b_1s+1}$$

How do we compute f_z, f_p?
And a_2, a_1, b_2, b_1?

TIME CONSTANT METHOD

$$T(s) = G_{T0} \frac{a_2s^2+a_1s+1}{b_2s^2+b_1s+1}$$

G_{T0} → DC gain

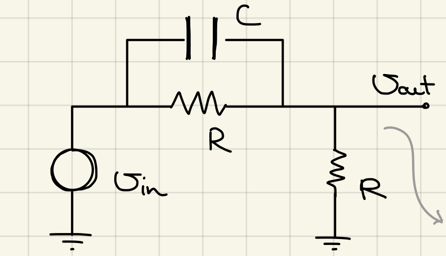
numerator is polinomial of order equal to the number of reactive capacitances when output is set to zero voltage

denominator is polinomial of order equal to the number of independent capacitances in the circuit

independent capacitances → number of capacitances across which I can freely set any voltage

be careful that in this case, C, C1 and C2 are DEPENDENT on each other

Example:



→ first order network

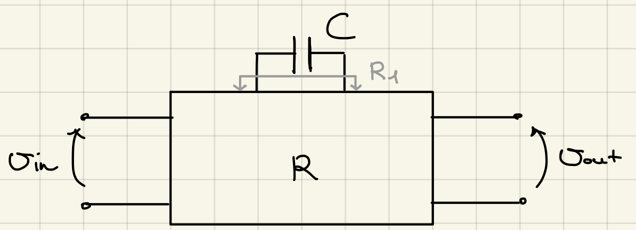
$$\frac{v_{out}}{v_{in}} = \frac{a_1 s + 1}{b_1 s + 1} \cdot G_0$$

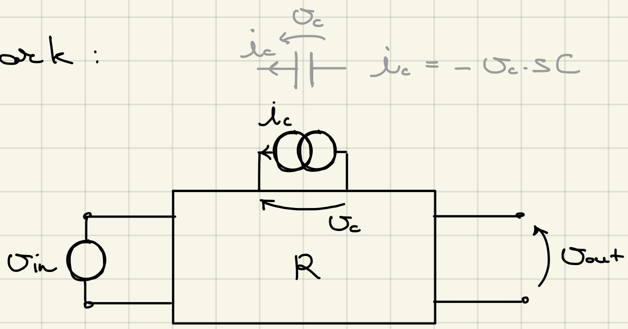
a_1 → C_1R

b_1 → $C_1R/2$

G_0 → $1/2$

Generalized first order network:





$$-v_c sC = i_c$$

$$\hookrightarrow \quad v_c = B_0 v_{in} - R_1 v_c sC$$

$$v_c(1 + sCR_1) = B_0 v_{in}$$

$$v_c = \frac{B_0 v_{in}}{1 + sCR_1}$$

$$\begin{cases} v_{out} = A_0 v_{in} + R_m i_c \\ v_c = B_0 v_{in} + R_1 i_c \end{cases}$$

$$\hookrightarrow \quad R_1 = \left.\frac{v_c}{i_c}\right|_{v_{in}=0}$$

$$\hookrightarrow \quad v_{out} = v_{in}\left[A_0 - \frac{sCR_mB_0}{1+sCR_1}\right]$$

$$= v_{in} \frac{A_0 + sC[R_1A_0 - R_mB_0]}{1+sCR_1}$$

$$\boxed{v_{out} = v_{in} A_0 \frac{1 + sC\left[R_1 - R_m\frac{B_0}{A_0}\right]}{1+sCR_1}}$$

R_{01}

→ 1 pole as expected

Let's better understand the expression of the zero:

$$\begin{cases} v_{out}|_0 = 0 = A_0 v_{in}|_0 + R_m i_c|_0 \\ v_c = B_0 v_{in} + R_1 i_c \end{cases} \longrightarrow \begin{cases} v_{in}|_0 = -\frac{R_m}{A_0} i_c|_0 \\ v_c|_0 = -R_m\frac{B_0}{A_0} i_c|_0 + R_1 i_c|_0 \end{cases}$$

$$\frac{v_c|_0}{i_c|_0} = R_1 - R_m\frac{B_0}{A_0} = R_{01}$$

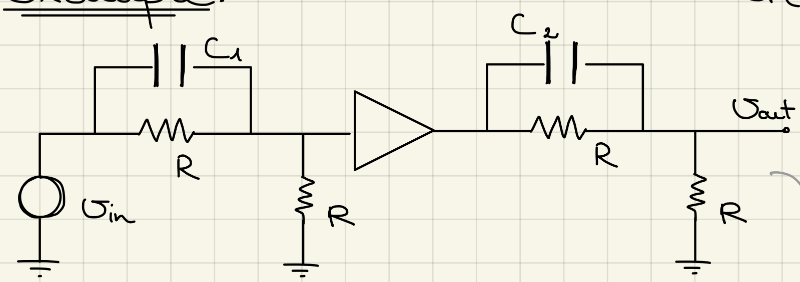
$$\Longrightarrow \quad \frac{v_{out}}{v_{in}} = \frac{a_1s+1}{b_1s+1} G_0$$

→ DC gain (with the capacitor open)

resistance seen from the capacitor when the output is at zero voltage (not ground!) times the capacitance

resistance seen from the capacitor when the input is turned off (shorted if voltage, open if current) times the capacitance

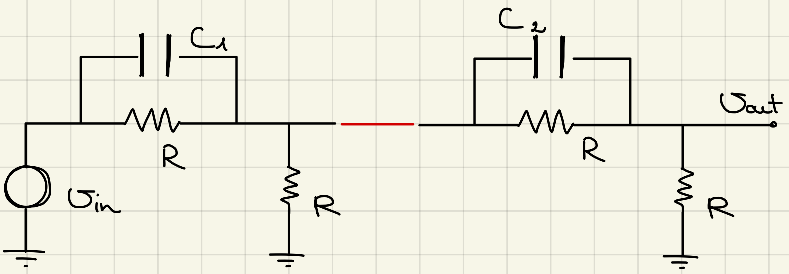
Example:



$$G(s) = \frac{1}{4}\frac{(1+sC_1R)}{(1+sC_1R/2)}\frac{(1+sC_2R)}{(1+sC_2R/2)} =$$

$$= \frac{1}{4}\frac{s^2\overset{a_2}{C_1C_2}R^2 + s(C_1R\overset{a_1}{+}C_2R)+1}{\underset{b_2}{s^2C_1C_2(R/2)^2} + \underset{b_1}{s(C_1R/2 + C_2R/2)}+1}$$

→ second order network



The capacitances are the same; the resistances seen from each capacitor will be different

↓

$$G(s) = G_0 \frac{s^2 C_1 C_2 \alpha_{12} + s[C_1\alpha_1 + C_2\alpha_2] + 1}{s^2 C_1 C_2 \beta_{12} + s[C_1\beta_1 + C_2\beta_2] + 1}$$

C₂ open

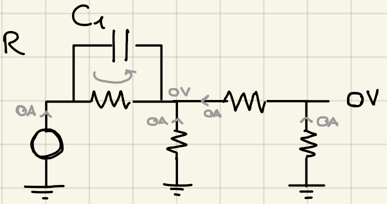
→ $C_2 = 0$ $\quad G(s) = G_0 \frac{sC_1\overbrace{\alpha_1}^{R_{o1}^{(0)}}+1}{sC_1\underbrace{\beta_1}_{R_1^{(0)}}+1}$

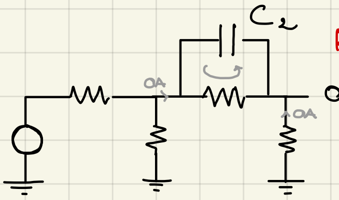
(first order networks)

→ $C_1 = 0$ $\quad G(s) = G_0 \frac{sC_2\overbrace{\alpha_2}^{R_{o2}^{(0)}}+1}{sC_2\underbrace{\beta_2}_{R_2^{(0)}}+1}$

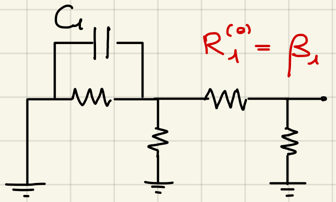
$R_x^{(0)}$:= resistance seen from capacitor x when all the other ones are open

$R_{o1}^{(0)} = \alpha_1 = R$



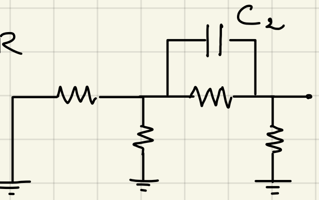


$R_{o2}^{(0)} = \alpha_2 = R$



$R_1^{(0)} = \beta_1 = R\,//\,R\,//\,2R = = \frac{2}{5}R$

$R_2^{(0)} = \beta_2 = \frac{3}{2}R\,//\,R = \frac{3}{5}R$



C₂ shorted

→ $C_2 \longrightarrow \infty$ $\quad G(s) \simeq G_0 \frac{s^2C_1\cancel{C_2}\alpha_{12} + \cancel{s}\cancel{C_2}\alpha_2}{s^2C_1\cancel{C_2}\beta_{12} + \cancel{s}\cancel{C_2}\beta_2} = G_0 \frac{sC_1\alpha_{12} + \alpha_2}{sC_1\beta_{12} + \beta_2}$

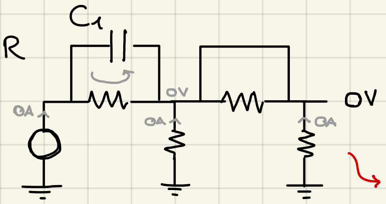
$$= G_0 \frac{\alpha_2}{\beta_2} \frac{sC_1\overbrace{\alpha_{12}/\alpha_2}^{R_{o1}^{(2)}} + 1}{sC_1\underbrace{\beta_{12}/\beta_2}_{R_1^{(2)}} + 1}$$

(first order networks)

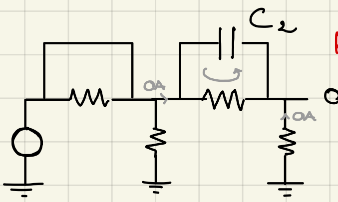
→ $C_1 \longrightarrow \infty$ $\quad G(s) \simeq G_0 \frac{\alpha_1}{\beta_1} \frac{sC_2\overbrace{\alpha_{12}/\alpha_1}^{R_{o2}^{(1)}} + 1}{sC_2\underbrace{\beta_{12}/\beta_1}_{R_2^{(1)}} + 1}$

$R_x^{(y)}$:= resistance seen from capacitor x when capacitor y is shorted and all the other ones are open

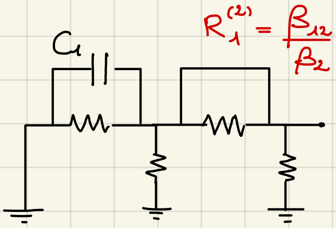
$R_{o1}^{(2)} = \frac{\alpha_{12}}{\alpha_2} = R$



↳ $\alpha_{12} = R^2$ ↲



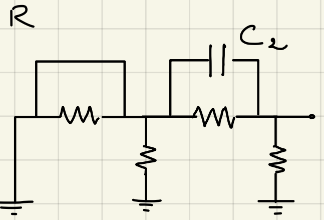
$R_{o2}^{(1)} = \frac{\alpha_{12}}{\alpha_1} = R$



$R_1^{(2)} = \frac{\beta_{12}}{\beta_2} = R \mathbin{/\!/} R \mathbin{/\!/} R = \frac{R}{3}$

$R_2^{(0)} = \frac{\beta_{12}}{\beta_1} = R \mathbin{/\!/} R = \frac{R}{2}$

$\beta_{12} = \frac{R^2}{5}$



Note that while you do need to compute both $R_1^{(0)}$ and $R_2^{(0)}$ ($R_{o1}^{(0)}$ and $R_{o2}^{(0)}$) to obtain β_1 and β_2 (α_1 and α_2), you do NOT need to compute both $R_1^{(2)}$ and $R_2^{(1)}$ ($R_{o1}^{(2)}$ and $R_{o2}^{(1)}$) to obtain β_{12} (α_{12}).
Computing just one of the two will suffice.

$$\boxed{\begin{array}{ll} \alpha_1 = R_{o1}^{(0)} & \beta_1 = R_1^{(0)} \\ \alpha_2 = R_{o2}^{(0)} & \beta_2 = R_2^{(0)} \\ \alpha_{12} = R_{o1}^{(2)} \cdot R_{o2}^{(0)} = R_{o2}^{(1)} \cdot R_{o1}^{(0)} & \beta_{12} = R_1^{(2)} R_2^{(0)} = R_2^{(1)} R_1^{(0)} \end{array}}$$

$$C_2 C_1 R_{o2}^{(1)} R_{o1}^{(0)} = C_1 C_2 R_{o1}^{(2)} R_{o2}^{(0)} \qquad C_1 R_{o1}^{(0)} + C_2 R_{o2}^{(0)}$$

$$G(s) = G_0 \frac{s^2 a_2 + s a_1 + 1}{s^2 b_2 + s b_1 + 1}$$

$$C_2 C_1 R_2^{(1)} R_1^{(0)} = C_1 C_2 R_1^{(2)} R_2^{(0)} \qquad C_1 R_1^{(0)} + C_2 R_2^{(0)}$$

What about third+ order networks?

$b_1 = \tau_1^{(0)} + \tau_2^{(0)} + \tau_3^{(0)} = C_1 R_1^{(0)} + C_2 R_2^{(0)} + C_3 R_3^{(0)}$

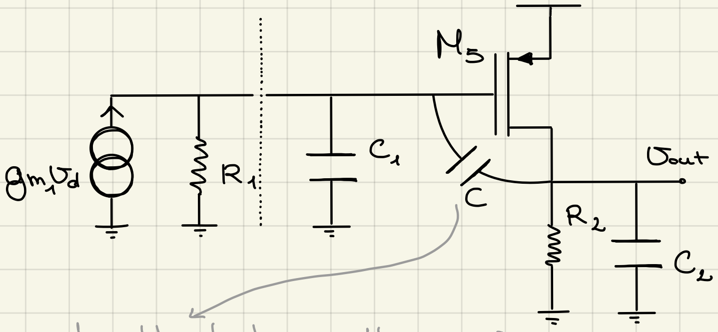
$b_2 = \tau_1^{(2)} \tau_2^{(0)} + \tau_1^{(3)} \tau_3^{(0)} + \tau_2^{(3)} \tau_3^{(0)} =$
$= \tau_2^{(1)} \tau_1^{(0)} + \tau_3^{(1)} \tau_1^{(0)} + \tau_3^{(2)} \tau_2^{(0)} = \dots$

$b_3 = \tau_1^{(2,3)} \tau_2^{(3)} \tau_3^{(0)} = \tau_2^{(1,3)} \tau_3^{(1)} \tau_1^{(0)} = \tau_3^{(1,2)} \tau_1^{(2)} \tau_2^{(0)} =$
$= \tau_1^{(2,3)} \tau_3^{(2)} \tau_2^{(0)} = \tau_2^{(1,3)} \tau_1^{(3)} \tau_3^{(0)} = \tau_3^{(1,2)} \tau_2^{(1)} \tau_1^{(0)}$

just choose one of these combinations

In general, you just have to follow the same calculation pattern of a simple second order network.

Let's now use this method to study the frequency response of our OTA.



note that here there is already a parasitic capacitance due to C_{gd} but it is too small to have significant effects on the circuit

$$G_0 = (g_{m_1} R_1)(g_{m_5} R_2)$$

$$T(s) = G_0 \frac{\cancel{s^3 a_3} + \cancel{s^2 a_2} + s a_1 + 1}{s^3 b_3 + s^2 b_2 + s b_1 + 1}$$

- only C_3 introduces a zero (that is not at infinite frequence)
- the three capacitances are dependent (that is, one of them has its voltage drop set by the other two and does not introduce a pole)

$$\boxed{T(s) = G_0 \frac{s a_1 + 1}{s^2 b_2 + s b_1 + 1}}$$

$$\{ C_3 = C \}$$

- $b_1 = C_1 R_1^{(0)} + C_2 R_2^{(0)} + C_3 R_3^{(0)}$

  $R_1^{(0)} = R_1 \qquad R_2^{(0)} = R_2 \qquad R_3^{(0)} = R_1 + R_2 + g_{m_5} R_1 R_2$

- $b_2 = C_1 C_2 R_1^{(0)} R_2^{(1)} + C_1 C_3 R_1^{(0)} R_3^{(1)} + C_2 C_3 R_2^{(0)} R_3^{(2)}$

  $R_2^{(1)} = R_2 \qquad R_3^{(1)} = R_2 \qquad R_3^{(2)} = R_1$

- $\cancel{b}_3 = C_1 C_2 C_3 R_1^{(0)} R_2^{(1)} R_3^{(1,2)} = 0$ since $R_3^{(1,2)} = 0$

- $a_1 = C_1 R_{o_1}^{(0)} + C_2 R_{o_2}^{(0)} + C_3 R_{o_3}^{(0)}$

  $\underline{R_{o_1}^{(0)} = 0} \qquad \underline{R_{o_2}^{(0)} = 0} \qquad R_{o_3}^{(0)} = -1/g_{m_5}$ → negative coefficient → positive zero

- $\cancel{a}_2 = C_1 C_2 \cancel{R_{o_1}^{(0)}} R_{o_2}^{(1)} + C_1 C_3 \cancel{R_{o_1}^{(0)}} R_{o_3}^{(1)} + C_2 C_3 \cancel{R_{o_2}^{(0)}} R_{o_3}^{(2)} = 0$

- $\cancel{a}_3 = 0$

Solve for the roots of the polinomial at the numerator or denominator of the transfer function to find the zeroes or poles of the network with no approximation:

$$s^2 b_2 + s b_1 + 1 = 0 \quad \begin{cases} \omega_L = 2\pi f_L \\ \omega_H = 2\pi f_H \end{cases}$$

Instead of solving this equation, we can consider the following approximations if f_L and f_H are far apart from each other (at least one decade):

s very low $\longrightarrow \cancel{s^2 b_2} + s b_1 + 1 \simeq s b_1 + 1 = 0 \longrightarrow \boxed{s_L \simeq -\frac{1}{b_1}}$

s very high $\longrightarrow$ $s^2b_2+sb_1+\cancel{1} \simeq s^2b_2+sb_1=0 \longrightarrow$ $\boxed{s_H \simeq -\frac{b_1}{b_2}}$

Middlebrook approximation $\quad f_L \simeq \frac{1}{2\pi\sum_i \tau_i^{(0)}} \qquad f_H \simeq \sum_i \frac{1}{2\pi\tau_i^{(\infty)}}$

$R_i^{(\infty)}$:= resistance seen from capacitance i when all other ones are shorted

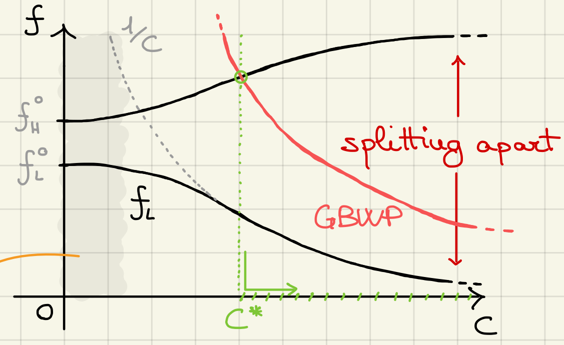
this is only valid if all the capacitances in the network are independent -

- for this example it is NOT valid since the three capacitances are dependent (must use $\omega_H=\frac{b_1}{b_2}$)

$$\Longrightarrow \quad f_L \simeq \frac{1}{2\pi\left[C_1R_1+C_2R_2+C\left(R_1+R_2+g_{ms}R_2R_1\right)\right]}$$

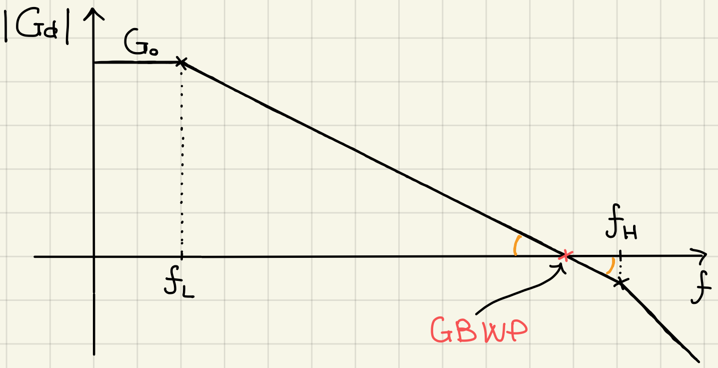
$$\Longrightarrow \quad f_H \simeq \frac{C_1R_1+C_2R_2+C(R_1+R_2+g_{ms}R_1R_2)}{2\pi\left[C_1C_2R_1R_2+C(C_1+C_2)R_1R_2\right]}$$

$$C\rightarrow\infty \quad \begin{cases} f_L=\dfrac{1}{2\pi C(R_1+R_2+g_{ms}R_1R_2)} \\[2ex] f_H=\dfrac{1+g_{ms}(R_1/\!/R_2)}{2\pi(C_1+C_2)(R_1/\!/R_2)} \end{cases}$$



✱ $f_L^0 \approx \dfrac{1}{2\pi\left[C_1R_1+C_2R_2\right]} = \dfrac{1}{2\pi\sum\tau_i^{(0)}}$

$f_H^0 \approx \dfrac{C_1R_1+C_2R_2}{2\pi\,C_1R_1C_2R_2} = \dfrac{1}{2\pi}\sum\dfrac{1}{\tau_i^{(\infty)}}$



We must have $\text{GBWP} \leqslant f_H \simeq \frac{1}{2\pi}\frac{g_{ms}}{C_1+C_2}$ to have a compensated amplifier.

But $\text{GBWP}=G_0f_L \simeq \frac{g_{m1}\cancel{R_1}\cancel{g_{ms}}\cancel{R_2}}{2\pi C\cancel{g_{ms}}\cancel{R_1}\cancel{R_2}} = \frac{g_{m1}}{2\pi C}$

So the GBWP is also dependent on C.

$$\Longrightarrow \quad \text{GBWP}=f_H \longrightarrow \frac{g_{m1}}{2\pi C^*}=\frac{g_{ms}}{2\pi(C_1+C_2)} \longrightarrow C^*=(C_1+C_2)\frac{g_{m1}}{g_{ms}}=$$

$$=\underbrace{(C_1+C_2)}_{\sim 3pF}\underbrace{\frac{2I_1}{V_{ov1}}\frac{V_{ovs}}{2I_s}}_{\sim 1}$$

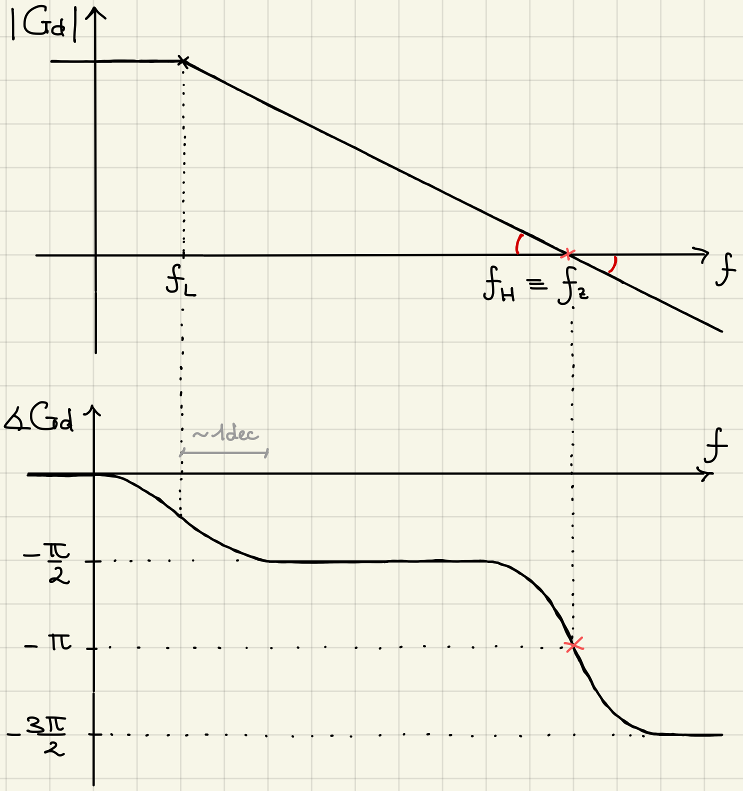
We also have a (positive) zero:

$$\Longrightarrow \quad f_z=\frac{g_{ms}}{2\pi C} \quad \text{but} \quad \text{GBWP}\simeq\frac{g_{m1}}{2\pi C} \quad \text{and} \quad g_{m1}=g_{ms}$$

so with our parameters we get GBWP $= f_z$.

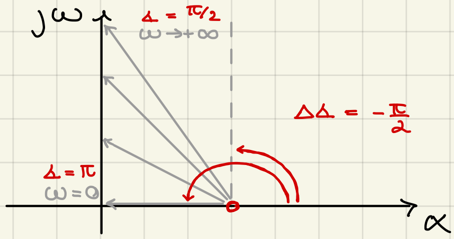
If we then set $C=C^*=(C_1+C_2)\frac{g_{m1}}{g_{ms}}$ it turns out that the zero is coincident with the high frequency pole, as well as with the GBWP frequency.

✱ watch out that for C = 0 the Middlebrook approx. doesn't hold anymore



Apparently, having pole and zero coinciding at GBWP seems to compensate the Bode plot of the absolute value by having a good closure angle.

However, since this is a positive zero it introduces a $-\pi/2$ phase shift which adds up with the pole phase shift causing the phase margin to be approximately 0°.



The signal in a negative feedback circuit with such amplifier would be fed back at the input with the same amplitude and in phase with the original signal, since the phase shift would be $-180° - \pi = -360° = 0°$ (insufficient phase margin), thus causing the output to grow with an unstable fashion.

$$f_z = \frac{1}{2\pi}\frac{\boxed{g_{m5}}}{\boxed{C}} \qquad GBWP = \frac{1}{2\pi}\frac{g_{m1}}{\boxed{C}} \qquad f_H = \frac{1}{2\pi}\frac{\boxed{g_{m5}}}{C_1+C_2}$$

To stabilize the amplifier in a negative feedback circuit we therefore need to move the POLE AND the ZERO at a frequency higher than GBWP, so we need to increase g_{m5} through using a higher current in the respective branch (higher power dissipation).

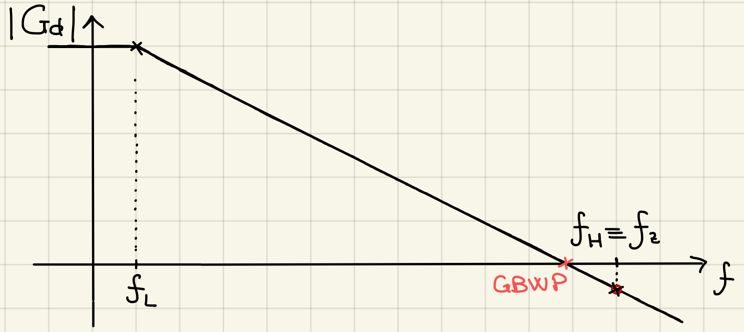
E.g.: $g_{m5} = \frac{2I_5}{V_{ov5}}$ $\qquad I_5^\circ = 150\mu A \longrightarrow I_5 = 300\mu A = 2I_5^\circ$

$$g_{m5}^\circ = 1{,}5\,\frac{mA}{V} \longrightarrow g_{m5} = 3\,\frac{mA}{V} = 2g_{m5}^\circ$$

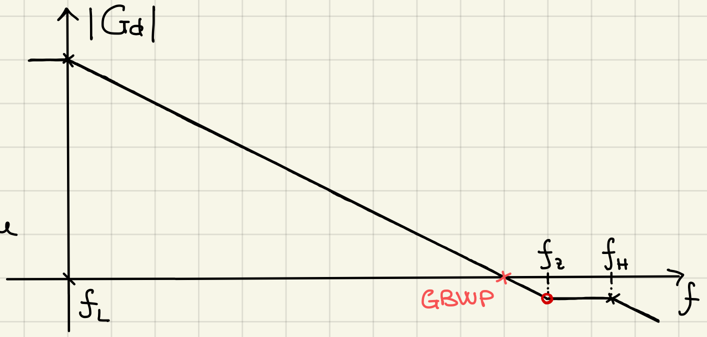
$$\frac{f_H}{GBWP} = 2\,\frac{g_{m5}^\circ}{g_{m1}}\,\frac{C^*}{C_1+C_2} = 2$$

$$\phi_m \simeq 180° - 90° - \overset{27°}{\arctan}\left(\frac{GBWP}{f_H}\right) - \overset{27°}{\arctan}\left(\frac{GBWP}{f_z}\right)$$
$$\simeq 35°$$! POOR !

This solution returns a good ϕ_m only with very high currents in M_5 and so with very high power dissipation.



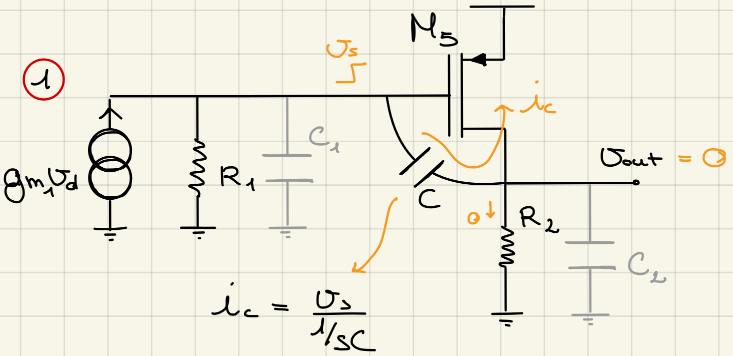
Another solution could then be to increase C above the minimum value C* to move both the GBWP AND the ZERO at lower frequencies (lower GBWP)



The phase margine should at this point be high enough (around 60°) but at the cost of more power dissipation and lower GBWP.

$$f_L \simeq \frac{1}{2\pi C g_{m5} R_1 R_2} \quad ② \qquad GBWP \simeq \frac{g_{m1}}{2\pi C} \qquad f_z \simeq \frac{g_{m5}}{2\pi C} \quad ① \qquad f_H \simeq \frac{g_{m5}}{2\pi (C_1+C_2)} \quad ③$$

Insights to better understand these numerical results - an intuitive way to calculate poles and zeroes:

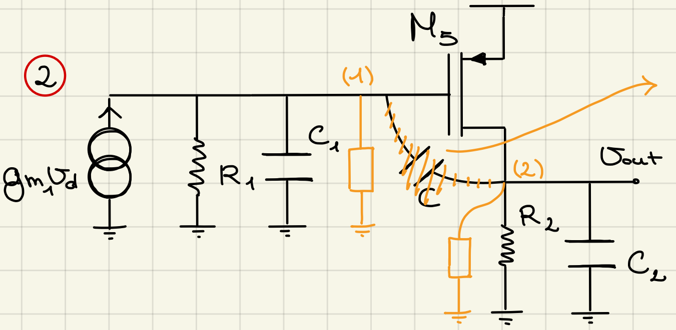


$$i_c = \frac{v_s}{1/sC}$$

$$sCv_s = g_{m5}v_s$$

$$s = \frac{g_{m5}}{C} \quad \text{(positive)}$$

$$f_z = \frac{g_{m5}}{2\pi C}$$



Is there a way to replace this capacitor with two separate equivalent impedences in order to consider C_1 and C_2 as independent capacitances?

⇓

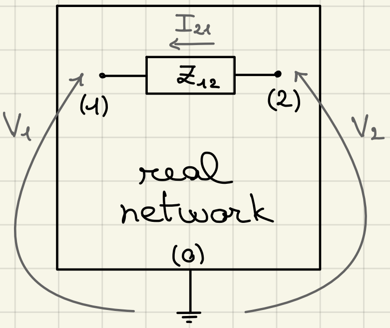
MILLER THEOREM

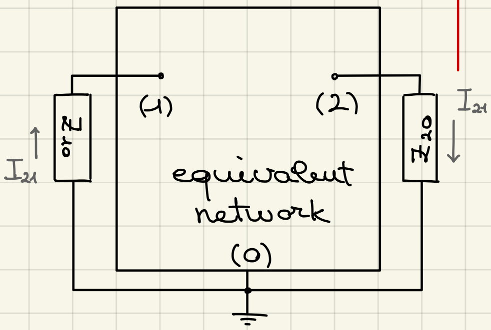
$$I_{21} = \frac{V_2 - V_1}{Z_{12}} = -\frac{V_1}{Z_{10}}$$

$$Z_{10} = Z_{12}\frac{-V_1}{V_2 - V_1} = Z_{12}\frac{1}{1 - V_2/V_1}$$

If $K(s) = \frac{V_2(s)}{V_1(s)}$ then

$$\boxed{Z_{10}(s) = Z_{12}(s)\frac{1}{1 - K(s)}}$$





$$I_{21} = \frac{V_2 - V_1}{Z_{12}} = \frac{V_2}{Z_{20}} \longrightarrow \boxed{Z_{20}(s) = \frac{K(s)}{K(s) - 1}}$$

To apply this theorem we would then need to know $\frac{V_2(s)}{V_1(s)} = K$.

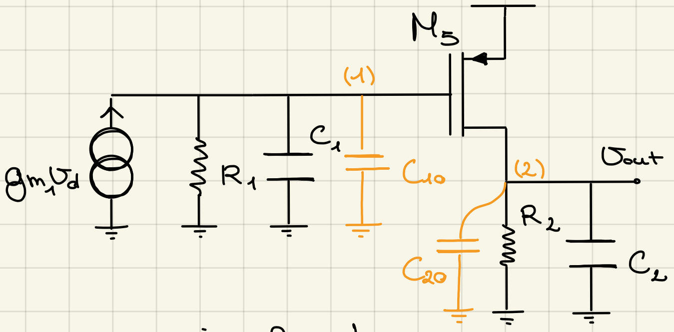
However in our problem $\frac{V_2(s)}{V_1(s)}$ is exactly the transfer function $T(s)$ between the first and second stage which is what we are trying to derive in the first place.

Nevertheless we can still apply the theorem for the low frequency pole by considering the value of $T(s) = K(s)$ approximately equal to the DC gain which is known.

$$K(s) = \frac{V_2(s)}{V_1(s)} = T(s) \simeq T(0) = -g_{m_5}R_2 = K(0)$$

$$\longrightarrow \quad Z_{10}(s) = Z_{12} \frac{1}{1-K(s)} \simeq \frac{1}{sC} \frac{1}{(1+g_{m_5}R_2)}$$

$$\longrightarrow \quad Z_{20}(s) = Z_{12}(s) \frac{K(s)}{K(s)-1} \simeq \frac{1}{sC} \frac{g_{m_5}R_2}{g_{m_5}R_2 \cancel{+1}} \simeq \frac{1}{sC}$$



equivalent network to derive the low frequency pole

$$C_{10} = C\,(1+\overbrace{g_{m_5}R_2}^{\text{2nd stage DC gain}})$$

$$C_{20} \simeq C$$

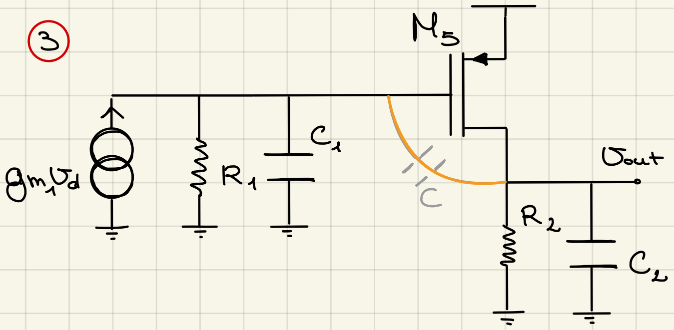
$$\boxed{f_L} = \frac{1}{2\pi \sum \tau_i^{(0)}} = \frac{1}{2\pi\left[(C_1+C_{10})R_1+(C_{20}+C_2)R_2\right]}$$

$$= \frac{1}{2\pi\left[C_1R_1 + C\underbrace{(R_1+R_2+g_{m_5}R_1R_2)}_{R_3^{(0)}} + C_2R_2\right]}$$

$$\simeq \boxed{\frac{1}{2\pi C g_{m_5}R_1R_2}}$$

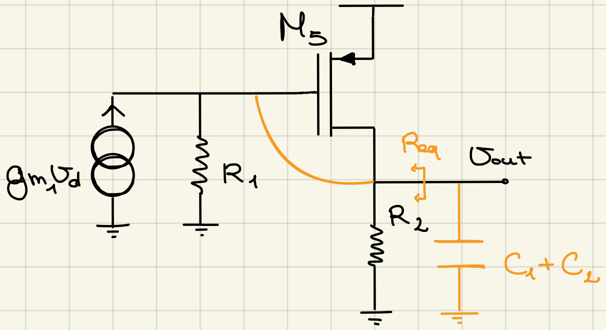
The Miller capacitance behaves like two separate capacitances: one at the output with the same size as the actual capacitance, and one at the second stage input with a size equal to the actual capacitance multiplied by the stage gain (Miller effect).

③



At high frequencies the first capacitance that will start behaving like a short circuit is C since it is related to the highest time constant (lowest frequency pole).

But then C_1 and C_2 can be considered in parallel, contributing equally to the high frequency pole.



$$R_{eq} = \frac{1}{g_{m5}} \,/\!/\, R_2 \,/\!/\, R_1 \simeq \frac{1}{g_{m5}}$$

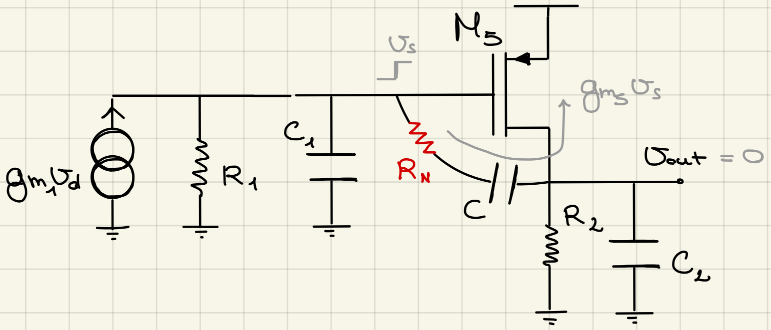
$$\boxed{f_H \simeq \frac{1}{2\pi (C_1+C_2) R_{eq}} = \frac{g_{m5}}{2\pi (C_1+C_2)}}$$

We observed that the Miller capacitance introduces a finite positive zero in the transfer function of our differential amplifier because of the feedforward current it enables between the two stages at higher frequencies.

This zero impairs the phase margin causing the amplifier to easily become unstable in a negative feedback loop. In order to compensate the phase shift introduced by the zero we needed to both increase the power consumption and decrease the bandwith.

Therefore we want to deal with this singularity without altering the performances of the amplifier

⟹ Add a nulling resistor in series with the Miller capacitance



$$g_{m5} \cancel{v_s} = \frac{\cancel{v_s}}{R_N + \frac{1}{sC}}$$

$$R_N + \frac{1}{sC} = \frac{1}{g_{m5}}$$

$$\frac{1}{sC} = \frac{1}{g_{m5}} - R_N$$

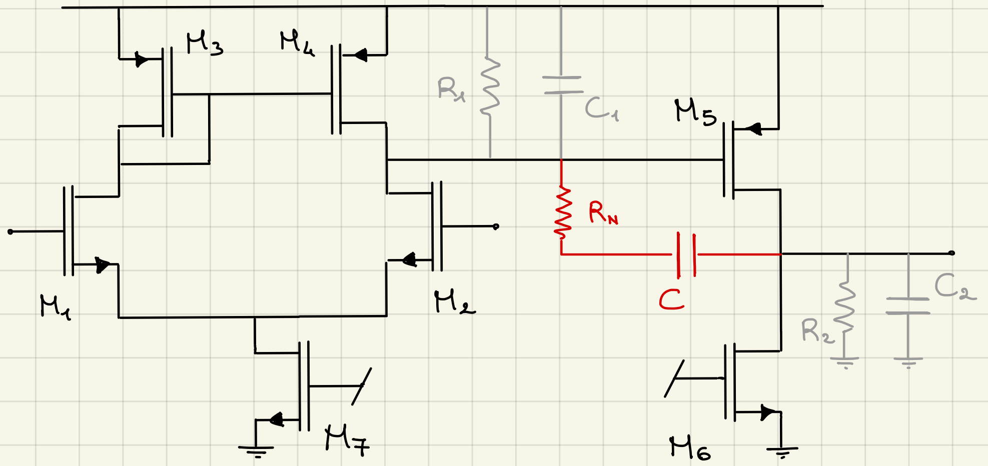
$$\Longrightarrow \boxed{s = \frac{1}{C\left(1/g_{m5} - R_N\right)}}$$

we can now move the zero however we like!

$R_N = \frac{1}{g_{m5}}$ NO zero (non-finite)

$R_N > \frac{1}{g_{m5}}$ NEGATIVE zero

Problem: adding the nulling resistor will move not only the zero but also any other pole



The three main capacitances are now independent.

↓

3 poles

(f_1, f_2, f_3)

$\{R_N$ in the order of $1/g_{m5}$ typically $2/g_{m5}\}$

The lower pole will be moved down, but just by a very negligible amount:

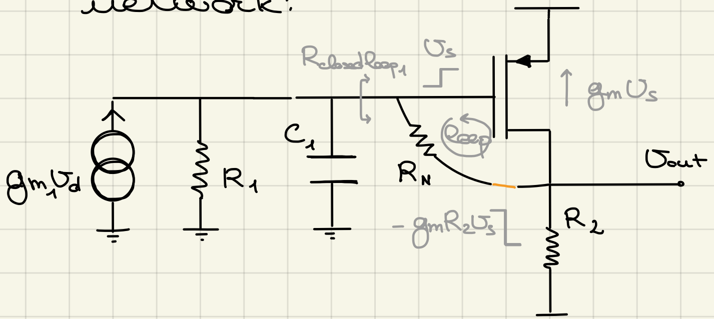
$$\Rightarrow f_1 \simeq \frac{1}{2\pi\sum\tau_i^{(0)}} = \frac{1}{2\pi[C_1R_1+C_2R_2+C(R_1+R_2+g_{m5}R_1R_2+\underbrace{R_N}_{\text{negligible}})]} \simeq \frac{1}{2\pi[Cg_{m5}R_1R_2]}$$

The higher pole will also be moved down from infinite frequency to a finite (but still high) frequency:

$$\Rightarrow f_3 \simeq \sum\frac{1}{2\pi\tau_i^{(\infty)}} = \frac{1}{2\pi}\left[\frac{1}{C_1(\underbrace{R_1/\!/R_N}_{R_N})}+\frac{1}{CR_N}+\frac{1}{C_2(\underbrace{R_2/\!/R_N}_{R_N})}\right] \simeq \frac{1}{2\pi R_N(\underbrace{C_1}_{1p}/\!/\underbrace{C}_{3p}/\!/\underbrace{C_2}_{2p})}$$

$$\simeq \frac{1}{2\pi R_N C_1}$$

To compute the middle pole, we can consider C shorted since it's related to the lower frequency pole, then calculate the low pole of the resulting second order network:



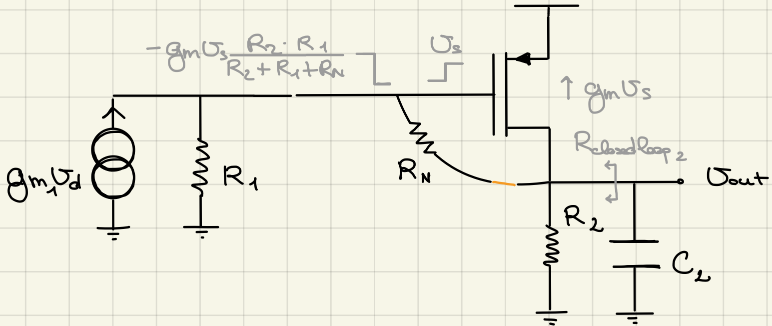
$$R_1^{(0)} = R_1 /\!/ \overbrace{\left(\frac{R_{openloop_1}}{1-G_{loop_1}}\right)}^{R_{closedloop_1}}$$

$$R_{openloop_1} = R_N + R_2$$

$$G_{loop_1} = \frac{-g_{m5}R_2v_s}{v_s} = -g_{m5}R_2$$

$$\Rightarrow R_{closedloop_1} = \frac{R_N+R_2}{1+g_{m5}R_2} \simeq \frac{1}{g_{m5}}$$

$$R_1^{(0)} \simeq R_1/\!/ {}^1\!/g_{m5} \simeq {}^1\!/g_{m5}$$



$$R_2^{(0)} = \frac{R_{openloop_2}}{1-G_{loop_2}}$$

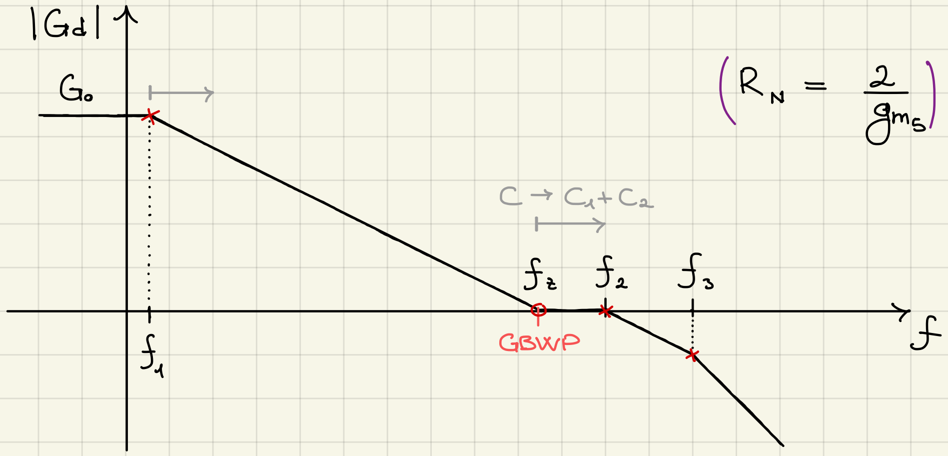
$$R_{openloop_2} = R_2/\!/(R_1+R_N) \simeq R_2/\!/R_1$$

$$G_{loop_2} = -g_{m5}\frac{R_2\cdot R_1}{R_2+R_1+R_N} \simeq -g_{m5}(R_1/\!/R_2)$$

$$R_2^{(0)} \simeq \frac{R_2/\!/R_1}{1+g_{m5}(R_2/\!/R_1)} \simeq {}^1\!/g_{m5}$$

$$\Rightarrow f_2 \simeq \frac{1}{2\pi[C_1/g_{m5}+C_2/g_{m5}]} = \frac{g_{m5}}{2\pi(C_1+C_2)}$$

The middle pole is approximately at the same frequency of the previous higher pole



$$\left(R_N = \frac{2}{g_{m5}}\right)$$

negative

$$G_0 = g_{m1} R_1 g_{m5} R_2 \qquad \boxed{f_1 \simeq \frac{1}{2\pi C g_{m5} R_1 R_2}} \qquad \boxed{GBWP \simeq \frac{g_{m1}}{2\pi C}} = \boxed{f_z = \frac{1}{2\pi C (R_N - 1/g_{m5})}}$$

$$\boxed{f_2 \simeq \frac{g_{m5}}{2\pi (C_1 + C_2)}} \qquad \boxed{f_3 \simeq \frac{1}{2\pi R_N (C_1 // C_2 // C)}}$$

lower power consumption

$$g_{m1} = g_{m5} = 150 \frac{\mu A}{V} \qquad C \geqslant C_1 + C_2 \qquad C_1 \simeq 1pF \qquad C_2 \simeq 2pF$$

higher Bandwith

If $C = C_1 + C_2$ the zero and the middle pole perfectly cancel out in both modulus and phase, thus having a fully compensated amplifier.

Note how the use of the nulling resistor allowed to avoid the use of a greater current and a bigger capacitor while still enabling the compensation of the OTA.

$$\left[FoM := \frac{GBWP \cdot C_L}{I_{tot}} \right]$$

Figure of Merit that represents how well an amplifier performs, given its load capacitance, total current consumption and Gain BandWidth Product. The higher it is, the better the amplifier.

In our example: $FoM \simeq \dfrac{g_{m1}/2\pi C \cdot C_2}{2 I_1 + I_5}$

without R_N: $g_{m1} = 150 \frac{\mu A}{V}$ $\quad I_1 = 75 \mu A \quad I_5 = 300 \mu A$

$C_2 \simeq 2pF \quad C = (C_1 + C_2) \cdot 2 \simeq 6pF$

$\longrightarrow$ $FoM = 0{,}18 \; [V^{-1}] \qquad GBWP = 4\,MHz$

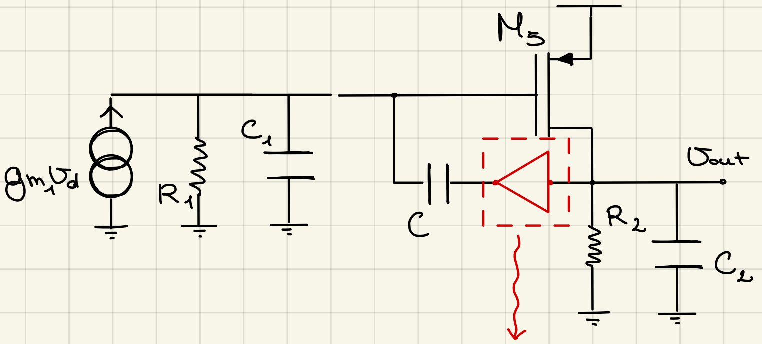
with R_N: $g_{m_1} = 150\,\frac{\mu A}{V}$ $I_1 = 75\mu A$ $I_5 = 150\mu A$

$C_2 \simeq 2pF$ $C = C_1 + C_2 \simeq 3pF$

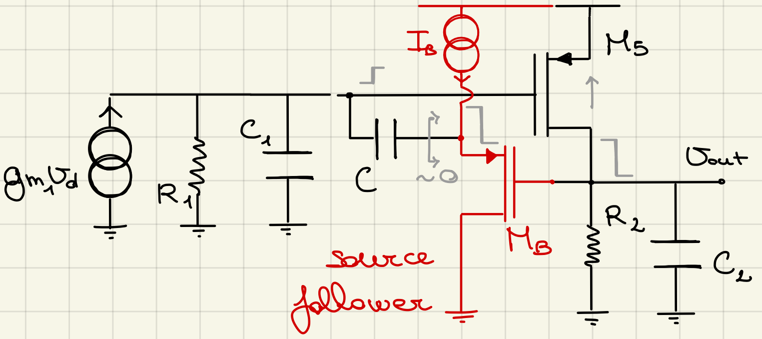
$\longrightarrow$ FoM $= 0{,}54\ [V^{-1}]$ GBWP $= 80\,MHz$

Two alternative ways to deal with the zero singularity without altering the performances of the amplifier

1. $\longrightarrow$ Add a voltage buffer in series after the Miller capacitance



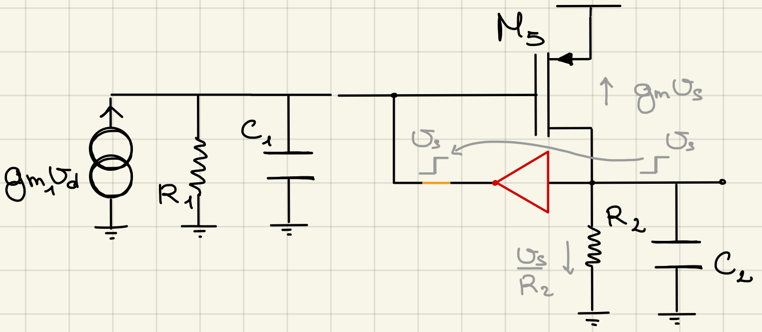
With this expedient, we maintain the Miller effect of capacitance C adding up with capacitance C_1 to attain the pole splitting, while completely avoiding the introduction of a zero (since there cannot be any feedforward current).



The three capacitors are dependent, therefore there are only two (finite) poles. There is now no (finite) zero

The lower pole is as before, since it's dominated by the Miller effect on capacitance C.
The GBWP is as before too, since the gain is also the same.

To compute the higher pole, we can consider C shorted and evaluate the lower pole of the corresponding circuit

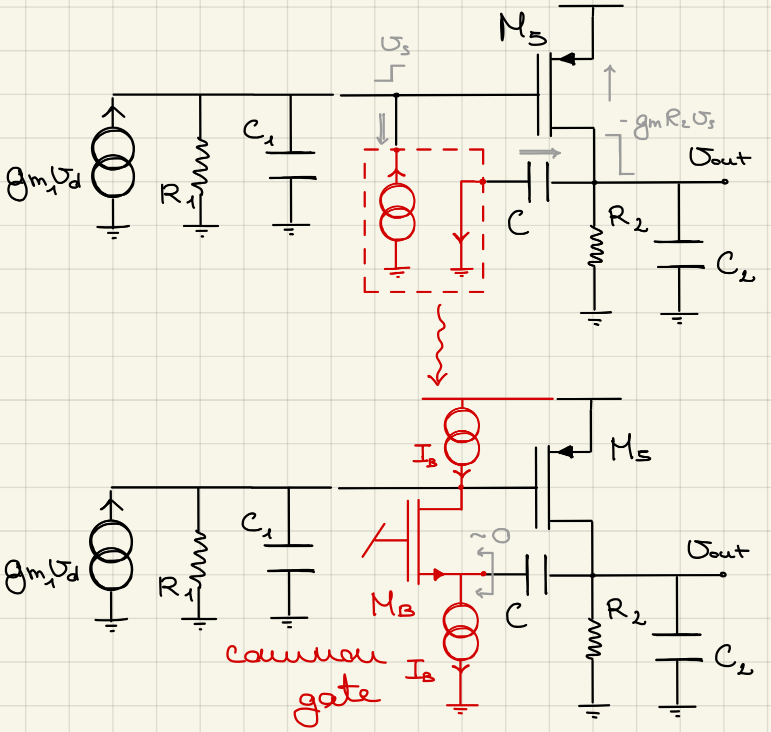


$R_1^{(0)} \simeq 0 \rightarrow$ high pole independent of inner cap.

$$R_2^{(0)} = \frac{1}{g_{m5} + \frac{1}{R_2}} \simeq \frac{1}{g_{m5}}$$

$$\Longrightarrow f_2 \simeq \frac{1}{2\pi \sum \tau_i^{(0)}} \simeq \frac{g_{m5}}{2\pi C_2}$$

2. $\longrightarrow$ Add a current buffer in series before the Miller capacitance

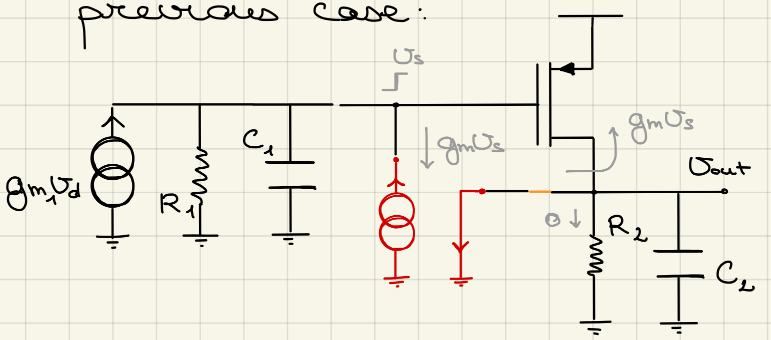


The Miller effect on C is still maintained, also there cannot be any feedforward current.

In a first order approximation, C and C_2 are in parallel, therefore there are only two independent capacitances and so only two poles.

The lower pole and the GBWP are as before.
There is now no (finite) zero.

We can compute the higher pole just like in the previous case:



$$R_1^{(0)} = R_1 // 1/g_{m5} \simeq 1/g_{m5}$$

$$R_2^{(0)} \simeq 0 \rightarrow \text{high pole independent of load cap.}$$

$$\Longrightarrow f_2 \simeq \frac{1}{2\pi \sum \tau_i^{(0)}} \simeq \frac{g_{m5}}{2\pi C_1}$$

great FoM

To summarize what we've got so far:

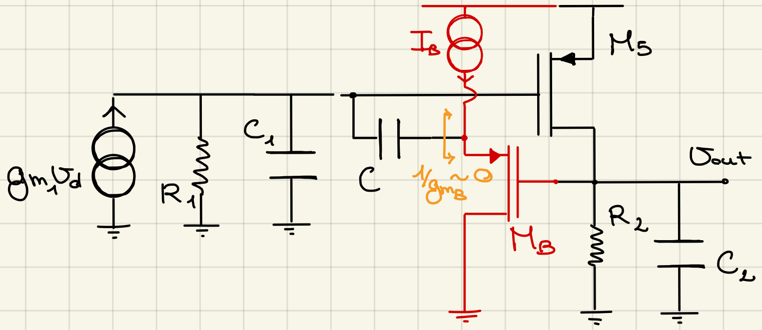
	GBWP	f_z	f_2	f_3
Miller	$\frac{g_{m1}}{2\pi C}$	RHP $\frac{g_{m5}}{2\pi C}$	$\frac{g_{m5}}{2\pi(C_1+C_2)}$	∞
Miller + Nulling resistor	$\frac{g_{m1}}{2\pi C}$	LHP $\frac{1}{2\pi C\left[R_N-\frac{1}{g_{m5}}\right]}$	$\frac{g_{m5}}{2\pi(C_1+C_2)}$	$\frac{1}{2\pi R_N(C_1 // C_2 // C)}$
Miller + voltage buffer	$\frac{g_{m1}}{2\pi C}$	LHP ∞ $\left(\frac{g_{mB}}{2\pi C}\right)$	$\frac{g_{m5}}{2\pi C_2}$	∞ $\left(\frac{g_{mB}}{2\pi C_1}\right)$
Miller + current buffer	$\frac{g_{m1}}{2\pi C}$	LHP ∞ $\left(\frac{g_{mB}}{2\pi C}\right)$	$\frac{g_{m5}}{2\pi C_1}$ $\left(\frac{1}{2\pi}\sqrt{\frac{g_{m5}g_{mB}}{C_1C_2}}\right)$ =	∞ $\left(\frac{1}{2\pi}\sqrt{\frac{g_{m5}g_{mB}}{C_1C_2}}\right)$

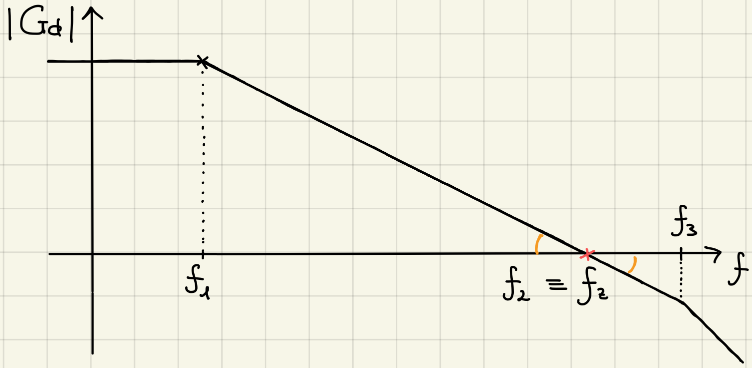
2nd order approx.

These results were obtained by considering the voltage and current buffers ideal.

By redoing the calculations, taking into account the non-null resistances of the buffers ($1/g_{mB}$), we can derive a more accurate value for the singularities of the transfer function:

1. Miller + voltage buffer





$$\Longrightarrow \quad f_z = \frac{g_{mB}}{2\pi C} \qquad f_1 \simeq \frac{1}{2\pi C g_{m5} R_1 R_2}$$

$$\Longrightarrow \left\{ \begin{array}{l} f_2 \simeq \dfrac{1}{2\pi \sum \tau_i^{(0)}} \simeq \dfrac{1}{2\pi \left[\cancel{\dfrac{C_1}{g_{m5} g_{mB} R_2}} + \dfrac{C_2}{g_{m5}} \right]} \simeq \dfrac{g_{m5}}{2\pi C_2} \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \wedge \\ f_3 \simeq \dfrac{1}{2\pi} \sum \dfrac{1}{\tau_i^{(\infty)}} = \dfrac{1}{2\pi} \left[\dfrac{g_{mB}}{C_1} + \dfrac{1}{C_2 R_2} \right] \simeq \dfrac{g_{mB}}{2\pi C_1} \end{array} \right\} \quad \underline{C \text{ shorted}}$$

GBWP = $f_2 = f_z$ to have zero and pole cancelling out and good phase margin.

$$\frac{g_{m1}}{2\pi C} = \frac{g_{m5}}{2\pi C_2} = \frac{g_{mB}}{2\pi C} \longrightarrow \left\{ \begin{array}{l} C = C_2 = 2\,pF \\ g_{m5} = g_{m1} \end{array} \right.$$

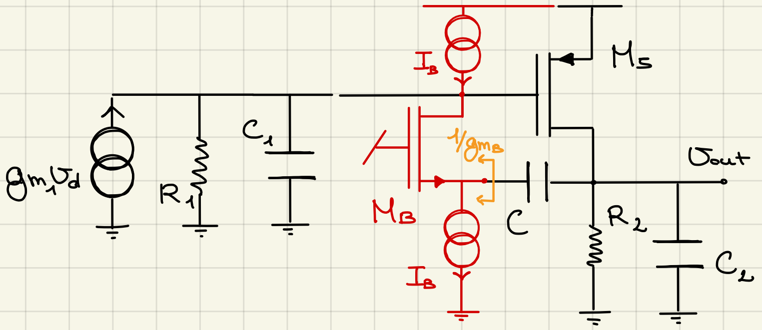
$$g_{mB} = g_{m1} = 150\,\frac{\mu A}{V}$$

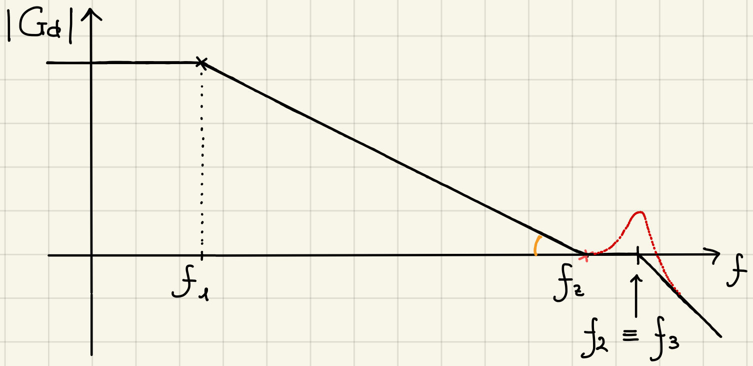
→ Similar frequency response of the nulling resistor, but FoM is impaired by the buffer current consumption.

Overall not an outstanding solution.

Note that the use of an active buffer implies that the power dissipation will be inherently higher in such configurations.

2. Miller + current buffer





$\Longrightarrow$ $f_z = \dfrac{g_{mB}}{2\pi C}$ $\qquad f_1 \simeq \dfrac{1}{2\pi C g_{m5} R_1 R_2}$

$$\Longrightarrow \begin{cases} f_2 \simeq \dfrac{1}{2\pi \sum \tau_i^{(0)}} \simeq \dfrac{1}{2\pi \left[\dfrac{C_1}{g_{m5}} + \dfrac{C_2}{g_{m5} g_{mB} R_1}\right]} \simeq \dfrac{g_{m5}}{2\pi C_1} \\ f_3 \simeq \dfrac{1}{2\pi} \sum \dfrac{1}{\tau_i^{(\infty)}} = \dfrac{1}{2\pi}\left[\dfrac{1}{C_1 R_1} + \dfrac{g_{mB}}{C_2}\right] \simeq \dfrac{g_{mB}}{2\pi C_2} \end{cases}$$

C shorted

$f_3 < f_2$!

$g_{m5} \simeq g_{mB}$
$C_2 > C_1$

the poles estimates yielded a meaningless result

↓

f_2 and f_3 must be really close to one another for the approximation not to work properly, maybe even overlapping (complex conjugate poles)

↓

must solve the second order equation: $[s^2 b_2 + s b_1 + 1 = 0]$

$$b_1 = C_1 R_1^{(0)} + C_2 R_2^{(0)} \simeq \frac{C_1}{g_{m5}} + \frac{C_2}{g_{m5} g_{mB} R_1}$$

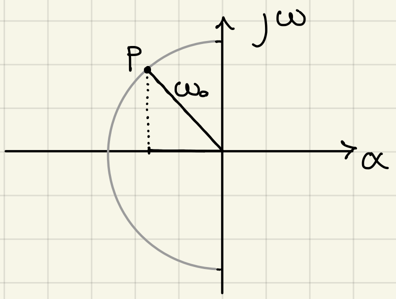
$$b_2 = C_1 C_2 R_1^{(0)} R_2^{(1)} \simeq \frac{C_1 C_2}{g_{m5} g_{mB}}$$

Instead of solving directly the equation, we can make a comparison with the general form:

$$\left[\frac{s^2}{\omega_0^2} + \frac{s}{\omega_0 Q} + 1 = 0\right]$$

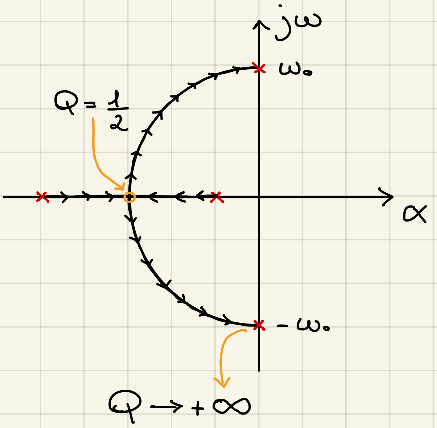
$\xi = \dfrac{Re|P|}{\underset{\omega_0 = }{}|P|}$ (form factor)

$Q = \dfrac{1}{2\xi}$ (quality factor)

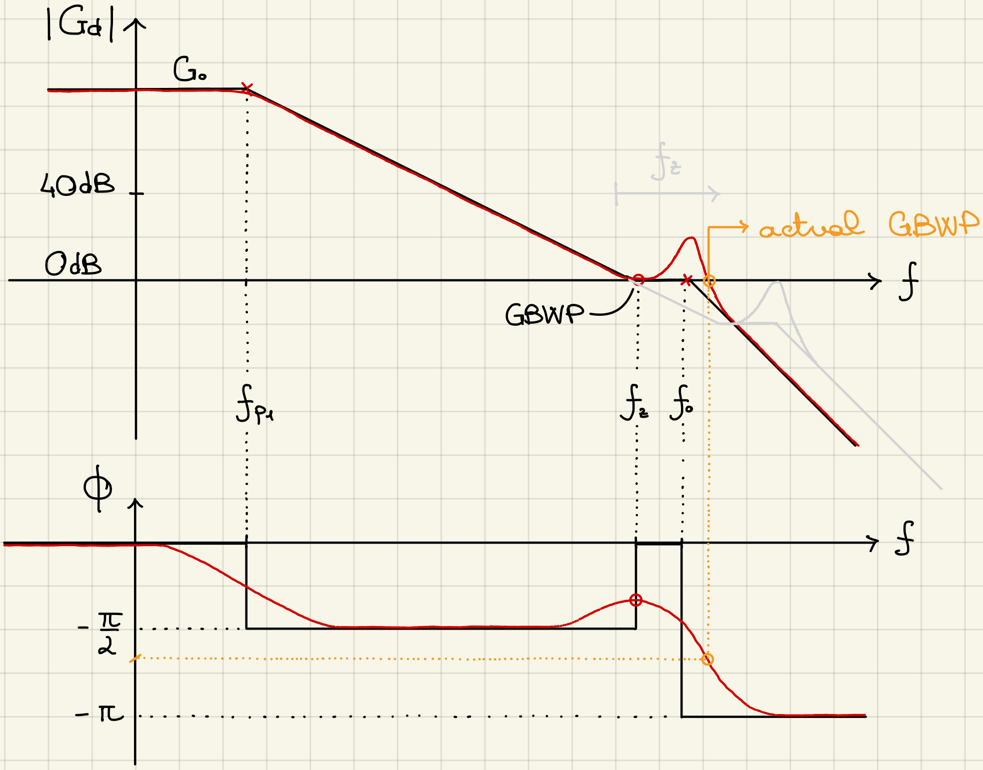


$$\Longrightarrow \omega_0 = \sqrt{\frac{g_{m5} g_{mB}}{C_1 C_2}} \qquad Q \simeq \frac{g_{m5}}{C_1 \omega_0} = \sqrt{\frac{g_{m5} C_2}{g_{mB} C_1}} \simeq \sqrt{2} > 1$$

If the Q factor is greater than 1/2, then the roots of the polinomial (the poles) are complex conjugate.



