BROOKHAVEN
NATIONAL LABORATORY
Instrumentation Division

Front-end design in CMOS for high resolution detectors

Gianluigi De Geronimo

degeronimo@bnl.gov
Brookhaven National Laboratory, Upton, NY 11973, USA

November 2013

Outline

I. Input MOSFET optimization

II. Amplifier design

III. Charge amplification

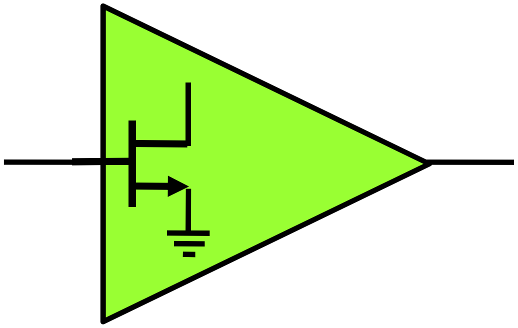
IV. Filter design

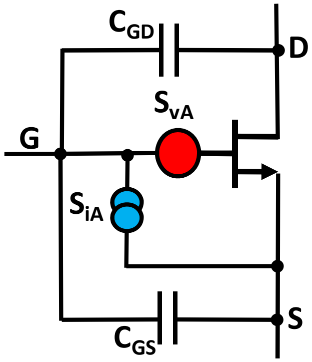
Part I

Input MOSFET optimization

Dominant noise sources

In a properly designed **amplifier**, the dominant noise sources are from the **input transistor**. The other noise sources (e.g. from the cascode, load, etc.) are made negligible.





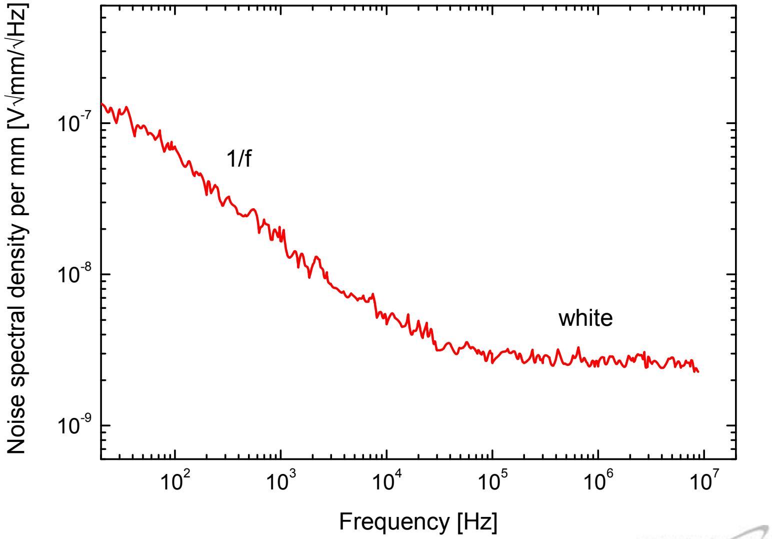
The parameters relevant to the resolution are the **equivalent input noise sources** (series and parallel) and the **input capacitance**.

Basic MOSFET model:

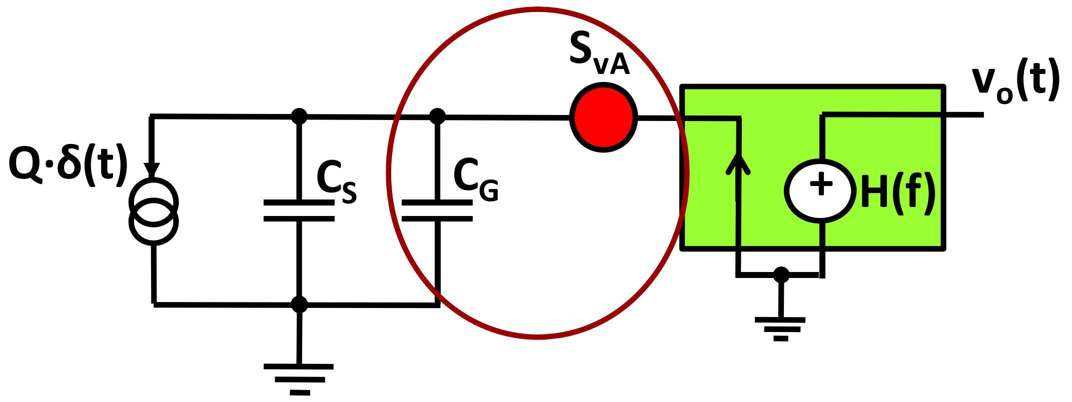
$$S_{vA} \approx \frac{S_{vf1}}{f} + \gamma n \frac{4kT}{g_m} \qquad \begin{cases} n = \dfrac{g_{mS}}{g_m} \approx 1.25 \approx \text{subth.slope} \\ \gamma = 1/2 - 2/3 \ (\text{WI} - \text{SI}) \end{cases}$$

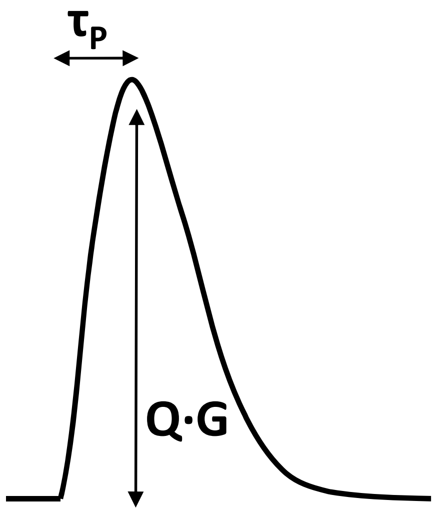
$$S_{iA} \approx \text{negligible}$$

$$C_G \approx C_{GS} + C_{GD}$$



Equivalent Noise Charge (ENC)





C_G depends on the transistor size and/or operating poin

C_S includes all other input capacitors (sensor, parasitics, feedback, ...)

Equivalent Noise Charge (ENC) = V_{orms}/G

$$ENC^2 = 2\pi A_{vfP} S_{vf1} (C_S + C_G)^2 + \frac{A_{vwP}}{\tau_P} \gamma n \frac{4kT}{g_m} (C_S + C_G)^2$$

Typical values **A_{vfP} ≈ 0.5, A_{vwP} ≈ 1** (for unilateral power spectral densities)

G. De Geronimo et al., IEEE TNS 52 (2005)

We must optimize the input MOSFET (L,W,i_{dw})

MOSFET model: densities

We model C_G and S_{VA} as **functions of** gate size (**L, W**) and drain current density **i_{Dw}**

$$C_G \approx C_{GS} + C_{GD} + C_{GB} \approx c_{GWL} WL \approx c_{ox} WL$$

c_{GWL} = gate capacitance per unit area
c_{ox} = gate oxide capacitance per unit area

$$S_{vA} \approx \frac{S_{vf1}}{f} + \gamma n \frac{4kT}{g_m} = \frac{K_f}{c_{ox} WL \cdot f} + \gamma n \frac{4kT}{g_{mW}(i_{DW}, L) W}$$

K_f = 1/f coeff. per unit area (weakly dep. on i_{DW} , L)
γ = gamma coefficient (function of i_{DW})
g_{mW} = g_m per unit W (function of i_{DW} , L)
i_{DW} = I_D per unit W (drain current density)

If we use the following **basic MOSFET equations**

$$I_D \approx \begin{cases} \dfrac{1}{2n} \mu c_{ox} \dfrac{W}{L} (V_{GS} - V_{th})^2 & \text{SI (strong inv.)} \\ \mu c_{ox} V_T^2 \dfrac{W}{L} \exp\left(\dfrac{V_{GS} - V_{th}}{nV_T} \right) & \text{WI (weak inv.)} \end{cases}$$

we can express the g_m as

$$g_m = g_{mW} W \approx \begin{cases} \sqrt{\dfrac{2\mu c_{ox}}{n} \dfrac{W}{L} I_D} = W \sqrt{\dfrac{2\mu c_{ox}}{n} \dfrac{i_{DW}}{L}} & \text{SI} \\ \dfrac{\mu c_{ox} V_T}{n} \dfrac{W}{L} \exp\left(\dfrac{V_{GS} - V_{th}}{nV_T} \right) = \dfrac{I_D}{nV_T} = W \dfrac{i_{DW}}{nV_T} & \text{WI} \end{cases}$$

Low-frequency noise vs L,W

To a first order, we can assume that the change **ΔI** in current due to **trapping / de-trapping** is **inversely proportional to the transit time** (speed of the device).

$$\Delta I \div \frac{1}{\tau} = \omega_T = 2\pi f_T$$

The superposition of the uncorrelated trapping / de-trapping events is, in power, **proportional to the area of the gate**, i.e. WL.

$$\Delta I^2 \div \omega_T^2 WL$$

Hence, the power spectrum (associated to the drain current) at the output of the FET has the same dependence.

$$S_{io} \div \omega_T^2 WL$$

We can **bring it to the input**, dividing by g_m^2 (note: $g_m \approx \omega_T C_G$):

$$S_{vi} = \frac{S_{io}}{g_m^2} \div \frac{\omega_T^2 WL}{\omega_T^2 W^2 L^2} = \frac{1}{WL} \quad \Longrightarrow \quad \frac{S_{vf1}}{f} = \frac{S_{vf1WL}(i_{DW}, L)}{WL \cdot f} \approx \frac{S_{vf1WL}}{WL \cdot f}$$

(i) Due to short channel effects, deep submicron MOSFETs may show a dependence of S_{vf1WL} on L and i_{DW}. (ii) Alternative theories consider the mobility fluctuation.

Optimization in absence of power constraint

The contribution of the MOSFET to the ENC can be written as follows:

$$ENC^2 = 2\pi A_{vfP}\frac{K_f}{c_{ox}L_{min}}\frac{\left(C_S + c_{ox}WL_{min}\right)^2}{W} + \begin{cases} \dfrac{A_{vwP}}{\tau_P}4kT\gamma n^{3/2}\sqrt{\dfrac{L_{min}}{2\mu c_{ox}}}\dfrac{1}{\sqrt{i_{DW}}}\dfrac{\left(C_S + c_{ox}WL_{min}\right)^2}{W} & \text{SI} \\ \dfrac{A_{vwP}}{\tau_P}4kT\gamma n^2 V_T\dfrac{1}{i_{DW}}\dfrac{\left(C_S + c_{ox}WL_{min}\right)^2}{W} & \text{WI} \end{cases}$$

- We operate the MOSFET in SI and select $\mathbf{i_{DW}=i_{DWmax}}$.($i_{DW}$ at maximum V_{GS}-V_{th})
- The series term is minimized for $\mathbf{L=L_{min}}$
- The white and 1/f terms are minimized for $W=C_S/c_{ox}L$ or $\mathbf{C_G=C_S}$ (**capacitive matching**)

The $\mathbf{ENC_{min}}$ (i.e. the ENC at maximum power and optimum W) can be written as:

$$ENC^2_{min} = 4C_S\left(\frac{A_{vwP}}{\tau_P}4kT\gamma n^{3/2}L^{3/2}_{min}\sqrt{\frac{c_{ox}}{2\mu}}\frac{1}{\sqrt{i_{DW\,max}}} + 2\pi A_{vfP}K_f\right)$$

Power prohibitively high for almost all of practical cases

With power constraint: moderate inversion

We impose a **limit to the power dissipation $I_D=i_{DW}W\leq I_{D0}$**. The ENC can be written as:

$$ENC^2 = 2\pi A_{vfP}\frac{K_f}{c_{ox}L_{min}}\frac{(C_S + c_{ox}WL_{min})^2}{W} + \begin{cases} \dfrac{A_{vwP}}{\tau_P}4kT\gamma n^{3/2}\sqrt{\dfrac{L_{min}}{2\mu c_{ox}}}\dfrac{1}{\sqrt{I_{D0}}}\dfrac{(C_S + c_{ox}WL_{min})^2}{\sqrt{W}} & \text{SI} \\ \dfrac{A_{vwP}}{\tau_P}4kT\gamma n^2 V_T\dfrac{1}{I_{D0}}(C_S + c_{ox}WL_{min})^2 & \text{WI} \end{cases}$$

- The 1/f term still has a minimum for $W=C_S/c_{ox}L_{min}$ or **$C_G=C_S$** (capacitive matching)
- The white term has a minimum for:
  - $W = C_S/3c_{GW}$ or **$C_G=C_S/3$** in strong inversion (SI)
  - W → 0 or **$C_G = 0$** in weak inversion (WI) **note: W→0 pushes back towards SI**

Most of the applications typically impose a limit of less (or much less) than 1 mW per pixel. With these constraints the input MOSFET frequently operates in **moderate inversion** (between weak and strong inversion).

The optimization requires a **model in the region of moderate inversion**.

MOSFET in moderate inversion: g_m

Transconductance g_{mW}

$$IC = \frac{i_{DW} L}{2nV_T^2 \mu c_{ox}}$$ inversion coefficient

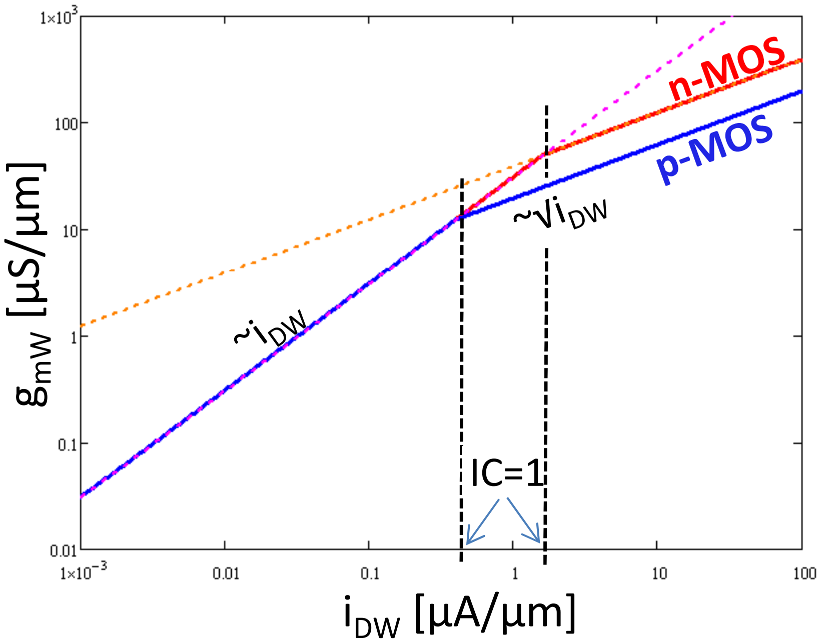
$$g_{mW} \approx \begin{cases} \sqrt{\dfrac{2\mu c_{ox}}{n}\dfrac{i_{DW}}{L}} & IC > 10 \quad \text{SI} \\ \dfrac{i_{DW}}{nV_T} & IC < 0.1 \quad \text{WI} \\ ?? & 0.1 < IC < 10 \quad \text{MI} \end{cases}$$

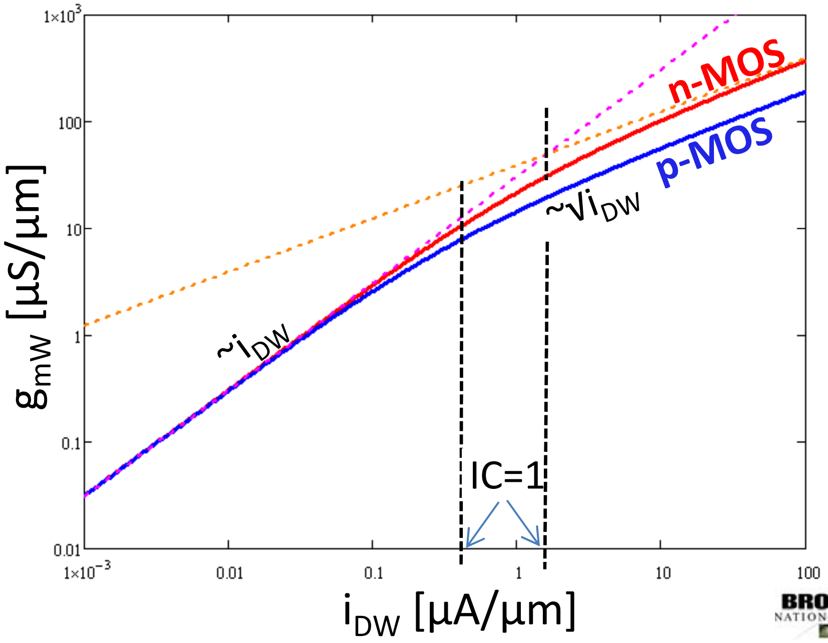
(EKV)

$$g_{mW}(IC) \approx \frac{i_{DW}}{nV_T}\frac{\sqrt{1+4\cdot IC}-1}{2\cdot IC} =$$

$$= \frac{V_T \mu c_{ox}}{L}\left(\sqrt{1+4\cdot IC}-1\right)$$

alternatively: extract from simulator



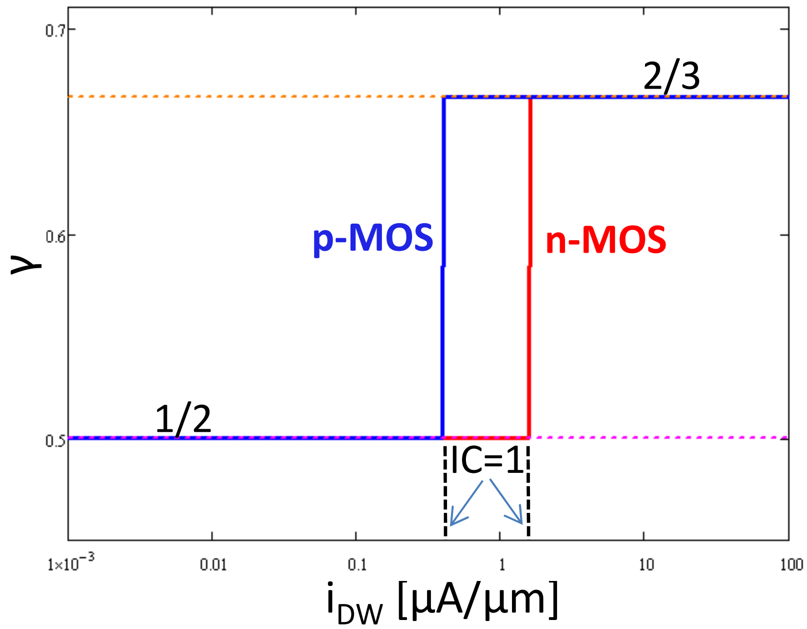


MOSFET in moderate inversion: γ

Gamma coefficient γ

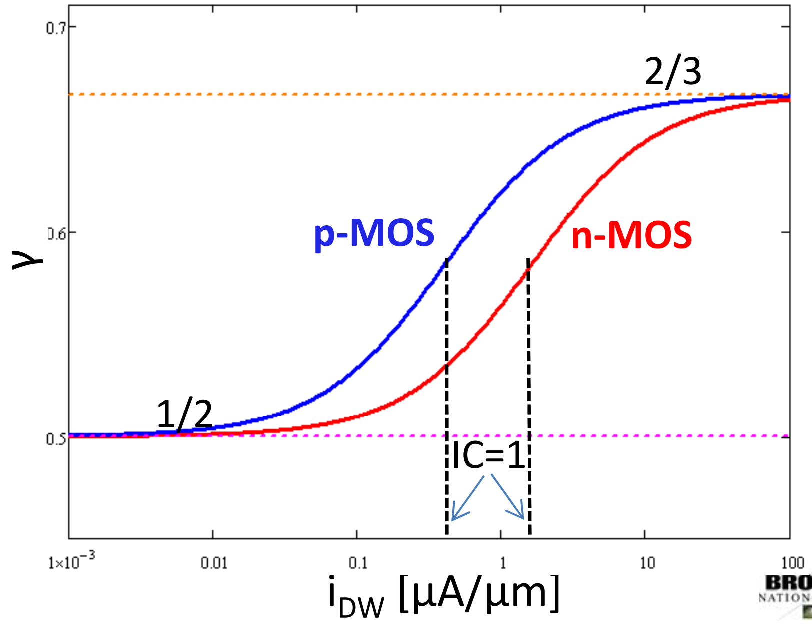
$$IC = \frac{i_{DW} L}{2nV_T^2 \mu c_{ox}}$$ inversion coefficient

$$\gamma \approx \begin{cases} \frac{2}{3} & IC > 10 \quad \text{SI} \\ \frac{1}{2} & IC < 0.1 \quad \text{WI} \\ ?? & 0.1 < IC < 10 \quad \text{MI} \end{cases}$$



