Designing of the front end electronics for silicon strip detectors in submicron technologies

Jan Kaplon

CERN PH/ESE

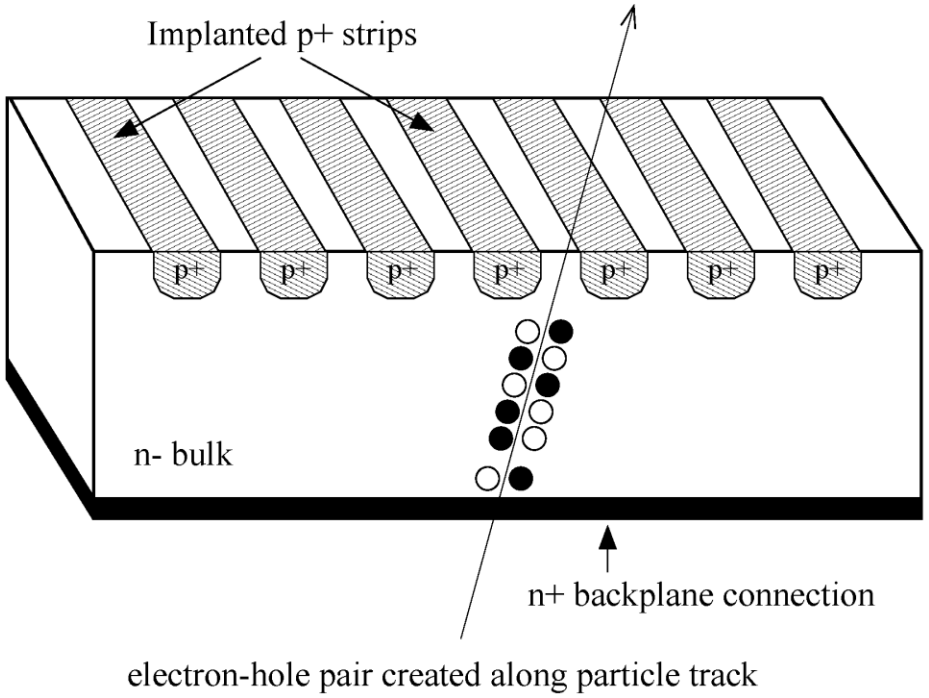
jan.kaplon@cern.ch

Outline

- Collection of signals from silicon detectors
  - Basic configuration of the preamplifier
  - Charge collection and cross talk signals → specifications for open loop gain and bandwidth of preamplifier stage
- General requirements for silicon strip detectors electronics
- Technology scaling and its consequences
- Improving the open loop gain of amplifier stages
  - Motivations
  - Methods
- Operating in weak inversion
  - Motivations
  - Costs
- Front end designed for SLHC silicon tracker (5 to 10pF detector capacitance) in 130 and 90nm process
  - Architecture
  - Performance
- Matching (provisional – low statistics from MPW runs)

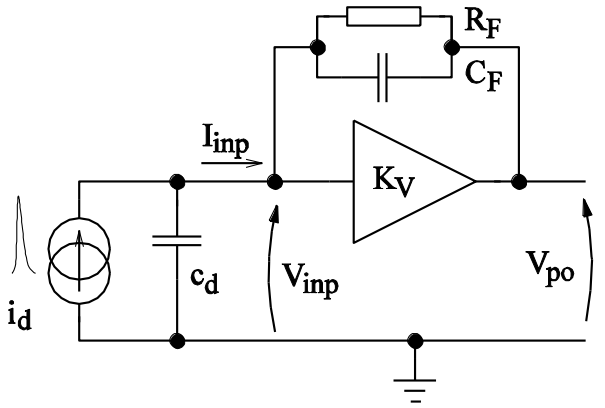
Principles of the silicon detectors for tracking applications

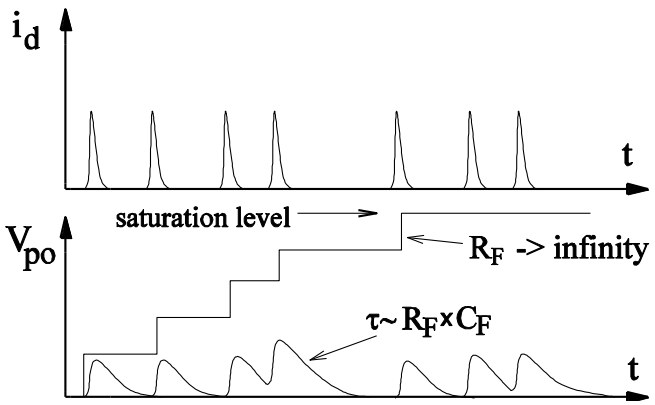
- p-n junction reverse biased forms the detection zone
- Ionization along the track of the high-energy particle
- For 300µm Si detector the most probable signal is around 3.5fC (non-irradiated detector)
- Electric field proportional to bias provides drifting of the created charge – induced current is readout by the front-end electronics
- Spatial resolution provided by the segmentation of the detector



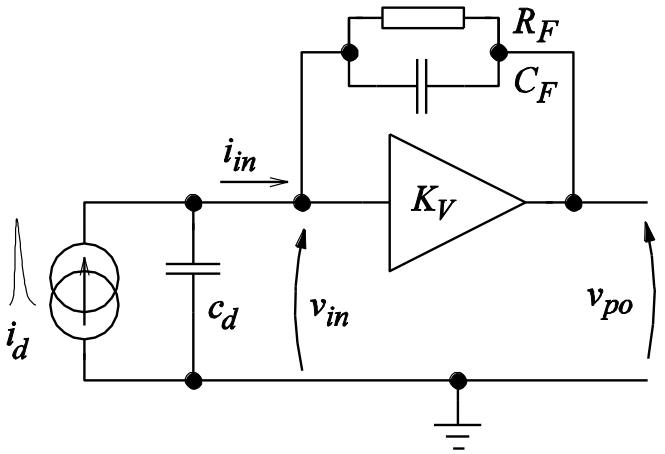
Reception of signals from silicon detector – basic configuration of the preamplifier

- Charge sensitive preamplifier
  - Delta-Dirac current pulses integrated on feedback capacitance
  - Discharge provided by the feedback resistor (prevents saturation)
- Mode of the preamplifier is defined by feedback time constant $\tau_F = R_F C_F$
  - τ_F comparable with the time constant of the shaper → transimpedance preamplifier
  - τ_F >> time constant of the shaper → charge preamplifier





Input impedance model (operator and frequency domain)



The input impedance in operator domain:

$$Z_{in}(s) = \frac{Z_F(s)}{1 + K_V(s)} \approx \frac{Z_F(s)}{K_V(s)}$$

Considering dominant pole only the open loop gain is:

$$K_V(s) = \frac{K_V}{1 + s \cdot \tau_{P0}}$$

Transimpedance mode:

$$Z_{in}(s) = \frac{R_F \cdot (1 + s \cdot \tau_{P0})}{K_V \cdot (1 + s \cdot \tau_f)}$$

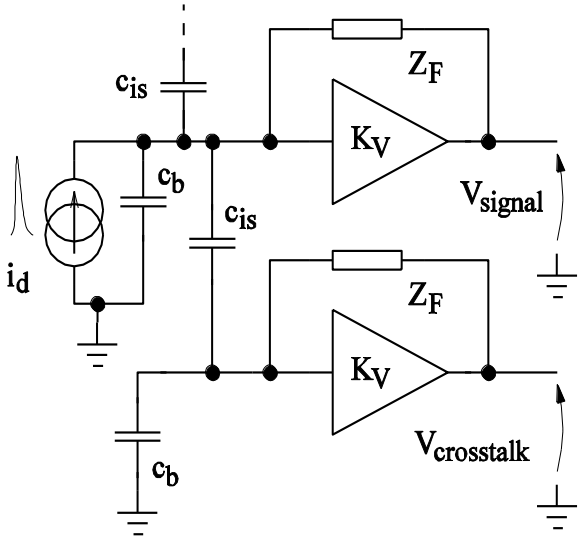
$$|Z_{in}| = \frac{R_F}{K_V} \cdot \frac{\sqrt{1 + \omega^2 \cdot \tau_{P0}^2}}{\sqrt{1 + \omega^2 \cdot \tau_f^2}}$$

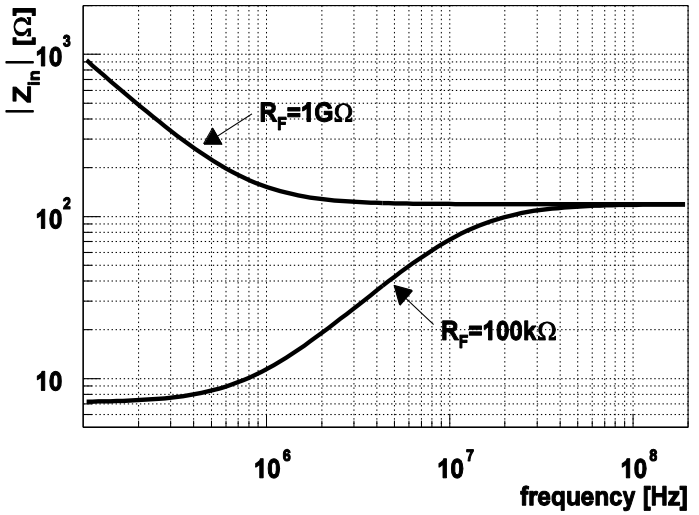
Charge preamplifier ($R_F \rightarrow \infty$):

$$Z_{in}(s) = \frac{1 + s \cdot \tau_{P0}}{K_V \cdot s \cdot C_F}$$

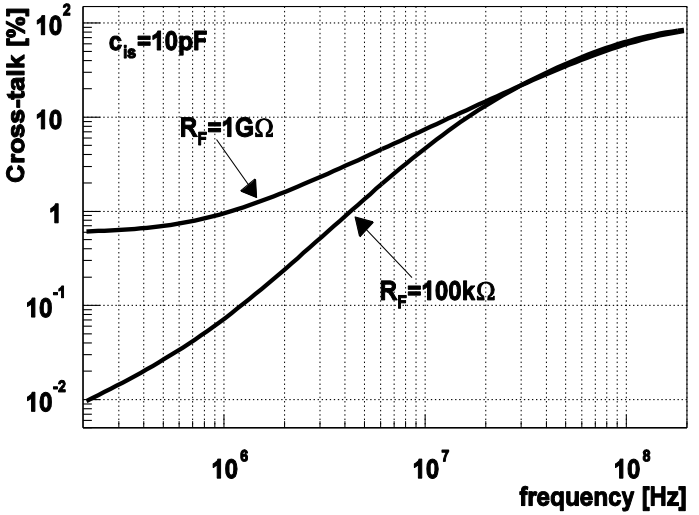
$$|Z_{in}| = \frac{\sqrt{1 + \omega^2 \cdot \tau_{P0}^2}}{\omega \cdot K_V \cdot C_F}$$

Input impedance and cross-talk signals for charge and transimpedance amplifiers





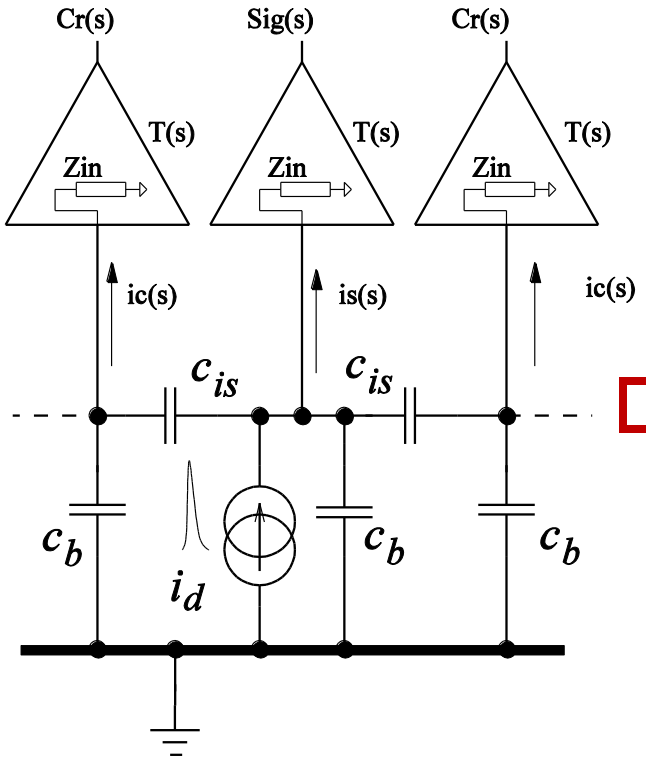
$$Cross\,Talk(s)=\frac{Z_{IN}(s)}{Z_{IN}(s)+Z_{IS}(s)}$$

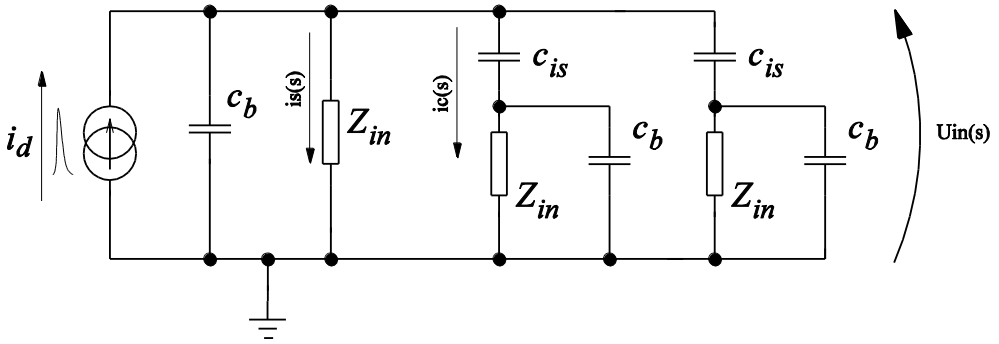


Input impedance and cross-talk for amplifier with 83dB gain and 1GHz Gain Bandwidth Product (GBP) working in charge and transimpedance configuration

But what is the cross talk in time domain?