The pole pair moves from real to complex as the value of Q increases, up to the point where they become entirely imaginary.
At resonance ($\omega = \omega_0$) the Bode plot displays a growing peak that grows proportionally with Q.

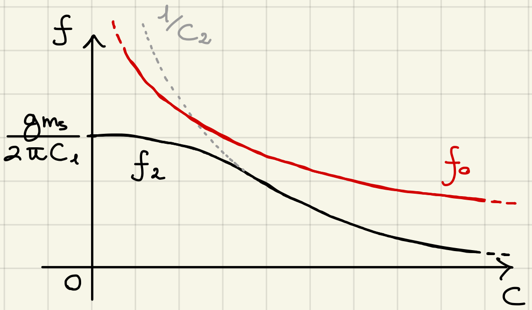


Note that:

- The actual phase margin might be quite different from the one obtained considering the GBWP as the 0dB crossing point, due to the amplitude increase in the resonance peak; a solution to avoid this problem would be to move the zero at a frequency higher than the GBWP

current buffer compensation

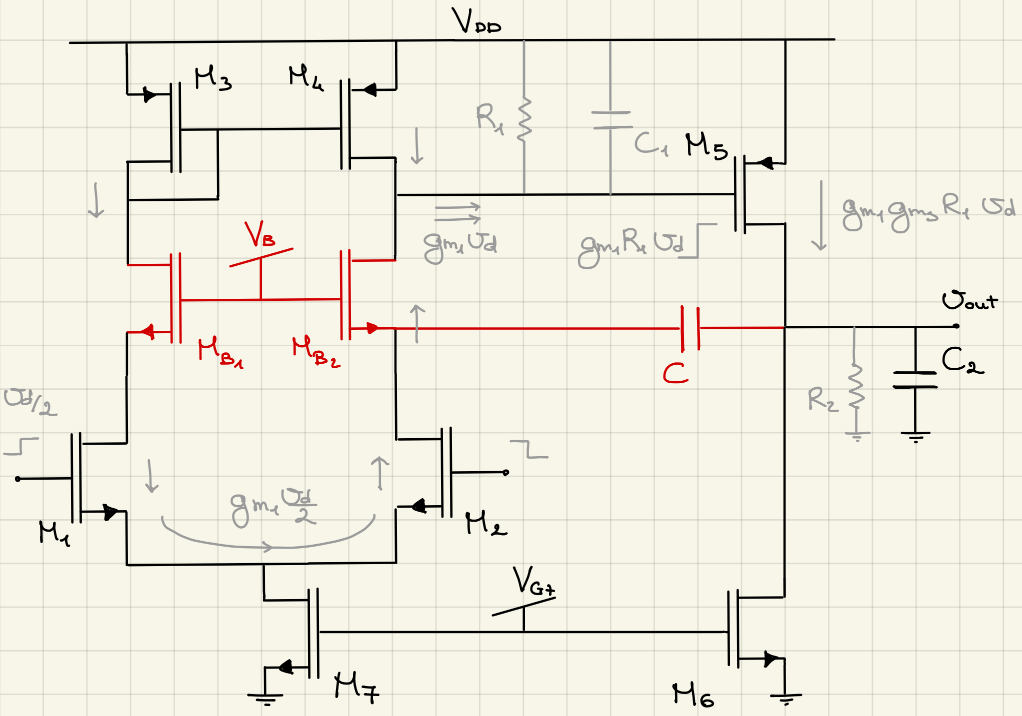
- The position of the second pole f_0 in the Ahuja configuration is less dependent on the value of C_2 (load capacitance) compared to the second pole f_z of a nulling resistor configuration.



$$\Longrightarrow \quad f_0 = \frac{1}{2\pi}\sqrt{\frac{g_{mS}\, g_{mB}}{C_1 C_2}} \propto \frac{1}{\sqrt{C_2}}$$

$$f_z = \frac{g_{mS}}{2\pi (C_1 + C_2)}$$

Is there a way to achieve the same result of the Ahuja compensation without having to supply the current buffer?



Ahuja - cascode structure

→ Use the bias current from the first stage to supply the buffer

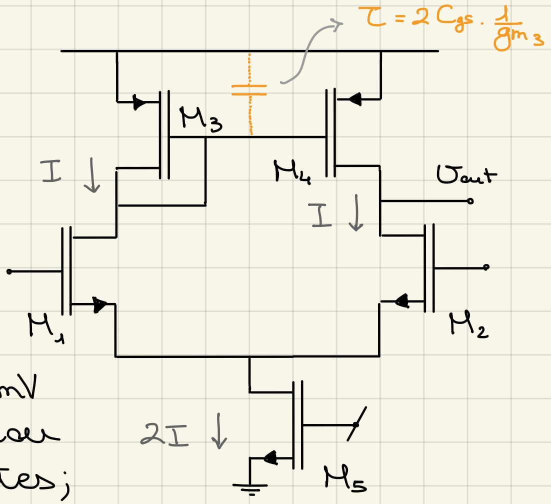
This configuration operates just like before, reducing power compensation while retaining the same DC gain and the Miller effect on capacitance C.

Single stage differential amplifiers

To obtain a good amplifier out of only one stage, using the same structure that we've used so far, as we have already seen requires both a low overdrive tension of the input transistors and, most importantly, a very long channel length:

$$G_d = g_{m_1}(r_{o_4} /\!/ r_{o_2}) = \frac{2I}{V_{ov_1}} \cdot \frac{V_{A_{2,4}}}{2I} =$$

$$= \frac{V_{A_{2,4}}}{V_{ov_1}} = \frac{V_A^0}{L_{min}} \cdot \boxed{\frac{L_{2,4}}{V_{ov_1}}}$$



However both decreasing Vov and increasing L have their limits: once the overdrive goes below ~50mV the transistor enters weak inversion and the transconductance saturates; on the other hand, if L increases then W has to increase by the same amount to maintain the form factor constant, determining a total increase of the transistor dimensions proportional to the square of the length increment. Too big dimensions will cause the oxide capacitances to become relevant and new poles will appear at lower frequencies thus impairing the frequency response of the amplifier.

E.g.: $V_A^\circ = 7V$ $\quad V_{ov_{in}} = 0,1V$ $\quad L_{min} = 0,35\mu m$ $\quad C'_{ox} = 5fF$

$g_{m_{in}} = 1,5 \frac{mA}{V}$ $\quad g_{m_M} = 0,75 \frac{mA}{V}$ $\quad (W/L)_{in} = 150$ $\quad (W/L)_M = 75$

- $L_o = 2L_{min} \longrightarrow G_{d_o} = 140 = 42,9dB$ poor, we want at least ~80dB

  $\hookrightarrow W_{M_o} = 75 L_4 = 52,5\mu m$

  $\frac{f_T}{2} = f_{M_o} = \frac{g_{mM}}{2\pi(2C_{gs_M})} = 1$ GHz pole introduced by the C_{gs} capacitance of the mirror transistor

- $L = 100 L_{min} \longrightarrow G_d = 7000 = 76,9dB$ good

  $\hookrightarrow W_M = 50 W_{M_o} = 2,625mm$ (huge!)

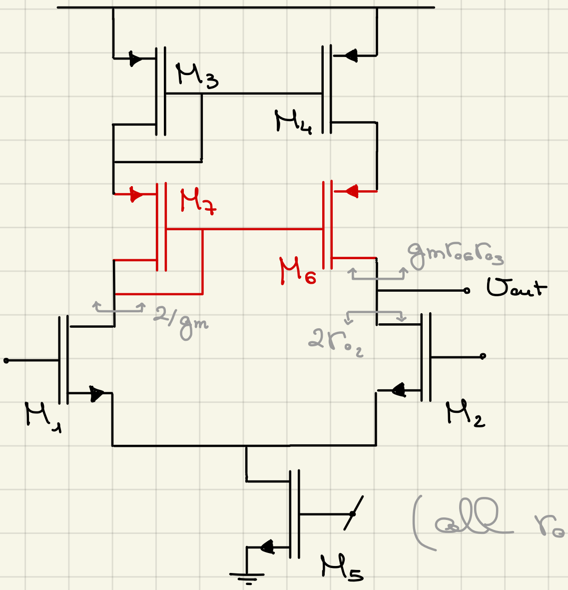
  $$f'_M = \frac{g_{mM}}{2\pi(2C_{gs_M})} = \frac{(WL)_M}{(WL)'_M} f_{M_o} = \frac{1}{50 \cdot 50} \cdot f_{M_o} = 4 \cdot 10^{-4} f_{M_o} = 400KHz$$

  ↑ too low

⟶ Trade-off between gain and bandwidth

We then need to change the amplifier structure to go beyond this limitation

⟶ Use a cascode configuration for the output resistance



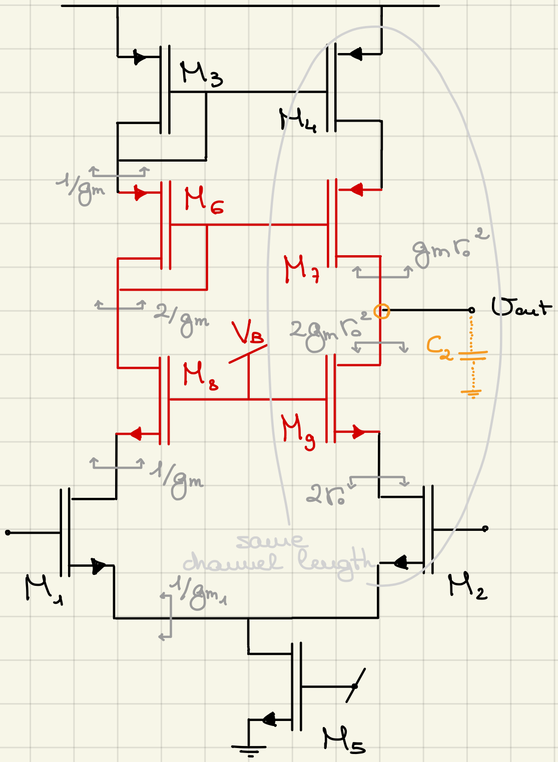
Improved output resistance with same channell length:

$$R_{out} = g_m r_o^2 // \frac{2r_o}{1-G_{loop}} = g_m r_o^2 // r_o \simeq r_o$$

improved only by a factor 2! (before it was $\frac{r_o}{2}$)

⟶ Need to increase lower branch resistance too.

(all r_o are the same)



telescopic cascode

$$R_{out} = g_m r_o^2 // \frac{2 g_m r_o^2}{1-G_{loop}} = \frac{g_m r_o^2}{2}$$

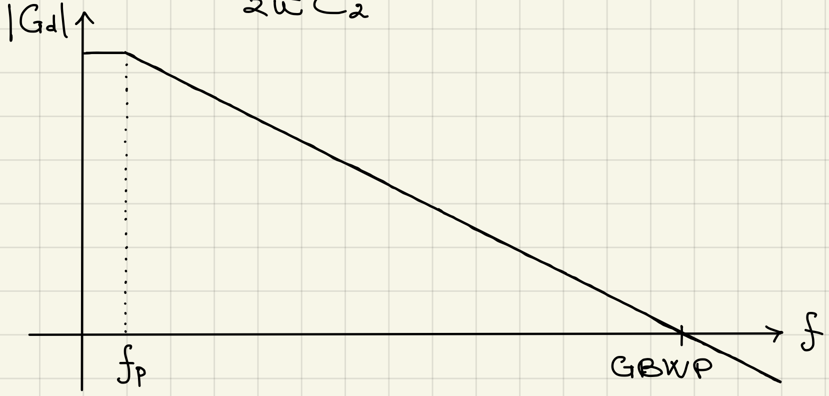
$$\Rightarrow G_d = g_{m_1} \frac{g_m r_o^2}{2} = \mu G_d^0 \simeq 90dB$$

↳ many orders of magnitude higher than the previous G_d^0! (same order of two-stage amplifier)

There is now only one high impedance node in the circuit (since it's single stage) and therefore only one low frequency pole

$$GBWP = G_0 f_p = g_{m_1} \frac{\cancel{g_m r_o^2}}{\cancel{2}} \cdot \frac{\cancel{2}}{2\pi C_2 \cancel{g_m r_o^2}} =$$

$$= \frac{g_{m_1}}{2\pi C_2}$$



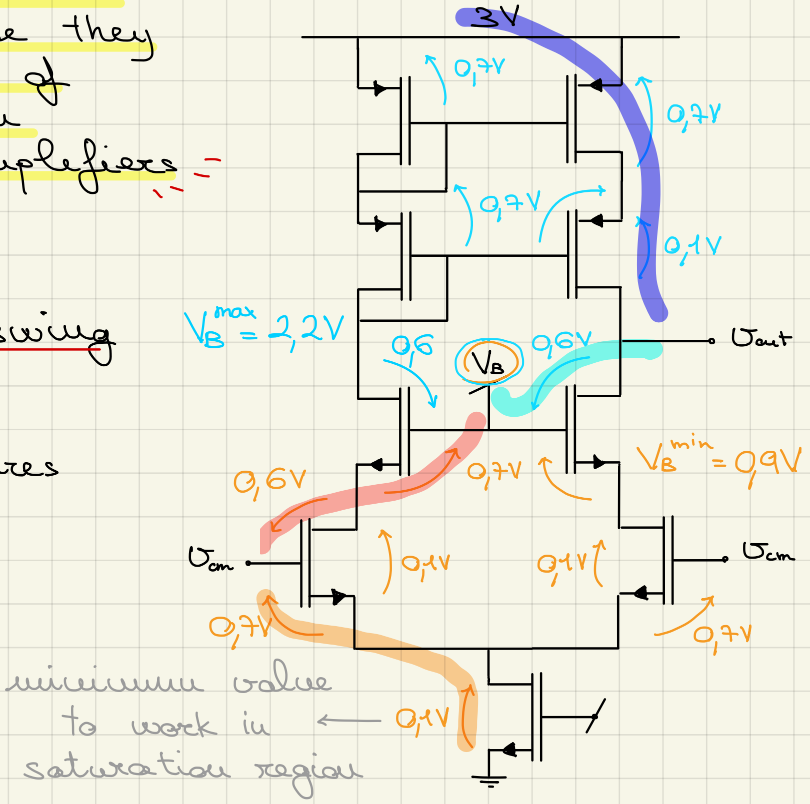
All other parasitic capacitances see very low resistances (once C_2 is shorted) and introduce poles at frequencies in the order of the f_T of the transistors.

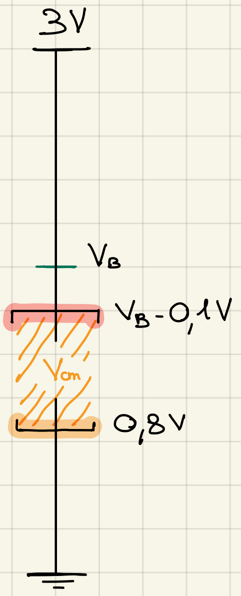
Single stage amplifiers have been introduced because they do not need any form of frequency compensation unlike multi-stage amplifiers

Issue: reduced voltage swing

Each added transistor requires a certain voltage drop in order to function properly.

$V_{ov} = 0,1V$ $V_T = 0,6V$





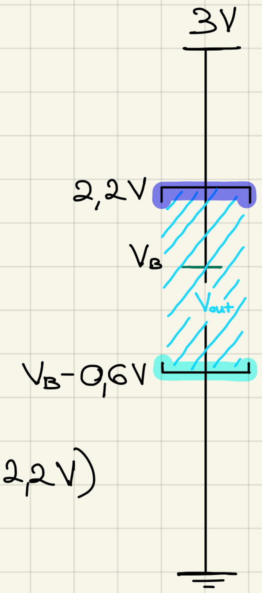
common mode input range

The input common mode voltage has a limited range of values, determined by the operating point of the tail generator and the cascode transistor.

This implies that V_B cannot be any lower than 0,9V to allow the input to have some swing.

The output also has a range, determined by the current mirror and the cascode transistor.

V_B cannot be greater than 2,8V so to have some output swing (actually, the left branch of the mirror limits V_B at no more than 2,2V)



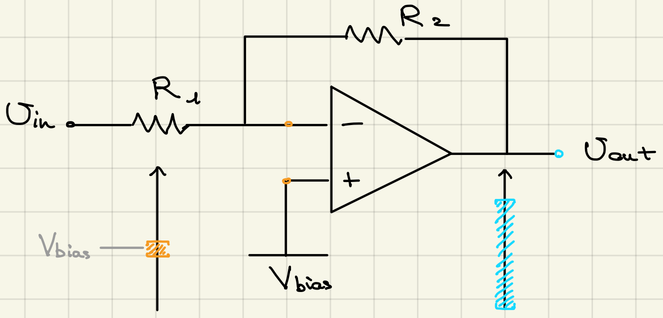
output range

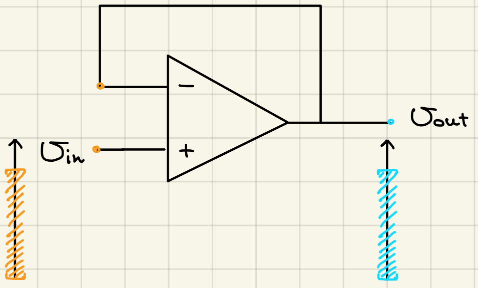
Depending on the value of V_B, the two ranges can be very diverse both in terms of mean value and swing width.

→ trade-off between input and output voltage swings

At least their values are somewhat overlapping.

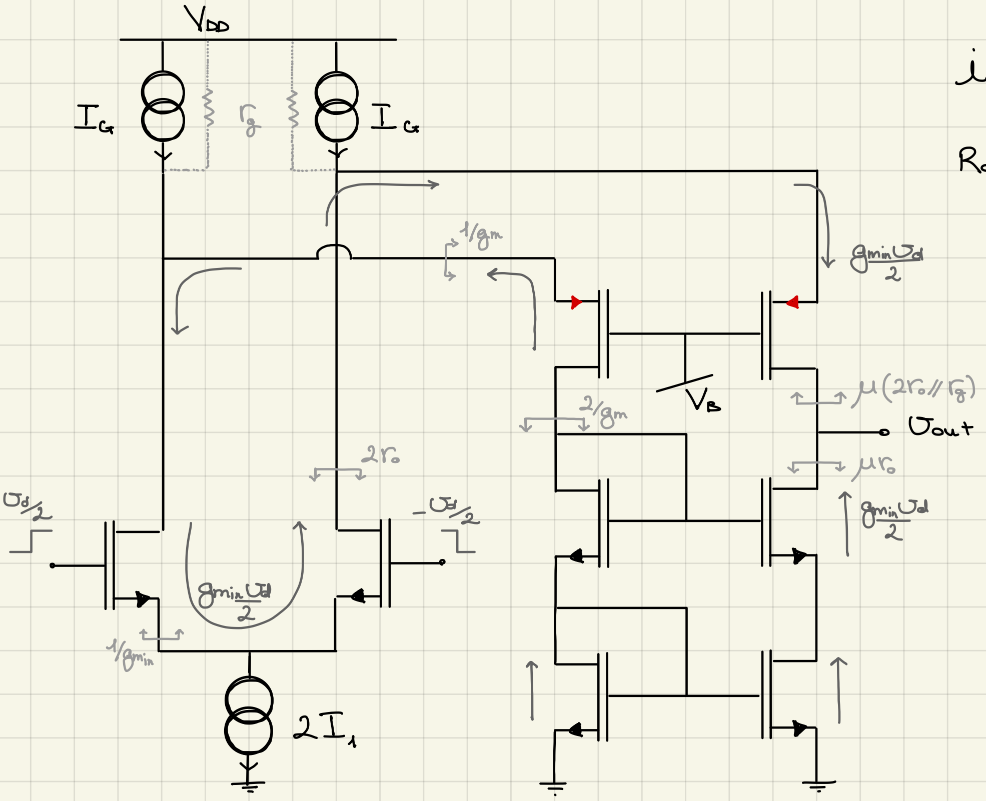
The application of the amplifier defines what the input and output voltage range should look like:





Solution: use p-type transistors "in parallel" to the main branch

("flip" the components above the input transistors so that they share the same voltage drop)



$$i_{cc} = g_{m\,in} v_d$$

$$R_{out} \simeq g_m r_o^2 // \frac{g_m r_o (2r_o // r_g)}{1 + \underbrace{\frac{r_g}{2r_o + r_g}}_{G_{loop}}}$$

$$\simeq \frac{g_m r_o^2}{3}$$

$r_g \simeq r_o$

$$\Longrightarrow G_d \simeq g_{m\,in} \frac{g_m r_o^2}{3}$$

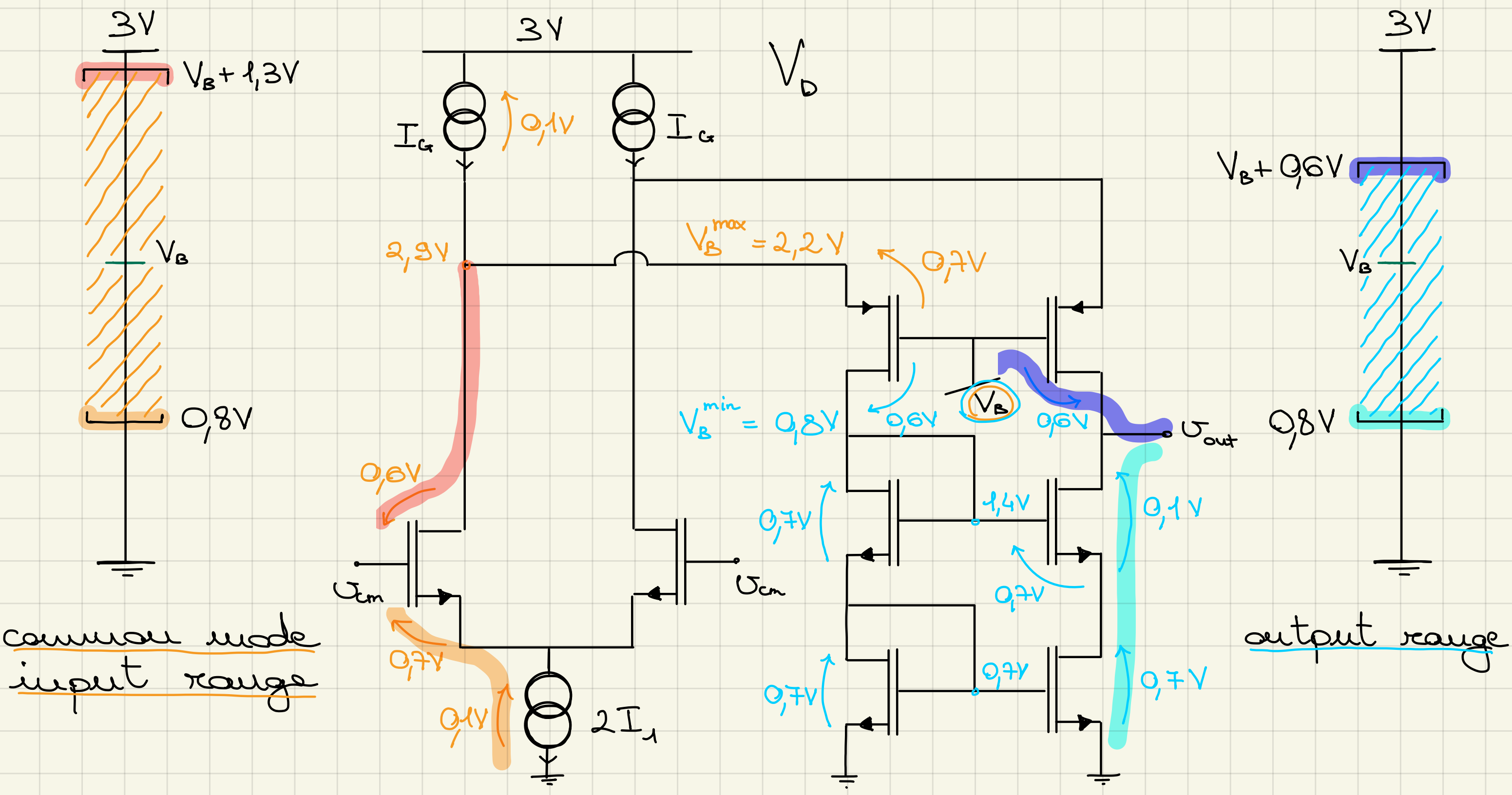
small reduction of the gain due to the non-ideal current generators

folded cascode

$$f_p = \frac{1}{2\pi C_L R_{out}}$$

$$GBWP = \frac{g_{m\,in}}{2\pi C_L}$$

frequency response is the same



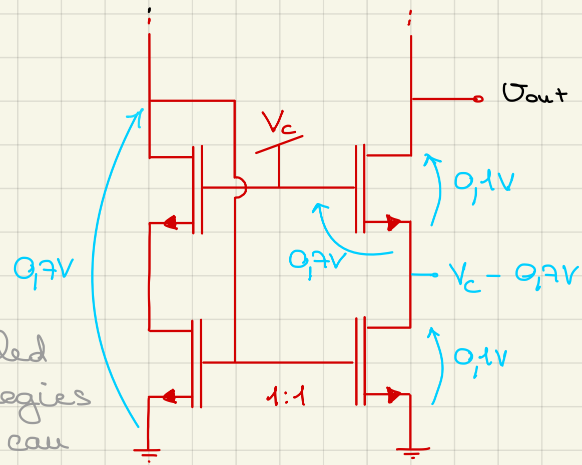
Input and output have almost the exact same voltage range, furthermore they both increase as V_B increases!

⟹ No more trade-off between input and output swings

In scaled technologies, the power supply is however much lower than 3V (typically around 1V)

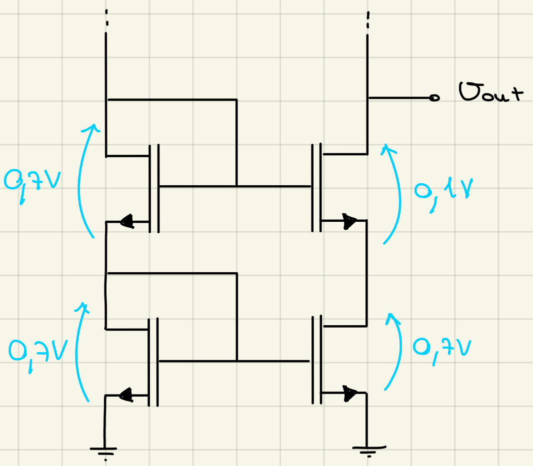
↳ Improve the output voltage swing even further, by lowering the minimum value (0,8V is too big compared to 1V)

↳ enhanced mirror vs. standard mirror



in scaled technologies the V_T can be reduced down to ~0,4V

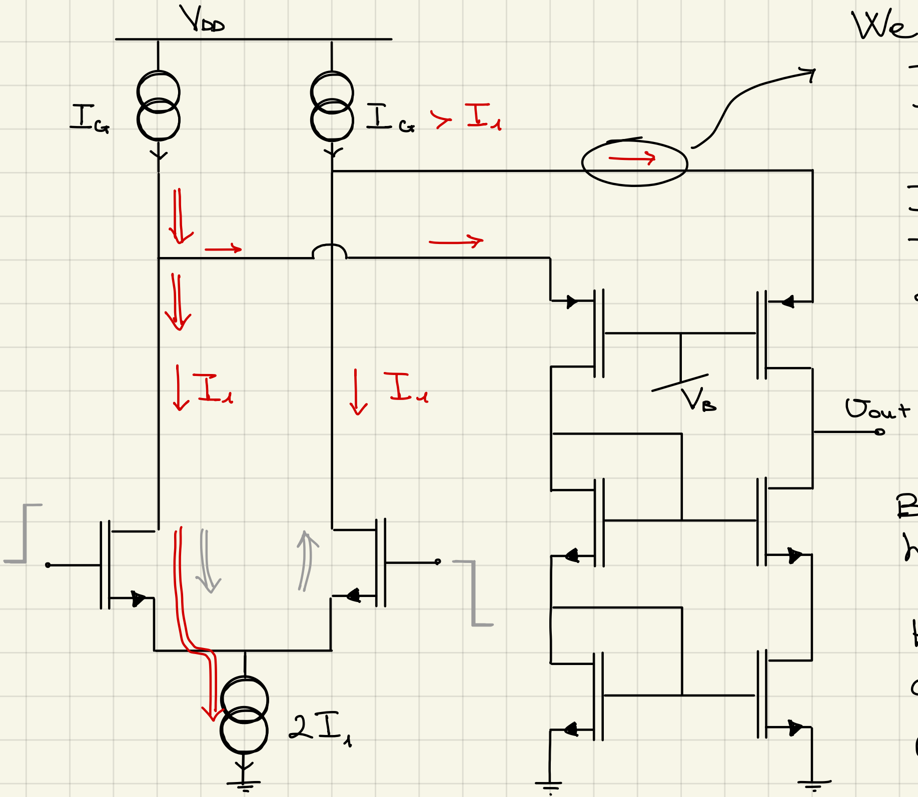
$v_{out}^{min} \leq 0,2V$



$v_{out}^{min} = 0,8V$

The cost to achieve this improved voltage dynamic is the use of an additional power supply V_c.

Issue: higher power consumption



We need some current in the mirror branches to set their bias

In a telescopic cascode the total current is $2I_1$, while in the folded cascode the total current is $2I_G > 2I_1$.

By how much does I_G have to be bigger than I_1?

↕

How much more power dissipation does the folded cascode entail with respect to the telescopic cascode?

In order to maintain the bias in both the mirror branches, the head generator (I_G) always has to provide more current than what could possibly be needed by the input transistor.
At the maximum differential input signal, all current from the tail generator ($2I_1$) will flow through just one input branch. The head generator of that branch will then have to supply more than $2I_1$ to allow some current to flow in the mirror (current cannot be drained from the mirror).

Therefore it must always be granted that $I_G > 2I_1$

↕

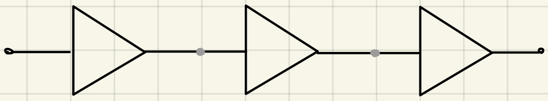
The folded cascode dissipates at least 2 times more than the telescopic cascode configuration

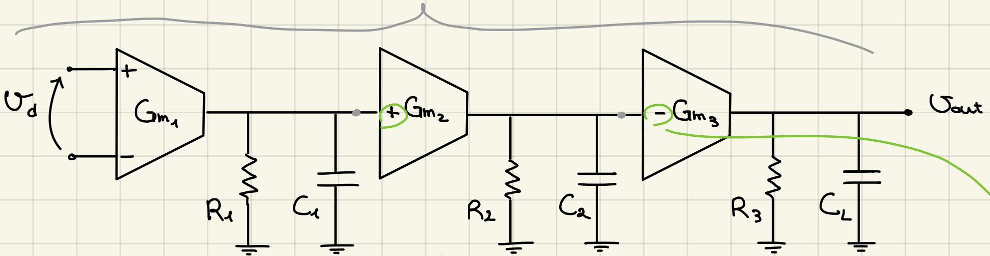
Multi-stage differential amplifiers

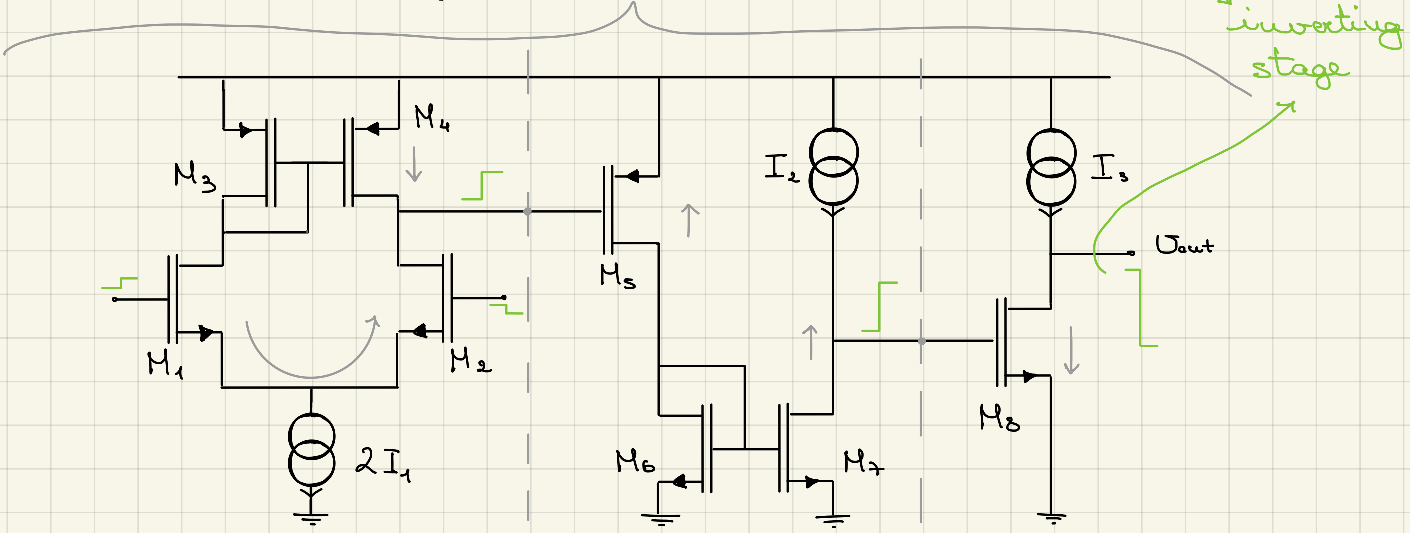
We want to achieve a even higher differential gain in the order of $> 100\,dB$

↳ must use more stages in cascade

in low bias implementations, cascode can only achieve so much







high G_{m1} to reduce noise

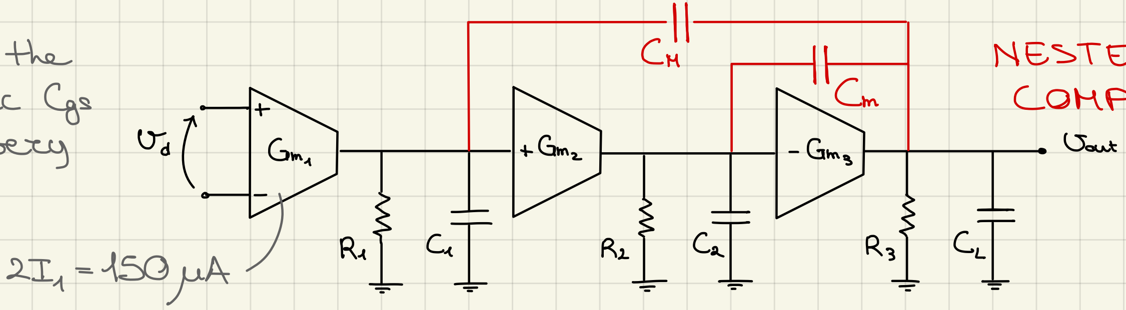
$G_{m1} = g_{m1} = 1{,}5 \frac{mA}{V}$ $\quad$ $G_{m2} = g_{m5} = 0{,}5 \frac{mA}{V}$ $\quad$ $G_{m3} = g_{m8} = 1 \frac{mA}{V}$

$R_1 = r_{o4} // r_{o2} = 47k\Omega$ $\quad$ $R_2 = r_{o7} // r_{g2} = 140k\Omega$ $\quad$ $R_3 = r_{o8} // r_{g3} = 70k\Omega$

$C_1 = 0{,}1\ pF$ $\quad$ $C_2 = 0{,}1 pF$ $\quad$ $C_L = 5pF$

they're the parasitic C_{gs} hence very small

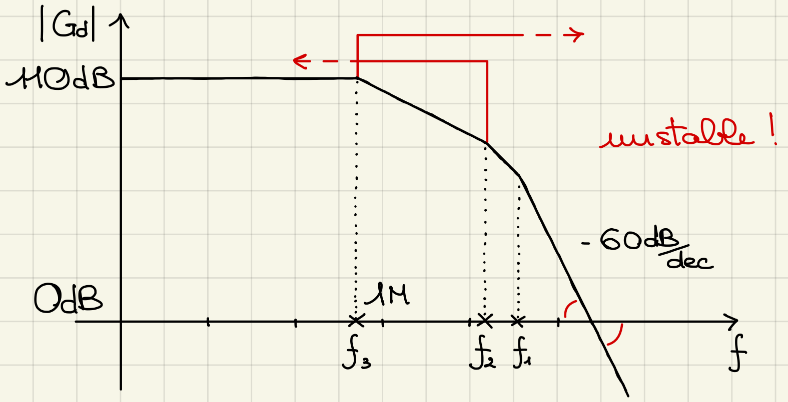
NESTED MILLER COMPENSATION



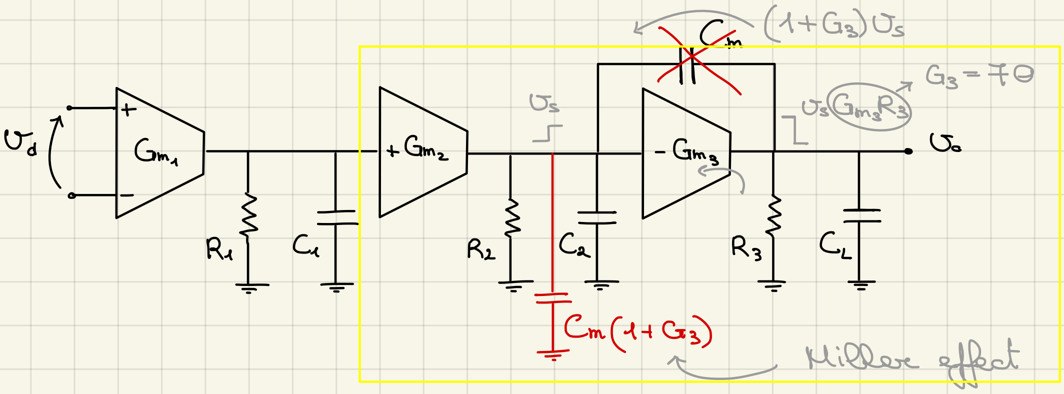
$$Gd_o = \underset{70}{g_{m1} R_1} \cdot \underset{70}{g_{m5} \cdot R_2} \cdot \underset{70}{g_{m8} R_3} = 3{,}4 \cdot 10^5 \simeq 110\,dB \checkmark$$

assuming $\mu = 140$

$$\boxed{f_1 = \frac{1}{2\pi C_1 R_1} = 34\,MHz \qquad f_2 = \frac{1}{2\pi C_2 R_2} = 11\,MHz \qquad f_3 = \frac{1}{2\pi C_L R_3} = 455\,kHz}$$



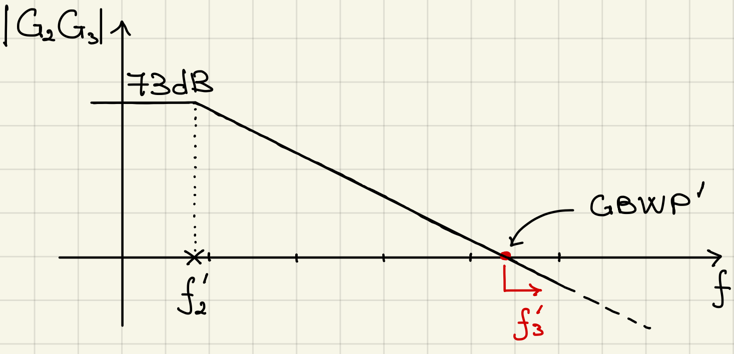
→ Miller compensation (m) to increase the time constant related to the C_2 node (and decrease the one related to C_3)



$$\Longrightarrow f_2' \simeq \frac{1}{2\pi R_2 C_m G_3}$$

$$f_3' > GBWP' \simeq \frac{G_{m2}}{2\pi C_m}$$

$f_3' = ?$ $\quad$ $C_m = ?$



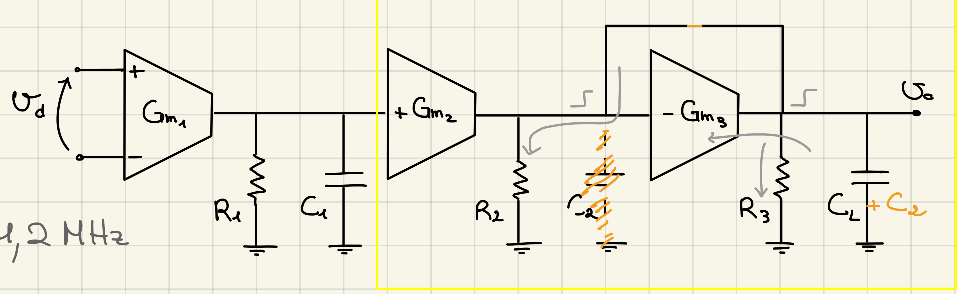
$$\downarrow$$

$$f'_3 = \frac{1}{2\pi C R_{eq}}\Bigg|_{C_m \text{ short}}$$

only one pole

$C = C_2 + C_L$

$R_{eq} = R_3 // R_2 // \frac{1}{G_{m_3}} \simeq \frac{1}{G_{m_3}}$



$$\Longrightarrow f'_3 = \frac{G_{m_3}}{2\pi(C_2+C_L)} = 31{,}2 \text{ MHz}$$

The Miller capacitor introduces a (RHP) zero as well:

$$\Longrightarrow f'_z = \frac{G_{m_3}}{2\pi C_m} \rightarrow 2\ GBWP'$$

27°

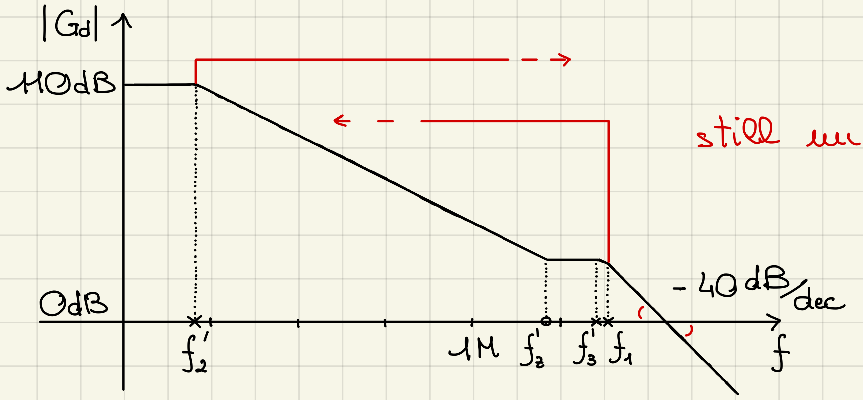
$$\phi'_m = 180° - 90° - \operatorname{arctg}\left(\frac{GBWP'}{f'_z}\right) - \operatorname{arctg}\left(\frac{GBWP}{f'_3}\right) \simeq 60°$$

↳ $f'_3 \simeq 10 GBWP'$

$$\Longrightarrow C_m = 10 \cdot \frac{G_{m_2}}{G_{m_3}}(C_2+C_L) = 25\text{pF}$$

$$f_1 = \frac{1}{2\pi C_1 R_1} = 34\text{MHz} \qquad f'_2 = \frac{1}{2\pi C_m G_{m_3} R_2 R_3} = 650\text{Hz} \qquad f'_3 = \frac{G_{m_3}}{2\pi(C_L+C_2)} = 31{,}2 \text{ MHz}$$

$$\left(f'_z = \frac{G_{m_3}}{2\pi C_m} = 6{,}24\text{MHz}\right)$$

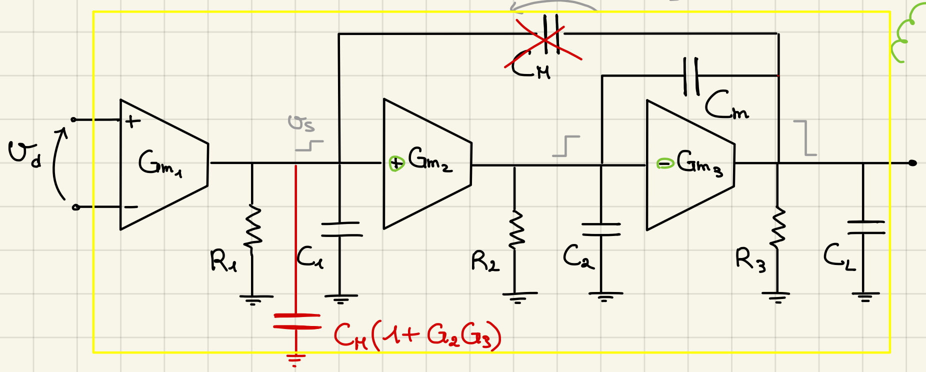


still unstable!

→ Miller compensation (M) to increase the time constant related to the C_1 node (and decrease the one related to C_2)

$C_m \updownarrow$

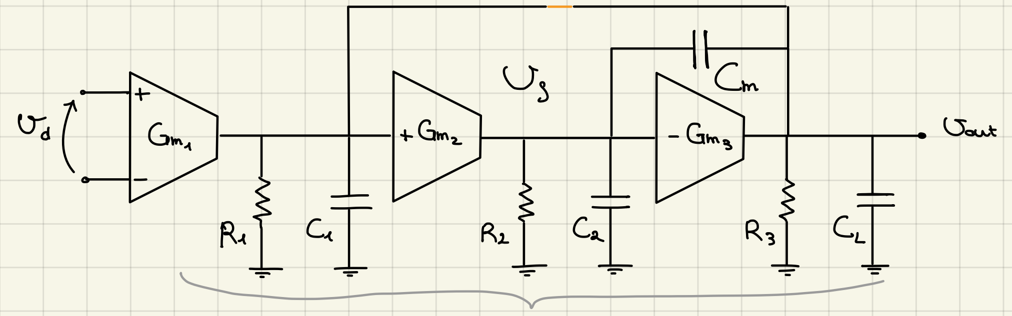
$(1+G_2G_3)v_s$



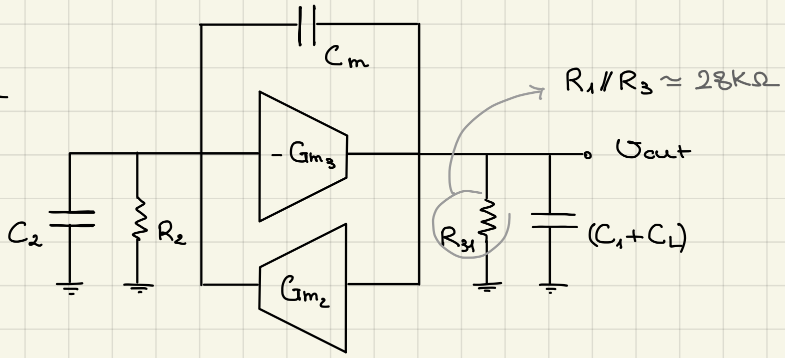
only G_3 is inverting so to retain the Miller effect

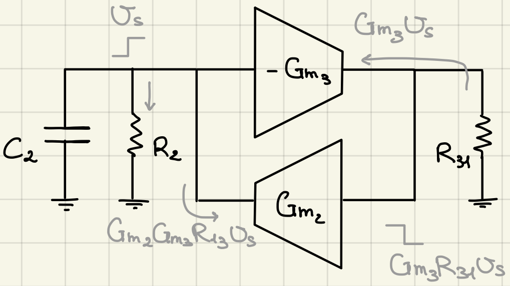
$$\Longrightarrow f''_1 \simeq \frac{1}{2\pi C_M R_1 G_2 G_3} \qquad G_d = G_1G_2G_3 \qquad GBWP'' \simeq \frac{G_{m_1}}{2\pi C_M}$$

$f_2'' > GBWP''$ $\qquad f_2'' \simeq f_L\Big|_{C_H \text{ shorted}} \simeq \frac{1}{2\pi \sum \tau_i^{(0)}}\Big|_{C_H \text{ shorted}}$



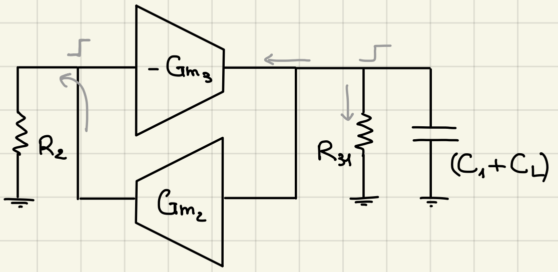
two independent capacitances
↓
two poles

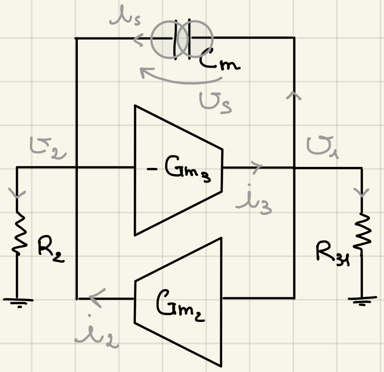




$$R_2^{(0)} = R_2 /\!/ \frac{1}{G_{m_2}G_{m_3}R_{13}} \simeq \frac{1}{G_{m_2}G_{m_3}R_{13}} = 70\Omega$$

$$R_{13}^{(0)} = R_{31} /\!/ \frac{1}{G_{m_2}G_{m_3}R_2} \simeq \frac{1}{G_{m_2}G_{m_3}R_2} = 14\Omega$$





$$R_m^{(0)} = \frac{v_s}{i_s} = \frac{v_2 - v_1}{i_s}$$

$$\begin{cases} \overset{\nearrow 0}{\frac{v_2}{R_2}} = i_s + i_2 = i_s + G_{m_2}v_1 \\ \overset{\nearrow 0}{\frac{v_1}{R_{31}}} + i_s = i_3 = -G_{m_3}v_2 \end{cases} \overset{*}{\approx} \begin{cases} v_1 \simeq -\frac{i_s}{G_{m_2}} \\ v_2 \simeq -\frac{i_s}{G_{m_3}} \end{cases}$$

$$\longrightarrow R_m^{(0)} \simeq -\frac{1}{G_{m_3}} + \frac{1}{G_{m_2}} = \frac{G_{m_3} - G_{m_2}}{G_{m_3}G_{m_2}} = 1K\Omega$$

$$v_1 = -\frac{1}{G_{m_2}}\left(\frac{1}{R_2 G_{m_3}}\left(\frac{v_1}{R_{31}} + i_s\right) + i_s\right)$$

$$v_1 + \frac{v_1}{G_{m_2}G_{m_3}R_2R_{31}} = -\left(\frac{i_s}{G_{m_2}} + \frac{i_s}{G_{m_2}G_{m_3}R_2}\right)$$

$$v_1\left(\frac{G_{m_2}G_{m_3}R_2R_{31} + 1}{\cancel{G_{m_2}}\cancel{G_{m_3}}\cancel{R_2}R_{31}}\right) = -i_s\left(\frac{G_{m_3}R_2 + 1}{\cancel{G_{m_2}}\cancel{G_{m_3}}\cancel{R_2}}\right)$$

$$v_1 = -i_s R_{31}\left(\frac{G_{m_3}R_2 + 1}{G_{m_2}G_{m_3}R_2R_{31} + 1}\right)$$

$$v_2 = -\frac{1}{G_{m_3}}\left(-i_s\left(\frac{G_{m_3}R_2 + 1}{G_{m_2}G_{m_3}R_2R_{31} + 1}\right) + i_s\right)$$

$$= \frac{i_s}{G_{m_3}}\left(\frac{G_{m_3}R_2 + 1}{G_{m_2}G_{m_3}R_2R_{31} + 1} - 1\right)$$

<u>Exact calculation</u> *

$$\Longrightarrow f_2'' \simeq \frac{1}{2\pi\left[\underbrace{C_2R_2^{(0)}}_{7ps} + \underbrace{C_mR_m^{(0)}}_{25ns} + \underbrace{(C_1+C_L)R_{13}^{(0)}}_{71ps}\right]} \simeq \frac{G_{m_3}G_{m_2}}{2\pi\, C_m(G_{m_3} - G_{m_2})} = 6{,}4MHz$$

There is a new zero as well:

$$\Longrightarrow \quad f_2'' \simeq \frac{\boxed{G_{m2}R_2G_{m3}}}{2\pi\, C_H} \gg \text{GBWP!} \quad \longrightarrow$$

it doesn't affect the frequency response too much (it's a LHP zero and will at most improve the ϕ_m by a small amount)

transconductance covered by the bridging capacitor C_H

Note: the value of the poles and zeroes is, at this point of network complexity, just an approximation; these values should give an idea of the sizing of compensating capacitors and the behaviour of the circuit, which should then be tested through simulations.

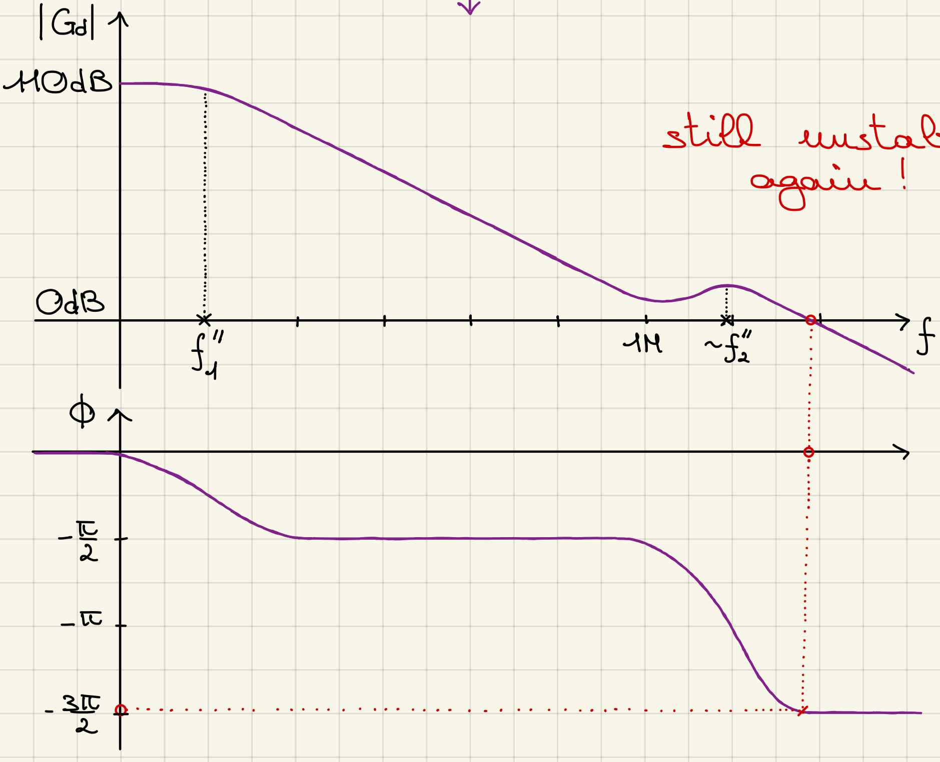
$$\frac{G_{m2}G_{m3}}{2\pi\, C_m(G_{m3}-G_{m2})} = f_2'' \gg \text{GBWP} = \frac{G_{m1}}{2\pi\, C_H}$$

$$\Longrightarrow \quad C_H > C_m \frac{G_{m1}(G_{m3}-G_{m2})}{G_{m2}G_{m3}} = 37{,}5\,pF \longrightarrow C_H = 75\,pF \qquad \text{GBWP} = 3{,}2\,MHz$$

$$f_1'' = \frac{1}{2\pi\, C_H R_1 G_{m2} G_{m3}} = 9{,}2\,Hz \qquad f_2'' = \frac{G_{m3}G_{m2}}{2\pi\, C_m(G_{m3}-G_{m2})} = 6{,}4\,MHz \qquad f_3'' = ?$$

$$\left(f_z'' = ? \right)$$

$\downarrow$ simulation



still unstable again!

Our approximations were not accurate enough.

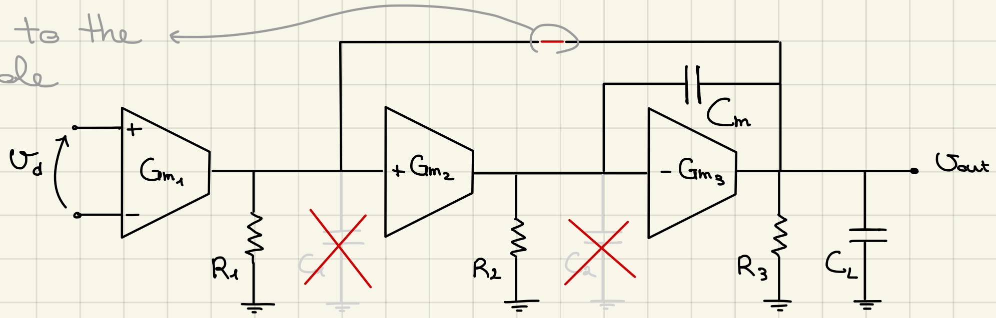
The low frequency pole was computed correctly.

The high frequency poles and zeroes have to be estimated more carefully

→ We can hold on to the result related to the low pole and study the circuit at higher frequencies, while also neglecting the parasitic capacitances that are negligible with respect to the corresponding Miller capacitance and load capacitance.

↳ the resulting network is a second order network

associated to the low pole



$$b_2 s^2 + b_1 s + 1 = 0$$

$\frac{1}{\omega_0 Q}$

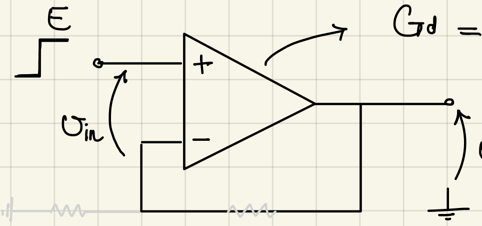
$$s^2 \frac{C_L C_m}{G_{m2} G_{m3}} + s \frac{G_{m3} - G_{m2}}{G_{m3} G_{m2}} C_m + 1 = 0$$

$\frac{1}{\omega_0^2}$

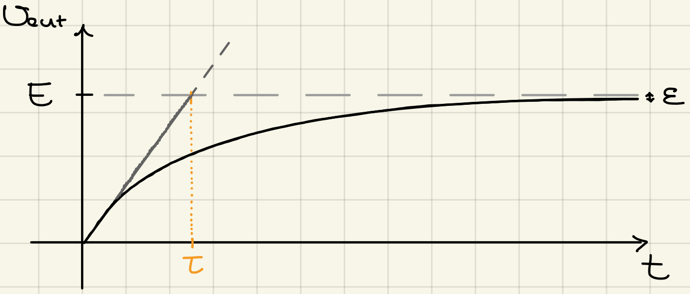
$$\left[\omega_0 = \sqrt{\frac{G_{m2} G_{m3}}{C_m C_L}} \qquad Q = \frac{1}{(G_{m3} - G_{m2})} \sqrt{G_{m3} G_{m2} \frac{C_L}{C_m}} \right]$$

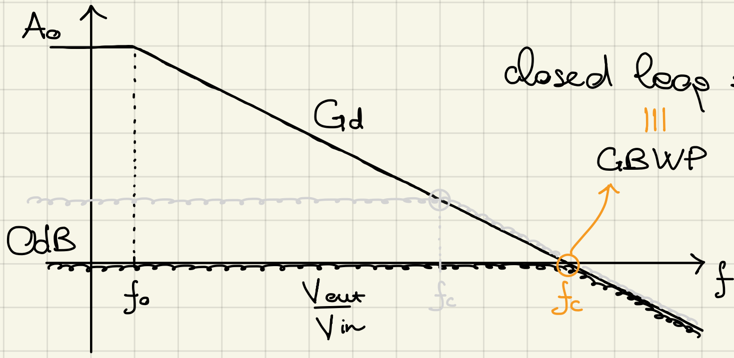
Note how changing C_H won't affect the position or the amplitude of the resonance peak.
Instead, changing G_{m3} or G_{m2} (or even C_m) will change the Q factor and thus the peak amplitude (since Q is directly proportional to the peak height), moving it below the 0dB axis.

Slew Rate and Settling time



$$G_d = \frac{A_0}{1+s\tau_0}$$





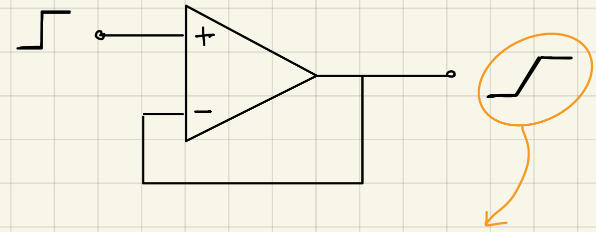
$$\underset{\underset{GBWP}{\equiv}}{\text{closed loop pole}} \longrightarrow f_c = GBWP = \frac{1}{2\pi\tau}$$

$$\Longrightarrow \tau = \frac{1}{2\pi\, GBWP}$$

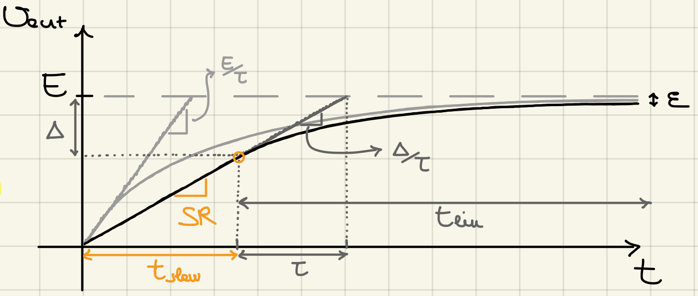
$$v_{out} = E\left(1-\underbrace{e^{-t/\tau}}_{\varepsilon}\right)$$

- $t = 3\tau \rightarrow \varepsilon = 5\%$
- $t = 5\tau \rightarrow \varepsilon = 1\%$
- $t = 7\tau \rightarrow \varepsilon = 1‰$

However closed loop pole is NOT the only limitation:



Ramp rate or Slew Rate (SR)



At the beginning of the response, if the initial exponential slope $\left(\frac{E}{\tau}\right)$ is steeper than the electronics slew rate (SR), then it will grow linearly according to the slew rate.
The response will then move from the slew rate limited region to the linear region (exponential growth) in such a way that the continuity of the derivative of the signal is retained. In other words, the link between the two branches happens when:

$$\left[\frac{\Delta}{\tau} = SR\right]$$

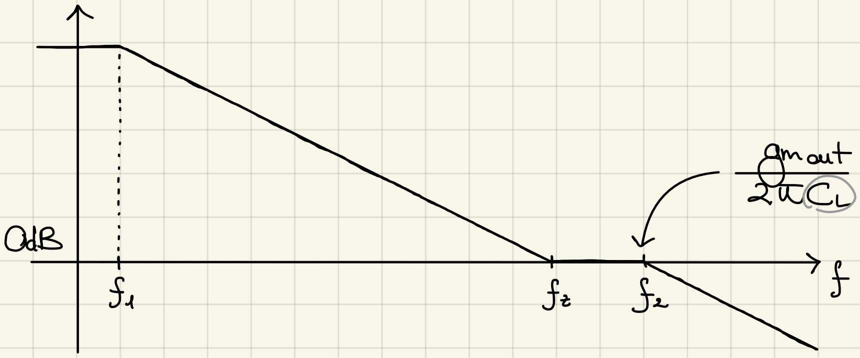
this relation returns the amplitude of Δ as well as the length of t_{slew}

settling time ← $t_s = t_{slew} + t_{lin}$

$$\boxed{t_{slew} = \frac{E-\Delta}{SR}}$$

$$E - \Delta e^{-t/\tau} = E(1-\varepsilon) \rightarrow \boxed{t_{lin} = \tau \ln\left(\frac{\Delta}{\varepsilon E}\right)}$$

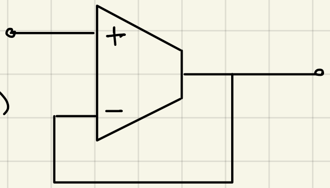
Consider now an amplifier with more than just one pole:

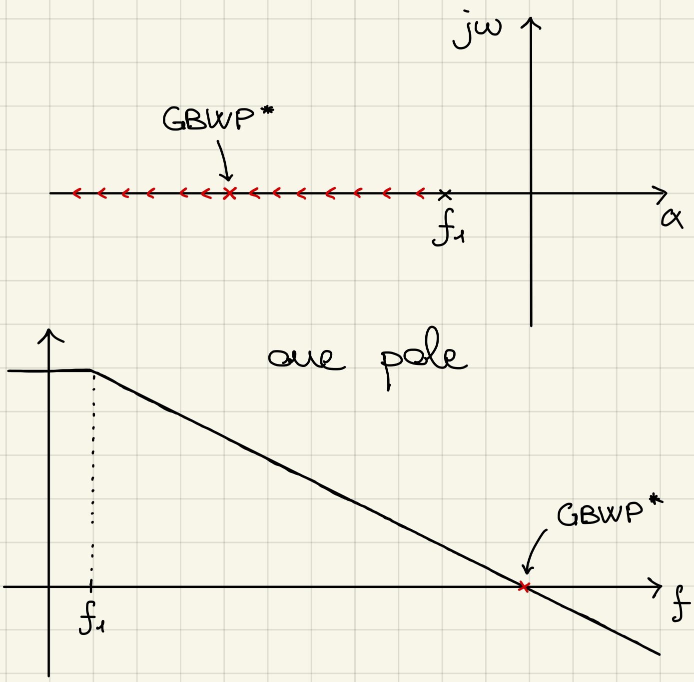


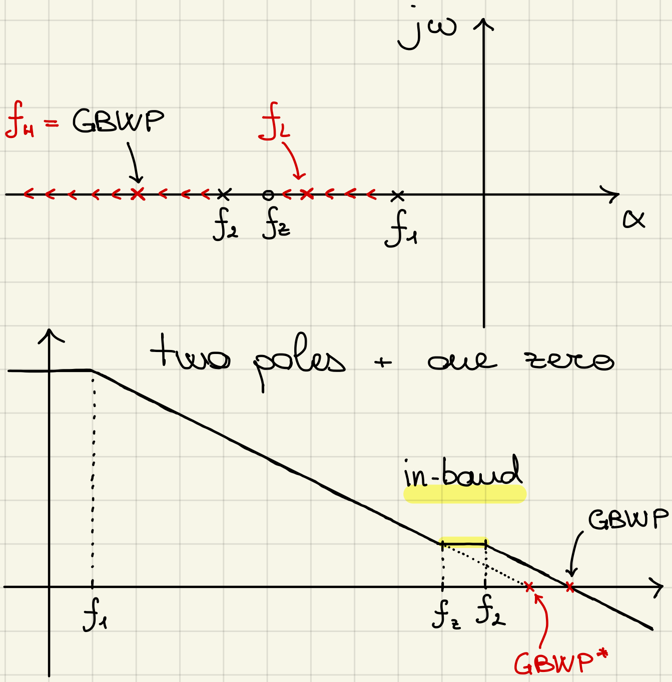
Wouldn't it be better to move f_z and f_2 at higher frequencies, above the 0dB axis ("in band"), so that the amplifier can better handle higher load capacitances?