(EKV)

$$\gamma(IC) \approx \frac{1}{1+IC}\left(\frac{1}{2} + \frac{2}{3} IC\right)$$



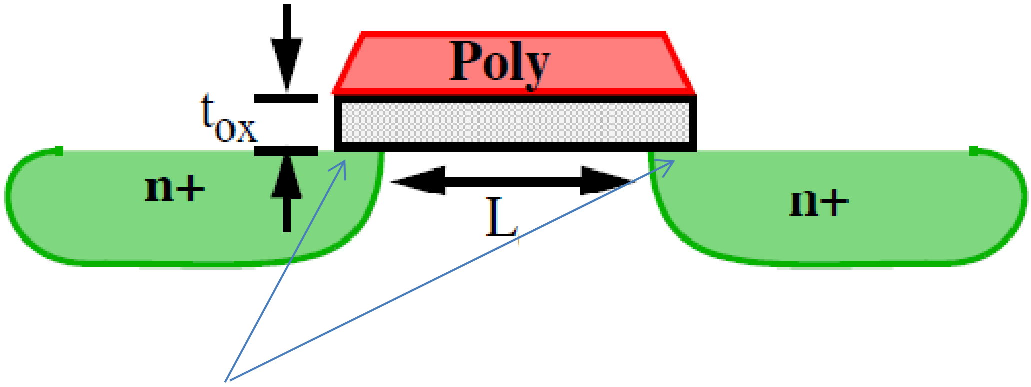
BROOKHAVEN
NATIONAL LABORATORY
Instrumentation Division

MOSFET in moderate inversion: C_G

Gate capacitance C_G

$$C_G \approx c_{ox} WL$$

$$C_G \approx 2c_{ov} W + \frac{2}{3} c_{ox} WL$$



c_{ov}=drain/source overlap capacitance per unit W

c_{ov}≈0.5 fF/µm for CMOS0.25µm

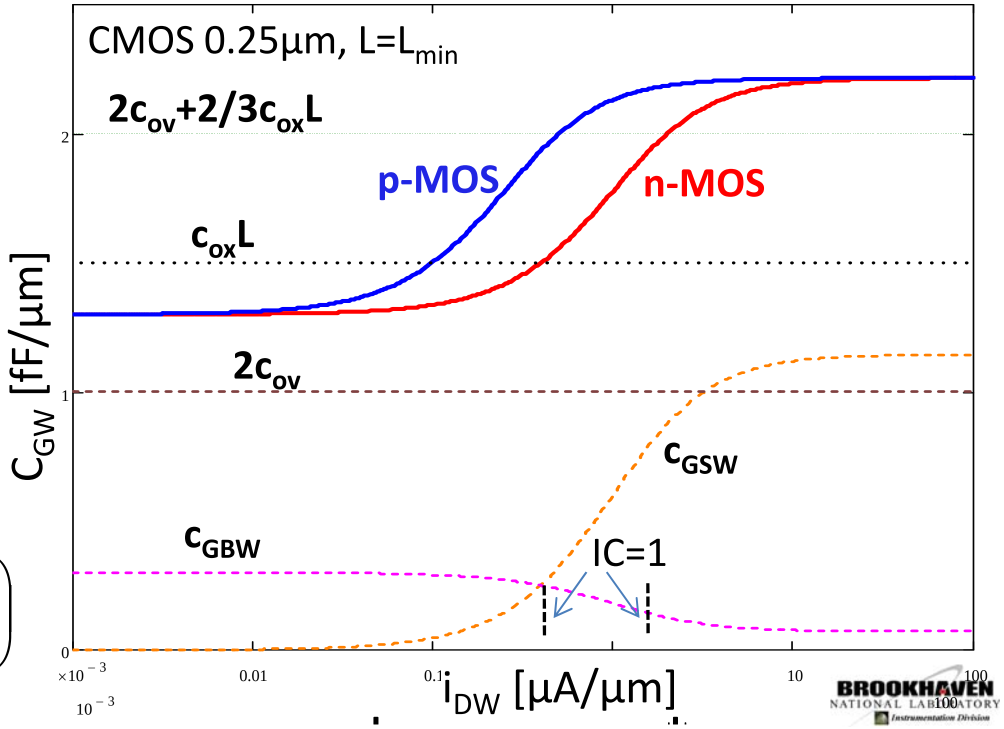
(EKV)

$$\gamma_C(IC) \approx \left(\frac{3}{2} + \frac{1}{3}\frac{\sqrt{1+4\cdot IC}+1}{IC^2}\right)^{-2/3}$$

$$c_{GSW} \approx c_{ox} L \gamma_C(IC)$$

$$c_{GBW} \approx c_{ox} L \frac{n-1}{n}\left[1-\gamma_C(IC)\right]$$

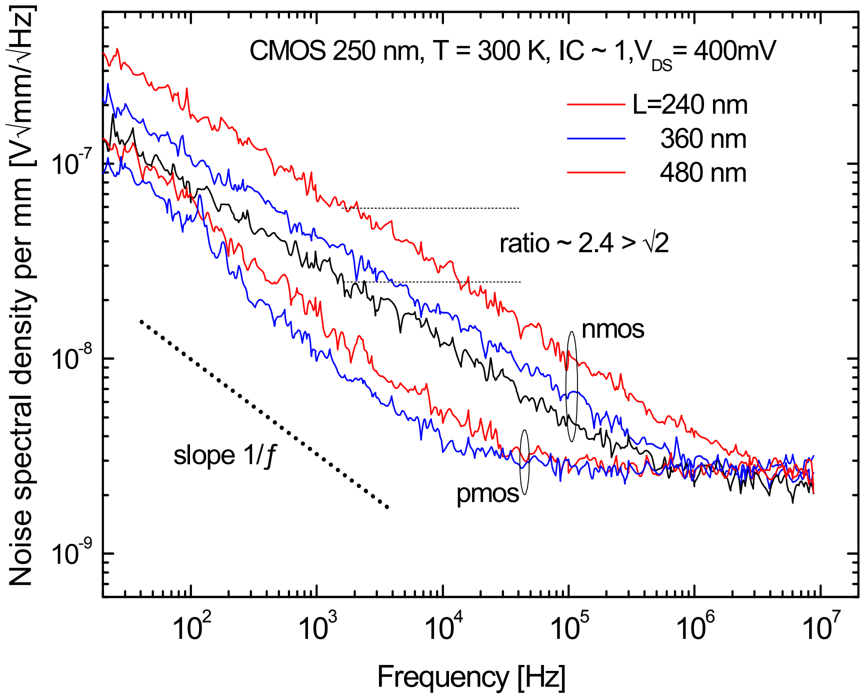
$$c_{GW} \approx 2c_{ov} + c_{ox} L\left(\gamma_C(IC) + \frac{n-1}{n}\left[1-\gamma_C(IC)\right]\right)$$



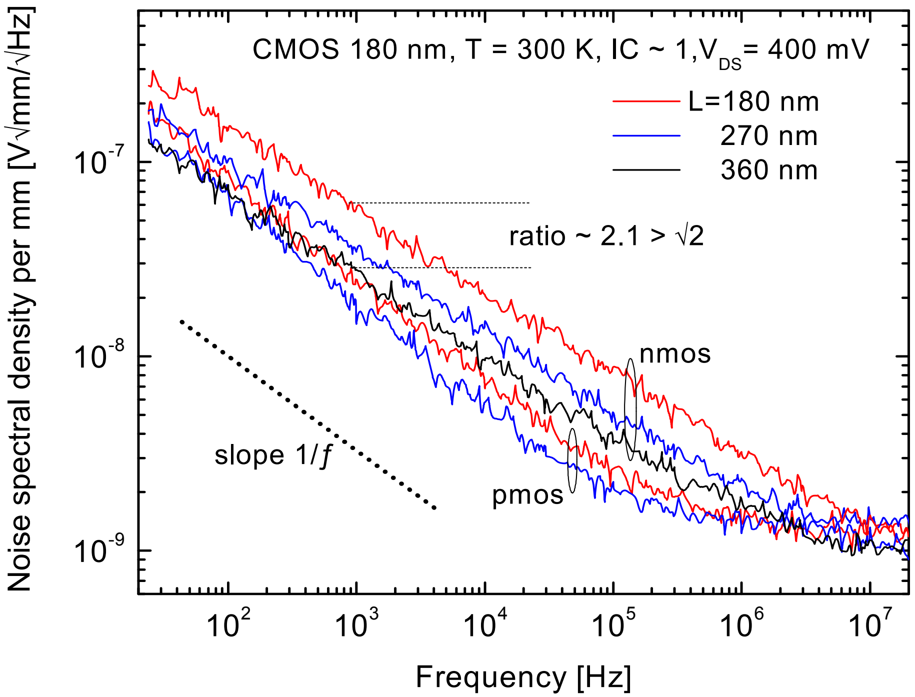
Low-frequency noise: amplitude and slope

The selected technology **should be characterized** for low-frequency noise

CMOS 250nm



CMOS 180nm



Note slope of pmos vs nmos

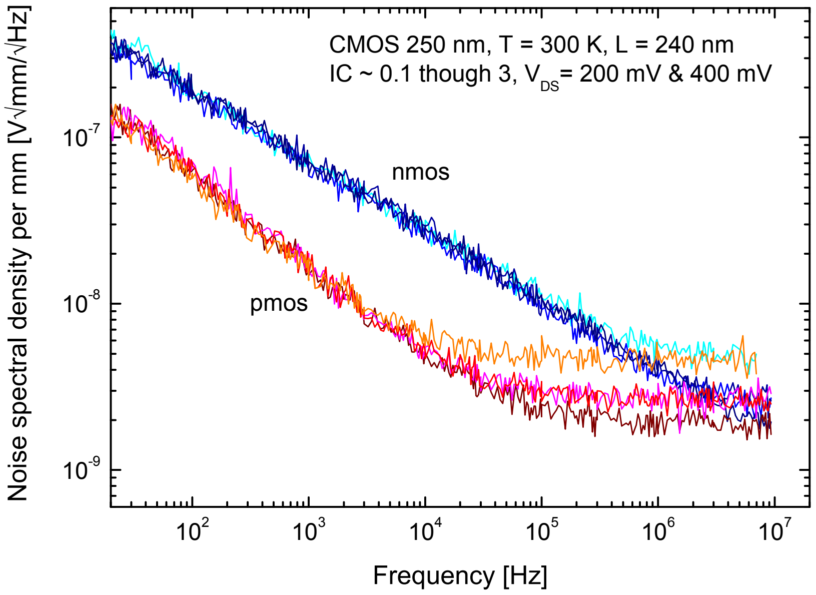
$$S_{vA\_lf} \approx \frac{K_f}{c_{ox}WL\cdot f} \longrightarrow S_{vA\_lf} \approx \frac{K_f(IC,L)}{c_{ox}WL\cdot f^{\alpha_f(IC,L)}}$$

$$ENC_{lf}^2 = 2\pi A_{vfP}\frac{K_f(IC,L)}{c_{ox}WL}(C_S+C_G)^2$$

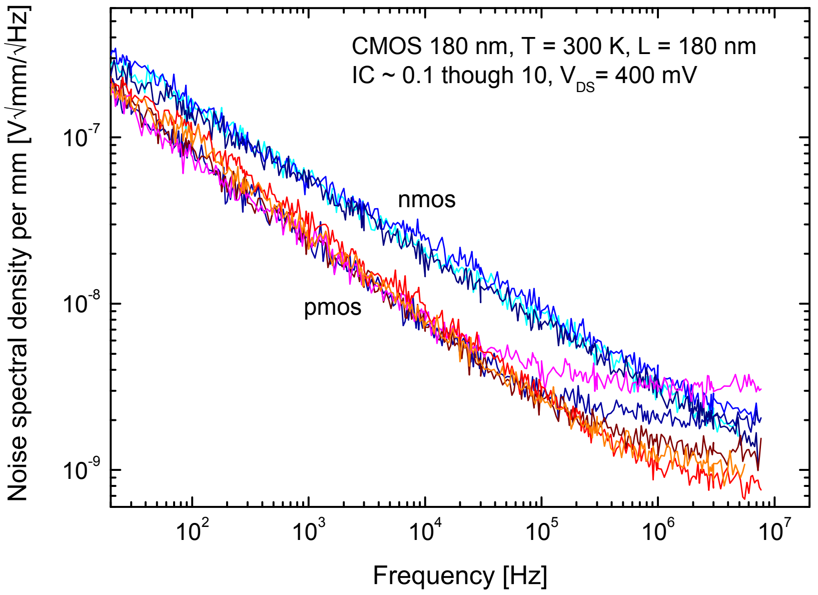
$$ENC_{lf}^2 = (2\pi)^{\alpha}\frac{A_{vfP}(\alpha)}{\tau^{1-\alpha_f(IC,L)}}\frac{K_f(IC,L)}{c_{ox}WL}(C_S+C_G)^2$$