Currents at front end inputs





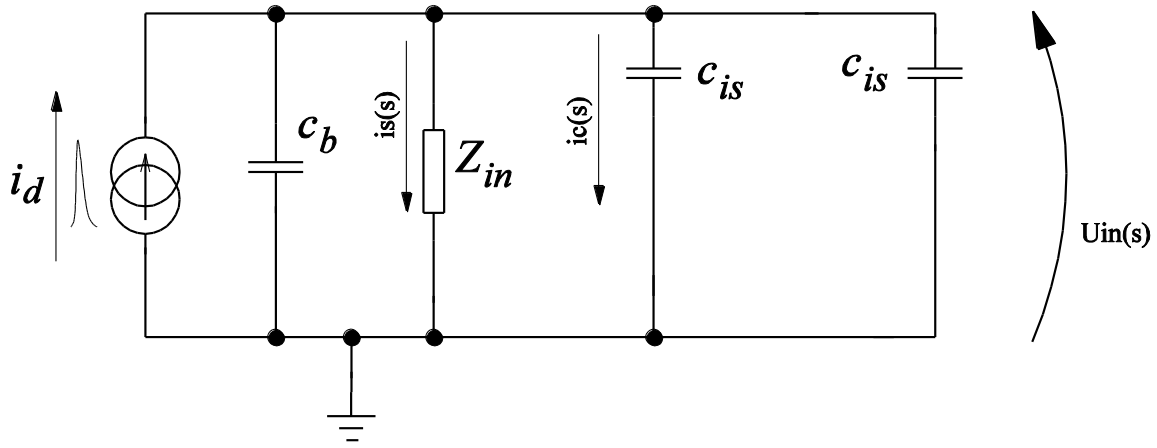
Using Kirchhoff law one can write:

$$i_d = u_{in} \cdot \left(s \cdot c_b + \frac{1}{Z_{in}} + \frac{2}{\frac{1}{s \cdot c_{is}} + \frac{Z_{in}}{s \cdot c_b \cdot Z_{in} + 1}} \right)$$

Unfortunately the use of this expression gives problems later during calculation of inverse Laplace functions. We have to simplify the model.

Currents at front end inputs

A reasonable trade off between accuracy and simplicity is shown below:



In this case we assume that input of the preamplifier is loaded with c_b and two c_{is} capacitances (neglecting input impedances of the neighbors). Using Kirchhoff law one can write:

$$i_d = u_{in} \cdot \left(s \cdot (c_b + 2 \cdot c_{is}) + \frac{1}{Z_{in}} \right)$$

Since we assume delta Dirac input we can write expression for voltage at the preamplifier input:

$$u_{in} = \frac{Z_{in}}{1 + s \cdot (c_b + 2 \cdot c_{is}) \cdot Z_{in}}$$

Currents at front end inputs

Expressions for current flowing into the input of readout channel:

$$i_s = u_{in} \cdot \frac{1}{Z_{in}}$$

For the expression of current flowing into neighboring channel we use simplified expression for u_{in} and expression for input impedance of neighboring channel connected in series with c_{is} capacitance (neglecting c_b):

$$i_c = u_{in} \cdot \frac{1}{Z_{in} + \frac{1}{s \cdot c_{is}}}$$

Front End transfer function and responses to signal and crosstalk

For calculation we will consider CR-RC^2 type of the shaper. The transfer function in operator domain is following;

$$T_{FE} = \frac{\tau_f}{(1+s\cdot\tau_f)^3}$$

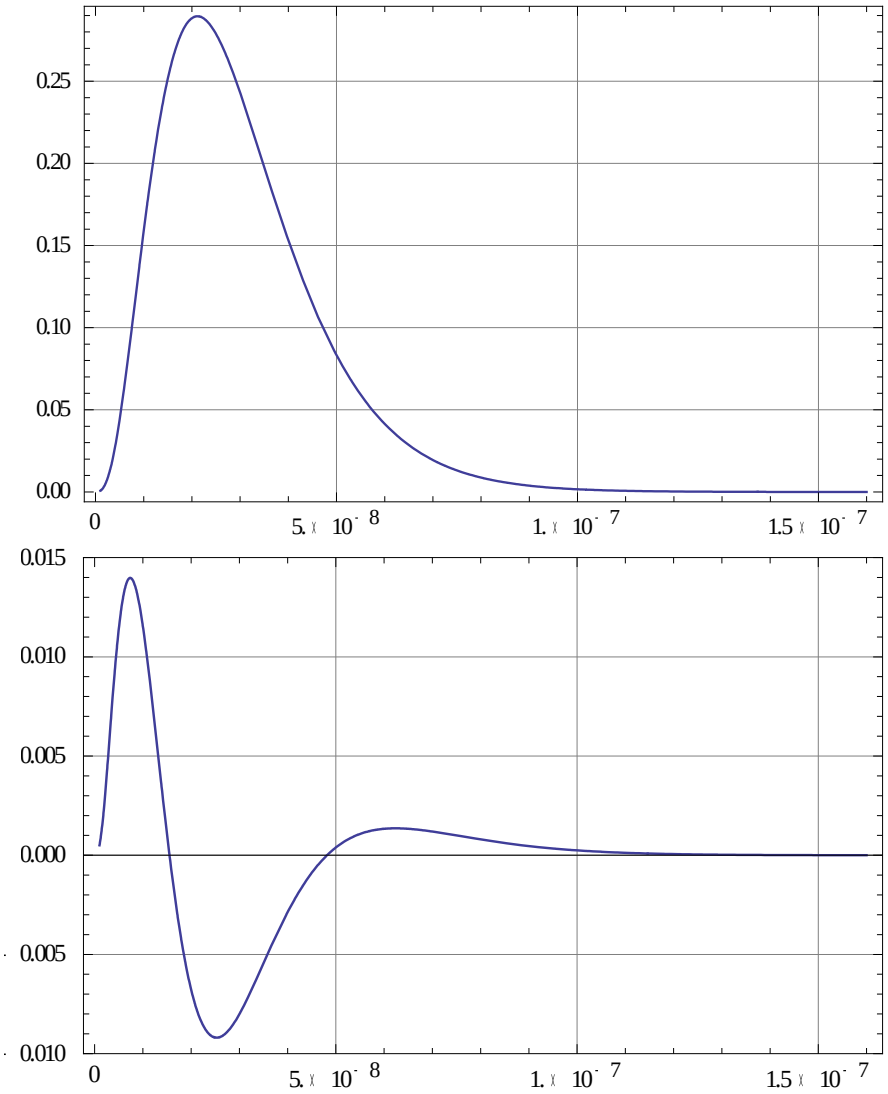
The response of Front End to delta Dirac function in time domain will be:

$$L^{-1}\left(T_{FE}\cdot i_s\right)$$

The crosstalk of first neighbor in time domain will be:

$$L^{-1}\left(T_{FE}\cdot i_c\right)$$

Example of calculation for transimpedance mode



Kv=83dB, τ_{p0} =200ns, GBP=1GHz *)

Detector; c_{is}=7pF, c_b=4pF (ATLAS SCT)

Response; Max=0.289 for t=21ns
(0.27 for 20ns without detector)

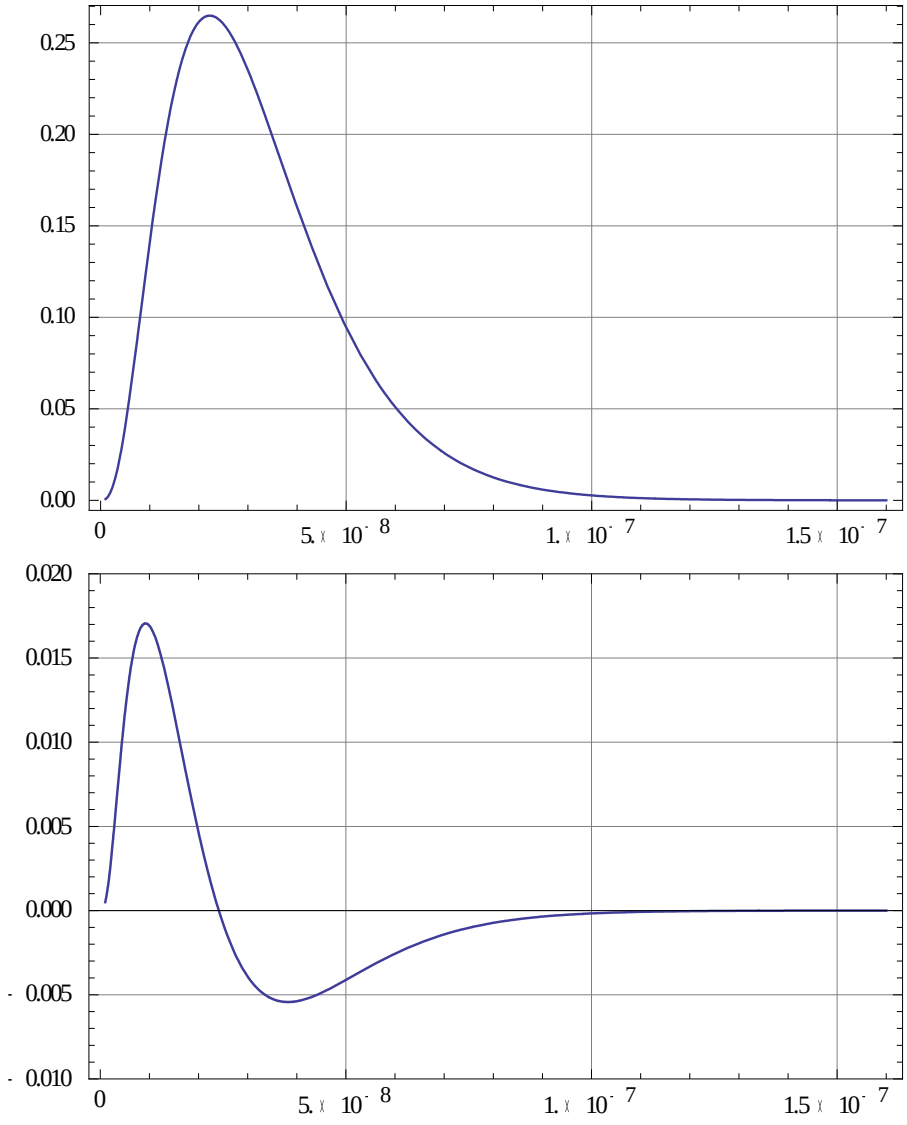
$$\int_0^\infty i_s(t)\,\partial t = 1$$

The overall charge readout by readout channel for transimpedance preamplifier is full!

Crosstalk; Max=0.0139 for t=7.4ns (5%)

Design presented in; J. Kaplon and W. Dabrowski, "Fast CMOS binary front end for silicon strip detectors at LHC Experiments," *IEEE Trans. Nucl. Sci.*, vol. 52, no. 6, pp. 2713–2720, Dec. 2005

Example of calculation for charge preamp



Kv=83dB, τ_{p0} =200ns, GBP=1GHz *)

Detector; c_{is}=7pF, c_b=4pF (ATLAS SCT)

Response; Max=0.265 for t=22.2ns
(0.27 for 20ns without detector)

$$\int_0^\infty i_s(t)\,\partial t = 0.99$$

Lost of charge related to finite open loop gain of the preamplifier!

Crosstalk; Max=0.017 for t=9.2ns (~6.5%)

Preamplifier stage the same as in the last slide but working in charge mode (very high RF, CR-RC2 filter build with shaper only)

What we have learned so far?