Let's study the behaviour of the closed loop singularities in a simple buffer configuration:

$$G_{loop}(s) = -G_0 \frac{1+s\tau_z}{(1+s\tau_1)(1+s\tau_2)}$$







$$T(s) = \frac{V_{out}}{V_{in}}(s) = \underbrace{\frac{G_{loop}(s)}{1 - G_{loop}(s)}}$$

$$1 - G_{loop}(s) = 0$$

$$+\frac{G_0(1+s\tau_z)}{(1+s\tau_1)(1+s\tau_2)} + 1 = 0$$

$$G_0(1+s\tau_z) + (1+s\tau_1)(1+s\tau_2) = 0$$

$$s^2\tau_1\tau_2 + s(\tau_1+\tau_2+G_0\tau_2) + (G_0+1) = 0$$

$$* \; p_L \simeq \frac{G_0+1}{G_0\tau_2+\tau_1+\tau_2} \xrightarrow{G_0\to\infty} \frac{1}{\tau_2}$$

$$p_L = \frac{1}{\tau_L} \Longrightarrow \tau_L = \frac{G_0\tau_2+\tau_1+\tau_2}{G_0+\cancel{1}}$$

$$\simeq \tau_2 + \frac{\tau_1+\cancel{\tau_2}}{G_0}$$

$$\simeq \tau_2 + \boxed{\frac{\tau_1}{G_0}} \rightarrow \frac{1}{2\pi\cdot GBWP^*}$$

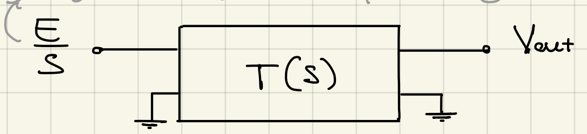
valid if $p_L << p_H$

$$* \; p_H \simeq \frac{\tau_1+\tau_2+G_0\tau_2}{\tau_1\tau_2} \simeq \frac{G_0\tau_2}{\tau_1\tau_2} = 2\pi GBWP^* \frac{f_2}{f_z}$$

$$\Longrightarrow GBWP = GBWP^* \frac{f_2}{f_z} \quad \text{(0dB crossover)}$$

laplace transform of a step as high as E

Step response:



$$V_{out}(s) = \frac{E}{S}T(s) = \frac{E}{S}\frac{(1+s\tau_z)}{(1+s\tau_L)(1+s\tau_H)} = \frac{E}{S}\left[\frac{A}{1+s\tau_L}+\frac{B}{1+s\tau_H}\right]$$

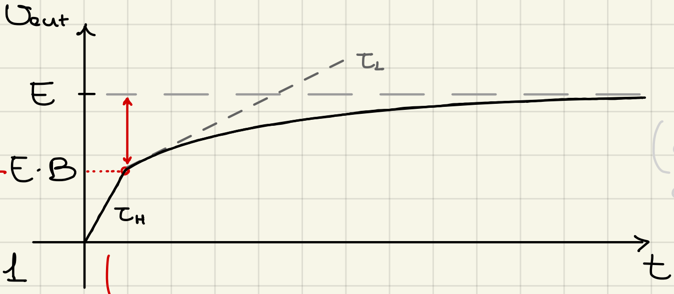
$$A = \lim_{s\to -1/\tau_L} \frac{1+s\tau_z}{(1+s\tau_L)(1+s\tau_H)}\cdot(1+s\tau_L) = \frac{1-\tau_z/\tau_L}{1-\tau_H/\tau_L} = \frac{\tau_L-\tau_z}{\tau_L-\tau_H}$$

$$B = \lim_{s\to -1/\tau_H} \frac{1+s\tau_z}{(1+s\tau_L)(1+s\tau_H)}(1+s\tau_H) = \frac{1-\tau_z/\tau_H}{1-\tau_L/\tau_H} = \frac{\tau_z-\tau_H}{\tau_L-\tau_H}$$

$$v_{out}(t) = E\left[A\left(1-e^{-t/\tau_L}\right) + B\left(1-e^{-t/\tau_H}\right)\right] =$$

$$= E\left[\underbrace{A+B}_{\frac{\tau_L-\tau_z+\tau_z-\tau_H}{\tau_L-\tau_H}=1} - Ae^{-t/\tau_L} - Be^{-t/\tau_H}\right] =$$

$$\Longrightarrow v_{out}(t) = E\left[1 - Ae^{-t/\tau_L} - Be^{-t/\tau_H}\right] \qquad \tau_H << \tau_L$$



(only considering time constant limitations, no SR)

when $e^{-t/\tau_H} \simeq 0$
e^{-t/τ_L} is still $\simeq 1$

the term related to τ_H "dies" much faster

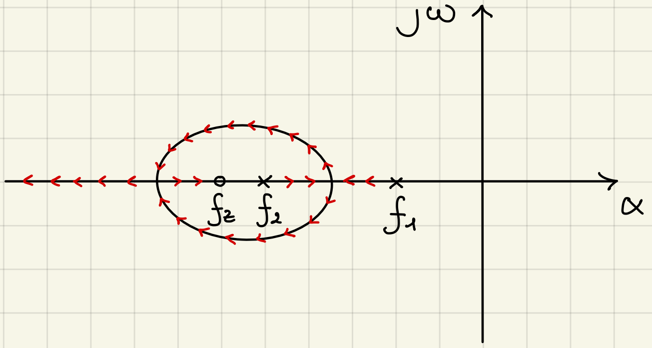
$$* \; A = \frac{\tau_L-\tau_z}{\tau_L-\cancel{\tau_H}} \simeq \frac{\tau_L-\tau_z}{\tau_L} \simeq \frac{\tau_1}{G_0\tau_z} = \frac{\boxed{f_z}}{GBWP^*}$$

⟹ Moving the zero (f_z) and the high frequency pole (f_2) at frequencies lower than the GBWP* while in a closed loop buffer configuration, will cause the closed loop pole (f_c = GBWP*) to split into a low frequency pole (f_L) and a high frequency pole (f_H = GBWP). As the zero moves towards lower frequencies, f_L will decrease accordingly.

Considering the response of such configuration to a step signal, it will have a first initial phase during which the exponential term related to the high pole f_H will rapidly reach the asymptotic value. The lower the zero, the shorter this phase and the higher its endpoint. However, it will also have a second following phase related to the exponential term of the low pole f_L, which will instead slowly reach the asymptote. The lower the zero, the longer this pase (since $\tau_L \propto \tau_z$).

For this reason, in order to have an overall faster step response, it is better not to have an amplifier with an in-band doublet.

The same reasoning can be applied to an amplifier with the second pole followed by the zero:



f_z decreases → less peaking but longer transient

What causes the Slew Rate

Buffer configuration. Step signal with amplitude E applied to v^+; v^- can be seen as fixed (feedback has not occurred yet).

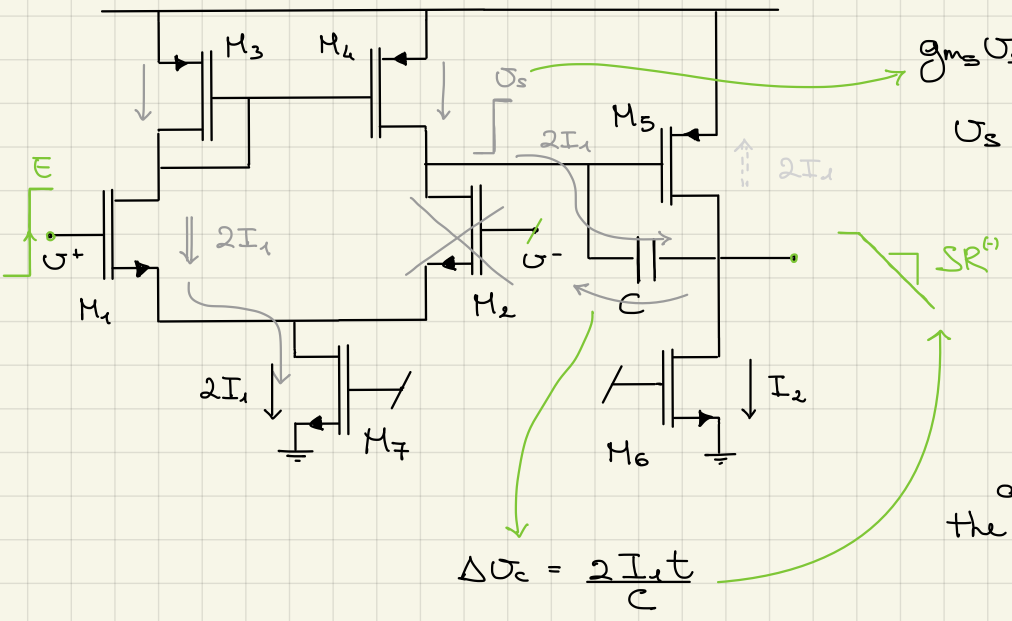
If E is high enough (at least > $(\sqrt{2}-1)V_{ov1}$) all current of the input stage ($2I_1$) will flow through only one branch.

Rising edge:

this is actually not so accurate since the small signal approximation does not hold anymore

$$g_{m5} v_s \simeq 2I_1$$

$$v_s \simeq \frac{2I_1}{g_{m5}}$$



$$\Longrightarrow \boxed{SR^{(-)} = \frac{2I_1}{C}} \simeq 50 \frac{V}{\mu s}$$

150µA

3pF

Note that while the input is positive, the output is negative, as the overall gain of the two stages is negative.

In order for the transistor M_5 to carry $2I_1$ signal current upward, it must be biased downward with an even grater current (i.e. the transistor cannot carry a total current = signal + bias that is negative otherwise it would turn off).

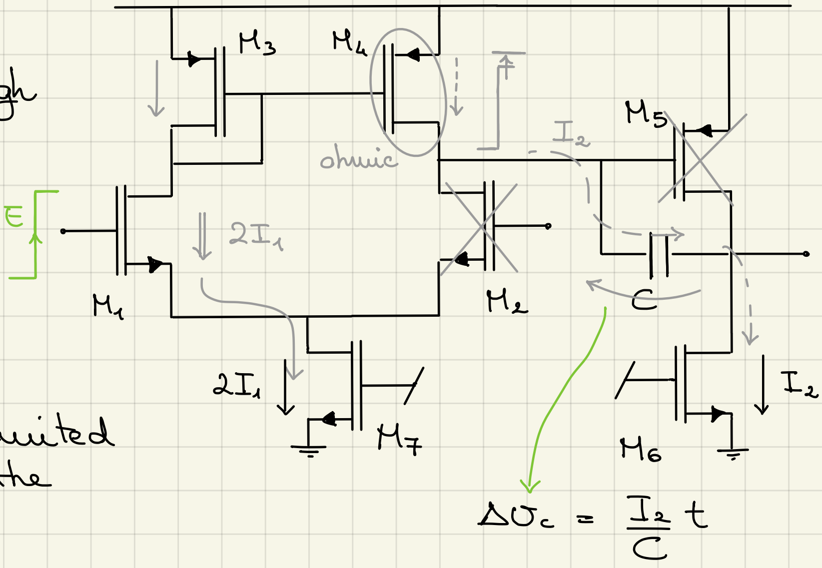
$$\Longrightarrow I_2 \geq 2I_1 = \Delta I_5$$

What happens if this condition is not met?

M_5 will indeed turn off. Furthermore, M_4 will have to go into ohmic region since it cannot carry $2I_1$ current anymore but instead has to match $I_2 < 2I_1$ forced by M_6, which is now the only path the current can flow through.

Hence the current through capacitor C will be I_2 instead of $2I_1$

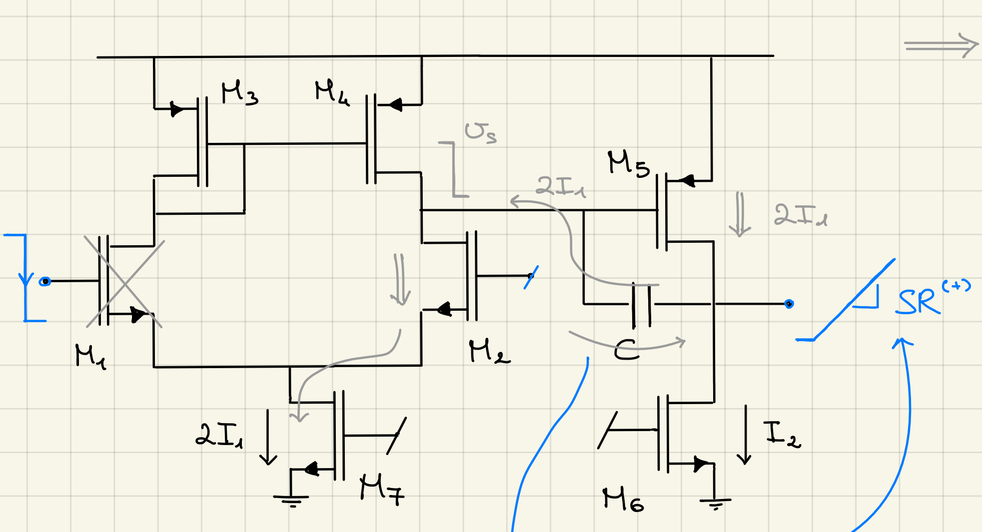
$$\Longrightarrow \boxed{SR^{(-)} = \frac{I_2}{C}}$$



The Slew Rate is now limited by the bias current of the second stage.

Falling edge:

$$\Longrightarrow \left[SR^{(+)} = \frac{2I_1}{C} \right]$$

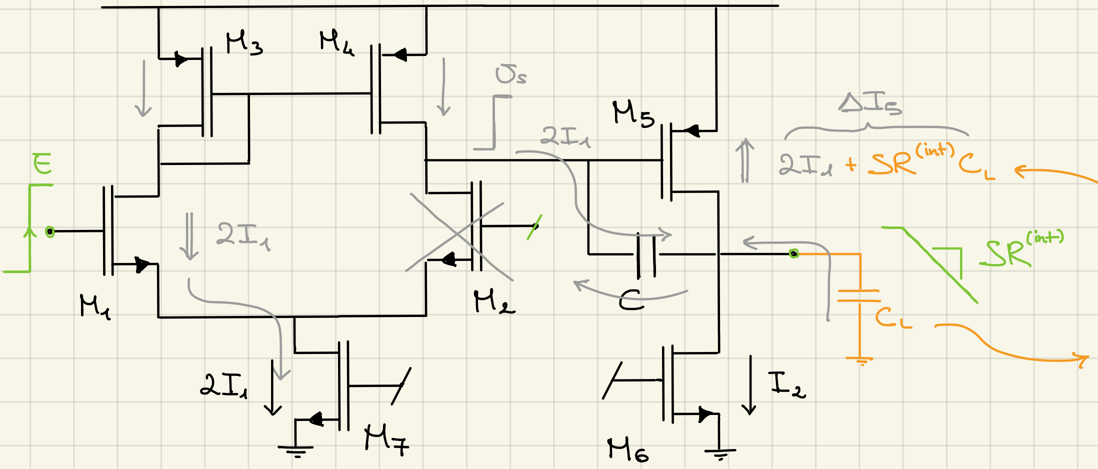


$$\Delta v_c = \frac{2I_1}{C} t$$

The circuit behaves symmetrically like before.

However, there is no "pathology", such as M_5 turning off, or M_2 going into ohmic (since its current is fixed by M_7).

What happens if we consider the load capacitance too?



$$I_{C_L} = \frac{dV_{C_L}}{dt} C_L$$
$$= SR^{(int)} C$$

$\Delta I_5 = 2I_1 + SR^{(int)} C_L \leq I_2$ to keep M_5 on.

$$2I_1 + \frac{2I_1}{C} C_L \leq I_2$$

$\Longrightarrow \quad I_2 \geq 2I_1 \frac{C+C_L}{C}$ more strict than before.

If the condition is not met, M_5 will turn off, M_4 will go into ohmic and M_6 will drain all the current (just like before)

$$I_2 = I_C + I_{C_L} =$$
$$= \frac{dV_C}{dt} C + \frac{dV_{C_L}}{dt} C_L =$$
$$= SR^{(ext)} C + SR^{(ext)} C_L$$

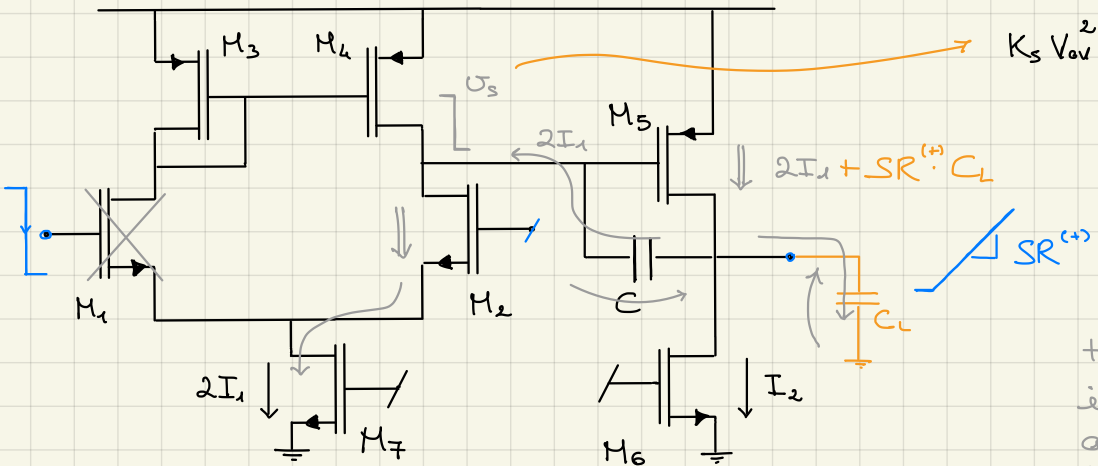
$V_C = v_S - v_{out}$ but v_S does not change (it instantly reaches the steady state value required by M_4)

$V_{C_L} = - v_{out}$

$$\Longrightarrow \boxed{SR^{(ext)} = \frac{I_2}{C + C_L}}$$

$$SR^{(-)} = \begin{cases} SR^{(int)} = \dfrac{2I_1}{C} & \text{if} \quad I_2 \geqslant 2I_1\left(1 + \dfrac{C_L}{C}\right) \\ SR^{(ext)} = \dfrac{I_2}{C + C_L} & " \quad I_2 < 2I_1\left(1 + \dfrac{C_L}{C}\right) \end{cases}$$

limited by both I_2 and C_L



$$K_S V_{ov}^2 = I_2 + \Delta I_S = \\ = I_2 + 2I_1\left(1 + \frac{C_L}{C}\right)$$

Similarly to the situation without C_L, there is no relevant issue related to a falling step signal (positive SR) even with a load capacitance

$$\Delta I_S = 2I_1 + SR^{(+)} C_L = 2I_1\left(1 + \frac{C_L}{C}\right)$$

$$SR^{(+)} = SR^{(int)} = \frac{2I_1}{C}$$

To summarize:
$$SR \begin{cases} \rightarrow SR^{(+)} = SR^{(int)} \\ \rightarrow SR^{(-)} = \begin{cases} SR^{(int)} = \dfrac{2I_1}{C} \\ SR^{(ext)} = \dfrac{I_2}{C + C_L} \end{cases} \text{if } I_2 \gtrless 2I_1\left(1 + \dfrac{C_L}{C}\right) \end{cases}$$

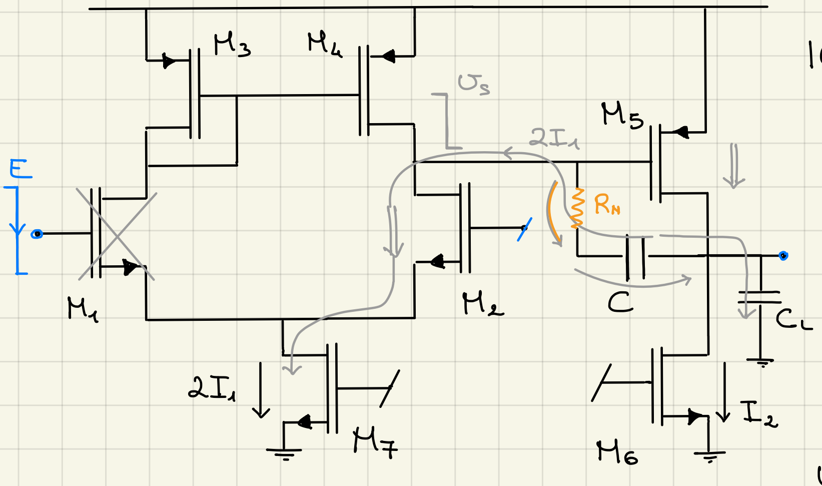
$\Longrightarrow$ New Figure of Merit: $\boxed{FoM := \dfrac{SR \cdot C_L}{I_{tot}}}$ SR - related

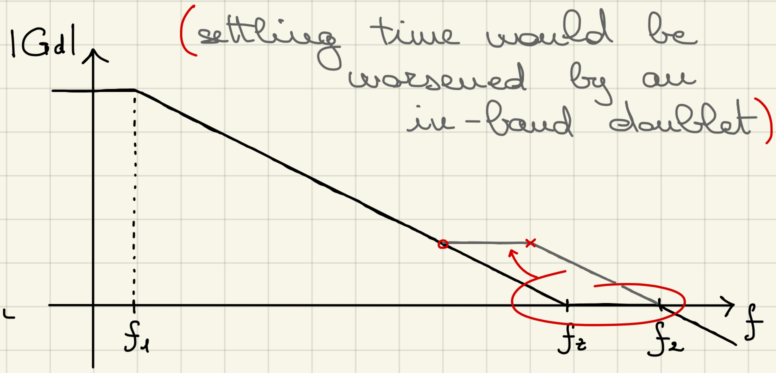
$\boxed{FoM := \dfrac{GBWP \cdot C_L}{I_{tot}}}$ GBWP - related

We generally want a <u>simmetric Slew Rate</u> ($SR^{(+)} = SR^{(-)}$), so having $I_2 \geqslant 2I_1\left(1 + \frac{C_L}{C}\right)$ is often mandatory, even if it means <u>more power consumption</u>.

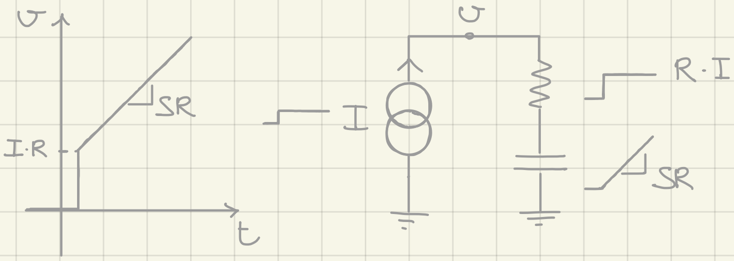
$$FoM = \frac{SR \cdot C_L}{I_{tot}} = \frac{2I_1}{C} \cdot \frac{C_L}{2I_1 + 2I_1\left(1 + \frac{C_L}{C}\right)} = \frac{C_L}{2C + C_L} \xrightarrow{C_L \to +\infty} 1$$

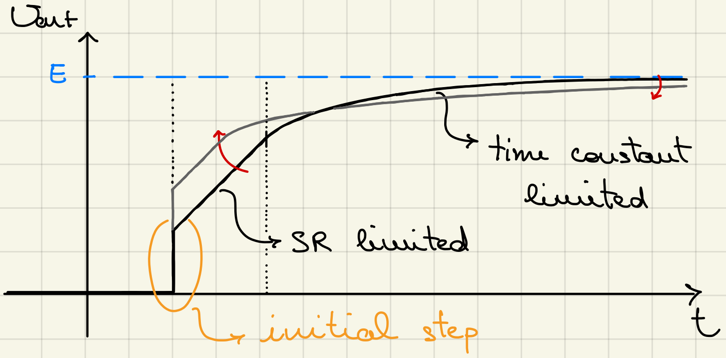
What happens (to the SR) if we consider a compensating <u>nulling resistor</u>?





(settling time would be worsened by an in-band doublet)

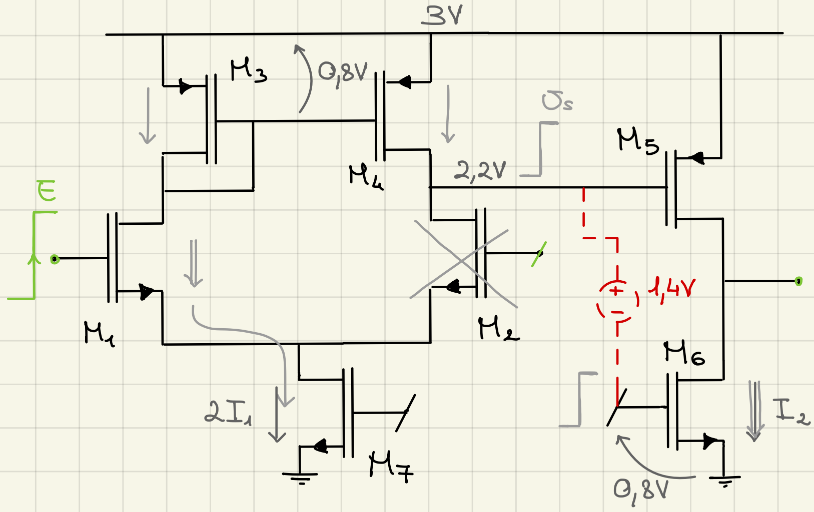




⟹ The nulling resistor introduces a step at the beginning of the response

Is it possible to improve the Slew Rate of the circuit, without impairing all other parameters, but most importantly without increasing the power consumption (that is, without using a large I_2)?

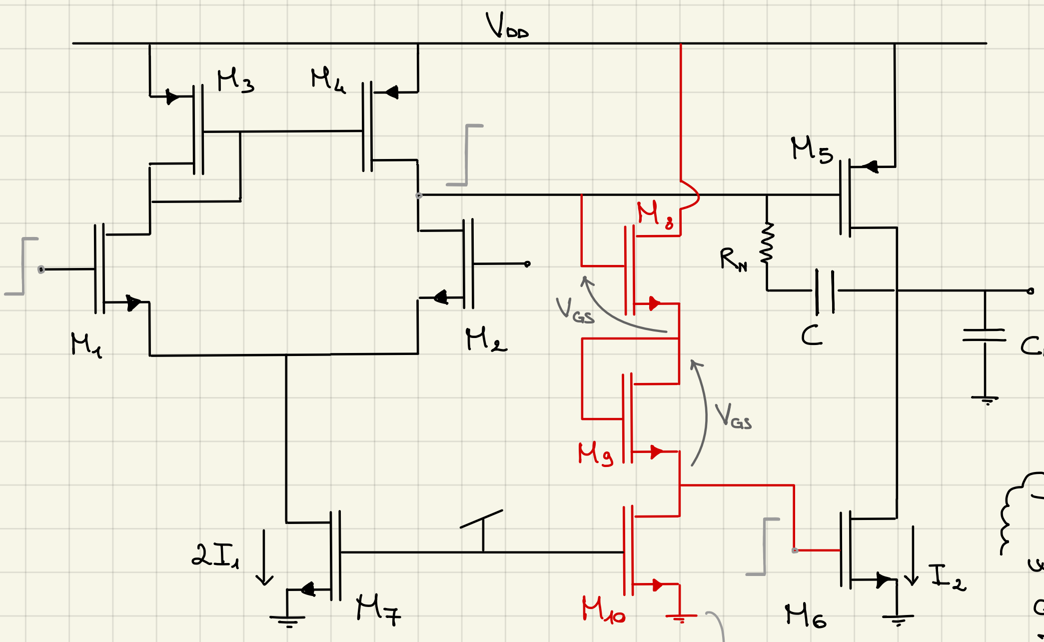
We should redesign the tail generator M_6 since we want it to carry large current only during the transient (when the SR limitation occurs) but a small current is sufficient during any other operating point.



⟹ Use the voltage increase at the drain of M_4 to pilot the gate of M_6 and therefore increase I_2.

We must fix and control the overdrive of M_6, so there needs to be a voltage offset between the two nodes.

How can we build this voltage offset? ⟹ Transdiodes



class A-B stage

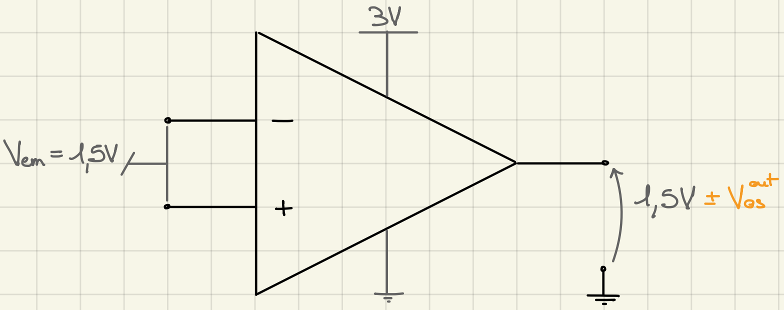
→ the added branch current is widely compensated by the reduced I_2 current

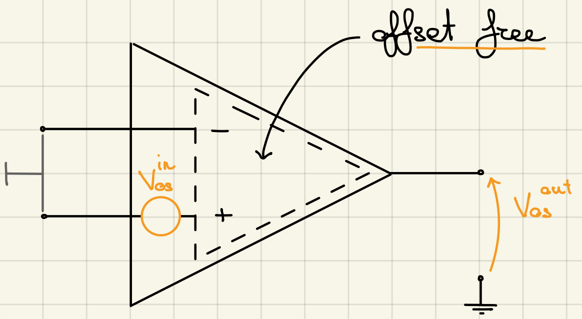
Allows to keep low bias current I_2 (therefore lower power consumption) while retaining the internal SR.

$$\uparrow FoM = \frac{SR \cdot C_L}{I_{tot}} \downarrow$$

This configuration will also (positively) affect the overall gain and phase margin

Input Referred Offset





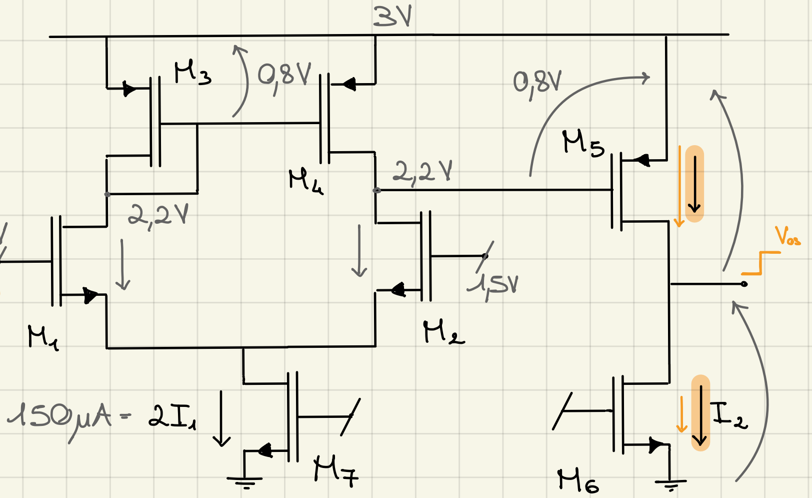
$$V_{os}^{out} = A_d \, V_{os}^{in}$$

statistic — should be kept below a resonable value

dete~~rministic~~ — should be 0, since it can be controlled by the designer

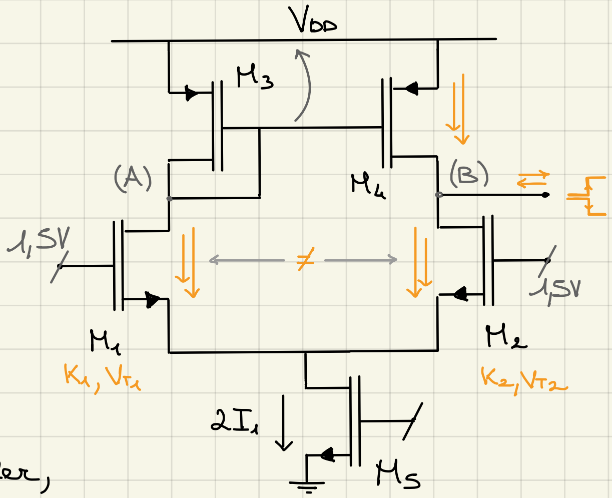
Deterministic offset

Transistors M5 and M6 must be sized so that their currents precisely match each other given a mid-range common mode input.



If not, their drain will increase or decrease accordingly to compensate, returning a voltage offset at the output

Statistic offset



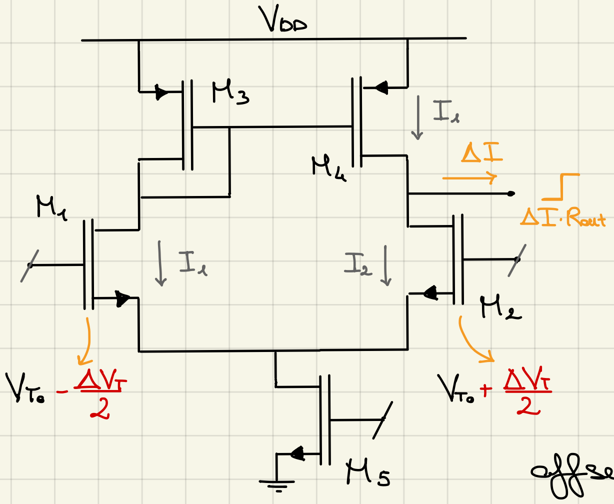
(A) and (B) should theoretically be at the same voltage level, provided that both transistor pairs are symmetrical.

In case of a mismatch in the transistors parameters (like different V_T or K) then (A) and (B) may differ, causing a different current flow in the two branches and therefore a residual, non-negligible current at the output of the stage (source of the offset).

Since the variation of the transistors parameters is a statistical matter, it can be represented as a gaussian function whose spread depends on a certain variance σ^2.

The objective is to find the expression of this σ and to find its relation with the variance of the output offset.

→ We superpose to the ideal DC voltage condition the variation due to the mismatch.

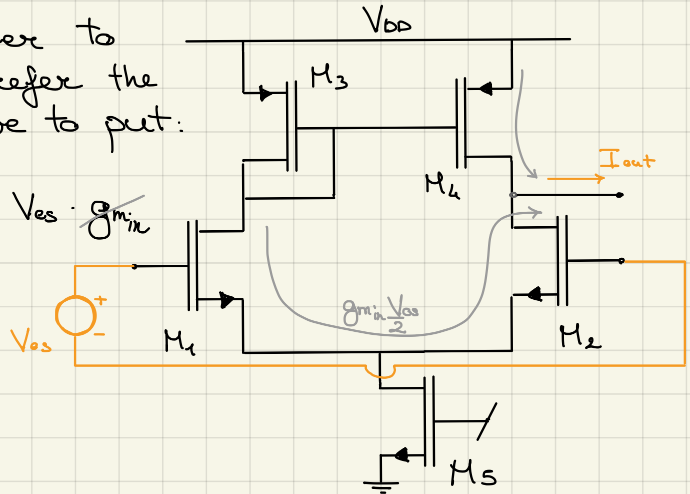


$$I_1 = K_{in}\left(\overbrace{V_{GS_1} - V_{T_0}}^{V_{ov}^{\circ}} + \frac{\Delta V_T}{2}\right)^2$$

$$I_2 = K_{in}\left(V_{GS_2} - V_{T_0} - \frac{\Delta V_T}{2}\right)^2$$

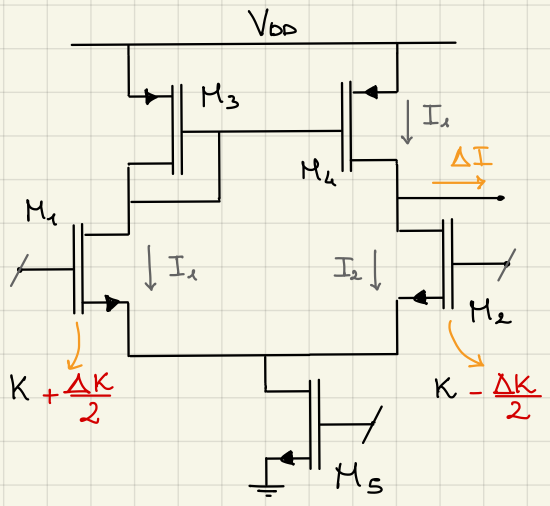
$$\hookrightarrow \Delta I = I_1 - I_2 = 2\,K_{in}V_{ov}^{\circ}\,\Delta V_T$$

In order to input-refer the offset we have to put:



$$\Delta I = I_{out}\Big|_{V_{os}} \longrightarrow \overset{g_{min}}{\cancel{2K_{in}V_{ov}^{\circ}}}\,\Delta V_T = V_{os}\cdot\cancel{g_{min}}$$

$$\Longrightarrow \boxed{V_{os}\Big|_{\Delta V_{T\,in}} = \Delta V_T}$$



$$I_1 = \left(K + \frac{\Delta K}{2}\right)\left(V_{ov_1}\right)^2$$

$$I_2 = \left(K - \frac{\Delta K}{2}\right)\left(V_{ov_2}\right)^2$$

$$\hookrightarrow \Delta I = I_1 - I_2 = \Delta K\, V_{ov}^2 = \overset{V_{os}\, g_m}{\overset{\parallel}{I_{out}}}\Big|_{V_{os}}$$

$$\longrightarrow V_{os} = \frac{\Delta K\, V_{ov}^2}{g_m} = \frac{\Delta K}{K}\,\frac{\overset{I}{V_{ov}^2 \cdot K}}{g_m}$$

$$= \frac{\Delta K}{K}\,\frac{\cancel{I}}{2\cancel{I}} \cdot V_{ov}$$

$$\Longrightarrow \boxed{V_{os}\Big|_{\Delta K_{in}} = \frac{\Delta K}{K}\,\frac{V_{ov}}{2}}$$

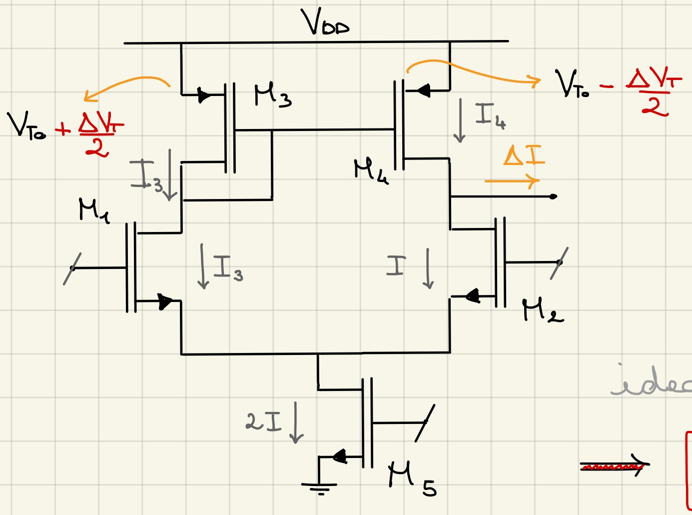
↳ related to the input transistors mismatch

So any mismatch will cause a definite contribution to the input-referred offset. However these mismatches are not a number whose value is deterministically known but they are a variance, that is a measure of the spread of the values the mismatch can assume.

$$V_{os_{in}} = \Delta V_T + \frac{\Delta K}{K} \cdot \frac{V_{ov}}{2}$$

$$\boxed{\sigma^2_{V_{os_{in}}} = \sigma^2_{\Delta V_T} + \sigma^2_{\frac{\Delta K}{K}} \cdot \left(\frac{V_{ov}}{2}\right)^2}$$

Mirror transistors can be a source of mismatch too:



$$I_3 = K_M\Big(\overbrace{V_{GS_3} - V_{T_0}}^{V_{ov}^{\circ}} + \frac{\Delta V_T}{2}\Big)^2 = I$$

$$I_4 = K_M\left(V_{GS_4} - V_{T_0} - \frac{\Delta V_T}{2}\right)^2$$

$$\hookrightarrow \Delta I = I_4 - I_3 = 2K_M V_{ov}^{\circ}\, \Delta V_T$$

$$= g_{m_M}\, \Delta V_T$$

ideal, without mismatch!

$$\Longrightarrow \boxed{V_{os}\Big|_{\Delta V_{T_M}} = \frac{g_{m_M}}{g_{m_{in}}}\, \Delta V_T = \frac{V_{ov\,in}}{V_{ov\,M}}\, \Delta V_T}$$

Same calculation can be done for the K factor:

$$\Longrightarrow \boxed{V_{os}\Big|_{\Delta K_M} = \frac{\Delta K}{K} \cdot \frac{V_{ov\,in}}{2}}$$

To sum up all contributions:

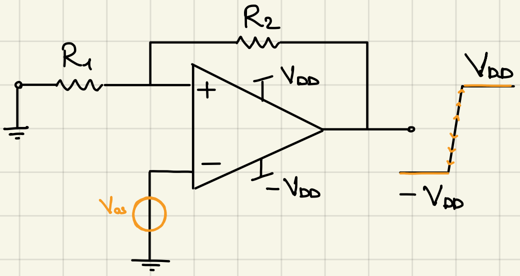
$$V_{os_{tot}} = \left(\Delta V_{T_{in}} + \Delta V_{T_M} \cdot \frac{V_{ov\,in}}{V_{ov_M}}\right) + \left(\frac{\Delta K_{in}}{K_{in}} + \frac{\Delta K_M}{K_M}\right) \cdot \frac{V_{ov\,in}}{2}$$

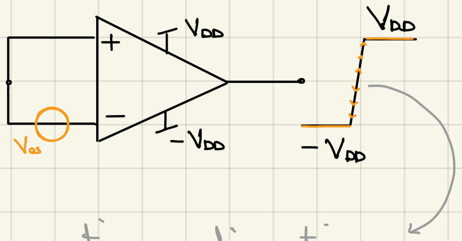
$$\left[\sigma_{V_{os}} = \sqrt{\sigma^2_{\Delta V_{T_{in}}} + \sigma^2_{\Delta V_{T_M}} \left(\frac{V_{ov\,in}}{V_{ov_M}}\right)^2 + \left(\sigma^2_{\frac{\Delta K}{K_{in}}} + \sigma^2_{\frac{\Delta K}{K_M}}\right) \cdot \left(\frac{V_{ov\,in}}{2}\right)^2}\right]$$

How do we control $\sigma^2_{\Delta V_T}$ and $\sigma^2_{\frac{\Delta K}{K}}$ in order to reduce the statistical offset?

The variance of these parameters is generally set by the technology and the size of the transistors.
For instance, K depends on the mobility μ, the oxide capacitance C'_{ox} and the form factor $\frac{W}{L}$, which all of them can fluctuate from their nominal value causing the mismatch in K.
In the same way, V_T is also a function of C'_{ox} and in general is dependent on the metal-oxide-semiconductor junctions and therefore on the MOS technology.
This arguments will be thoroughly discussed later on.

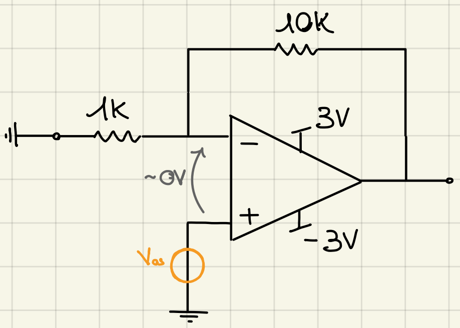
Note how the offset (which is typically in the order of few mV) causes the amplifier to saturate every time it is in a positive feedback or in a feedback-less configuration, even with no input whatsoever.





the saturation direction is non-deterministic!

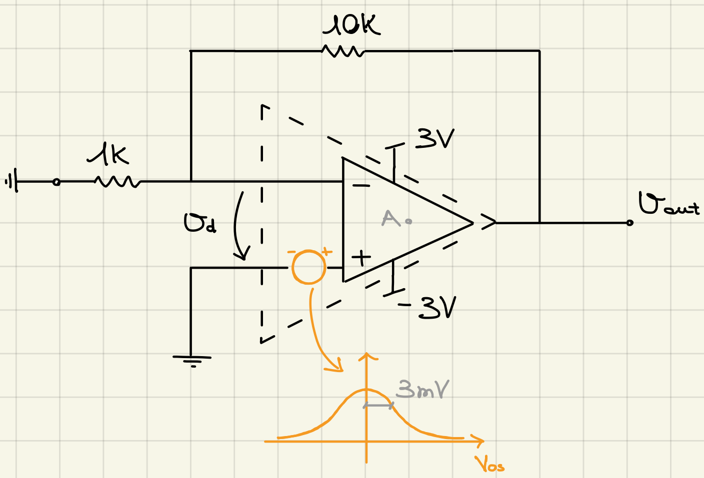
Only a negative feedback can allow to have a stable, non-saturated output (with, however, a fixed offset).

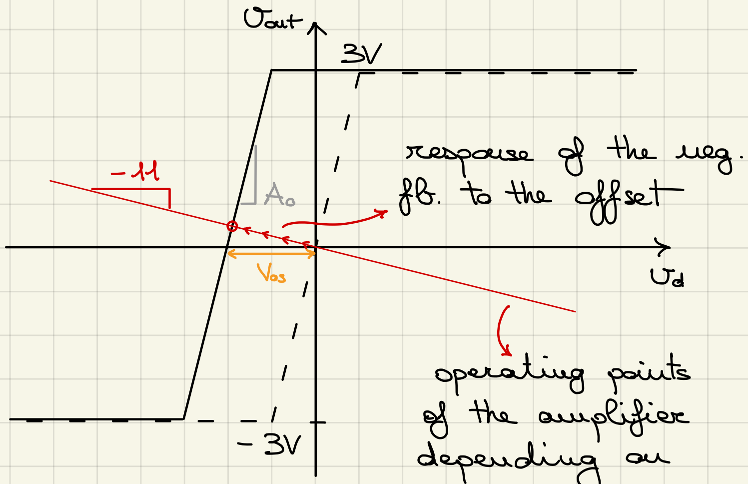


r.m.s. values

$$v_{out} = \overset{\pm 3mV}{V_{os}} \left(1 + \frac{10k}{1k}\right) \simeq \pm 33\ mV$$

The negative feedback basically operates to balance the internal mismatch of the OPAMP (that is, the offset) by varying the output voltage and therefore adjusting the input signal to achieve compensation.



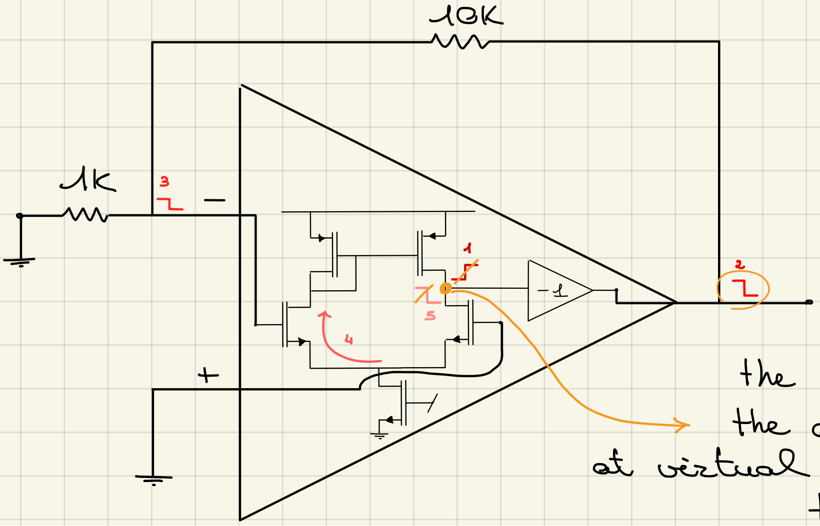


response of the neg. fb. to the offset

operating points of the amplifier depending on the offset

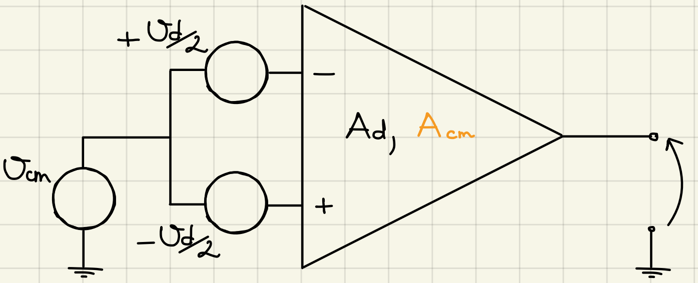
$$v_d = -v_{out}\frac{1K}{1K+10K} \longrightarrow v_{out} = -v_d\left(1+\frac{10K}{1K}\right) = -11$$

To have a deeper insight into understanding the effect of the offset and the stabilization of the neg. fb.:



the internal node from which the offset is generated is kept at virtual ground by the feedback through an output offset

|| Common Mode Rejection Ratio ||

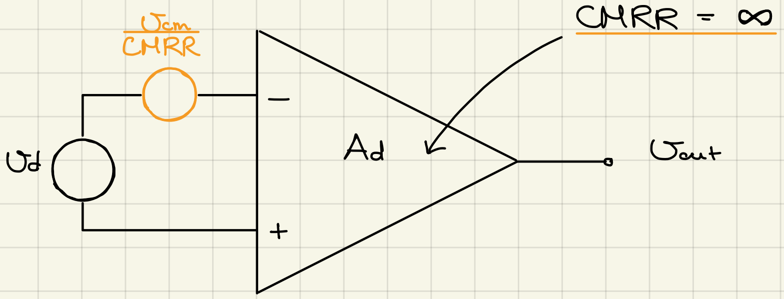


$$v_{out} = A_d v_d + A_{cm} v_{cm}$$
$$= A_d\left(v_d + \boxed{\frac{A_{cm}}{A_d}}\, v_{cm}\right)$$

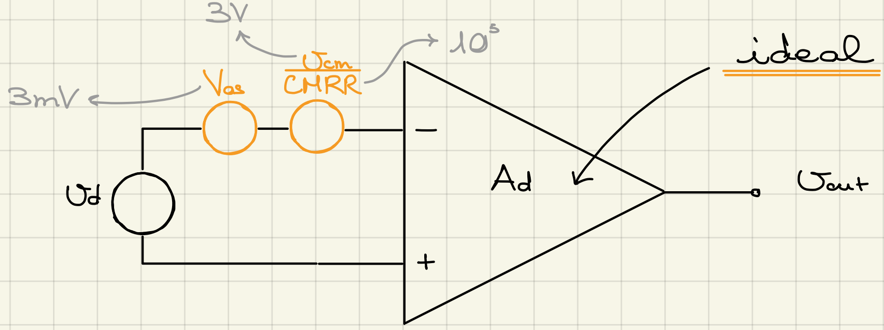
$$\boxed{CMRR = \frac{A_d}{A_{cm}}}$$

$$\longrightarrow \quad v_{out} = A_d\left(v_d + \frac{v_{cm}}{CMRR}\right)$$

The presence of a finite CMRR can be also modeled as an input-referred offset whose value depends on the common mode signal.



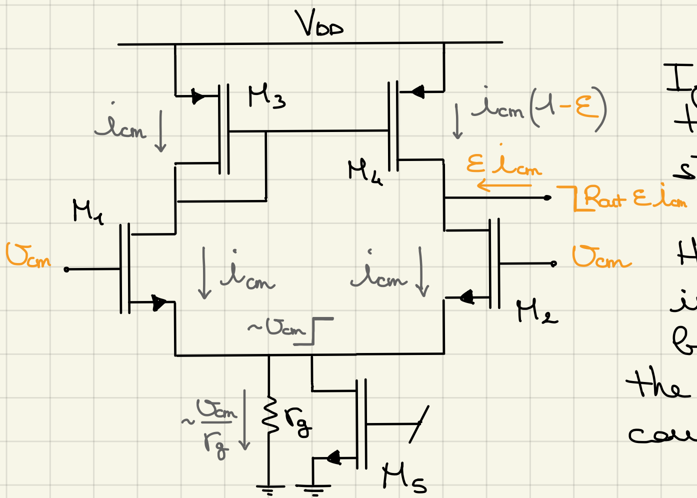
The effect of the CMRR occurs on top of the already discussed internal offset.



At a first glance, it might seem that the CMRR offset contribution is negligible compared to the internal offset (V_{os} is in the order of mV while $v_{cm}/CMRR$ is in the order of tens of μV). This is incorrect for 2 main reasons:

- the CMRR offset is time dependent, as v_{cm} can vary over time depending on the input and so its effects on the output are also variable, while the internal offset is typically constant (drifts only with temperature)
- the CMRR is itself frequency dependent, as it is a function of the amplifier gain; the CMRR can therefore degrade as frequency grows meaning its offset will not be so negligible anymore.

For these reasons it is necessary to better understand the CMRR and its defining factors.



If the stage is perfectly symmetric the common mode gain of the stage is ideally zero.

However in case of a mismatch in the current mirror, there will be some current flowing through the output branch, yelding a non-zero common mode gain.

$$v_{out}\Big|_{cm} = v_{cm}\, G_{cm} = \varepsilon\, i_{cm} R_{out} =$$

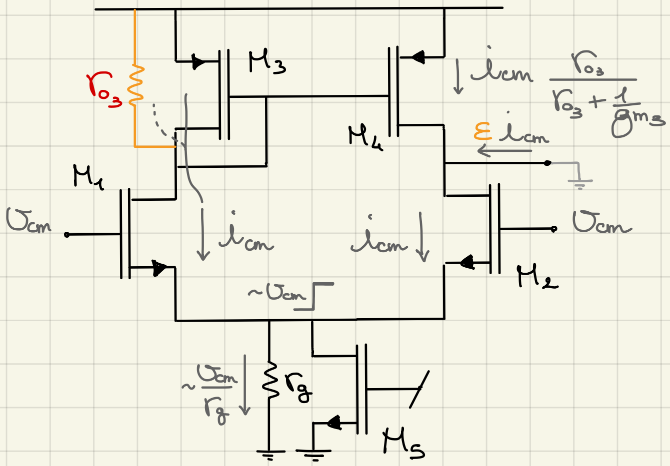
neglect the sign

$$= \frac{\varepsilon\, v_{cm} R_{out}}{2\, r_g} \longrightarrow G_{cm} = \frac{\varepsilon R_{out}}{2\, r_g}$$

$$\Longrightarrow \quad CMRR = \frac{G_d}{G_{cm}} = \frac{g_{m_{in}} R_{out}}{\frac{\varepsilon R_{out}}{2\, r_g}} = 2\,\frac{g_{m_{in}} r_g}{\varepsilon}$$

(the lower the error, the better the CMRR)

What is the source of this error?



$$\varepsilon\, i_{cm} = i_{cm} - i_{cm}\frac{r_{o_3}}{r_{o_3}+\frac{1}{g_{m_3}}} =$$

$$= i_{cm}\left(\frac{1/g_{m_3}}{r_{o_3}+1/g_{m_3}}\right)$$

$$\Longrightarrow \quad \varepsilon = \frac{1}{1+g_{m_3} r_{o_3}} \simeq \frac{1}{\mu_3}$$

1st deterministic contribution

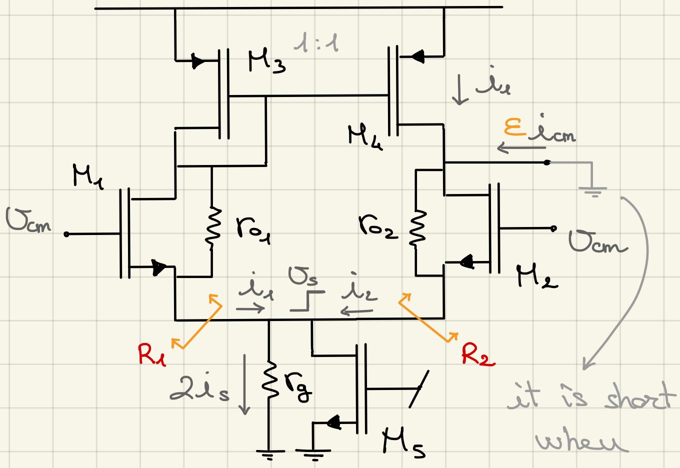
Moreover, there is another source of deterministic error that is due to the asymmetry of the stage seen from the tail generator.

$r_{o_{in}}$

$$R_1 \simeq \frac{1/g_{m_3} + r_{o_1}}{1+g_{m_1} r_{o_1}} \qquad R_2 = \frac{r_{o_2}}{1+g_{m_2} r_{o_2}}$$

μ_{in}

$$i_1 = \frac{v_s}{r_g}\,\frac{R_2}{R_1+R_2} \qquad i_2 = \frac{v_s}{r_g}\,\frac{R_1}{R_1+R_2}$$



$$\varepsilon\, i_{cm} = i_2 - i_1 = \frac{v_s}{r_g}\,\frac{R_1 - R_2}{R_1+R_2} =$$

$$= \frac{v_s}{r_g}\,\frac{1/g_{m_M}}{1/g_{m_M} + 2 r_{o_{in}}} \simeq \frac{v_s}{r_g}\,\frac{1}{2 g_{m_M} r_{o_{in}}}$$

it is short when computing the output current using Norton theorem (it was a short before too)

$$i_{cm} = \frac{v_{cm}}{2\, r_g} \qquad i_s = \frac{v_s}{2\, r_g} \qquad v_s \approx v_{cm}$$

$$\downarrow$$

$$\frac{v_s}{v_{cm}} = \alpha \lesssim 1$$

$$\rightarrow \quad \varepsilon \frac{v_{cm}}{2r_g} = \frac{v_s}{r_g}\,\frac{1}{2 g_{mM} r_{oin}} \Longrightarrow \varepsilon = \frac{\alpha}{g_{mM} r_{oin}} \leq \underset{\substack{\uparrow \\ \text{2nd deterministic contribution}}}{\frac{1}{g_{mM} r_{oin}}}$$

$$\Longrightarrow \boxed{\varepsilon_{det} = \frac{1}{g_{mM} r_{oM}} + \frac{1}{g_{mM} r_{oin}}} \approx 2\%$$

$$CMRR_{det} = \frac{2\, g_{min} r_g}{\varepsilon_{det}} \approx 2 \cdot 10^4 = 86dB$$

So far we have only discussed deterministic sources of finite CMRR, which in our differential stage cannot be completely cancelled out but can indeed be controlled through the circuit parameters.

There is also a statistical source of error, again due to process fabrication non-uniformities, that causes a finite CMRR.

Mirror pair transcond. mismatch

$$\varepsilon\, i_{cm} = i_{cm} - i_{cm}\frac{g_{m4}}{g_{m3}}$$

$$= i_{cm}\left[1 - \frac{g_{m4}}{g_{m3}}\right]$$

$$= i_{cm}\frac{\Delta g_{mM}}{g_{m3}}$$

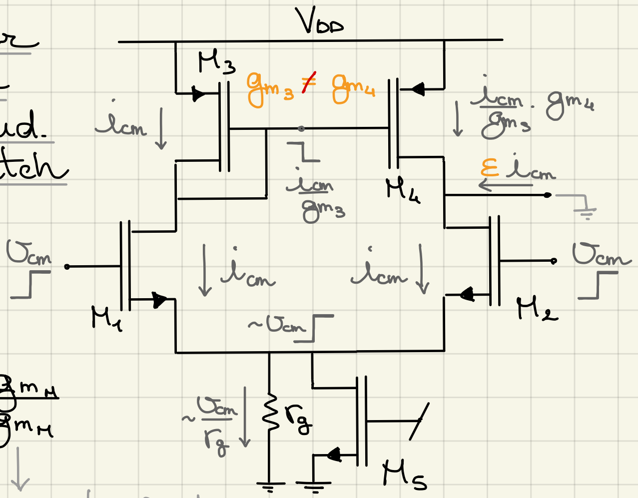
$$\simeq i_{cm}\frac{\Delta g_{mM}}{g_{mM}} \Longrightarrow \varepsilon = \frac{\Delta g_{mM}}{g_{mM}}$$

nominal transcond.

$\uparrow$

$g_{mM} = \frac{g_{m3}+g_{m4}}{2} \simeq g_{m3}$

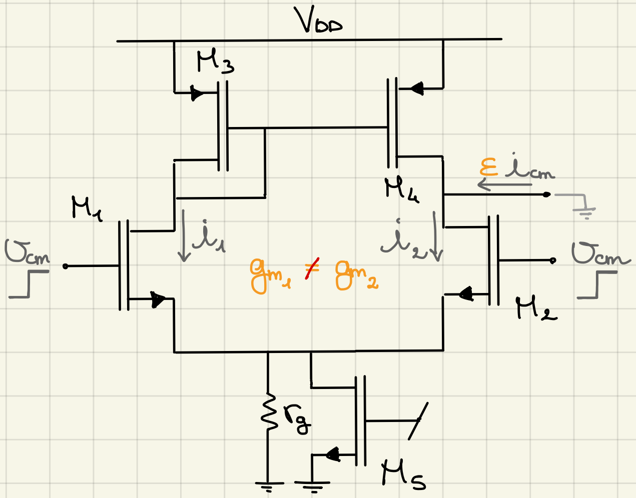
if Δg_{mM} is small



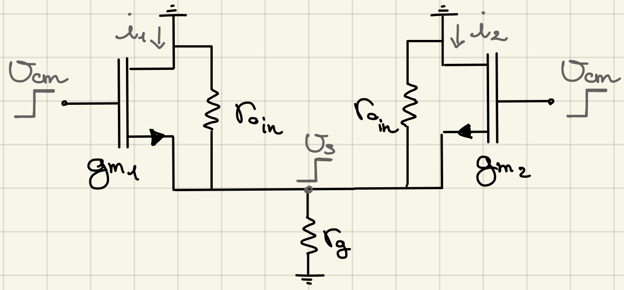
$\downarrow$

1st statistical contribution

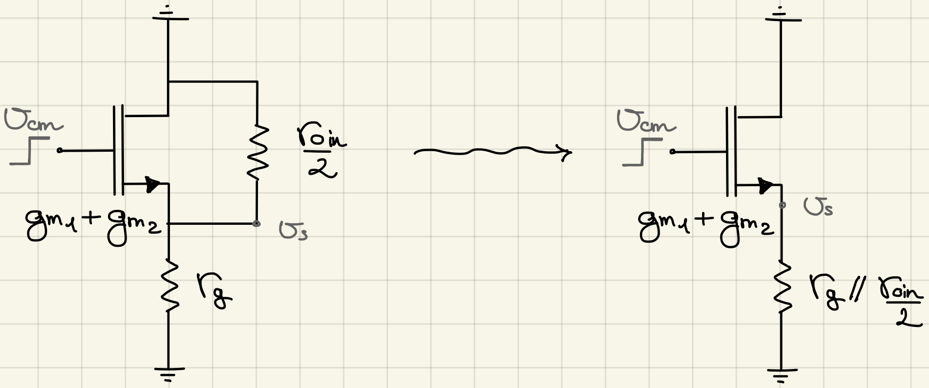
($\frac{\Delta g_{mM}}{g_{mM}}$ is represented by its variance)



Input pair transconductance mismatch



We can "fold" the circuit by taking advantage of its symmetry thus obtaining the new equivalent circuit:



$$v_s = v_{cm} \frac{\overbrace{r_g // \frac{r_{oin}}{2}}^{r_s^*}}{r_g // \frac{r_{oin}}{2} + \frac{1}{\underbrace{g_{m1}+g_{m2}}_{g_m^*}}}$$

Returning back to the previous circuit, we can now compute i_1 and i_2:

$$i_1 = \frac{v_s}{r_{oin}} + g_{m1}(v_{cm} - v_s) \qquad i_2 = \frac{v_s}{r_{oin}} + g_{m2}(v_{cm} - v_s)$$

$$\varepsilon\, i_{cm} = i_2 - i_1 = (g_{m2} - g_{m1})(v_{cm} - v_s) = \Delta g_{min} \cdot v_{cm}\left(1 - \frac{r_s^*}{r_s^* + \frac{1}{g_m^*}}\right) =$$

$$= \Delta g_{min}\, v_{cm}\left(\frac{1/g_m^*}{r_s^* + 1/g_m^*}\right) \simeq \frac{\Delta g_{min}}{g_m^*}\, v_{cm}\, \frac{1}{r_s^*} \simeq$$

$2 g_{min} \simeq g_{m1} + g_{m2}$ ↳ nominal transcond.

$$r_s^* \to \frac{r_g + r_{oin}/2}{r_g\, r_{oin}/2} = \frac{2r_g + r_{oin}}{r_g\, r_{oin}}$$

$$\simeq \frac{\Delta g_{min}}{2 g_{min}}\, v_{cm}\, \frac{2r_g + r_{oin}}{r_g\, r_{oin}} = \frac{\Delta g_{min}}{g_{min}}\, \underbrace{\frac{v_{cm}}{2r_g}}_{i_{cm}} \left(1 + \frac{2r_g}{r_{oin}}\right)$$

$$\Longrightarrow \varepsilon = \frac{\Delta g_{min}}{g_{min}}\left(1 + \frac{2r_g}{r_{oin}}\right) \longrightarrow$$ 2nd statistical contribution

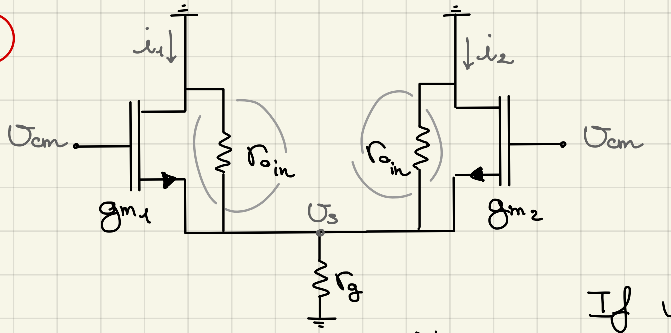
$$\Longrightarrow \boxed{\varepsilon_{stat} = \frac{\Delta g_{mM}}{g_{mM}} + \frac{\Delta g_{min}}{g_{min}}\left(1 + \frac{2r_g}{r_{oin}}\right)}$$ ①

$$\boxed{CMRR = \frac{2\, g_{min}\, r_g}{\varepsilon_{tot}} = \frac{2 g_{min} r_g}{\varepsilon_{det} + \varepsilon_{stat}}}$$ ②

① Why is the statistical contribution to the CMRR error related to the input pair grater, by a factor $\frac{2r_g}{r_{oin}}$, than the one related to the mirror pair?

② How can we add together ε_{det} and ε_{stat}, since the former is a definite number while the latter is a spread of values?

①



$$\mathcal{E}^{in}_{stat} = \frac{\Delta g_{m\,in}}{g_{m\,in}}\left(1 + \frac{2r_g}{r_{o\,in}}\right)$$

$$CMRR^{in}_{stat} = \frac{2 g_{m\,in} r_g}{\mathcal{E}^{in}_{stat}}$$

If we wanted to improve the CMRR, it would seem a good idea to increase r_g (increase channel length of M_3):

$$\lim_{r_g \to \infty} CMRR^{in}_{stat} = \lim_{r_g \to \infty} \frac{2 g_{min} r_g}{\frac{\Delta g_{min}}{g_{min}}\left(1+\frac{2r_g}{r_{oin}}\right)} = \frac{2 g_{min}}{\frac{\Delta g_{min}}{g_{min}}} \cdot \frac{r_{oin}}{2} < \infty!$$

However the CMRR does not tend to infinity by having an ideal tail generator as one would expect.

This is due to the <u>finite output resistance</u> of M_1 and M_2 which allows the current mismatch to have an additional path toward ground even when the tail transistor is ideal.

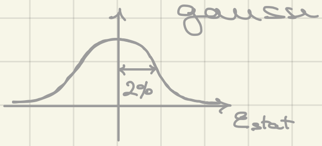
The additional $\frac{2r_g}{r_{oin}}$ factor related to the input pair is therefore needed to ensure that the CMRR will grow only if both the tail transistor AND the input transistors are built with an higher output impedance.

②

$$CMRR = \frac{2 g_{min} r_g}{\mathcal{E}_{tot}} \qquad \mathcal{E}_{tot} = \mathcal{E}_{det} + \mathcal{E}_{stat}$$

$\mathcal{E}_{det}$ → 2%

$\mathcal{E}_{stat}$ → $0 \pm 2\%$ → <u>root mean square</u> and → <u>mean value</u> of the gaussian distribution (#)



$$\mathcal{E}_{stat} = \frac{\Delta g_{mM}}{g_{mM}} + \frac{\Delta g_{m\,in}}{g_{m\,in}}\left(1 + \frac{2r_g}{r_{o\,in}}\right)$$

$$\sigma_\varepsilon^2 = \sigma^2_{\frac{\Delta g_m}{g_m}M} + \sigma^2_{\frac{\Delta g_m}{g_m}in}\left(1 + \frac{2r_g}{r_{o\,in}}\right)^2$$

$$\sigma^2_{\frac{\Delta g_m}{g_m}} = ?$$

$$g_m = 2K(V_{GS} - V_T) \qquad \Delta g_m = dg_m = 2dK(V_{GS} - V_T) - 2K\,dV_T$$

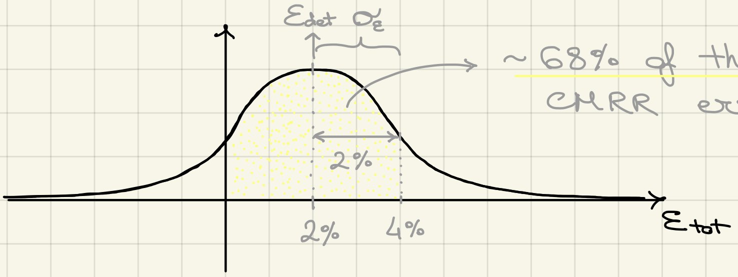
$$\frac{\Delta g_m}{g_m} = \frac{2dK(V_{GS}-V_T) - 2K\,dV_T}{2K(V_{GS}-V_T)} = \frac{dK}{K} - \frac{dV_T}{V_{ov}}$$

$$\Rightarrow \boxed{\sigma^2_{\frac{\Delta g_m}{g_m}} = \sigma^2_{\frac{\Delta K}{K}} + \sigma^2_{\Delta V_T}\frac{1}{V_{ov}^2}}$$

$$V_T = \overset{V_T^0}{0{,}6} \pm \overset{\sigma_{V_T}}{10\,mV}$$

$$K = \underset{K^0}{50\,\frac{\mu A}{V^2}} \pm \underset{\sigma_{\frac{\Delta K}{K}}}{10\%}$$

Adding $\mathcal{E}_{det}$ to $\mathcal{E}_{stat}$ means <u>shifting</u> the <u>gaussian</u> of the statistical error by an amount equal to the deterministic error.