Low-frequency noise vs i_{DW}

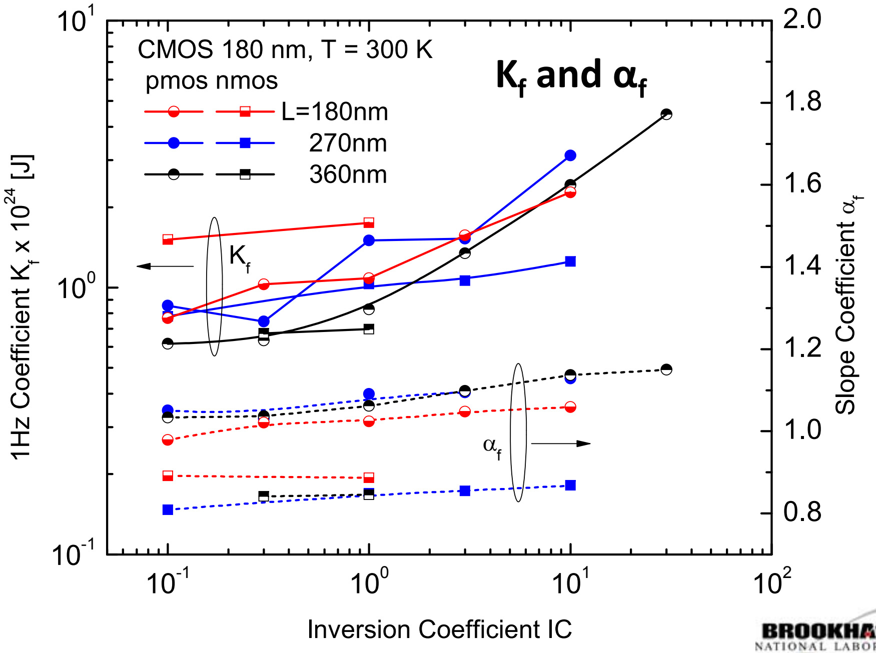
CMOS 250nm



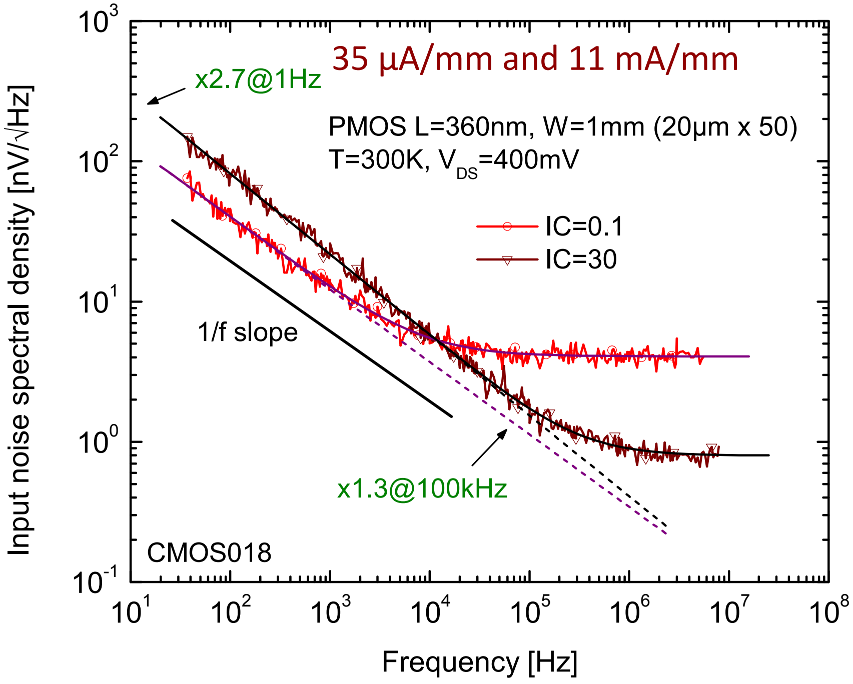
CMOS 180nm

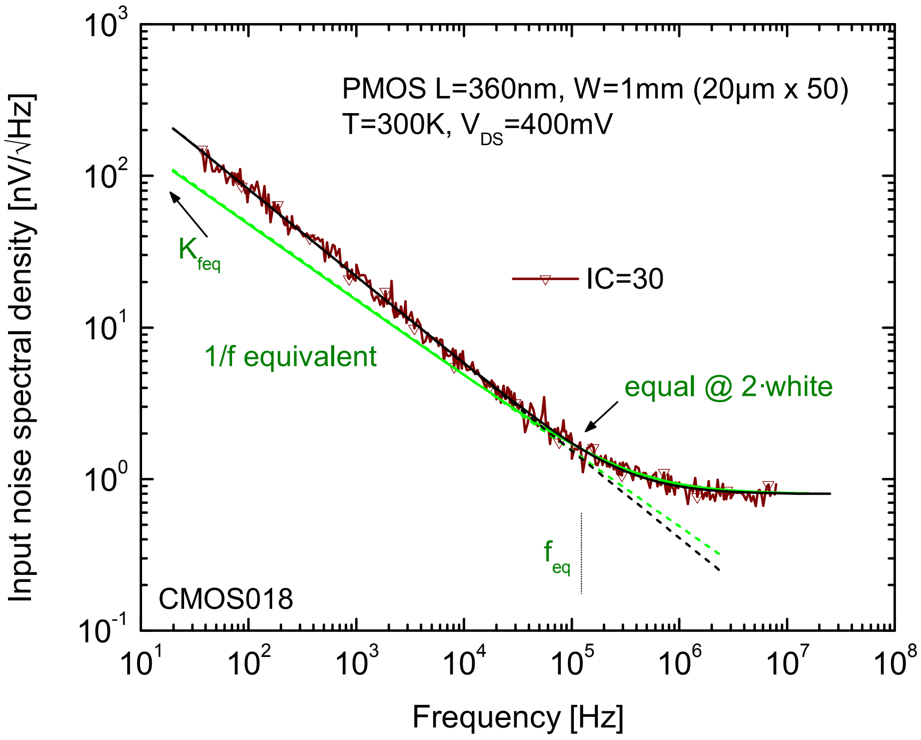


- The dependence of K_f and α_f on the **operating point** is usually small, more visible in deeper submicron technologies
- Note that:
  - **K_f increases** with IC
  - **also α_f increases** with IC



The 1/f-equivalent model



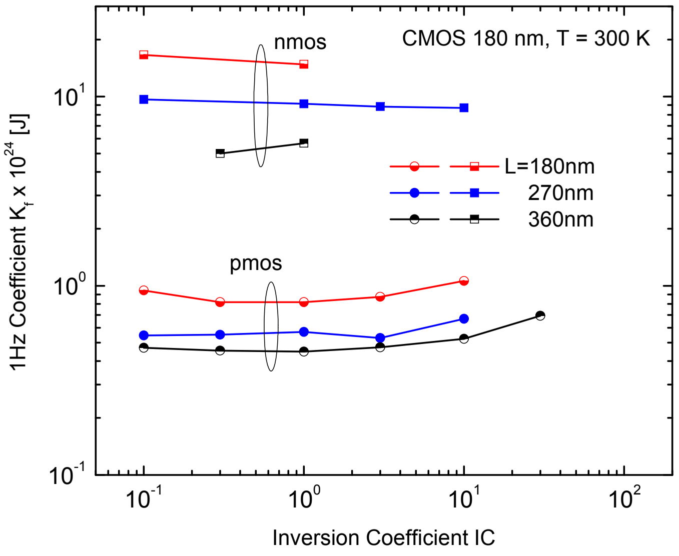


<u>Equivalent 1/f</u>: **equal value at twice the white component** (four times in power)

$$\frac{K_{feq}}{C_{ox}WLf_{eq}} + \text{white} = S_v\left(f_{eq}\right)$$

$$\text{where} \quad S_v\left(f_{eq}\right) = 4 \times \text{white}$$

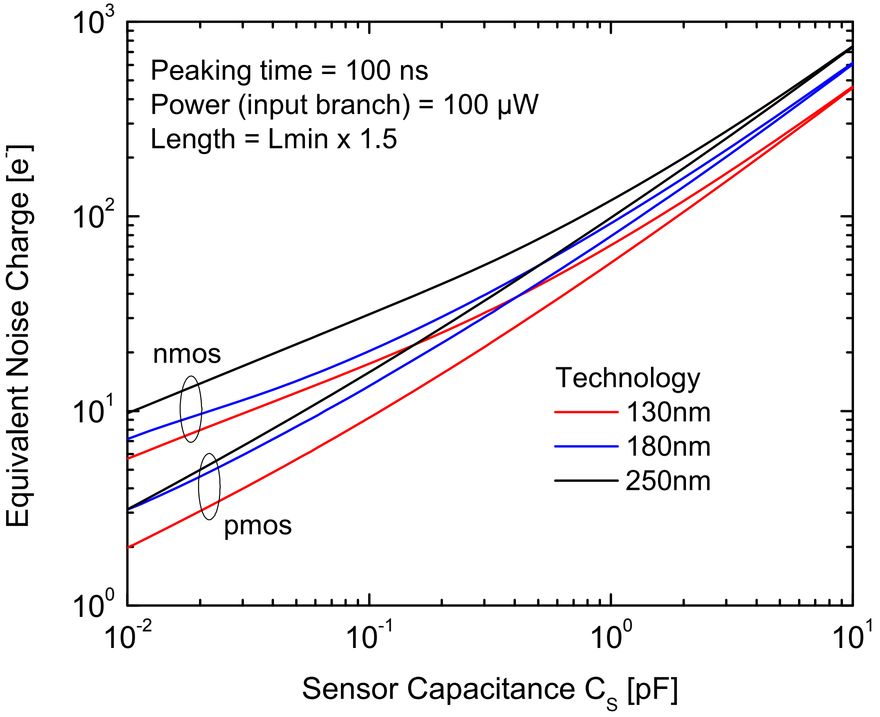
$$\Rightarrow K_{feq} = C_{ox}WLf_{eq}3 \times \text{white}$$



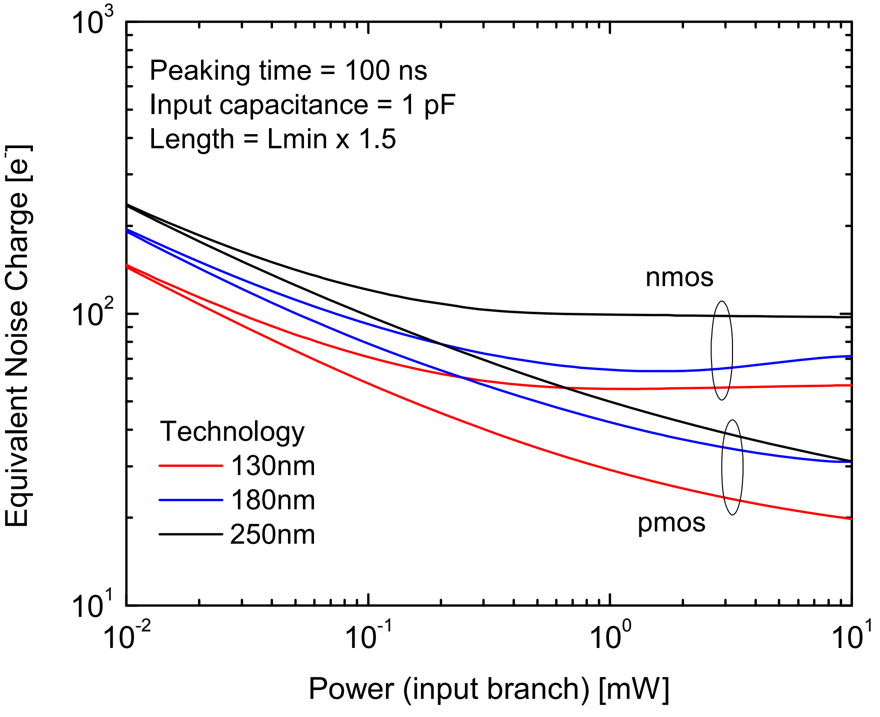
G. De Geronimo et al., IEEE TNS 58 (2011)

Example of ENC vs input capacitance and power

ENC_{min} vs C_S



ENC_{min} vs P_d



- **PMOS** offer typically a better resolution due to higher slope of low-frequency noise.

- **Smaller technology nodes** offer better resolution at equal dissipated power (due to higher drain current from lower voltage).

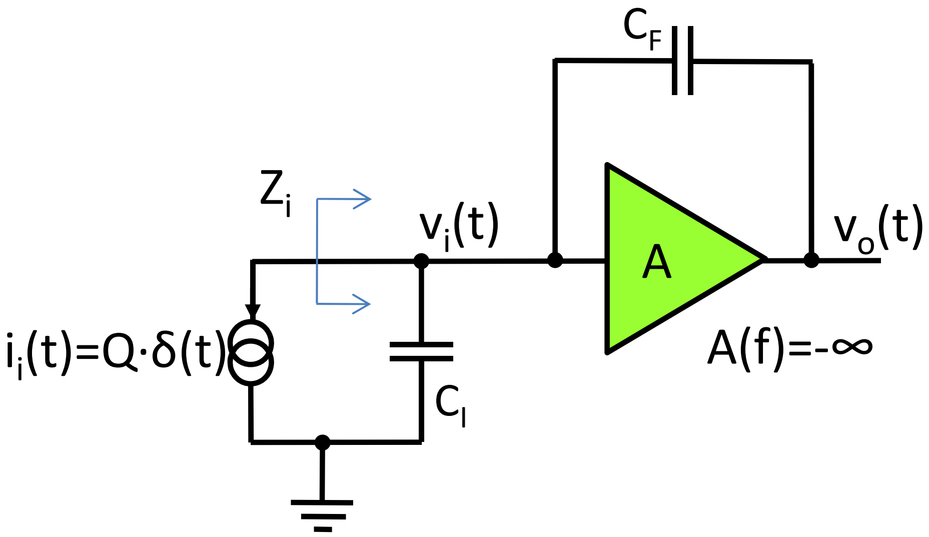
- The resolution **flattens with power** due to the low-frequency noise. A **minimum** may occur due to increase in capacitance density when entering strong inversion with low-frequency dominant (i.e. nmos).

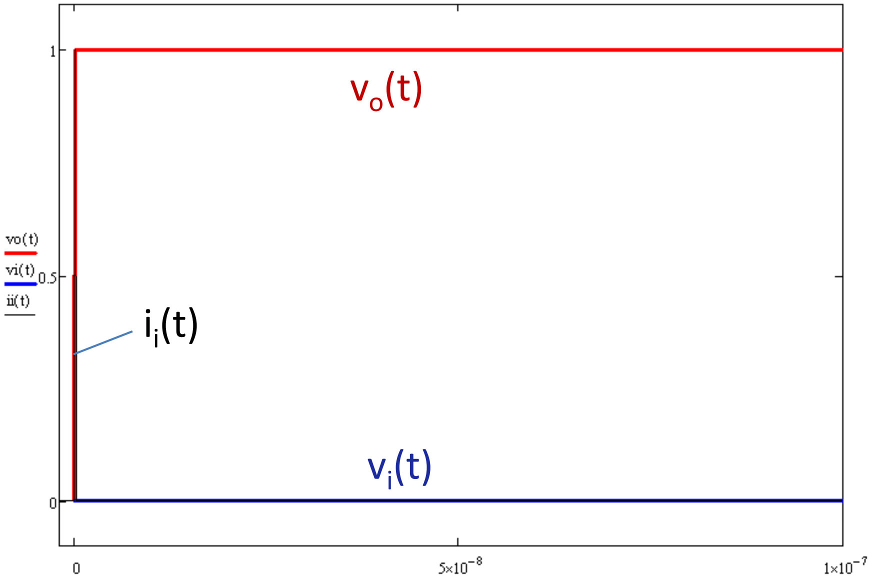
Part II

Amplifier design

Charge amplifier response

We start assuming a charge amplifier realized using an **ideal** voltage amplifier with infinite gain and bandwidth, i.e. A(f)=-∞.



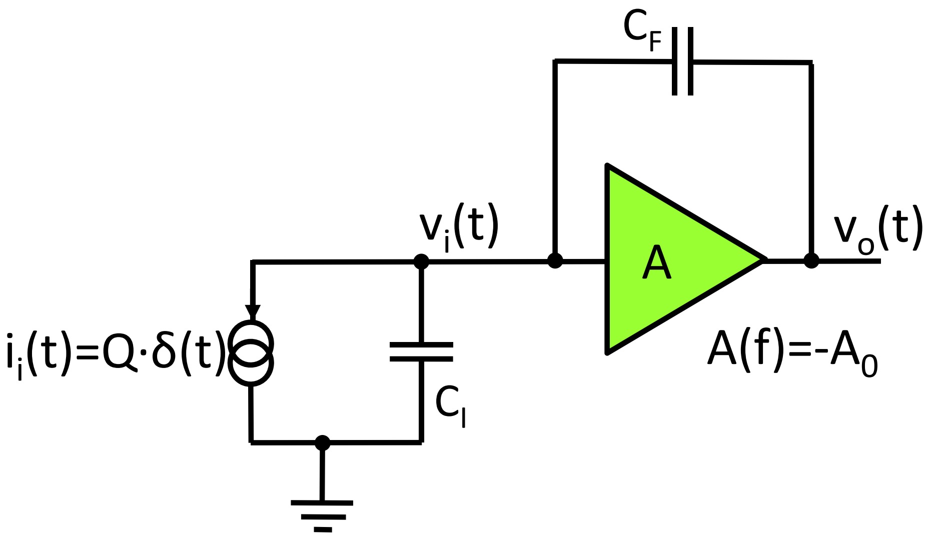


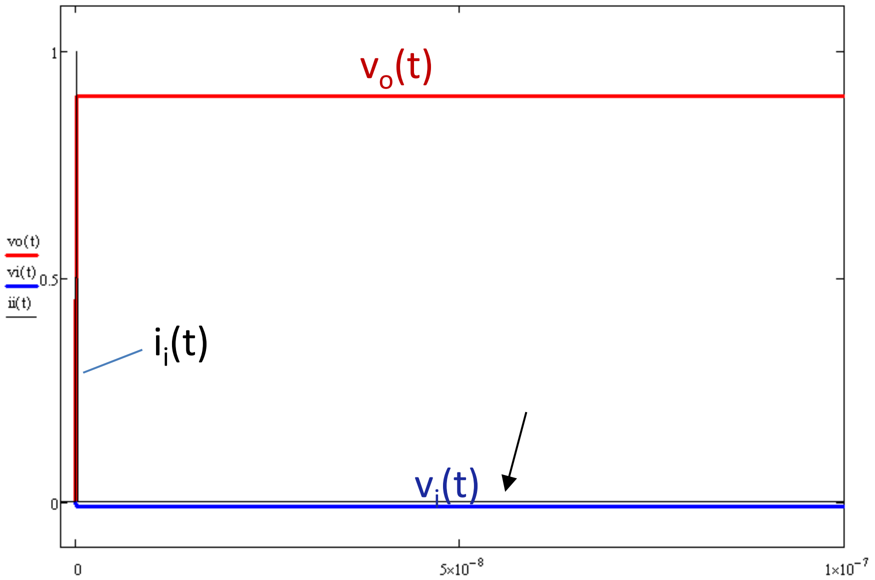
The response $v_o(t)$ of the system to a delta current δ(t) (area Q) is a step, while the input node is steady at the virtual ground. The impedance $Z_i(f)$ seen by the signal source is zero. The loop gain $G_L(f)$ is infinite.

$$v_o(t) = \Phi(t)\cdot\frac{Q}{C_F} \qquad v_i(t) = 0 \qquad Z_i(f) = 0 \qquad G_L(f) = \left| A(f)\cdot\frac{C_F}{C_I + C_F} \right| = \infty$$

Constraints on dc gain for charge amplifiers

In a **semi-realistic** case, A(f) has **finite dc gain** -A_o and **infinite bandwidth**, i.e**. A(f)=-A_o**.





