Front-end electronics for silicon trackers

Valerio Re

Università di Bergamo

Dipartimento di Ingegneria e Scienze Applicate

INFN

Sezione di Pavia

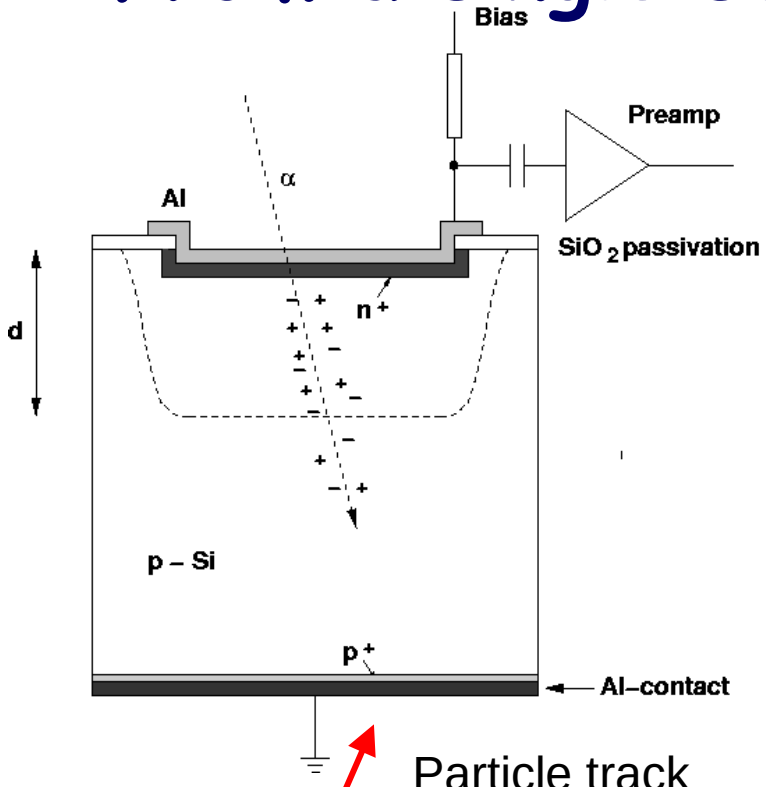
VI Scuola Nazionale "Rivelatori ed Elettronica per Fisica delle Alte Energie, Astrofisica, Applicazioni Spaziali e Fisica Medica“ & RADFAC 2015 thematic day

23-27 Marzo 2015
INFN Laboratori Nazionali di Legnaro (Padova)

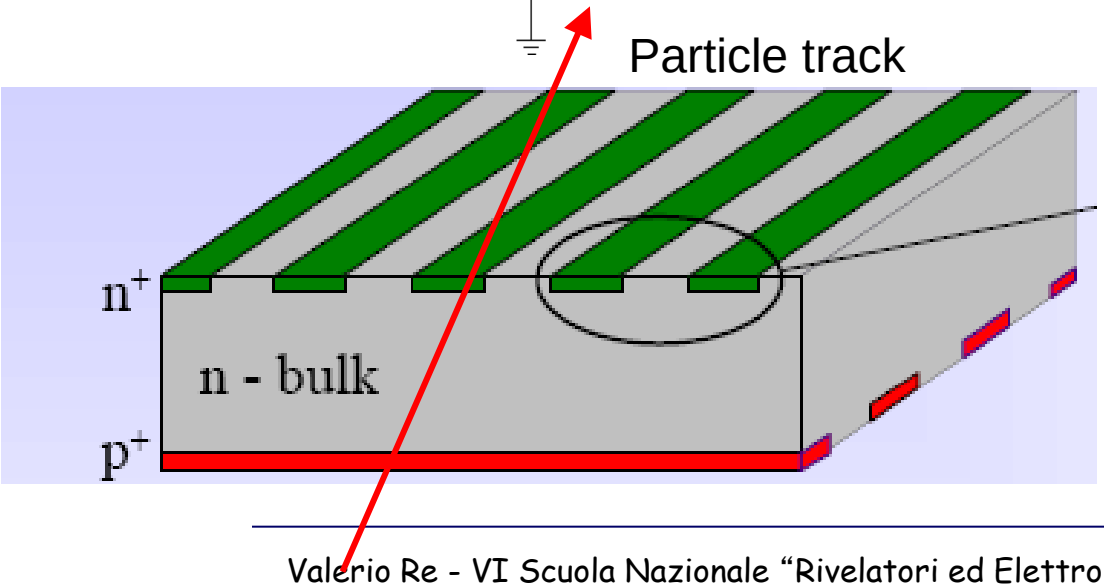
Outline

- Processing of signals from semiconductor detectors: general concepts (amplification, shaping) and electronic noise
- Discussion of fundamental design parameters of front-end electronics for silicon trackers: signal-to-noise ratio, speed, power dissipation, radiation hardness,…
- Architecture of mixed-signal integrated circuits for the readout of **silicon pixel detectors** for tracking and vertexing in high energy physics experiments

From a single semiconductor sensor…



Ionization sensor converts the energy deposited by a particle to an electrical signal. In a fully-depleted semiconductor sensor, electron-hole pairs are swept to electrodes by an electric field, inducing an electrical current.



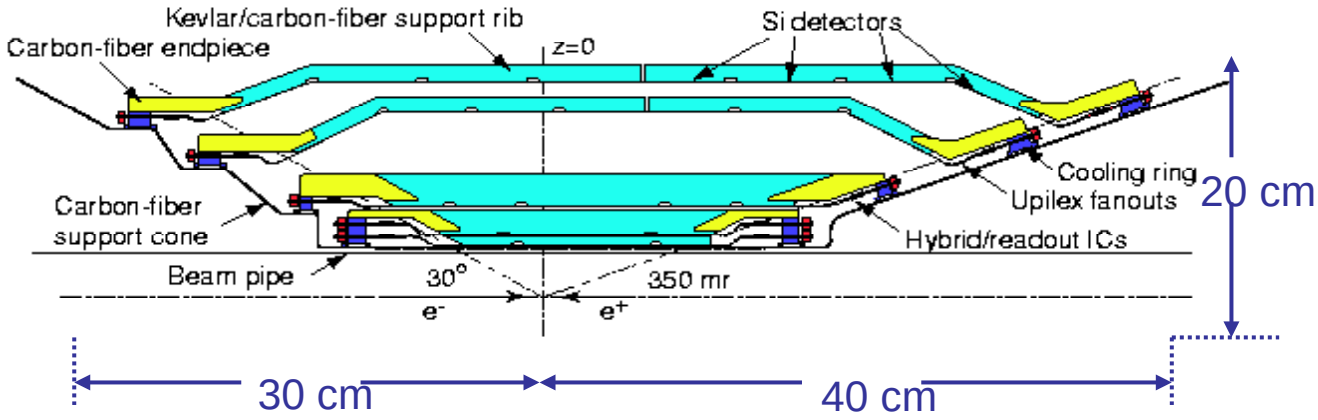
Position-sensitive detector:

Information about the coordinates of the interaction point in a segmented region (presence of a hit, amplitude measurement, timing)