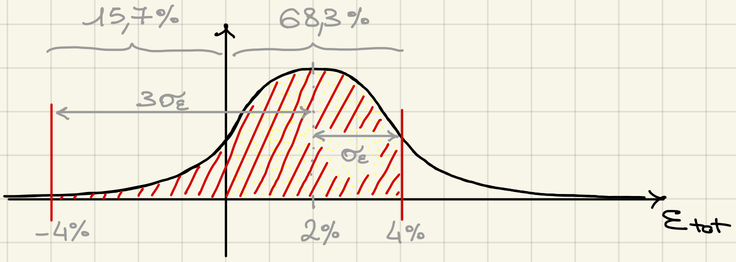
~68% of the samples (amplifiers) will have a CMRR error between 0% and 4%

Depending on the specifications on the CMRR requested by the user, the amplifier should match those specs by having an appropriate error spread.

E.g.: $|CMRR| > 80dB = 10^4$

$$\frac{\underbrace{2 g_{min} r_0}_{400}}{|\varepsilon_{tot}|} > 10^4$$

$$\Longrightarrow \quad |\varepsilon_{tot}| < 4\%$$



Roughly 84% of the samples will match the specification

In the end, how do we reduce the total CMRR mismatch?

As already said, the deterministic error can be reduced (to move the centroid of the error distribution around zero) by increasing the channel length of the transistors.

On the other hand, how can we reduce the statistical error (to narrow the error distribution)?
We need to quantify $\boxed{\sigma_{\frac{\Delta K}{K}}}$ and $\boxed{\sigma_{\Delta V_T}}$ to understand how they can be controlled (this is an important matter also for the computation of the amplifier offset).

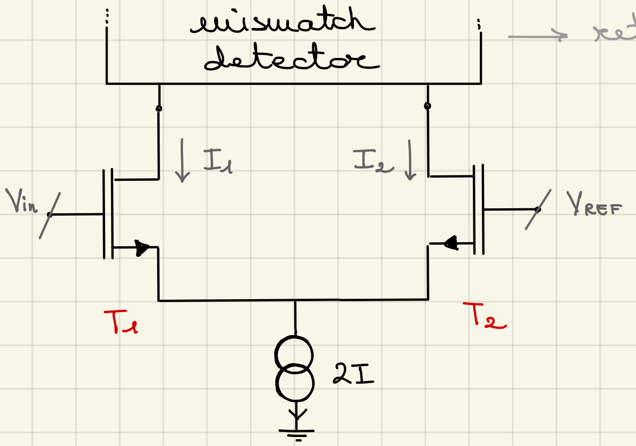
Note that $\frac{\Delta K}{K}$ and ΔV_T might be characterized by both a deteterministic term and a statistical term.

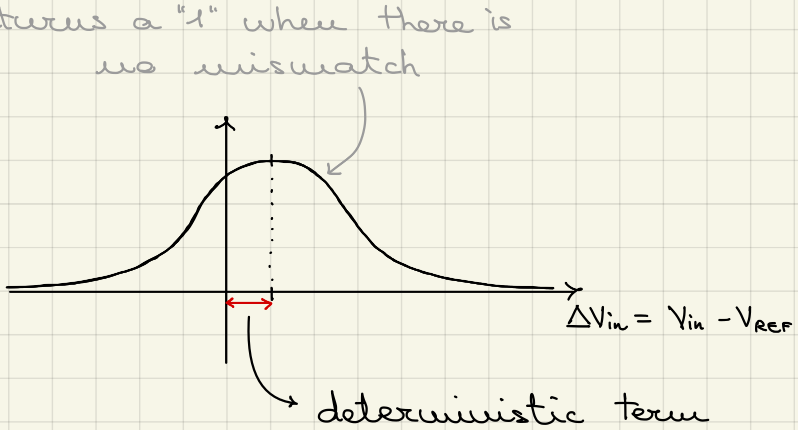
1. The ~~deterministic term~~ represents the offset of their gaussian distribution and is caused by a known, definite set of non-uniformities in the fabrication of the transistors.
2. The statistical term represents the spread of their distribution and is caused by random, unpredictable differences in the fabrication of many transistors.

1. We generally want the deterministic contribution of these mismatches to be as low as possible(#), since their causes

(#) this is why so far we assumed it to be nihil are known and their effects can be computed and compensated accordingly in advance.

Assume we wanted to measure the mismatch between the two input transistors:

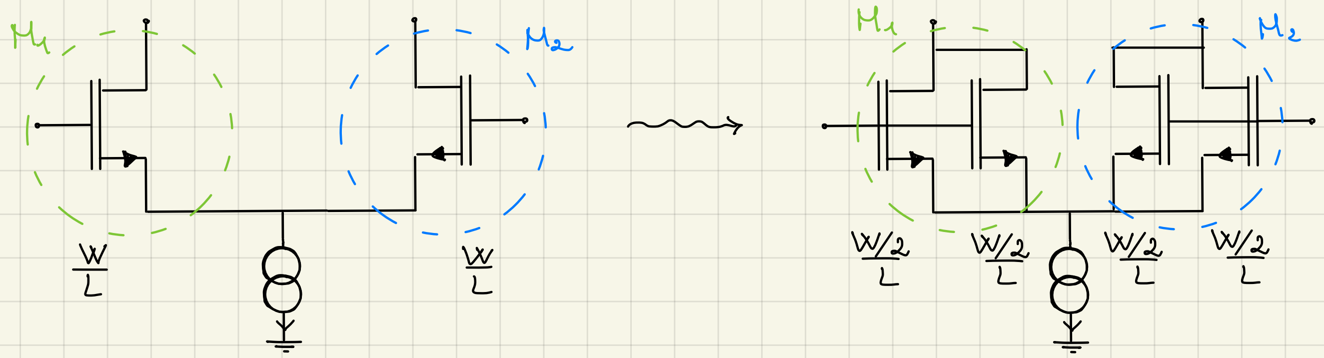


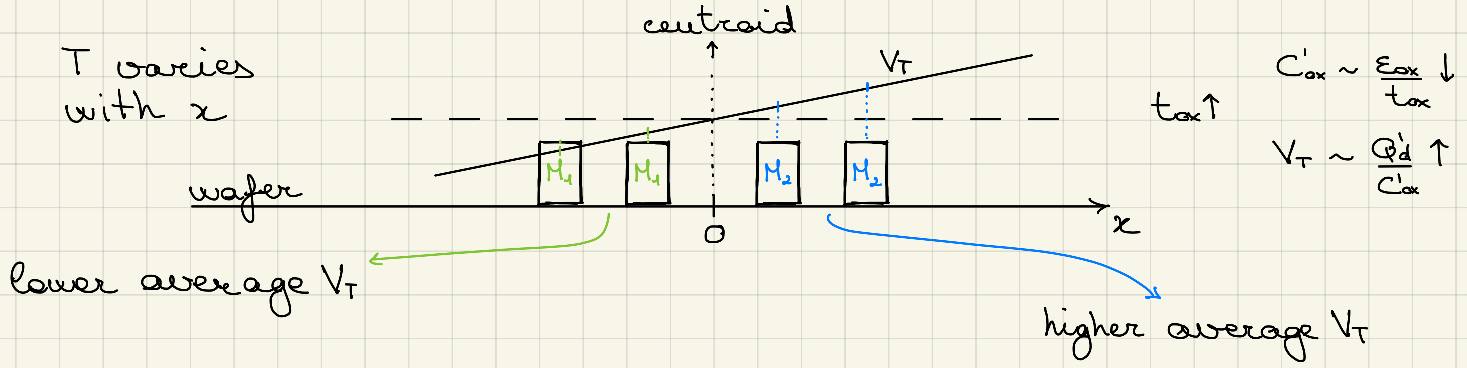


The deterministic term, which we want to eliminate, arises from deterministic differences in the process fabrication parameters, such as temperature. Temperature along a wafer is not uniform but tends to be higher in the inner part and lower in the outer part. This temperature difference determines a different growth rate of the oxide and therefore a different threshold voltage and K factor.

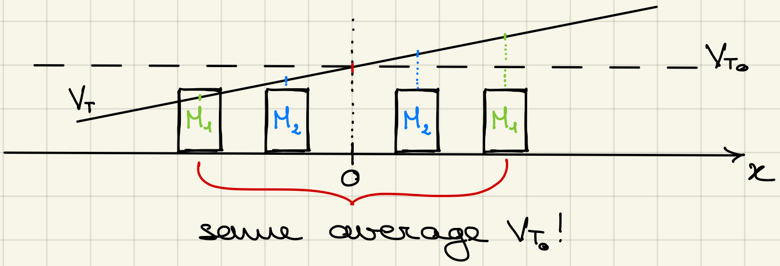
In order not to have this temperature difference during fabrication, the two transistors should be placed very close to each other, even better if inside one another.

How can two transistors be fabricated inside one another?





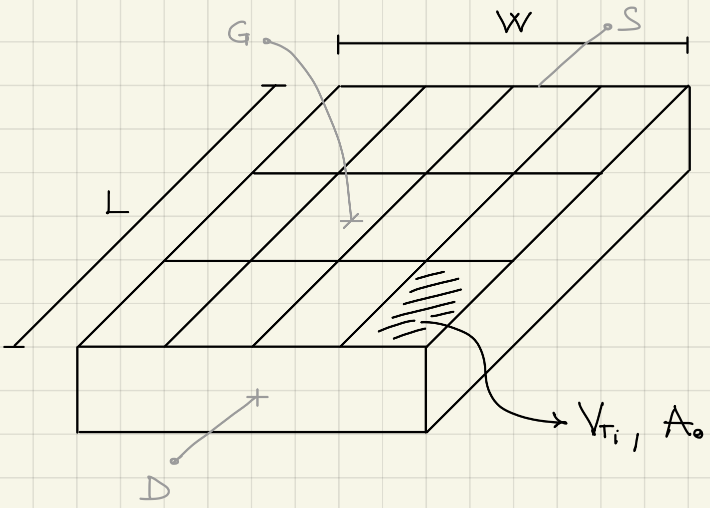
common centroid geometry



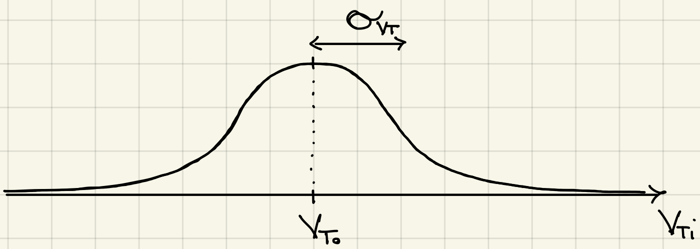
This expedient allows to cancel out any difference in the threshold voltage caused by non-uniform temperature, assuming that the size of each transistor is negligible with respect to the variation rate of the parameters.

On a further note, the variation rate of the parameter is not just the derivative along one direction of the wafer but rather a gradient along its entire surface. Therfore the common centroid technique should be applied with respect to both the x-axis and the y-axis of the wafer.

2. We now need to reduce the statistical contribution of ΔV_T (and $\Delta K/K$) to effectively narrow down the CMRR error distribution.



Imagine to split the transistor's surface into N smaller transistors, each with the same threshold voltage gaussian distribution, centered around a nominal value V_{T_0}:



We then compute the average threshold voltage and the variance of the entire transistor:

$$\bar{V}_{T_0} = \frac{\sum^{N} V_{Ti}}{N} \underset{\uparrow}{\simeq} \frac{N\, V_{T_0}}{N} = V_{T_0} \qquad \sigma^2_{\bar{V}_T} \underset{\uparrow}{=} \frac{\sum^{N} \sigma^2_{V_T}}{N^2} \simeq \frac{N\, \sigma^2_{V_T}}{N^2} = \frac{\sigma^2_{V_T}}{N}$$

for very high N

$$\sigma^2(\bar{V}_T) = \sigma^2\left(\frac{\sum V_{T_i}}{N}\right) = \frac{1}{N^2}\sigma^2\left(\sum V_{T_i}\right) = \frac{\sum \sigma^2(V_{T_i})}{N^2} = \frac{\sum \sigma^2_{V_T}}{N^2}$$

This shows that the variance of the entire transistor is smaller for a larger N

However we don't know how much is N nor $\sigma^2_{V_{T0}}$.
Nevertheless, it is obvious that a larger N requires a larger transistor surface.
If each smaller transistor has a fixed A_0 surface, then their number depends on how many of them can fit in the entire surface:

$$N = \frac{W\cdot L}{A_0}$$

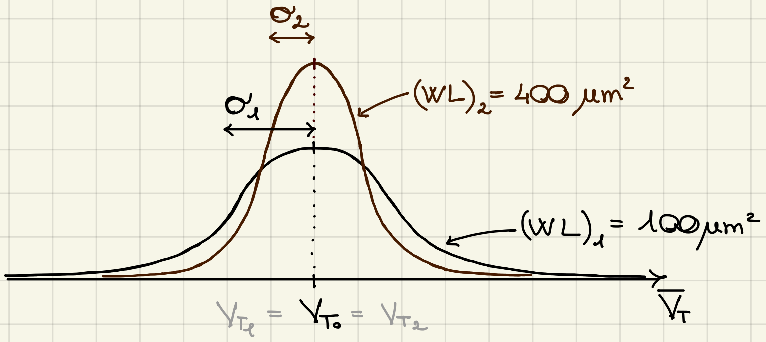
$$\Downarrow$$

$$\sigma^2_{\bar{V}_T} = \frac{\boxed{\sigma^2_{V_T}\cdot A_0}}{W\cdot L} = \frac{K^2_{V_T}}{W\cdot L}$$

With this result we can conclude that, if we consider a transistor with a given cross-section W·L, we would expect the spread of the average threshold value to be proportionally dependent on $1/\sqrt{W\cdot L}$

$$\sigma_{\bar{V}_T} = \sqrt{\sigma^2_{\bar{V}_T}} = \sqrt{\frac{K^2_{V_T}}{W\cdot L}} = \frac{K_{V_T}}{\sqrt{W\cdot L}}$$

This means that to a larger transistor corresponds a smaller variability of its parameters.



$$\frac{\sigma_2}{\sigma_1} = \frac{\sqrt{(WL)_1}}{\sqrt{(WL)_2}} = \frac{\sqrt{100}}{\sqrt{400}} = \frac{1}{2}$$

From a macroscopic point of view, increasing the transistor area is equivalent to adding many, tiny contributions whose parameters fluctuate with a certain spread; the more of these contributions, the better they can compensate each other with their own fluctuations, returning an overall spread of the device parameters that is lower than the "local" spread.

So to put everything together: if we were to look at the distribution of V_T out of many transistors (samples) of the same fabrication process, we would expect to see a negligible (thanks to the common centroid techique) deterministic shift, and a spread that decreases as the cross-section of the transistor increases.

<u>Note:</u> so far we have only considered $\sigma_{\bar{V}_T}$, but what we were initially interested in was actually $\sigma_{\Delta V_T}$

expected value = mean = center of the distribution

$$E(\Delta V_T) = (\underset{V_{T0}}{\underset{\downarrow}{V_{T1}}} - \underset{V_{T0}}{\underset{\downarrow}{V_{T2}}}) = 0$$

the two spreads are uncorrelated

typically ~4 mV·μm

$$\sigma^2(\Delta V_T) = \sigma^2_{V_{T1}} + \sigma^2_{V_{T2}} = 2\sigma^2_{\bar{V}_T} = \frac{2K^2_{V_T}}{W\cdot L} = \frac{(K_{\Delta V_T})^2}{W\cdot L}$$

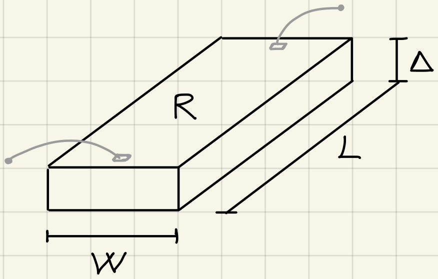
the two MOSFET should be equally sized

(⟶ There is just a factor $\sqrt{2}$ difference between $\sigma_{\bar{V}_T}$ and $\sigma_{\Delta V_T}$.)

This whole discussion can be repeated this time with respect to the conductivity parameter K.
We therefore need to find the: 1. ~~deterministic~~ and 2. <u>statistical</u> contribution of its relative variability $\frac{\Delta K}{K}$.

Since K gives a measure of the resistivity of the transistor channel, it is possible to compare the matching of the K parameter of two transistor with the matching of two resistors.
We will therefore consider resistor matching for now, and then apply the same argument to transistors.

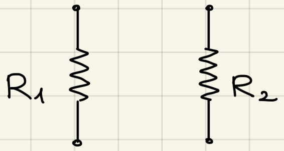
A resistor is a stripe of conductive layer that is characterized by a certain sheet resistivity $R_\square$ as well as a spread parameter $K_{\frac{\Delta R}{R}}$



$$R = \rho\cdot\frac{L}{\Delta\cdot W} = \frac{\rho}{\Delta}\cdot\frac{L}{W} = R_\square\cdot\frac{L}{W}$$

given by the process features

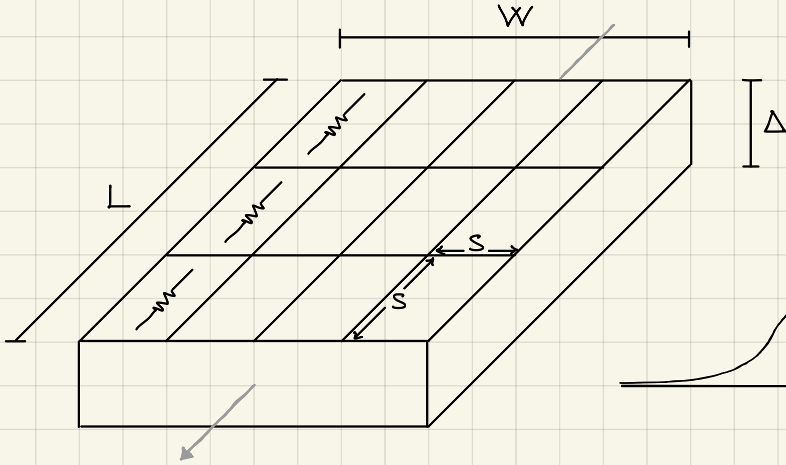
$$\Delta R = (R_1 - R_2) \qquad \left[\frac{\sigma_{\Delta R}}{R} = \frac{K_{\Delta R/R}}{\sqrt{W\cdot L}}\right]$$

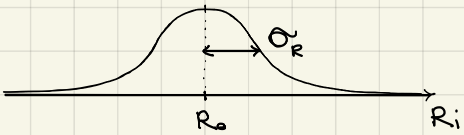


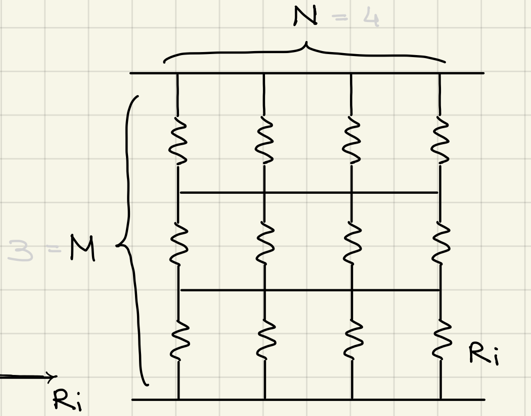
<u>Pelgrom's formula</u>

In addition to the statistical spread $\sigma_{\Delta R}$ there could also be a deterministic term affecting ΔR, which can be conveniently <u>cancelled out</u> through a common centroid geometry approach during fabrication.

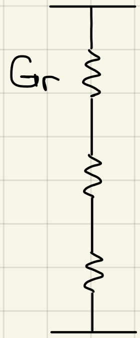
It is possible to derive Pelgrom's formula in the same way we previously computed $\sigma^2_{\bar{V}_T}$.







Each of the small resistors R_i is taken from a gaussian distribution with a nominal (mean) value R_0 and a spread (root mean square) σ_R.
Let's compute the total resistance mean value $\underset{\wr\, V_{T_0}}{R_{T_0}}$ and its spread $\underset{\wr\, \sigma_{V_T}}{\sigma_{R_T}}$.



We can consider each row independently (it is easier to use the conductance $G_0 = \frac{1}{R_0}$):

$G_{r_0} = \sum^{N} G_0 = N G_0$ average value

$\sigma^2_{G_r} = \sum^{N} \sigma_G^2 = N \sigma_G^2$ variance

$\hookrightarrow$ how much is this?

$$dG_0 = -\frac{dR_0}{R_0^2}$$

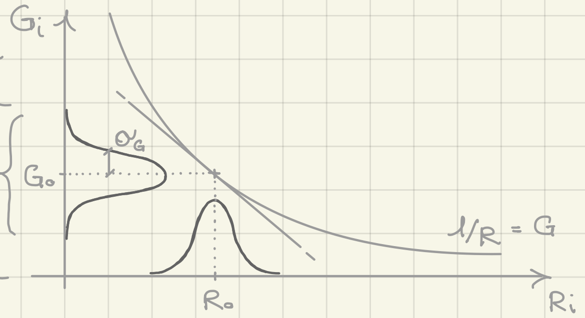
$$\frac{dG_0}{G_0} = -\frac{dR_0}{R_0^2} \cdot R_0 = -\frac{dR_0}{R_0}$$

$$\downarrow$$

$$\boxed{\frac{\sigma_G^2}{G_0^2} = \frac{\sigma_R^2}{R_0^2}}$$

given that σ_G is small compared to G_0

if σ_G is too large the tails of the gaussian will get distorted due to the non-linear relation



The total resistance is then the sum of the resistance of each row:

$$R_{T_0} = \sum^{M} R_{r_0} =$$
$$= M R_{r_0} =$$
$$= M \cdot \frac{1}{G_{r_0}} =$$
$$= \frac{M}{N} \frac{1}{G_0} = \frac{M}{N} R_0$$

$$\sigma^2_{R_T} = \sum^{M} \sigma^2_{R_r} = M \sigma^2_{R_r}$$

$$\longrightarrow \frac{\sigma^2_{R_T}}{R^2_{T_0}} = \frac{M \sigma^2_{R_r}}{M^2 R^2_{r_0}} = \frac{1}{M} \frac{\sigma^2_{R_r}}{R^2_{r_0}} = \frac{1}{M} \frac{\sigma^2_{G_r}}{G^2_{r_0}} = \frac{1}{M} \frac{N \sigma_G^2}{N^2 G_0^2} = \frac{1}{M \cdot N} \frac{\sigma_G^2}{G_0^2} = \frac{1}{MN} \frac{\sigma_R^2}{R_0^2}$$

$$N = \frac{W}{S}, \quad M = \frac{L}{S} \longrightarrow MN = \frac{WL}{S^2} \Longrightarrow \frac{\sigma^2_{R_T}}{R^2_{T_0}} = \frac{S^2 \sigma_R^2}{(WL) R_0^2} = \frac{K^2}{WL}$$

$$\frac{\Delta R_T}{R_T} = \sqrt{\frac{\sigma_{R_T}^2}{R_{T_0}^2}} = \sqrt{\frac{K^2}{WL}} = \frac{K}{\sqrt{W \cdot L}}$$ associated to a single resistor

$$R_{1_0} = R_{2_0} = R_{T_0}$$

$$\Delta R = \overset{\uparrow}{R_1} - \overset{\uparrow}{R_2} = 0 \qquad \frac{\sigma_{R_1}^2}{R_1^2} = \left(\frac{K}{\sqrt{(WL)_1}}\right)^2 = \frac{\sigma_{R_2}^2}{R_2^2} = \left(\frac{K}{\sqrt{(WL)_2}}\right)^2$$

$$\sigma^2(\Delta R) = \sigma^2(R_1) + \sigma^2(R_2)$$

$$= 2\sigma^2(R) = \frac{2K^2R^2}{(WL)}$$

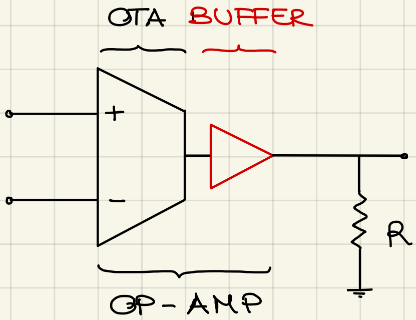
$$\left[\frac{\Delta R}{R} = \sqrt{\frac{\sigma^2(\Delta R)}{R^2}} = \sqrt{\frac{2K^2}{WL}} = \frac{\sqrt{2}\,K}{\sqrt{WL}} = \frac{K_{\Delta R/R}}{\sqrt{W \cdot L}}\right]$$ associated to the difference between two resistors

This formula can now be applied to the ΔK between two transistors:

$$\boxed{\frac{\Delta K}{K} = \sigma_{\frac{\Delta K}{K}} = \frac{K_{\Delta K/K}}{\sqrt{W \cdot L}}}$$

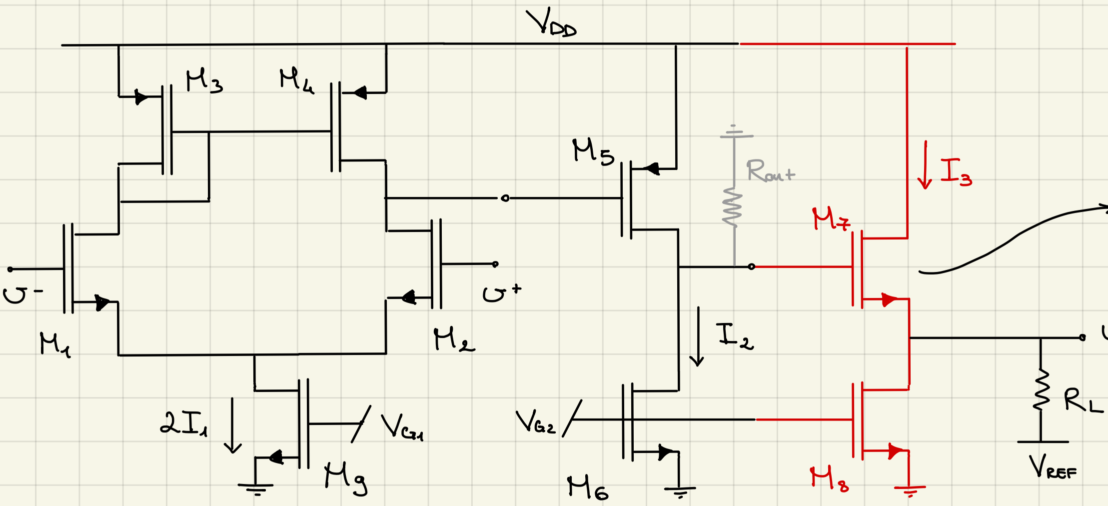
$$\boxed{\Delta V_T = \sigma_{\Delta V_T} = \frac{K_{\Delta V_T}}{\sqrt{W \cdot L}}}$$

Output Stages



In order to build an operational amplifier we need a buffer stage at the output to avoid a reduction of the gain due to the (low) load resistance directly linked to a high impedance node of the OTA.

A basic buffer configuration can be for example a source-follower stage:



it decouples the low load resistance R_L from the high output impedance R_{out} of the OTA

class A buffer

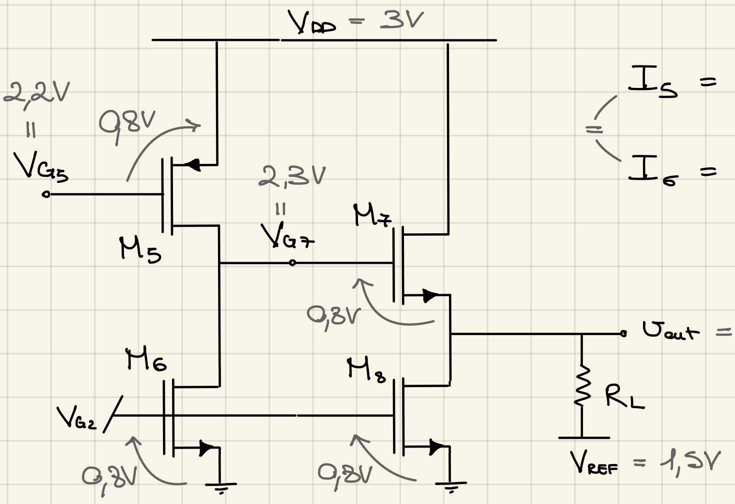
With the added output buffer stage, we will now point out what is its distortion contribution to the output signal and how it affects the efficiency of the amplifier, after having properly set the bias of the stage

↓

the output should be biased so that it sits at mid-range

e.g.: $V_{DD} = 3V \longrightarrow v_{out} = 1{,}5V \longrightarrow V_{REF} = 1{,}5V$

however this will move the output of the OTA at a higher voltage (it was previously at mid-range).



$V_{DD} - V_{GS7} - V_{REF} = 0{,}7V$

$$I_5 = K_5 V_{ov5}^2 \left[1 + \frac{V_{SD5} - V_{ov5}}{V_{A5}}\right]$$

$$=$$

$$I_6 = K_6 V_{ov6}^2 \left[1 + \frac{V_{DS6} - V_{ov6}}{V_{A6}}\right]$$

$V_{GS7} + V_{REF} = 2{,}3V$

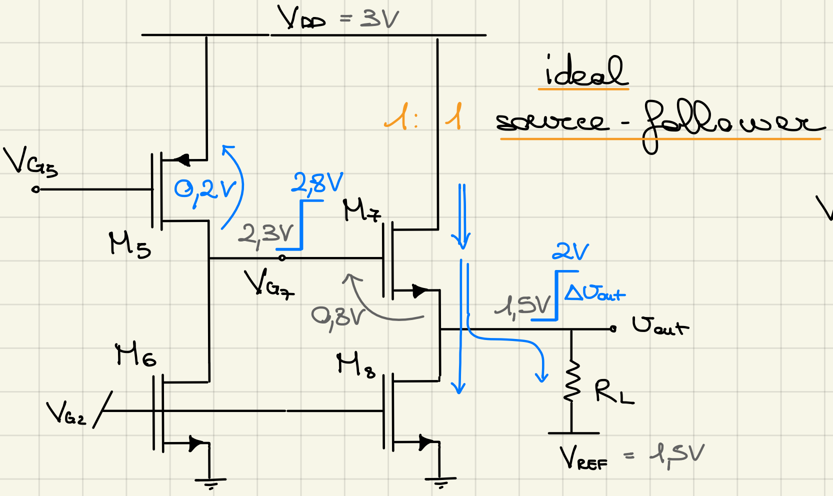
We need to set $V_{G7} = 2{,}3V$ to have $v_{out} = 1{,}5V$.

$$\Longrightarrow \overset{(W/L)_5}{K_5}\, \cancel{V_{ov}^2} \left[1 + \frac{0{,}5V}{V_{A5}}\right] = \overset{(W/L)_6}{K_6}\, \cancel{V_{ov6}^2} \left[1 + \frac{2{,}1V}{V_{A6}}\right]$$

$$W_6 = \frac{L_6}{L_5} W_5 \frac{\left[1 + \frac{0{,}5V}{V_A}\right]}{\left[1 + \frac{2{,}1V}{V_A}\right]}$$ to avoid systematic offset.

Let's now study the output swing:

positive swing

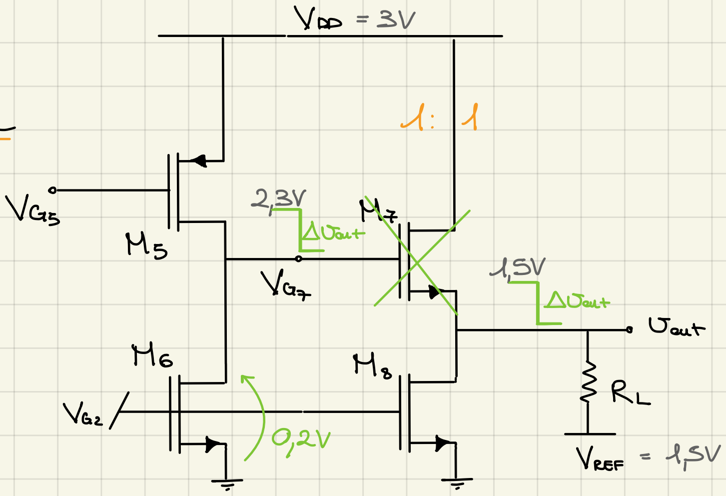


$$I_7 = I_8 + \frac{\Delta V_{out}}{R_L}$$

$$\Delta V_{out} = 0{,}5V$$

Limited by
M_5 entering ohmic

negative swing



There are 2 limits for the negative output swing:

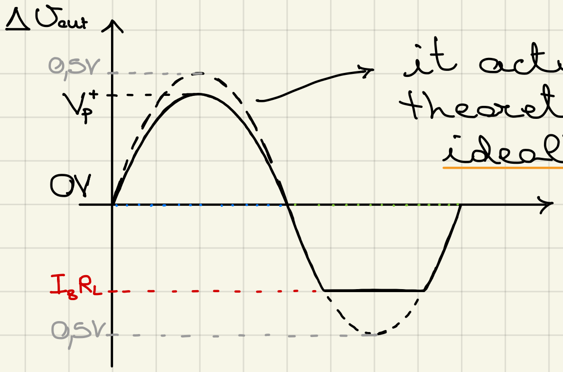
M_7 turning off OR M_6 entering ohmic

M_7 turning off ↓ $\frac{\Delta V_{out}}{R_L} = I_8$ ⇓ typically this is the most restrictive condition

M_6 entering ohmic ↓ $\Delta V_{out} = 2{,}1V$ $(V_{G7} = 0{,}2V)$

Note that to have a symmetric output swing I_8 must match a precise value that is dependent on R_L
e.g.: $R_L = 0{,}5K\Omega$, $\Delta V_{out}^+ = 0{,}5V \Longrightarrow I_8 = \frac{\Delta V_{out}}{R_L} = \frac{\Delta V_{out}^+}{R_L} = 1mA$

If the condition is not reached then the full swing output signal will be clamped at the negative end!



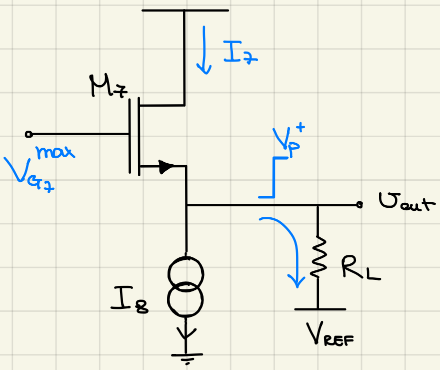
it actually does NOT reach the full theoretical swing because of other non-idealities (real source-follower, distortions etc.)

↓

Let's compute what is the real peak value V_p^+

The two main reasons of the reduced peak value are:

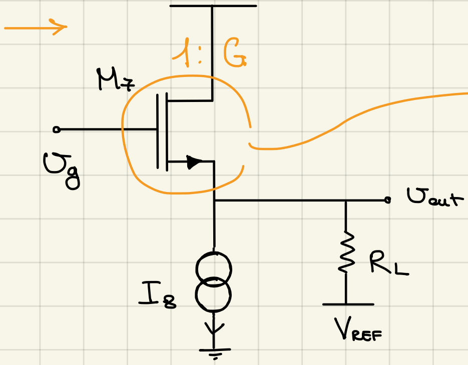
- non-1:1 transfer of the buffer
- non-linear characteristic of the transistor

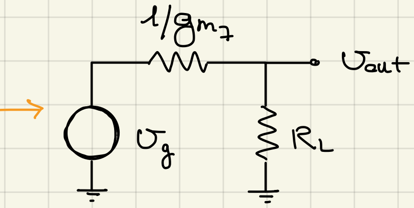


→ $I_7 = K_7\left[V_{G7}^{max} - V_{REF} - V_p^+ - V_{T7}\right]^2 = I_8 + \frac{V_p^+}{R_L}$

$$a V_p^{+2} + b V_p^+ + c = 0 \rightarrow V_p^+ < 0{,}5V$$

→

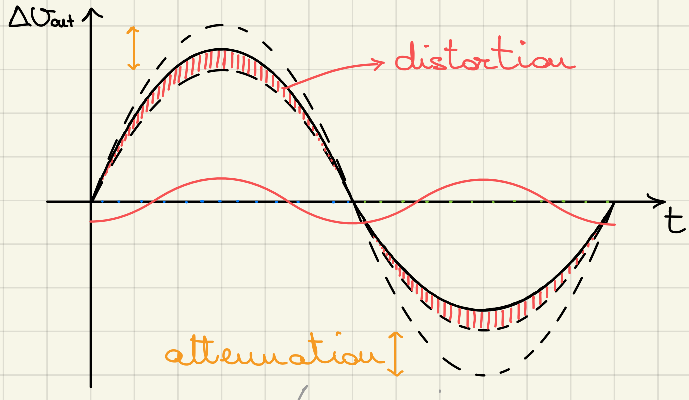




$$G = \frac{R_L}{1/g_{m7} + R_L} = \frac{g_{m7}R_L}{1 + g_{m7}R_L} \leq 1$$

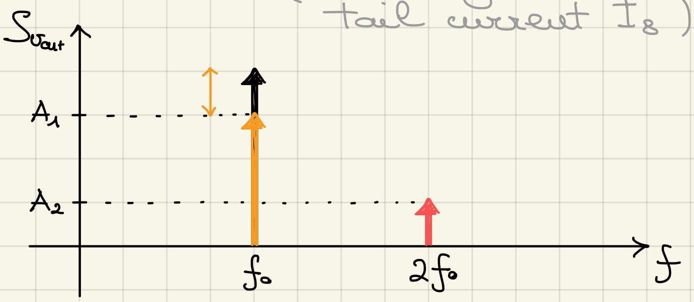
$$\rightarrow V_p(v_{out}) \leq V_p(v_g)$$

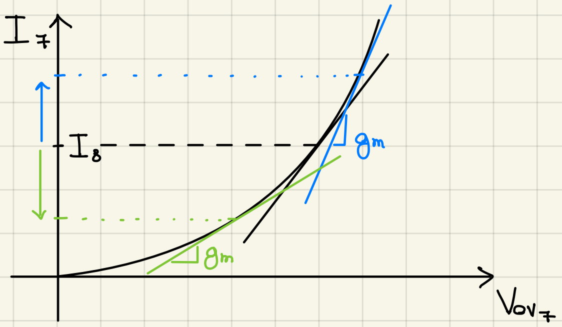
Note that while the real source-follower causes an attenuation of the output signal (amplitude reduction of all of its spectral components), the non-linear characteristic of the transconductance causes a distortion (addition of spectral components that are not present at the input).



(assuming an appropriate tail current I_8)

→ When the output increases, more current flows through M_7 therefore its transconductance slightly increases.
Viceversa, when the output decreases g_{m7} decreases.
This translates into a distorted waveform.





$$\boxed{HD_2 = \frac{A_2}{A_1}}$$ second harmonic distortion

Note how increasing g_{m7} (higher I_8, more power consumption) will benefit both the distortion and attenuation of the output signal.

Let's now compute the power efficiency η of the stage.

$$\eta_{max} = \frac{P_L}{P_{DC}}$$

power delivered to the load

power dissipated across the stage

$$= \frac{V_p^2/2R_L}{V_{DD} \cdot I_8}$$

average current across the stage

$$= \frac{V_p^2}{2R_L V_{DD} I_8}$$

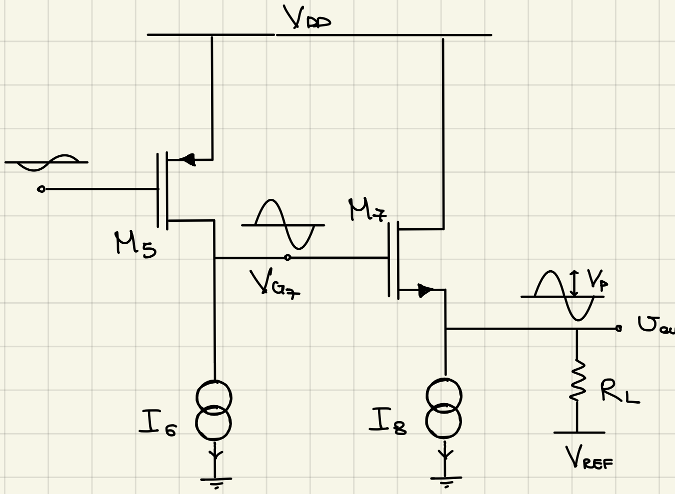
$R_L I_8 \geq V_p$

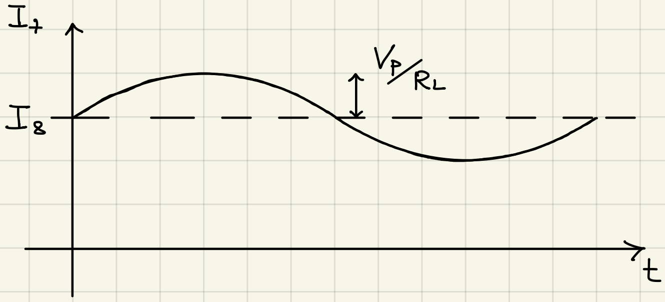
$$\leq \frac{V_p^2}{2 V_p V_{DD}}$$

$V_p < \frac{V_{DD}}{2}$

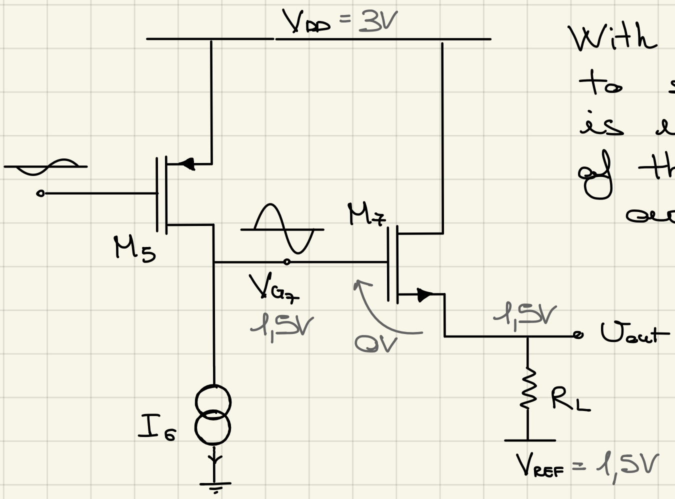
$$\leq \frac{V_p^2}{4 V_p^2}$$

$$\Longrightarrow \boxed{\eta_{max} \leq \frac{1}{4} = 25\%}$$ poor !

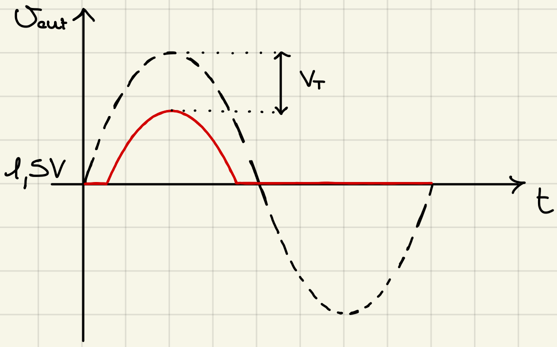




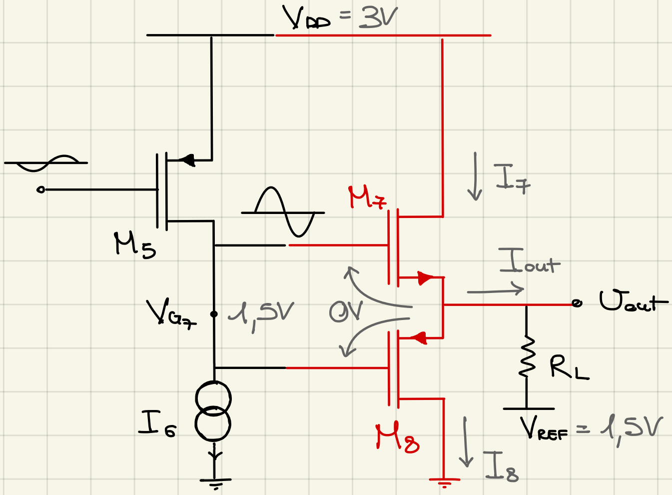
We need another architecture to improve power efficiency



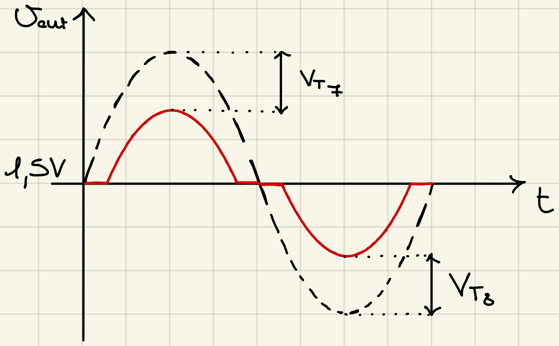
With this configuration there is no need to supply the buffer stage so there is no impairing of the power efficiency of the whole amplifier, however the output waveform is heavily altered.



→ Add a pMOS to complement the negative swing transition



class B (push-pull) buffer



The distortion around crossover is still present though.

Let's first try to understand what type of distortion we are dealing with.

$$I_7 = A_0 + A_1 \sin(\omega_0 t + \varphi_1) + A_2 \sin(2\omega_0 t + \varphi_2) + A_3 \sin(3\omega_0 t + \varphi_3) + \dots$$

I_8 is equivalent to I_7 shifted by $\frac{T}{2}$ (given M_7 and M_8 have the same parameters):

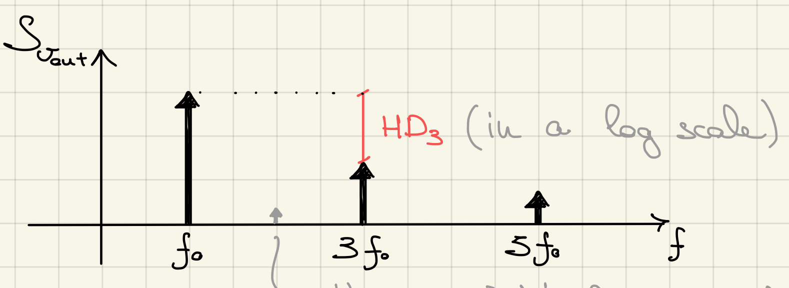
$$I_8 = A_0 + A_1 \sin\left[\omega_0\left(t - \frac{T}{2}\right) + \varphi_1\right] + A_2 \sin\left[2\omega_0\left(t - \frac{T}{2}\right) + \varphi_2\right] + $$
$$+ A_3 \sin\left[3\omega_0\left(t - \frac{T}{2}\right) + \varphi_3\right] + \dots$$

$$\left(\omega_0 \frac{T}{2} = \frac{2\pi}{T} \cdot \frac{T}{2} = \pi, \quad \sin(\varphi + \pi) = -\sin(\varphi)\right)$$

$$= A_0 - A_1 \sin[\omega_0 t + \varphi_1] + A_2 \sin[2\omega_0 t + \varphi_2] - A_3 \sin[3\omega_0 t + \varphi_3] + \dots$$

$$I_{out} = I_7 - I_8 = 2A_1 \sin(\omega_0 t + \varphi_1) + 2A_3 \sin(3\omega_0 t + \varphi_3) + \dots$$

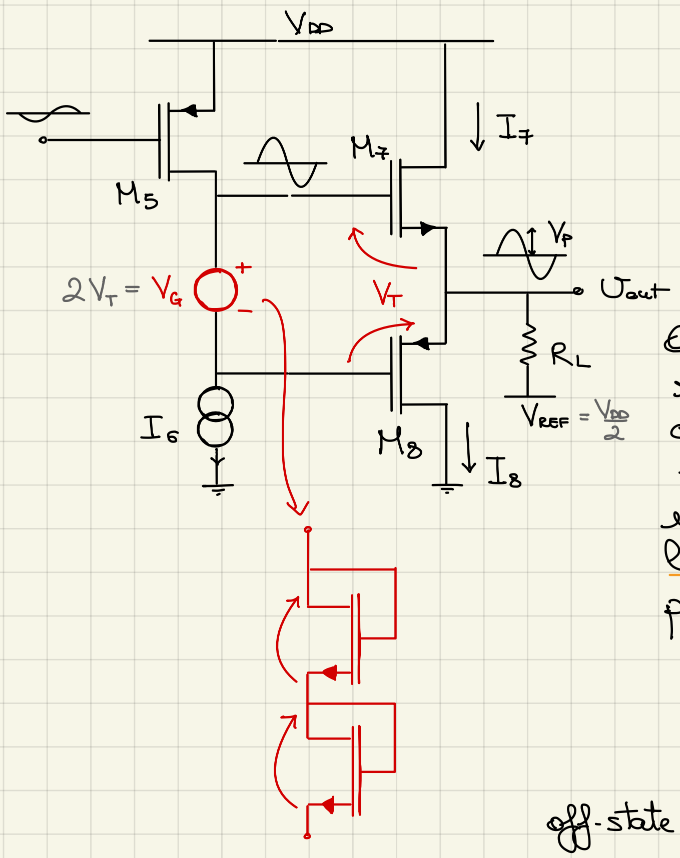
$\Longrightarrow$ All odd harmonics are maintained



there might be some further even harmonic in case of transistors mismatch

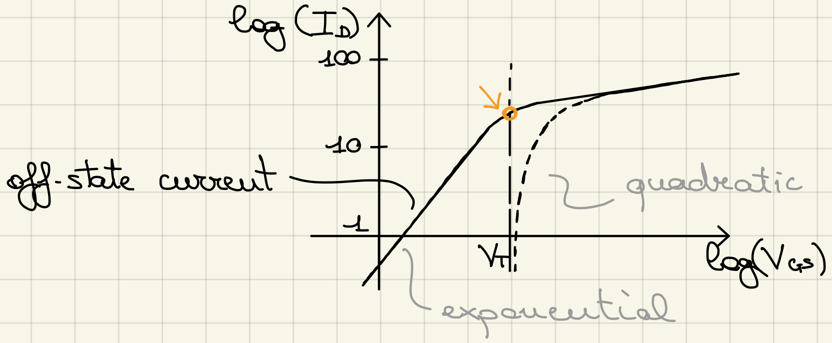
The distortion arises from the fact that in bias condition the output transistors are left with zero driving voltage ($V_{GS_7} = V_{SG_8} = 0V$). When a signal is applied at the input, the gate of the two transistors must first rise above threshold before the output can move. This solution guarantees no current consumption of to the buffer stage but as we've seen it impairs the output frequency spectrum.

An idea to fix the distortion caused by the stage would then be to fix the driving voltage bias of the output transistors exactly at threshold, so that there is no "dead zone" during which they need to turn on.



We have already seen that a voltage shifter with virtually no resistance can be obtained through MOSFETs in transdiode configuration.

Of course this expedient to reduce distortion comes with a cost: having the transistors of the buffer stage biased close to threshold means that there will be some leakage current which will cause power dissipation.



→ Trade-off between distortion and power efficiency

Let's compute the power efficiency of this stage.

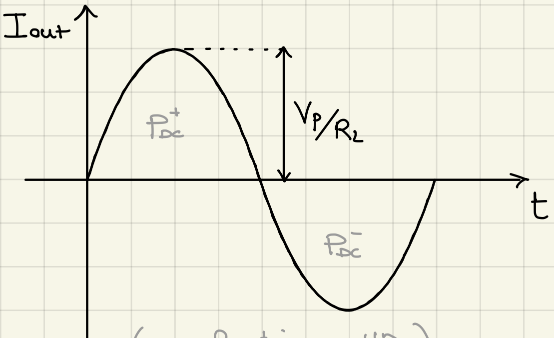
$$\eta = \frac{P_L}{P_{DC}} = \frac{V_P^2/2R_L}{\frac{P_{DC}^+ + P_{DC}^-}{2}}$$

$$P_{DC}^+ = P_{DC}^- = (V_{DD} - V_{REF}) \cdot \bar{I}_{out} = \frac{V_{DD}}{2} \bar{I}_{out}$$

$$\bar{I}_{out} = \frac{2}{T} \int_0^{T/2} \frac{V_P}{R_L} \sin(\omega t)\, dt =$$

$$= \frac{2}{T} \frac{T}{2\pi} \frac{V_P}{R_L} \int_0^{T/2} \sin(\omega t)\, dt \cdot \frac{2\pi}{T} =$$

$\omega t = \theta$ $\quad dt \cdot \frac{2\pi}{T} = dt \cdot \omega = d\theta$ $\quad t = \frac{T}{2} \rightarrow \theta = \pi$



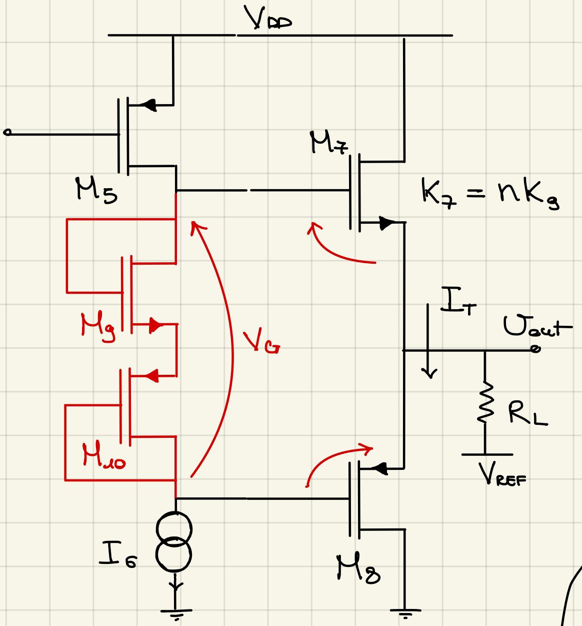
(neglecting HD_2)

$$= \frac{1}{\pi}\frac{V_P}{R_L}\int_0^{\pi}\sin\vartheta\, d\vartheta = \frac{2}{\pi}\frac{V_P}{R_L} = \bar{I}_{out}$$

$$P_{DC}^{+} = \frac{V_{DD}}{2}\bar{I}_{out} = \frac{V_{DD}\cdot V_P}{\pi R_L}$$

$$\Longrightarrow \boxed{\eta_{max} = \frac{V_P^2/2R_L}{P_{DC}^{+}} = \frac{V_P^2}{2R_L}\cdot\frac{\pi R_L}{V_{DD}\cdot V_P} = \frac{\pi}{2}\frac{V_P}{V_{DD}} \overset{\overset{V_P \le \frac{V_{DD}}{2}}{\downarrow}}{\le} \frac{\pi}{4} \simeq 78\%}$$ great!

The efficiency of the push-pull stage is roughly 3x times better than the source-follower stage!



class A-B stage

As we have already discussed, this solution greatly benefits the distortion of the output stage at the cost of some power dissipation, which is due to the off-state current of M_7 and M_8.

Since the off-state current of a transistor varies exponentially with its sub-threshold V_{GS}, it is relevant to accurately set V_G in order for I_T not to differ too much from its estimated value (which translates to a different value of power efficiency).

For this reason it is important to use one pMOS and one nMOS to fix V_G (instead of two nMOS or two pMOS):

$$V_G = V_{GS_7} + V_{SG_8} = V_{GS_9} + V_{SG_{10}}$$

$$\underset{\underset{V_{Tn}}{\downarrow}}{\cancel{V_{T_7}}} + \sqrt{\frac{I_T}{K_7}} + \underset{\underset{V_{Tp}}{\downarrow}}{\cancel{V_{T_8}}} + \sqrt{\frac{I_T}{K_8}} = \underset{\underset{V_{Tn}}{\downarrow}}{\cancel{V_{T_9}}} + \sqrt{\frac{I_6}{K_9}} + \underset{\underset{V_{Tp}}{\downarrow}}{\cancel{V_{T_{10}}}} + \sqrt{\frac{I_6}{K_{10}}}$$

The threshold voltage of p-type and n-type transistors are typically slightly different. Using one pMOS and one nMOS for the voltage shifter allows to neglect this mismatch when computing the value of I_T.

$$\sqrt{\frac{I_T}{I_6}} = \frac{\frac{1}{\sqrt{K_9}} + \frac{1}{\sqrt{K_{10}}}}{\frac{1}{\sqrt{K_7}} + \frac{1}{\sqrt{K_8}}}$$

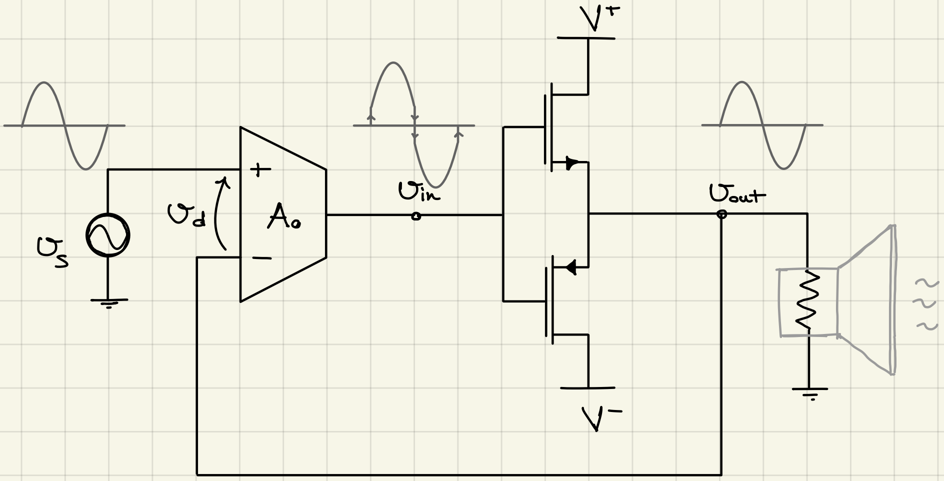
$$\frac{I_T}{I_6} = \left[\frac{\frac{1}{\sqrt{K_9}} + \frac{1}{\sqrt{K_{10}}}}{\frac{1}{\sqrt{K_7}} + \frac{1}{\sqrt{K_8}}}\right]^2 = \frac{\frac{1}{K_9}}{\frac{1}{K_7}}\left[\frac{1+\sqrt{\frac{K_9}{K_{10}}}}{1+\sqrt{\frac{K_7}{K_8}}}\right]^2 = \frac{K_7}{K_9}\left[\frac{1+\sqrt{\frac{K_9}{K_{10}}}}{1+\sqrt{\frac{K_7}{K_8}}}\right]^2 = n\left[\frac{1+\sqrt{\frac{K_9}{K_{10}}}}{1+\sqrt{\frac{K_7}{K_8}}}\right]^2$$

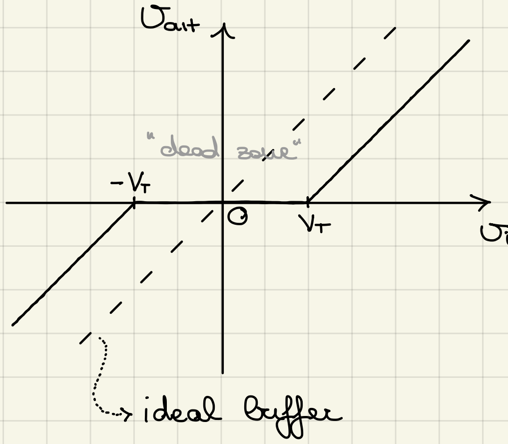
The current of the buffer I_T is proportional to the form

factor ratio n between the buffer transistors and the transdiode transistors.
In order not to have a too high current the transdiodes must be therefore sufficiently large to have a lower n.

$n \uparrow \quad I_T \uparrow \quad HD \downarrow \quad \eta \downarrow$

Negative feedback effects on distortion



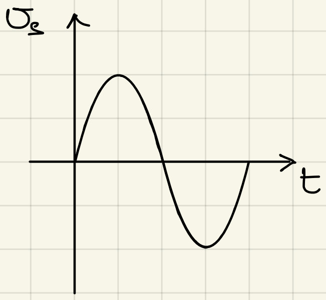


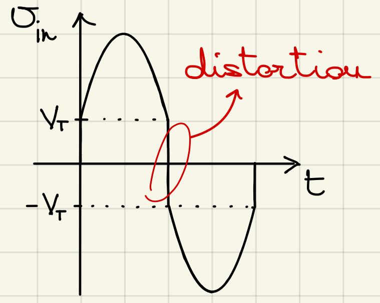
Assume: $G_{loop} \to \infty \Leftrightarrow A_0 \to \infty \Rightarrow v_d \to 0 \Rightarrow v_{out} \to v_s$

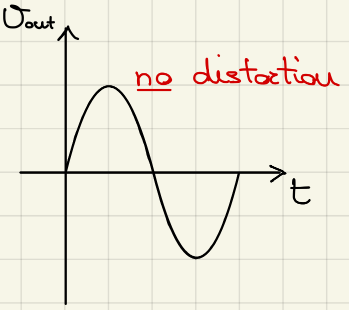
even if the signal is distorted!

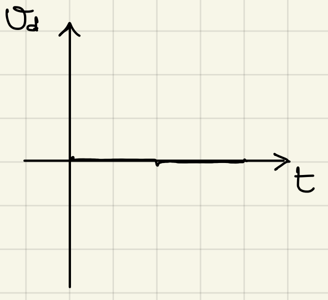
How can the feedback deal with distortion?

Thanks to the (ideally) infinite gain of the OTA, any non-zero signal at its input (v_d) will cause its output (v_{in}) to clamp at maximum voltage (V^+ or V^-). During the initial transition, when $0 < v_s < V_T$ but $v_{out} = 0$ because of the "dead zone" of the non-ideal buffer, v_d is momentarily non-null therefore v_{in} skips to a value such that $v_{out} = v_s$ and therefore $v_d = 0$.
This means that v_{out} will always be following v_s without (ideally) any distortion, while in turn v_{in} will be the one distorted to compensate the further distortion introduced by the buffer!



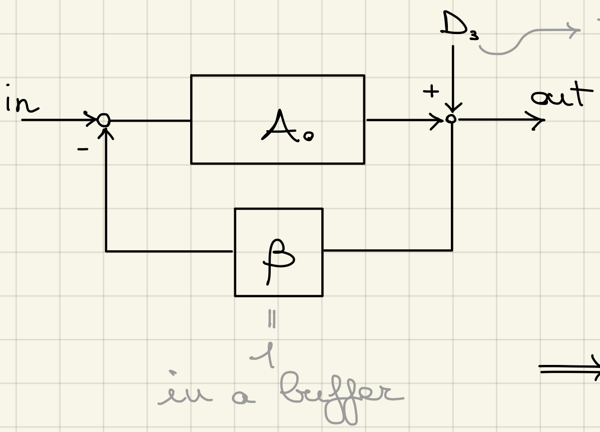






→ The non-linearity of the buffer stage is cancelled out by pre-destorting the signal driving the stage

To better understand this concept:



the distortion caused by the push-pull buffer is a third (odd) harmonic distortion

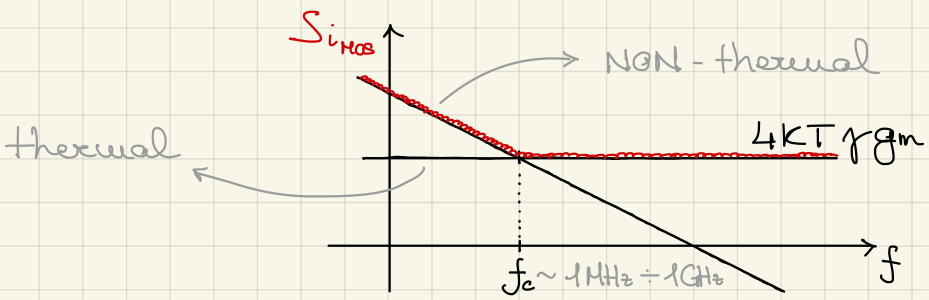
$$D_3^{out} = D_3 - \boxed{\beta A_0 D_3^{out}} \rightarrow$$ compensating distortion term generated by the loop

$$D_3^{out}(1+\beta A_0) = D_3$$

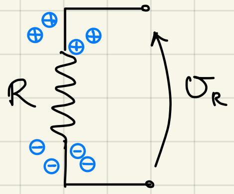
$$\Longrightarrow D_3^{out} = \frac{D_3}{\underbrace{1+\beta A_0}_{\to +\infty}} \to 0$$

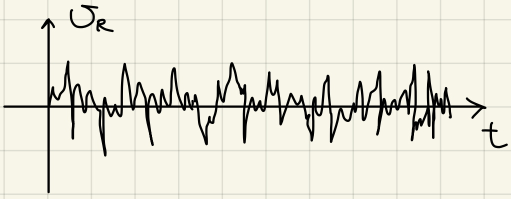
|| Noise Models ||

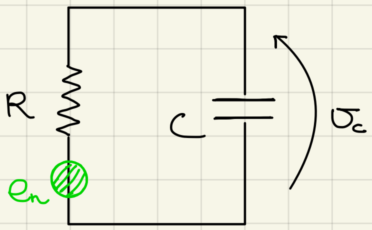
So far we have only considered the presence of thermal noise in electronic circuits. However there exist more types of electronic noise, especially regarding transistors, that are very relevant due to their murkier origins and their frequency behaviour which is not necessarily constant (white noise) but instead varies with frequency.
They are therefore harder to deal with and require a deeper understanding.



Let's first revisit thermal noise.







$$\boxed{S_{v_R}(f) = 4KTR}$$

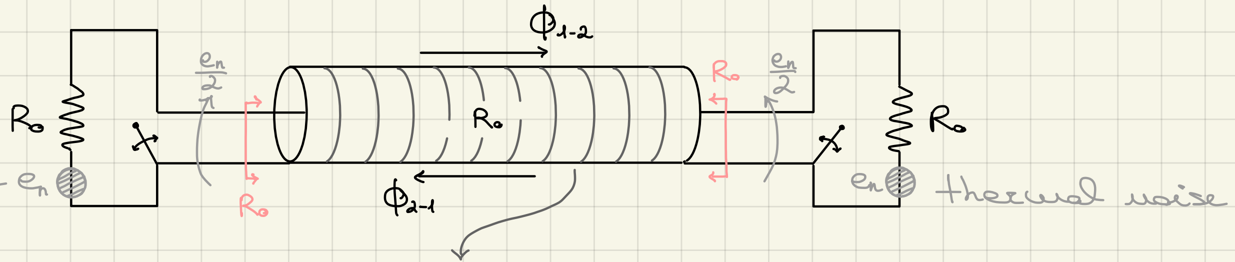
Boltzmann's law: $\mathcal{E}_c(v_c) = KT \times \frac{\text{degrees of freedom}}{2}$

$$\frac{1}{2}C\langle v_c^2\rangle = KT\cdot\frac{1}{2}$$

$$S_{v_R}\cdot\frac{1}{4RC} = \int_0^{+\infty}\frac{S_{v_R}\,df}{1+(\omega RC)^2} = \frac{S_{v_R}\cdot\Delta f}{1+(\omega RC)^2} = \frac{e_n^2}{|1+j\omega RC|^2} = \langle v_c^2\rangle = \frac{KT}{C}$$

train of deltas over t
i.e. constant over f

Nyquist demonstration for thermal noise of a resistor:



r.m.s. value: $[e_n] = V_{r.m.s}$

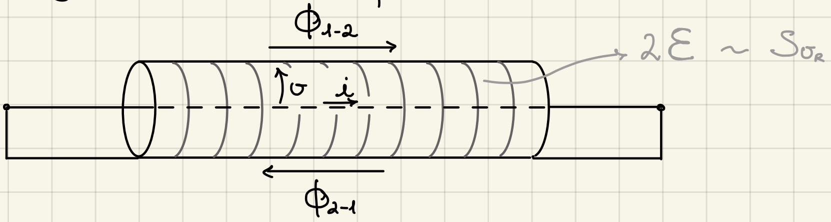
coaxial cable with:

$$Z_c = \frac{\vec{V}}{\vec{I}} = \sqrt{\frac{L}{C}} = R_0 \text{ characteristic impedance}$$

↳ adapted load (i.e. no reflections of $\vec{V}$ and $\vec{I}$ and Φ)

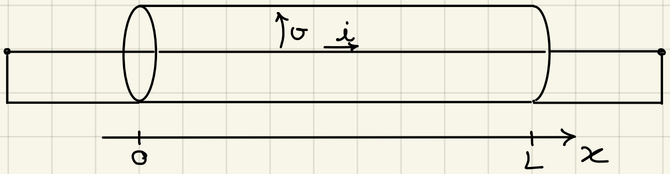
Φ is the energy flux generated by v_n travelling across the transmission line (electromagnetic - tension/current wave).