The **response** $v_o(t)$ of the system to a delta current δ(t) (area Q) is a step, but attenuated by $(1+1/G_{L0})$, while the **input node** is an opposite step, A_0 times smaller.

$$V_o(f) = \frac{Q}{sC_F\left(1+1/G_{L0}\right)} \qquad V_i(f) = \frac{V_o(f)}{A(f)} = \frac{-Q}{sC_F A_0\left(1+1/G_{L0}\right)} \qquad G_L(f) = \left|-A_0\cdot\frac{C_F}{C_I+C_F}\right|$$

$$v_o(t) = \Phi(t)\frac{Q}{C_F\left(1+1/G_{L0}\right)} \qquad v_i(t) = -\frac{v_o(t)}{A_0} \qquad G_{L0} = G_L(0)$$

What is the **impact of low dc gain A_0** on the front-end performance**?**

Constraints on dc gain

The attenuation at $v_o(t)$ can be recovered by increasing the gain in the next stages, but a low dc gain A_0 can have two relevant effects:

(1) - **Dependence** of charge gain **on A_0 (i.e. on active elements, temperature, ...).**

We minimize this dependence by increasing the dc loop gain G_{L0}. As a rule of thumb, a minimum loop gain of 100 should be achieved:

$$G_{L0} = \left| -A_0 \cdot \frac{C_F}{C_I + C_F} \right| > 100$$

(2) - Increase in **cross-talk between channels**.

Assuming an inter-pixel capacitance C_C,
a voltage step at the input injects, at the input of the neighbor channel, a charge:

$$Q_C \approx \frac{Q}{A_0 C_F} C_C$$

We impose

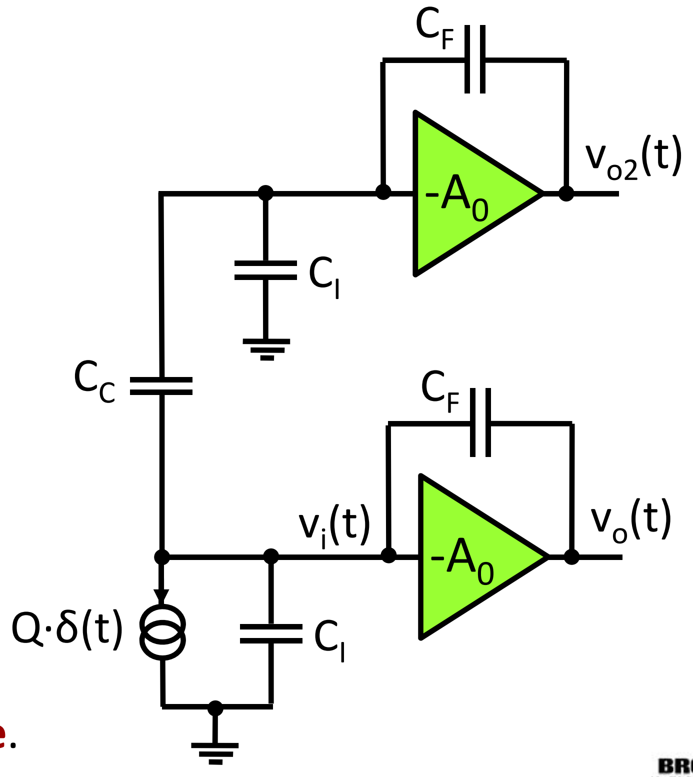
$$Q_{C\,max} \approx \frac{Q_{max}}{A_0} \frac{C_C}{C_F} < ENC$$

i.e.

$$A_0 > \frac{Q_{max}}{ENC} \frac{C_C}{C_F} = DR \frac{C_C}{C_F}$$

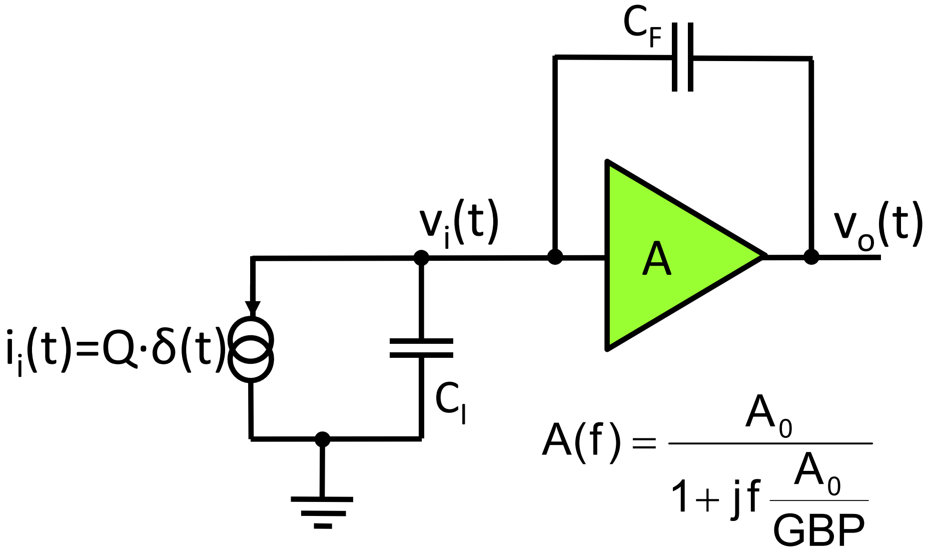
Example:
DR≈500, C_C/C_F≈0.2
-> A_0>100

where **DR=Q_{max}/ENC** is the **analog dynamic range**.

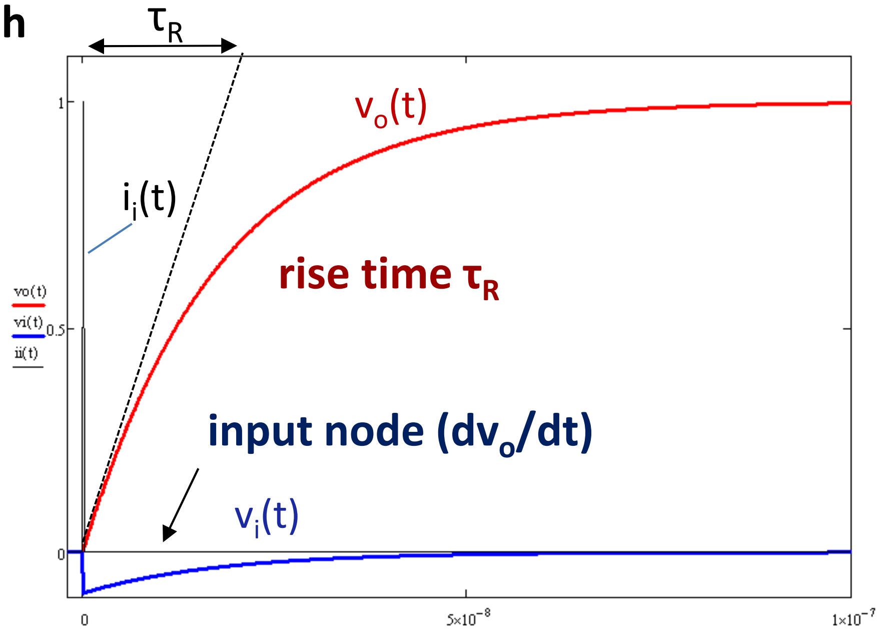


Constraints on bandwidth

In a **more realistic** case, A(f) has **finite bandwidth**



$$A(f)=\frac{A_0}{1+jf\frac{A_0}{GBP}}$$



rise time τ_R

input node (dv_o/dt)

Assuming $G_{L0} >> 1$ we can write:

$$V_o(f)\approx\frac{Q}{sC_F}\frac{1}{1+j2\pi f\tau_R} \qquad V_i(f)=\frac{V_o(f)}{A(f)} \qquad \tau_R\approx\frac{\tau_A}{G_{L0}}=\frac{1}{2\pi GBP}\frac{C_I+C_F}{C_F}$$

$$v_o(t)\approx\Phi(t)\frac{Q}{C_F}\left(1-e^{-\frac{t}{\tau_R}}\right) \qquad v_i(t)\approx-\frac{Q}{C_FA_0}\left[1-\left(1-\frac{\tau_A}{\tau_R}\right)e^{-\frac{t}{\tau_R}}\right]\approx-\frac{Q}{C_I+C_F}e^{-\frac{t}{\tau_R}}$$

Constraints on bandwidth

A low GBP can have two relevant effects:

(1) - **Dependence** of shaped output **on rise time (i.e. on active elements, temperature, ...).**

We minimize this dependence by increasing the GBP. As a rule of thumb, the rise time should be < 0.1 x peaking time:

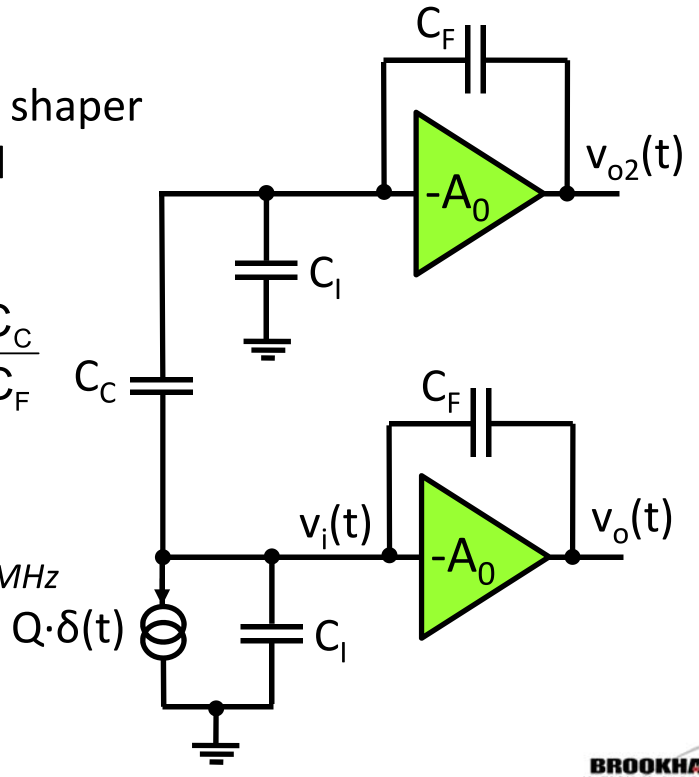
$$\tau_R < 0.1\tau_P \quad \rightarrow \quad GBP > \frac{10}{2\pi\tau_P}\frac{C_I + C_F}{C_F}$$

(2) - Increase in **cross-talk between channels**.

Assuming an inter-pixel capacitance C_C, the additional exponential voltage at the input injects, at the input of the neighbor channel, a current with zero area.

The current generates at the output of the neighbor shaper a signal whose peak amplitude can be approximated with (K = shaper gain):

$$v_{o2} \approx \frac{QK}{C_F}\frac{\tau_R}{\tau_P}\frac{C_C}{C_I} \approx \frac{QK}{2\pi GBP\tau_P}\frac{C_C}{C_F^2} \quad \rightarrow \quad Q_{Ceq} \approx \frac{Q}{2\pi GBP\tau_P}\frac{C_C}{C_F}$$



We impose

$$Q_{Ceqmax} < ENC$$

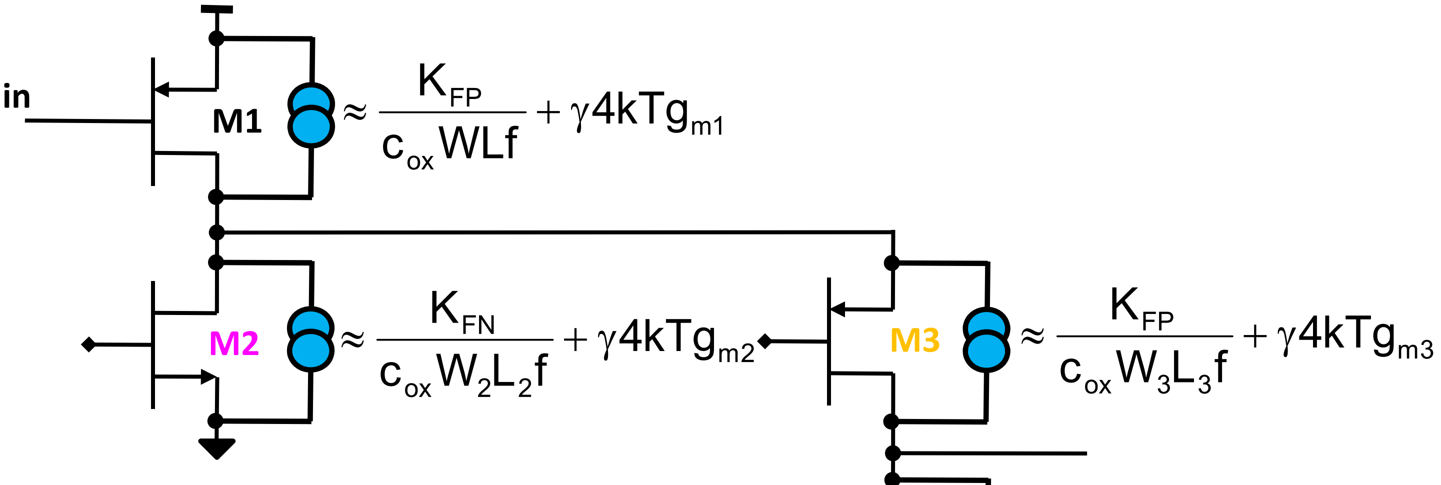
Example:
DR≈500, C_C/C_F≈0.2,
τp=100ns -> GBP>150 MHz

which gives:

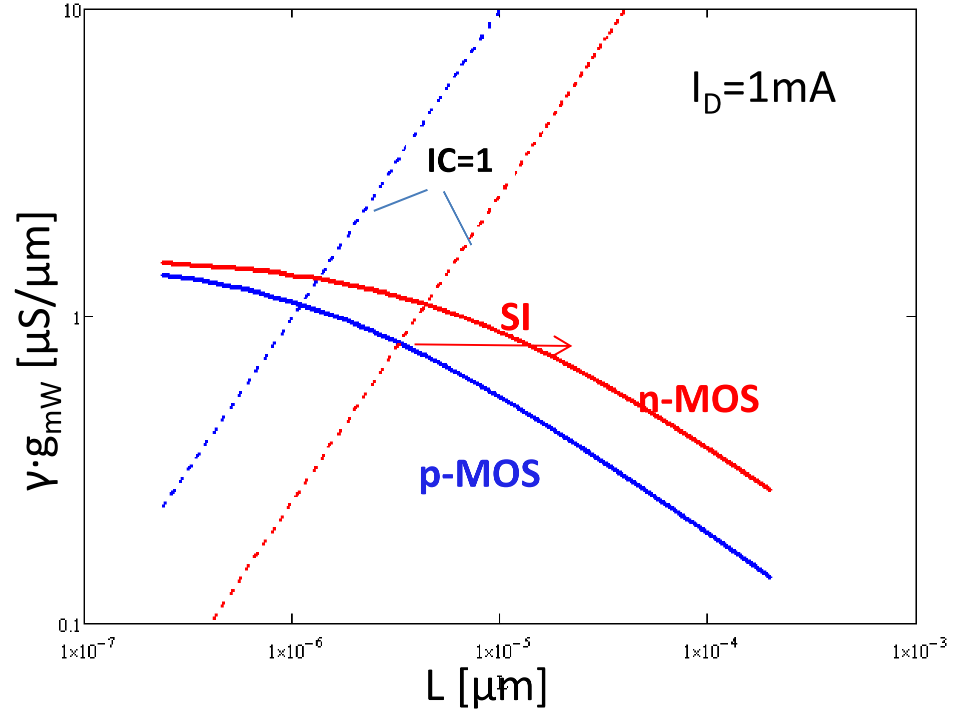
$$\frac{Q_{max}}{2\pi GBP\tau_P}\frac{C_C}{C_F} < ENC \quad \rightarrow \quad GBP > \frac{DR}{2\pi\tau_P}\frac{C_C}{C_F}$$

Noise from current sources

We use **approximate equations** for the noise spectral densities. First we compare the white terms, then the low-frequency terms. Let’s start with the **sources M2**, **M4**.



in
M1
$\approx \frac{K_{FP}}{c_{ox}WLf} + \gamma 4kTg_{m1}$
M2
$\approx \frac{K_{FN}}{c_{ox}W_2L_2f} + \gamma 4kTg_{m2}$
M3
$\approx \frac{K_{FP}}{c_{ox}W_3L_3f} + \gamma 4kTg_{m3}$
M4
$\approx \frac{K_{FN}}{c_{ox}W_4L_4f} + \gamma 4kTg_{m4}$



Observe g_m vs L for a given drain current: **the higher the channel length L, the lower the white noise (g_m/I_D).**

In order to minimize the white noise the MOSFET should operate **at the highest possible L (i.e. strong inversion)**.

Noise from current sources

To achieve this result, we size the load MOSFET for a given V_{DS0} as follows:

1. **set V_{GS}** at the **highest possible value** and **W** at the **minimum value**
2. **increase L** until $V_{DSAT} = V_{DS0}$ (i_{DW} is now set)
3. **increase W** to obtain the desired current $I_D = i_{DW}W$

At this point the MOSFET is biased at an inversion as strong as possible, and the **white noise term is minimized**.

Next we **minimize the low frequency term** by increasing both L and W while maintaining a fixed W/L ratio, i.e. a fixed g_m and operating point (beware of the parasitics!).

The W and L are increased until:

$$\frac{K_{F2,4}}{c_{ox}W_{2,4}L_{2,4}}g_{m2,4} << \frac{K_{FP}}{c_{ox}WL}g_{m1}$$

Note that, in principle, the low frequency term can be made negligible while the white term cannot.

The higher the available V_{DS0} (compared with the threshold), the lower the white term.

Noise from cascode

Concerning the noise contribution from the **cascode** M_3, it is usually negligible.