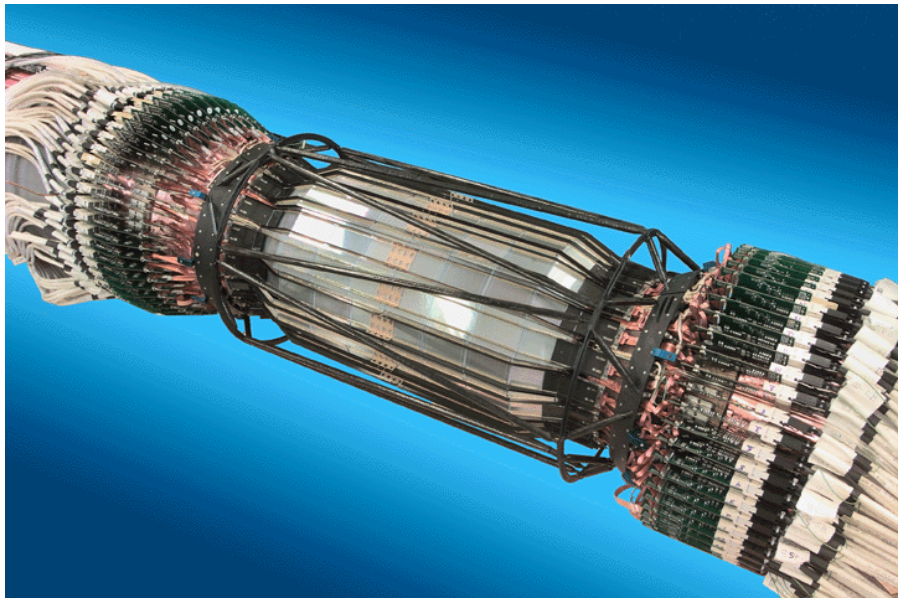
(single-sided or double-sided strip detector, pixel sensors)

…… to a full silicon tracker

Multiple layers of segmented detectors (pixel, strips) provide space points to reconstruct particle trajectories

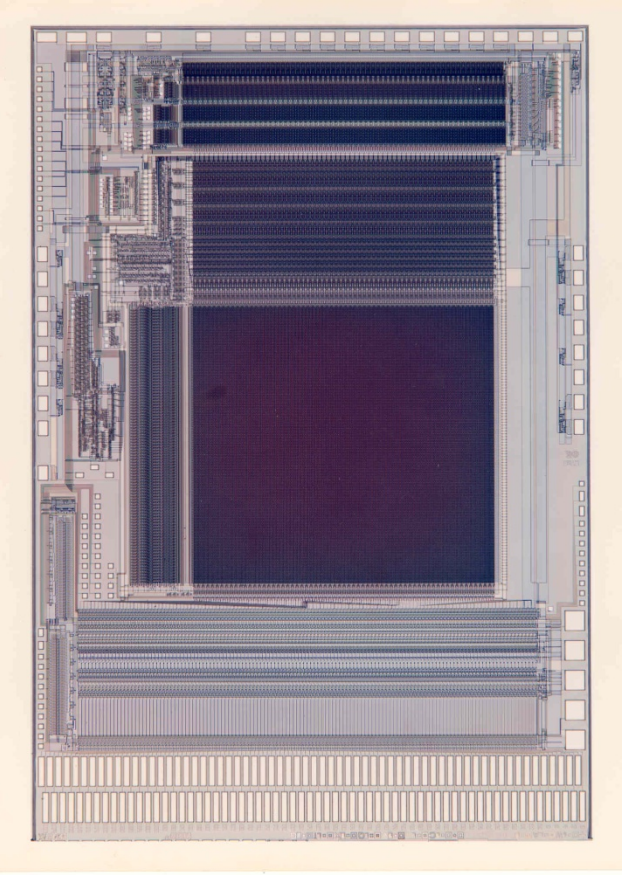


BaBar Silicon Vertex Tracker at the Stanford Linear Accelerator Center, 1999-2008: CP violation in B meson decay



Readout electronics

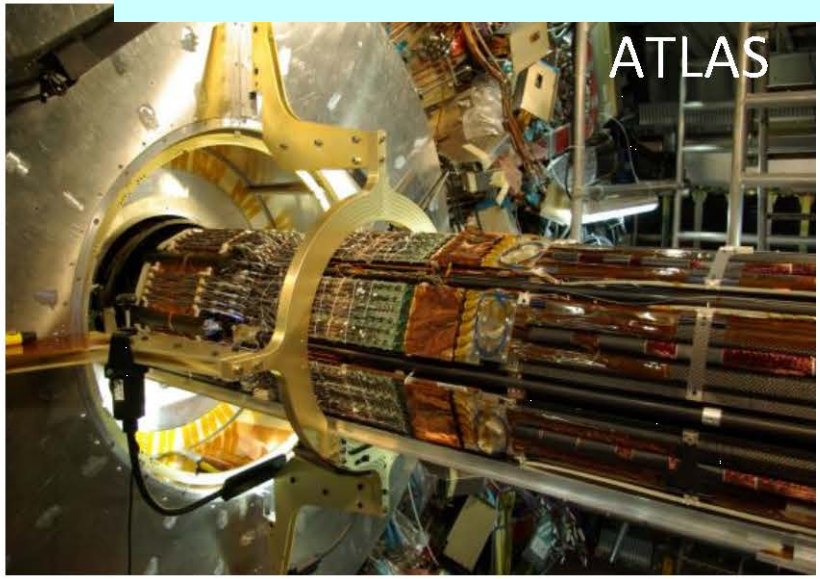
- Silicon strip detectors need miniaturization of front-end electronics
- They were the driving force for the development of integrated circuits for these applications



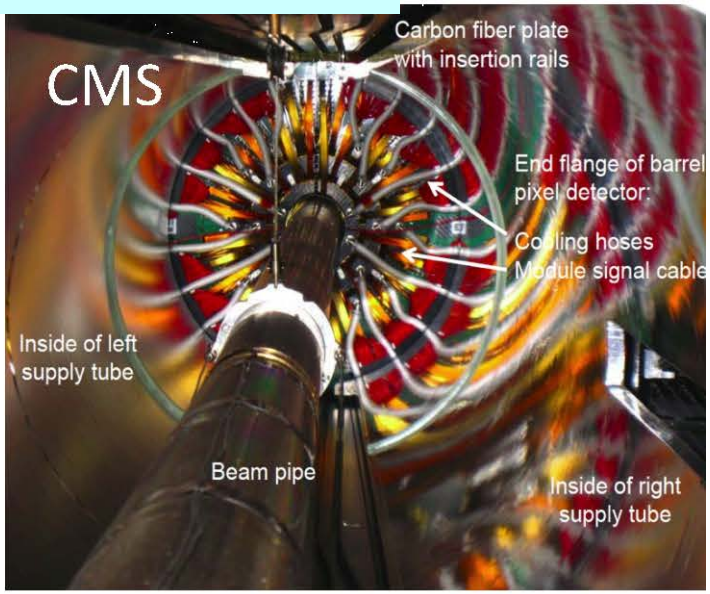
This is a mixed-signal chip, with 128-channel analog processing, A/D conversion, data storage and serial data transmission.

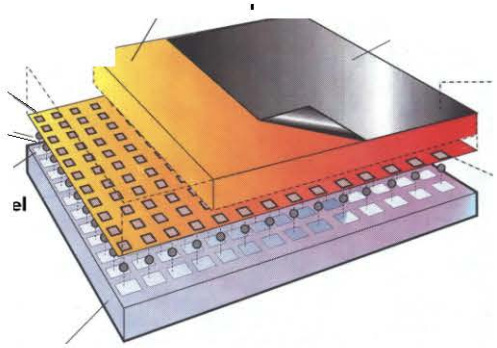
The AToM chip was fabricated in Honeywell rad-hard 0.8 μm CMOS (300k transistors) for the readout of the BaBar SVT (1998).