At a certain instant the two switches are closed thus isolating the three parts of the circuit.



The energy generated by the thermal noise of the two resistors, with all its spectral components, is trapped within the coaxial cable.

If we get to know how much is the energy contained within the coaxial cable we will then know the energy related to its source in the first place, that is the PSD of the thermal noise of the two resistors.



The tension and current in the cable must abide the wave equation (as well as Maxwell's equations).

$$\frac{\partial^2 v(x,t)}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 v(x,t)}{\partial t^2} \qquad \frac{\partial^2 i(x,t)}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 i(x,t)}{\partial t^2}$$

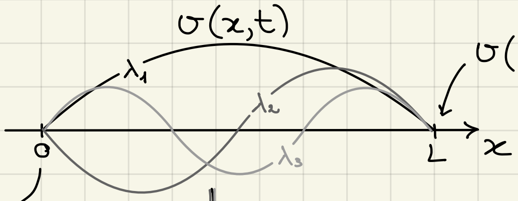
$$\Longrightarrow v(x,t) = \sin(\omega t + \varphi)\left[A\sin(Kx) + B\cos(Kx)\right]$$

$$\omega = Kc$$

$$\frac{2\pi}{f} = \frac{2\pi}{\lambda}c$$

$$\underline{c = \lambda f} \quad (\underline{\text{dispersion equation}})$$

By shorting the ends of the cable we are forcing some boundary conditions (null tension at $x=0$ and $x=L$):



$v(L,t)=0 \Longrightarrow KL = \pi + n\pi$

$B=0 \Longleftarrow v(0,t)=0$

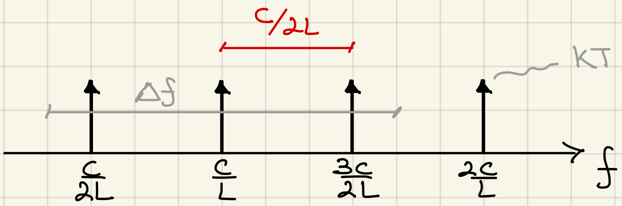
$\downarrow$

fundamental solutions:

$$KL = \pi + n\pi \longrightarrow \frac{2\pi}{\lambda}\cdot L = \pi + n\pi \Longrightarrow \lambda = \frac{2L}{1+n}$$

$$\lambda_1 = 2L \qquad \lambda_2 = L \qquad \lambda_3 = \frac{2L}{3} \qquad \lambda_4 = \frac{L}{2} \quad \dots$$

→ discrete number of electromagnetic waves that exist inside the coaxial cable



How much is the energy that each of these EM waves has at equilibrium?

→ Use Boltzmann's law

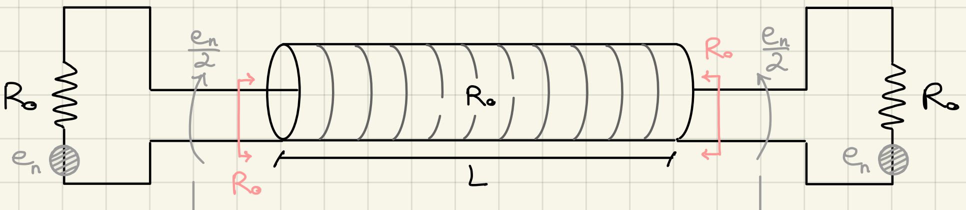
2 degrees of freedom

$$\mathcal{E}_{EM} = \mathcal{E}_{EM}(\vec{E},\vec{H}) = KT \cdot \frac{2}{2} = KT$$

Energy per frequency interval: $\mathcal{E}_{\Delta f} = KT \cdot \frac{\Delta f}{c/2L}$

From this derives that the PSD is constant over all the frequency spectrum (hence we're dealing with white noise) because the number of EM modes per frequency range is everywhere the same, and they all carry the same energy.

Now, how much was the power we injected into the cable before closing the switch?



$$P = \left(\frac{e_n}{2}\right)^2 \frac{1}{R_0}$$

$$\mathcal{E}_{tot} = 2P \cdot \Delta t = 2P \cdot \frac{L}{c}$$

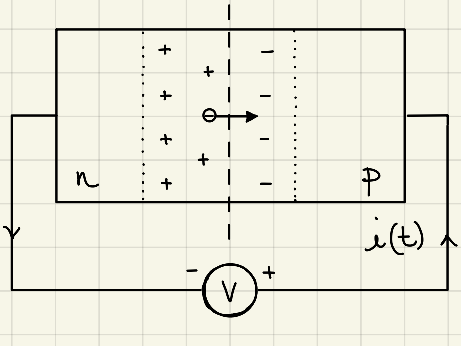
Remember that e_n is the r.m.s. value: $S_{v_R} = \frac{e_n^2}{\Delta f}$

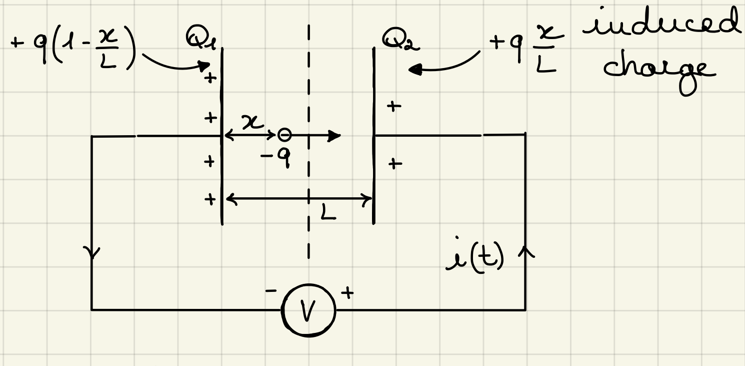
$$\Longrightarrow \quad \mathcal{E}_{tot} = \cancel{2} \cdot \frac{S_{v_R} \cdot \cancel{\Delta f}}{4} \cdot \frac{1}{R_0} \cdot \frac{\cancel{L}}{\cancel{c}} = \mathcal{E}_{\Delta f} = KT \cdot \frac{\cancel{\Delta f}}{\cancel{c}} \cdot \cancel{2}\cancel{L}$$

$$\longrightarrow \quad S_{v_R} = 4KTR_0$$

Shot noise

Shot noise is due to the granularity of the electrical charge.





$$Q_1 = Q_{01} + q\left(1 - \frac{x}{L}\right) \qquad Q_2 = Q_{02} + q\frac{x}{L}$$

$$\frac{dQ_1}{dt} = -\frac{q}{L}\frac{dx}{dt} \qquad \frac{dQ_2}{dt} = +\frac{q}{L}\frac{dx}{dt}$$

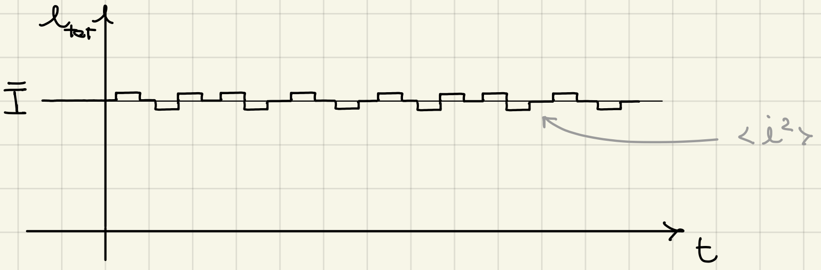
For each carrier crossing the space charge region:

$$i(t) = \frac{q}{L} \cdot v(t)$$ velocity

If average current is $\overline{I}$ then average number of carriers (electrons or holes) crossing this region per second (rate λ) is:

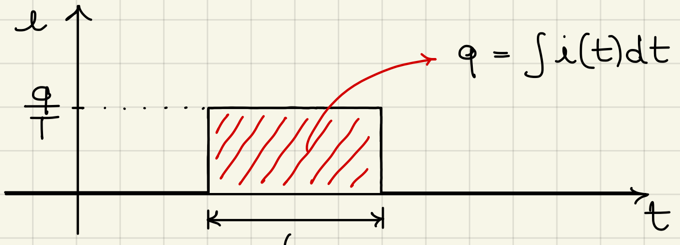
$$\lambda = \frac{\overline{I}}{q}$$

But if we looked microscopically we would see:



because of each individual carrier travelling through the region.

For each single charge we can then depict its current waveform as:



$$\frac{1}{T} \text{rect}\left(\frac{t}{T}\right)$$

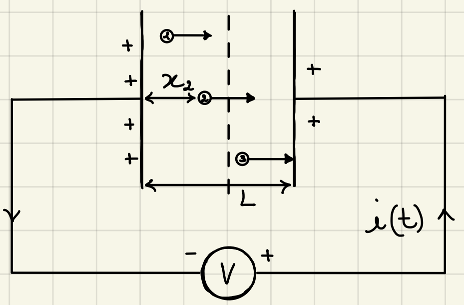
$$q = \int i(t)dt \Rightarrow i(t) = q \cdot h(t)$$

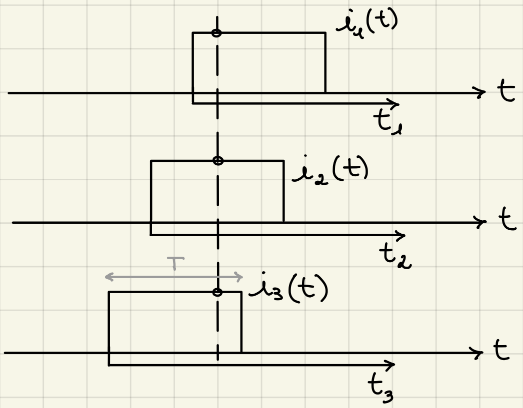
where $\int_{-\infty}^{+\infty} h(t) = 1$

↳ travelling time $T = \frac{L}{v_{sat}}$

We consider a square waveform as elementary waveform of the granular current since in a p-n junction carrier velocity saturates, hence $i = q/L \cdot v_{sat}$ is constant.

Our task now is to find a model to derive the average square value $\langle i^2 \rangle$ of the superposition of these current fluctuations which can effectively be seen as a form of noise (shot noise).





$$i(t) = \sum_j q h_j(t) = q h_1(t) + q h_2(t) + q h_3(t) + \dots = \sum_j q h(t_j)$$

Moving to a continuous series of pulses:

$$\langle i(t)\rangle = \int_{-\infty}^{+\infty} q\, h(t)\cdot \boxed{\lambda dt} \rightarrow$$ number of pulses starting in the elementary timeframe dt

$$\int_{-\infty}^{+\infty} h(t)\,dt = 1$$

$$\downarrow$$

$\bar{I} = q\cdot\lambda$ which is exactly the equation we previously stated, therefore so far everything seems to be coherent.

If now we were able to compute $\langle i^2(t)\rangle$ we could then derive the value of the variance $\underset{\text{noise}}{\sigma_i^2}$, provided the relation $\sigma_i^2 = \langle i^2\rangle - \langle i\rangle^2$.

$$i^2(t) = (q\,h(t_1) + q\,h(t_2) + q\,h(t_3) + \dots)^2 =$$

$$= \underline{q^2 h_1^2 + q^2 h_2^2 + q^2 h_3^2 + \dots} + \underline{2q^2 h_1 h_2 + 2q^2 h_1 h_3 + \dots}$$

Moving to a continuum: they represent two different pulses

$$\langle i^2(t)\rangle = \underline{q^2 \int_{-\infty}^{+\infty} h^2(t)\lambda\,dt} + q^2 \underline{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} [h(x)\lambda\,dx]\cdot[h(y)\lambda\,dy]}$$

$$= q^2\lambda \int_{-\infty}^{+\infty} h^2(t)\,dt + q^2\lambda^2$$

$$= q^2\lambda \int_{-\infty}^{+\infty} h^2(t)\,dt + \langle i(t)\rangle^2$$

Since $\sigma_i^2 = \langle i^2(t)\rangle - \langle i(t)\rangle^2$ we can immediately conclude that:

$$\sigma_i^2 = q^2\lambda \int_{-\infty}^{+\infty} h^2(t)\,dt$$

To obtain $S_i(f)$ we can use Parseval's theorem:

$$\boxed{\int_{-\infty}^{+\infty} h^2(t)\,dt = \int_{-\infty}^{+\infty} |H(f)|^2\,df}$$

$$\int_0^{+\infty} S_i(f)\,df = \sigma_i^2 = q^2\lambda \int_{-\infty}^{+\infty} h^2(t)\,dt \overset{\downarrow}{=} q^2\lambda \int_{-\infty}^{+\infty} \underset{\text{even function}}{|H(f)|^2}\,df = 2q^2\lambda \int_0^{+\infty} |H(f)|^2\,df$$

$$\Longrightarrow S_i(f)\,df = 2q^2\lambda\,|H(f)|^2$$

$$= 2q\bar{I}\,|H(f)|^2$$

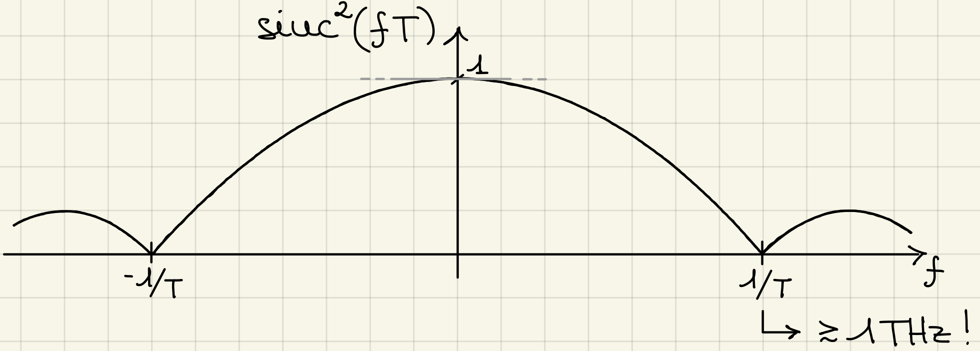
How much is $|H(f)|$?

In our model we depicted h(t) to be a rectangular pulse:

$$h(t) = \frac{1}{T}\,\mathrm{rect}\left(\frac{t}{T}\right)$$

whose Fourier transform is a cardinal sine:

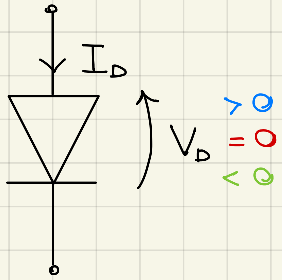
$$H(f) = \frac{\sin(\pi f T)}{\pi f T} = \mathrm{sinc}(fT)$$



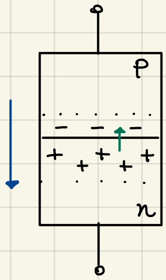
Since we don't usually work with frequencies beyond GHz we can assume the $|H(f)|^2$ to be constant $(=1)$ in our range of interest (remember that $T = \frac{L}{v_{sat}}$ → very small!)

$$\Longrightarrow \boxed{S_I(f) = 2qI}$$

In a diode there are two independent current contributions:



$$I_D = I_s\left(\underbrace{e^{\frac{qV_D}{kT}}}_{\substack{I_D + I_s \\ \text{diffusion} \\ \text{current}}} - \underbrace{1}_{\substack{I_s \\ \text{reverse} \\ \text{current}}}\right)$$



one contribution due to the concentration gradient outside the space charge region (diffusion) and one contribution due to the weak voltage difference across the region (reverse).

Since they are two independent current fluxes their contributions sum up in terms of shot noise spectral density:

$$S_I = 2q(I_D + I_s) + 2qI_s$$

forward bias: $S_I = 2qI_D$
$(I_D \gg I_s)$

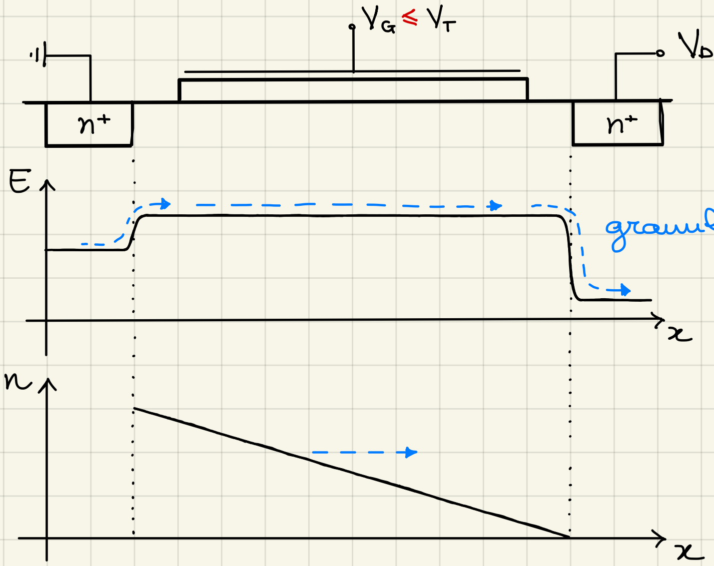
reverse bias: $S_I = 2qI_s$
$(I_D = -I_s)$

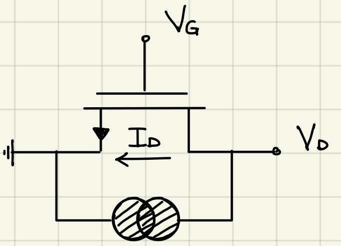
equilibrium : $S_I = 4qI_s$
$(I_D = 0)$

Note that at equilibrium shot noise and thermal noise are equal (they're actually the same thing):

$$S_I\Big|_{eq} = 4qI_s = 4q\frac{I_s}{kT}\cdot kT = 4kTg_D = S_{i_{thermal}}$$

This argument is also applicable to a transistor in weak inversion:





$$\frac{4kT}{R_{ch}} \simeq 4kT\gamma g_m$$

thermal noise current spectral density

$I_D = I_0\, e^{\frac{qV_{GS}}{nkT}}$ weak inversion current

$g_m = q\cdot\frac{I_D}{nkT}$ weak inversion transconductance

shot noise current spectral density

$$\Longrightarrow S_I = 2qI_D$$

$$\overset{|}{=} 2q\frac{I_D}{nkT}nkT = 4kT\frac{n}{2}g_m = 4kT\gamma g_m$$

$$\Longrightarrow S_{i_T} = 4kT\gamma g_m \begin{cases} \gamma = 2/3 & \text{saturation } (\gamma = 2 \text{ short channel}) \\ \gamma = 1 & \text{ohmic region} \\ \gamma = \frac{n}{2} & \text{weak inversion} \end{cases}$$

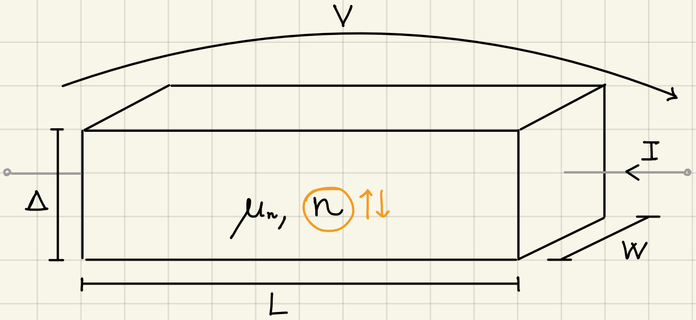
$$n = 1 + \frac{C'_D}{C'_{ox}}$$

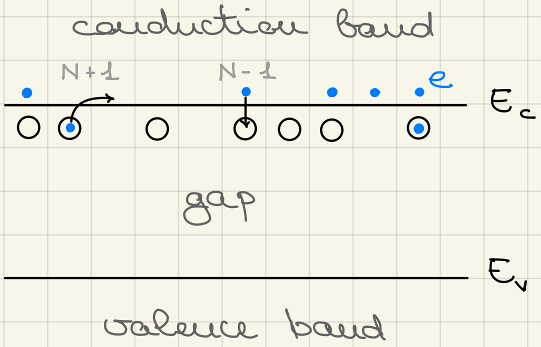
C'_D: depleted region capacitance

C'_{ox}: oxide capacitance

RTN noise

RTN noise (Random Telegraph Noise, also called burst or pop-corn noise) is related to capture and emission processes of electrons, that are caused by non-idealities of the devices' fabrication.





$$I = G \cdot V = \underbrace{q \mu_n n}_{\sigma} \frac{W\Delta}{L} V = q\mu_n \boxed{n \frac{W \cdot \Delta \cdot L}{L^2}} \cdot V = q \mu_n \frac{\boxed{N}}{L^2} V$$

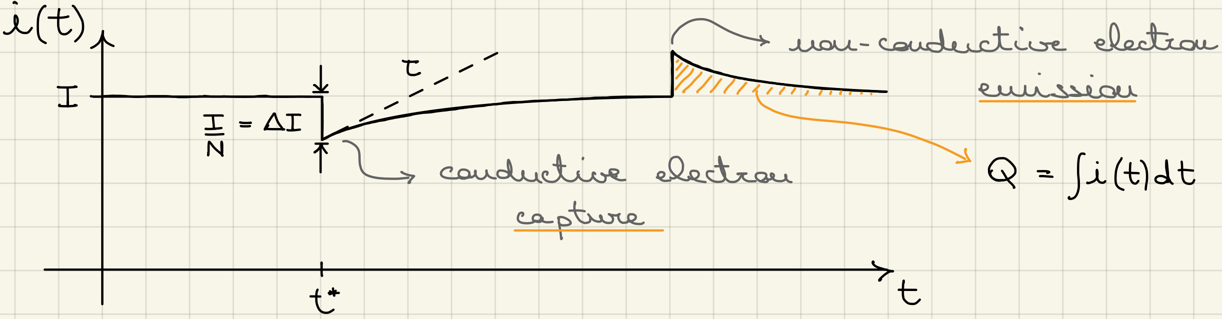
Because of the trapping or releasing of charges (electrons) their concentration n might not be constant, resulting in a modulation of the device conductivity σ and therefore in a variation of the current I.

The capture or release of an electron means that the total number of charges N will go down or up by one unit, causing the aforementioned variation in current.

$$\Delta N \longrightarrow \Delta I = q \mu_n \frac{V}{L^2} \cdot \Delta N$$

$$\Longrightarrow \frac{\Delta I}{I} = \frac{\Delta N}{N}$$

$$\Longrightarrow \Delta I = \frac{I}{N} \quad (\text{since } \Delta N = 1)$$



We assume the "capture-release waveform" to have an exponential behaviour: a conductive electron is captured, causing an instantaneous current variation; the current is then expected to recover the steady state value since the electron will eventually be ejected. Our assumption is that this recovery transient is characterized by a time constant which is the average time needed for each carrier to be released (we will see that this is true only if we consider the superposition of many electrons being captured at the same time - of course the capture and ejection of a single carrier would indeed have a square-like waveform, not an exponential one*).

We therefore expect the current to be somehow affected by pulses with a negative step followed by a positive recovery transient with an exponential-like behaviour. This could actually happen to an electron that sits outside the conductive band as well: it can be ejected causing an increment in the total number of carriers and in the current (positive step) and will then be absorbed back in its original state (negative recovery transient).

$i(t) = \frac{I}{N} \cdot e^{-t/\tau} \text{step}(t)$ current waveform of capture-emission of electrons

$i(t) = Q \, h(t) = Q \left(\frac{1}{\tau} e^{-t/\tau} \cdot \text{step}(t) \right)$

$q = \int i(t) dt = \int q \, h(t) dt$

$\int h(t) dt = 1$

$\int_0^{+\infty} \frac{1}{\tau} e^{-t/\tau} dt = 1$

$\Longrightarrow \frac{I}{N} e^{-t/\tau} = Q \frac{1}{\tau} e^{-t/\tau} \Longrightarrow Q = \frac{I}{N} \cdot \tau$

Now that we know $i(t)$ in the form of $Q\,h(t)$ it is possible to re-use the same result previously obtained for shot noise:

Anytime the current is given by the series of many pulses in the form of $i(t) = Q\,h(t)$ where $h(t)$ is an elementary waveform, the resulting overall PSD is:

$$\underline{S_I(\omega) = 2 Q^2 \lambda \, |H(\omega)|^2}$$

$\longrightarrow \lambda = \lambda_e = \lambda_c$ same rate for emission and capture in steady-state conditions

$\longrightarrow$ two processes (emission and capture) that are independent

$\longrightarrow \frac{1}{\tau} e^{-t/\tau} \text{step}(t) \xrightarrow{\mathcal{F}} \frac{1}{\tau} \frac{1}{\frac{1}{\tau} + j\omega}$

$$\Longrightarrow S_I = 4Q^2\lambda\frac{1}{1+\omega^2\tau^2} = 4\frac{I^2}{N^2}\tau^2\lambda\frac{1}{1+\omega^2\tau^2}$$

How much is λ? We know it represents the rate at which capture and emission phenomena happen, so it can be put in a relation with the recovery time constant τ and with the number of traps (impurities) N_T in the device:

$$\lambda \approx \frac{N_T}{\tau}\cdot\beta \rightarrow$$

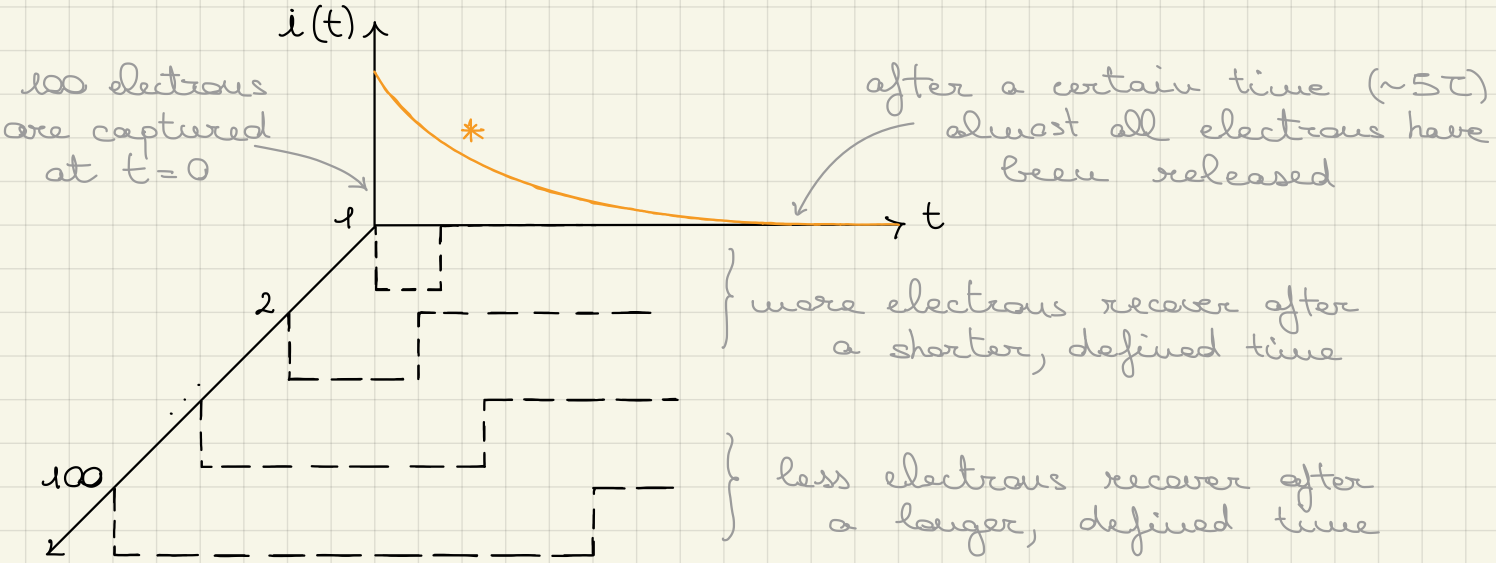
proportionality factor (pure number)

$$\Longrightarrow S_I = 4\frac{I^2}{N^2}\frac{N_T}{\cancel{\tau}}\beta\frac{\tau^{\cancel{2}}}{\underline{1+\omega^2\tau^2}} \rightarrow$$

Lorentzian shape

This frequency dependance on $1/f^2$ can be actually seen through appropriate experiments to estimate the noise PSD.
If we were to look instead at the time-domain behaviour of current, we would not see exponential-like pulses such as the ones we used for our calculation but we would see square-like bumps representing the real capture and emission of carriers due to traps. After all, an electron cannot be ejected in fractions but only in discrete quantities. Then how come our frequency-domain model was still correct?

The reason is that when looking at noise we're not looking at each single event but instead we're taking into account all the events taking place in parallel.



100 electrons are captured at $t=0$

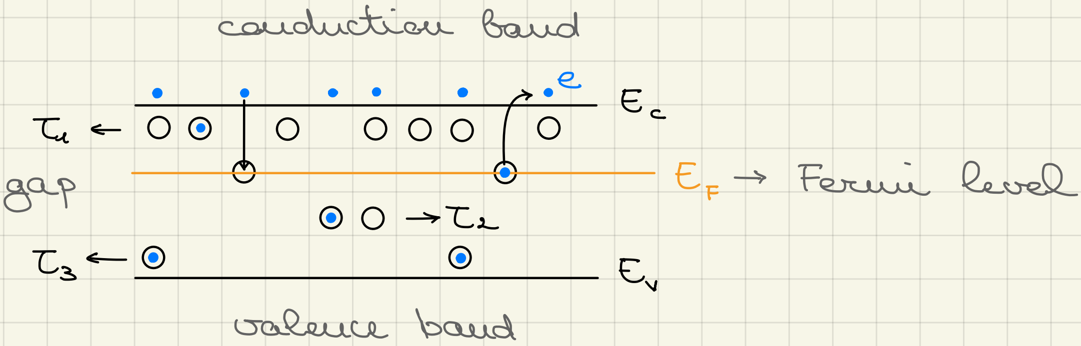
after a certain time ($\sim 5\tau$) almost all electrons have been released

more electrons recover after a shorter, defined time

less electrons recover after a longer, defined time

The superposition of many square-like pulses that obey a certain time constant law returns an overall exponential-like pulse

We now just need a value for β.



In a real device there would be many different traps at many different energy levels in a non-ideal device, each of them having its own time constant. This result in a overall time constant τ that is the superposition of many different ones.

We can simplify this considering that all traps below Fermi level are always occupied, while all traps above Fermi level are always free, which means that the population of traps contributing to the capture and emission of electrons only includes those around Fermi level.

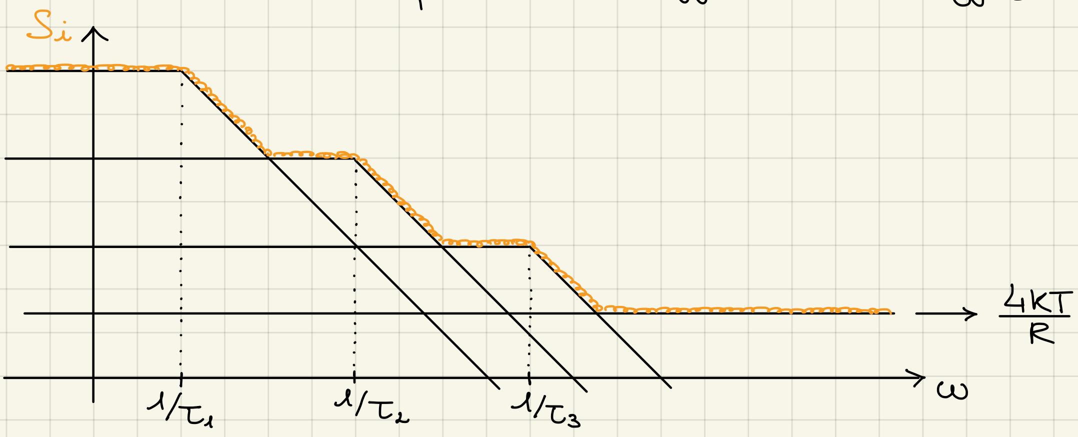
If we limit ourselves to just consider processes happening around Fermi level it can be demostrated that

$$\left[\beta \simeq \frac{1}{4}\right]$$

time constant of traps at Fermi level

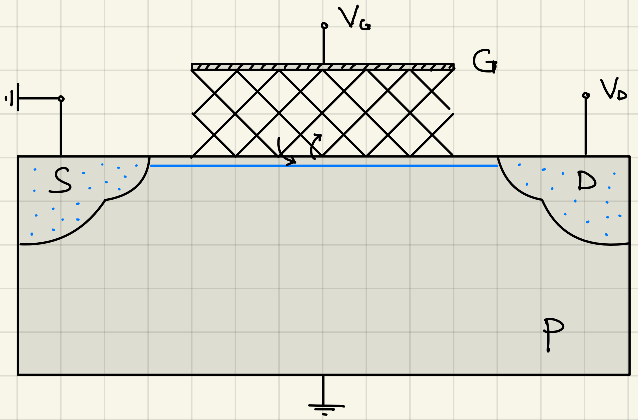
$$\Longrightarrow \boxed{S_I(\omega) = \left(\frac{I}{N}\right)^2 \cdot \frac{N_T \tau}{1 + \omega^2 \tau^2}}$$

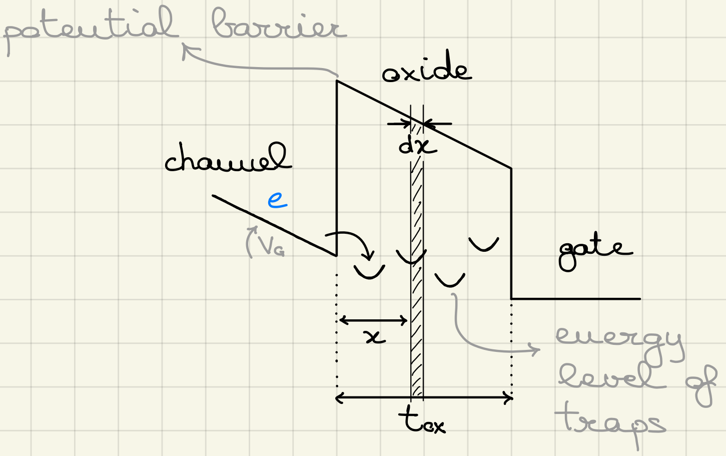
Note that in presence of more than one family of traps, with different time constant at a different energy level (e.g. there are different devices in the circuit all of them affected by RTN), each respective PSD will sum up resulting in a "stair" looking shape (superposition of many Lorentzian shapes with different cut-off frequencies).



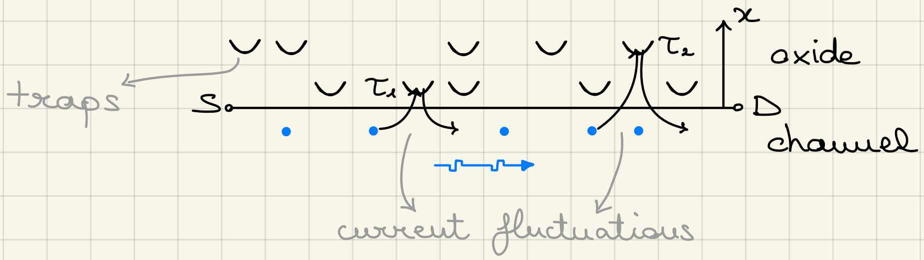
Flicker noise

Flicker noise ("1/f noise") is typically related to transistors and is due to the non-ideal junction between the semiconductor and the oxide, which can be place of capture and emission of the channel charges, similarly to RTN noise.





The oxide is characterized by some relevant traps. Carriers flowing in the channel can easily be trapped by them, also thanks to the electrostatic pressure due to the voltage difference between the base and the gate.



In order for an electron to jump inside the oxide it has to tunnel through the oxide potential barrier (the oxide is an isolating material), which is allowed only in terms of quantum mechanics: each electron has a certain probability to overcome the potential difference and reach the trap (tunnel effect). This likelihood decreases exponentially the farther the trap is.

Also the time constant associated with the capture and emission of the electron from a trap will be exponentially dependent on the distance from the junction. In particular, we expect it to increase the farther the trap is, since the electron will be deeper inside the oxide.

$$[\tau = \tau_0 e^{\gamma x}]$$

accounts for the energy barrier height

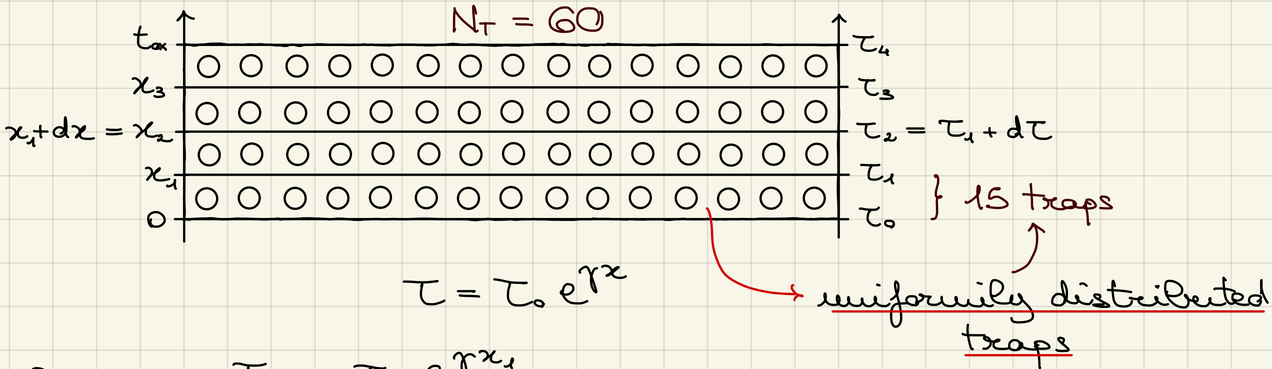
Since the current fluctuations are caused by capture-emission of carriers caused by traps, we can use the same result previously obtained for RTN; this time, however, we won't have just one single time constant characterizing the fluctuations, instead we are going to have a distribution of time constants, each of them contributing with a certain weight $g(\tau)$ to the overall PSD:

$$\Longrightarrow \quad S_I = N_T\left(\frac{I}{N}\right)^2\int_{\tau_{min}}^{\tau_{max}}\frac{\tau\, g(\tau)\, d\tau}{1+\omega^2\tau^2}$$

It is possible to derive $g(\tau)$ by considering the (uniform) distribution of traps through the oxide thickness:

$$\underbrace{N_T\cdot\frac{dx}{t_{ox}}}_{\text{number of traps in the elementary slice } dx \text{ of the oxide}} = \underbrace{N_T\, g(\tau)\, d\tau}_{\text{number of traps characterized by a time constant between } \tau \text{ and } \tau+d\tau}$$

Let's clarify this better:



$$\tau = \tau_0 e^{\gamma x}$$

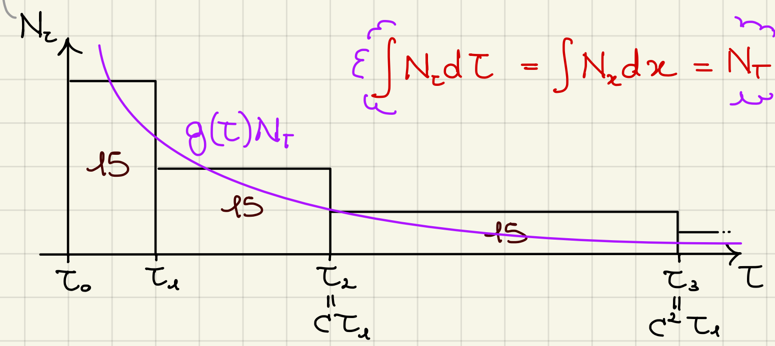
$$x_n = n\cdot x_1 \longrightarrow \tau_1 = \tau_0 e^{\gamma x_1}$$

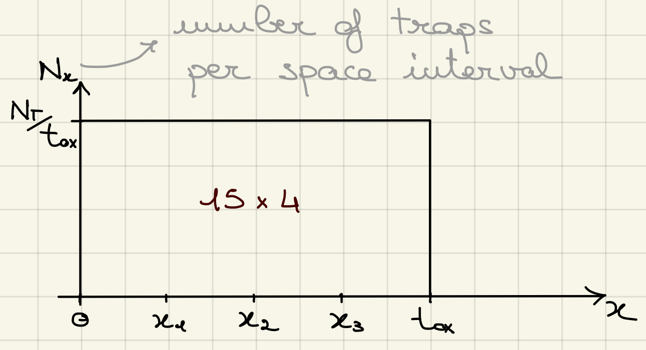
$$\tau_2 = \tau_0 e^{\gamma x_2} = \tau_0 e^{\gamma 2 x_1}$$

$$\left.\begin{aligned}\frac{\tau_2}{\tau_1} &= \frac{\tau_0 e^{2\gamma x_1}}{\tau_0 e^{\gamma x_1}} = e^{\gamma x_1}\\ \frac{\tau_3}{\tau_2} &= \frac{\tau_0 e^{3\gamma x_1}}{\tau_0 e^{2\gamma x_1}} = e^{\gamma x_1}\end{aligned}\right\} \tau_{n+1} = e^{\gamma x_1}\tau_n = c\cdot\tau_n$$

$g(\tau)\cdot N_T$
$\parallel$
number of traps per time constant

$$\left\{\int N_\tau d\tau = \int N_x dx = N_T\right\}$$





$$\int \underset{\downarrow}{N_\tau}\, d\tau = \int \underset{\downarrow}{N_x dx} = N_T$$

$$N_T \cdot g(\tau)\, d\tau = \frac{N_T}{t_{ox}}\, dx$$ (what we had written before!)

$$\Longrightarrow g(\tau) = \frac{dx}{d\tau}\frac{1}{t_{ox}} = \frac{dx}{\gamma \tau dx}\frac{1}{t_{ox}} = \frac{1}{t_{ox}\gamma}\cdot\frac{1}{\tau}$$

$$d\tau = d(\tau_0 e^{\gamma x}) = \tau_0 \gamma e^{\gamma x} dx = \gamma \tau dx$$

$$\Longrightarrow S_I = \frac{N_T}{t_{ox}\gamma}\left(\frac{I}{N}\right)^2 \int_{\tau_{min}}^{\tau_{max}} \frac{d\tau}{1+\omega^2\tau^2} = \frac{N_T}{t_{ox}\gamma}\left(\frac{I}{N}\right)^2 \frac{1}{\omega}\left[\operatorname{arctg}(\omega\tau_{max}) - \operatorname{arctg}(\omega\tau_{min})\right]$$

To evaluate the finite integral we should consider how much is ω in our range of interest vs. τ_{min} and τ_{max}. Since ω_{min} and ω_{max} (maximum and minimum observation time) range from Hz to GHz, while τ_{max} and τ_{min} range from years to ps, it is safe to assume that:

$$\left.\begin{array}{l} \omega_{max} \ll \frac{1}{\tau_{min}} \\ \omega_{min} \gg \frac{1}{\tau_{max}} \end{array}\right\} \quad \frac{1}{\tau_{max}} \ll \omega \ll \frac{1}{\tau_{min}}$$

$$\downarrow$$

$$\omega\tau_{min} \ll 1$$
$$\omega\tau_{max} \gg 1$$

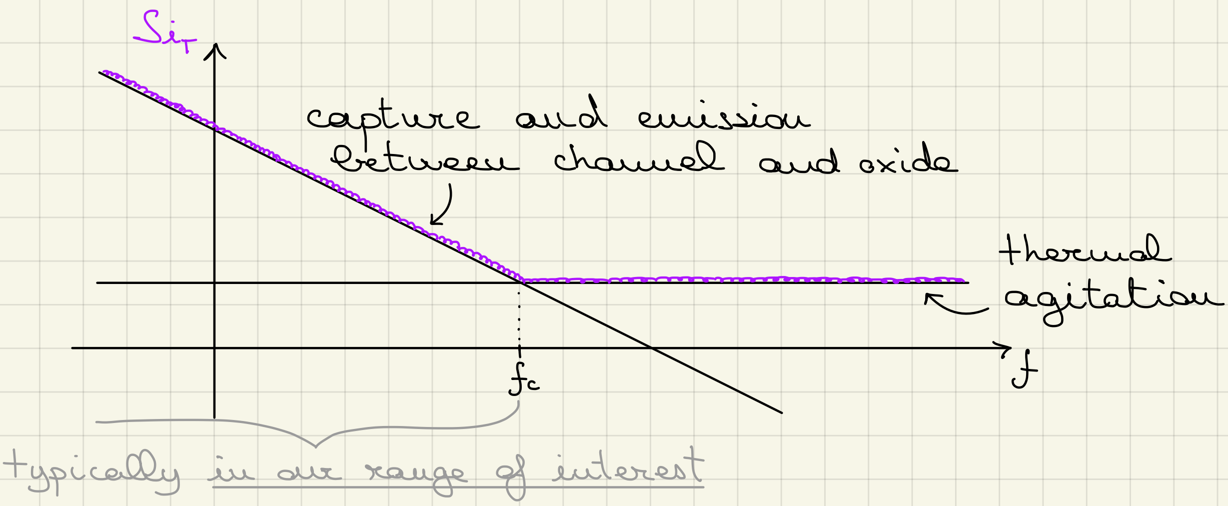
and therefore we can write:

$$\operatorname{arctg}(\omega\tau_{min}) \simeq 0 \qquad \operatorname{arct}(\omega\tau_{max}) \simeq \frac{\pi}{2}$$

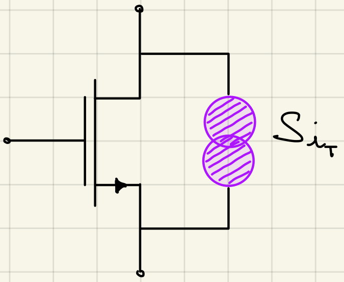
$$\Longrightarrow S_I(\omega) = \frac{N_T}{t_{ox}\gamma}\left(\frac{I}{N}\right)^2 \frac{1}{\omega}\cdot\frac{\pi}{2}$$

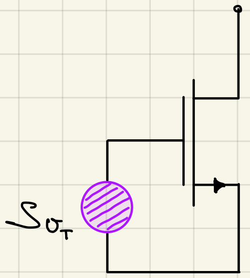
$$\Longrightarrow \boxed{S_I(f) = \frac{N_T}{4\, t_{ox}\gamma}\left(\frac{I}{N}\right)^2 \frac{1}{f}}$$

→ Mc Worther form for the 1/f noise in MOSFETs



Let's see what parts of the expression of the $\frac{1}{f}$ noise can be controlled from a designer's perspective.





$$S_I = \frac{N_T}{4 t_{ox} \gamma} \left(\frac{I}{N}\right)^2 \cdot \frac{1}{f} =$$

$N_T = n_T \cdot Vol_{ox} = n_T (WL t_{ox})$ n. of trap in oxide

$N = \frac{C'_{ox}}{q} \cdot WL \cdot V_{ov}$ n. of carriers in channel

$$= \frac{n_T \cancel{WL} \cancel{t_{ox}}}{4 \cancel{t_{ox}} \gamma} \cdot \frac{k \cancel{V_{ov}^2} \cdot I}{C'^2_{ox} (WL)^{\cancel{2}} \cancel{V_{ov}^2}} \cdot \frac{q^2}{f} = \frac{q^2 n_T}{4 \gamma} \frac{1/2 \mu_n \cancel{C'_{ox}} \cancel{W}/L \cdot I}{C'^{\cancel{2}}_{ox} (\cancel{W} \cdot L)} \cdot \frac{1}{f} =$$

$$= \underbrace{\frac{q^2 n_T \mu_n}{8 \gamma C'_{ox}}}_{\text{set by technology}} \frac{I}{L^2} \cdot \frac{1}{f} \implies \boxed{S_I = K_I^{(1/f)} \frac{I}{L^2} \cdot \frac{1}{f}}$$

Input-referred voltage noise: $S_V \cdot g_m^2 = S_I$

$$S_V = \frac{K_I^{(1/f)}}{4 k \cancel{I}} \cdot \frac{\cancel{I}}{L^2} \cdot \frac{1}{f}$$

$$= \frac{K_I^{(1/f)}}{{}^{2}\cancel{4} \cdot 1/\cancel{2} \mu_n C'_{ox} \frac{W}{\cancel{L}} \cdot L^{\cancel{2}}} \cdot \frac{1}{f}$$

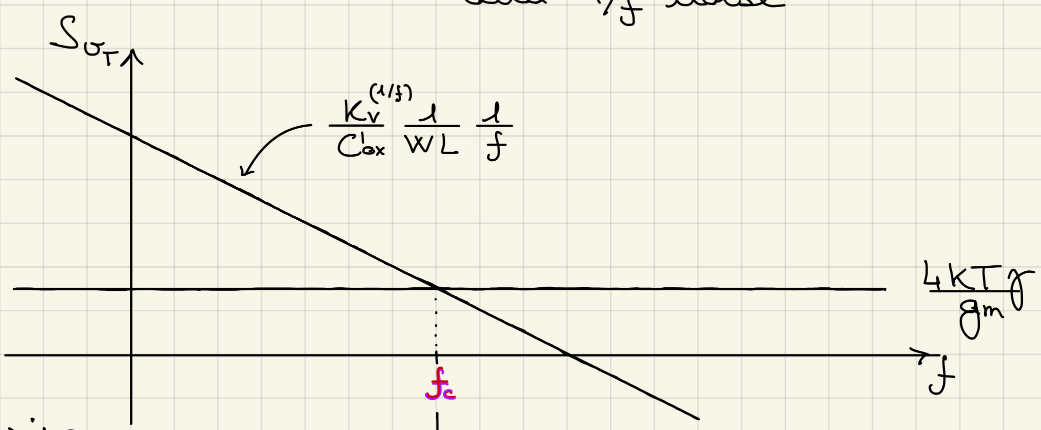
$$= \frac{K_I^{(1/f)}}{2 \mu_n C'_{ox} W \cdot L} \cdot \frac{1}{f}$$

$\downarrow$ $K_V^{(1/f)}$

$$\implies \boxed{S_V = \frac{K_V^{(1/f)}}{C'_{ox} W \cdot L} \cdot \frac{1}{f}} \longrightarrow \text{Tvidis formula}$$

↳ does NOT depend on bias!

Noise corner frequency: crossover between thermal noise and 1/f noise



↓

the lower, the better

↓

~~reduce g_m~~ or increase transistor size

increasing thermal noise not a good option

<u>Note</u> how both $1/f$ noise PSD and the transistors mismatch variance (= power) are both proportional to $1/W \cdot L$.

This is not a coincidence and it can be explained considering the effects of capture/emission of an electron in/from the oxide on the threshold voltage and the transconductance factor.

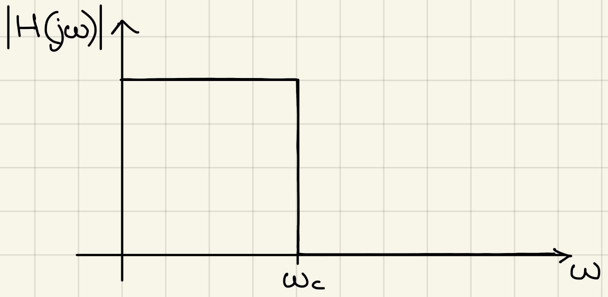
Every time a capture/emission process occurs, the local threshold voltage of the transistor varies, as well as the oxide capacitance and therefore the k factor.

So the $1/f$ noise could actually be seen as noise related to fluctuations of the transistors parameters.

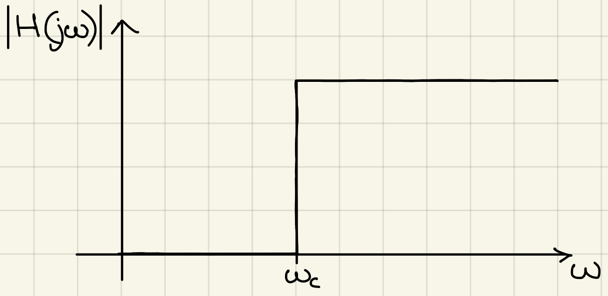
As explained for the computation of the variability terms $\sigma^2_{\Delta V_T}$ and $\sigma^2_{\frac{\Delta k}{k}}$, a larger area of the transistor allows to even out all this fluctuations as their contributions cancel out more easily when there are many of them.

For this same reason, a larger area allows for more capture/emission processes to happen simultaneously thus reducing their overall contribution, resulting in a smaller σ^2, that is, a smaller PSD.

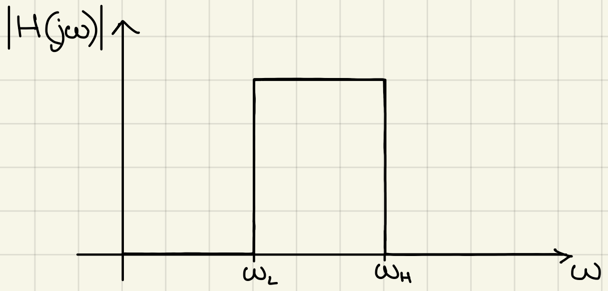
ANALOG FILTERS



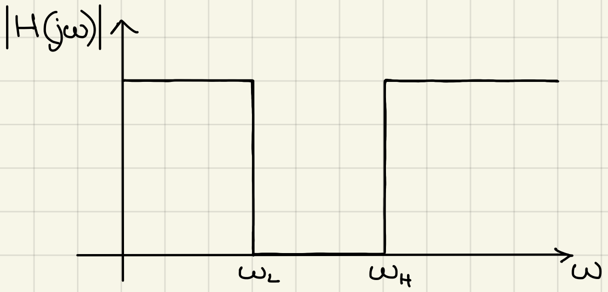
Low-Pass Filter



High-Pass Filter



Band-Pass Filter

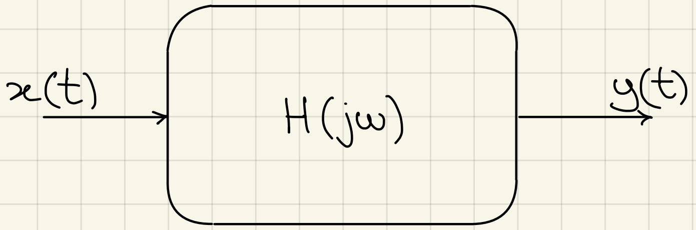


Band-Stop Filter

These *ideal* filters are described by a "brick-wall" transfer function, which in reality cannot be implemented.
Why?

The transfer function of a filter is associated to the delta-pulse frequency response of the system.
It can be shown that in order to have a transfer function with a very sharp, instantaneous transition, you need a pulse response that *forfeits the laws of causality* - that is, the system would need to respond to the pulse before the pulse has even arrived.
Of course such time behaviour cannot be obtained in the real world.

We therefore need to be somewhat tolerant with our real filters and design them to meet some specific requirements.



E.g. we'd like $|H(j\omega)|$ to be as constant as possible over a certain frequency range.

Let's see what are the ideal requirements first.

Assume that $x(t) = A\sin(\omega_1 t) + B\sin(\omega_2 t)$ with both frequency components in-band.

Then $y(t) = A|H(j\omega_1)|\sin(\overset{\omega_1(t-\tau)}{\omega_1 t + \varphi_1}) + B|H(j\omega_2)|\sin(\overset{\omega_2(t-\tau)}{\omega_2 t + \varphi_2})$

In order to properly filter the signal we would need:

$|H(j\omega_1)| = |H(j\omega_2)|$ AND $\varphi_1 = -\omega_1\tau, \quad \varphi_2 = -\omega_2\tau$

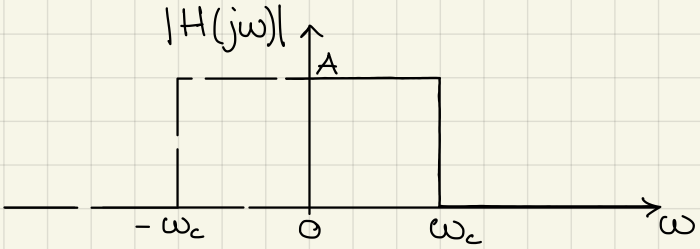
so both frequencies in our band of interest must be amplified and shifted by the same amount.

$|H(j\omega)| = \text{const.}$
amplitude stays constant with frequency

$\varphi = -k\omega$
phase grows proportional with frequency

In the ideal filter shown before, this is easily achieved.

Let's show why in practice it cannot be implemented.

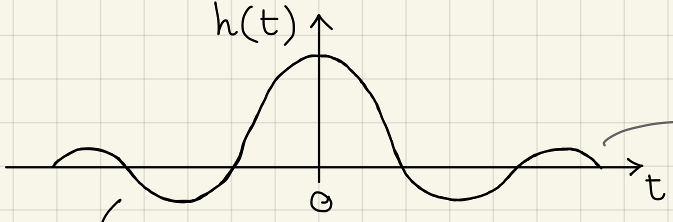


$$h(t) = \mathcal{F}^{-1}[H(j\omega)] = \int_{-\infty}^{+\infty} H(j\omega)\, e^{j\omega t}\, d\omega = \int_{-\omega_c}^{\omega_c} A e^{j\omega t} \frac{d\omega}{2\pi} =$$

$$= \frac{A}{2\pi}\frac{1}{jt}\left[e^{j\omega_c t} - e^{-j\omega_c t}\right] = \frac{\omega_c A}{\pi}\left[\frac{e^{j\omega_c t} - e^{-j\omega_c t}}{2j\omega_c t}\right] =$$

$$= \frac{\omega_c A}{\pi}\,\frac{\sin(\omega_c t)}{\omega_c t} = \frac{\omega_c A}{\pi}\,\text{sinc}(\omega_c t)$$

(non normalized)



needs a theoretically infinite observation time

non-causal response!
filter is excited before the delta-pulse (which sits at t = 0) has even happened

↓

cannot be implemented

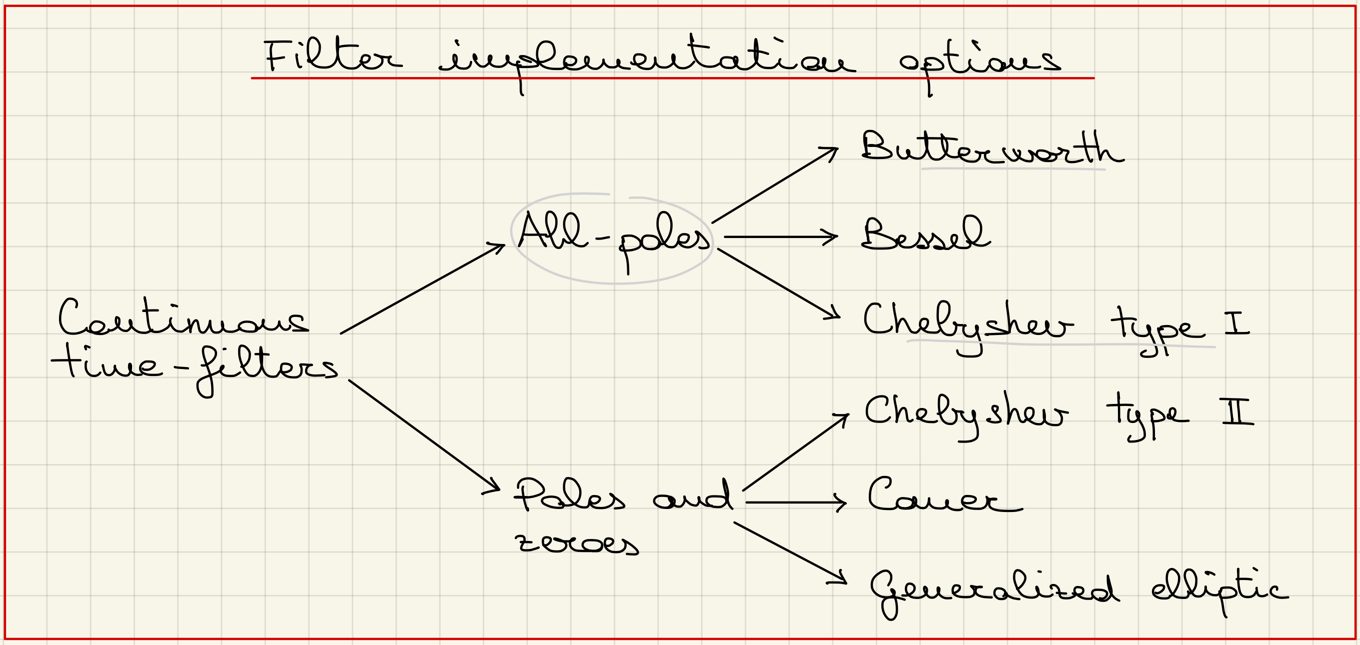
→ Approximated filters

We should accept a maximum in band attenuation (A_{BP}), a minimum out of band attenuation (A_{SP}) and two different values for the cut-off frequency to allow for a smooth transition between the band-pass region ($\omega < \omega_{BP}$) and the stop-band region ($\omega > \omega_{SB}$). Of course we should also accept to have a non-constant band-pass amplification (and non-constant stop-band attenuation).

The approximation requirements can be implemented through an appropriate transfer function:

$$T(s) = T_0 \frac{s^n + a_{n-1}s^{n-1} + \ldots + a_1 s + a_0}{s^n + b_{n-1}s^{n-1} + \ldots + b_1 s + b_0}$$

We're going to see how to mathematically build this transfer function for a LPF, and then extend the same method to the HPF and the BPF through appropriate variable transformations.

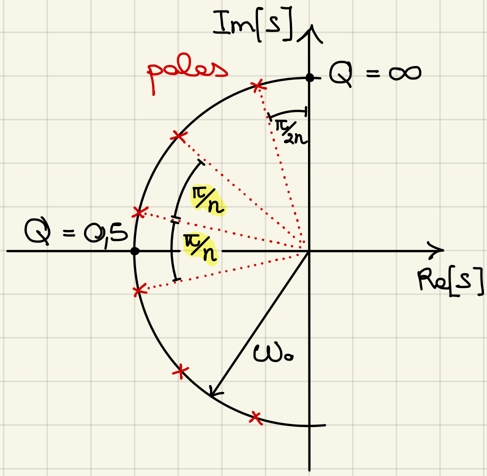


All-poles transfer functions implement low-pass filters and, through the associated transformations, high/band-pass filters.

Poles and zeroes transfer functions are useful in those cases where the elimination of a specific frequency tone is needed.

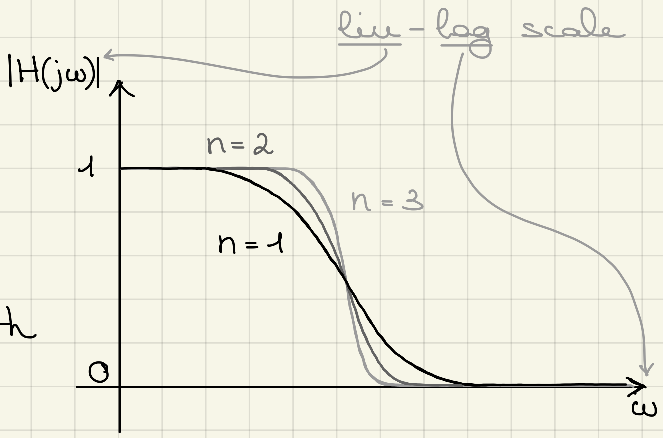
We're going to deal only with all-poles transfer functions for our purposes, namely the Butterworth and the Chebyshev implementations.

Butterworth

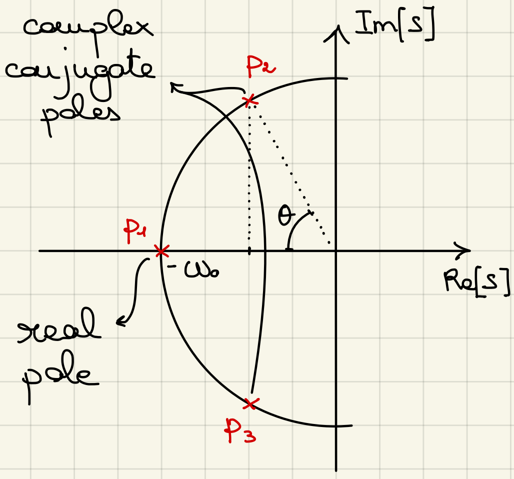


n = 6 (even)

$$T(s) = \frac{\gamma}{D_n(s)}$$
$$= \frac{1}{B_n(s)}$$



The Butterworth transfer function is characterized by just poles (all-poles transfer function).
The number of poles depends on the filter order, which sets the sharpness of the band cut-off.
The poles in the Gauss plane are situated on a circle with a characteristic frequency ω_0, that is related (but not equal) to the band-pass frequency, and their angular distance is set by the filter order as π/n.



n = 3 (odd)

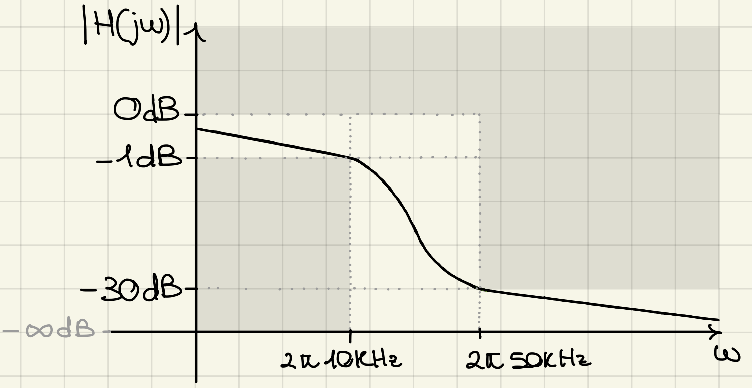
$$\text{E.g.:}\quad T(s)\Big|_{n=3} = \frac{\gamma}{(s+\omega_0)\left(s^2+\frac{s\omega_0}{Q_{23}}+\omega_0^2\right)}$$

$$\left[\,Q = \frac{1}{2\xi} = \frac{1}{2\cos\theta} = \frac{|P|}{2\,\mathrm{Re}[P]}\,\right]$$

$$Q_{23} = \frac{|P_2|}{2\mathrm{Re}|P_2|} = \frac{\omega_0}{2\cdot\omega_0/2} = 1$$

Example: LP FILTER

$$\left.\begin{matrix}\omega_{BP} = 2\pi\cdot 10\,\text{KHz} & A_{BP} = 1\,\text{dB}\\ \omega_{SB} = 2\pi\cdot 50\,\text{KHz} & A_{SB} = 30\,\text{dB}\end{matrix}\right\}\xrightarrow{\text{Butterworth}} H(s)$$



$$H(j\omega) = \frac{\gamma}{D_n(j\omega)}$$

It can be easily demonstrated that:

$$\left[\, |H(j\omega)|^2 = \frac{1}{1+\left(\frac{\omega}{\omega_0}\right)^{2n}} \,\right]$$

Note that the attenuation is the absolute value of the denominator of the transfer function itself:

I. if $\omega \leq \omega_{BP}$ then $\frac{1}{\sqrt{1+\left(\frac{\omega}{\omega_0}\right)^{2n}}} \geq \frac{1}{A_{BP}}$

II. if $\omega \geq \omega_{SB}$ then $\frac{1}{\sqrt{1+\left(\frac{\omega}{\omega_0}\right)^{2n}}} \leq \frac{1}{A_{SB}}$

To grant these conditions, since $|H(j\omega)|$ is monotonously decreasing, it is sufficient to verify them for $\omega = \omega_{BP}$ and $\omega = \omega_{SB}$ respectively.