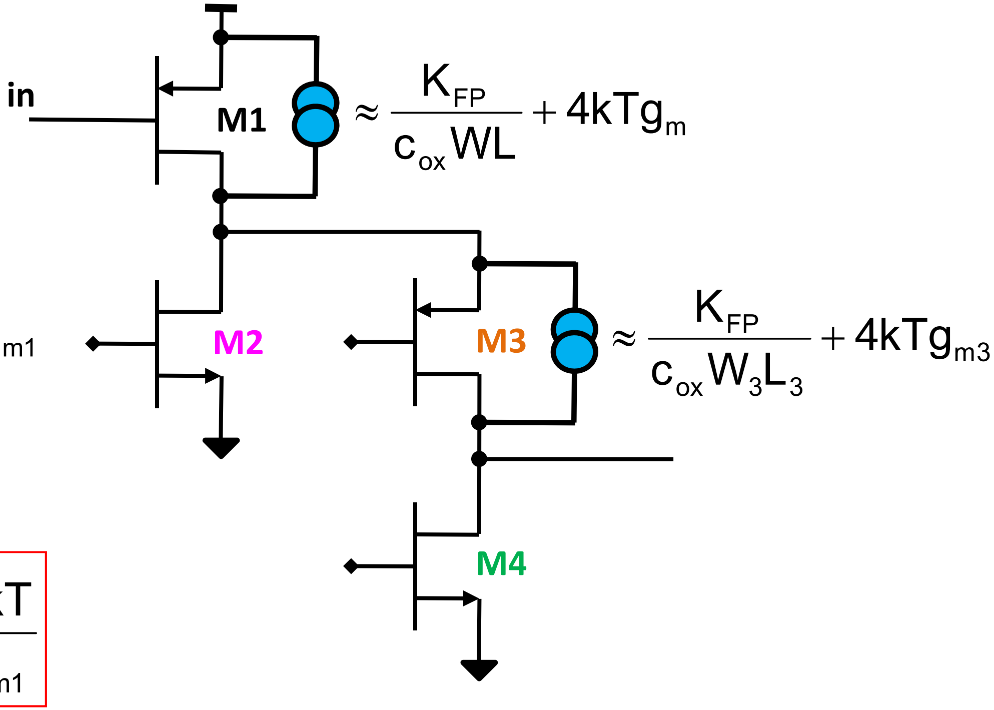
We can use Blakesley's theorem (assume thermal only):

$$4kTg_{m3}\left(\frac{1-\lambda}{\lambda}\right)^2 \approx 4kTg_{m3}\left(\frac{1}{g_{m3}r_{o1}}\right)^2 \Leftrightarrow \lambda^2 4kTg_{m1}$$

where: $\lambda = \dfrac{g_{m3}r_{o1}}{1+g_{m3}r_{o1}} \approx 1$

It follows:

$$\boxed{\frac{4kT}{g_{m3}}\left(\frac{1}{g_{m1}r_{o1}}\right)^2 \Leftrightarrow \frac{4kT}{g_{m1}}}$$



$$\approx \frac{K_{FP}}{c_{ox}WL} + 4kTg_m$$

$$\approx \frac{K_{FP}}{c_{ox}W_3L_3} + 4kTg_{m3}$$

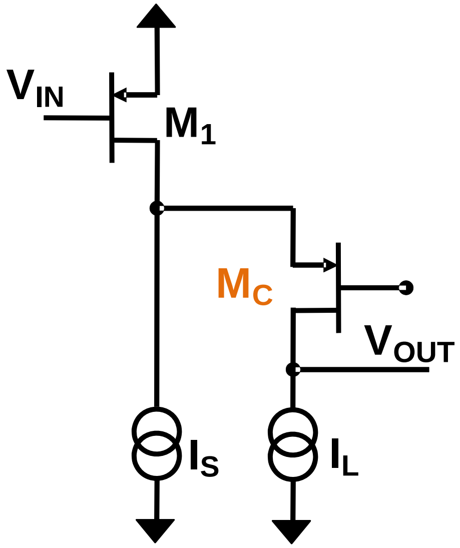
Note that r_{o1} must be **compared with** $\dfrac{C_i}{g_{m1}C_{gd}}$, whichever is the lowest: $\boxed{\dfrac{4kT}{g_{m3}}\left(\dfrac{C_{gd}}{C_i}\right)^2 \Leftrightarrow \dfrac{4kT}{g_{m1}}}$

We conclude that M_3 should have the **minimum L** (maximum g_m) while the width W is set as a trade-off between the dc gain and the secondary pole at the source of M_3.

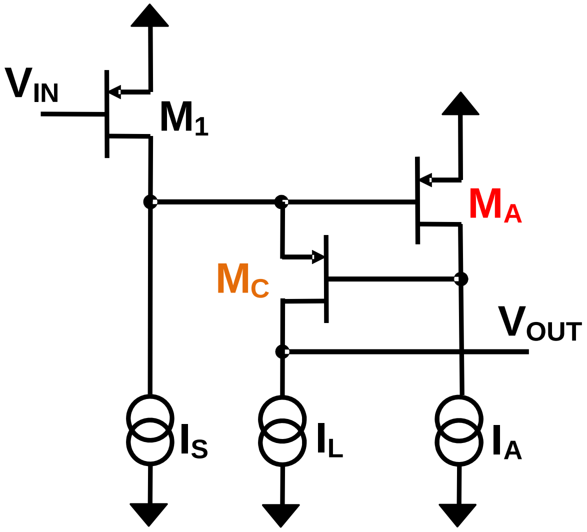
Note also that the **resistance at the drain** of M_1 is (cascode efficiency):

$$r \approx r_{o1}\frac{r_{o4}+r_{o3}}{r_{o4}+r_{o3}\left(1+g_{m3}r_{o1}\right)} \approx \begin{cases} \dfrac{r_{o1}}{1+g_{m3}r_{o1}} \approx \dfrac{1}{g_{m3}} & \text{small } r_{o4} \\ r_{o1} & \text{large } r_{o4} \quad \text{(more realistic)} \end{cases}$$

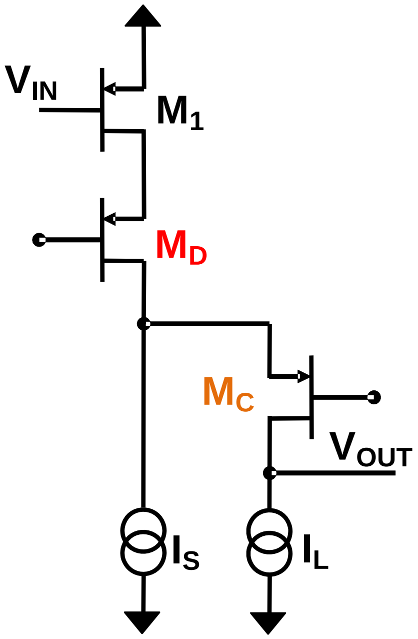
Cascoding for large input capacitance



single cascode



amplified cascode



dual cascode

Issues: dc gain, secondary pole, charge gain

+ cascode *impedance reduced by M_A gain*

+ adds third pole, but both poles are at high freq.

± additional voltage drop set by threshold of M_A

– *requires additional power (noise from M_A)*

– real poles if:

$$g_{mA} > 4g_{mC} \frac{C_{gd1} + C_{gsA}}{C_{gdA} + C_{gsC}}$$

G. De Geronimo et al., IEEE TNS 55 (2008)

+ cascode impedance *reduced through M_D*

+ add third pole, but both poles are at high freq.

± additional voltage drop controlled by bias circuit

+ *does not require additional power*

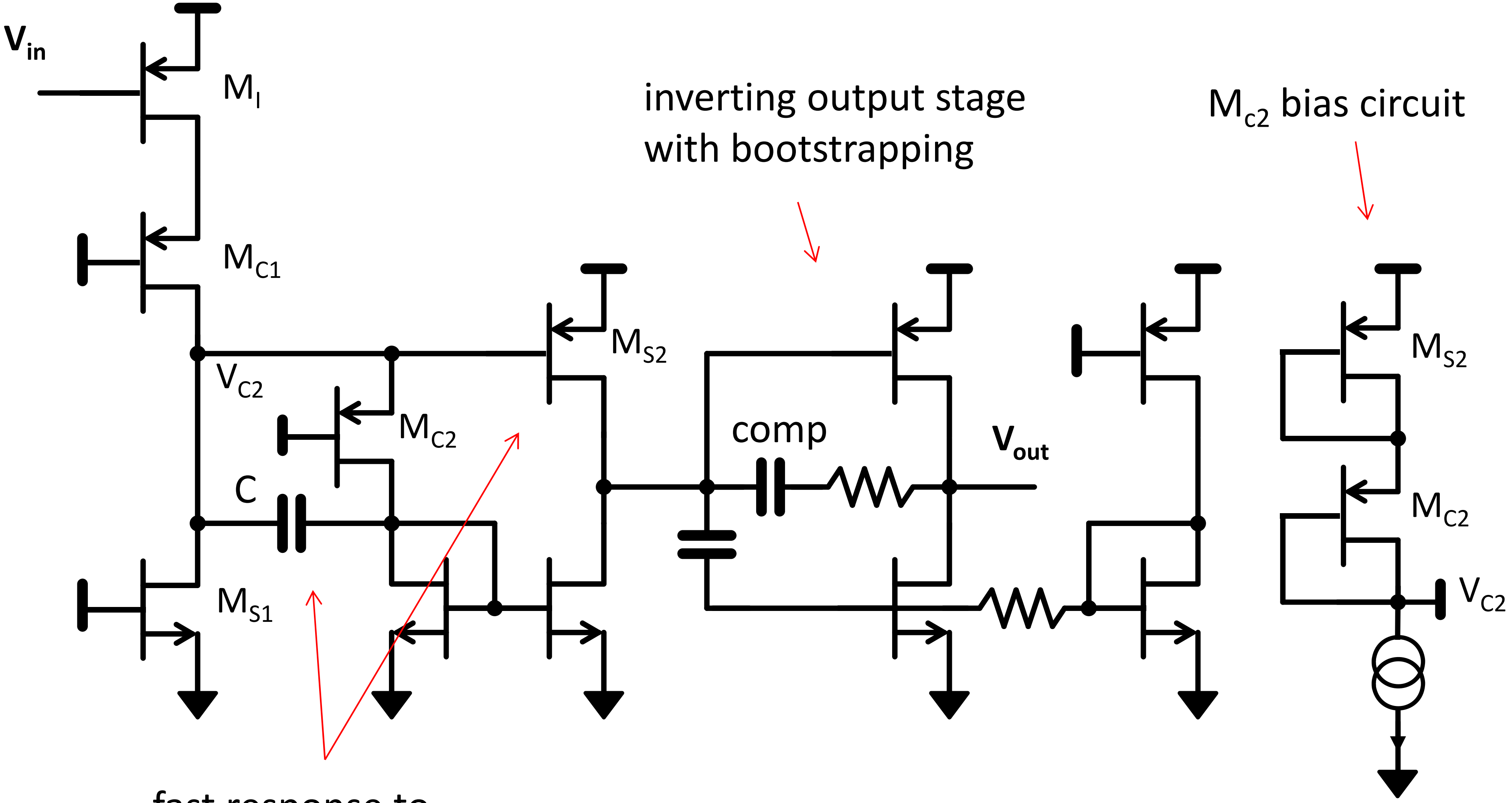
+ always real poles

+ *optimum size ≈ 1/3 to 1/4 of M_1*

– *slightly higher node resistance*

BROOKHAVEN
NATIONAL LABORATORY
Instrumentation Division

Rail-to-rail voltage amplifier



inverting output stage
with bootstrapping

M_{c2} bias circuit

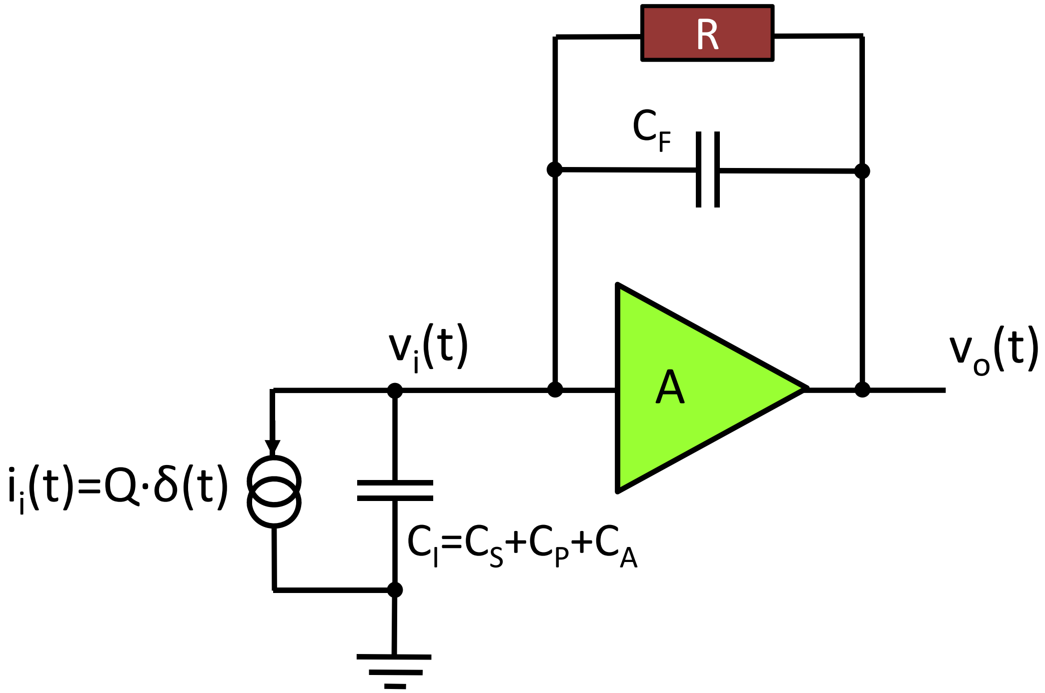
fast response to
positive charge

Part III

Charge amplification

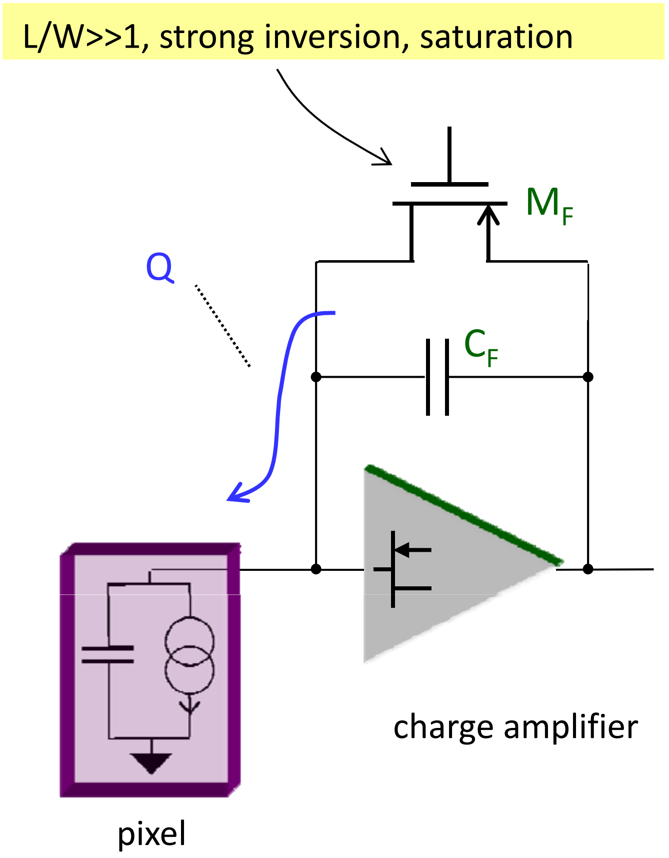
Charge amplifier

A charge amplifier, in its **classical definition**, is composed of a voltage amplifier A and a feedback capacitor C_F. The **capacitor integrates** the current $Q \cdot \delta(t)$ released by the sensor. It provides a "**virtual ground**" at the input node, thus stabilizing the potential of the sensor electrode along with providing **low-noise charge-to-voltage conversion**.



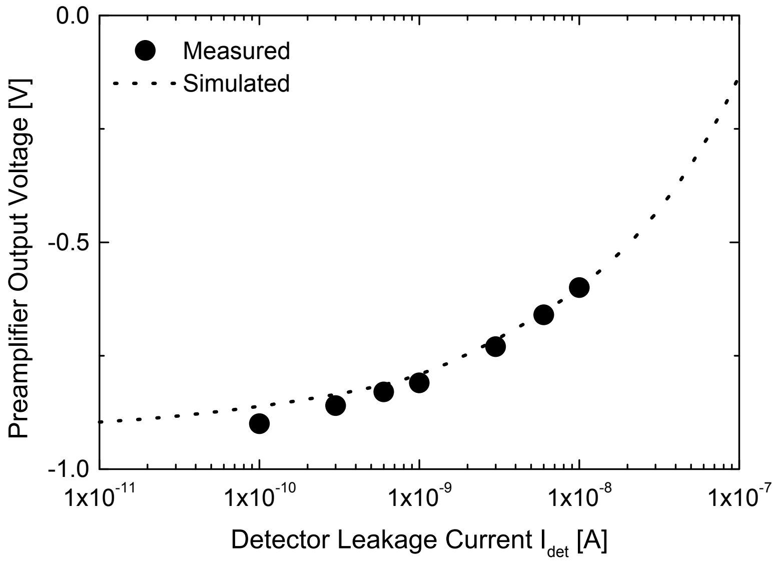
Charge amplifiers require an additional low-frequency **network R** (known as "**reset**") in feedback for (i) stabilization of the bias point and (ii) discharge (continuous or switched) of the feedback capacitor C_F. A **properly designed reset** has negligible effect on the signal processing (R is very large) and, in most cases, little effect on the resolution.