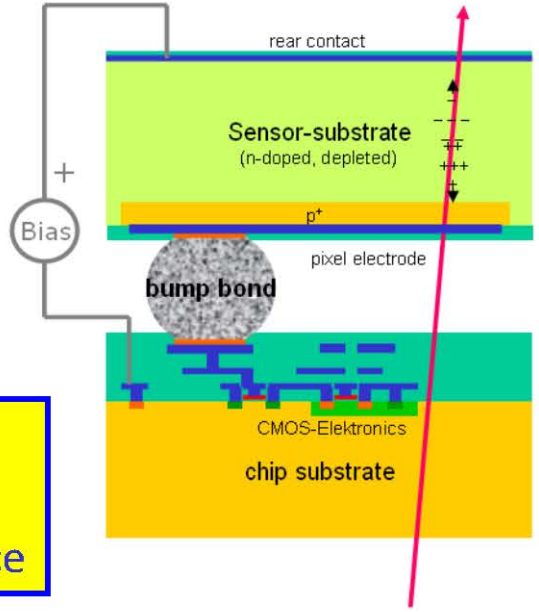
Pixel detectors at the LHC



all based on
"Hybrid Pixels"



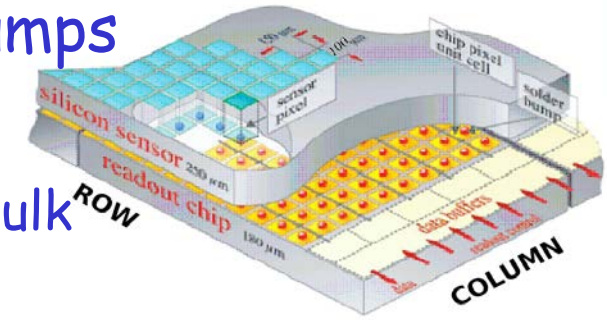




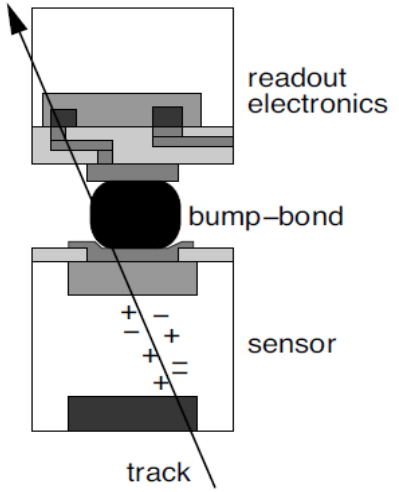
- amplification by a dedicated R/O chip
- 1-1 cell correspondence

Hybrid pixel sensors

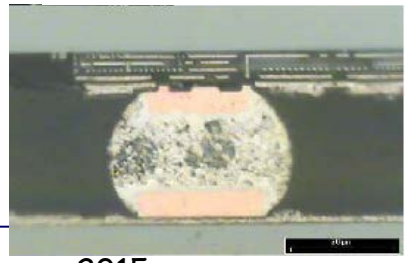
- A pixellated sensor chip is connected to a matching readout chip by an array of solder bumps



- **Sensors**
  - Particle sensitive volume is a high resistivity silicon bulk (1-10 kΩcm, 250 μm typical thickness, thinner for higher radiation hardness and less material), can be fully depleted for fast charge collection by drift
  - Typical pixel dimensions at LHC: 50 μm × 400 μm (will decrease for HL-LHC, probably to 50 μm × 50 μm)
  - Radiation-hard to 50 Mrad at LHC (1 Grad for HL-LHC)



- **Front-end chips (Nanoscale CMOS)**
  - For any event (particle hit in the sensor) provide pixel position, timing, pulse amplitude
  - Only a small number of pixels are hit in any event
  - Analog pre-amplification, discrimination, time stamping, digitization, zero suppression (sparsification)…



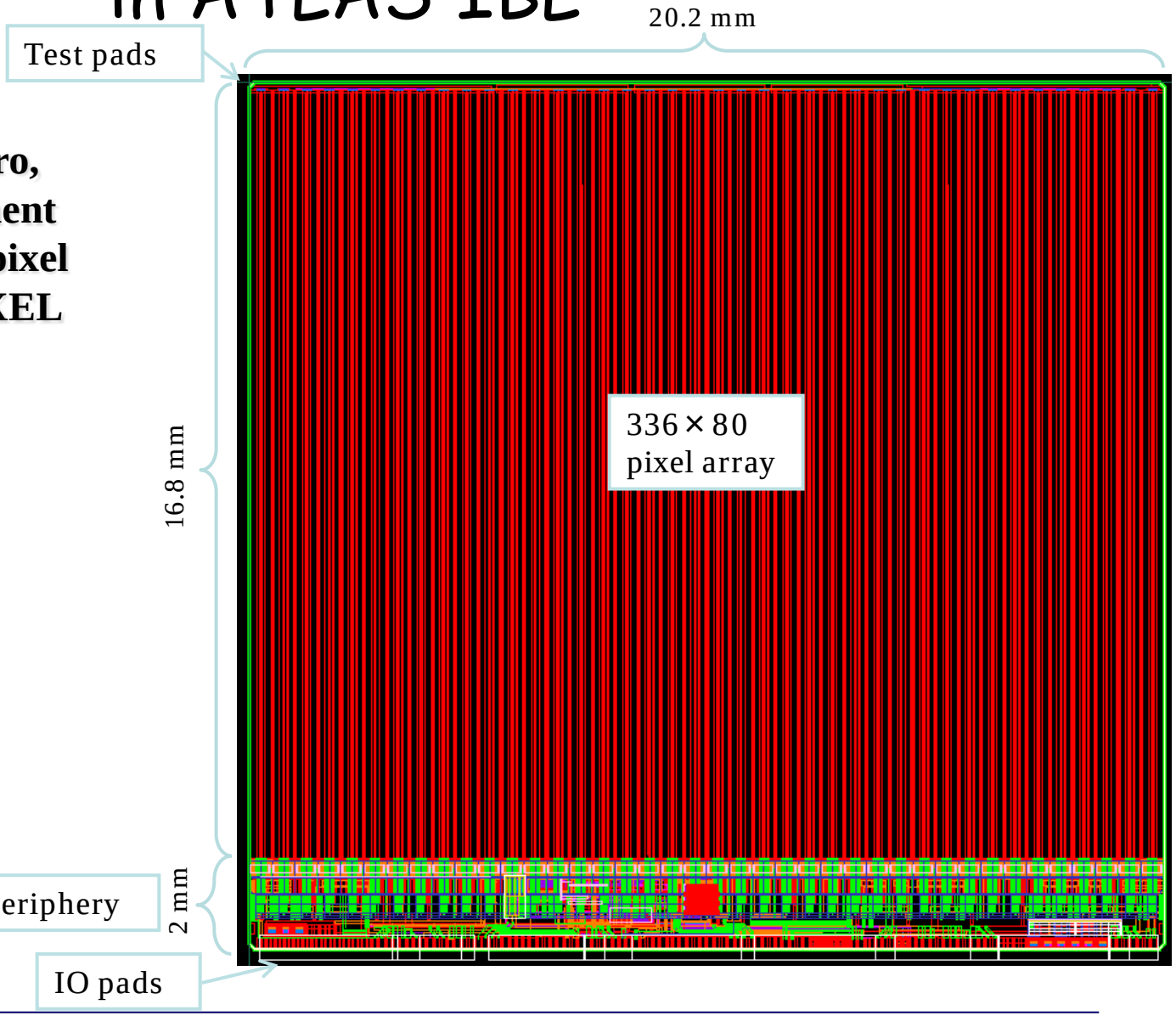
FE-I4 readout chip for pixel sensors in ATLAS IBL