- Charge preamplifier worse than transimpedance preamplifier in terms of;
  - charge collection efficiency
  - cross talk signal amplitude
- For few pF detector capacitances as planned for SLHC the optimal open loop gain of the preamplifier should be around **70 to 80dB** (in order to provide cross talk less than 5%)
- For peaking time around 20 ns as for SLHC the GBP should be above **1 GHz**

Another requirements concerning front end circuits

- Low power (<300uW/channel), low noise (S/N>15 → ENC < 1000e-)
  - optimization of power for a minimum affordable noise level → influence on the architecture (single ended)
- Collisions of particles every 25 ns → data time tagging to the given BCO (peaking times <25ns)
- Stability → required phase margin above 85 to 90 degree
- Optimum PSRR (large systems, difficult to provide clean power supply)
- Radiation hardness – doses >2×10^{14} N/cm^2 (1MeV) and >10MRad (CMOS front end preferred)

Comparison of basic analogue parameters for three generations of IBM CMOS processes

IBM CMOS	250nm RF	130nm RF	90nm LP (low power)
t_{ox} physical/effective	5nm/6.2nm	2.2nm/3.12nm	2.1nm/2.8nm
K_a ($C_{ox}\cdot\mu$) NMOS	330 uA/V^2	720 uA/V^2	800 uA/V^2
Vdd	2.5V	1.2V (1.5V)	1.2V
g_m/g_{ds} Weak Inv.	70	30	18
Peak ft	35 GHz	94 GHz	105 GHz

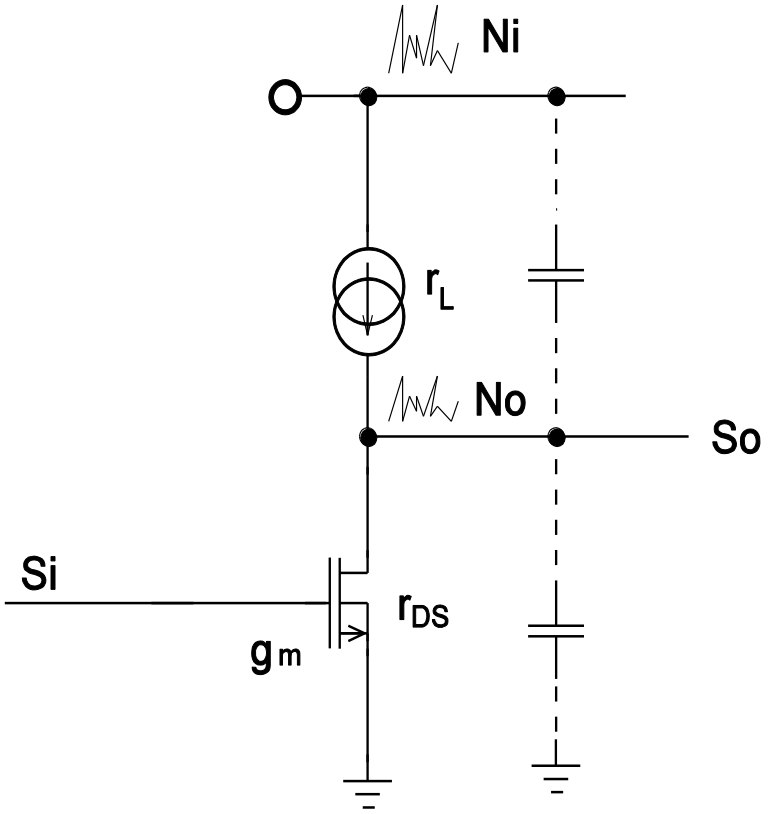
Scaling advantages; higher ft, higher Ka

Challenges for front end; lower Vdd (lower dynamic range), **lower intrinsic transistor gain**

Motivations to increase open loop gain

- Lower input impedance of preamplifier;
  - better charge collection efficiency
  - lower cross talk
- Optimizing feedback impedance (i.e. pulse gain of the preamplifier) versus input impedance
- **PSRR (all single ended stages)**

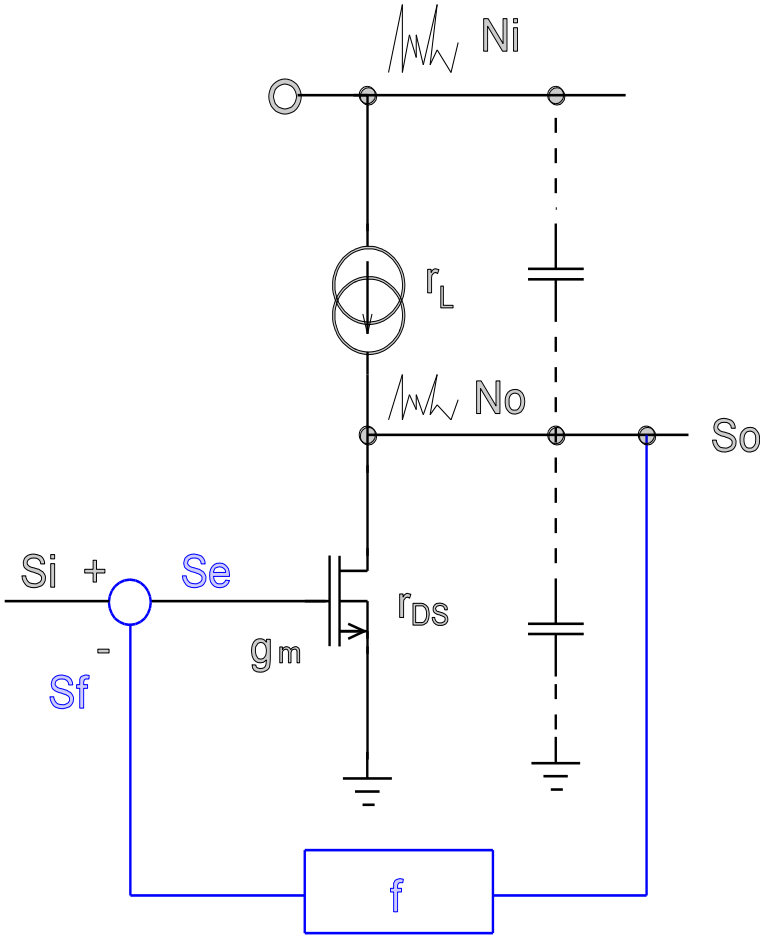
PSRR for single ended stage (1)



$$K_U = \frac{So}{Si} = g_m \cdot \left(r_L \parallel r_{DS} \right)$$

$$N_O = N_i \cdot \frac{Z_{DS}(s)}{Z_{DS}(s) + Z_L(s)}$$

PSRR for single ended stage (2)



$$S_O = K_U \cdot S_e + N_O$$

$$S_f = f \cdot S_O$$

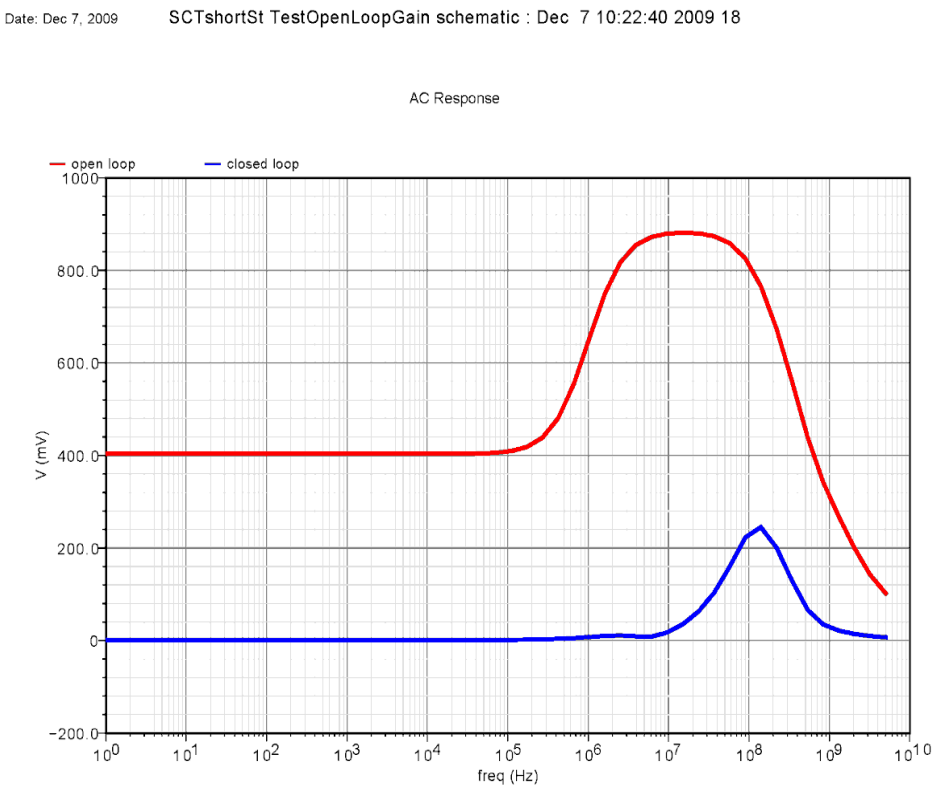
$$S_e = S_i - S_f$$

Loop gain

$$S_O = S_i \frac{K_U}{1 + K_U \cdot f} + N_O \frac{1}{1 + K_U \cdot f}$$

Driving K_U improves PSRR.
All single ended stages should be designed as feedback amplifiers with high open loop gain.

PSRR for single ended stage (3)

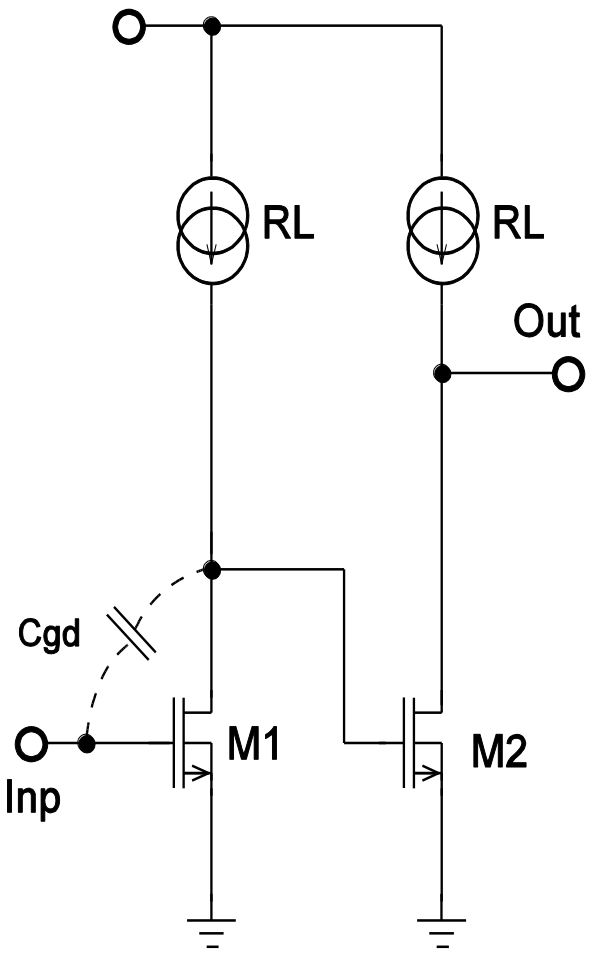


Power supply disturbance (1V) seen at cascode output working in open loop configuration (red) and in transimpedance preamplifier (blue). 130nm version of front end.

Basic configurations for gain boosting

Intrinsic gain in 130nm ~30 V/V → we need 70 to 80dB (2000 to 10000 V/V)…

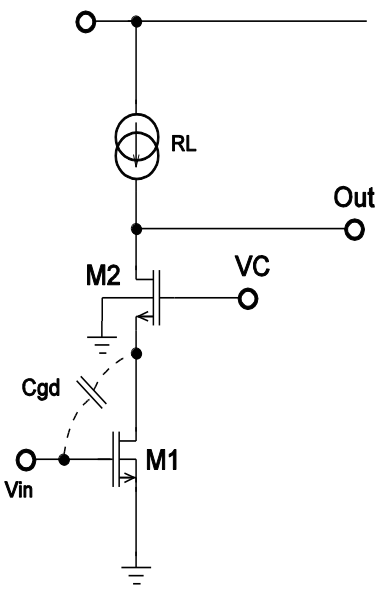
Cascade

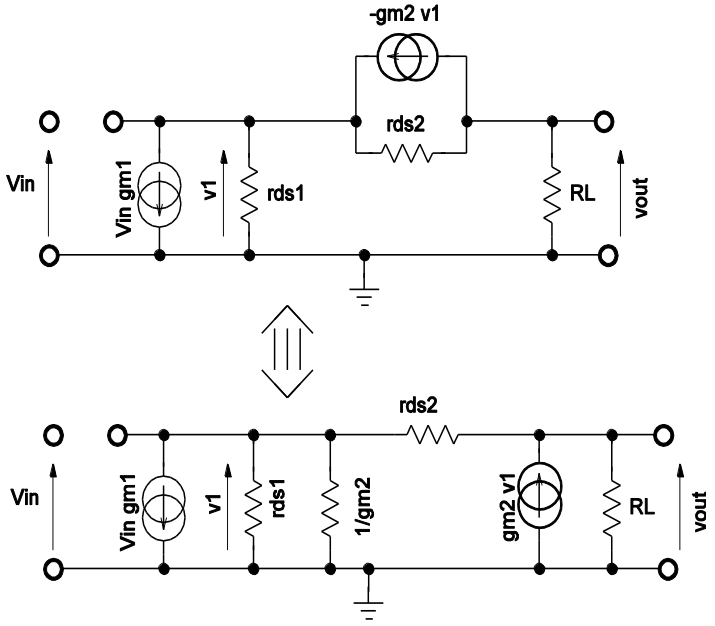


$$K_U = \frac{g_{m1}}{g_L + g_{DS1}} \cdot \frac{g_{m2}}{g_L + g_{DS2}}$$

- Two stage i.e. two pole circuit; needs to be stabilized
- Significant gain after first stage; Miller effect in case of driving from high impedance (as for silicon detector) → not used as an preamplifier stage
- In 90 nm the gain of cascade is significantly degraded because of intrinsic transistor gain, some circuits which works in 250nm version shows bad PSRR characteristic