I. $1 + \left(\frac{\omega_{BP}}{\omega_0}\right)^{2n} \leq A_{BP}^2$

II. $1 + \left(\frac{\omega_{SB}}{\omega_0}\right)^{2n} \geq A_{SB}^2$

(Note also that $|H(j\omega_0)| = \frac{1}{\sqrt{2}} = -3dB$.)

I. $\left(\frac{\omega_{BP}}{\omega_0}\right)^n \leq \left[\sqrt{A_{BP}^2 - 1} = \varepsilon_{BP}\right]$

II. $\left(\frac{\omega_{SB}}{\omega_0}\right)^n \geq \left[\sqrt{A_{SB}^2 - 1} = \varepsilon_{SB}\right]$

attenuation coefficient

I. $\frac{\omega_{BP}}{\omega_0} \leq \varepsilon_{BP}^{1/n}$

II. $\frac{\omega_{SB}}{\omega_0} \geq \varepsilon_{SB}^{1/n}$

$$\frac{\omega_{BP}}{\omega_{SB}} \leq \left(\frac{\varepsilon_{BP}}{\varepsilon_{SB}}\right)^{1/n}$$

$$\downarrow$$

$$K \leq (K_\varepsilon)^{1/n}$$

$\left[K = \frac{\omega_{BP}}{\omega_{SB}}\right]$ selectivity index

$\left[K_\varepsilon = \frac{\varepsilon_{BP}}{\varepsilon_{SB}}\right]$ discrimination index

$K < 1$ and $K_\varepsilon < 1$ always

$$\ln k \leq \frac{1}{n} \ln k_\varepsilon$$

$$n \ln k \leq \ln k_\varepsilon$$

$\ln k < 0 \leftarrow k < 1$

$$\boxed{n \geq \frac{\ln k_\varepsilon}{\ln k}}$$ order of a Butterworth filter

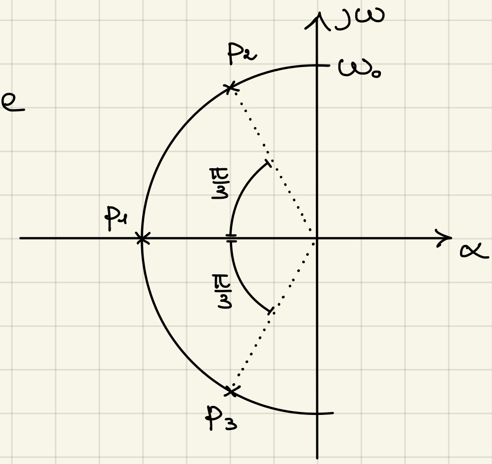
$$K = \frac{\omega_{BP}}{\omega_{SB}} = \frac{10k}{50k} = 0{,}2$$

$$\varepsilon_{BP} = \sqrt{A_{BP}^2 - 1} = \sqrt{10^{1/10} - 1} = 0{,}509$$

$$\varepsilon_{SB} = \sqrt{A_{SB}^2 - 1} = \sqrt{10^{30/10} - 1} = 31{,}607$$

$$K_\varepsilon = \frac{\varepsilon_{BP}}{\varepsilon_{SB}} = \frac{0{,}509}{31{,}607} = 0{,}016$$

$$\longrightarrow n \geq \frac{\ln k_\varepsilon}{\ln k} = \frac{\ln 0{,}016}{\ln 0{,}2} = 2{,}57 \Longrightarrow \underline{\underline{n = 3}}$$



We now need to find a proper value for ω₀. Remember the result previously obtained:

I. $\frac{\omega_{BP}}{\omega_0} \leq \varepsilon_{BP}^{1/n} \longrightarrow \omega_0 \geq \frac{\omega_{BP}}{\varepsilon_{BP}^{1/n}}$

II. $\frac{\omega_{SB}}{\omega_0} \geq \varepsilon_{SB}^{1/n} \longrightarrow \omega_0 \leq \frac{\omega_{SB}}{\varepsilon_{SB}^{1/n}}$

Hence ω₀ should be set within an interval range given by:

$$\boxed{\frac{\omega_{BP}}{\varepsilon_{BP}^{1/n}} \leq \omega_0 \leq \frac{\omega_{SB}}{\varepsilon_{SB}^{1/n}}}$$ characteristic frequency of a Butterworth filter

Having a range of values for ω₀ enables a further degree of freedom when designing the filter (such as having a specific attenuation at a certain frequency).

$$\frac{\omega_{BP}}{\varepsilon_{BP}^{1/n}} = 2\pi \cdot \frac{10kHz}{0{,}509^{1/3}} = 2\pi \cdot 12{,}5kHz$$

$$\Longrightarrow \underline{\underline{12{,}5kHz \leq f_0 \leq 15{,}8kHz}}$$

$$\frac{\omega_{SB}}{\varepsilon_{SB}^{1/n}} = 2\pi \cdot \frac{50KHz}{31{,}607^{1/3}} = 2\pi \cdot 15{,}8KHz$$

$|T(0)| = 1$

$$\Longrightarrow \left[\, T(s) = \frac{\gamma}{(s-\omega_0)\left(s^2+\frac{s\omega_0}{Q_{23}}+\omega_0^2\right)} = \frac{\omega_0^3}{(s-\omega_0)(s^2+s\omega_0+\omega_0^2)} \,\right]$$

$\frac{1}{2\cos\frac{\pi}{3}} = 1$

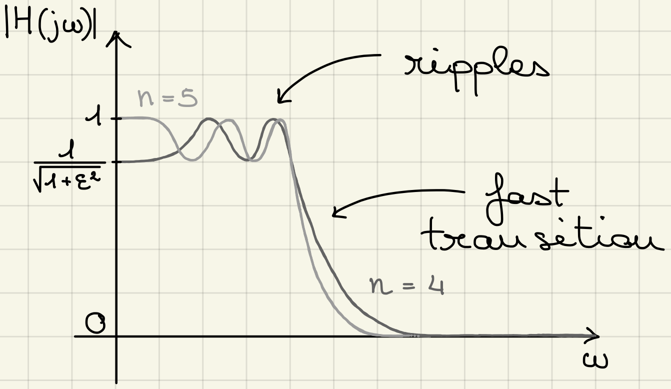
Butterworth transfer functions are extremely flat in the in-band region (they do not have any ripples) and are very regular across the entire spectrum.
They are intended for a <u>maximally flat response</u>

Chebyshev type I

The Chebyshev-I approximant is characterized by a steeper transition (with respect to the Butterworth) for the same order of the filter. However, its in-band behaviour is not as regular and has some ripples (the higher the order, the more the ripples).

When designing a filter one might have to choose between a Butterworth model, of a higher order but more regular, and a Chebyshev-I model, of a lower order but less regular.

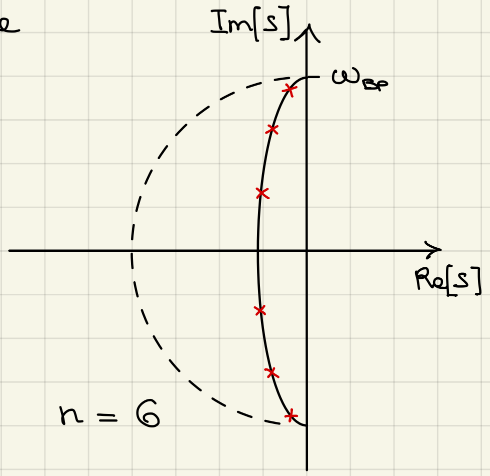
⟹ TRADE-OFF between order (= complexity/cost) and regularity



Note that the number of transitions in-band that cause the ripples is exactly equal to the filter order.

This means that for an even order the DC gain is slightly less than 1 (but still within attenuation requirements).

The poles of the transfer function are placed around an elliptical shape within the reference band-pass circle and they therefore have different radial frequencies.



We must find the filter order and, after that, the ω_0 and Q of each single pole pair.

$n \geq \frac{\ln(K_\varepsilon^{-1})}{\ln(K^{-1})}$ ↔ $\boxed{n \geq \frac{Ch^{-1}(K_\varepsilon^{-1})}{Ch^{-1}(K^{-1})}}$ order of a Chebyshev-I filter

Butterworth

$$\left[\Gamma = \left(\frac{1+\sqrt{1+\varepsilon_{BP}^2}}{\varepsilon_{BP}}\right)^{1/n}\right]$$

$$\boxed{s_m = -\sin\left[(2m-1)\frac{\pi}{2n}\right]\frac{\Gamma^2-1}{2\Gamma} + j\cos\left[(2m-1)\frac{\pi}{2n}\right]\frac{\Gamma^2+1}{2\Gamma}; \quad m = 1, \dots, 2n}$$

poles of a <u>normalized</u> Chebyshev-I filter

↓

the resulting values shall then be multiplied by ω_{BP}

<u>Example</u>: LP FILTER

$\omega_{BP} = 2\pi \cdot 10\,\text{KHz}$ $\quad A_{BP} = 1\text{dB}$

$\omega_{SB} = 2\pi \cdot 50\,\text{KHz}$ $\quad A_{SB} = 30\text{dB}$

$\Big\}\ \xrightarrow{\text{Chebyshev-I}} H(s)$

$\varepsilon_{BP} = 0{,}509 \qquad \varepsilon_{SB} = 31{,}607 \qquad K = 0{,}2 \qquad K_\varepsilon = 0{,}016$

$\longrightarrow n \geq 2{,}1 \implies n = 3$

$\longrightarrow \Gamma = 1{,}61 \longrightarrow s_1 : s_6 \longrightarrow$ take only those with $Re[s_m] < 0$

$$\implies \begin{cases} s_1 = -0{,}247 + j\,0{,}966 \\ s_2 = -0{,}494 + j\,0 \\ s_3 = -0{,}247 - j\,0{,}966 \end{cases}$$

→ real pole (since n is odd)

complex conjugate poles

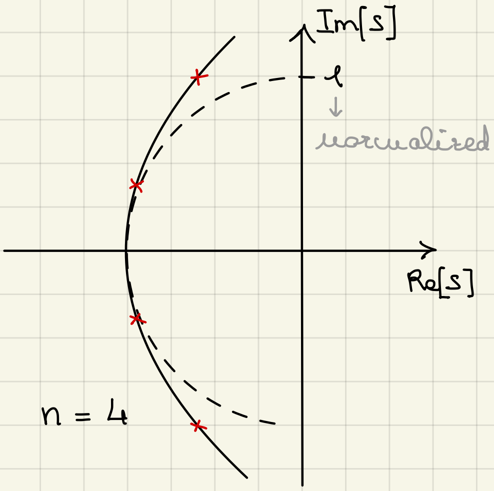
$\implies |p_{13}| = 0{,}997, \qquad Q_{13} = 2{,}018$

$|p_2| = 0{,}494, \qquad Q = 0{,}5$

$$\implies H(j\omega) = \frac{\gamma}{[s + 0{,}494 \cdot \omega_{BP}]\left[s^2 + s\frac{0{,}997\,\omega_{BP}}{2{,}018} + (0{,}997 \cdot \omega_{BP})^2\right]}$$

Note how in this example n(Butterworth) = n (Chebyshev). This means that for these particular requirements there is no big advantage in using a Chebyshev-I filter.

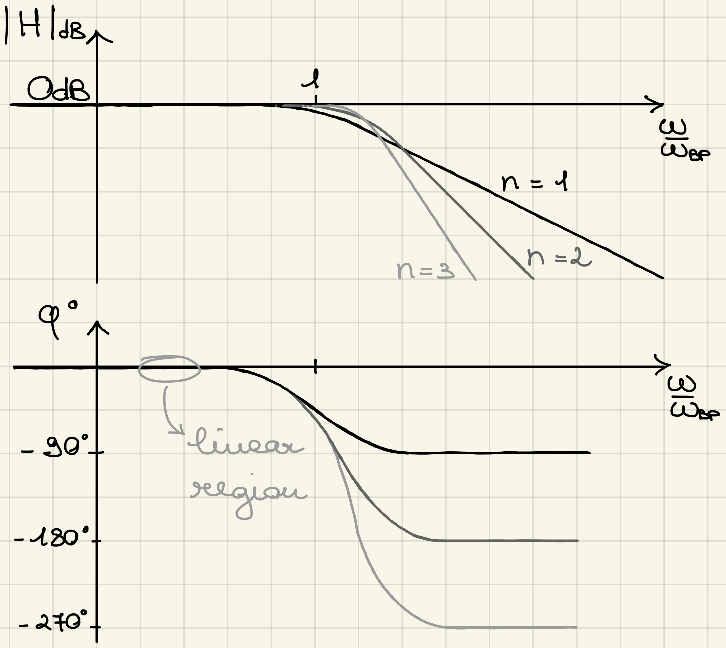
Bessel



The poles in a Bessel approximant are located on a parabola outside the reference band-pass circle; their radial frequencies, just like for Chebyshev, have to be of different values.

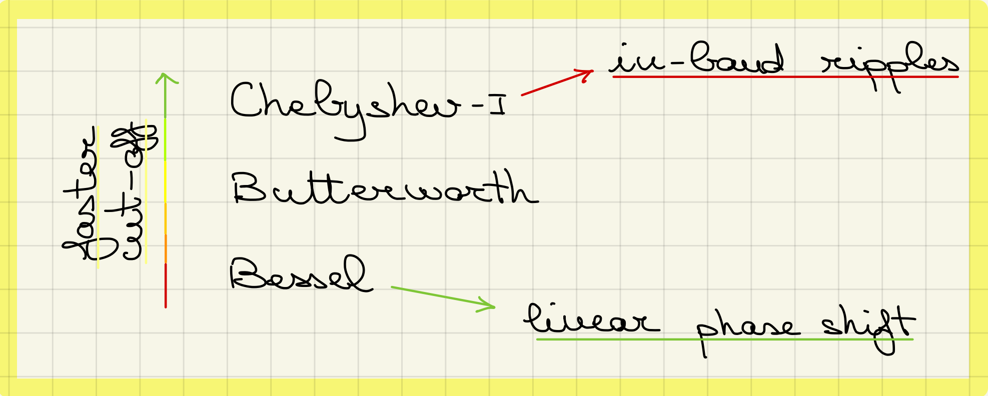
The advantage of using a Bessel model is to have a very linear phase shift.

The Bessel transfer function has its poles at a higher radial frequency (than ω_{BP}) so that the phase shift in-band is better approximated by the linear relation $\varphi = -k\omega$, since they move the non-linear region further from ω_{BP}. However, this approach also moves the cut-off at higher frequencies so the band-pass to stop-band transition will be slower.



Hence the disadvantage of using a Bessel model is to sacrifice a sharp cut-off.

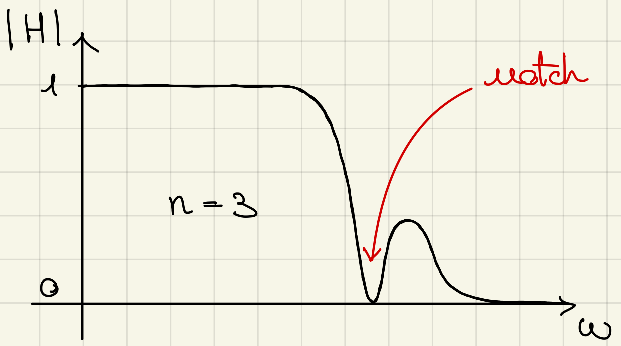
There exist no open form to compute the filter order and poles position. In fact, a table with the typical values is generally used.



Note: the absence of peaks (ripples) in the Butterworth and Bessel transfer functions, even in presence of complex conjugate poles, is due to the attenuation that weak resonance poles (those closer to the Re axis, that do not produce a visible peak) exert on the peaks of strong resonance poles (those closer to the Im axis, that do indeed produce a peak).

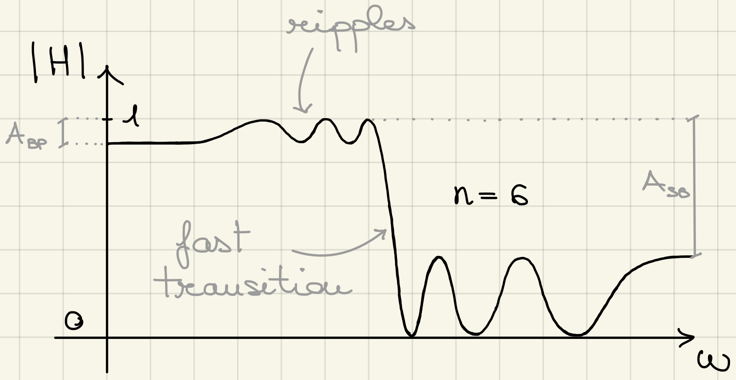
Poles and zeroes continuous time filters

As already stated, poles and zeroes transfer functions are useful to eliminate specific frequency tones out-of-band; they are therefore characterized by the presence of a notch in correspondence with these tones.



Chebyshev type II

No in-band ripples.
Out-of-band ripples
Moderately sharp cut-off
Elliptical placing of the poles.



Cauer

In-band ripples.
Out-of-band ripples.
Very sharp cut-off.
! Relative attenuation is ! limited: $\frac{A_{BP}}{A_{SB}} > K$

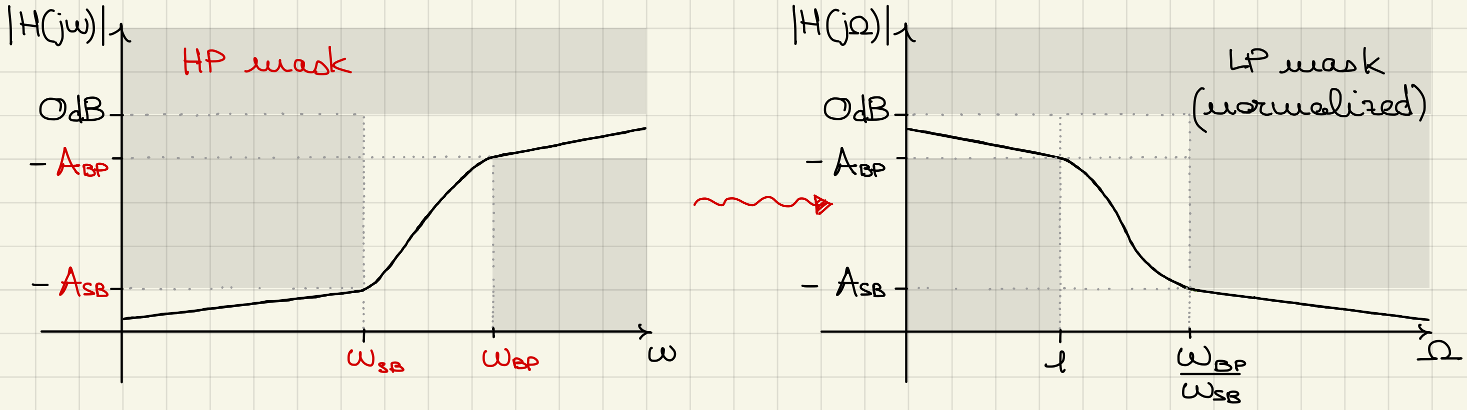
Generalized elliptic solves this issue

So far we have only considered Low-Pass Filters. As already said, it is possible to transform and normalize any High-Pass and Band-Pass Filter into a Low-Pass, find the transfer function parameters with the appropriate model, then denormalize and anti-Transform the result into the original filter type.

High-Pass Filter synthesis

We want to transform high-pass filter mask into low-pass filter mask.

$$\boxed{\Omega = \frac{\omega_{BP}}{\omega}}$$

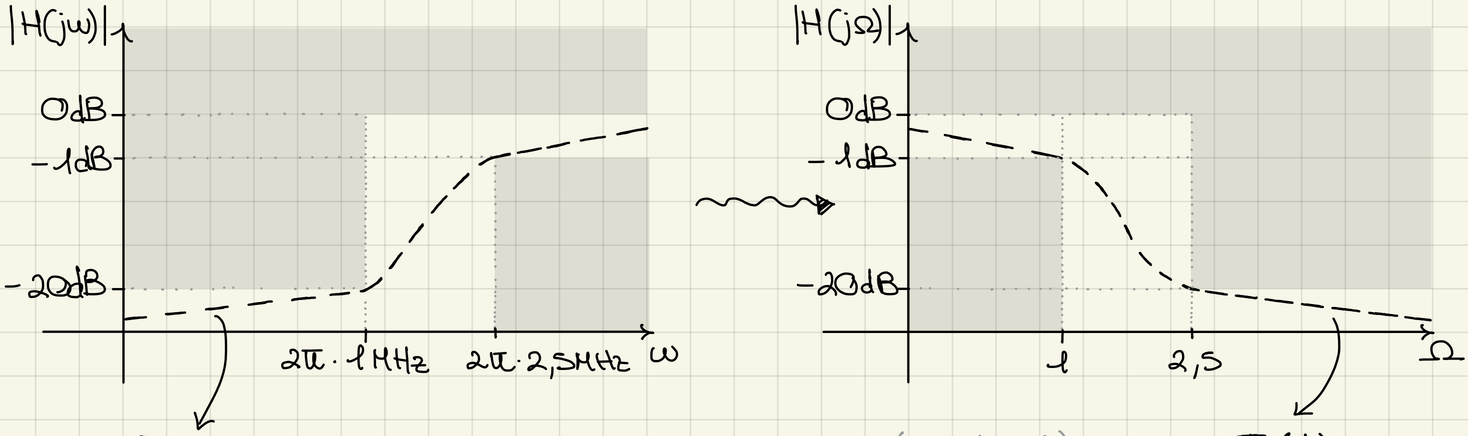


$$\text{High-Pass Filter} \longrightarrow \Omega = \frac{\omega_{BP}}{\omega} \longrightarrow \text{Low-Pass Filter}$$

Example: HP FILTER

$\omega_{BP} = 2\pi \cdot 2{,}5\,\text{MHz}$ $\quad A_{BP} = 1\,\text{dB}$

$\omega_{SB} = 2\pi \cdot 1{,}0\,\text{MHz}$ $\quad A_{SB} = 20\,\text{dB}$



$T(s)$
$\parallel$
$j\omega$

$\Omega = -\frac{\omega_{BP}}{\omega}$ (bilateral)

$$\boxed{\hat{s} = \frac{\omega_{BP}}{s}}$$

$T(\hat{s})$
$\parallel$
$j\Omega$

$\varepsilon_{BP} = \sqrt{10^{1/10} - 1} = 0{,}509$ $\qquad \varepsilon_{SB} = \sqrt{10^{20/10} - 1} = 9{,}94$

$K = \frac{1}{2{,}5} = 0{,}4$ $\qquad K_\varepsilon = \frac{0{,}509}{9{,}94} = 0{,}051$

Butterworth: $n \geq \frac{\ln k_\varepsilon}{\ln k} = 3{,}4 \longrightarrow n = 4$

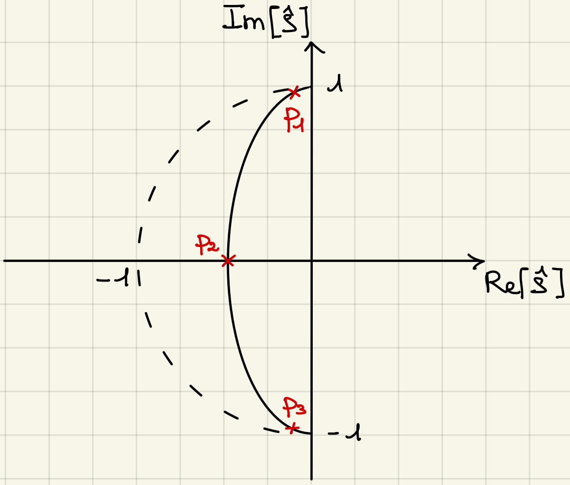
$\underline{\underline{\text{Chebyshev - I}}}$: $n \geq \frac{Ch^{-1}(k_\varepsilon^{-1})}{Ch^{-1}(k^{-1})} = 2{,}33 \longrightarrow \underline{\underline{n = 3}}$

With these requirements, adopting a Chebyshev - I model has the advantage of a lower filter order (circuital implementation will be less complex)

$\longrightarrow \Gamma = 1{,}61$

$\longrightarrow p_{13} = -0{,}247 \pm j\,0{,}966 \qquad |p_{13}| = 0{,}997 \qquad Q_{13} = 2{,}018$

$p_2 = -0{,}494 + j0 \qquad |p_2| = 0{,}494 \qquad (Q_2 = 0{,}5)$



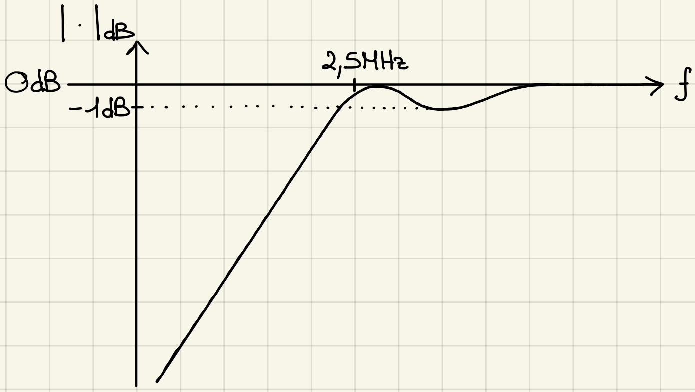
it's a LPF!

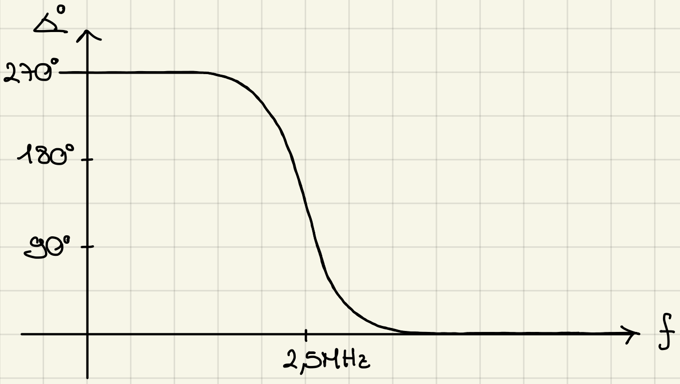
$$\Longrightarrow T(\hat{s}) = \frac{\gamma}{(\hat{s} + |p_2|)\left(\hat{s}^2 + \frac{\hat{s}\,|p_{13}|}{Q_{13}} + |p_{13}|^2\right)} = \frac{|p_2|\,|p_{13}|^2}{(\hat{s} + |p_2|)\left(\hat{s}^2 + \frac{\hat{s}\,|p_{13}|}{Q_{13}} + |p_{13}|^2\right)}$$

$$\Longrightarrow T(s) = \frac{|p_2|\,|p_{13}|^2}{\left(\frac{\omega_{BP}}{s} + |p_2|\right)\left(\frac{\omega_{BP}^2}{s^2} + \frac{\omega_{BP}|p_{13}|}{s\,Q_{13}} + |p_{13}|^2\right)} = \frac{|p_2|\,|p_{13}|^2 \cdot s^3}{(\omega_{BP} + s|p_2|)\left(\omega_{BP}^2 + \frac{\omega_{BP}|p_{13}|s}{Q_{13}} + s^2|p_{13}|^2\right)}$$

it's a HPF!

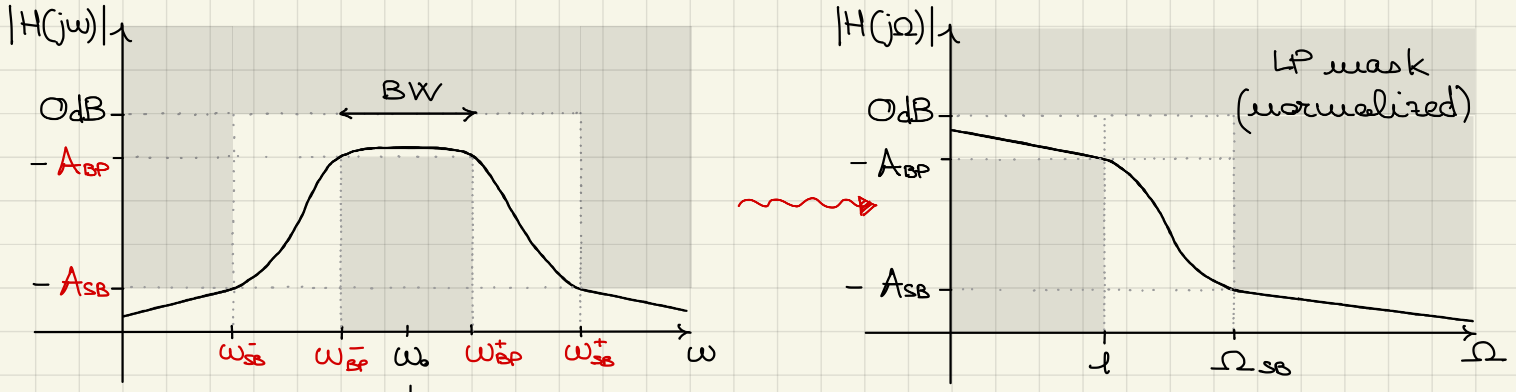
$$= \frac{s^3}{\left(s + \frac{\omega_{BP}}{|p_2|}\right)\left(s^2 + \frac{s\,\omega_{BP}}{|p_{13}|\,Q_{13}} + \frac{\omega_{BP}^2}{|p_{13}|^2}\right)}$$





Band-Pass Filter synthesis

We want to transform band-pass filter mask into low-pass filter mask.



$$\log \omega_0 = \frac{\log \omega_{BP}^{+} + \log \omega_{BP}^{-}}{2} = \frac{1}{2} \log \omega_{BP}^{+} \omega_{BP}^{-}$$

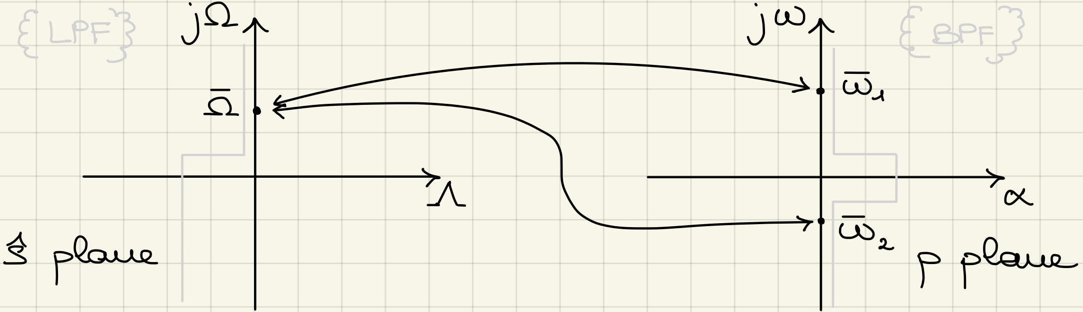
$$\Longrightarrow \boxed{\omega_0 = \sqrt{\omega_{BP}^{-} \omega_{BP}^{+}}}$$ geometric center of the band

$$\boxed{BW = \omega_{BP}^{+} - \omega_{BP}^{-}}$$ bandwidth

$$= \frac{\omega_0}{Q} \Longrightarrow \boxed{Q = \frac{\omega_0}{BW}}$$ Q-factor of the band-pass filter → measure for the filter selectivity

Let's try the following transformation:

$$\boxed{\hat{s} = p + \frac{1}{p}}$$



$$\hat{s} = \frac{p^2 + 1}{p}$$

$$p\hat{s} = p^2 + 1$$

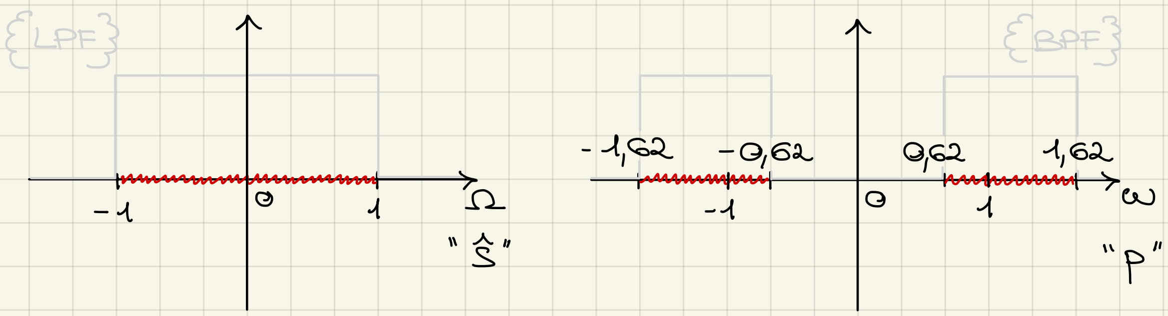
$$p^2 - p\hat{s} + 1 = 0$$

$$p_{1-2} = \frac{\hat{s} \pm \sqrt{\hat{s}^2 - 4}}{2} = \frac{\Lambda + j\Omega \pm \sqrt{(\Lambda + j\Omega)^2 - 4}}{2}$$

We are only considering ŝ points on the imaginary axis:

$$\Lambda = 0 \longrightarrow p_{1-2} = \frac{j\Omega \pm j\sqrt{\Omega^2+4}}{2} = j\,\frac{\Omega \pm \sqrt{\Omega^2+4}}{2} = j\overbrace{\left(\frac{\Omega}{2} \pm \sqrt{\left(\frac{\Omega}{2}\right)^2 + 1}\right)}^{\omega}$$

This transformation links a imaginary value with two new imaginary values.

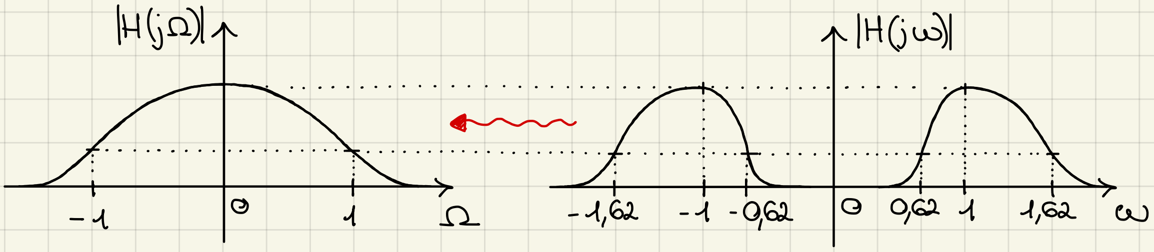


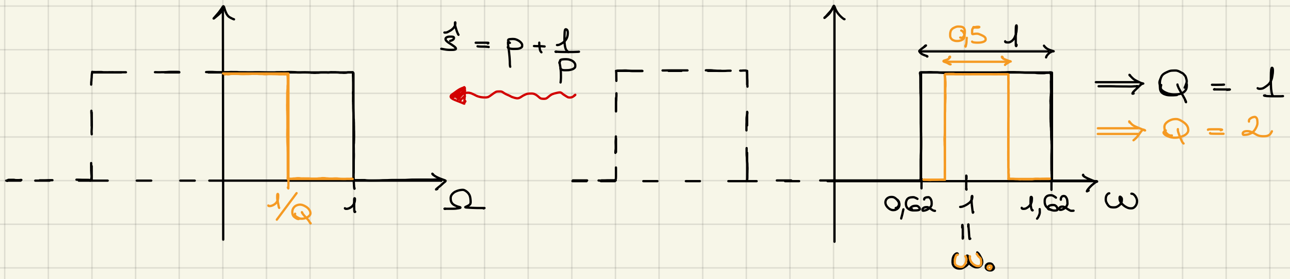
$$\Omega = -1 \qquad \omega = -\frac{1}{2} \pm \frac{\sqrt{5}}{2} = \begin{cases} 0,62 \\ -1,62 \end{cases}$$

$$\Omega = 0 \qquad \omega = \pm 1$$

$$\Omega = 1 \qquad \omega = \frac{1}{2} \pm \frac{\sqrt{5}}{2} = \begin{cases} 1,62 \\ -0,62 \end{cases}$$

$\Longrightarrow$ This is a valuable transformation to map a bilateral band-pass requirement into a bilateral low-pass requirement





This transformation however can only account for a Q-factor equal to 1 (the equivalent LPF gets otherwise denormalized).

$\longrightarrow$ We need to expand the mapped LP transfer function

$$\Longrightarrow \boxed{\hat{s} = Q\left[p + \frac{1}{p}\right]}$$

Of course we first need to normalize the BPF so that the center frequency ω_0 is effectively at 1.

$$\boxed{p = \frac{s}{\omega_0}}$$

$$\longrightarrow \quad \hat{s} = Q\left[\frac{s}{\omega_0} + \frac{\omega_0}{s}\right]$$

$$\boxed{\hat{s} = Q\left[\frac{s^2+\omega_0^2}{s\,\omega_0}\right]}$$

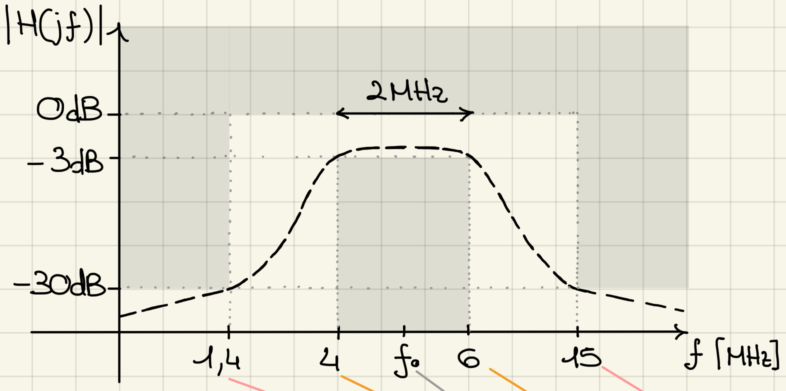
$$\hat{s} = j\Omega, \quad s = j\omega$$

$$\boxed{\Omega = Q\left[\frac{\omega^2-\omega_0^2}{\omega\,\omega_0}\right]}$$

<u>Example</u>: BP FILTER

$\omega_{BP}^{-} = 2\pi \cdot 4\text{MHz}$ $\quad \omega_{BP}^{+} = 2\pi \cdot 6\text{MHz}$ $\quad A_{BP} = 3\text{dB}$

$\omega_{SB}^{-} = 2\pi \cdot 1{,}6\text{MHz}$ $\quad \omega_{SB}^{+} = 2\pi \cdot 15\text{MHz}$ $\quad A_{SB} = 30\text{dB}$



BW = 2MHz

$f_0 = \sqrt{4 \cdot 6}$ MHz = 4,9MHz

$Q = \frac{f_0}{BW} = 2{,}45$

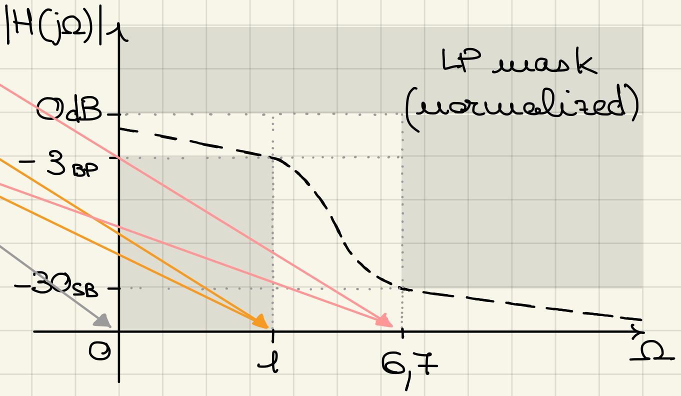
$$\Omega_{BP} = Q\left(\frac{6^2-4{,}9^2}{6\cdot 4{,}9}\right)$$

$$= 2{,}45 \cdot 0{,}41 = 1$$

↳ obvious

$$\Omega_{SB} = 2{,}45\left(\frac{15^2-4{,}9^2}{15\cdot 4{,}9}\right)$$

$$= 6{,}7$$



Note that we are mapping both sides of the BPF to one single HPF. How can this be done?

In principle, for one BP mask there should be two separate LP masks, one for each side.
Both masks must use the same transformation:

$$\Omega = Q\left[\frac{\omega^2-\omega_0^2}{\omega\omega_0}\right]$$

(the left side of the BPF will get negative values for the parameters of its LP mask, which can be neglected and converted into positive values - remember that the transformation is bilateral).

In this specific example, the two masks are exactly the same since we have a geometrically symmetric BPF, that is a filter whose central frequency is the same for both band-pass and stop-band frequencies and whose attenuation is the same on both sides:

$$\omega_0 = \sqrt{\omega_{BP}^{+}\omega_{BP}^{-}} = 2\pi \cdot 4{,}9\text{MHz} = \sqrt{\omega_{SB}^{+}\omega_{SB}^{-}} = 2\pi \cdot 4{,}9\text{MHz}$$

$$A_{BP} = A_{BP}^{-} = A_{BP}^{+} = 3\text{dB} \qquad A_{SB} = A_{SB}^{-} = A_{SB}^{+} = 30\text{dB}$$

For this reason we can use one single mask to map the entire BPF.

$$\Omega_{BP} = Q\left(\frac{\omega_{BP}^{+2}-\omega_0^2}{\omega_{BP}^{+}\omega_0}\right) = Q\left|\frac{\omega_{BP}^{-2}-\omega_0^2}{\omega_{BP}^{-}\omega_0}\right| = 1$$

$$\Omega_{SB} = Q\left(\frac{\omega_{SB}^{+2}-\omega_0^2}{\omega_{SB}^{+}\omega_0}\right) = Q\left|\frac{\omega_{SB}^{-2}-\omega_0^2}{\omega_{SB}^{-}\omega_0}\right| = 6{,}7$$

When the filter is not symmetric, one should consider the mask with the tougher requirements.

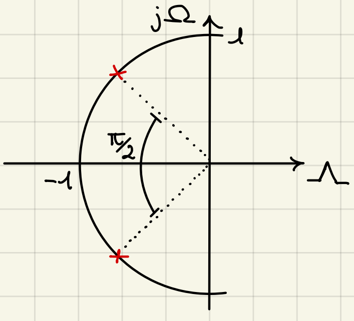
$$\varepsilon_{BP} = \sqrt{10^{3/10}-1} = 0{,}998 \qquad \varepsilon_{SB} = \sqrt{10^{30/10}-1} = 31{,}51$$

$$K = \frac{1}{6{,}7} = 0{,}149 \qquad K_\varepsilon = \frac{0{,}998}{31{,}51} = 0{,}032$$

Butterworth: $n \geqslant \dfrac{\ln 0{,}032}{\ln 0{,}149} = 1{,}81 \longrightarrow \underline{\underline{n=2}}$

Chebyshev - I: $n \geqslant \dfrac{Ch^{-1}(31{,}25)}{Ch^{-1}(6{,}7)} = 1{,}6 \longrightarrow n = 2$

$$Q = 1/\sqrt{2}$$



$$\Longrightarrow \quad T(\hat{s}) = \frac{\gamma}{\hat{s}^2 + \frac{\hat{s}}{Q} + 1} = \frac{1}{\hat{s}^2 + \hat{s}\sqrt{2} + 1}$$

$$\hat{s} = Q\left(\frac{s^2+\omega_0^2}{s\omega_0}\right)$$

$$\Longrightarrow \quad T(s) = \frac{1}{\left[Q\left(\frac{s^2+\omega_0^2}{s\omega_0}\right)\right]^2 + \sqrt{2}\,Q\left(\frac{s^2+\omega_0^2}{s\omega_0}\right) + 1} = \frac{s^2}{D_4(s)}$$

Use Matlab functions:

[0 0 1]   [1 √2 1]

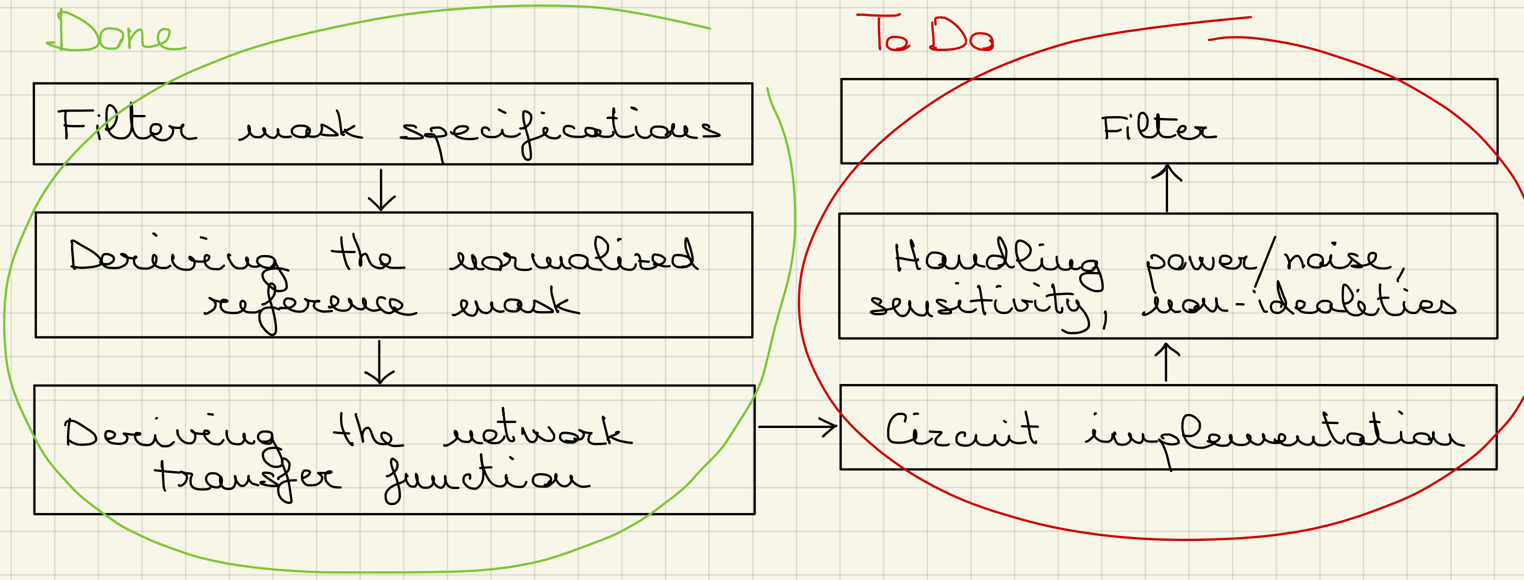
$$[\text{bp\_num},\ \text{bp\_den}] = \text{lp2bp}(\underset{\uparrow}{\text{lp\_num}},\ \underset{\uparrow}{\text{lp\_den}},\ \omega_0,\ BW)$$

bptf = tf(bp_num, bp_den)

pzmap(bptf)

bode(bptf)

$LP(\omega) \longrightarrow LP(\Omega)$	$\hat{s} = \frac{s}{\omega_{BP}}$
$HP(\omega) \longrightarrow LP(\Omega)$	$\hat{s} = \frac{\omega_{BP}}{s}$
$BP(\omega) \longrightarrow LP(\Omega)$	$\hat{s} = Q\,\frac{s^2+\omega_0^2}{s\,\omega_0}$



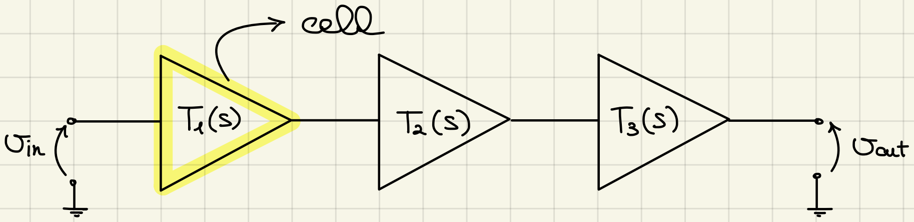
So far we have analyzed the mathematical procedure to obtain the filter transfer function starting from the mask requirements.

We now have to proceed to the electronic implementation of the filter.

Generally speaking, we would expect the filter transfer function to have the following typical form:

$$T(s) = \frac{\gamma \cdot (s+\omega_{z1})(\cdot\cdot\cdot)}{(s+\omega_1)\left(s^2+\frac{s\,\omega_2}{Q}+\omega_2^2\right)(\cdot\cdot\cdot)}$$

Such rational transfer function, whose nominator and denominator are composed by the product of first and second order terms only, suggests to use the cascade of many "Cells" to implement this type of filter.



The idea is that using proper amplifiers it is possible to deliver a signal across the cell independently of the impedance seen at the input or at the output of the cell. So if the cells are properly decoupled from one another we can write the overall transfer function as the product of each single transfer:

$$T(s) = T_1(s)\,T_2(s)\,T_3(s)$$

Each cell therefore has to implement just one singularity at a time: either real (first order) or complex conjugate (second order) singularities (poles or zeroes).

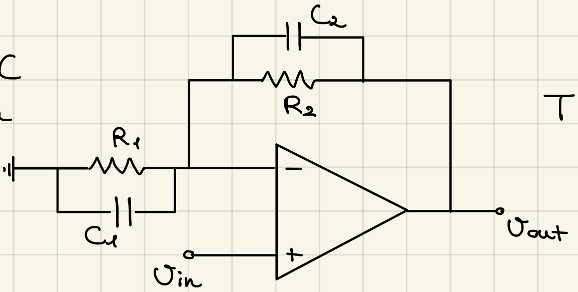
For example: $T(s) = \dfrac{\gamma}{(s+\omega_1)\left(s^2+\frac{\omega_2 s}{Q}+s\right)} = T_1(s)\,T_2(s)$

$$\Longrightarrow \begin{cases} T_1(s) = \dfrac{\gamma_1}{s+\omega_1} \\ T_2(s) = \dfrac{\gamma_2}{\left(s^2+\frac{\omega_2 s}{Q}+s\right)} \end{cases}$$

"Biquad cell"

First order cell

(a simple RC network can also work)

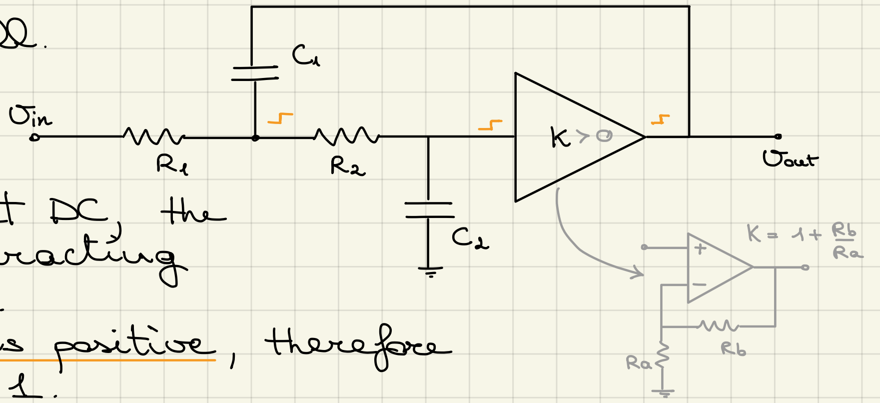


$$T(s) = \underset{\downarrow \left(1+\frac{R_2}{R_1}\right)}{T_0} \frac{1+s(C_1+C_2)(R_1 /\!/ R_2)}{1+sC_2R_2}$$

Sallen Key cell

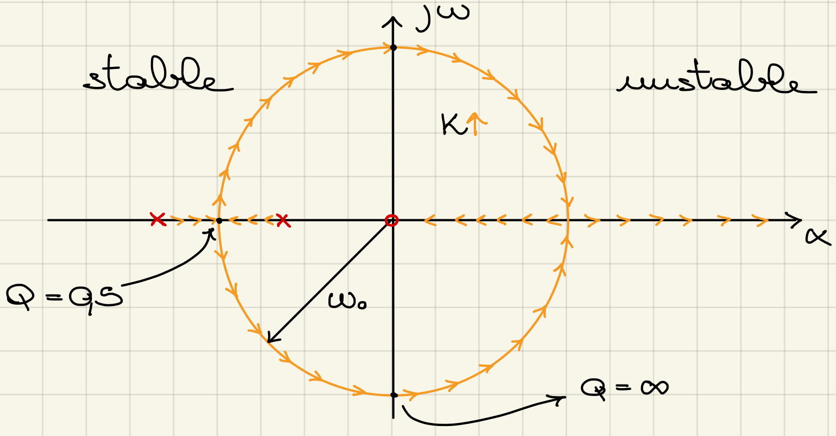
{this cell implements a LPF; switch C with R to implement a HPF}

It's a second order cell.



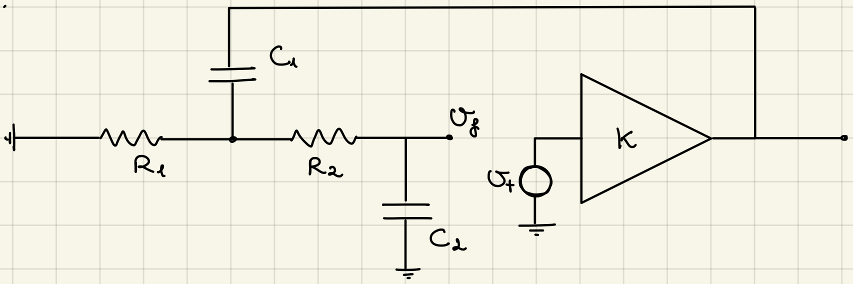
DC gain is equal to K.
C1 introduces a zero at DC, the two capacitors are interacting and introduce two poles.
Note that the feedback is positive, therefore G_{loop} must be lower than 1.

(we assume the GBWP of the inner loop to be much larger than the poles of the entire circuit)



This configuration is very convenient because it allows with just one amplifier to have a fixed radial frequency ω_0 while freely setting the Q factor of the pole pair (it decouples Q from ω_0).

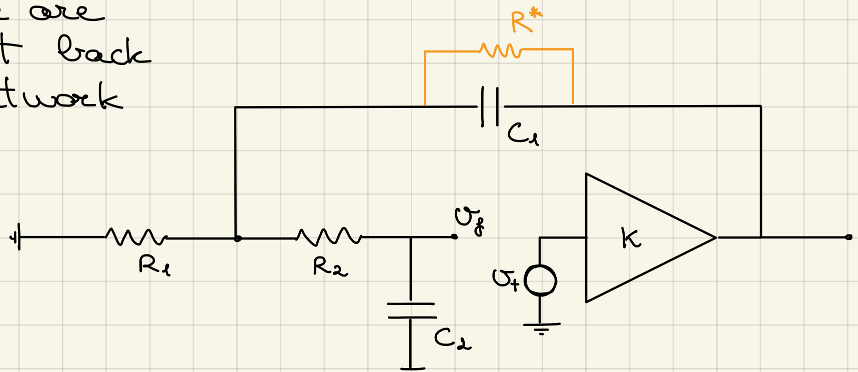
Let's compute Gloop to exactly determine the position of the poles!



$$G_{loop}(s) = \gamma \frac{a_1 s \cancel{+1}}{b_1 s^2 + b_1 s + 1}$$

Watch out for the zero in DC! the usual form only accounts for finite zeroes (there should be no +1)

Need to modify the circuit so that the usual form still holds, but in such a way that we are then able to revert back to the original network



$$G_{loop}(s) = \lim_{R^* \to \infty} G^*_{loop}(s)$$

$$\gamma = K \frac{R_1}{R_1 + R^*}$$

$$b_1 = (R^* /\!/ R_1) C_1 + (R_2 + R_1 /\!/ R^*) C_2$$

$$b_2 = C_1 C_2 (R_1 /\!/ R^*) R_2$$

$$a_1 = R^* C_1$$

$R^* \to \infty$

$$\Longrightarrow \quad G^*_{loop} = K \frac{R_1}{\cancel{R_1 +} R^*} \frac{\cancel{1+} s C_1 \cancel{R^*}}{(R_1 \cancel{/\!/ R^*}) R_2 C_1 C_2 s^2 + [(R_1 \cancel{/\!/ R^*})(C_1 + C_2) + R_2 C_2] s + 1}$$

$$\Longrightarrow \quad G_{loop} = K \frac{s C_1 R_1}{R_1 R_2 C_1 C_2 s^2 + [R_1 C_1 + (R_1 + R_2) C_2] s + 1}$$

Closed loop poles ⟷ $G_{loop}(s) = 1$

$$s^2 R_1 R_2 C_1 C_2 + s [(1-K) C_1 R_1 + C_2 (R_1 + R_2)] + 1 = 0$$

compare it with the canonical form:

$$\frac{s^2}{\omega_0^2} + \frac{s}{\omega_0 Q} + 1 = 0$$

$$\Longrightarrow \quad \boxed{\omega_0 = \frac{1}{\sqrt{R_1 R_2 C_1 C_2}}} \qquad \boxed{Q = \frac{\sqrt{R_1 R_2 C_1 C_2}}{(1-K) C_1 R_1 + C_2 (R_1 + R_2)}}$$

Note how ω_0 is directly dependent on the value of resistors and capacitors of the circuit, whose actual value can fluctuate due to tolerance issues.

This is a typical problem of analog filters, since their parameters depend on the real values of their components. Temperature variation, process non-uniformities and so on can cause the target specifications to differ from the implemented performances.

What is usually done to cope with this is to implement an auxiliary system whose role is to check what is the actual value of the resonance frequency, compare it to the desired one and accordingly fix it so that they match (in a sort of negative feedback fashion).

The Sallen Key cell, however, has the merit of having a Q factor of its poles that is very robust with respect to variability of its components:

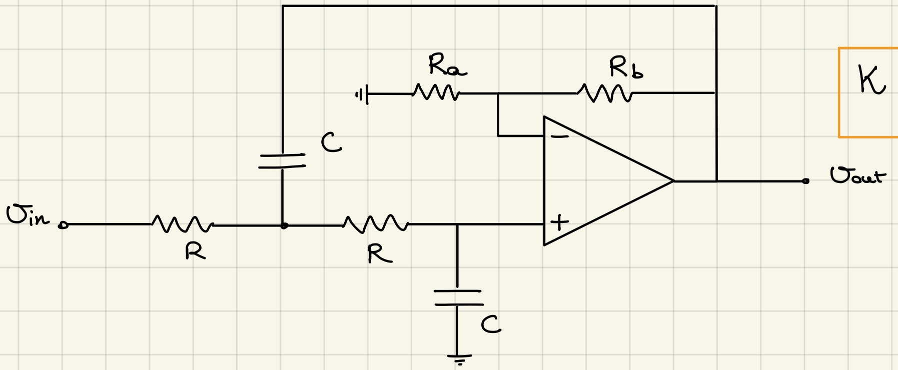
$$Q = \frac{1}{(1-K)\left[\sqrt{\frac{C_1R_1}{C_2R_2}} + \sqrt{\frac{C_2}{C_1}}\left(\sqrt{\frac{R_1}{R_2}} + \sqrt{\frac{R_2}{R_1}}\right)\right]}$$

In this form it can be easily seen that Q depends solely on the relative variation between each component (rather than the absolute variation, like it was for ω_0) which can be easily be controlled through proper fabrication techniques - such as the common centroid geometry - and is therefore much more reliable.

A disadvantage of this cell is that it has quite many components (higher cost). It would be nice if they could somehow be reduced.

→ Use only one type of resistor and capacitor ($R_1 = R_2 = R$ and $C_1 = C_2 = C$)

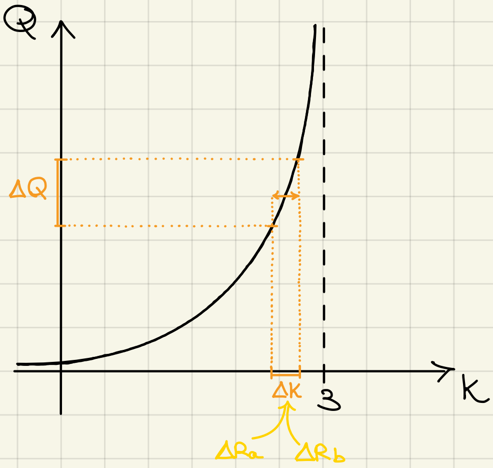
$$\Rightarrow \boxed{\omega_0 = \frac{1}{RC}} \qquad \boxed{Q = \frac{1}{3-K}}$$

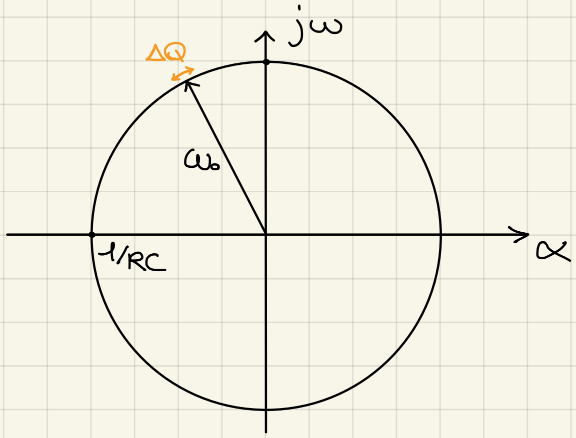


$$\boxed{K = 1 + \frac{R_b}{R_a}}$$

Mind that even if Q would seem to have lost any dependency on analog components (and their variability) it is still a function of K which depends on the ratio of two resistances.

When $K = 1+\frac{R_b}{R_a}$ gets closer to 3, Q tends to infinity and a small fluctuation of K can cause a huge variation of Q.





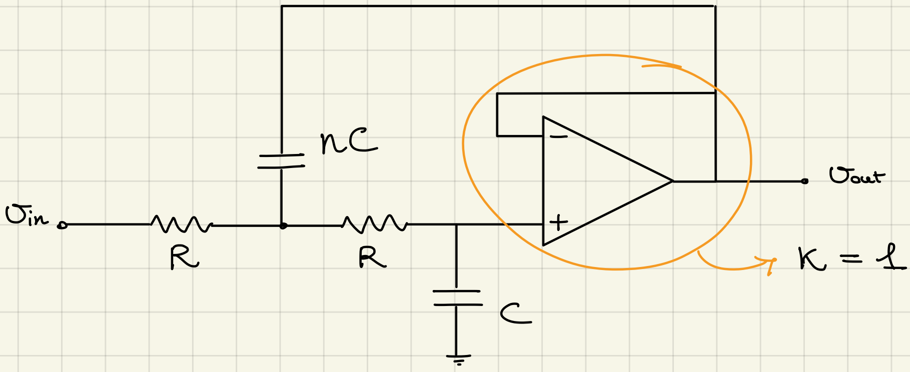
This filter is, for this reason, not reliable in those cases when a Q factor bigger than ~1 (K>2) is needed.

An alternative solution would then be to remove the dependance of Q over K, by setting $K \equiv 1$ fixed, and allow for different values of resistors and/or capacitors (so to still have a degree of freedom to set Q and ω_0):

$$Q = \frac{\sqrt{R_1 R_2 C_1 C_2}}{\cancel{(1-\underset{1}{\underset{\|}{K}})C_1R_1} + C_2(R_1+R_2)} = \sqrt{\frac{C_1}{C_2}}\,\frac{\sqrt{R_1R_2}}{(R_1+R_2)}$$

We can choose to have the capacitors ratio equal to a fixed constant n ($C_1 = nC$, $C_2 = C$) while the resistors are exactly equal to each other ($R_1 = R_2 = R$)

$$\Longrightarrow \quad \boxed{\omega_0 = \frac{1}{RC\sqrt{n}}} \qquad \boxed{Q = \frac{\sqrt{n}}{2}}$$



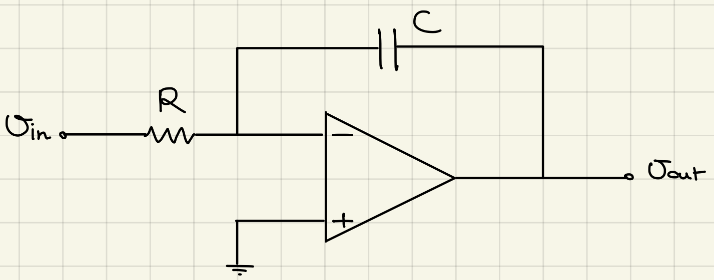
In this way Q only depends on the square root of n, which means that it is quite robust to the variability of the filter components.

This kind of approach is best suited when a pole pair with a large Q factor is required

Anyhow there is a drawback with this design which is the increasing area of the capacitor needed to match a larger value of Q (since $n = 4Q^2$, a quality factor twice as big requires a capacitor four times as big).

Universal cell

Let's consider the ideal integrator configuration:



$$\frac{v_{out}}{v_{in}} = -\frac{1}{sRC} = -\frac{\omega_0}{s}$$

$$\omega_0 = \frac{1}{RC}$$

As a standalone piece of circuitry, the ideal integrator doesn't do much because of non-idealities such as voltage offset and current bias causing the output to inevitably saturate.

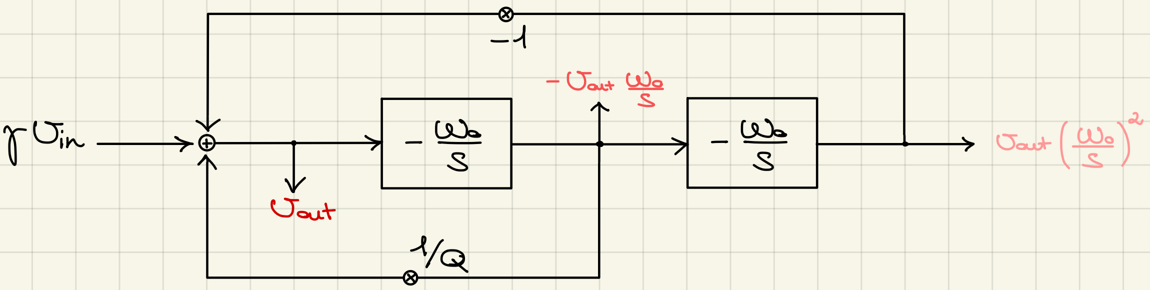
Nonetheless it can still be very useful when adopted as a building block for transfer functions.

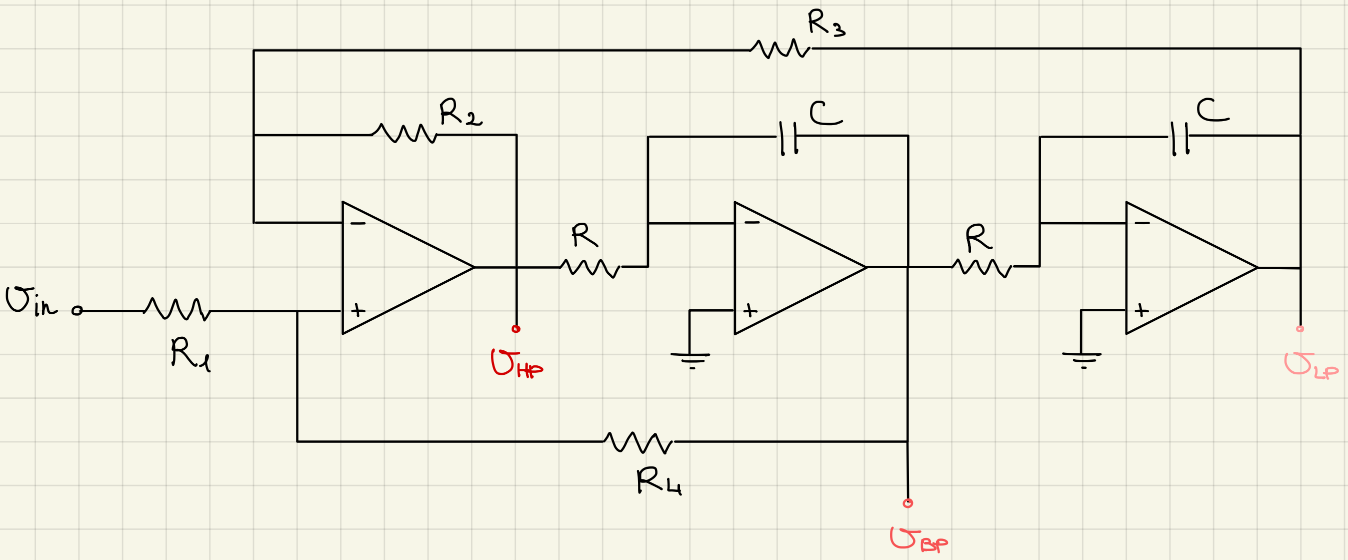
E.g.: $T(s) = \frac{\gamma s^2 / s^2}{\left(s^2 + \frac{s\omega_0}{Q} + \omega_0^2\right)/s^2}$ transfer function of a HPF

$$v_{HP} \leftrightarrow \frac{v_{out}}{v_{in}} = \frac{\gamma}{1 + \frac{\omega_0}{Qs} + \frac{\omega_0^2}{s^2}}$$

$$v_{out}\left(1 + \frac{\omega_0}{Qs} + \frac{\omega_0^2}{s^2}\right) = v_{in}\gamma$$

$$v_{out} = v_{in}\gamma - v_{out}\left(\frac{\omega_0}{sQ}\right) - v_{out}\left(\frac{\omega_0^2}{s^2}\right)$$





The network provides three different output each corresponding to a different filter shape of the common input.
This is why it's called Universal cell.