(from Marlon Barbero, "FE-I4 chip development for upgraded ATLAS pixel detector at LHC", PIXEL 2010 workshop)

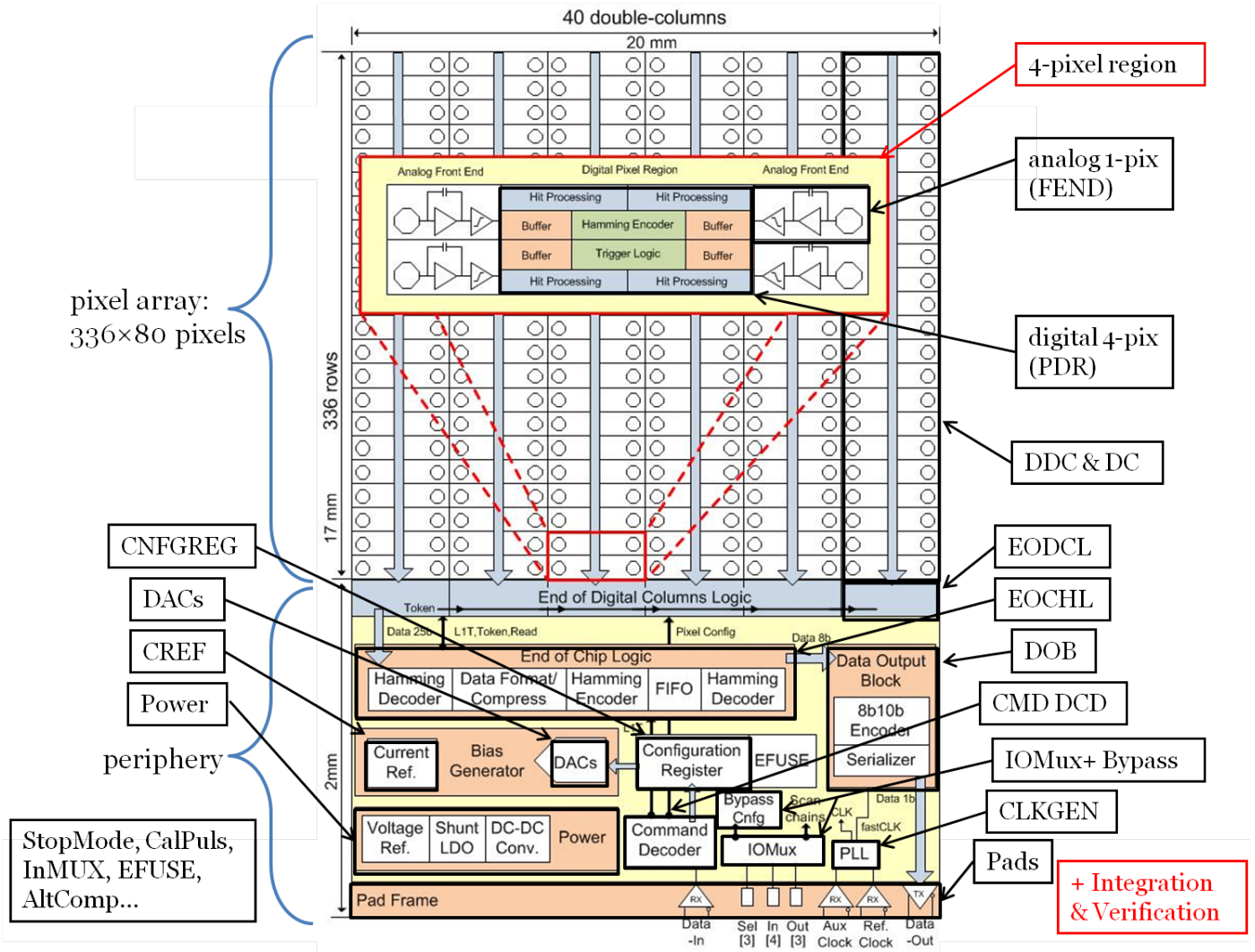
CMOS 130 nm mixed-signal chip

Digital readout of hit pixels with analog information (signal amplitude) and time stamp

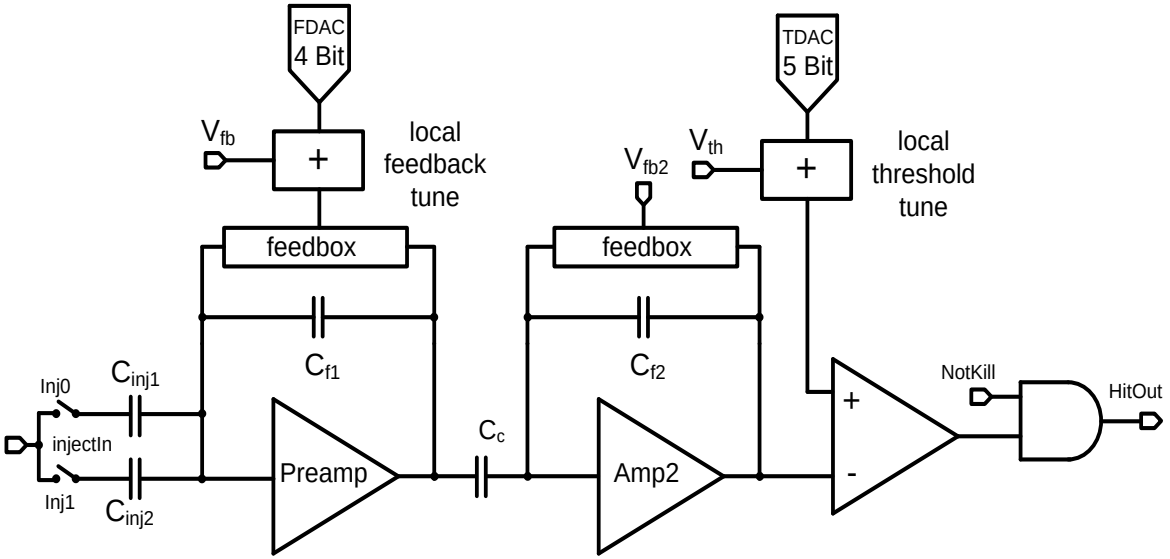
Store hits locally in region until L1T.