Cascode; common source – common gate amplifier



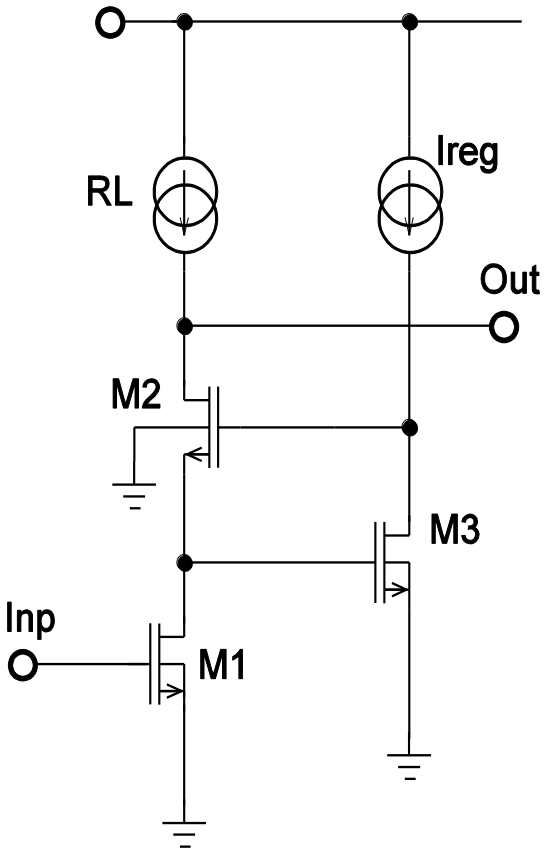


$$K_U \cong \frac{-g_{m1}}{g_{DS1}\dfrac{g_{DS2}}{g_{m2}} + g_L}$$

$$GBW = \frac{g_{m1}}{2\pi\, C_{OUT}}$$

- single stage amplifier; one dominant pole
- no Miller effect (low gain of common source stage)

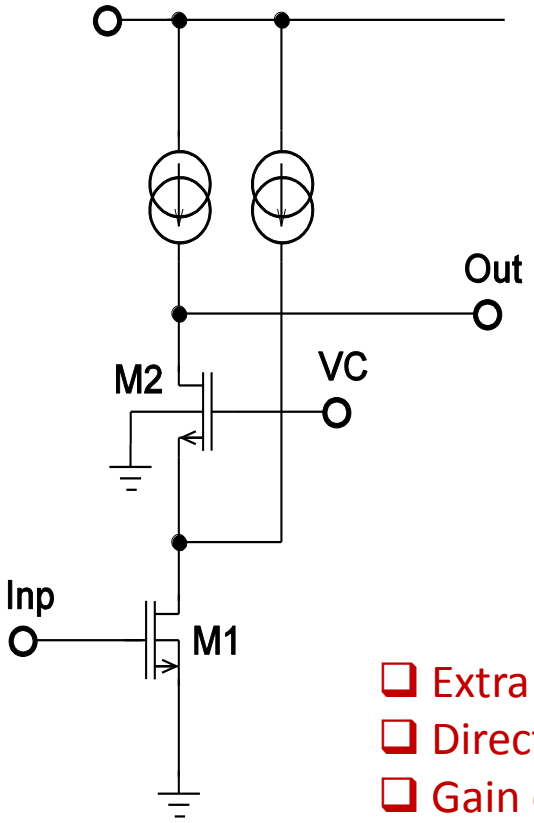
Regulated cascode



$$K_U \cong \frac{-g_{m1}}{\frac{g_{DS1} \cdot g_{DS2}}{g_{m2}} \cdot \frac{g_{DS3}}{g_{m3}} + g_L}$$

- Cascode transistor controlled with common source amplifier
- Higher output conductance of cascode; possible higher gain
- GBW the same as for simple cascode

Boosting bandwidth and gain in cascode

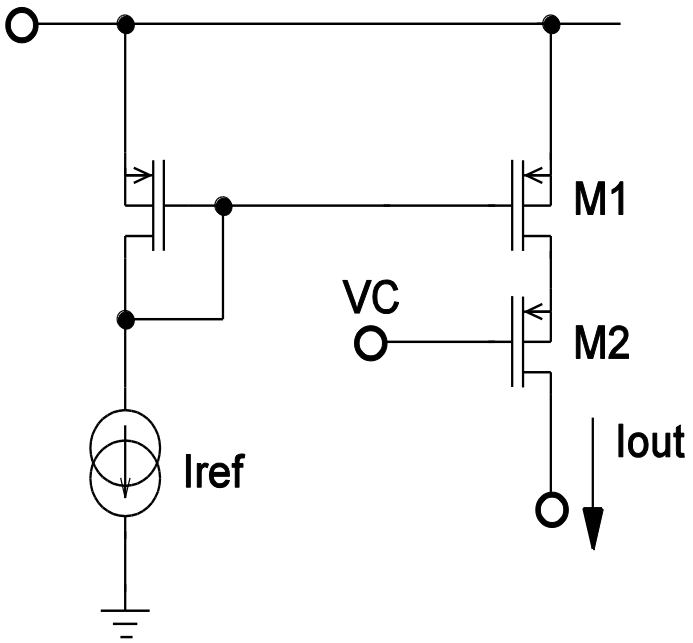


$$GBW = \frac{g_{m1}}{2\pi\, C_{OUT}}$$

$$K_U \cong \frac{-g_{m1}}{\frac{g_{DS1} \cdot g_{DS2}}{g_{m2}} + g_L}$$

- Extra current source to drain of M1 → increase of g_{m1}
- Direct impact on gain bandwidth
- Gain changed according to output conductance of cascode and active load

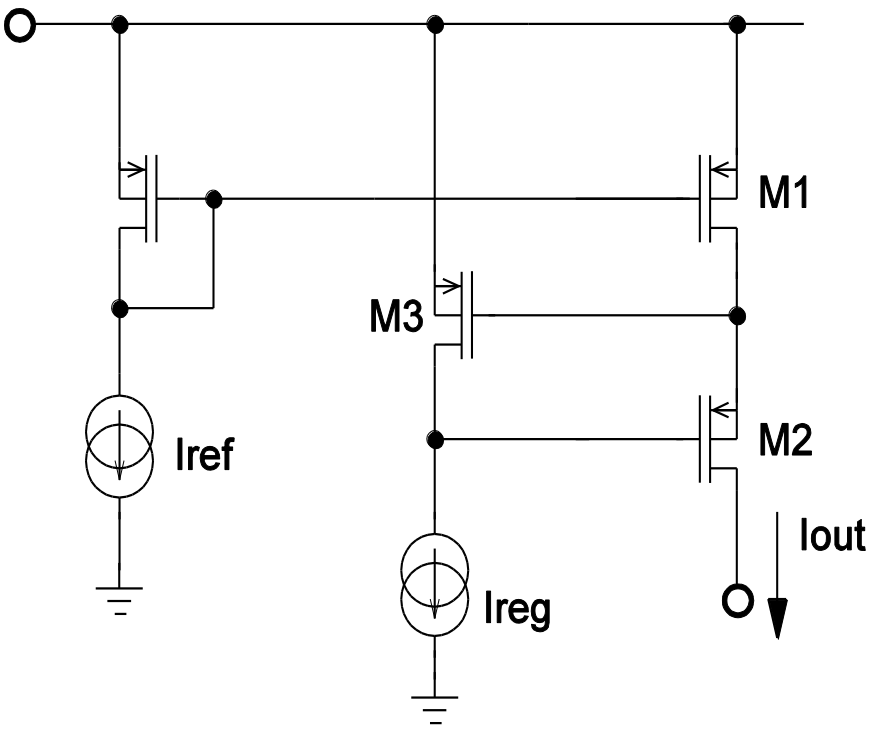
Active load for cascode stage (cascode load)



$$r_{OUT} \cong g_{m2} \cdot r_{DS2} \cdot r_{DS1}$$

- Amplification of r_{DS1} by g_{m2}
- For short SSD application; OK for 250nm, not sufficient for 130 & 90nm

Active load for cascode stage (regulated cascode load)

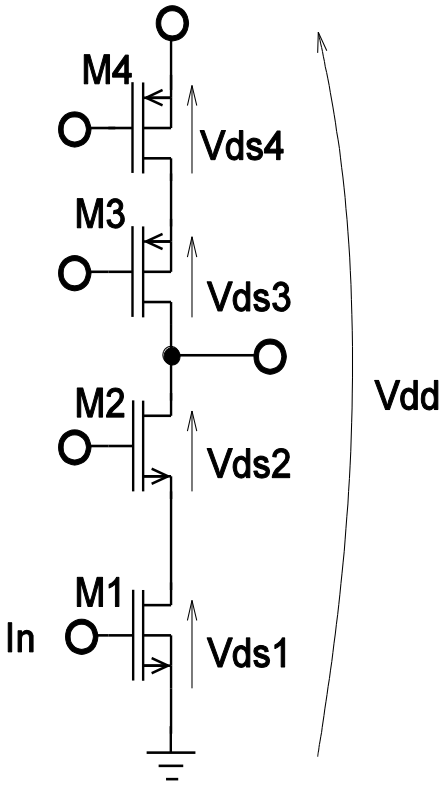


$$r_{OUT} \cong g_{m2} \cdot r_{DS2} \cdot g_{m3} \cdot r_{DS3} \cdot r_{DS1}$$

- Amplification of r_{DS1} by g_{m2} and g_{m3}
- Used in 130 & 90nm versions of preamplifiers

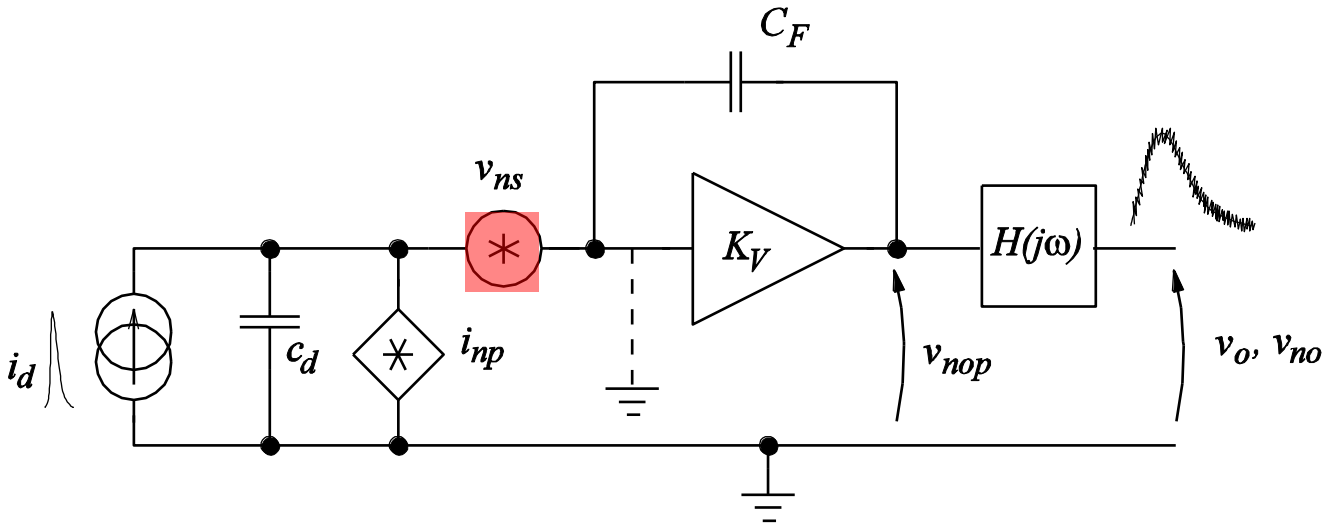
Biasing transistors in weak inversion; motivations

Biasing of the cascode



- All, four, transistors must be in the saturation ($V_{DS} \geq V_{GS} - V_T$)
- Technology scaling → Vdd diminished from 2.5V in 250nm to 1.2V in 130nm and 90nm CMOS → possible problems with dynamic range
- **Solution → subthreshold operation ($V_{GS} \approx V_T$)**
- Minimum $V_{DS\,SAT}$ for weak inversion roughly 5 U_T (125mV)

Noise optimization in CR-RC^n filters for multi-channel FE electronics



$$ENC[C] = F_V \cdot \frac{\overline{v_{ns}}}{\Delta f} \cdot c_d \Big/ \sqrt{t_{peak}} \oplus F_i \cdot \frac{\overline{i_{np}}}{\Delta f} \cdot \sqrt{t_{peak}}$$

$$\frac{\overline{v_{n\,thermal}}}{\Delta f} = \sqrt{\frac{4 \cdot k \cdot T \cdot n \cdot \gamma}{g_m}}$$