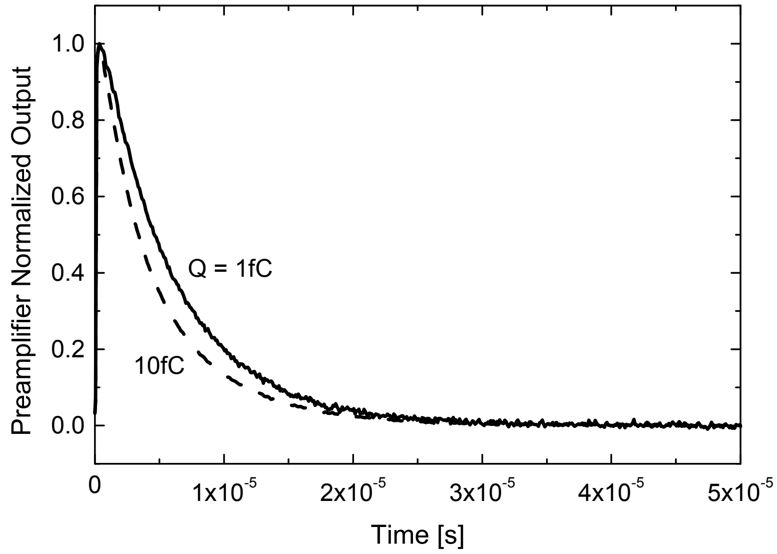
Adaptive continuous reset



shot noise: $2qI_{det}$

equivalent: $4kT/R = 2qI_{eq}$



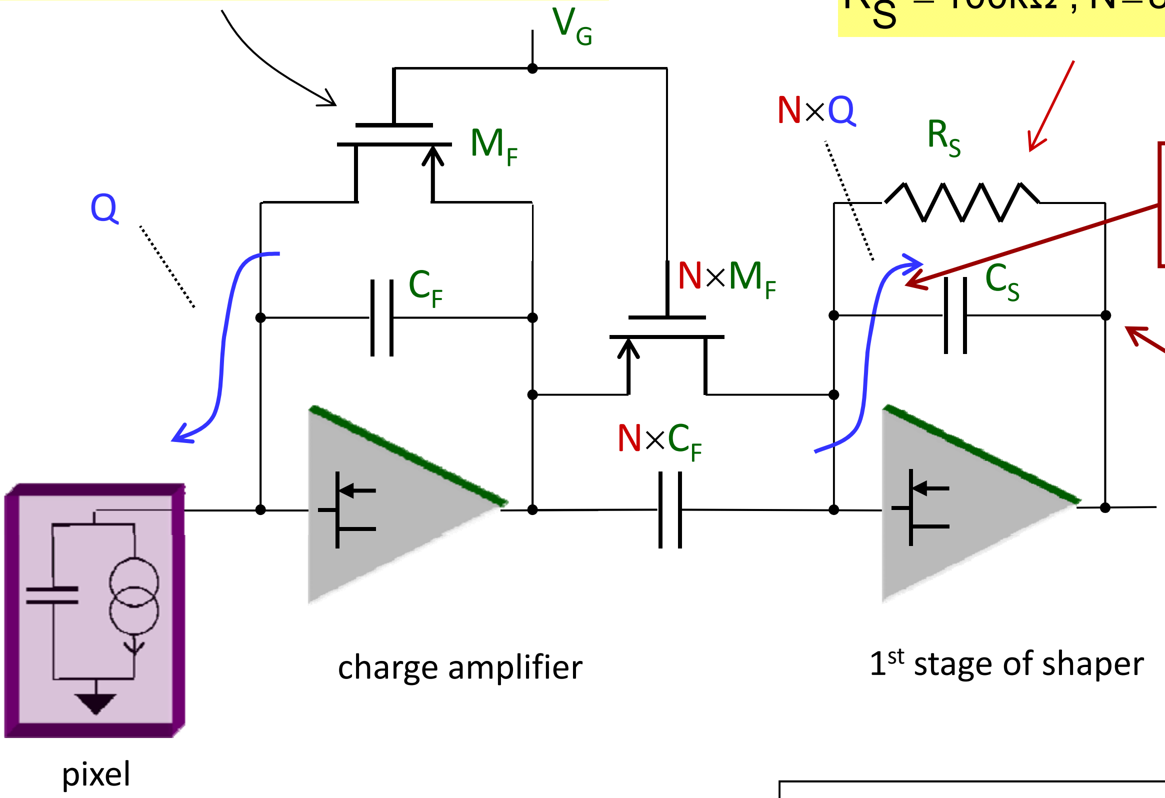


Compensated adaptive reset: charge amplification

L/W>>1, strong inversion, saturation

~ 100 e$^-$ at 100ns !

$R_S = 100k\Omega$, $N=8 \rightarrow I_{eq} \approx 8nA$



Linear charge amplification: charge gain is equal to N

Linear charge-to-voltage conversion is at this node

- Charge gain N
- High linearity (charge)
- Low noise
- Adaptive

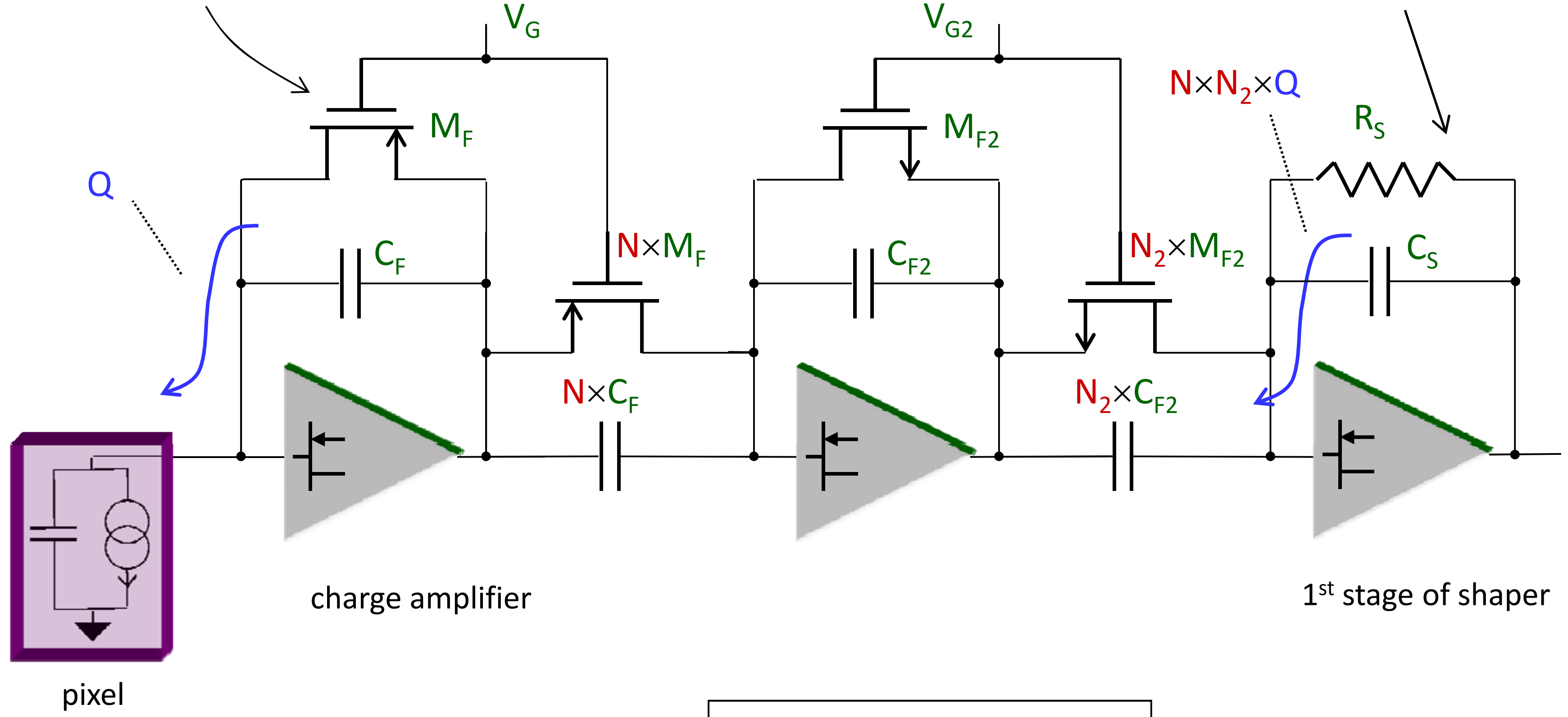
G. De Geronimo et al., IEEE TNS 47 (2000)

Doubling the reset

~ 11 e$^-$ at 100ns !

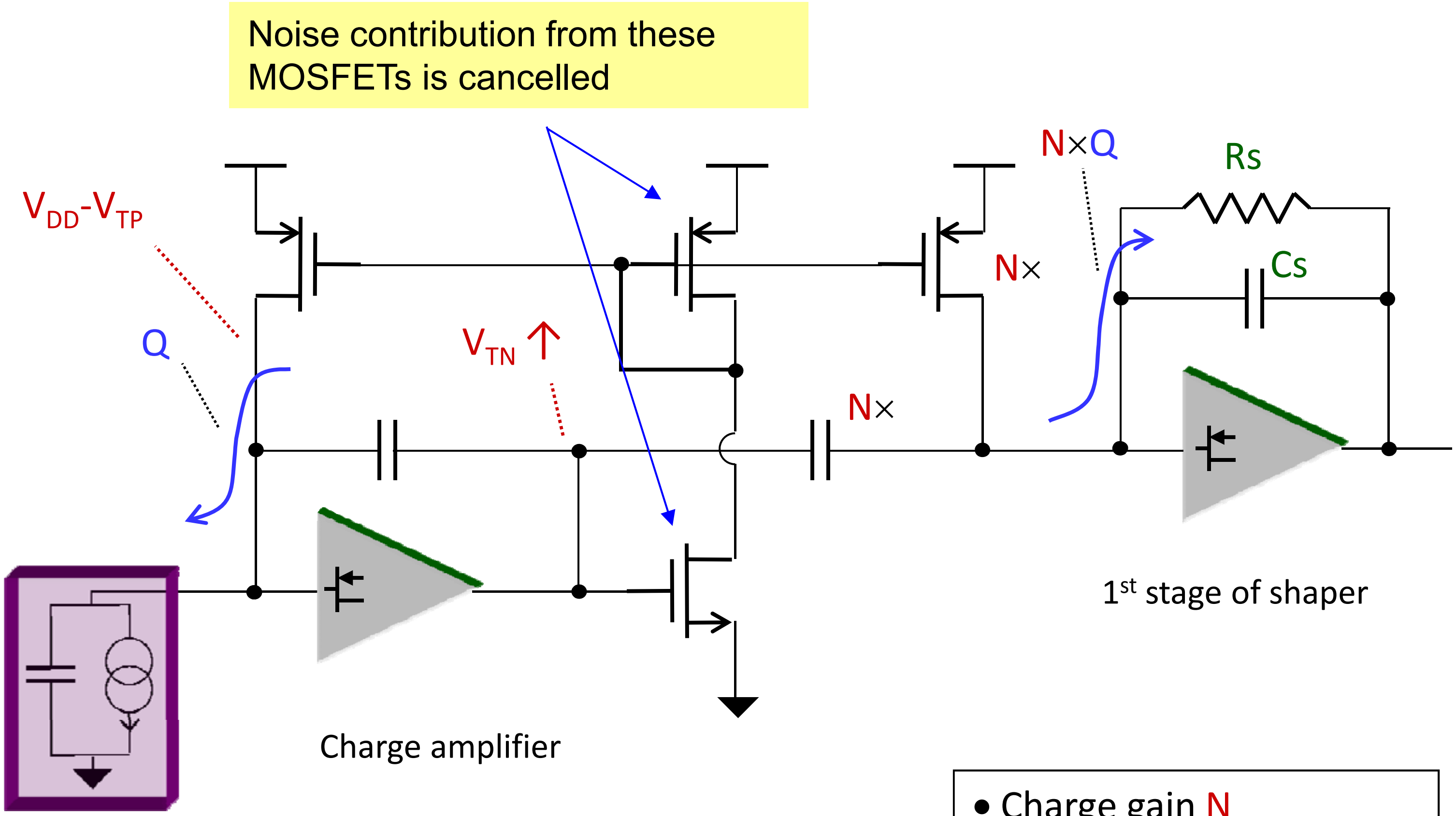
$R_S = 100k\Omega$, $N \times N_2 = 64$ → $I_{eq} \approx 100pA$

L/W>>1, strong inversion, saturation



- Charge gain NxN$_2$
- High linearity (charge)
- Low noise
- Adaptive
- Limited swing

Low-voltage configuration



- Charge gain N
- High linearity (charge)
- Low noise
- Adaptive
- High swing

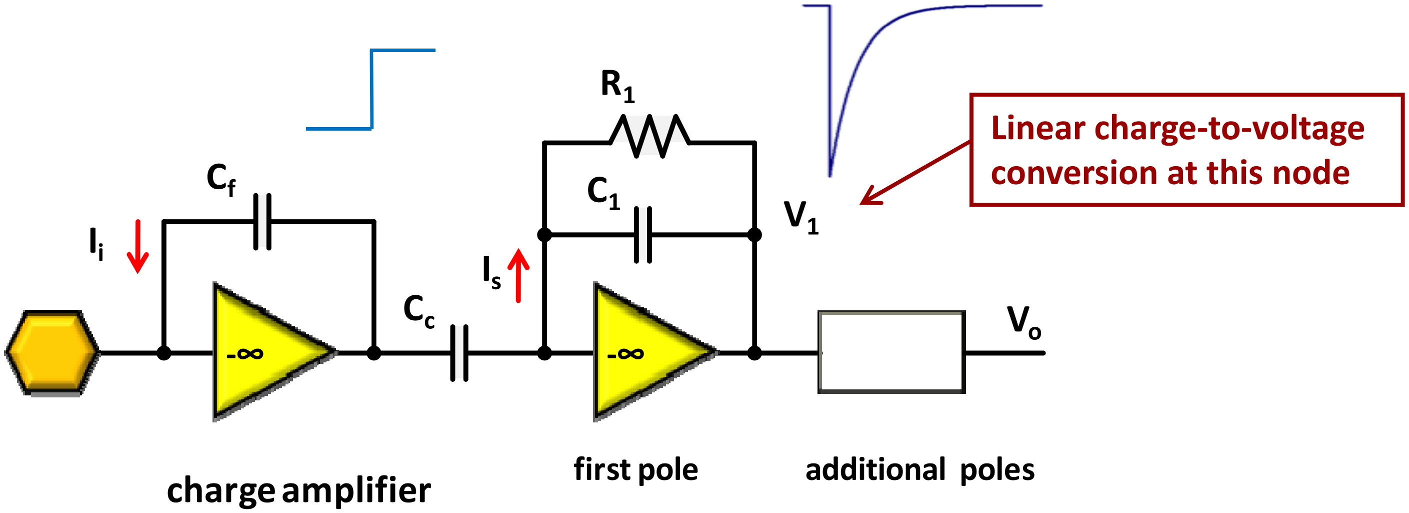
G. De Geronimo et al., IEEE TNS 54 (2007)

Part IV

Filter design

(analog dynamic range)

Maximum charge



Linear charge-to-voltage conversion at this node

$$I_s = \frac{C_c}{C_f} I_i = A_c I_i$$

shaper

$$V_1 = Z_1 I_s = \frac{A_c R_1}{1 + sR_1C_1} I_i$$

The **voltage linearity at the output node of the charge amplifier is not required**, as long as the desired linear charge amplification **$A_c = C_c / C_f$** is achieved at v_o.

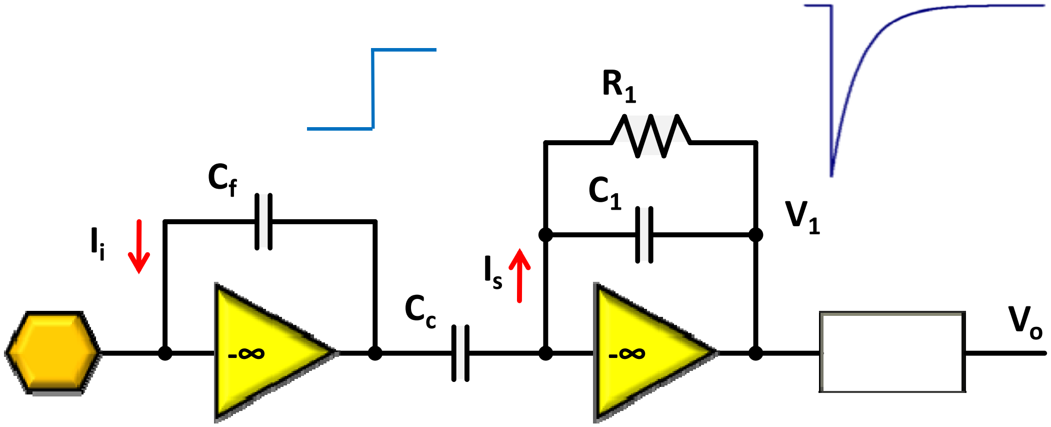
However, the **outputs of the filtering stages (e.g. v_1) must be linear**.

The **maximum charge** can be calculated as

$$Q_{max} = \frac{C_1 V_{1max}}{A_c}$$

Given a maximum charge Q_{max} we must select A_c , C_1 and V_{1max}

Shaper noise



The contribution to the ENC from the noisy (i.e. dissipative) element R_1 can be calculated from the **equivalent parallel noise at the input**

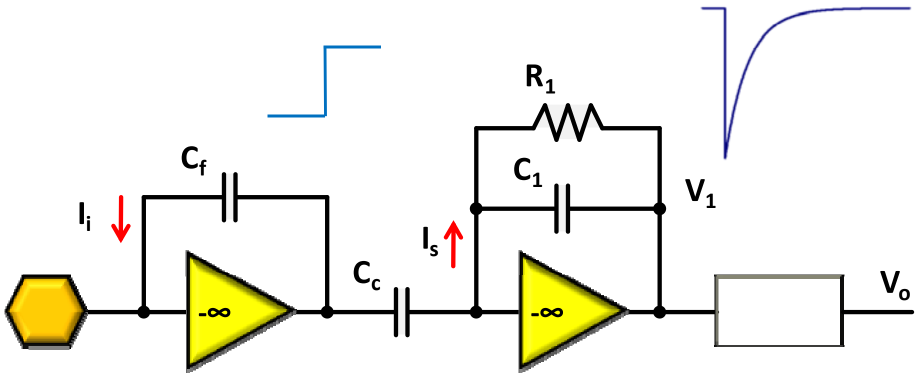
$$ENC_{s1}^2 = \frac{A_{iwp}}{A_c^2}\frac{4kT}{R_1}\tau_p = \frac{A_{iwp}}{A_c^2}\frac{4kT}{R_1}\eta_p R_1 C_1 = A_{iwp}\eta_p 4kT\frac{C_1}{A_c^2}$$

The ratio $\boldsymbol{\eta_p = \tau_p / R_1 C_1}$ is defined by the type of filter (shaper).