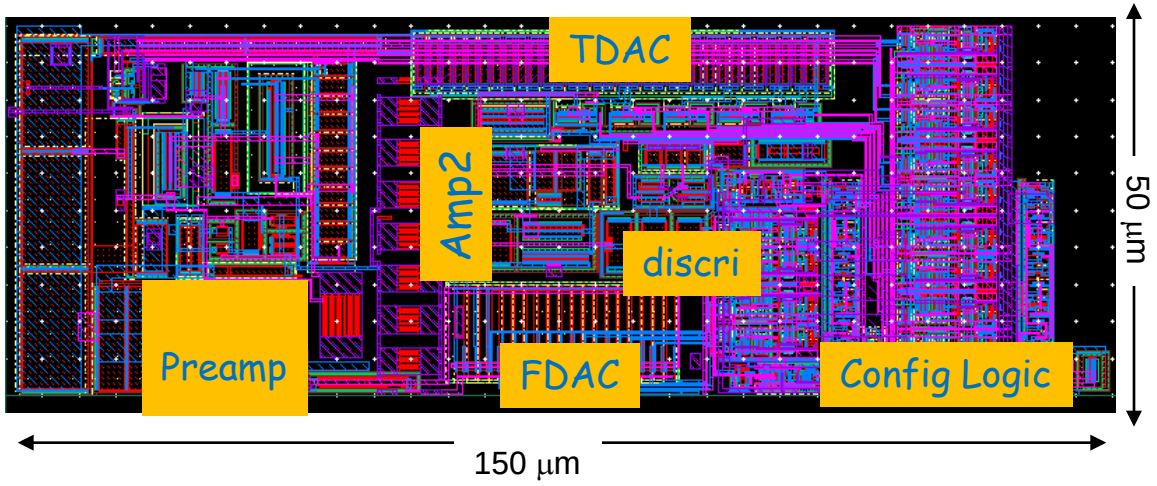


FE-I4 readout chip for pixel sensors in ATLAS IBL

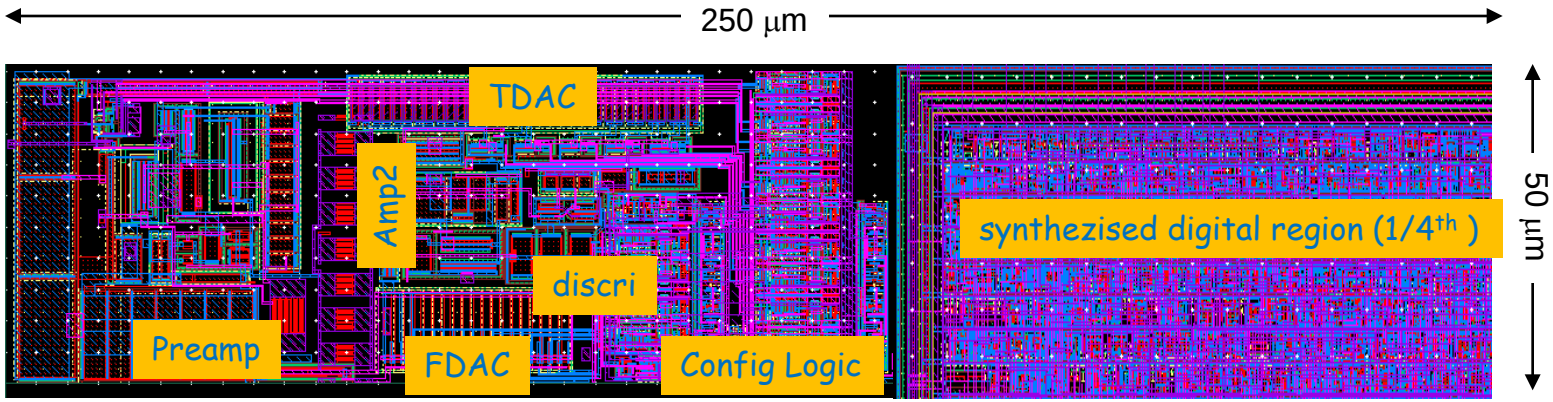


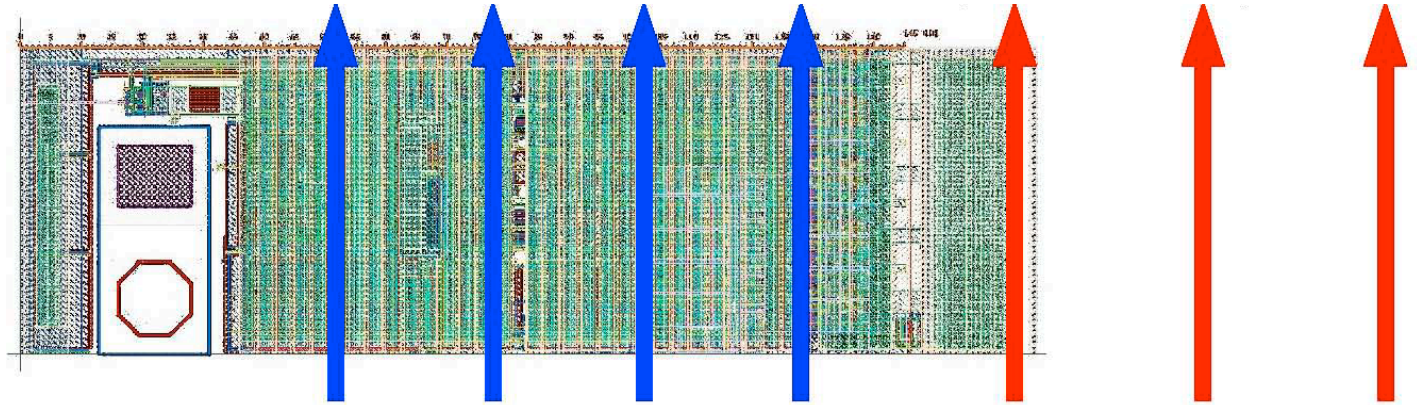
FE-I4 readout chip for pixel sensors in ATLAS IBL: the analog pixel cell