Each filter output depends on the same radial frequency $\omega_0 = 1/RC$ and quality factor Q.

In order to have a feedback branch from LP to HP with a gain equal to -1, as demanded by the calculations, R2 and R3 must be equal $\left(v_{HP}\Big|_{LP} = -\frac{R_2}{R_3} \cdot v_{LP} \right)$.

To compute the value of R1 and R4, we apply the superposition effect:

$$v_{HP}\Big|_{in} = \left(\frac{R_4}{R_1+R_4}\right)\left(1+\frac{R_2}{R_3}\right) v_{in} = \frac{2R_4}{R_1+R_4} v_{in}$$

$$v_{HP}\Big|_{BP} = \left(\frac{R_1}{R_1+R_4}\right)\left(1+\frac{R_2}{R_3}\right) v_{in} = 2\frac{R_1}{R_1+R_4} v_{BP}$$

$$v_{HP}\Big|_{LP} = \left(-\frac{R_2}{R_3}\right) v_{LP} = -v_{LP}$$

$$\Longrightarrow \underset{\overset{\updownarrow}{v_{out}}}{v_{HP}} = \underbrace{\frac{2R_4}{R_1+R_4}}_{\gamma} v_{in} + \underbrace{\frac{2R_1}{R_1+R_4}}_{1/Q} v_{BP} - v_{LP}$$

$$\longrightarrow \boxed{Q = \frac{1+R_4/R_1}{2} \qquad \gamma = \frac{R_4}{R_1}\cdot\frac{1}{Q}}$$

→ once Q is set through the ratio R_4/R_1, γ is also set and cannot be changed!

There is only one degree of freedom $\left(\frac{R_4}{R_1}\right)$ for two variables (Q and γ).

! γ and Q cannot be independently set !

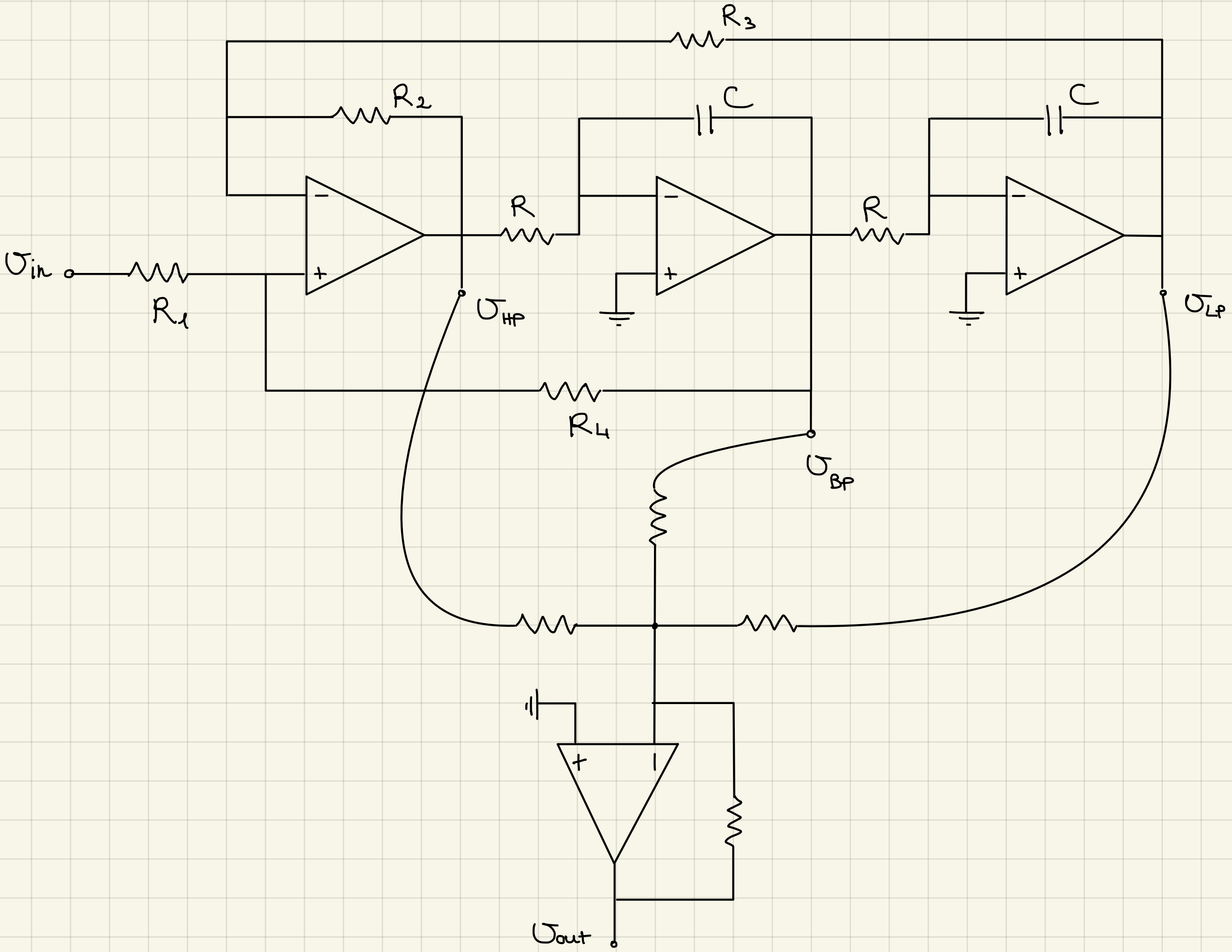
This configuration has the great advantage of providing a high-pass, band-pass and low-pass filter in parallel while using just three OPAMPs; it has however the minor drawback that the gain of the filter cannot be set (which is not a big deal since an amplifying circuit can do it in its place).

Another advantage of this cell is that it is useful for the implementation of poles and zeroes transfer functions.

$$T(s) = \frac{s^2 + s\,\omega_z/Q_z + \omega_z^2}{(s^2 + s\,\omega_0/Q + \omega_0^2)} =$$

$$= \frac{s^2}{(\cdots)} + \frac{s\,\omega_z/Q_z}{(\cdots)} + \frac{\omega_z^2}{(\cdots)}$$

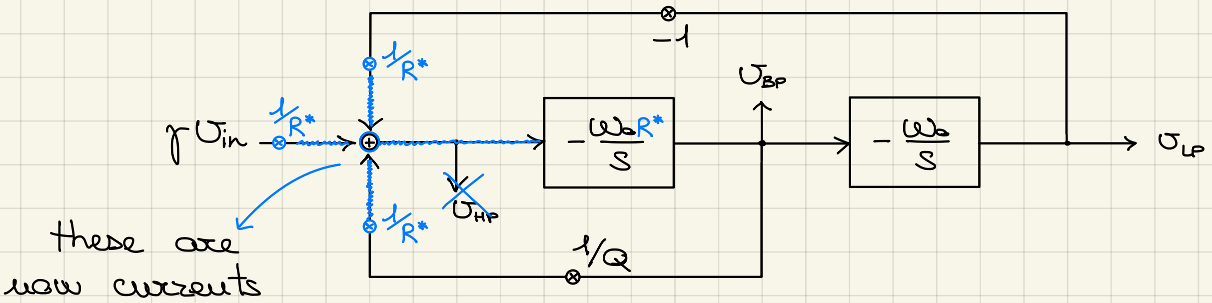
$$v_{out} = v_{HP} + v_{BP}\beta_1 + v_{LP}\beta_2$$



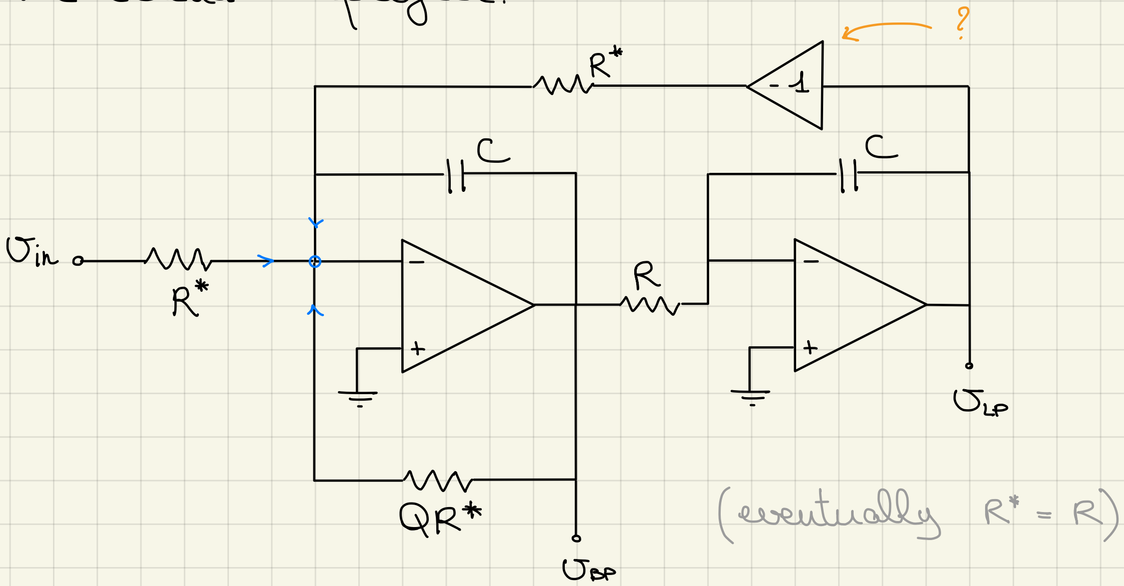
We can now think about how to improve this cell. One way could be to remove the first amplifier (which is used as a voltage summing node) and replace it with a current summing node

Tow Thomas cell

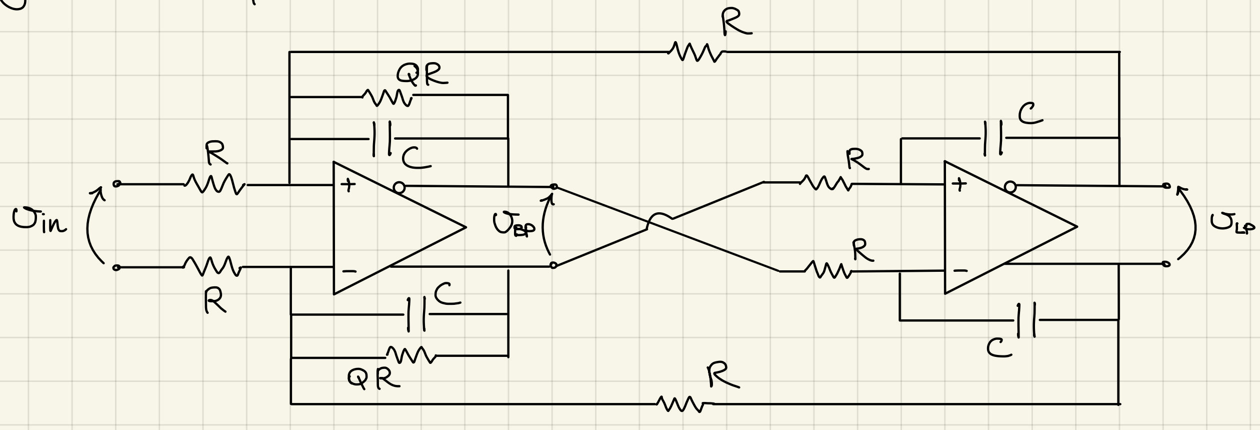
The idea is to sum the various contribution of v_{HP} in the Universal cell in the form of current instead of tension, so to avoid using an OPAMP.



The current summing node can be the virtual ground of the second amplifier:



Apparently this solution does not offer any advantage since now the gain of v_{LP} has to be implemented through an inverting stage (i.e. another amplifier).
This issue doesn't actually exist however: commercial OPAMPs typically have a differential output (fully differential amplifier), so in order to achieve a gain equal to -1 it is sufficient to cross the polarities of the output.

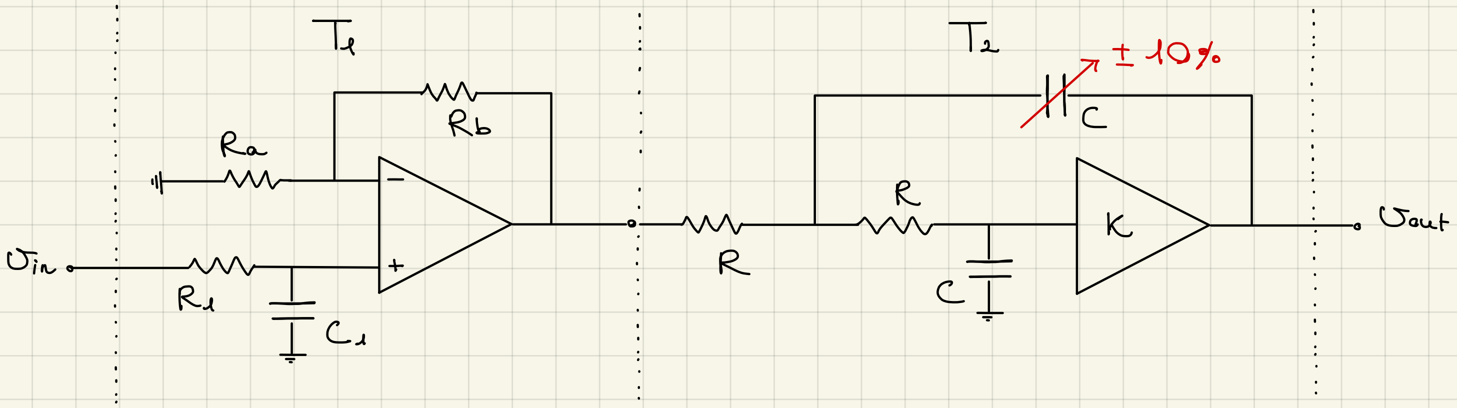


An alternative to active cells for the building of a filter transfer function is Ladder Networks.

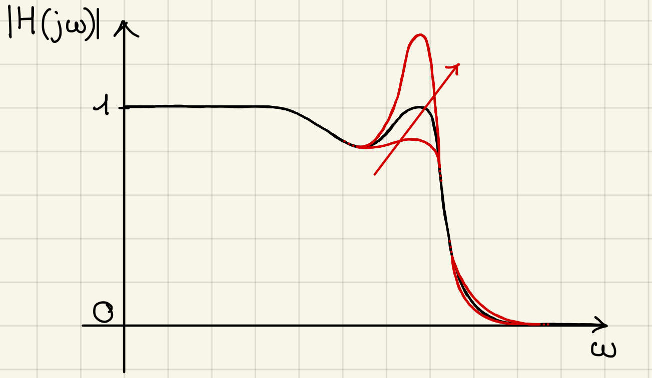
The advantage of using ladder networks instead of cells is the improved robustness of the resulting filter with respect to components variability.

Example: consider a Chebyshev type I transfer function of order n = 3; using active cells, the filter can be implemented with the cascade of, for instance, one first order cell and one Sallen key cell:

$$\frac{v_{out}}{v_{in}} = T(s) = \gamma \frac{1}{(s+\omega_1)} \frac{1}{\left(s^2 + \frac{s\,\omega_2}{Q} + \omega_2^2\right)} = T_1(s) \cdot T_2(s)$$



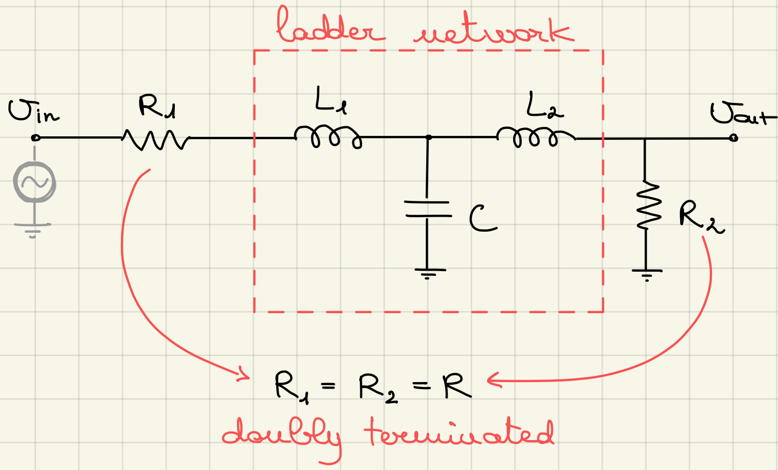
However, this solution strongly depends on the tolerance of its components. If we were to consider a variability of ± 10 % in the value of the Sallen key feedback capacitor, the resulting filter shape would be greatly impaired:

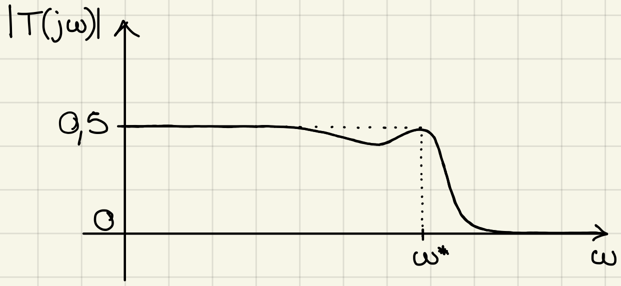


Adopting a ladder network topology for the filter implementation allows, as we will see, for a much more limited variation of the filter transfer function when one of its components has some fluctuations.

The idea behind ladder networks comes from passive networks, that are made only of resistors, capacitors and inductors.

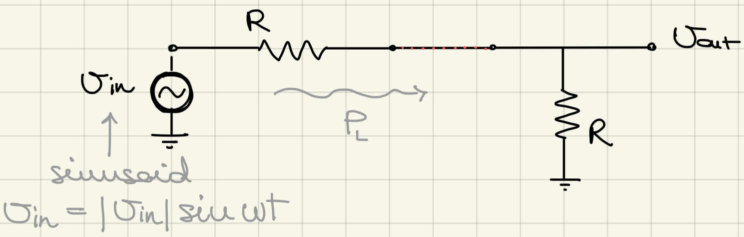
Using a passive network to implement the filter from the previous example, it could be built as follows:





The ladder network (the reactive part of the circuit) adds three poles to the TF, which can be adequately adjusted to match the filter specs, while having a DC gain exactly equal to 1 (also the gain at $\omega = \omega^*$ will be equal to 1 since we are implementing a Chebyshev-I TF with in-band ripples)

It is important for the network to be doubly terminated as it allows to have the maximum possible power transfer at DC (and at $\omega = \omega^*$) from input to output.



@ DC or $\omega = \omega^*$:

max. value of the sinusoid

$$V_{out} = \frac{V_{in}}{2}$$

$$P_L = \left(\frac{V_{in}}{2}\right)^2 \frac{1}{2R} = \boxed{\frac{|V_{in}|^2}{8R} = P_{L_{max}}}$$

maximum deliverable power for a resistive network

So generally speaking, for any doubly terminated ladder network, there will be some frequencies (in this case, $\omega = 0$ and $\omega = \omega^*$) that grant maximum power transfer from input to output and for which the TF reaches its peak value.

Let's consider the dependency of the output power on the frequency of the input signal:

$$P_L(\omega) = \frac{|V_{out}|^2}{2} \cdot \frac{1}{R} = \frac{|V_{in}|^2 |T(j\omega)|^2}{2R}$$

$$\left[\frac{\partial P_L}{\partial \omega}\bigg|_{\omega=\omega^*} = 0\right] \longrightarrow \frac{|V_{in}|^2}{2R} \cdot 2\,|T(j\omega^*)|\,\frac{\partial |T(j\omega)|}{\partial \omega}\bigg|_{\omega=\omega^*} = 0$$

max. at $\omega = \omega^*$

$$\Longrightarrow \left[\frac{\partial |T(j\omega)|}{\partial \omega} \bigg|_{\omega=\omega^*} = 0 \right]$$

As it was expected, the peak in the delivered power corresponds to the peak in the transfer function.

Even though the result is obvious, it must be noted that it holds for a TF that is a function of its network parameters as well:

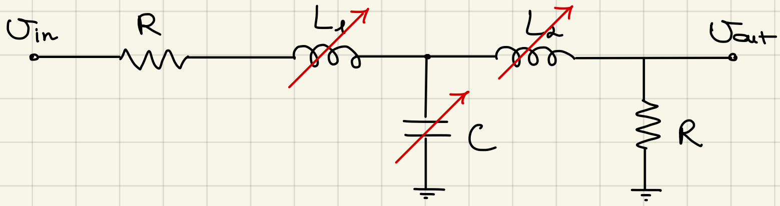
$$T = T(j\omega, x)$$

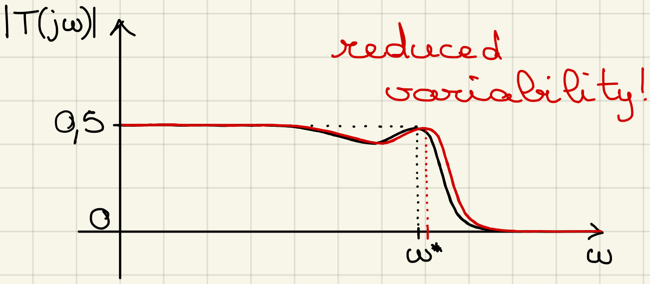
L_1, L_2, C

$$P_L = P_L(\omega, x)$$

$$\frac{\partial P_L}{\partial x}\bigg|_{\omega=\omega^*} = 0 \Longrightarrow \left[\frac{\partial |T(j\omega, x)|}{\partial x}\bigg|_{\omega=\omega^*} = 0 \right]$$ Orchard theorem

This means that if the capacitor or the inductors were to slightly differ from their nominal value, the transfer function would not be changing much around ω^* (at most it would shift a little since ω^* would move depending on L and C).





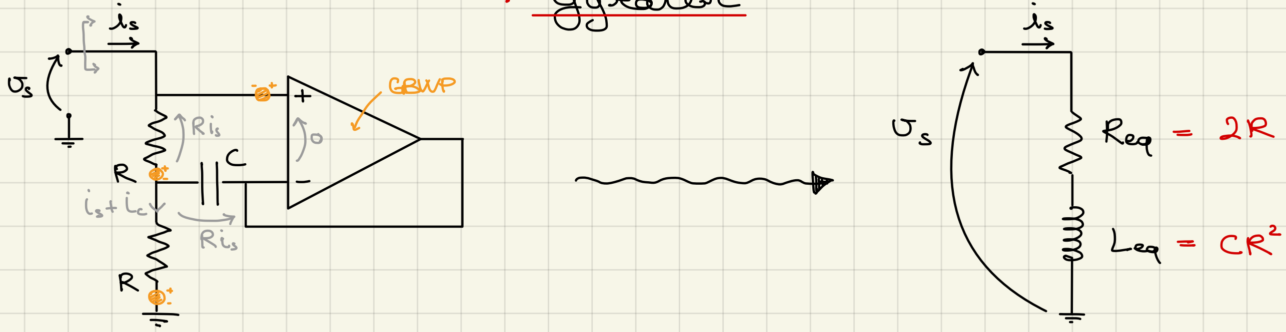
A more intuitive explanation for the reduced variability of the transfer function can be understood considering that in a ladder network all components are coupled and interacting with one another, so the variation of one parameter won't affect just one pole, causing the TF to deform, but rather it will affect the TF in its entirety (causing the aforeseen shift); whereas in a cells cascade fluctuations of a single cell parameters won't be "seen" by other cells causing the TF to deform in those points where the fluctuating cell placed its singularities.

Therefore, for high order filters, it is usually recommended to adopt a ladder network implementation since the use of many active cells could heftly impair the variability of the resulting transfer function.

Issue: inductors in integrated circuits

We need to implement ladder networks without using inductors (which are practically impossible to have in integrated technologies).

We can mimic the behaviour of an inductive impedance through an active network → Gyrator



$$Z_{in} = \frac{v_s}{i_s}$$

$$v_c = R i_s \qquad i_c = sC v_c = sCR i_s$$

$$v_s = R i_s + (i_s + i_c)R = 2R i_s + sCR(i_s R)$$
$$= i_s R(2 + sCR)$$

$$\longrightarrow Z_{in} = 2R + sCR^2 = R_{eq} + sL_{eq}$$

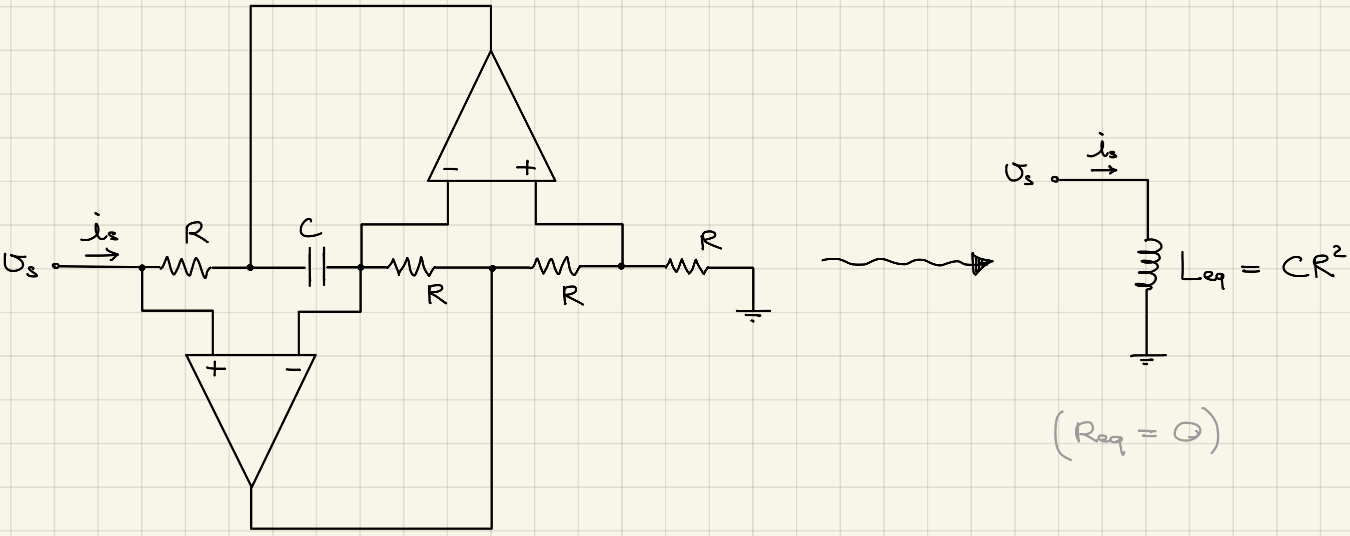
E.g.: $R = 10k\Omega$ $C = 10pF$

$\longrightarrow R_{eq} = 20k\Omega$

$\longrightarrow L_{eq} = 1mH$ ↓ Huge!

The gyrator can mimic an inductive impedance whose size could never be obtained with real inductors (in integrated circuits).

This is not the only gyrator topology but there exist many more with different characteristics:



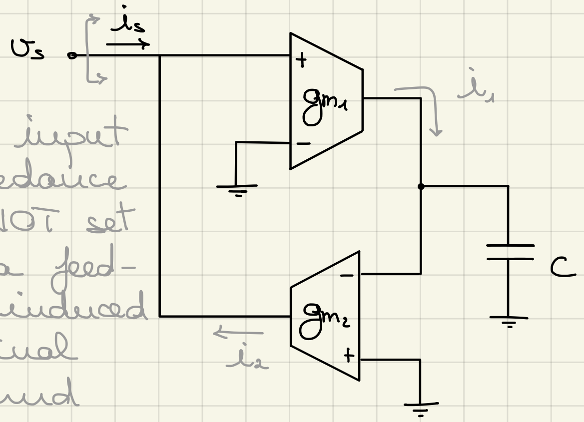
There are of course some limitations of using an active network instead of a proper inductor:

1) the inherent bandwidth limitation of a feedback circuit; the active network must work with frequencies much below the GBWP otherwise it loses its inductive behaviour

2) the noise introduced by the non-reactive components; an ideal inductor would be noiseless, while the gyrator has resistors and amplifiers both contributing with their own noise

To solve the first issue we should look for "feedback-less" gyrator topologies, such as the following one:

i.e. w/o virtual ground



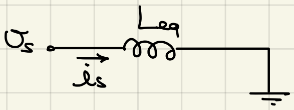
the input impedance is NOT set by a feed-back induced virtual ground

$$i_1 = g_{m_1} v_s \qquad v_2^- = \frac{i_1}{sC}$$

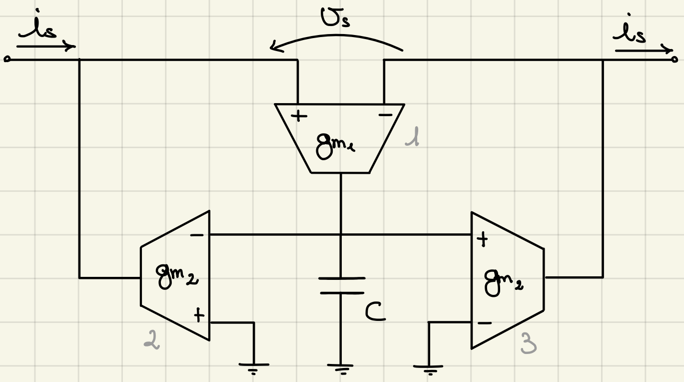
$$i_2 = -g_{m_2} v_2^- = -\frac{g_{m_1} g_{m_2}}{sC} v_s = -i_s$$

$$\longrightarrow Z_{in} = \frac{v_s}{i_s} = \frac{sC}{g_{m_1} g_{m_2}}$$

$$\rightsquigarrow L_{eq} = \frac{C}{g_{m_1} g_{m_2}}$$



To obtain an equivalent inductor between two nodes (so far it was only between one node and ground) the following topology can be used:

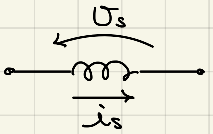


$$v_c = \frac{i_1}{sC} = \frac{v_s g_{m_1}}{sC}$$

$$i_s = g_{m_2} v_c = v_s \frac{g_{m_1} g_{m_2}}{sC}$$

$$\longrightarrow Z_{in} = \frac{sC}{g_{m_1} g_{m_2}}$$

$$\rightsquigarrow L_{eq} = \frac{C}{g_{m_1} g_{m_2}}$$

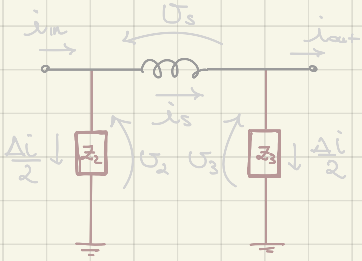


The major problem with this configuration is that the transconductances of the two OTAs must exactly match. In case of a mismatch, the equivalent impedance will not be just an inductor:

$$g_{m_2} = g_m + \frac{\Delta g_m}{2} \qquad i_{in} = g_{m_2} v_c = \underbrace{v_s \frac{g_{m_1} g_m}{sC}}_{i_s} + \underbrace{v_s \frac{g_{m_1} \Delta g_m}{sC\,2}}_{\Delta i/2}$$

$$g_{m_3} = g_m - \frac{\Delta g_m}{2} \qquad i_{out} = g_{m_3} v_c = v_s \frac{g_{m_1} g_m}{sC} - v_s \frac{g_{m_1} \Delta g_m}{sC\,2}$$

$$i_{in} - i_{out} = v_s \frac{g_m \Delta g_m}{sC} = \Delta i$$



$$\Delta i = i_s \cdot \frac{\Delta g_m}{g_m} = \frac{v_s}{sL_{eq}} \frac{\Delta g_m}{g_m}$$

$$Z_2 = \frac{v_2}{\Delta i/2} \qquad Z_3 = \frac{v_3}{\Delta i/2}$$

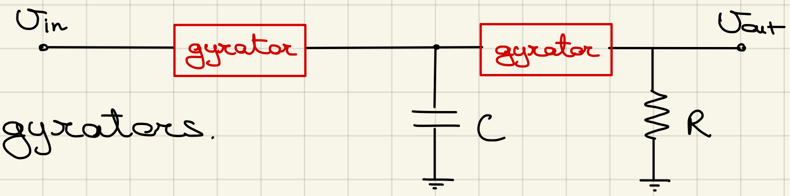
$$v_2 = v_{cm} + \frac{v_s}{2} \qquad v_3 = v_{cm} - \frac{v_s}{2}$$

$$Z_2 = \frac{v_{cm}}{\Delta i/2} + \frac{v_s}{\Delta i} = s \frac{v_{cm}}{v_s} \cdot \frac{2L_{eq}}{\Delta g_m/g_m} + \frac{sL_{eq}}{\Delta g_m/g_m}$$

$$Z_3 = s \frac{v_{cm}}{v_s} \cdot \frac{2L_{eq}}{\Delta g_m/g_m} - \frac{sL_{eq}}{\Delta g_m/g_m}$$

Other issues of these gyrator configurations is the finite output resistance of the OTAs (so for instance the inductive zero will not be exactly in the origine but at a low, finite frequency)

Doubly terminated ladder network from previous example using gyrators.



So far, we assumed that a ladder network could only be implemented as a passive network, hence our discussion about gyrators and imitation of inductive impedances.
Is it possible to obtain a circuit that operates in the exact same way as a ladder network, retaining the same transfer function as well as its robustness with respect to the variability of its parameters, but that does not make use of inductors at all?

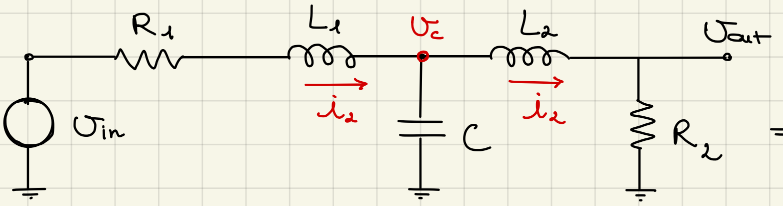
The objective is to use only resistors, capacitors and active components where needed to implement a whole new network whose transfer function is the exact same as that of a ladder network.
The starting point of this approach to filter synthesis is the derivation of the links between the state variables of the original ladder network.

↓

electrical variables related to the energy stored in the network

Energy of a network ⟷ voltage across capacitors, current along inductors

state variables



$$\mathcal{E} = \mathcal{E}(i_1, i_2, v_c)$$

⟹ We need to find 3 independent equations that link the 3 state variables of the system: i_1, i_2 and v_c.

1. $i_1 = \dfrac{v_{in} - v_c}{R + sL_1}$

2. $\dfrac{v_c - v_{out}}{sL_2} = i_2$

3. $i_1 - i_2 = v_c sC$

anything related to the network performance and to the overall transfer function is within these links

We can try to obtain these equations using ideal integrators as building blocks (similarly to what we did for the universal cell).

It is better to first convert all variables to the same physical quantity (voltage for instance):

$$\left[i_1 = \frac{v_1}{R^*} \right] \qquad \left[i_2 = \frac{v_2}{R^*} \right]$$

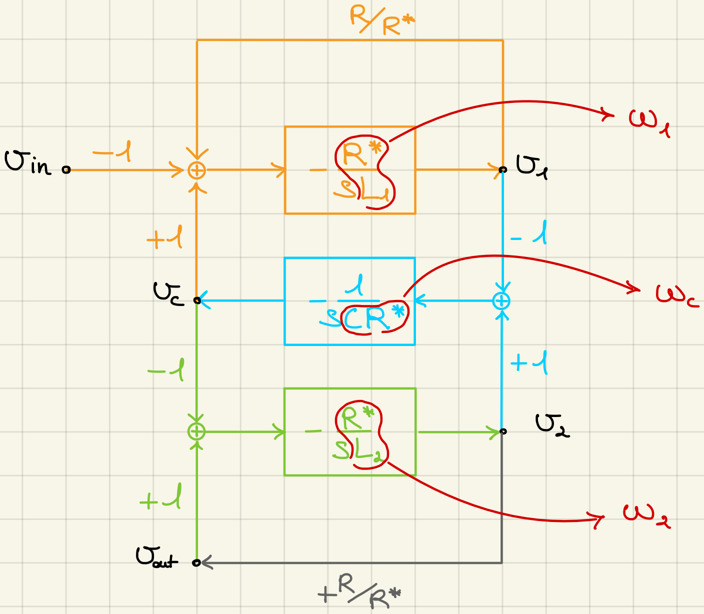
{ v_1, v_2 and R^* are just a mathematical expedient to ease the dissertation, they do not appear in the original network but will be needed for the synthesis of the new one }

1. $\dfrac{v_1}{R^*} = \dfrac{v_{in} - v_c}{R + sL_1}$ 2. $\dfrac{v_c - v_{out}}{sL_2} = \dfrac{v_2}{R^*}$ 3. $\dfrac{v_1}{R^*} - \dfrac{v_2}{R^*} = v_c sC$

(4.) $v_{out} = i_2 R = \dfrac{v_2}{R^*} \cdot R$ → this equation is needed to derive the transfer function but does NOT give any information about the energy of the network (in fact, it does not link two state variables and depends on where the output is taken from)

1. $v_{in} - v_c = v_1 \dfrac{R + sL_1}{R^*} = v_1 \dfrac{R}{R^*} + v_1 \dfrac{sL_1}{R^*} \longrightarrow \boxed{v_1 = \dfrac{R^*}{sL_1}\left(v_{in} - v_c - v_1 \dfrac{R}{R^*}\right)}$

2. $\boxed{v_2 = \dfrac{R^*}{sL_2}\left(v_c - v_{out}\right)}$ 3. $\boxed{v_c = \dfrac{1}{sCR^*}\left(v_1 - v_2\right)}$



Since this circuit holds the same state equations as the original ladder network, we expect the two transfer functions to be exactly the same (and so their dependency on their components and the reduced variability, which is what matters after all).

The original parameter x that was related to L_1, L_2 and C is now related to the radial frequency of the integrator blocks:

$$L_1 \longrightarrow \omega_1 \qquad L_2 \longrightarrow \omega_2 \qquad C \longrightarrow \omega_c$$

$$\left[\frac{\partial |T(j\omega, x)|}{\partial x} \Big|_{\omega=\omega^*} = 0 \right]$$

$$x \rightarrow \omega_1, \omega_2, \omega_c$$

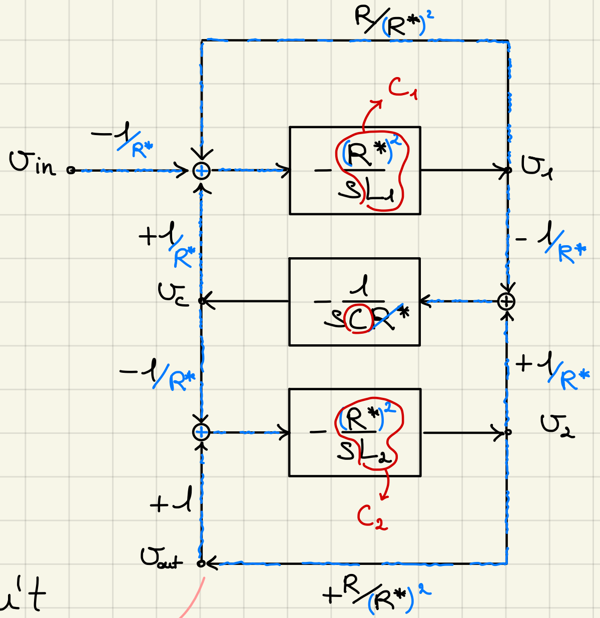
therefore the robustness w.r.t. the components tolerance is correctly retained.

Note: the sensitivity w.r.t. R (and R*) is NOT limited; nonetheless, this problem was already present in the ladder network: in fact, the transfer function is indeed robust against the variability of the reactive components, but it is not necessarily so for the resistors tolerance $\left(\frac{\partial T}{\partial R}\Big|_{\omega=\omega^*} \neq 0 \right)$.

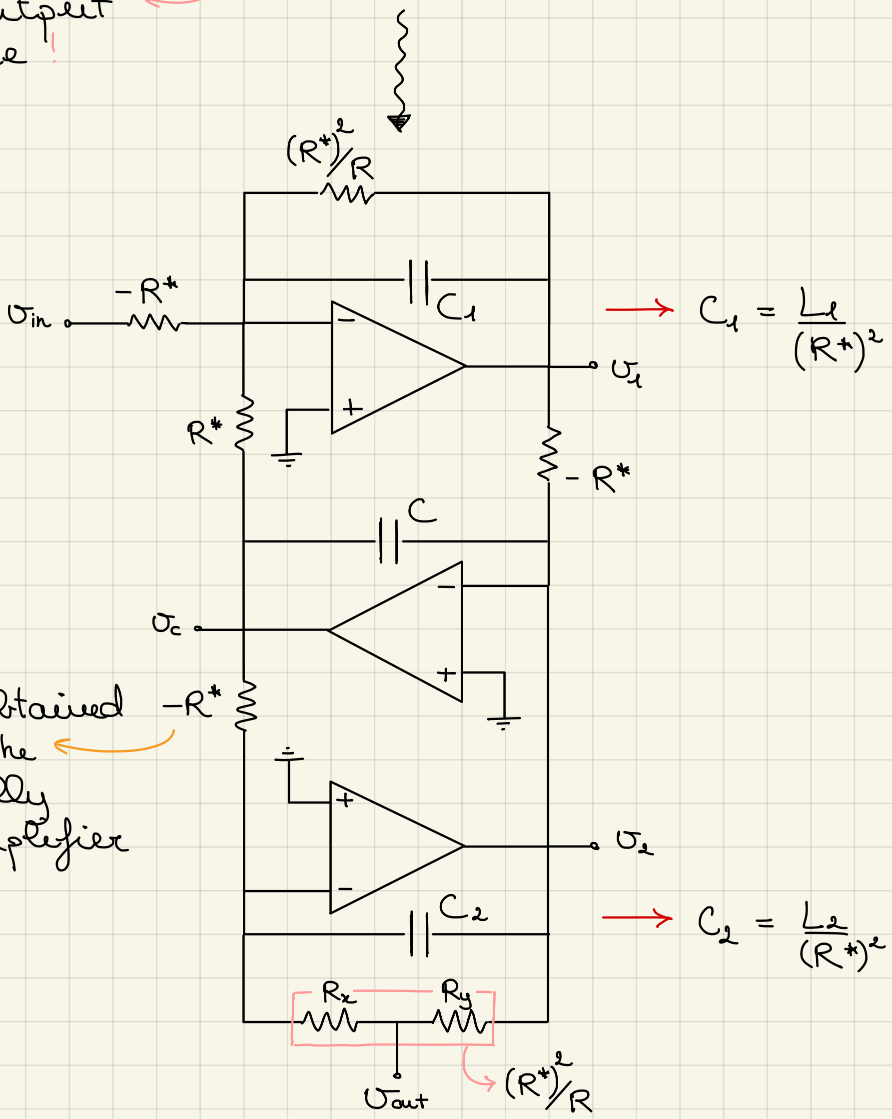
Anyways this problem can be dealt with, both in the original network and in this new synthesized network, since the sensitivity happens to be dependent on the RATIO of two resistor: a common centroid technique helps reducing any possible mismatch.

We should now ask ourselves: how many amplifiers are needed for such implementation?

At least 3 amplifiers are mandatory to build the three integrators.
The three summing nodes can also be implemented through amplifiers, however the cheaper approach (as seen for the universal cell) is summing currents instead of voltages using the virtual ground of the integrators



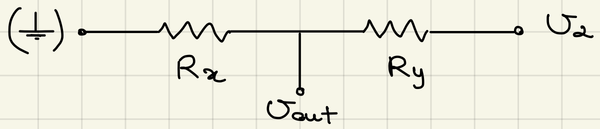
now we don't have our output anymore!



the minus sign can be obtained by crossing the wires of a fully differential amplifier

⇒ To provide the proper voltage output it has to be:

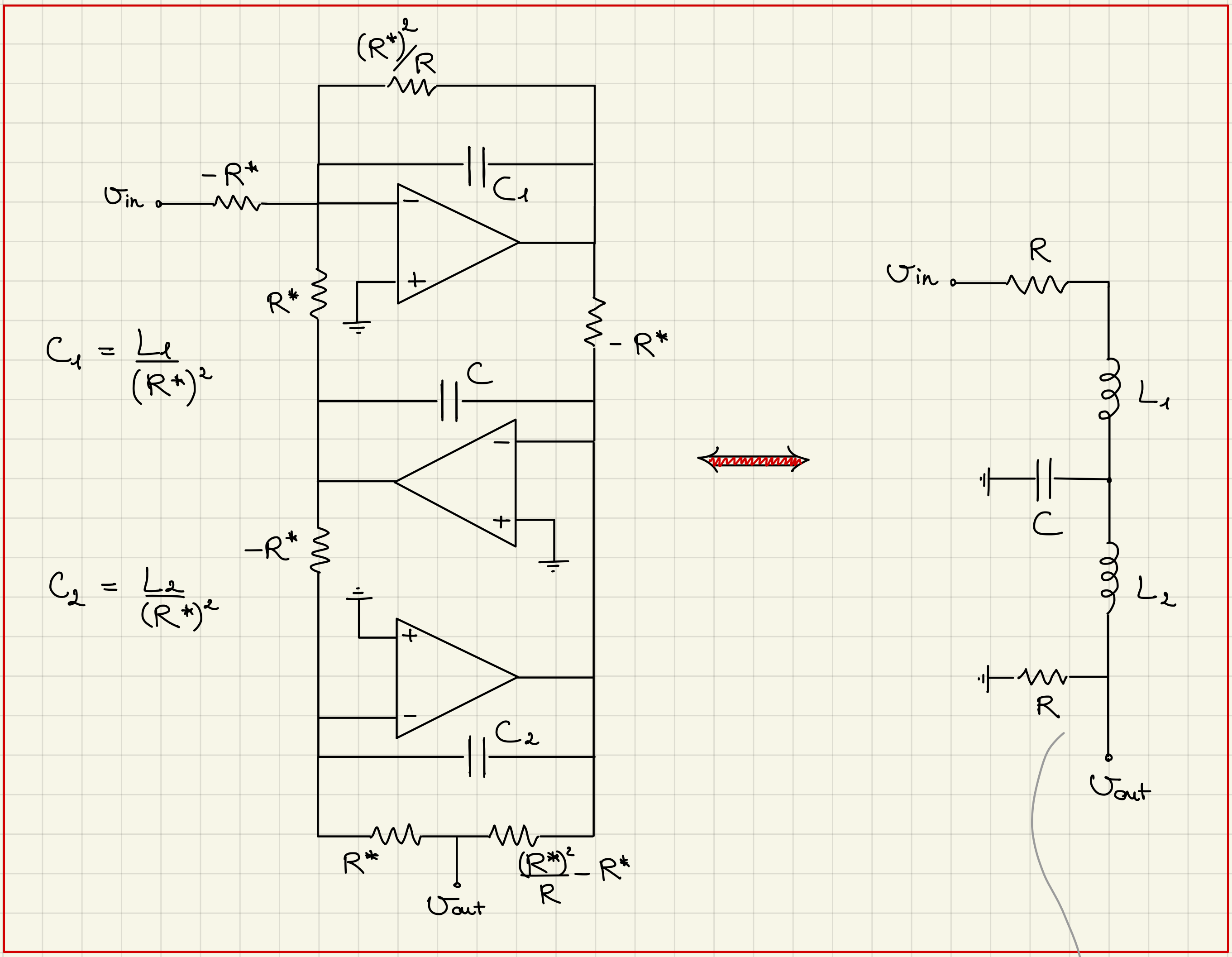
$$v_{out} = v_2 \frac{R}{R^*}$$



$$v_{out} = \frac{R_x}{R_x+R_y} v_2 = \frac{R}{R^*} v_2 \quad \text{and} \quad R_x + R_y = \frac{(R^*)^2}{R}$$

$$\Longrightarrow \begin{cases} \dfrac{R_x}{R_x+R_y} = \dfrac{R}{R^*} \\ R_x + R_y = \dfrac{(R^*)^2}{R} \end{cases}$$

$$\Longrightarrow \begin{cases} R_x = R^* \\ R_y = \dfrac{(R^*)^2}{R} - R^* \end{cases}$$

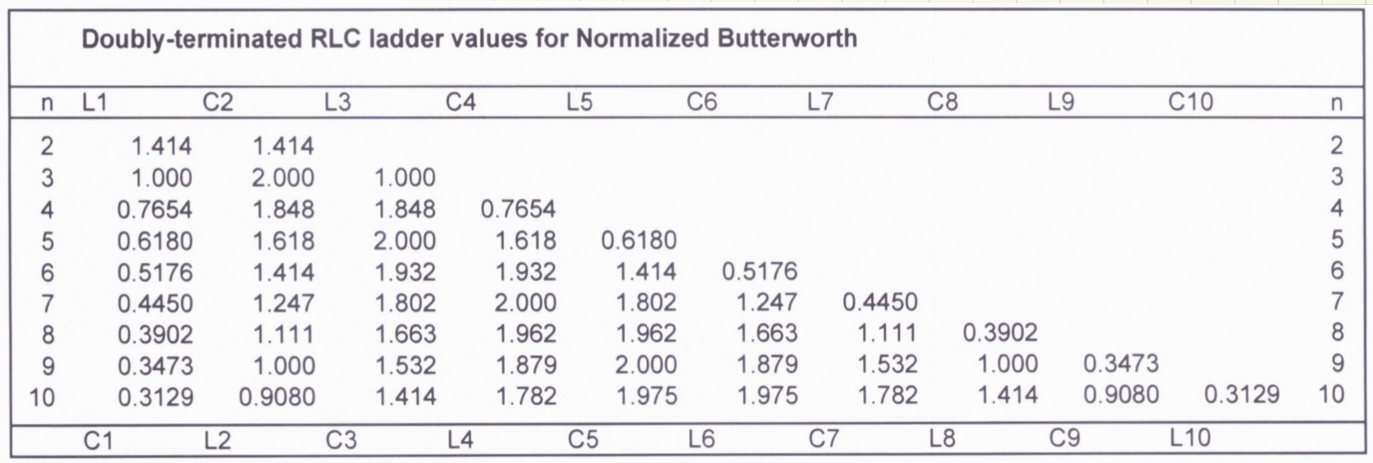


$$C_1 = \frac{L_1}{(R^*)^2}$$

$$C_2 = \frac{L_2}{(R^*)^2}$$

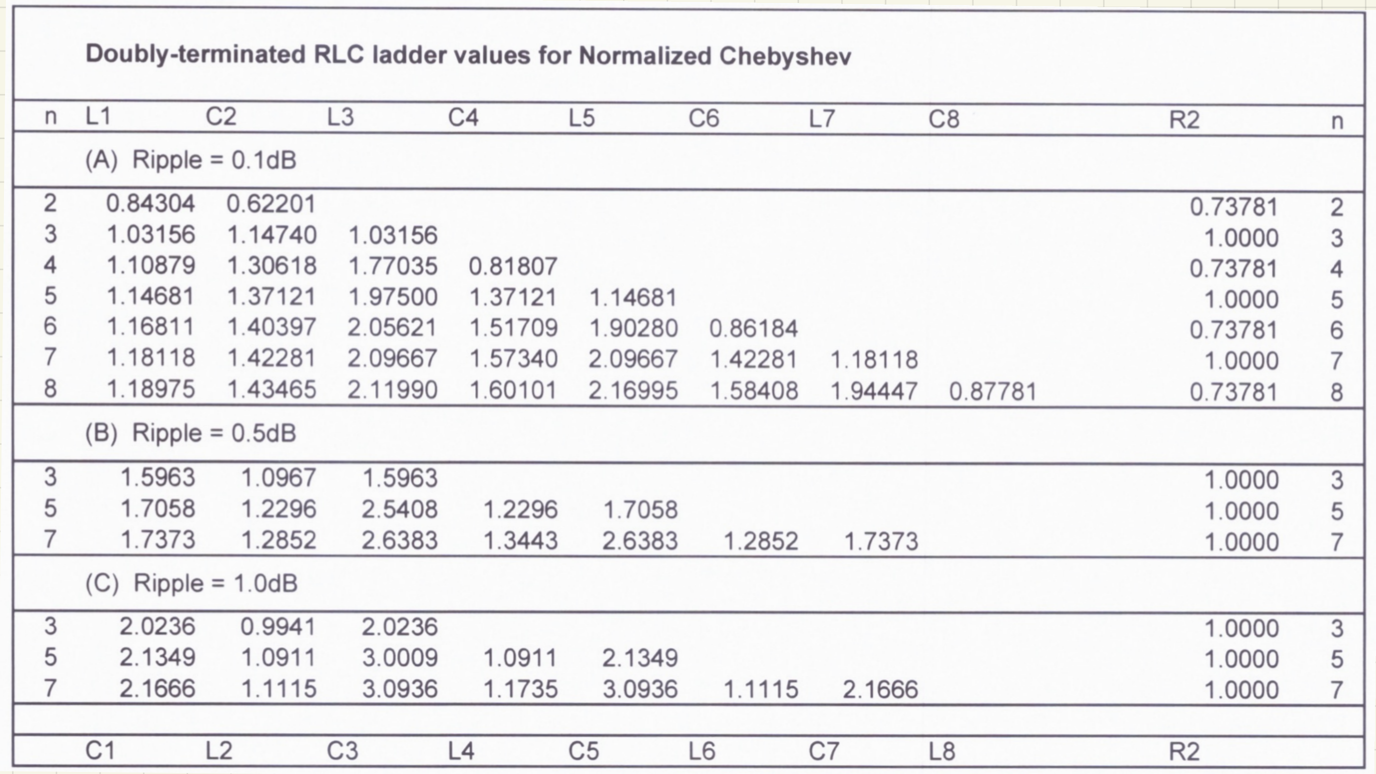
How do we obtain this circuit's parameters in the first place?

The normalized values of inductances, capacitances and resistances of a ladder network low-pass filter are generally given by the proper table of values



Doubly-terminated RLC ladder values for Normalized Butterworth

n	L1	C2	L3	C4	L5	C6	L7	C8	L9	C10	n
2	1.414	1.414									2
3	1.000	2.000	1.000								3
4	0.7654	1.848	1.848	0.7654							4
5	0.6180	1.618	2.000	1.618	0.6180						5
6	0.5176	1.414	1.932	1.932	1.414	0.5176					6
7	0.4450	1.247	1.802	2.000	1.802	1.247	0.4450				7
8	0.3902	1.111	1.663	1.962	1.962	1.663	1.111	0.3902			8
9	0.3473	1.000	1.532	1.879	2.000	1.879	1.532	1.000	0.3473		9
10	0.3129	0.9080	1.414	1.782	1.975	1.975	1.782	1.414	0.9080	0.3129	10
	C1	L2	C3	L4	C5	L6	C7	L8	C9	L10	



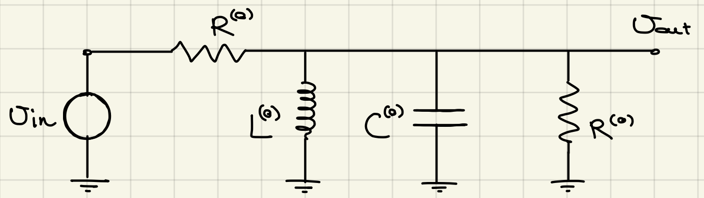
Doubly-terminated RLC ladder values for Normalized Chebyshev

n	L1	C2	L3	C4	L5	C6	L7	C8	R2	n
	(A) Ripple = 0.1dB									
2	0.84304	0.62201							0.73781	2
3	1.03156	1.14740	1.03156						1.0000	3
4	1.10879	1.30618	1.77035	0.81807					0.73781	4
5	1.14681	1.37121	1.97500	1.37121	1.14681				1.0000	5
6	1.16811	1.40397	2.05621	1.51709	1.90280	0.86184			0.73781	6
7	1.18118	1.42281	2.09667	1.57340	2.09667	1.42281	1.18118		1.0000	7
8	1.18975	1.43465	2.11990	1.60101	2.16995	1.58408	1.94447	0.87781	0.73781	8
	(B) Ripple = 0.5dB									
3	1.5963	1.0967	1.5963						1.0000	3
5	1.7058	1.2296	2.5408	1.2296	1.7058				1.0000	5
7	1.7373	1.2852	2.6383	1.3443	2.6383	1.2852	1.7373		1.0000	7
	(C) Ripple = 1.0dB									
3	2.0236	0.9941	2.0236						1.0000	3
5	2.1349	1.0911	3.0009	1.0911	2.1349				1.0000	5
7	2.1666	1.1115	3.0936	1.1735	3.0936	1.1115	2.1666		1.0000	7
	C1	L2	C3	L4	C5	L6	C7	L8	R2	

The table provides values for the reference low-pass filter with $\omega_{BP} = 1\,\frac{rad}{s}$ and $R_1 = 1\Omega$
The values must then be properly denormalized to derive the actual parameters of the ladder network (L_1, L_2, C, R_1 and R_2 in our example) or rather the equivalent parameters of the integrators (C_1, C_2, C, R_1, and R_2).

The process of denormalization typically leaves a few degrees of freedom when sizing the components, so it has to be merged with whatever power/noise/sensitivity constraint to define the optimal filter implementation, as we will see.

Example: ladder network band-pass filter



$$\omega_0 = \frac{1}{\sqrt{LC}} \sim \frac{Q}{RC} \sim \frac{R}{QL}$$

$$Q = \sqrt{\frac{C}{L}} \cdot R \sim \omega_0 CR \sim \frac{R}{\omega_0 L}$$

$$\gamma = \frac{1}{2}$$

$$T(s) = \frac{\gamma\, s\, L/R}{s^2 LC + s\, L/R + 1}$$

Assume $R^{(0)} = 1\Omega$ $L^{(0)} = 1H$ $C^{(0)} = 1F$ normalized values.

We need to denormalize ω_0 to a certain radial frequency N:

$$\omega_0^{(0)} = 1 \frac{rad}{s} \longrightarrow \omega_0 = N \frac{rad}{s} \quad \text{(target BP frequency)}$$

$\Longrightarrow$ Divide both $L^{(0)}$ and $C^{(0)}$ by N:

$$L^{(1)} = \frac{L^{(0)}}{N} \qquad C^{(1)} = \frac{C^{(0)}}{N}$$

We must check that Q did not change during the ω_0 transformation:

$$Q = \frac{R^{(0)}}{\omega_0 L^{(1)}} = \frac{R^{(0)}}{(\omega_0^{(0)} \cdot \cancel{N}) \cdot (L^{(0)}/\cancel{N})} = \frac{R^{(0)}}{\omega_0^{(0)} \cdot L^{(0)}} \quad \checkmark \qquad Q \text{ remained const.}$$

However if N is not large enough (at least $\sim 10^3$!) we might get a value for $C^{(1)} = \frac{C^{(0)}}{N} = \frac{1F}{N}$ too large to be practically implemented.

$\Longrightarrow$ Multiply $R^{(0)}$ and $L^{(1)}$ by a factor M and divide $C^{(1)}$ by the same M

$$R = R^{(0)} \cdot M \qquad L = \frac{L^{(0)}}{N} \cdot M \qquad C = \frac{C^{(0)}}{N \cdot M}$$

Check that ω_0 and Q stayed the same:

$$\omega_0 = \frac{1}{\sqrt{L \cdot C}} = \frac{N}{\sqrt{(L^{(0)}\cancel{M})(C^{(0)}/\cancel{M})}} = N \frac{rad}{s} \quad \checkmark$$

$$Q = \frac{R}{\omega_0 \cdot L} = \frac{R^{(0)} M}{(\omega_0^{(0)} \cdot N)(L^{(0)} \frac{M}{N})} = \frac{R^{(0)}}{\omega_0^{(0)} \cdot L^{(0)}} \quad \checkmark$$

Now it seems that, while C can be low enough by adjusting factor M (which also determines the value of R), L might be too high.

This is not a problem since the inductance L is not actually implemented: it is either replaced by a gyrator or identified by the capacitance of an integrator block. In the latter case, we know from previous computations that the capacitance that will eventually get implemented is proportional to said inductance:

$$C_1 = \frac{L_1}{(R^*)^2}$$

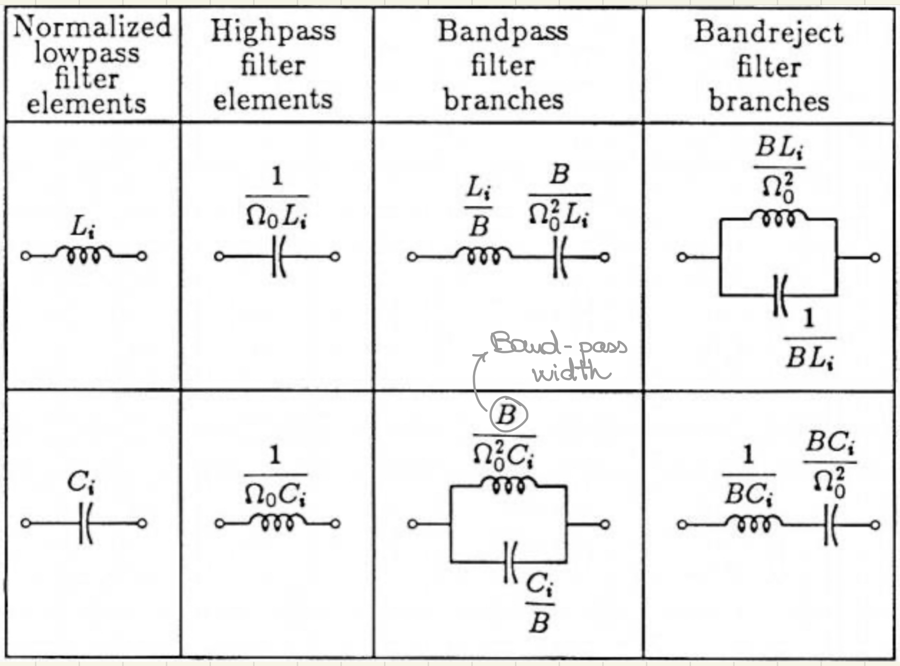
$\Longrightarrow$ Size R^* to obtain a proper value for the implemented capacitances of the integrator blocks

Normalized values	Band-pass frequency (N)	Resistance value (M)
$R^{(0)}$	$\times 1$	$\times M$
$C^{(0)}$	$\times 1/N$	$\times 1/NM$
$L^{(0)}$	$\times 1/N$	$\times M/N$

Some additional comments to clarify a few things

For starters, as reported in the tables, some Chebyshev-I configurations are not doubly terminated; it can be demonstrated that the Orchad theorem (and all the discussion held so far) is also valid for non-doubly terminated ladder networks.

Now, a crucial point we haven't covered yet is how to implement high-pass and band-pass filters with a ladder network, starting from the aforeseen normalized low-pass values.
We know any filter mask can be converted to a normalized low-pass mask, for which we have seen the table of values of the corresponding ladder network. Given these values, one can revert back to the original filter type through the following transformations:



Normalized lowpass filter elements	Highpass filter elements	Bandpass filter branches	Bandreject filter branches
L_i	$\frac{1}{\Omega_0 L_i}$	$\frac{L_i}{B}$ $\frac{B}{\Omega_0^2 L_i}$ Band-pass width	$\frac{BL_i}{\Omega_0^2}$ $\frac{1}{BL_i}$
C_i	$\frac{1}{\Omega_0 C_i}$	$\frac{B}{\Omega_0^2 C_i}$ $\frac{C_i}{B}$	$\frac{1}{BC_i}$ $\frac{BC_i}{\Omega_0^2}$

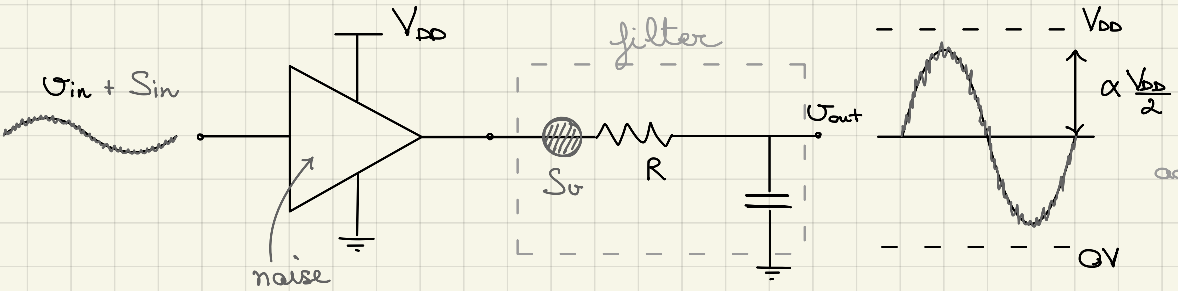
So, for example, the denormalized high-pass filter is derived from the normalized low-pass ladder network by swapping inductors with capacitors and viceversa, with the proper denormalizing transformation.
This table does not take into account eventual denormalizations of R_1 and R_2. As we've seen, sizing the terminating resistances to the target M value simply entails scaling up all inductances and scaling down all capacitances by the same factor M.

In general, to obtain the ladder network that implements a certain filter, these steps should be taken in order:

filter specifications
↓
filter mask
↓
normalized low-pass mask
↓
Butterworth/Chebyshev order
↓
RLC values for normalized low-pass Butterworth/Chebyshev
↓
denormalized RLC values
↙ ↘
implementation through gyrators — implementation through integrators
↓
choice of R* value

So far it seemed like we could choose the values of the resistances and capacitances in almost any way we wanted (for instance, we could set huge R and R* values to minimize C and C_1).

However, the setting of resistors and capacitors is also influenced by the filter non-idealities: noise, finite gain distortion, etc.



($\alpha \leq 1$ takes into account possible additional limitations to the output dynamic)

The Signal-to-Noise ratio considering only the filter noise is given by:

$$\left(\frac{S}{N}\right)^2 = \frac{\left(\alpha \frac{V_{DD}}{2}\right)^2 \cdot \frac{1}{2}}{S_v \cdot BW} = \frac{\left(\alpha \frac{V_{DD}}{2}\right)^2 \cdot \frac{1}{2}}{4KTR \cdot \frac{1}{4RC}} = \frac{\left(\alpha \frac{V_{DD}}{2}\right)^2 \cdot \frac{1}{2}}{KT/C}$$

All the additional noise (coming from the source, early stages, etc) can be taken into account by adding the noise figure term:

→ is the ratio between a specific noise over all other noise sources (F); sometimes it is also indicated as the ratio between a specific noise over all noise sources (1+F).

$$\left(\frac{S}{N}\right)^2_{tot} = \frac{\left(\alpha \frac{V_{DD}}{2}\right)^2 \cdot \frac{1}{2}}{\frac{KT}{C}(1+F)}$$

$$\left(\frac{S}{N}\right)^{\uparrow}_{tot} = \alpha V_{DD}\sqrt{\frac{C\uparrow}{8KT(1+F)}}$$

Note that to reduce the noise, the capacitance value should be increased. However, the band-pass frequency of the filter must not be altered, therefore the resistance should also be decreased to compensate ($BW \propto \frac{1}{RC}$).

This statement goes against the criteria we adopted during denormalization: if we choose smaller capacitances (and bigger resistances) we reduce the silicon occupation but we increase the noise.

⟹ Trade-off between silicon real estate and noise

Not only noise, but also power dissipation can be an issue when choosing the size of capacitors and resistors.

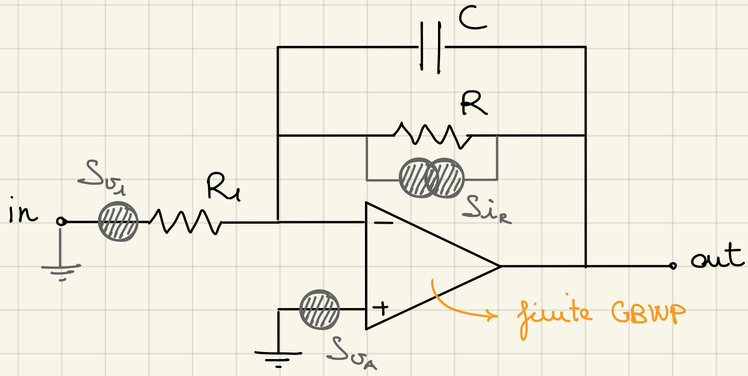
To compute the power dissipation of the system, we need to ask ourselves how much energy is drawned from the supply during each cycle:

$$\downarrow P_d = \frac{\mathcal{E}}{T} = \underset{\substack{\downarrow \\ \text{charge collected from p.s.}}}{(\alpha V_{DD} C)} V_{DD} \cdot f = \alpha \downarrow C \, V_{DD}^2 \, f$$

So a larger capacitance will cause higher power consumptions.