Transconductance in MOS transistor (EKV model)

Specific current

$$I_S = 2 \cdot n \cdot K_P \cdot \frac{W}{L} \cdot U_T^2$$

$$U_T = \frac{k \cdot T}{q}$$

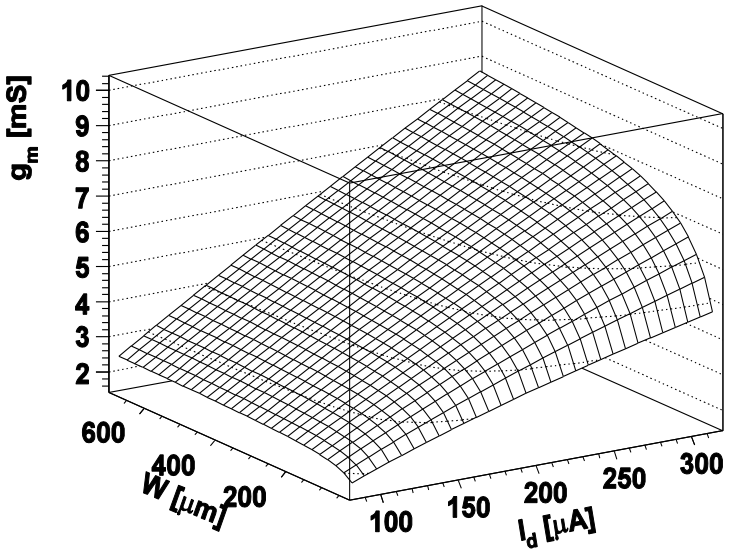
WI/SI interpolation for $I_f = I_D / I_S$

$$G(I_f) = \frac{1}{\sqrt{I_f + \frac{1}{2} \cdot \sqrt{I_f} + 1}}$$

Transconductance:

$$g_m = G(I_f) \cdot \frac{I_D}{n \cdot U_T}$$

g_m in weak inversion

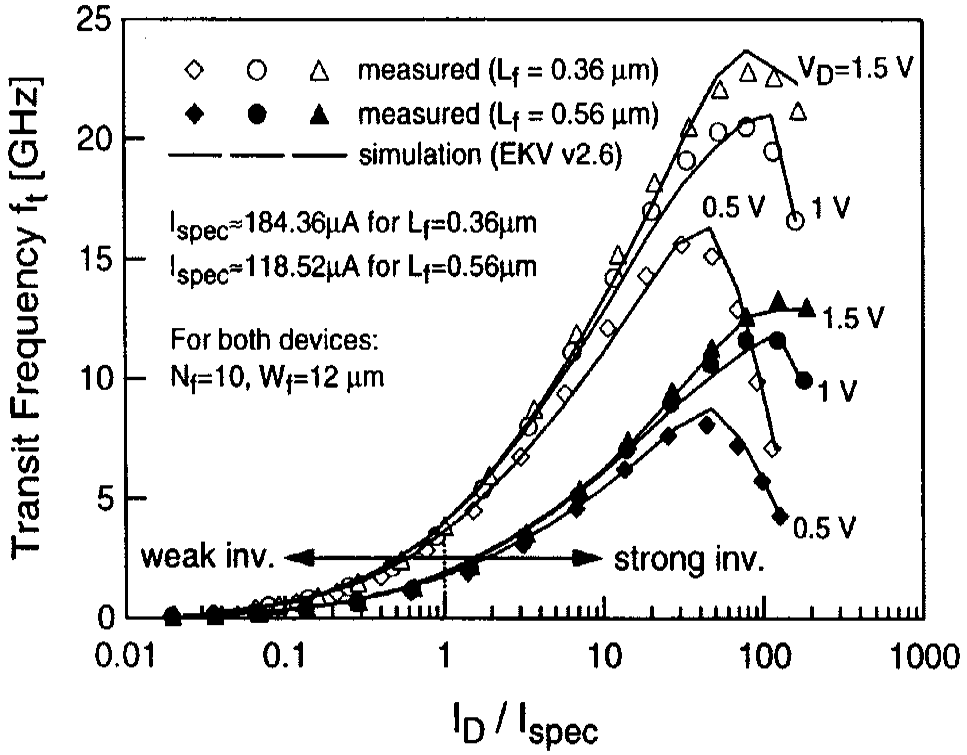


IBM 130nm
NMOS
L=300nm

- Weak inversion provides highest transconductance at a given bias current
- Some technologies report excess noise for devices in strong inversion
- **Conclusion; weak inversion in input transistor is good from the standpoint of power consumption/noise optimization**

Biasing transistors in weak inversion; some consequences

Impact of the inversion order on the speed of CMOS circuit

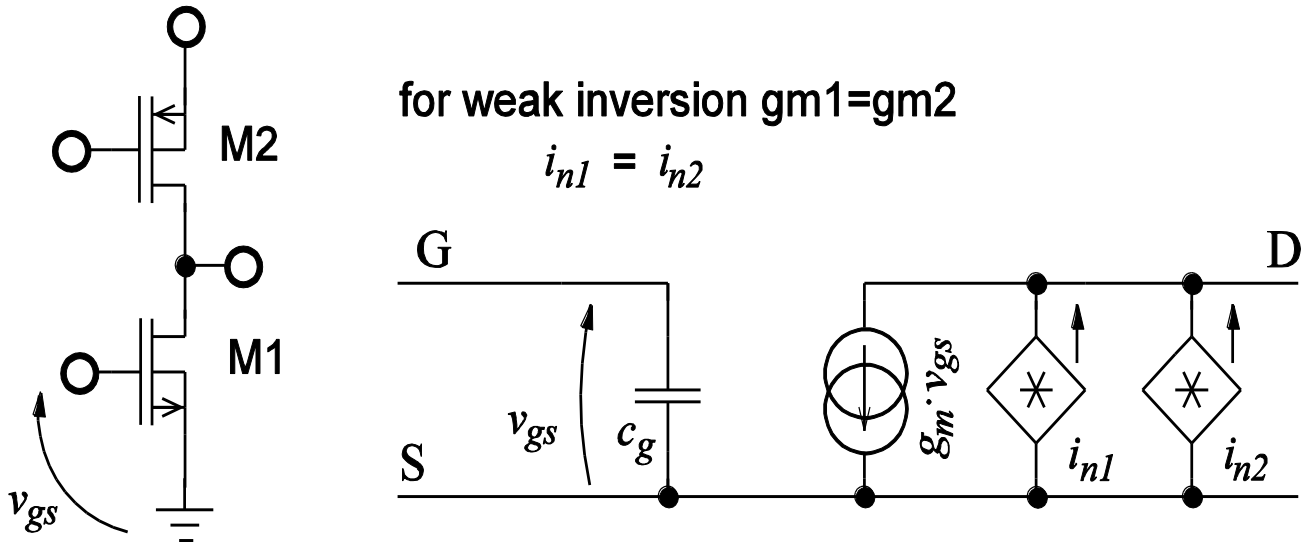


Transit frequency ft as a function of inversion order for 250nm CMOS technology *

For devices biased in weak inversion we never obtain highest possible speed of a given technology

* C.Enz, "MOS transistor modeling for RF IC design", *IEEE J.Solid-State Circ.,* vol. 35, no. 2, pp.186-201)

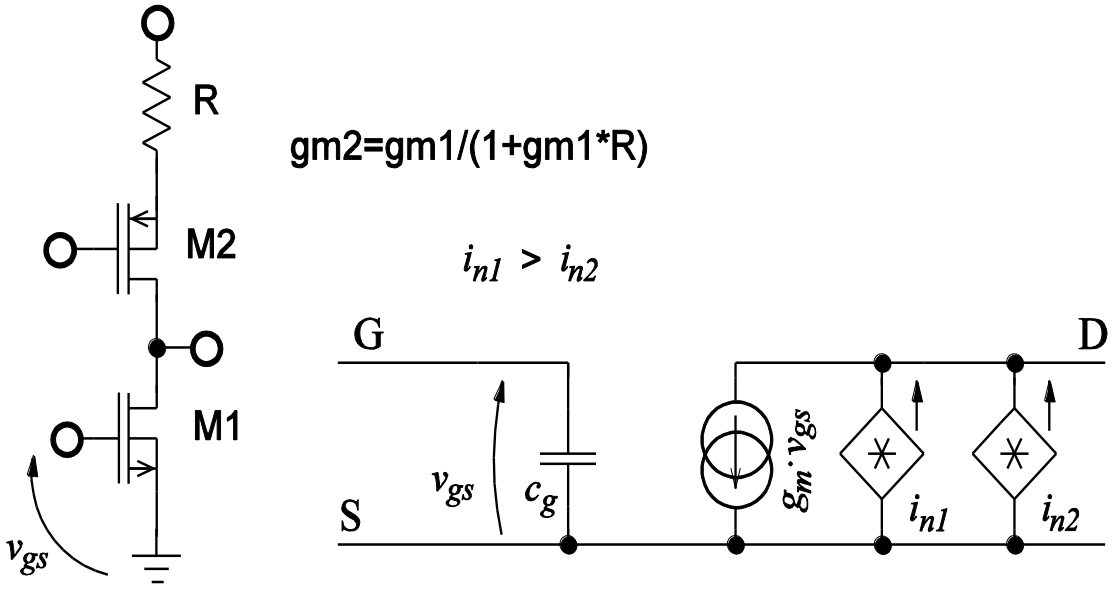
Noise of the active load (1)



If all transistor in weak inversion the gm is defined only by current → all gm the same

Increase of input series noise by ~40%!

Noise of the active load (2)

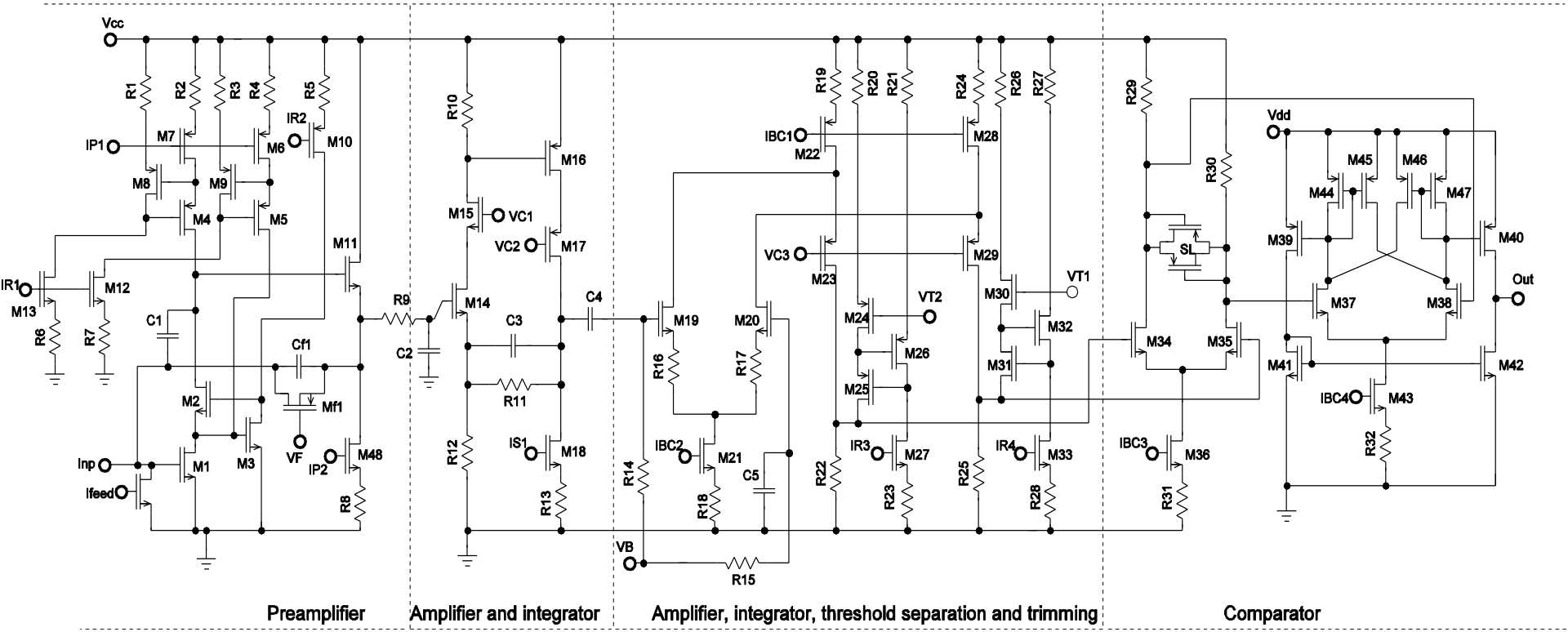


Resistive degeneration of gm works

But we have to spend another ~100mV taken out from Vdd…

Architecture of front ends implemented in 130 & 90nm processes

Front end channel in 130nm & 90nm technology (SCT short strips)