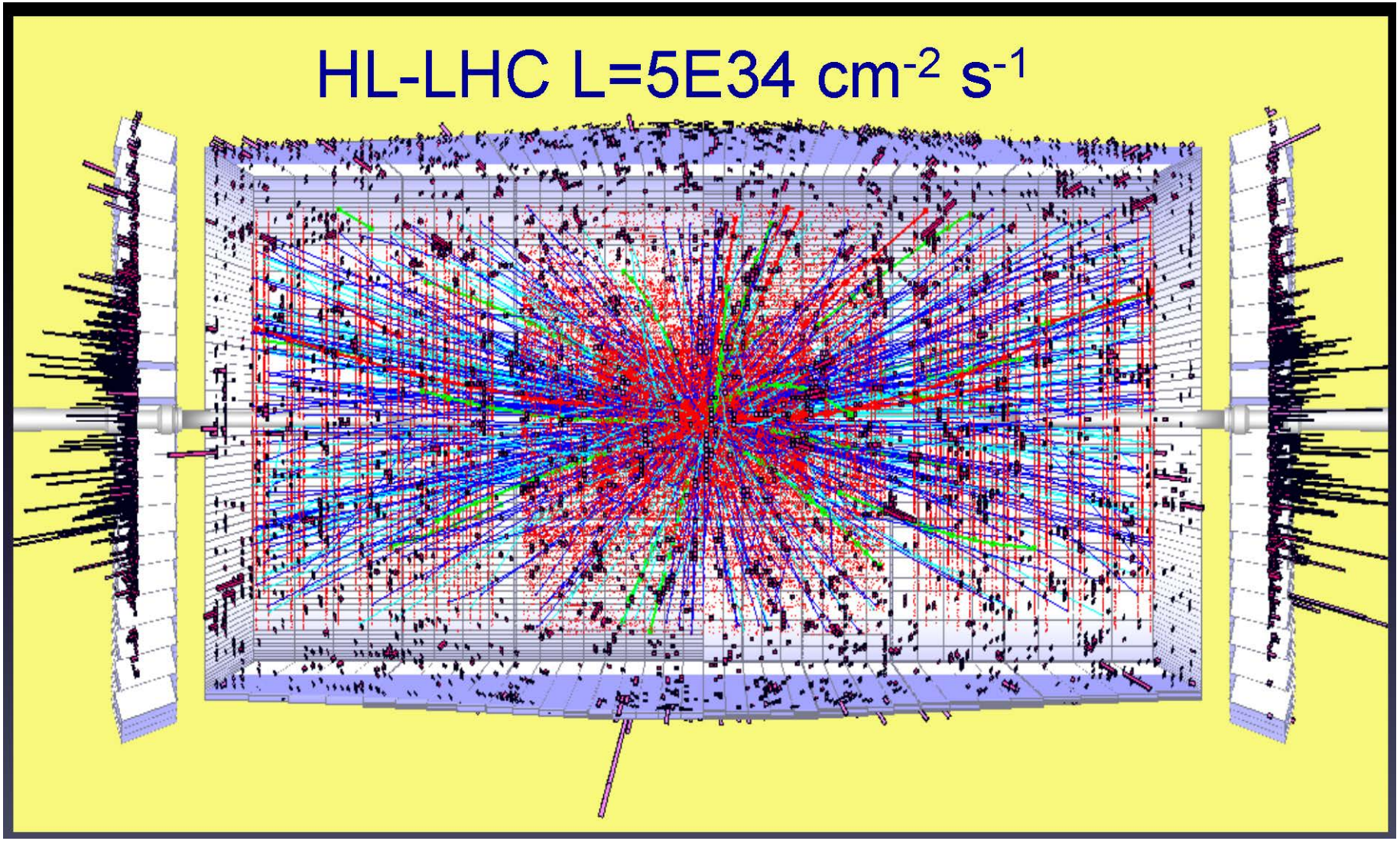
FE-I4 pixel cell layout





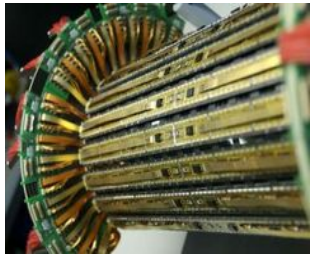
Power distribution and shield on top metals. Only vertical - no analog/digital crossing

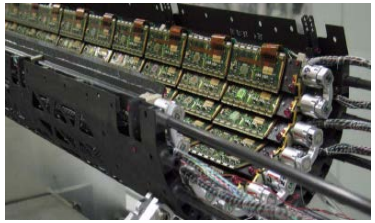
The challenge of the phase-II upgrade at the High Luminosity LHC

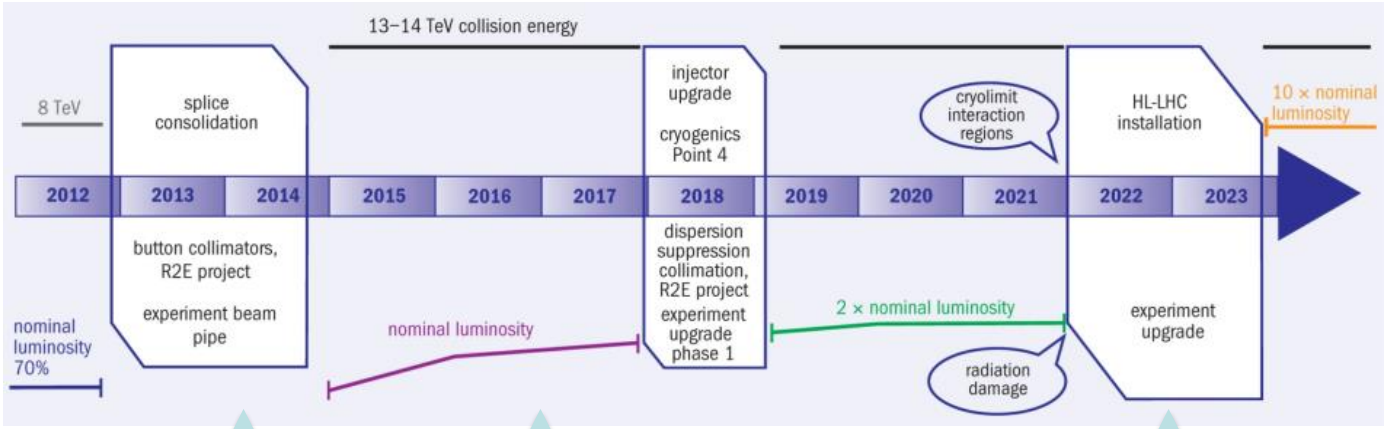


LHC pixel upgrades

- Current LHC pixel detectors have clearly demonstrated the feasibility and power of pixel detectors for tracking in high rate environments
- Phase1 upgrades: Additional pixel layer, ~4 x hit rates
  - ATLAS: Addition of Inner B Layer (IBL) with new 130nm pixel ASIC (FEI4)
  - CMS: New pixel detector with modified 250nm pixel ASIC (PSI46DIG)
- **Phase2 upgrades**: ~16 x hit rates, ~4 x better resolution, 10 x trigger rates, 16 x radiation tolerance, Increased forward coverage, less material, , ,
  - Installation: ~ 2022
  - Relies fully on significantly improved performance from next generation pixel chips.







Advanced pixel detectors and readout microelectronics

Particle tracking at LHC- Phase II:

- Very high hit rates (**1-2 GHz/cm^2**), need of an intelligent pixel-level data processing
- Very high radiation levels (**1 Grad** Total Ionizing Dose, 10^{16} neutrons/cm^2)
- Small pixel cells to increase resolution and reduce occupancy (~50x50μm^2 or 25x100 μm^2)

➔ Large chips: > 2cm x 2cm, **½ - 1 Billion transistors**

X-ray imaging at free electron laser facilities:

- Reduction of pixel size (100x100 μm^2 or even less), presently limited by the need of complex electronic functions in the pixel cell
- Larger memory capacity to store more images (at XFEL, ideally, 2700 frames at 4.5 MHz every 100 ms)
- Advanced pixel-level processing (1 – 10000 photons dynamic range, 10-bit ADC, 5 MHz operation)

Front-end electronics and nanoscale MOSFETs: CMOS 65 nm and the Low Power flavor

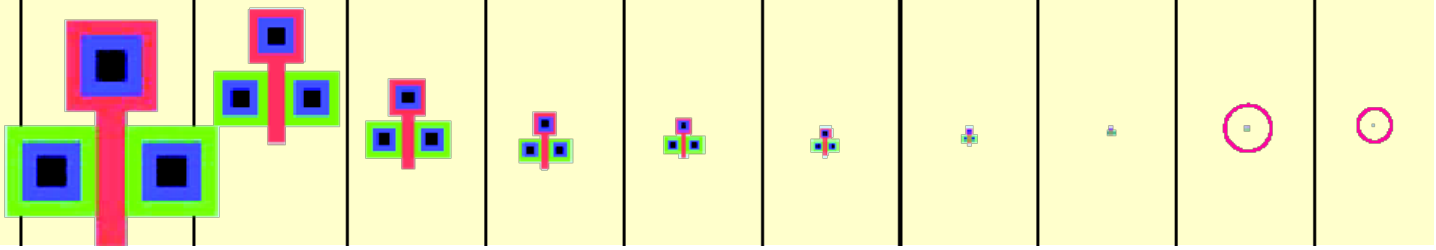
- After the 250 nm (LHC) and the 130 nm node (LHC upgrades, XFEL,…), our community appears to be very interested in the 65 nm CMOS generation: several prototypes have been already fabricated and tested.