$\Longrightarrow$ Trade-off between power dissipation and noise

Example: band-pass filter output noise PSD



$$T_1(s) = -\frac{R}{R_1}\frac{1}{1+sCR} = -G_1 \frac{1}{1+\frac{s}{\omega_0}}$$

$$\Longrightarrow \langle n_{out}^2 \rangle_1 = \int_0^{+\infty} 4kTR_1 \, |T_1(j\omega)|^2 \, df = 4kTR_1 G_1^2 \int_0^{+\infty} \frac{df}{1+\left(\frac{\omega}{\omega_0}\right)^2} =$$

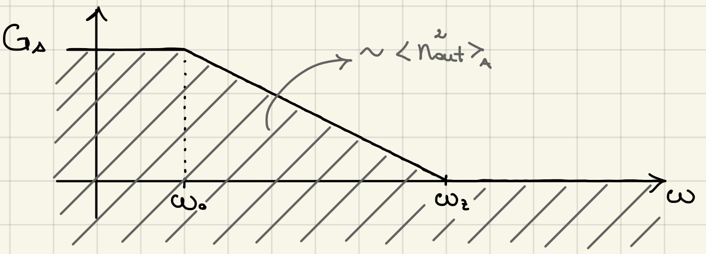
$$= 4kTR_1 G_1^2 \frac{\omega_0}{4}$$

$$T_R(s) = -R \frac{1}{1+\frac{s}{\omega_0}} \qquad \Longrightarrow \qquad \langle n_{out}^2 \rangle_R = \frac{4kT}{R} \cdot R^2 \frac{\omega_0}{4}$$

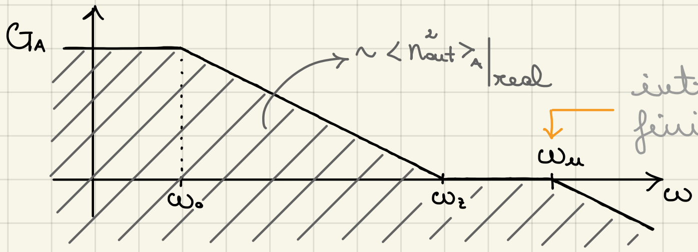
$$T_A(s) = \left(1+\frac{R}{R_1}\right)\frac{1+sC(R_1 /\!/ R)}{1+sCR} = G_A \frac{1+\frac{s}{\omega_z}}{1+\frac{s}{\omega_0}}$$

$$\langle n_{out}^2 \rangle_A = \int_0^{+\infty} S_{v_A} G_A^2 \frac{1+\left(\frac{\omega}{\omega_z}\right)^2}{1+\left(\frac{\omega}{\omega_0}\right)^2} df \longrightarrow +\infty \quad \text{divergent !}$$

we erroneously considered the transfer function T_A to be ideal



however the cut-off at higher frequencies due to the finite GBWP of the OPAMP should be taken into account



introduced by the finite GBWP when $G_{loop} = 1$

$$\langle n_{out}^2 \rangle_A \Big|_{real} = \int_0^{+\infty} S_{v_A} G_A^2 \frac{\left[1+\left(\frac{\omega}{\omega_z}\right)^2\right]}{\left[1+\left(\frac{\omega}{\omega_0}\right)^2\right]\left[1+\left(\frac{\omega}{\omega_u}\right)^2\right]} \, df$$

This integral is not easy to solve in closed form; nonetheless there are tables that give the result for some standard functions of this type. In this case:

$$\int_0^{+\infty} \left| \frac{1+\frac{s}{\omega_z}}{\frac{s^2}{\omega_0^*}+\frac{s}{\omega_0^* Q}+1} \right|^2 df = \frac{\omega_0^* Q}{4}\left[1+\left(\frac{\omega_0^*}{\omega_z}\right)^2\right] = BW$$

where $\left(1+\frac{s}{\omega_0}\right)\left(1+\frac{s}{\omega_u}\right) = 1 + s\left(\frac{1}{\omega_0}+\frac{1}{\omega_z}\right) + \frac{s^2}{\omega_0\omega_u} = 1 + \frac{s}{\omega_0^* Q} + \frac{s^2}{\omega_0^2}$

$\longrightarrow \quad \omega_0^* = \sqrt{\omega_0 \omega_u} \qquad Q = \frac{\sqrt{\omega_0\omega_u}}{\omega_0+\omega_u} \longleftarrow$

$$\Longrightarrow BW = \frac{\sqrt{\omega_0\omega_u}}{4} \cdot \frac{\sqrt{\omega_0\omega_u}}{\cancel{\omega_0+}\,\omega_u}\left[1+\frac{\omega_0\omega_u}{\omega_z^2}\right] =$$

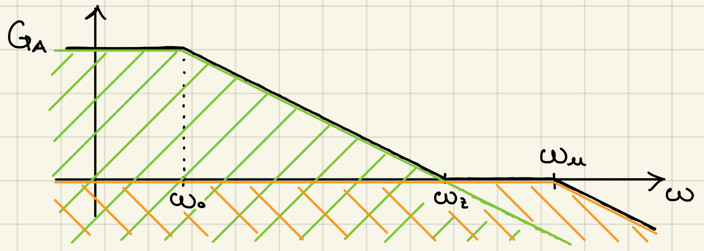
$\omega_0 G_A = \omega_z$

$$\overset{\omega_u \gg \omega_0}{\simeq} \frac{\omega_0\omega_u}{4\omega_u} + \frac{(\omega_0\omega_u)^2}{4\omega_u\omega_z^2} = \frac{\omega_0}{4} + \frac{\omega_u}{4}\left(\frac{\omega_0^2}{\omega_z^2}\right) = \frac{\omega_0}{4} + \frac{\omega_u}{4}\frac{1}{G_A^2}$$

$$\Longrightarrow \langle n_{out_A}^2 \rangle \Big|_{real} \simeq S_{v_A} G_A^2 \cdot BW = S_{v_A} G_A^2 \frac{\omega_0}{4} + S_{v_A}\frac{\omega_u}{4}$$

A faster, intuitive way to obtain the same result without incurring in all these calculations is to consider the filtering effects of ω_0 and ω_u as separate contributions.
Given that $\omega_0 \ll \omega_u$, we can assume the total output

noise to be the sum of the noise PSD integrated up to ω_0, plus the noise PSD integrated up to ω_u:



$$\Longrightarrow \langle n_{out}^2 \rangle_\Delta \Big|_{real} \lesssim \langle n_{out}^2 \rangle + \langle n_{out}^2 \rangle = \int_0^{+\infty} S_{v_A} G_A^2 \frac{1}{1+\left(\frac{\omega}{\omega_0}\right)^2} df + \int_0^{+\infty} S_{v_A} \frac{1}{1+\left(\frac{\omega}{\omega_u}\right)^2} df$$

$$= S_{v_A} G_A^2 \frac{\omega_0}{4} + S_{v_A} \frac{\omega_u}{4}$$

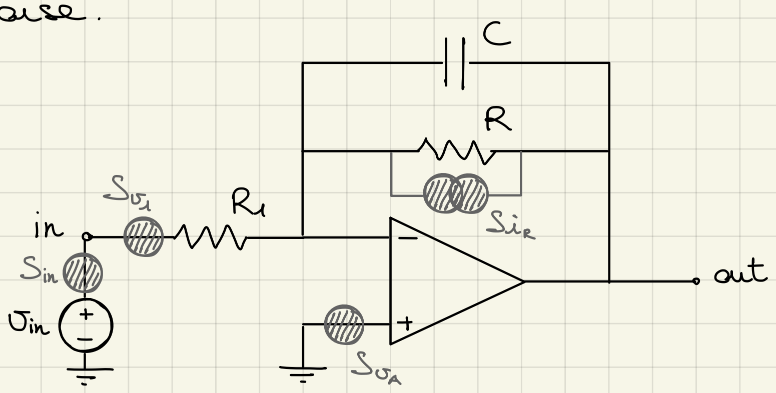
$$\Longrightarrow \langle n_{out}^2 \rangle = 4KTR_1 \left(\frac{R}{R_1}\right)^2 \frac{\omega_0}{4} + 4KTR \frac{\omega_0}{4} +$$

$$+ S_{v_A} \left(1 + \frac{R}{R_1}\right)^2 \frac{\omega_0}{4} + S_{v_A} \frac{\boxed{\omega_u}}{4}$$

an ideally infinite GBWP would cause an infinite output noise

Note that the GBWP of the amplifier appears in the expression of the output noise (through ω_u). <u>The higher the GBWP, the noisier the output</u> (since $\langle n_{out}^2 \rangle \propto \omega_u$). For this reason having a too large GBWP can harshly impair the performance of the filter. If lowering it is not an option, then additional poles should be placed at the filter output to limit the overall noise transfer

In this example we've only considered the noise introduced by the filter. However the source signal comes itself with some noise.



$G_1 = \frac{R}{R_1} = G$

$G_A = 1 + \frac{R}{R_1} = 1 + G$

$$\langle n_{out}^2 \rangle_{tot} = S_{in} G_1^2 \frac{\omega_0}{4} + 4KTR_1 G_1^2 \frac{\omega_0}{4} + 4KTR \frac{\omega_0}{4} +$$

$$+ S_{v_A} G_A^2 \frac{\omega_0}{4} + S_{v_A} \frac{\omega_u}{4}$$

$$\Longrightarrow \langle n_{out}^2 \rangle_{tot} = S_{in}\, G^2 \frac{\omega_0}{4}\left[1 + \underbrace{\frac{4KTR_1}{S_{in}} + \frac{4KTR}{S_{in}G^2} + \frac{S_{v_A}}{S_{in}}\frac{(1+G)^2}{G^2} + \frac{S_{v_A}}{S_{in}G^2}\frac{\omega_u}{\omega_0}}_{F}\right]$$

The designer's objective is to reduce the noise figure: since he has no control over the source noise, he must make things so that the filter/amplifier noise is negligible compared to it – that is, so that the noise figure F is as low as possible.

→ High gain $G = \frac{R}{R_1}$. In this way the signal noise gets amplified and the filter noise is overshadowed.

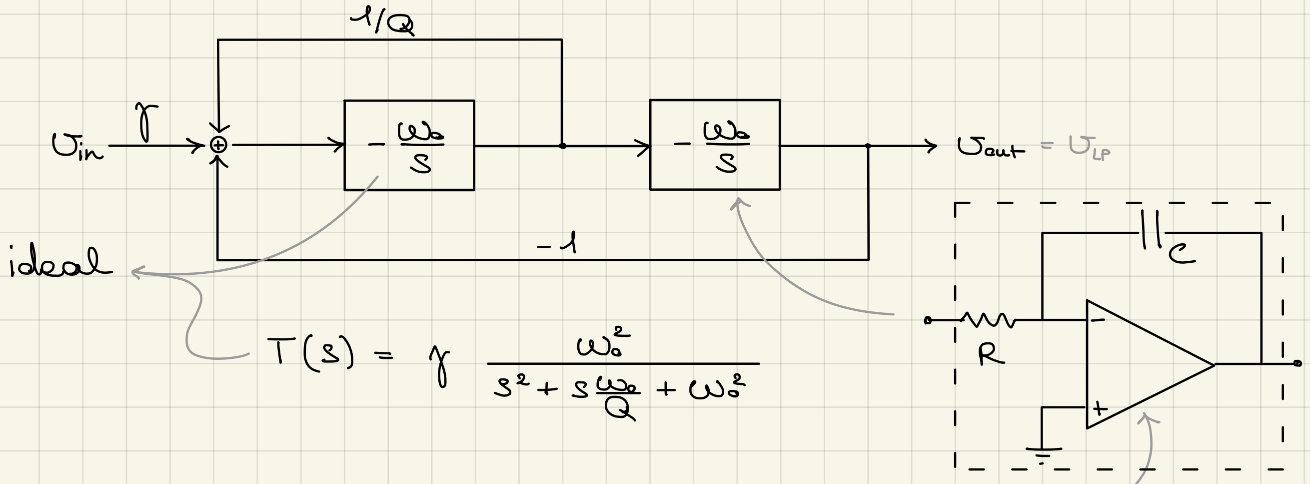
→ Low resistance R_1. The thermal noise of R_1 is directly comparable with the source noise, so a lower value for the front end resistor is better in order not to produce a noise greater than the input one

→ Low input referred noise of the OPAMP $S_{v_A} \sim \frac{8KT\gamma}{g_m}$. A proper input bias of the amplifier should be adopted so to have a low input referred noise.

→ Low GBWP $\sim \omega_u$. As already discussed, a larger GBWP allows for more noise of the OPAMP to reach the filter output; either a lower GBWP or an additional filtering action at the output is therefore recommended.

Let's now see how the finite gain of an amplifier can affect the filter transfer function.

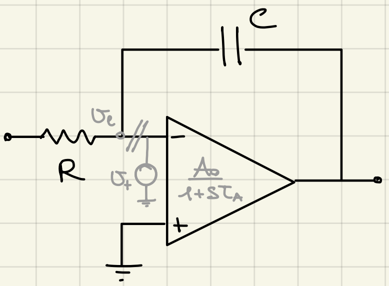
Consider the following biquad universal cell:



$$T(s) = \gamma \frac{\omega_0^2}{s^2 + s\frac{\omega_0}{Q} + \omega_0^2}$$

finite gain and bandwidth ← $\frac{A_0}{1+s\tau_A}$

$$H_{id}(s) = -\frac{1}{sRC} = -\frac{\omega_0}{s} \qquad H_{real}(s) = \frac{H_{id}(s)}{1 - \frac{1}{G_{loop}(s)}}$$

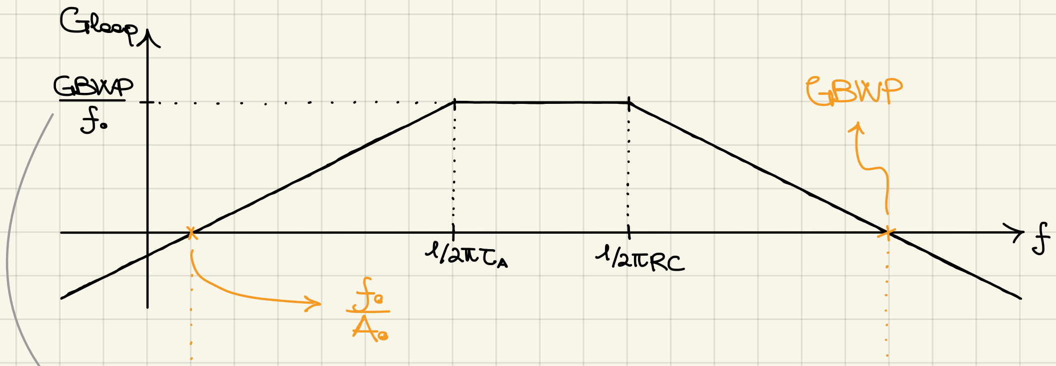


$$G_{loop}(s) = \frac{v_e}{v_+} = -\frac{A_0}{1+s\tau_A}\,\frac{R}{R+1/sC} = -\frac{A_0}{1+s\tau_A}\,\frac{sCR}{1+sCR}$$

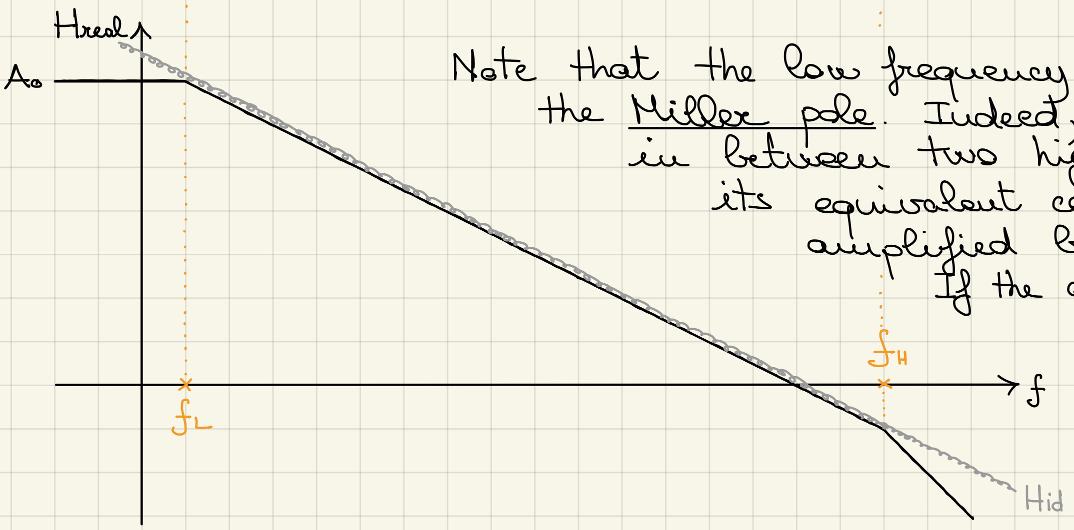
$G_{loop}(s) = 1 \longleftrightarrow$ <u>closed loop poles</u>

$$GBWP = \frac{A_0}{2\pi\tau_A}$$

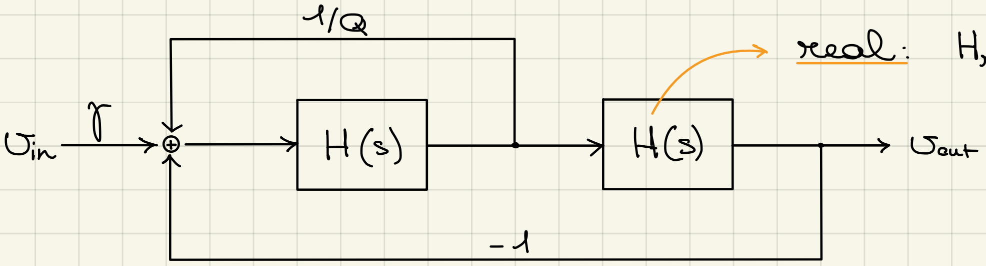
$$f_0 = \frac{1}{2\pi RC}$$



to have a good feedback (i.e. a high Gloop) the GBWP of the amplifier should be much larger than the characteristic frequency of the filter $\left(\frac{A_0}{\tau_A} >> \frac{1}{RC}\right)$



Note that the low frequency pole is actually the <u>Miller pole</u>. Indeed, since C is placed in between two high gain nodes, its equivalent capacitance is amplified by the Miller effect. If the amplifier was ideal, then the Miller effect would make the C capacitance virtually infinite, moving its pole to the origin (ideal integrator); since the amplifier is not ideal, the pole is instead at a low but non-zero frequency.



<u>real</u>: $H_{real}(s) = -\frac{A_0}{(1+\frac{s}{\omega_L})(1+\frac{s}{\omega_H})}$

The ideal filter transfer function $T_{id}(s) = \gamma \frac{\omega_0^2}{s^2 + s\frac{\omega_0}{Q} + \omega_0^2}$ will suffer from a shift of both ω_0 and Q due to the different expression of the real integrators.

While the shift of ω_0 can be adjusted (as already pointed out) through an auxillary network that controls the actual radial frequency of the circuit and fixes it accordingly, the shift of Q is not controllable and can therefore heavily affect the filter's performance.

Let's compute how much different the real ω_0' and, more importantly, the real Q' are going to be with respect to the ideal target values ω_0 and Q.

$$T_{id}(s) = \gamma \frac{\omega_0^2/s^2}{\frac{\omega_0^2}{s^2} + \frac{\omega_0}{sQ} + 1} = \gamma \frac{\left(-\frac{\omega_0}{s}\right)^2}{\left(-\frac{\omega_0}{s}\right)^2 - \left(-\frac{\omega_0}{s}\right)\frac{1}{Q} + 1} = \gamma \frac{H_{id}^2(s)}{H_{id}^2(s) - \frac{H_{id}(s)}{Q} + 1}$$

$$\longrightarrow \quad T_{real}(s) = \gamma \frac{H_{real}^2(s)}{H_{real}^2(s) - \frac{H_{real}(s)}{Q} + 1} \quad \text{where} \quad \begin{cases} H_{real}(s) = -\dfrac{A_0}{\left(1+\frac{s}{\omega_L}\right)\left(1+\frac{s}{\omega_H}\right)} \\ \omega_L = \dfrac{\omega_0}{A_0} = \dfrac{1}{A_0 RC} \\ \omega_H = 2\pi \, GBWP = \dfrac{A_0}{\tau_A} \end{cases}$$

The non-idealities affecting the real transfer function are:

1. finite gain of the amplifiers (causing ω_L pole)
2. finite bandwidth of the amplifiers (causing ω_H pole)

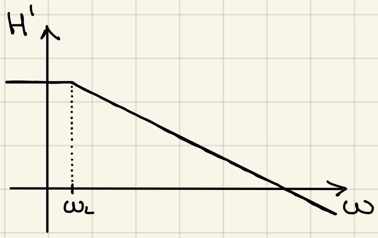
In order to ease the study of this problem, it is better to split the two non-idealities and consider their effects separately.

1. $H'(s) = -\dfrac{A_0}{\left(1+\frac{s}{\omega_L}\right)}$ real integrator with finite gain

$$T'(s) = \frac{\gamma \dfrac{A_0^2}{(1+s/\omega_L)^2}}{\dfrac{A_0^2}{(1+s/\omega_L)^2} + \dfrac{A_0}{(1+s/\omega_L)Q} + 1} = \gamma \frac{A_0^2}{(1+s/\omega_L)^2 + \frac{A_0}{Q}(1+s/\omega_L) + A_0^2} =$$

$$= \gamma \frac{A_0^2}{\frac{s^2}{\omega_L^2} + s\left(\frac{2}{\omega_L} + \frac{A_0}{Q\omega_L}\right) + A_0^2 + \frac{A_0}{Q} + 1} =$$

$$= \gamma \frac{A_0^2 \omega_L^2}{s^2 + s\omega_L\left(2 + \frac{A_0}{Q}\right) + \omega_L^2\left(A_0^2 + \frac{A_0}{Q} + 1\right)}$$



$\omega_L = \frac{\omega_0}{A_0}$

$$= \gamma \frac{\omega_0^2}{s^2 + s\omega_0\underbrace{\left(\frac{2}{A_0} + \frac{1}{Q}\right)}_{\sim 1/Q'} + \underbrace{\omega_0^2\left(1 + \frac{1}{A_0 Q} + \frac{1}{A_0^2}\right)}_{\sim \omega_0'^2}}$$

compare with $T_{id}(s) = \gamma \dfrac{\omega_0^2}{s^2 + s\frac{\omega_0}{Q} + \omega_0^2}$

real characteristic frequency: $$\boxed{\omega_0' = \omega_0\sqrt{\frac{1}{A_0^2}+\frac{1}{A_0Q}+1}}$$

real quality factor: $$\omega_0\left(\frac{2}{A_0}+\frac{1}{Q}\right) = \frac{\omega_0'}{Q'} \simeq \frac{\omega_0}{Q'}$$

with finite gain

$$\frac{1}{Q'} \simeq \frac{2}{A_0}+\frac{1}{Q}$$

$$\frac{1}{Q'}-\frac{1}{Q} \simeq \frac{2}{A_0}$$

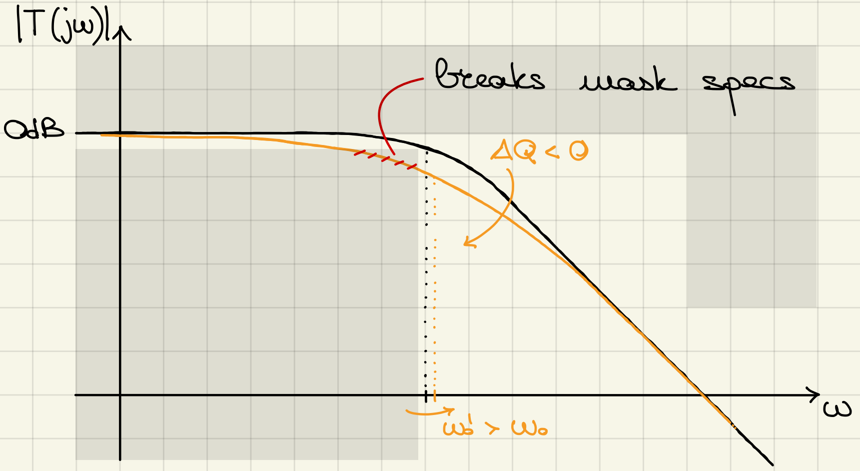
$$\frac{Q-Q'}{QQ'} \simeq \frac{2}{A_0}$$

$$-\frac{\Delta Q}{QQ'} \simeq \frac{2}{A_0}$$

relative shift of the quality factor

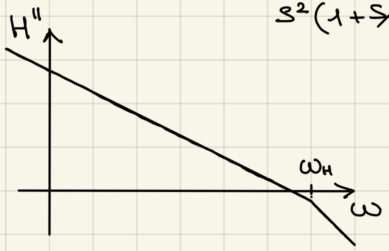
$$\boxed{\frac{\Delta Q}{Q} \simeq -\frac{2Q}{A_0}}$$

Note how the finite gain of the amplifier will cause a lower Q factor than expected for all pole pairs of the filter cell, which might bring the resulting implementation off the required filter mask.



2. $$H''(s) = -\frac{\omega_0}{s}\frac{1}{\left(1+\frac{s}{\omega_H}\right)}$$ real integrator with finite BW

$$T''(s) = \frac{\gamma\,\frac{\omega_0^2}{s^2(1+s/\omega_H)^2}}{\frac{\omega_0^2}{s^2(1+s/\omega_H)^2}+\frac{\omega_0}{Qs(1+s/\omega_H)}+1} = \gamma\frac{\omega_0^2}{s^2\left(1+\frac{s}{\omega_H}\right)^2+\frac{\omega_0 s}{Q}\left(1+\frac{s}{\omega_H}\right)+\omega_0^2}$$



$$= \gamma\frac{\omega_0^2}{\left(1+\frac{s}{\omega_H}\right)^2\left[s^2+\frac{\omega_0 s}{Q\left(1+\frac{s}{\omega_H}\right)}+\frac{\omega_0}{\left(1+\frac{s}{\omega_H}\right)^2}\right]}$$

As already said, the GBWP should be much larger than the frequencies of interest.

Therefore, we can afford the following semplifications:

$$\omega \ll \omega_H \implies \frac{1}{1+s/\omega_H} \simeq 1 - \frac{s}{\omega_H}$$

$2\pi\, GBWP = \omega_H$

$$\implies T''(s) \simeq \frac{\gamma}{\cancel{\left(1+\frac{s}{\omega_H}\right)^2}} \frac{\omega_0^2}{\left[s^2 + s\frac{\omega_0}{Q}\left(1-\frac{s}{\omega_H}\right) + \omega_0^2\left(1-\frac{s}{\omega_H}\right)^2\right]}$$

negligible

$$\simeq \gamma \frac{\omega_0^2}{s^2\left(1-\frac{\omega_0}{Q\omega_H}+\frac{\omega_0^2}{\omega_H^2}\right) + s\left(\frac{\omega_0}{Q} - \frac{2\omega_0^2}{\omega_H}\right) + \omega_0^2}$$

$$= \gamma \frac{\omega_0^2}{s^2 + \frac{s\,\omega_0}{\left(1-\frac{\omega_0}{Q\omega_H}+\frac{\omega_0^2}{\omega_H}\right)}\underbrace{\left(\frac{1}{Q} - \frac{2\omega_0}{\omega_H}\right)}_{\sim 1/Q'} + \underbrace{\frac{\omega_0^2}{\left(1-\frac{\omega_0}{Q\omega_H}+\frac{\omega_0^2}{\omega_H}\right)}}_{\sim \omega_0'^2}}$$

real characteristic frequency:

$$\boxed{\omega_0' = \frac{\omega_0}{\sqrt{1-\frac{\omega_0}{Q\omega_H}+\frac{\omega_0^2}{\omega_H^2}}}}$$

real quality factor:

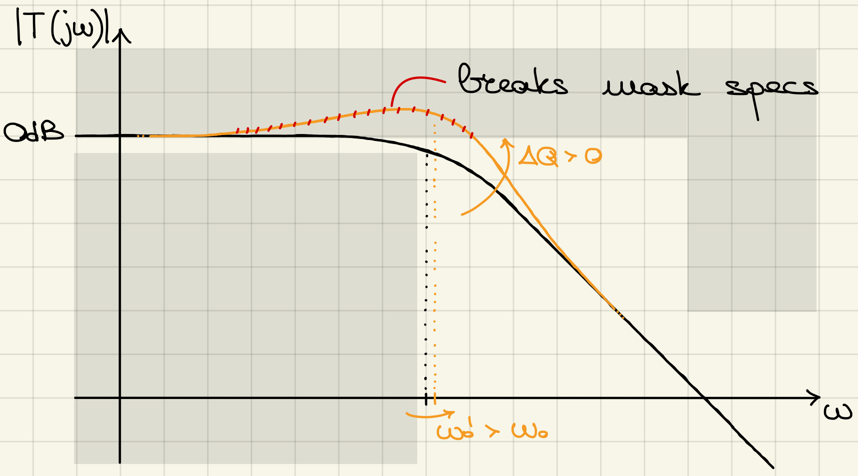
$$\frac{\omega_0'}{Q'} = \frac{\omega_0}{\left(1-\frac{\omega_0}{Q\omega_H}+\frac{\omega_0^2}{\omega_H}\right)}\left(\frac{1}{Q} - \frac{2\omega_0}{\omega_H}\right) \simeq \omega_0'\left(\frac{1}{Q} - \frac{2\omega_0}{\omega_H}\right)$$

with finite bandwidth

$$\frac{1}{Q'} \simeq \frac{1}{Q} - \frac{2\omega_0}{\omega_H}$$

$$\boxed{\frac{\Delta Q}{Q} \simeq 2\frac{Q\,\omega_0}{\omega_H}}$$

Note how this time the finite bandwidth of the amplifier will cause a higher Q factor than expected, which might impair the filter's performance by not abiding the mask specifications!



(These results can be generalized to all filters implemented by active integrators)

The two non-idealities (finite gain and BW) affect the filter at the same time and, even though their effects on the Q factor seem to somewhat cancel out each other, they should be anyway always taken into account.

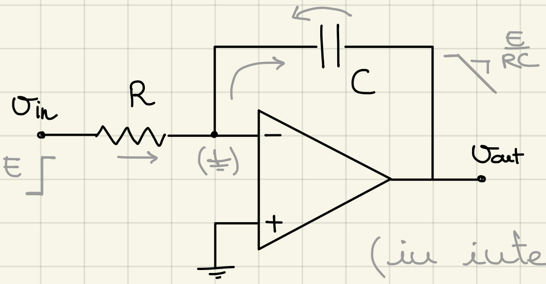
All these constraints, that we've seen arise from noise, power dissipation, finite GBWP (and distortion), are crucial when designing a filter since they give information about what the most fitting components (resistors, capacitors and amplifiers) will be for our task and how well they are required to perform (hence the choice for a proper amplifier design).

Switched Capacitors

The concept of switched capacitors was firstly used by James Clerk Maxwell in its introduction to the fundations of electromagnetism. Switched capacitors have then been used to implement filters in the entire audio range, thanks to their merit of being able to imitate the working principle of large resistors with just a small capacitor.

Let's see where and how this merit takes place.

Assume we have to implement an audio filter with a bandwith of 10 KHz (the full audio range is 20 ÷ 20k Hz). We then need a cell to build the filter, which can be made up by integrator blocks whose radial frequency has to match that of the filter.



$$\omega_0 = 2\pi\, 10\text{KHz} = \frac{1}{RC}$$

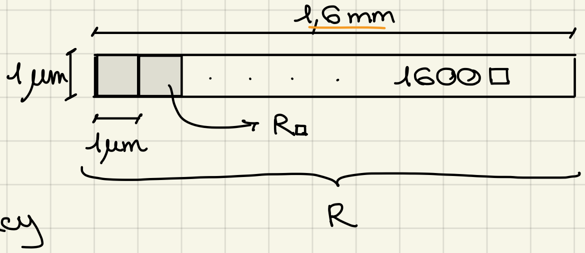
$$C \simeq 1\text{pF} \longrightarrow R = \frac{1}{2\pi\, 10\text{KHz}\, 1\text{pF}} \simeq \underline{16\text{M}\Omega}$$

(in integrated circuits)

$$R_\square \lesssim 10\frac{\text{K}\Omega}{\square}$$

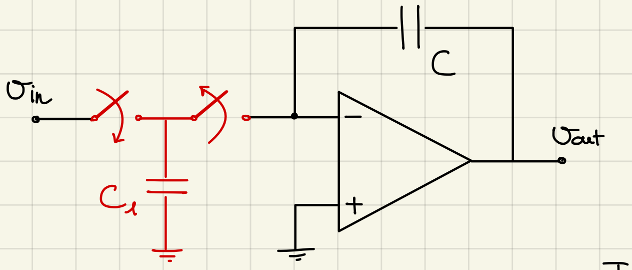
too large !

The problem is that, when we move to the low frequency range, in order to obtain the desired radial frequency



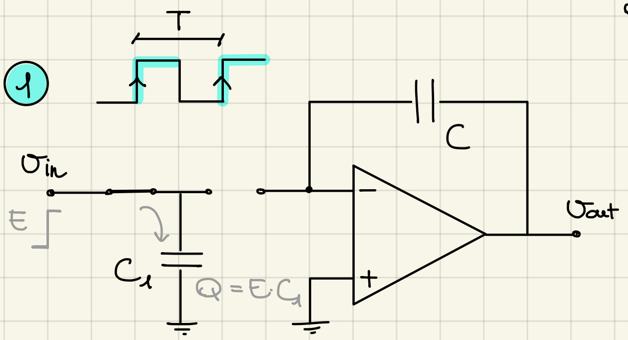
in an integrated technology we need huge resistance values that would take up too much of the available chip area, if implemented in a standard way.

An alternative, more efficient way to obtain large resistances for the implementation of low frequency filters is precisely the use of switched capacitors.
The role of switched capacitors is in fact to mimic the behaviour of the resistors in the aforeseen circuit. What the resistor does in an integrator is simply convert the voltage signal E into a current signal E/R which can be integrated by C.

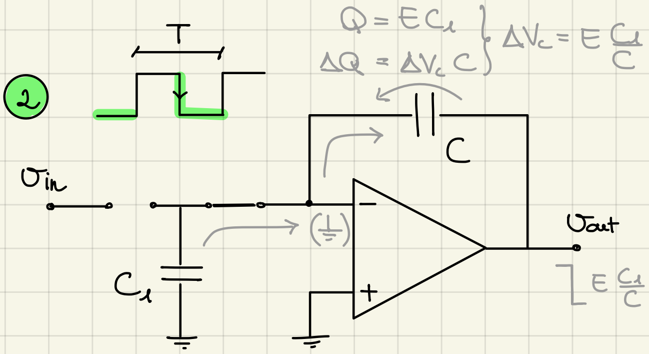


switched capacitor

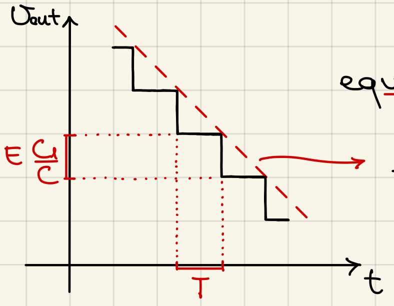
In the equivalent switched capacitor configuration, a capacitor with one end to ground is placed instead of resistor and whose other end is connected to the circuit through two switches, that are closed and open at alternate times over a period T.



In the first part of the period, the left switch is closed while the other is open; the switched capacitor is charged up by the input signal.



In the second part of the period, the right switch is closed while the other is open; the switched capacitor discharges the accumulated charge into the virtual ground, thus integrating it against capacitance C ideally over just an instant (current pulse).
The resulting voltage variation at the output is therefore a step (integral of the current pulse) proportional to the input signal from the first phase.
Over many clock cycles, the output waveform corresponding to a constant input signal with amplitude E will then be:



equivalent ramp rate

$$\frac{\Delta V}{T} = \frac{EC_1}{CT} = \frac{E}{CR_{eq}}$$

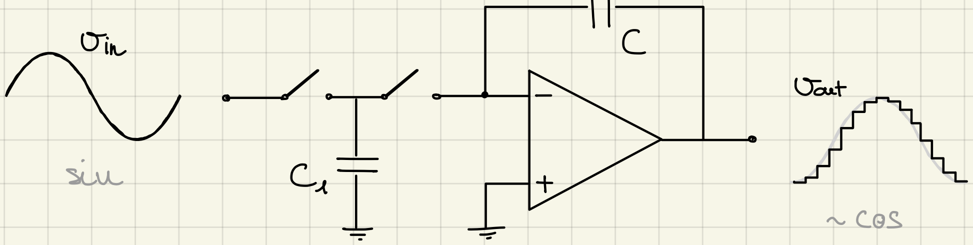
In the continuous time implementation, a resistor is making current flow into virtual ground proportionally to the input voltage and the output is a linear ramp.
This current is equal to $I = \frac{E}{R}$

In the discrete time implementation, a switched capacitor is taking charge from the input voltage and is giving it to the integrating capacitance during each cycle, making an average current flow from input to output.
This average current is equal to $\bar{I} = \frac{Q}{T} = \frac{EC_1}{T}$

It is now clear that the switched capacitor is mimicking an equivalent resistance equal to:

$$\boxed{R_{eq} = \frac{T}{C_1}}$$

The main difference from a continuous time approach will be the staircase-shaped waveform at the output, instead of a linear one.



The discrete approximation will be good enough provided that the switching time is much lower than the period of the input waveform — or, to be more correct, the clock frequency is much larger than the bandwidth of the signal.

Now what is the advantage of this solution?

If you consider the previous audio filter ($f_0 = 10KHz$), we firstly need to ensure that $\frac{1}{T} = f_{clk} \gg f_0$.
This is easily done by setting $f_{clk} = 1MHz$ ($T = 1\mu s$), which is a common value for clock frequencies.
This means that, in order to obtain the required resistance $R = 16M\Omega$ (computed before) with a switched capacitor, we would then need a capacitance C_1 as large as:

$$R = R_{eq} = \frac{T}{C_1} \longrightarrow C_1 = \frac{T}{R} = \frac{1\mu s}{16M\Omega} = \underline{62{,}5\ fF} \quad \text{good !}$$

The switched capacitor allows to approximate the behaviour of a very large resistance with just a small capacitance (and some switches).

Not only this: the switched capacitor has another advantage.

In a standard implementation the radial frequency of the filter is dependent on the absolute value of its components:

$$\omega_0 = 1/RC$$

and thus suffers from tolerance and variability issues.

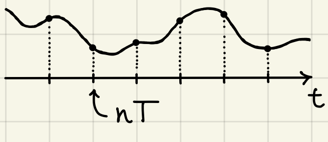
In a switched capacitor implementation instead the radial frequency is dependent on the relative value of the components

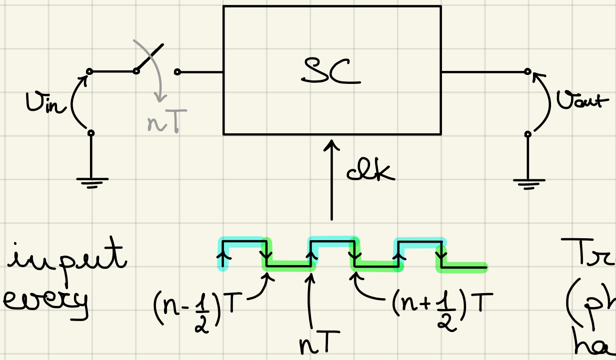
$$\omega_0 = \frac{1}{R_{eq} C} = \frac{C_1}{T \cdot C} = f_{clk} \frac{C_1}{C}$$

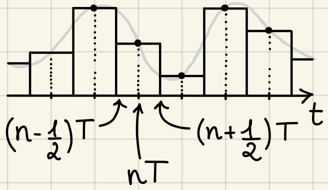
Since the frequency clock can be controlled and is very stable, the only source of error is the ratio of the two capacitors, whose variability can be greatly improved with the proper fabrication layout technique (e.g. common centroid).

The switched capacitor allows for a more reliable effective value of the radial frequency of the filter.

Let's now take a closer look at the implications of dealing with a discrete time system







Sampling of the input (phase 1) happens every full period

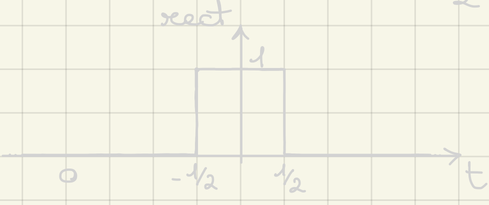
Transition of the output (phase 2) happens every half period

The output waveform can be expressed as:

$$\boxed{v_{out}(t) = \sum_{n=0}^{\infty} v_{out}(nT) \cdot \left\{ step\left[t-(n-\tfrac{1}{2})T\right] - step\left[t-(n+\tfrac{1}{2})t\right] \right\}}$$

$$= \sum_{n=0}^{\infty} v_{out}(nT) \cdot \left\{ rect\left[\frac{t-nT}{T}\right] \right\}$$

where $\mathrm{step}(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases}$ and $\mathrm{rect}(x) = \begin{cases} 1 & |x| < \frac{1}{2} \\ 0 & |x| > \frac{1}{2} \end{cases}$



Laplace transform

Fourier transform

$$\mathcal{L}[\mathrm{step}(t)](s) = \frac{1}{s}$$

$$\mathcal{F}[\mathrm{rect}(t)](f) = \frac{\sin(\pi f)}{\pi f} = \underset{\text{normalized}}{\mathrm{sinc}} f$$

Our task here is to understand what is the link between the Fourier transform (i.e frequency spectrum) of the output signal and the spectrum of the input signal.

$$V_{out}(s) = \mathcal{L}[v_{out}(t)](s) = \sum_{n=0}^{\infty} v_{out}(nT) \cdot \left\{ \frac{1}{s} e^{-s(n-\frac{1}{2})T} - \frac{1}{s} e^{-s(n+\frac{1}{2})T} \right\}$$

remember that $\mathcal{L}[f(x-x_0)](s) = \mathcal{L}[f(x)](s) \cdot e^{-sx_0}$

and also $\mathcal{F}[f(x-x_0)](f) = \mathcal{F}[f(x)](f)\, e^{-j2\pi f x_0}$

$\mathcal{F}[f(x \cdot T)](f) = \frac{1}{T} \mathcal{F}[f(x)](\frac{f}{T})$

$$V_{out}(s) = \sum_{n=0}^{\infty} v_{out}(nT) \frac{1}{s} e^{-snT} \left\{ e^{s\frac{T}{2}} - e^{-s\frac{T}{2}} \right\}$$

$$\Big\downarrow s = j\omega$$

$$V_{out}(j\omega) = \sum_{n=0}^{\infty} v_{out}(nT) \frac{1}{j\omega} e^{-j\omega nT} \left\{ e^{j\omega\frac{T}{2}} - e^{-j\omega\frac{T}{2}} \right\}$$

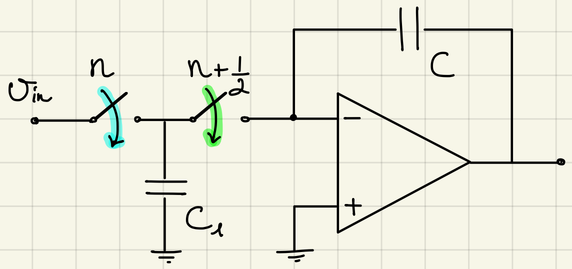
$$= \sum_{n=0}^{\infty} v_{out}(nT)\, e^{-j\omega nT}\, T \left\{ \frac{e^{j\omega T/2} - e^{-j\omega T/2}}{2j \cdot \omega T/2} \right\}$$

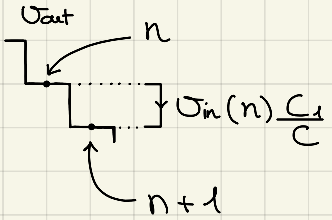
$$\mathcal{F}[v_{out}(t)](\omega) = \sum_{n=0}^{\infty} v_{out}(nT)\, e^{-j\omega nT}\, T \underset{\substack{\text{non} \\ \text{normalized}}}{\mathrm{sinc}} \left(\frac{\omega T}{2}\right) \dashrightarrow \mathcal{F}\left[\mathrm{rect}\left(\frac{t-nT}{T}\right)\right](f)$$

$$= \underbrace{\sum_{n=\infty}^{\infty} v_{out}(nT)\, z^{-n}}\Big|_{z=e^{j\omega T}} \cdot T \mathrm{sinc}\left(\frac{\omega T}{2}\right)$$

like a "discrete Laplace transform" ↔ zeta-transform of v_{out} := $V_{out}(z)$

$$1 \longrightarrow \quad V_{out}(\omega) = V_{out}(z)\Big|_{z=e^{j\omega T}} \cdot T \mathrm{sinc}\left(\frac{\omega T}{2}\right)$$





$$\Longrightarrow \quad v_{out}(n+1) = v_{out}(n) - v_{in}(n)\frac{C_1}{C}$$

zeta - transform

$$V_{out}(z)\cdot z = V_{out}(z) - V_{in}(z)\frac{C_1}{C}$$

$$V_{out}(z)(z-1) = -V_{in}(z)\frac{C_1}{C}$$

$$2 \Longrightarrow \quad \frac{V_{out}(z)}{V_{in}(z)} = -\frac{C_1}{C}\,\frac{1}{z-1} = H(z) =$$ transfer function of the discrete (sampled) time filter

$$1 \;\&\; 2 \Longrightarrow \quad \left[V_{out}(\omega) = V_{in}(z)\,H(z)\Big|_{z=e^{j\omega T}} \cdot T\,\mathrm{sinc}\left(\frac{\omega T}{2}\right) \right]$$

$$V_{in}(z) := \sum_{n=0}^{\infty} v_{in}(nT)\,z^{-n}$$

$$V_{in}(z)\Big|_{z=e^{j\omega T}} = \sum_{n=0}^{\infty} v_{in}(nT)\,e^{-j\omega nT}$$

$$= \int_{-\infty}^{+\infty} v_{in}(t)\sum_{n=0}^{\infty}\delta(t-nT)\cdot e^{-j\omega t}\,dt$$

$$= \mathcal{F}\left[v_{in}(t)\cdot\sum_{n=0}^{+\infty}\delta(t-nT)\right](\omega) \longleftrightarrow$$ Discrete Fourier Transform

" The expression $V_{in}(z)\big|_{z=e^{j\omega T}}$ represents the Fourier transform of the input waveform $v_{in}(t)$ sampled every $t = nT$ i.e. multiplied by $\delta(t-nT)$ "

$$V_{in}(z)\Big|_{z=e^{j\omega T}} = \mathcal{F}\left[v_{in}(t)\cdot\sum_{n}\delta(t-nT)\right](\omega) =$$

$$= V_{in}(\omega) * \mathcal{F}\left[\sum_{n}\delta(t-nT)\right](\omega) =$$

$$= V_{in}(\omega) * \sum_{k}\frac{2\pi}{T}\,\delta\left(\omega - \frac{2\pi}{T}k\right)$$

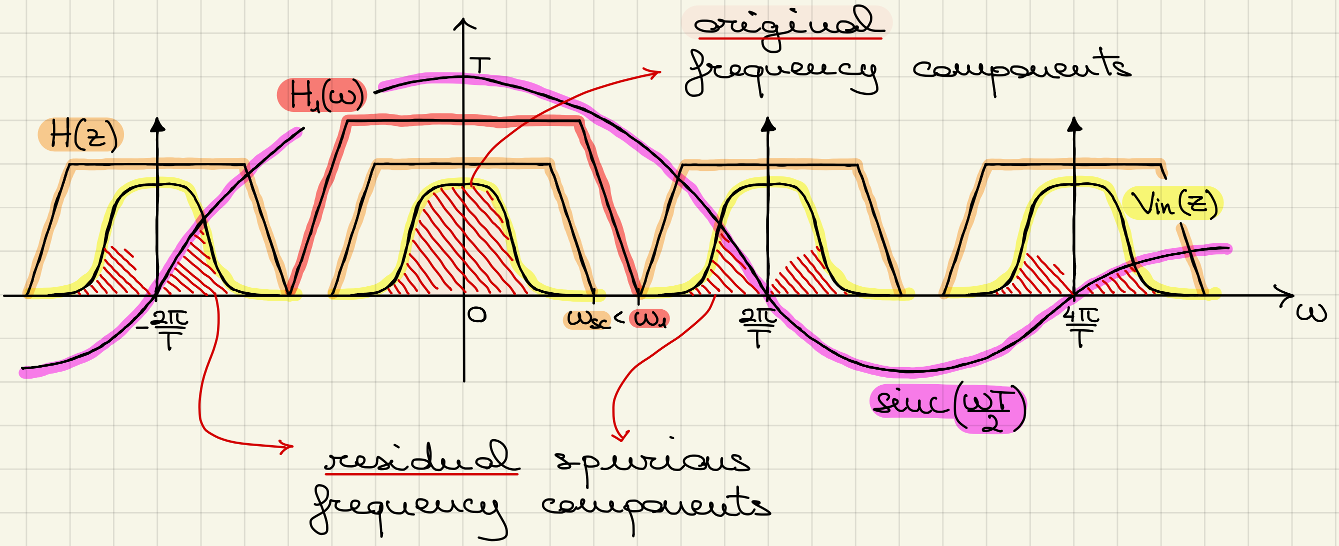
$$\Longrightarrow \quad \boxed{V_{out}(\omega) = \Big[\frac{2\pi}{T}V_{in}(\omega) \underset{ⓐ}{*} \sum_{k}\delta\left(\omega - \frac{2\pi}{T}k\right)\Big]\cdot\Big[H(z)\Big|_{z=e^{j\omega T}}^{ⓑ}\Big]\cdot\Big[T\,\underset{ⓒ}{\mathrm{sinc}}\left(\frac{\omega T}{2}\right)\Big]}$$

Let's see what each term in the final equation means:

ⓐ The first term simply says that due to sampling we are replicating the input spectrum around all the harmonics of the clock (radial) frequency ← aliasing

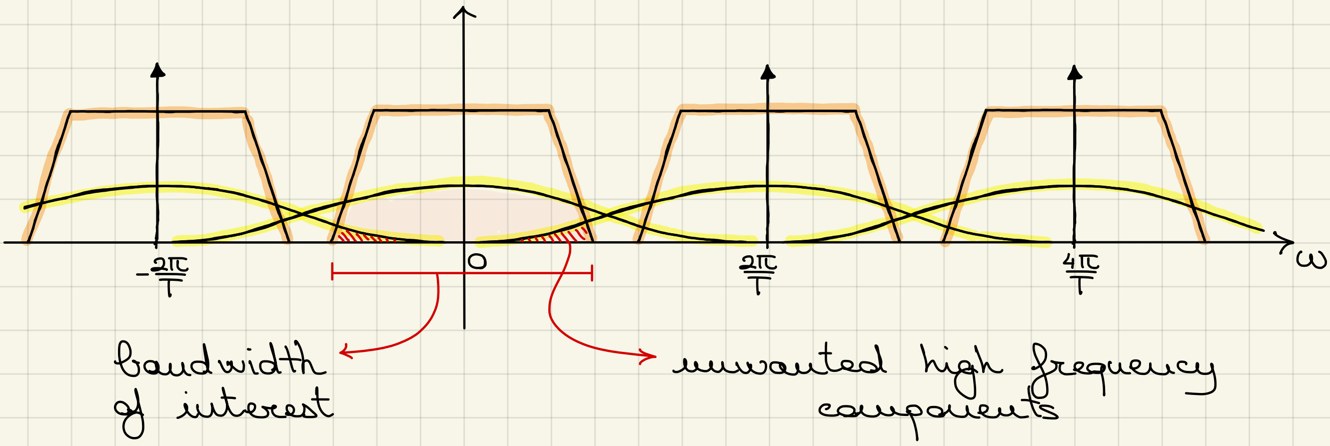
ⓑ The second term yields to the operation of the SC filter itself (an integrator in our case) in the discrete time domain.
Note that $H(z)\big|_{z=e^{j\omega T}} = -\frac{C_1}{C}\frac{1}{z-1} = -\frac{C_1}{C}\frac{1}{e^{j\omega T}-1}$ is a periodic function in ω.
In fact $e^{j\omega T}$ is a periodic function itself with period $\frac{2\pi}{T}$.
Therefore $H(z)\big|_{z=e^{j\omega T}}$ can be seen as a "periodic filter" that acts on each replica of the input spectrum

ⓒ The third term is a cardinal sine centered around the origin and whose zeroes coincide exactly with the clock harmonics ($\text{sinc}\frac{\omega T}{2} = 0 \rightarrow \frac{\omega T}{2} = \pi \rightarrow \omega = \frac{2\pi}{T}$).
Its effect is to amplify the original input spectrum while attenuating other replicas.



In order to improve the fidelity of the output signal we need to kill all residuals at high frequency that are caused by the sampling.
Therefore an additional low-pass filter $H_1(\omega)$ should be placed after the switched capacitor filter to filter off these residual harmonics ("reconstructing filter").
This is not an issue since the needed cut-off of the additional filter is very close to clock frequency ($\geqslant$ 1MHz) hence it can be easily implemented with a standard RC network (remember that the switched capacitor implementation was required only for low frequency cut-off; high frequency filters can be built with normal resistors in the continuous time domain and of course do not suffer from aliasing issues).

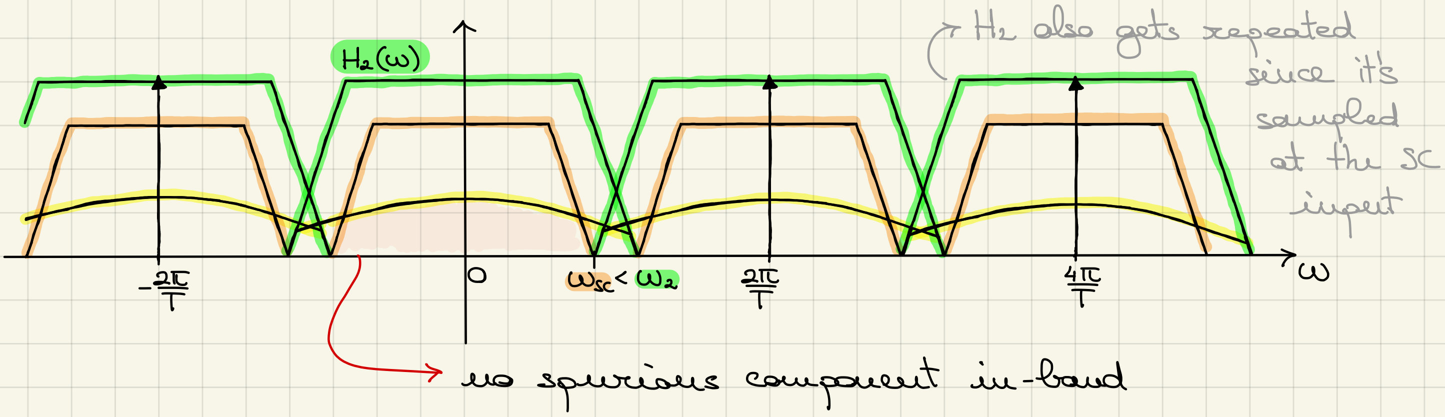
Another issue could be caused by input aliasing when the input spectrum is not just a narrow band but also has some *unwanted* high frequency components, which due to sampling will be brought close to base-band.



bandwidth of interest

unwanted high frequency components

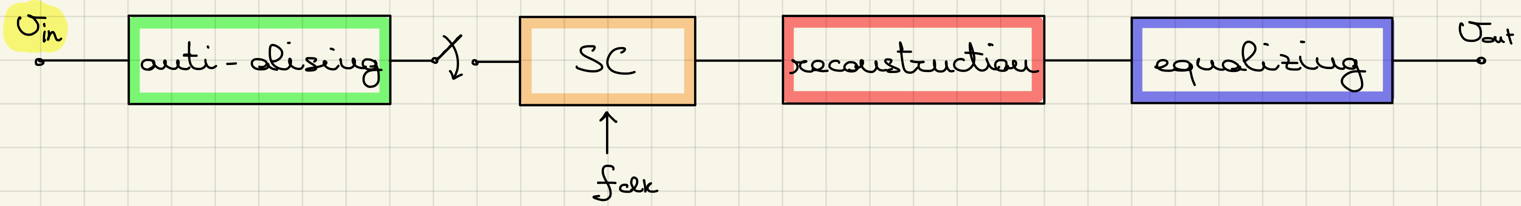
In order to avoid this problem a low-pass filter $H_2(\omega)$ should be placed *before* the switched capacitor filter to remove these high frequency harmonics from the input ("anti-aliasing filter").
This again is not an issue since the newly added cut-off needs to remove only high order harmonics hence it can be much higher than the frequency range of interest and a standard RC implementation is feasible.



H_2 also gets repeated since it's sampled at the SC input

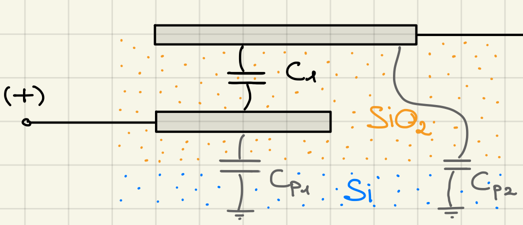
no spurious component in-band

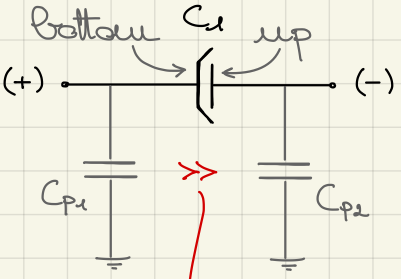
In addition to these two analog filters, a third one is typically used at the end of the filtering chain whose purpose is to compensate the spectrum shape alteration due to the cardinal sinc term ("equalizing filter").



Stray insensitive topologies

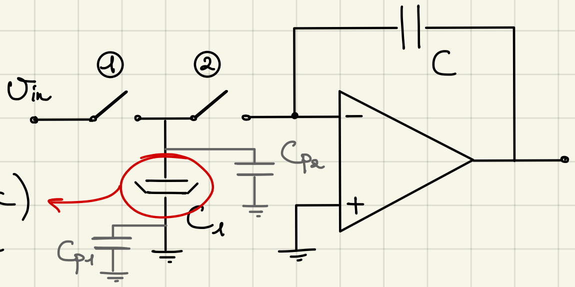
About capacitor structure, parasitic capacitances and their effects on switched capacitors.





because of different distance from conductive substrate

Depending on the configuration, it is more convenient to place the capacitor in the circuit in one way instead of the other:



so that bottom (i.e. larger parasitic) is shorted between grounds

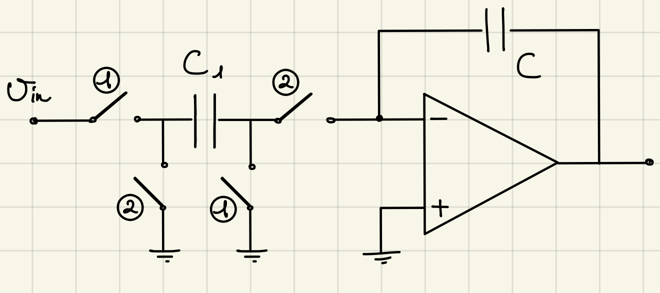
Note that now the charge transfer of the switched capacitor is directly dependent on C_{p2}, which is in parallel with C_1.

$$R_{eq} = \frac{T}{C_1 + C_{p2}}$$

Req now suffers from the high variability of C_{p2}.

How can we avoid this issue?

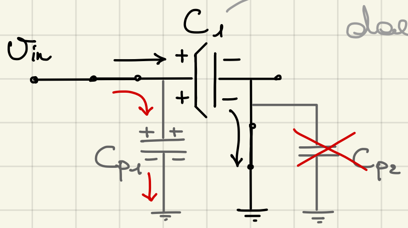
⟶ Stray insensitive configuration



This is one of many topologies that allow to remove the contribution of both parasitic capacitances

Phase 1:

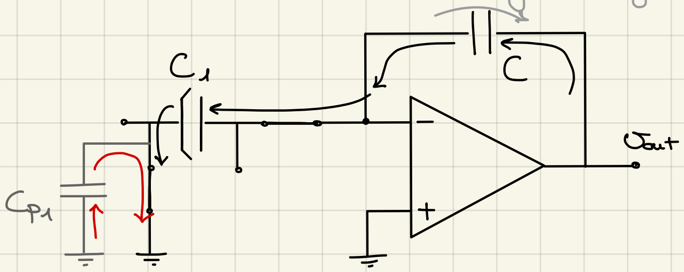
the orientation does not matter in the end.



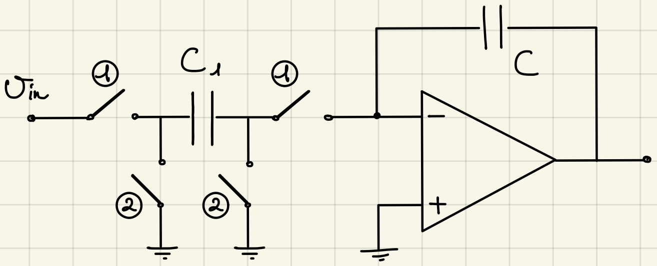
During sampling, C_{p2} is always shorted and doesn't gather charge so it won't contribute to the output. C_{p1} instead gets charged up just like C_1.

Phase 2:

note this is a non-inverting integrator



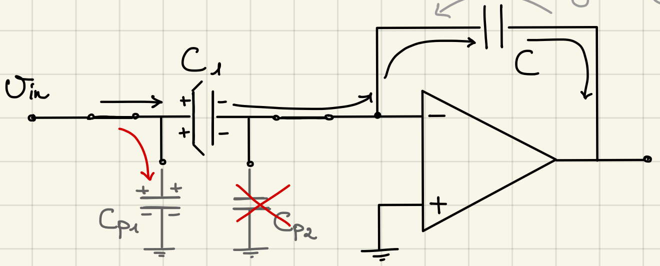
During transfer, C_{p1} is now shorted to ground so it discharges without affecting the output.



Just by inverting the phase of the switches one can obtain a new stray insensitive configuration

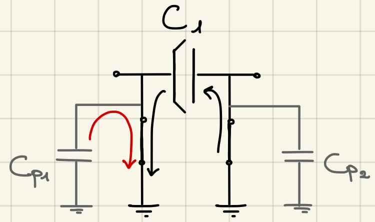
Phase 1:

this is instead an inverting stage



Sampling and transfer now occur together. C_{p2} is always between grounds, while C_{p1} is charged up but does not interact with the circuit.

Phase 2:

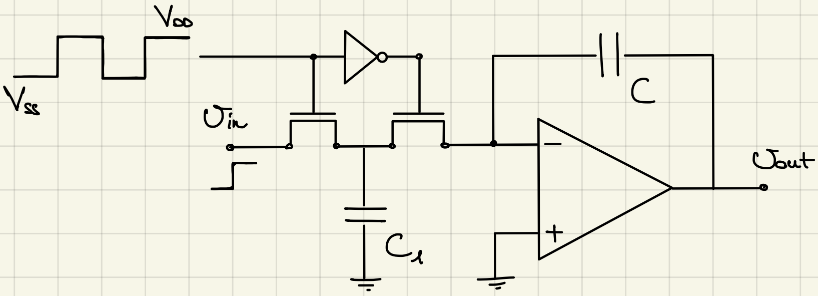


During this "discharge phase" both C_1 and C_{p1} lose the accumulated voltage of the previous half period.

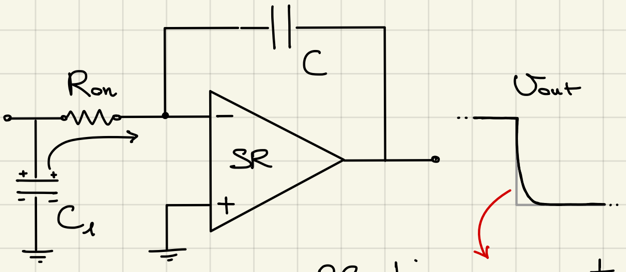
Downside of stray insensitive topologies: more switches are required

Clock feedthrough

About switches structure, their non-idealities and how they affect switched capacitors.



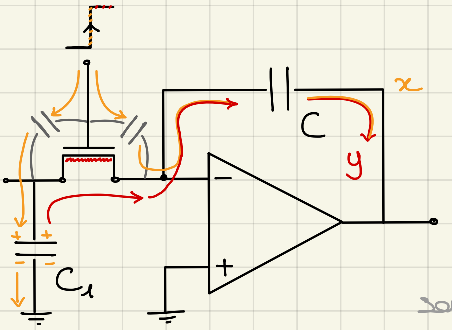
Phase 2:



RC time constant or slew rate limited

The transient due to switch or amplifier non-idealities should be much lower than the clock period (like in the order of nanoseconds) so to have a good sample of the input (i.e. a "nice staircase" at the output).

To reduce the $R_{on} = 1/g_m$ we need to increase the form factor W/L or the overdrive of the transistor.
However, increasing W will result in larger stray capacitances of the switching transistor, which play a relevant role during transitions.



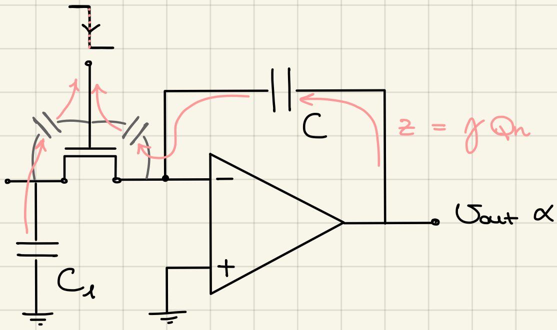
$$v_{out} \propto -(C_1 v_{in} + \underbrace{x+y}_{\sim Q_n})$$

source of error

For instance, during the phase 1 → 2 transition, at the transition edge current is injected both in C_1 and, more importantly, in C causing an immediate change in the output voltage. Then, when the transition is over and the transistor turns on, the charge that was injected in C_1 flows through virtual ground to the output, effectively altering the previously sampled value of v_{in}.

⟶ The entire oxide capacitance of the transistor (both source and drain) contributes to the error at the output.

What about the falling edge, that is, phase 2 → 1 transition?



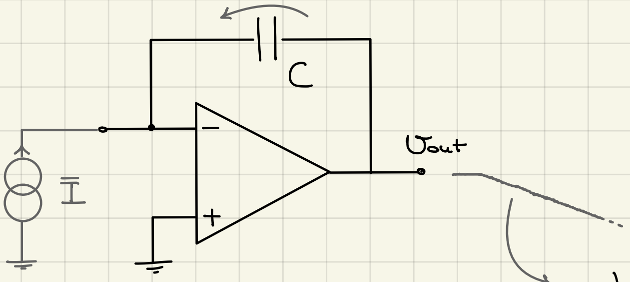
$$V_{out} \propto -(C_1 V_{in} + Q_n(1-\gamma)) \quad \gamma < 1$$

Similarly to before, charge is taken from both C_1 and C during the transition. However, while charge taken from C means a bump up (since it's an inverting stage) of the output, charge taken from C_1 cannot affect the output since the switch will then open (transistor turning off).

So during the trailing edge it is taking out some residual charge from C that is just a portion of the charge injected during the leading edge.
Overall, over a clock period, there will always be some additional charge deposited on C that won't come from the input but from the transistor's gate.

Since this phenomenon occurs at each clock cycle, it resembles the bias current of a continuous time filter: the charge deposited at each cycle basically corresponds to an equivalent current flowing through the feedback capacitor:

$$I = \frac{Q(1-\gamma)}{T} \sim I_{BIAS}$$



output droop caused by the clock feedthrough