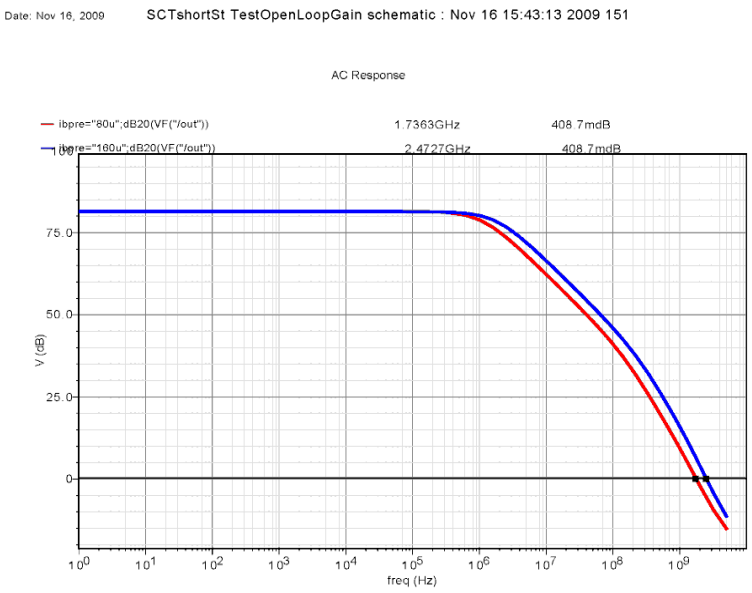


5.5 mV/fC
tp 8 ns

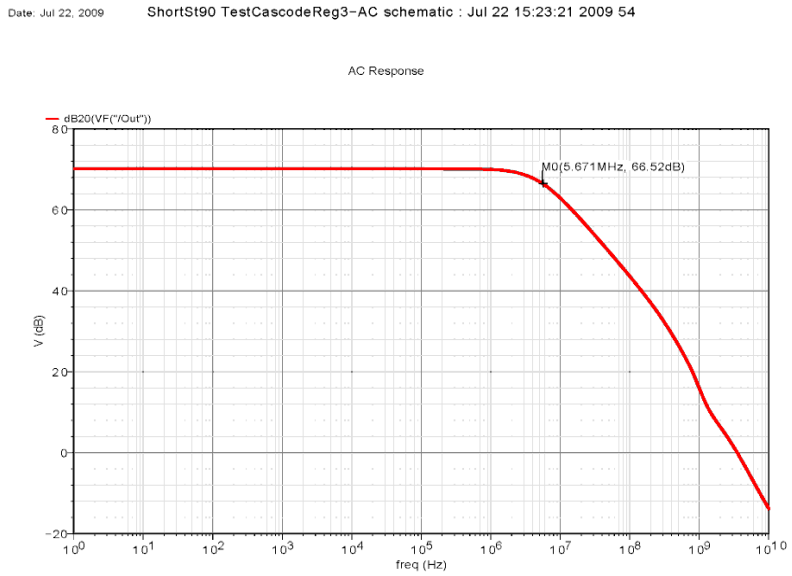
30 mV/fC
tp 18 ns

100 mV/fC
tp 22 ns

Preamplifiers open loop gain

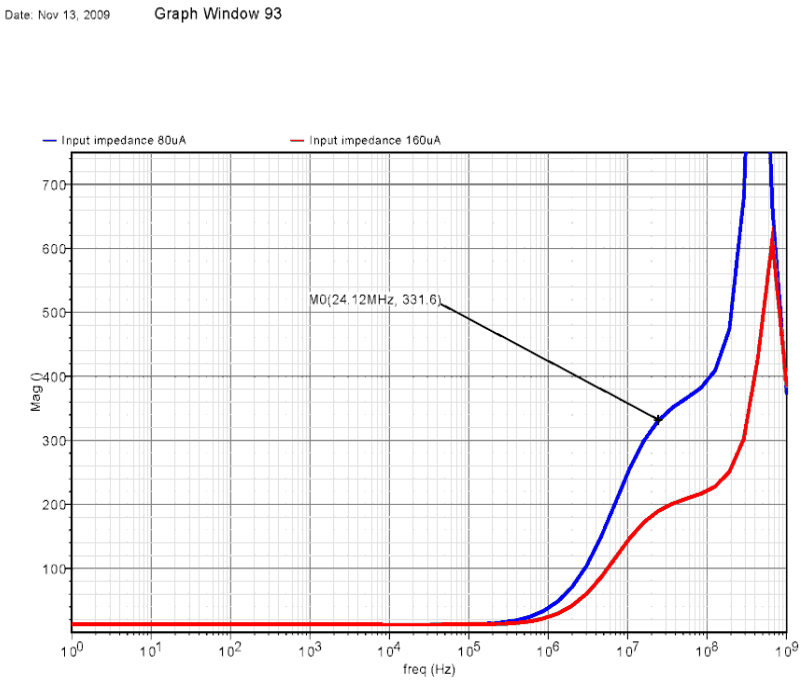


130nm, 80dB, GBW=2GHz
Iin=80uA

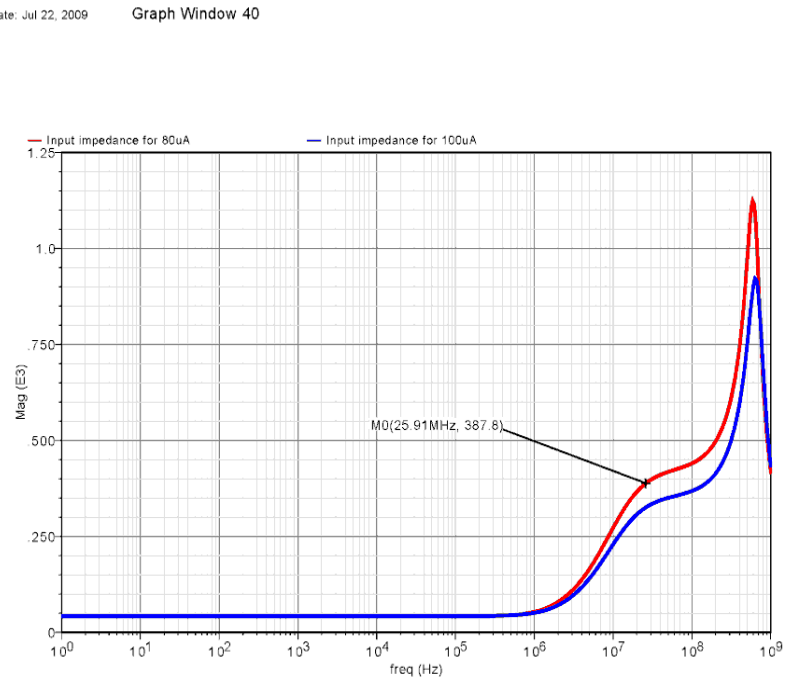


90nm, 70dB, GBW=3.5GHz
Iin=80uA

Preamplifiers input impedances



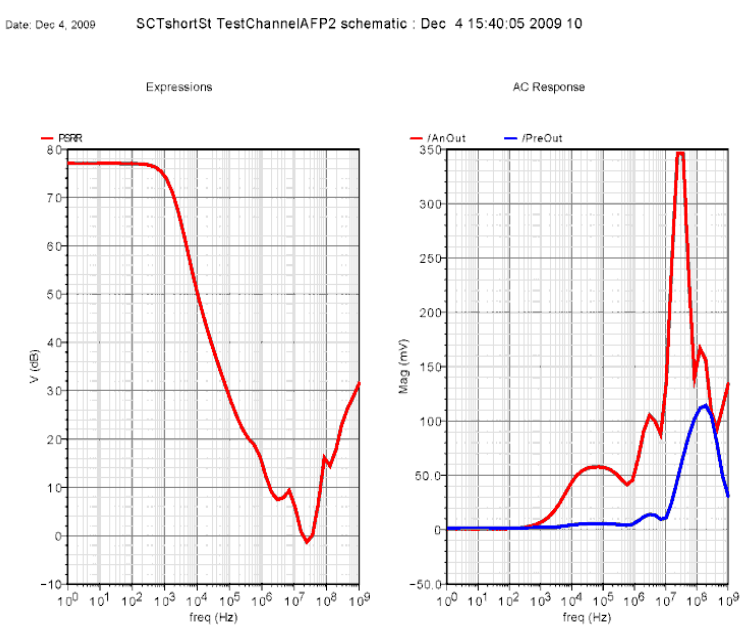
130nm, 330Ω @25MHz
Iin=80uA



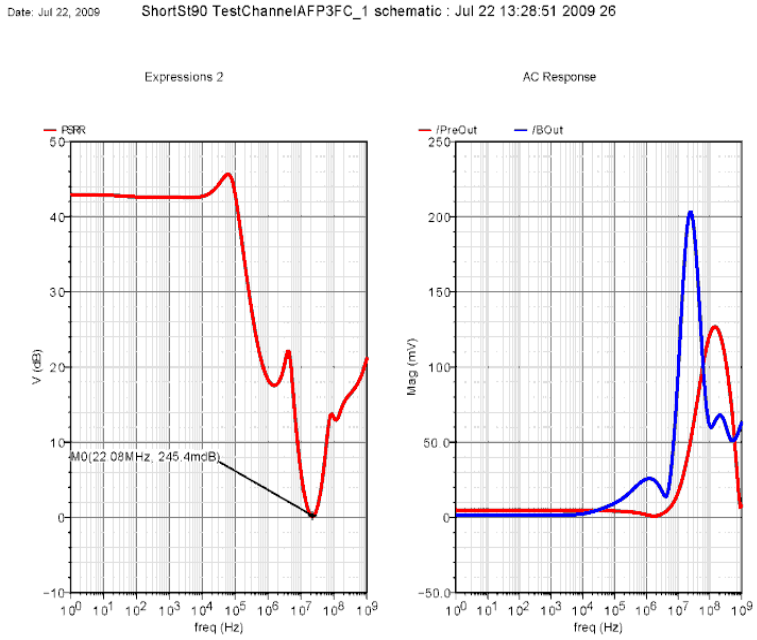
90nm, 380Ω @25MHz
Iin=80uA

In both cases the cross talk signals less than 3%
Detector 1.5 pF to bulk + 2x 1.6 pF to neighbor

PSRR *



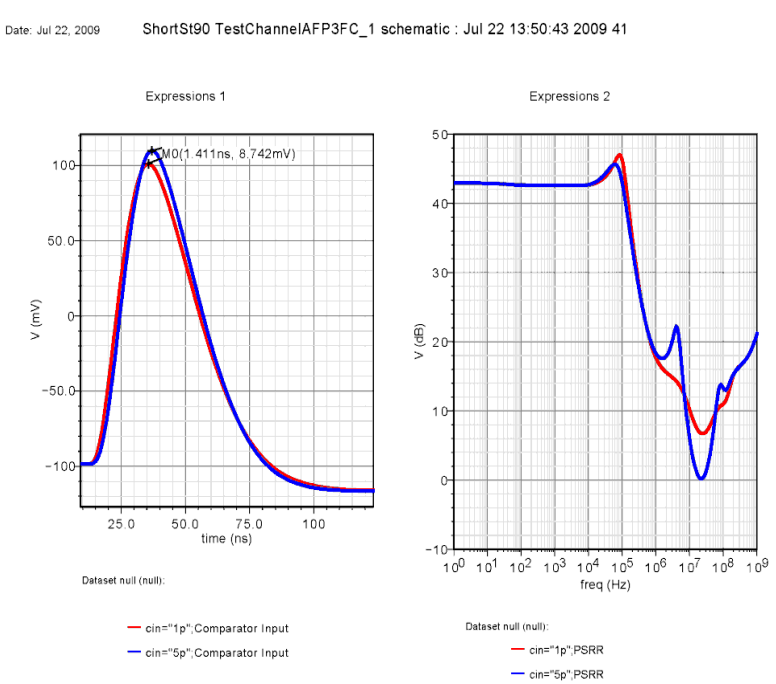
130nm, -0.5dB @ 25MHz
Iin=80uA, Cin=5pF to GND

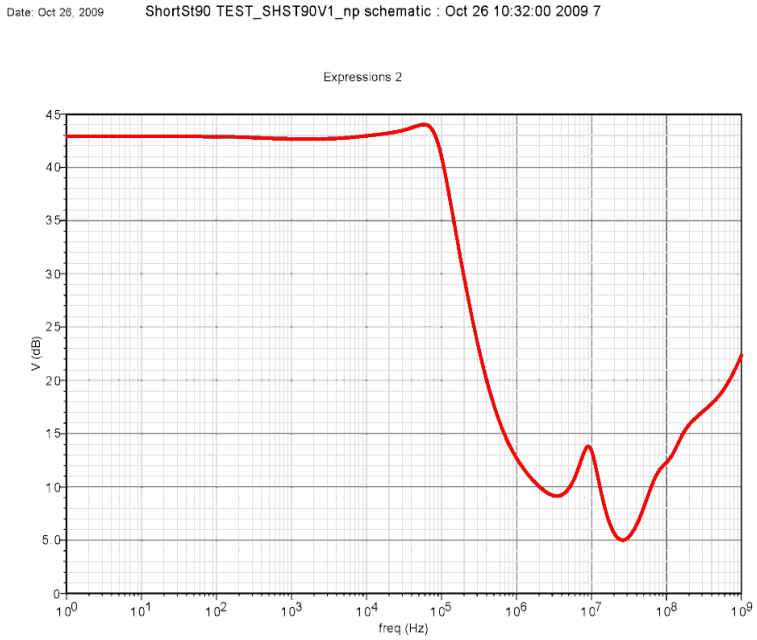


90nm, +0.5dB @ 25MHz
Iin=80uA, Cin=5pF to GND

* PSRR defined as the ratio of the 1V signal at the power supply line to the signal at the output. For two different front end one should also look at the charge gain!