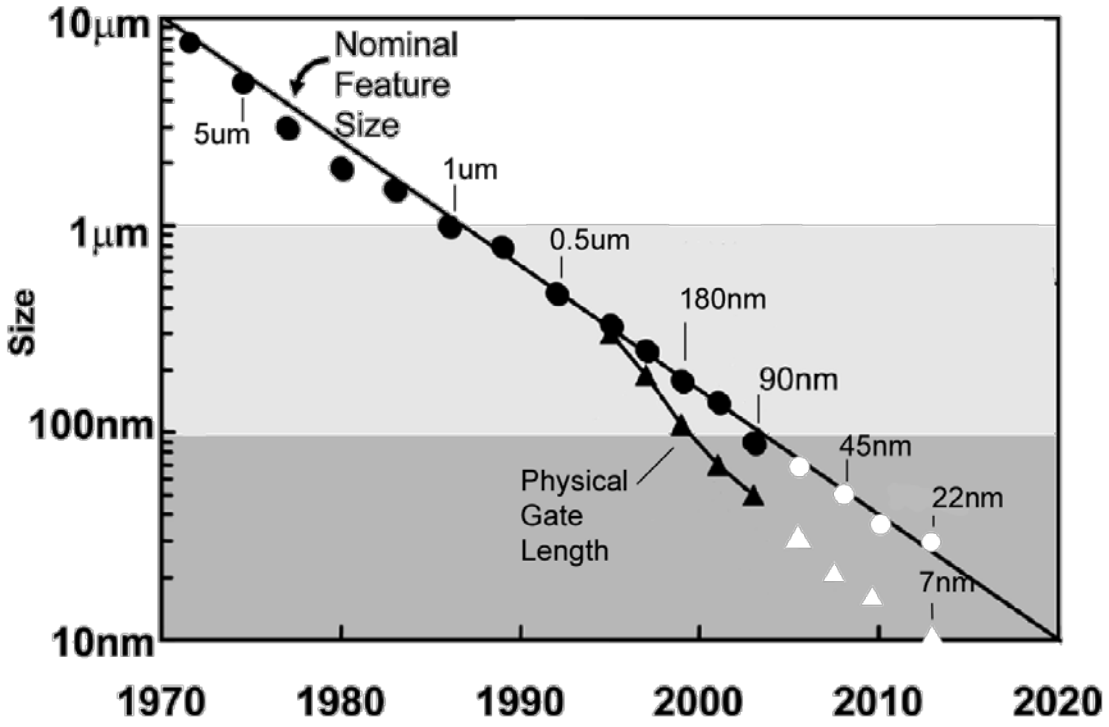
- Among the wide choice of options of this technology, the Low Power flavor is less aggressive than other variants (thicker gate oxide, smaller gate current, higher voltage), and is more attractive for mixed-signal chips where analog performance is an essential feature.

- **65nm LP (Low Power) transistors (V_{DD} = 1.2 V)** are optimized for a reduced leakage (different level of gate oxide nitridation with respect to other flavours, different silicon stress,…).

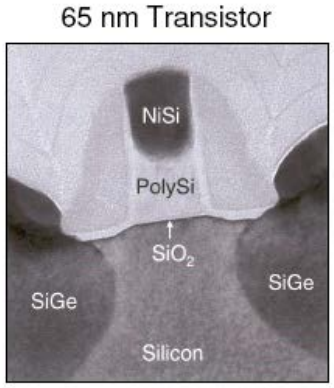
- Wireless and consumer based market rely on the use of LP transistors for mobile devices (low standby power), whereas high speed applications such as microprocessors have used dedicated HP (High Performance) technology.

Industry Scaling Roadmap for CMOS

Feature Size [nm]	2000	1200	800	500	350	250	130	65	35	20
Minimum NMOS										



Nanoscale MOSFETs



Lewyn et al, "Analog circuit design in nanoscale CMOS technologies" , Proc. IEEE, Vol. 97, no. 10, Oct. 2009.