Typically $A_{iwp}\eta_p$ ~ 1 to 2.

Next stages can be included in the analysis, see G. De Geronimo et al., IEEE TNS 58 (2011)

Analog dynamic range



The **analog dynamic range** is defined as

$$DR = \frac{Q_{max}}{\sqrt{ENC_{ca}^2 + ENC_{s1}^2}}$$

The design proceeds with the following steps:

1) Maximize the front-end resolution (minimize ENC_{ca})

2) Optimize dynamic range (**minimize ENC_{s1}**)

$$DR = \frac{Q_{max}}{\sqrt{ENC_{ca}^2 + ENC_{s1}^2}} = \frac{Q_{max}}{\sqrt{\rho \cdot ENC_{s1}^2}} \qquad \rho = \frac{ENC_{ca}^2 + ENC_{s1}^2}{ENC_{s1}^2}$$

we must set the value of the coefficient **ρ (>1)**

Optimizing the dynamic range

$$DR \approx \frac{Q_{max}}{\sqrt{\rho \cdot ENC_{s1}^2}} = \frac{\dfrac{C_1 V_{1max}}{A_c}}{\sqrt{\rho \cdot \dfrac{A_{iwp}\eta_p}{A_c^2} 4kTC_1}} = \frac{V_{1max}\sqrt{C_1}}{\sqrt{4kT\rho \cdot A_{iwp}\eta_p}}$$

- DR is proportional to V_{1max} → use **rail-to-rail**
- DR does not depend on the peaking time τ_p or resistor R_1
- **ρ** is a **key design parameter** (which defines C_1 and A_c for a given Q_{max}) $\quad Q_{max} = \dfrac{C_1 V_{1max}}{A_c}$

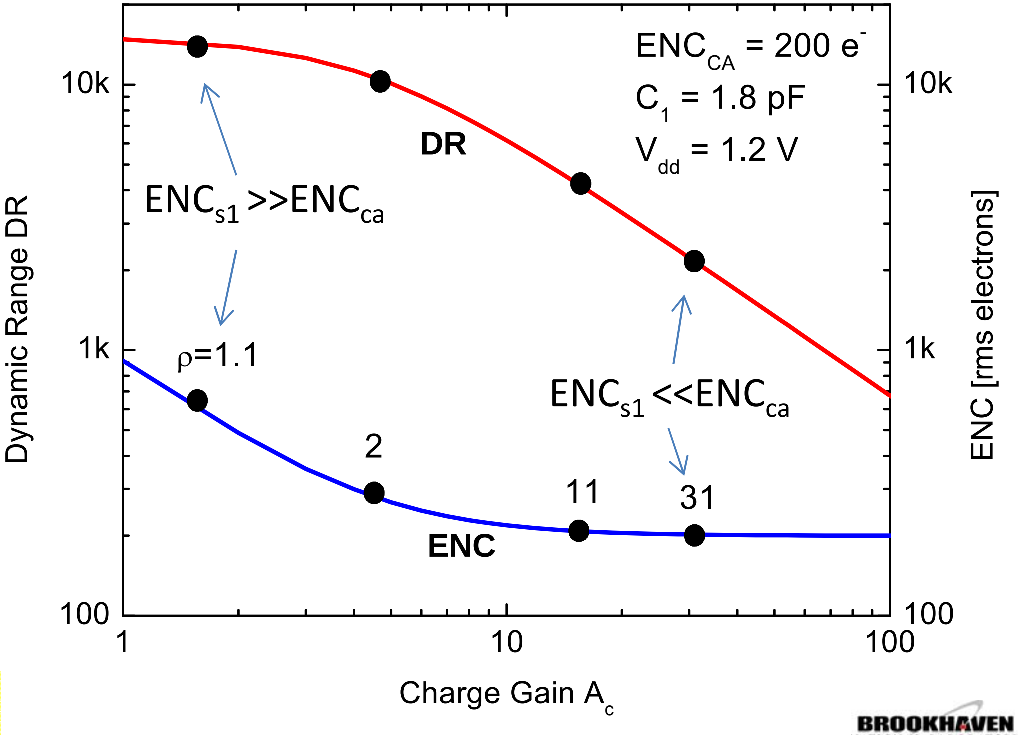
The dynamic range can be increased (i.e. A_c decreased) at the expense of the ENC.

Values of ρ **lower than 1.1** (ENC dominated by ENC_{s1}) would **not benefit much the DR but would further limit the resolution** by increasing the total ENC (ρ cannot be lower than 1).

Values of ρ **higher than 30** (ENC dominated by ENC_{ca}) would **not benefit much the ENC but would further limit the DR.**

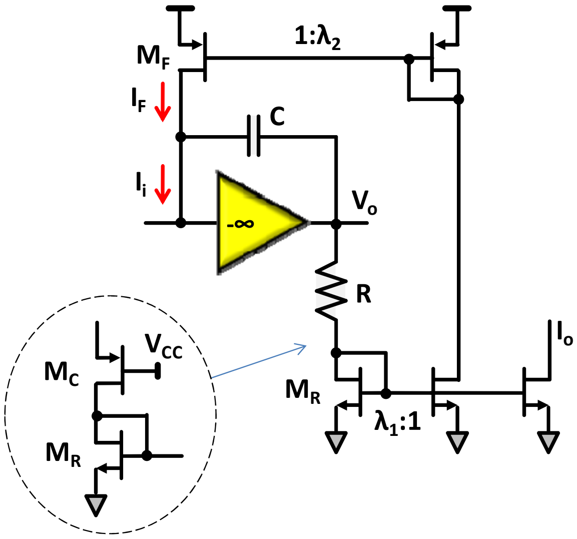
A convenient choice is **ρ ≈ 5.76** where **ENC$_{s1}$ contributes to the rms at 10%** (ENC_{s1}/ENC_{ca} ≈0.46).

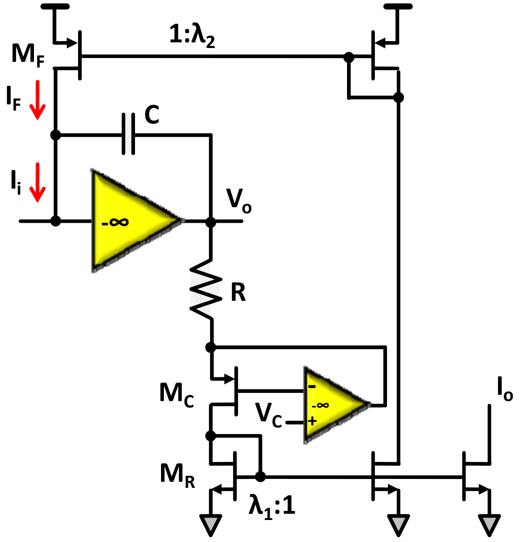
Example: CMOS 130nm (1.2V), 1.8pF, → DR ~ 6,000

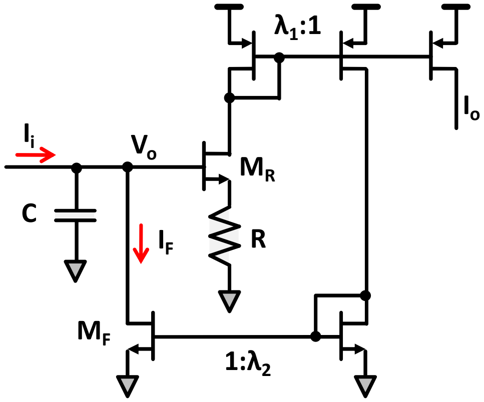


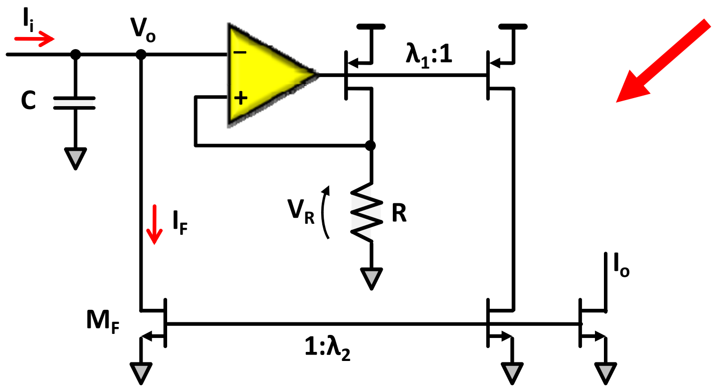
Can we decrease $\eta_p = \tau_p / R_1C_1$?

Scaling R with mirrors









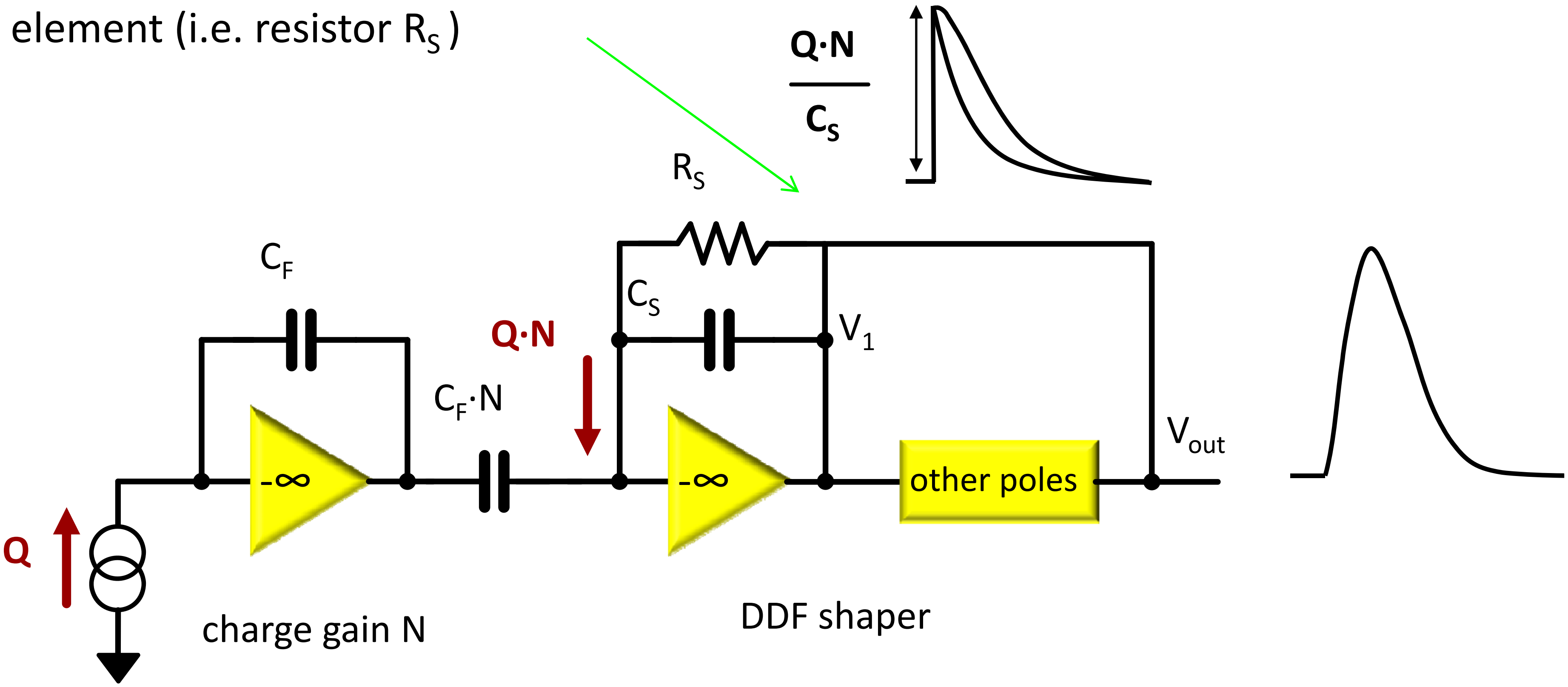
Noise contribution from R: $S_{nR} = \frac{4kT}{\lambda R_{eq}}$

Noise **contribution from M_F** (imposing linearity with $g_{mR}R$ >> 1): $S_{nMF} = 2qI_F = \frac{2qI_R}{\lambda} >> \frac{2qnV_T}{R\lambda} = \frac{2kT}{R_{eq}}$

Signal through active components affect **linearity** and introduce **non-stationary noise** $S/N \sim \surd(N_e A_C)$

Delayed Dissipative Feedback (DDF)

delay feedback of dissipative element (i.e. resistor R_S)



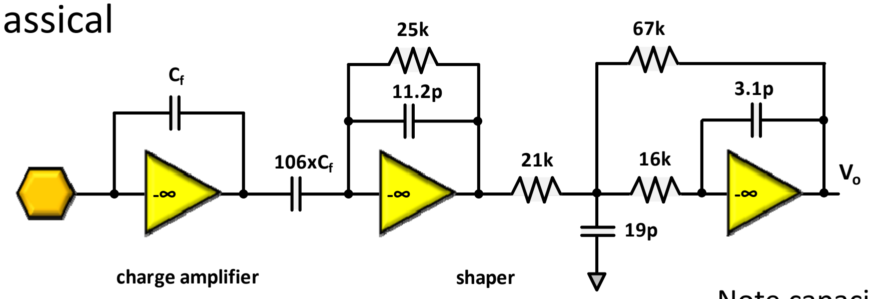
higher analog dynamic range

$$DR_a = \frac{Q_{max}}{\sqrt{ENC_{CA}^2 + ENC_S^2}}$$

Applies also to the other stages of the shaper
see G. De Geronimo and S. Li, TNS 58, Oct. 2011

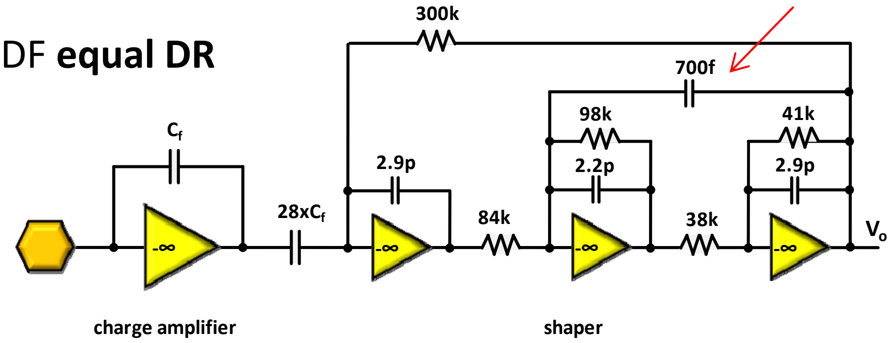
Delayed Dissipative Feedback (DDF)

Classical

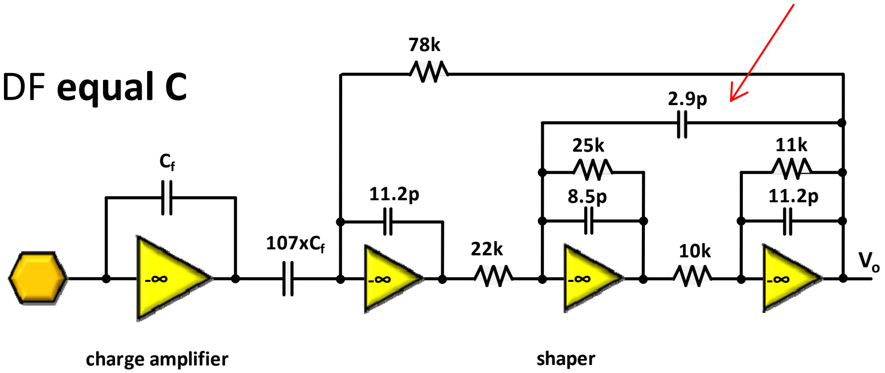


Note capacitance in positive feedback

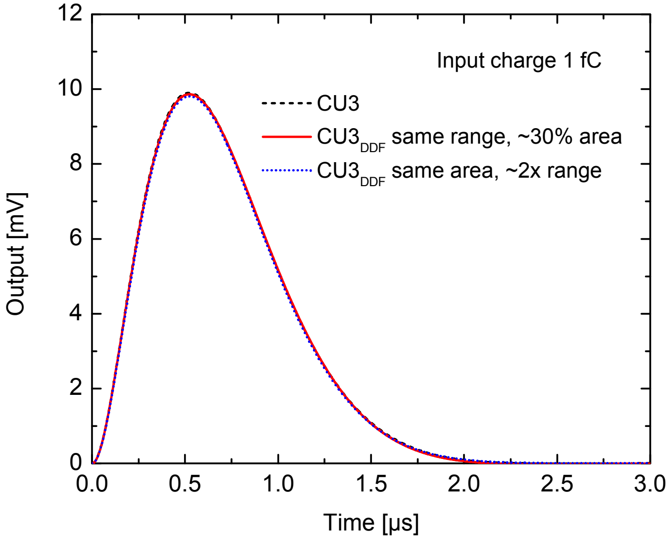
DDF **equal DR**

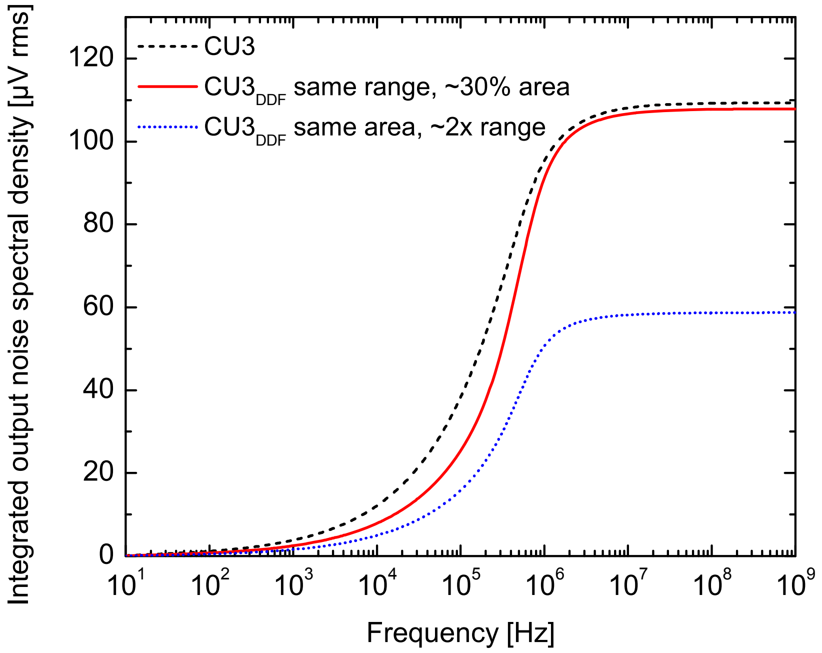


DDF **equal C**



Example: CMOS 130nm (1.2V), 1.8pF, → DR ~ 12,000





G. De Geronimo et al., IEEE TNS 58 (2011)

Higher dynamic range in limited area requires **multi-range** or **charge subtraction** techniques