PSRR (2)

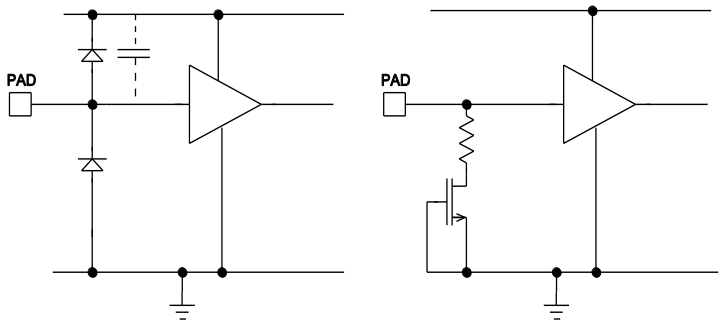


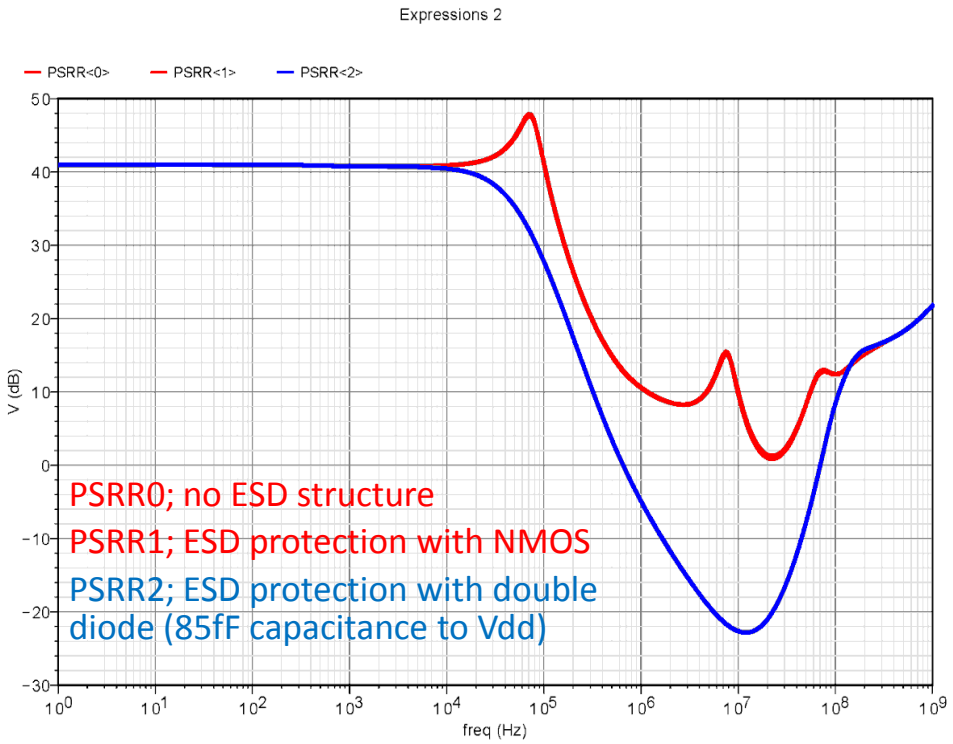


PSRR improves when:

- Cin decreases (also in case of real detector when part of the detector capacitance is connected to neighboring channel)
- Bias current increases (GBW increases → loop gain increases)

PSRR (3)

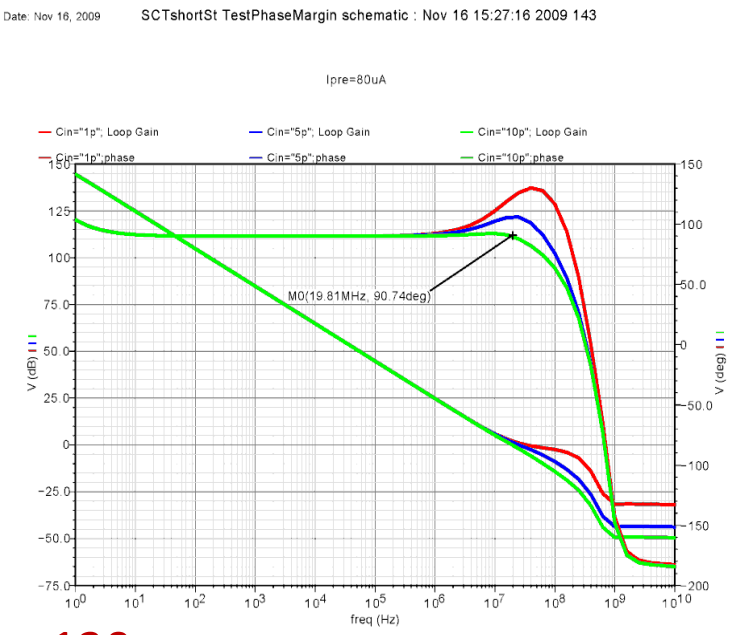




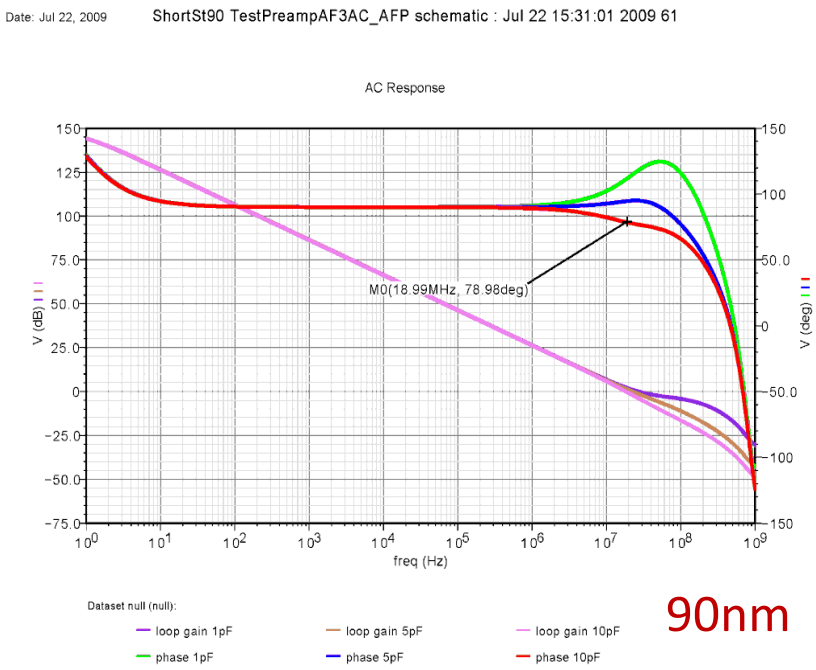
PSRR might be broken by input protections:

- ❑ double diode structure not admitted
- ❑ preferable structure; silicide blocked NMOS
  - ❑ drawback → higher capacitance

Phase margin



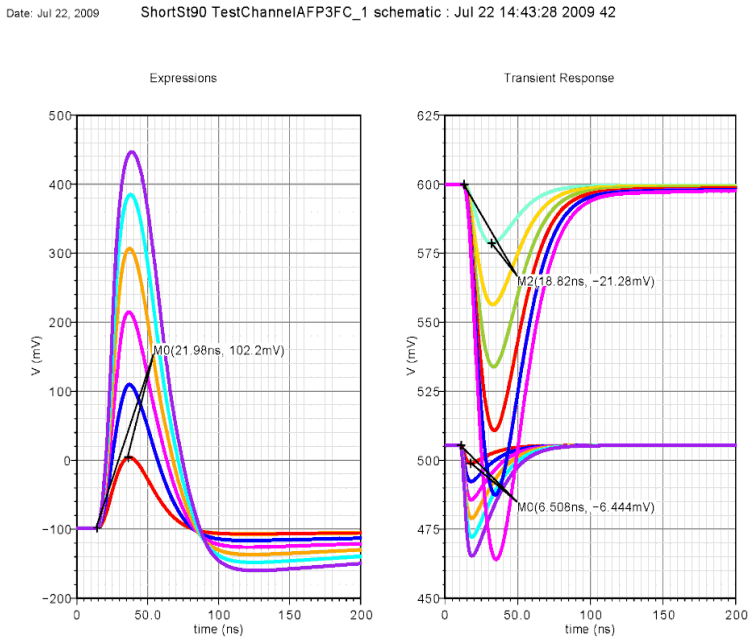
130nm

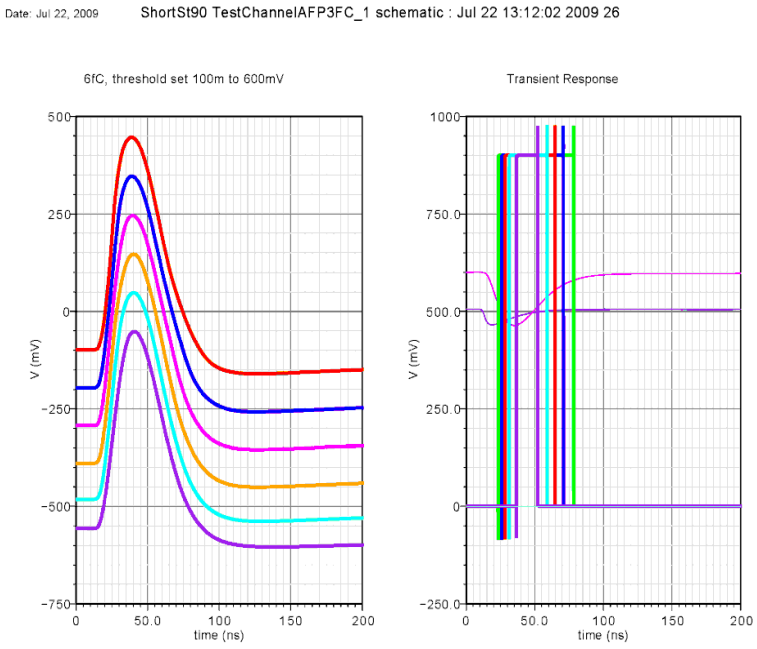


90nm

We want to have 90 degree for nominal input capacitance (5pF), this has impact on input impedance and PSRR but safety first.

Linearity and dynamic range



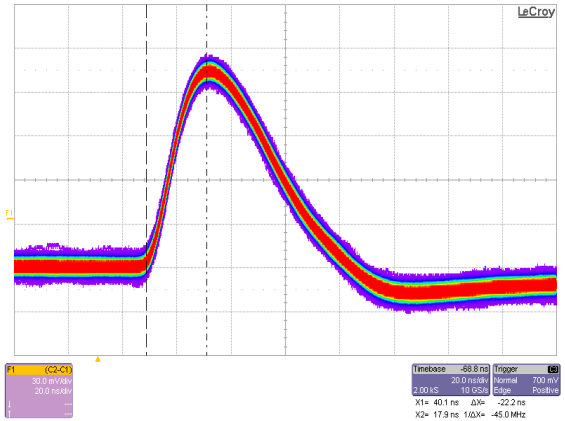


In 130nm and 90nm versions dynamic range up to 6 fC (good linearity up to 4fC)
Same 1.2V Vdd *)

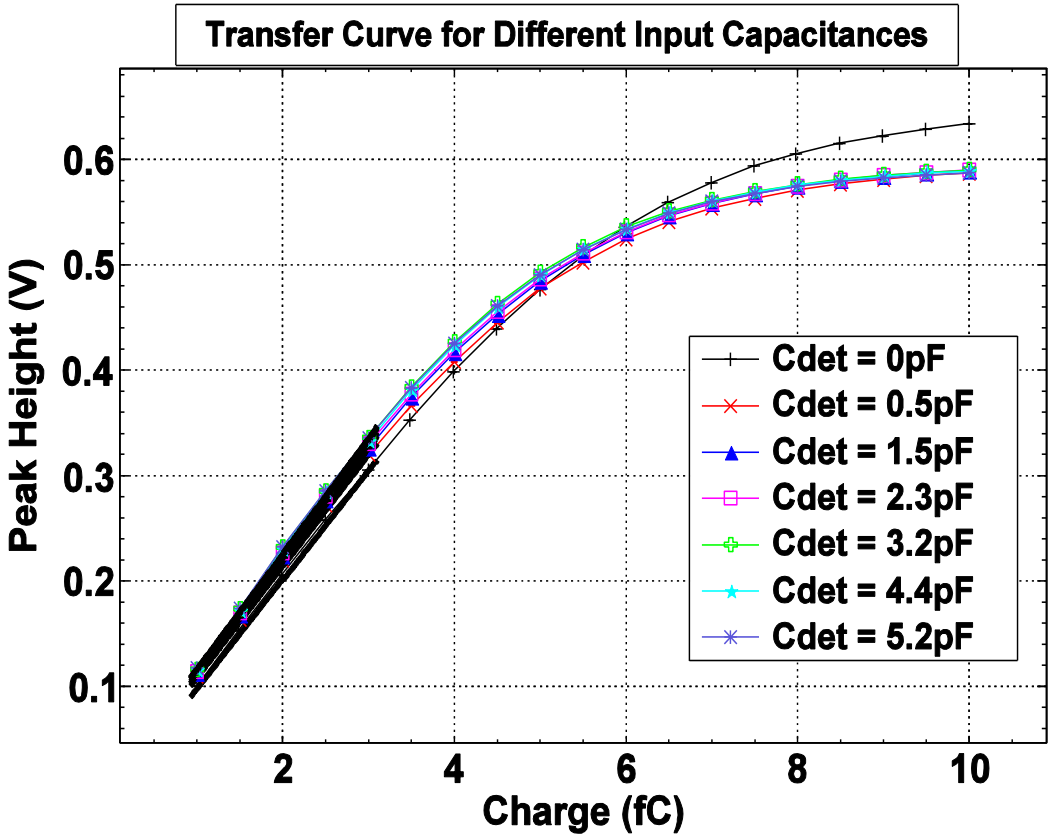
*) in 250nm version the dynamic range (limit in discriminator stage - might be adjusted) is about 12fC

Selected results from 90nm and 130nm front end for short strips silicon detectors

Dynamic range and linearity (90nm)



Peaking time 22ns
Gain 100mV/fC

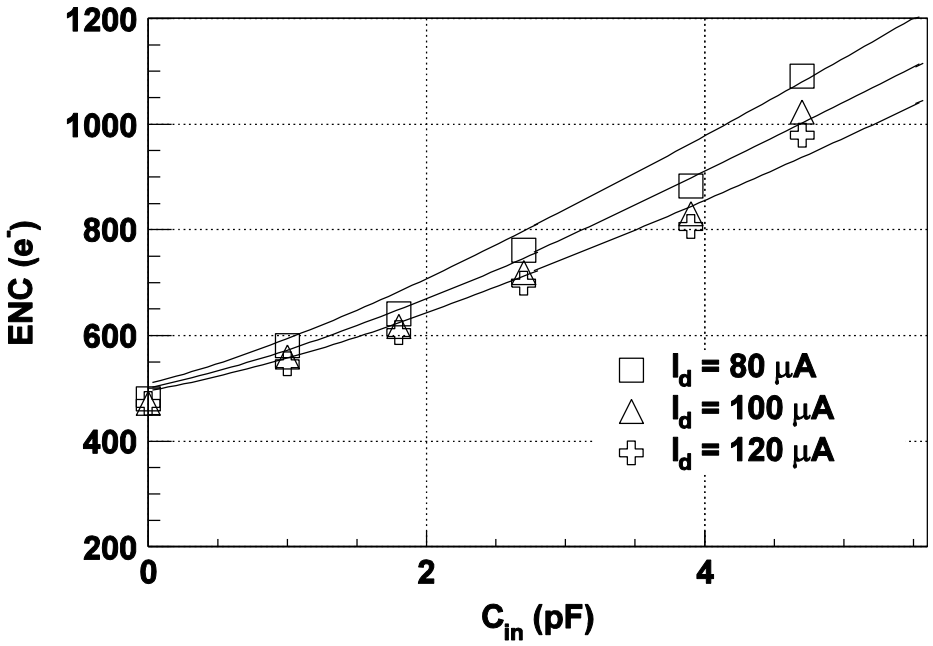


Good linearity up to 4fC (400mV signal range)
Dynamic range up to 6fC (limit set by the bias of the differential stage).
Good agreement with simulation → the same figures for 130nm version

Noise performance of 130 nm version

As predicted by theoretical model → no excess noise in 130 nm for NMOS devices with gate length of 300nm

Noise performance of 90 nm version



Markers are measurement points, lines are theoretical fit with excess noise set to **3**!.

High excess noise for regular transistors. Analog transistors not available at the time of submission.

Comparison of power consumption at constant ENC for Cdet=5pF and ENC = 800e-

	250nm (ABCN25)	**130nm**	**90nm**
$\mathbf{I_{input}}$	140uA	80uA	160uA
$\mathbf{I_{total}}$	280uA	160uA	240uA
Vdd	2.5V (2.2V)	1.2V	1.2V

Excess noise!

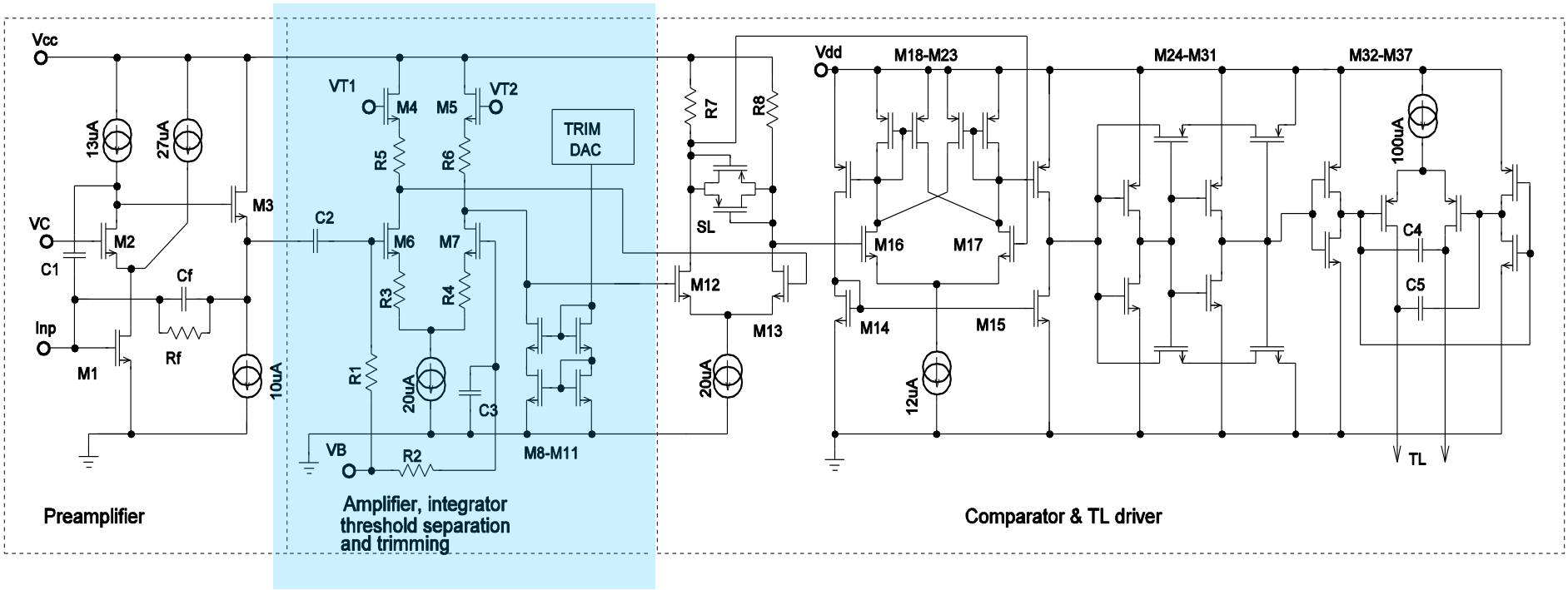
measurements

First look at matching in 130 and 90nm IBM processes

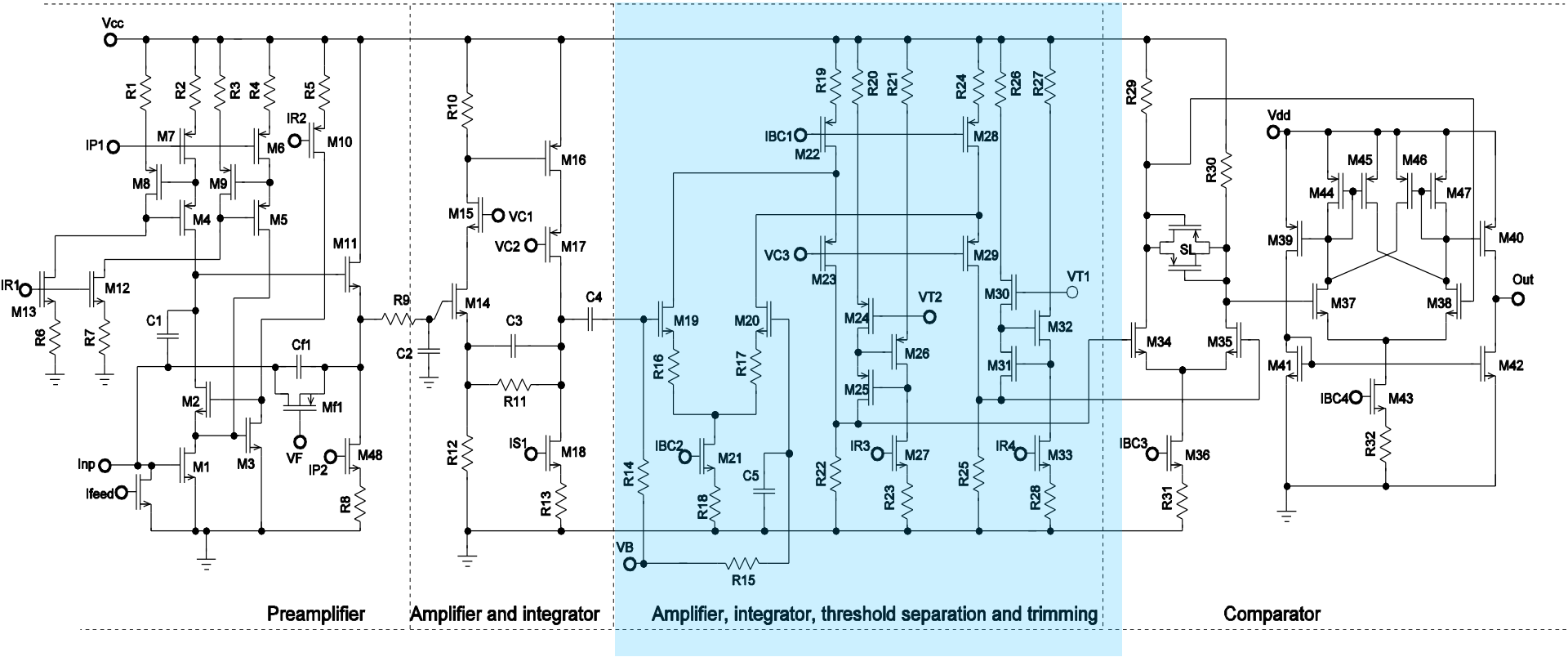
- Detailed study of matching issues requires high statistics (engineering runs with high number of samples)
- In our circumstances we limit statistic to some number of multichannel chips submitted to one or two MPW
- Data for matching in 130nm available also from GTK front end chip (discriminator with voltage threshold distribution)
- One should stress that our architecture is sometimes sensitive to matching.
  - we rely on matching of devices placed over the whole chip area
  - for the presented designs generation of filling structures have been done by IBM.

Matching data for 130nm process; GTK Front End

- Front end for silicon pixel 300x300um (250fF detector capacitance)
  - Transimpedance preamplifier/shaper 5ns peaking time / ENC 180e-
  - Comparator working in voltage mode



Comparator for short strips (130 & 90nm)



Matching, comparison between MC and measurements

	Gain [mV/fC] Measured*/Simulated	RMS Gain [mV/fC] Measured*/MC	RMS Offset [mV] Measured*/MC
130nm FE GTK	72 ←→ 70	1.5 ←→ 1.5	10 ←→ 6
90nm FE	97 ←→ 100	7.5 ←→ 2	18 ←→3
130nm FE	90 ←→ 100	3 ←→ 2	10 ←→ 8

- Discrepancy between MC and data → MC models are too optimistic or problems are related to non optimum layout? Better estimates for 130 nm process.
- High mismatch for 90nm FE gains partially understood (related to matching of feedback current in active feedback preamplifier). High value of RMS due to non Gaussian statistics; pk-pk values for 130 and 90 nm FE much closer (60mV aand 80mV resectively)
- Best matching of gains for GTK FE → related to the fact that preamp uses resistive feedback

*average from measurements of various samples

Summary

- For tracking applications, the technology scaling can provide improvements regarding to lower power and better PSRR (in general due to higher ft)
- Analog (especially noise) parameters does not scaled with technologies; short channel effects might caused excess noise for some technologies/device type
  - The excess noise in 90 nm process should be investigated more deeply if one want to consider it for front end application (with special look on analog transistors available now)
- With the presented architecture, one can obtain reasonable dynamic range and linearity even with low, 1.2V, voltage supply
- For present developments (front end for strips or pixels of reasonable size) the 130nm process is the most competitive technology; cheap, no excess noise, reasonable matching, a lot of nice features (RF metals, good substrate separation)