Summary

I. Input MOSFET optimization

- model and operation in moderate inversion
- low-frequency noise and 1/f equivalent
- resolution vs technology

II. Amplifier design

- rules of thumb for gain and bandwidth
- dealing with secondary noise sources
- advanced cascoding

III. Charge amplification

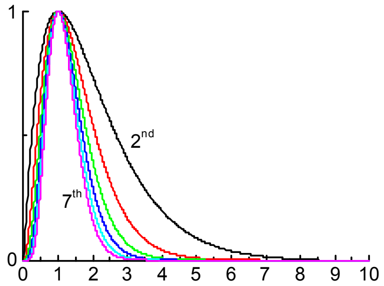
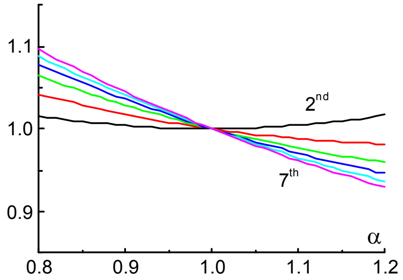
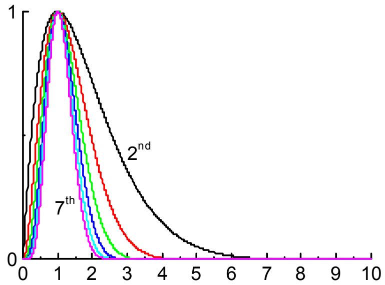
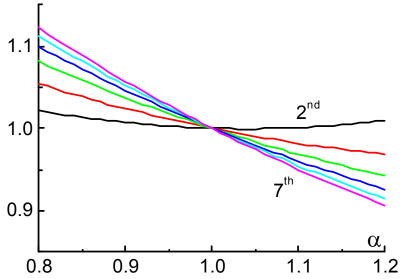
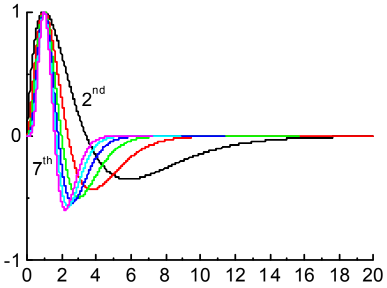
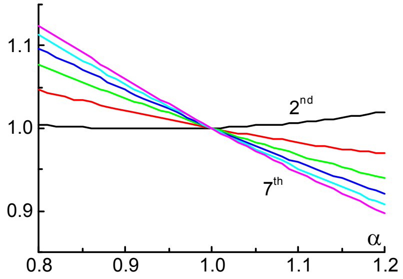
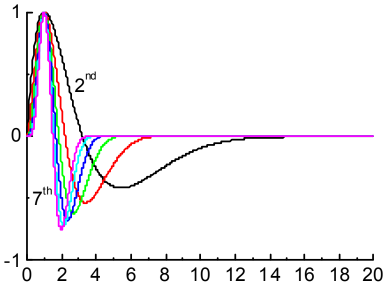
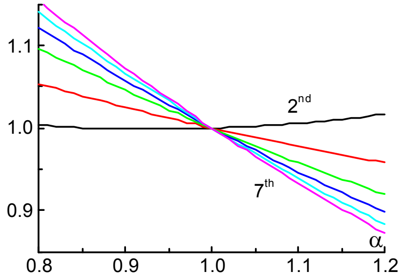
- charge gain and adaptive continuous reset

IV. Filter design

- impact on analog dynamic range
- optimization : area and voltage
- delayed dissipative feedback (DDF)

Backup slides

ENC (τ_p) coefficients for most common filters

Filter	Shape	a_w	$a_f(1)$	a_p	$\rho_f(\alpha_f)=a_f(\alpha_f)/a_f(1)$	τ_w/τ_p
RU-2		0.92	0.59	0.92		7.49
RU-3		0.82	0.54	0.66		5.04
RU-4		0.85	0.53	0.57		4.17
RU-5		0.89	0.52	0.52		3.72
RU-6		0.92	0.52	0.48		3.46
RU-7		0.94	0.51	0.46		3.28
CU-2		0.93	0.59	0.88		6.17
CU-3		0.85	0.54	0.61		3.92
CU-4		0.91	0.53	0.51		3.16
CU-5		0.96	0.52	0.46		2.84
CU-6		1.01	0.52	0.42		2.66
CU-7		1.04	0.52	0.40		2.55
RB-2		1.03	0.75	1.01		16.6
RB-3		1.11	0.78	0.76		9.87
RB-4		1.30	0.81	0.66		7.67
RB-5		1.47	0.85	0.62		6.61
RB-6		1.61	0.87	0.59		5.96
RB-7		1.74	0.90	0.57		5.53
CB-2		1.08	0.80	1.02		12.9
CB-3		1.27	0.86	0.76		7.29
CB-4		1.58	0.93	0.67		5.58
CB-5		1.87	0.98	0.62		4.80
CB-6		2.10	1.03	0.60		4.39
CB-7		2.33	1.06	0.57		4.10

R = real coincident poles, C = complex-conjugate poles, U = unipolar, B = bipolar

G. De Geronimo et al., IEEE TNS 52 (2005)

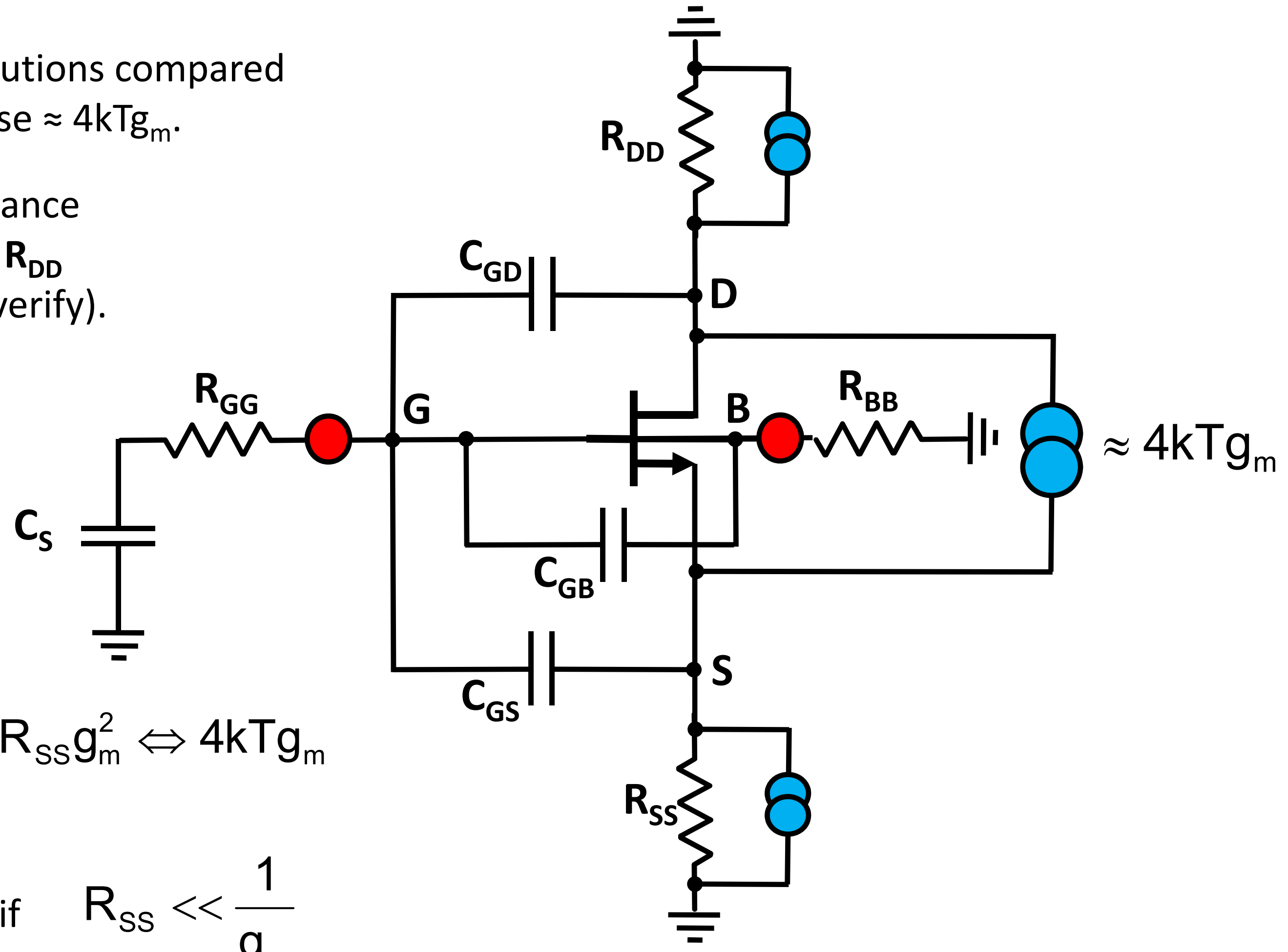
FET parasitic resistances

Consider the four parasitic resistors R_{GG}, R_{BB}, R_{SS}, and R_{DD}, each contributing with thermal noise. These parasitic resistors come mainly from the physical layout. Assume that the resistor values are low enough to have no impact on the signal response. Assume for simplicity a low impedance at the drain.

Let's analyze the noise contributions compared to the FET channel's white noise ≈ $4kTg_m$.

Assuming a large output resistance the relative contribution from $\mathbf{R_{DD}}$ is negligible (use Blakesley to verify).

The relative contribution from $\mathbf{R_{SS}}$ in the frequency range within f_T can be approximated with:



$$\frac{4kT}{R_{SS}}\left(\frac{R_{SS}g_m}{1+R_{SS}g_m}\right)^2 \approx 4kTR_{SS}g_m^2 \Leftrightarrow 4kTg_m$$

and it can be made negligible if $R_{SS} << \dfrac{1}{g_m}$

which can be obtained in the layout by **increasing the Source diffusion size, its number of contacts, and the width of the interconnection**.

FET parasitic resistances

The gate and bulk parasitic resistors require more attention.

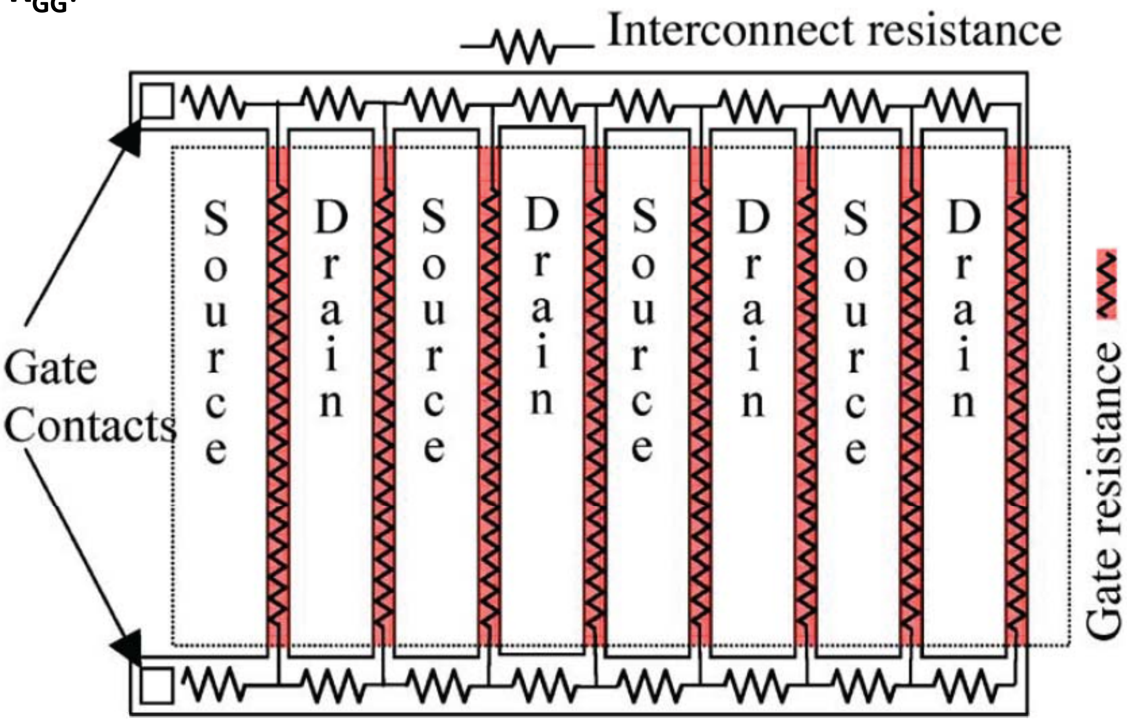
Let's consider first the contribution from $\mathbf{R_{GG}}$.

With arguments similar to the base spreading resistance in the BJT, the relative contribution from R_{GG} can be approximated as:

$$4kTR_{GG}g_m^2\gamma_{GG}^2\lambda \Leftrightarrow 4kTg_m$$

$$\gamma_{GG} = \frac{C_S}{C_S + C_G}$$

$$\lambda \approx \begin{cases} 1/3 & \text{single side gate contact} \\ 1/12 & \text{dual side gate contact} \end{cases}$$



The R_{GG} contribution can be made negligible if $R_{GG} << \dfrac{1}{g_m\gamma_{GG}^2\lambda}$

which can be obtained in the layout by **increasing the number of fingers**.
Note: the interconnect resistance can be made negligible with contacts and metals.

Ref: R. P. Jindal, "Compact noise model s for MOSFETs", IEEE TED 53, pp. 2051-2061, 2006

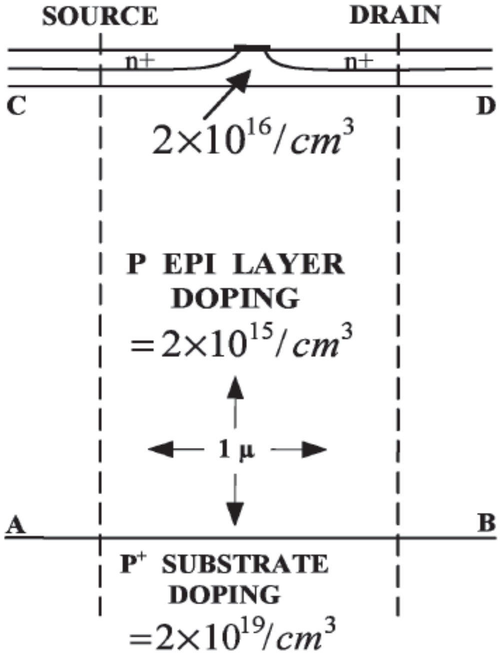
FET parasitic resistances

Finally let's consider the contribution from **R_{BB}**.

This contribution can be non negligible, especially in technologies that use epitaxial layer, which is characterized by a resistivity higher than the substrate. The relative contribution from R_{BB} can be approximated as:

$$4kTR_{BB}g_{mB}^2 \Leftrightarrow 4kTg_m$$

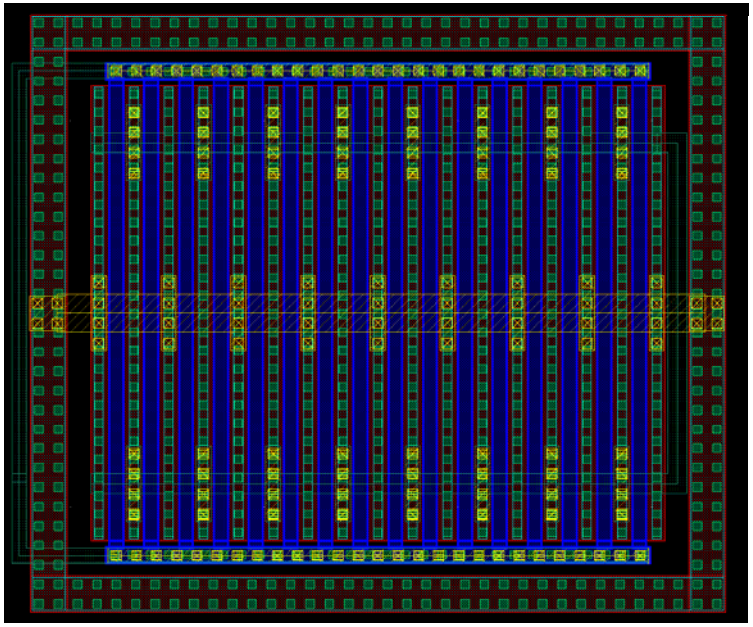
and it can be made negligible if $R_{BB} << \frac{g_m}{g_{mB}^2} = \frac{1}{g_m(n-1)^2}$



which can be partially obtained by minimizing the distance between the channel and the guard ring diffusion:

- making the layout more rectangular than square by **increasing the number of fingers**;
- **extending the guard ring diffusion** as close as possible to the channel

Note: the limit in the number of fingers is set by the real estate and parasitic capacitances at the gate and drain



Example of layout