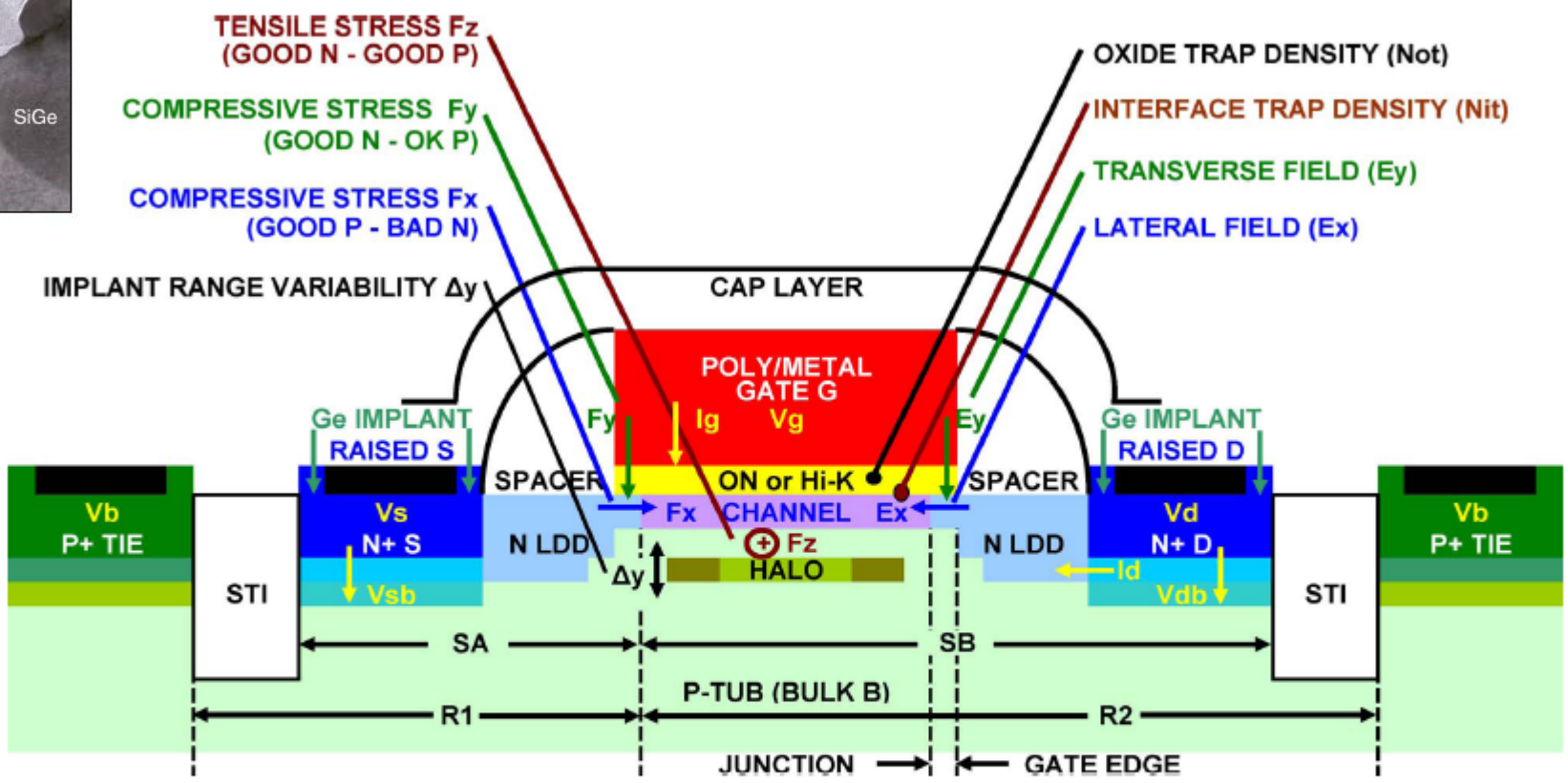


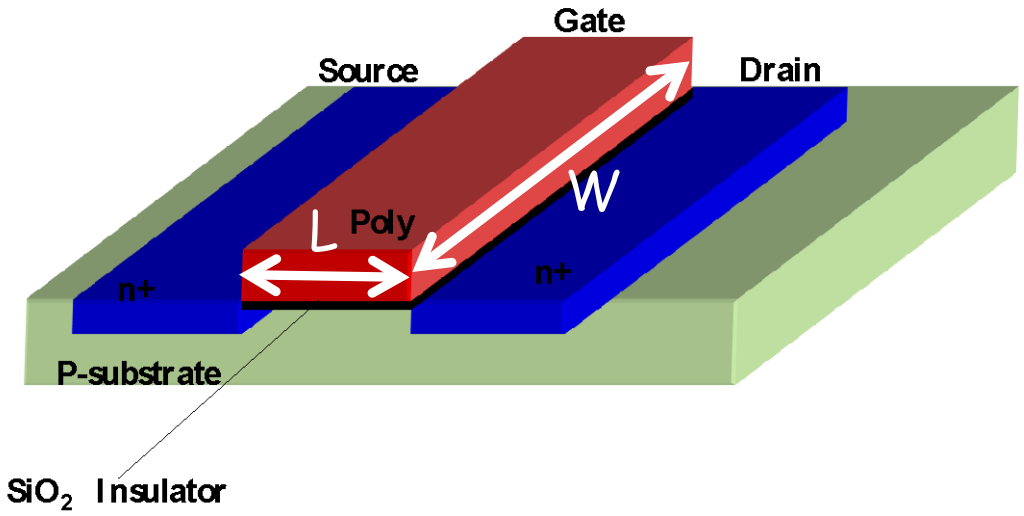
Fig. 2. *NMOS cross-section. In addition to stress from cap layers and Ge raised source-drain (S-D) implants, device dimensions such as distance from source-channel boundary to nearby STI (SA and SB), proximity and regularity of overlying metal patterns, and short distances to other device patterns within the local (< 2 μm) stress field induce transverse (F_y) and lateral (F_x and F_z) stress components, which affect threshold and mobility. Increasing the distance to P+ ties increases local tub (bulk) resistance components R1 and R2, which isolate the device MOS model substrate node from the device subcircuit symbol V_b node and degrade HF performance. Hot carrier reliability stress is dependent on the sum of transverse and lateral fields E_y and E_x. These fields are increased near the drain by increasing source to bulk (V_{sb}) and drain (V_d) to gate (V_g) or source (V_s) voltages in various combinations. As hot carrier stress increases, damage to channel from interface trap density (N_{it}) affects threshold and mobility, while gate oxynitirde (ON) or high-dielectric-constant (Hi-K) insulator trap density (N_{ot}) affects threshold and gate leakage.*

CMOS scaling: why?

- Industrial microelectronic technologies are today well beyond the 100 nm frontier, bringing CMOS into the nanoscale world
- Digital performances (speed, density, power dissipation) are driving the evolution of CMOS technologies towards a continuous shrinking of physical feature sizes.
- **Analog performance** remains essential for the processing of signals delivered by semiconductor detectors.
- Front-end electronics may benefit from scaling in terms of functional density (small pitch pixels) and digital performance. Analog design is a challenge (reduced supply voltage and dynamic range, statistical doping effects, ……)
- For a full integration of analog and digital circuits in the most modern semiconductor technologies, design advances are needed to exploit the full potential: digital signal processing may be used to overcome analog limitations, analog circuits may be used to monitor the performance of digital circuits and their power consumption.

CMOS scaling: why?

A. Marchioro, INFIERI 2013



W, L = gate width and length

- Smaller dimensions give:
  - Shorter transit time
    - Faster circuit
  - Lower parasitic capacitance
    - Faster circuit
    - Lower power ($P \propto C V^2$)
  - More transistors/unit area
    - More functionality can be built into chips

CMOS scaling: how

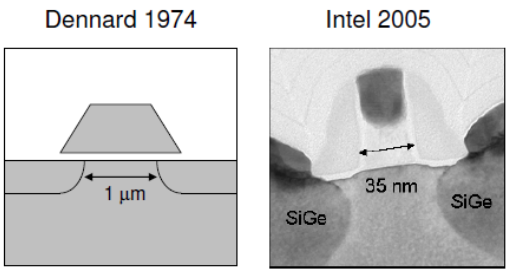
- **Shrinking of gate length** leads to an increase in speed and circuit density. To avoid short-channel effects, drain and source depletion regions are made correspondingly smaller by **increasing substrate doping concentration** and decreasing reverse bias (**reduction of the supply voltage**)

- Increasing substrate doping increases the device threshold voltage: this is overcome by **decreasing the gate oxide thickness**.

- Classical scaling ended because of gate oxide thickness limits: in very thin oxides, direct tunneling of carriers leads to a large gate leakage current.

as first described in 1974

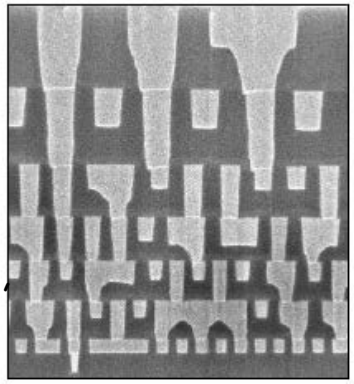
30 Years of MOSFET Scaling



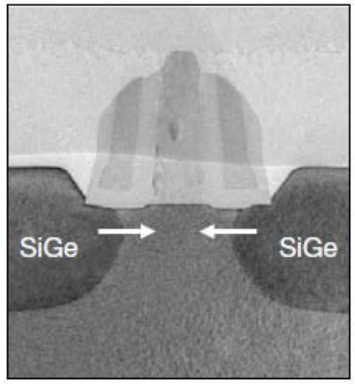
	Dennard 1974	Intel 2005
Gate Length:	1.0 µm	35 nm
Gate Oxide Thickness:	35 nm	1.2 nm
Operating Voltage:	4.0 V	1.2 V

The New Era of Device Scaling

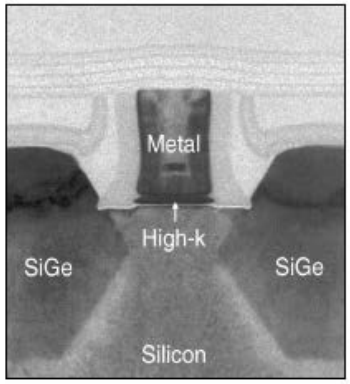
M. Bohr (Intel), "The new era of scaling in a SoC world", ISSCC 2009.



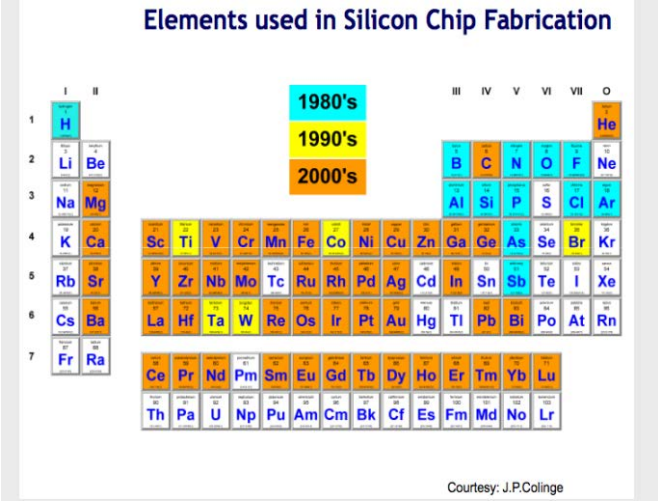
Copper + Low-k



Strained Silicon



High-k + Metal Gate



Modern CMOS scaling is as much about material and structure innovation as dimensional scaling

- Mechanical stress (compressive or tensile strain) is introduced in the silicon channel to enhance carrier mobility and drive current.
- Gate dielectric is made thicker (still increasing gate capacitance) by using materials with higher dielectric constant than SiO_2.

RD53: an ATLAS-CMS-LCD collaboration

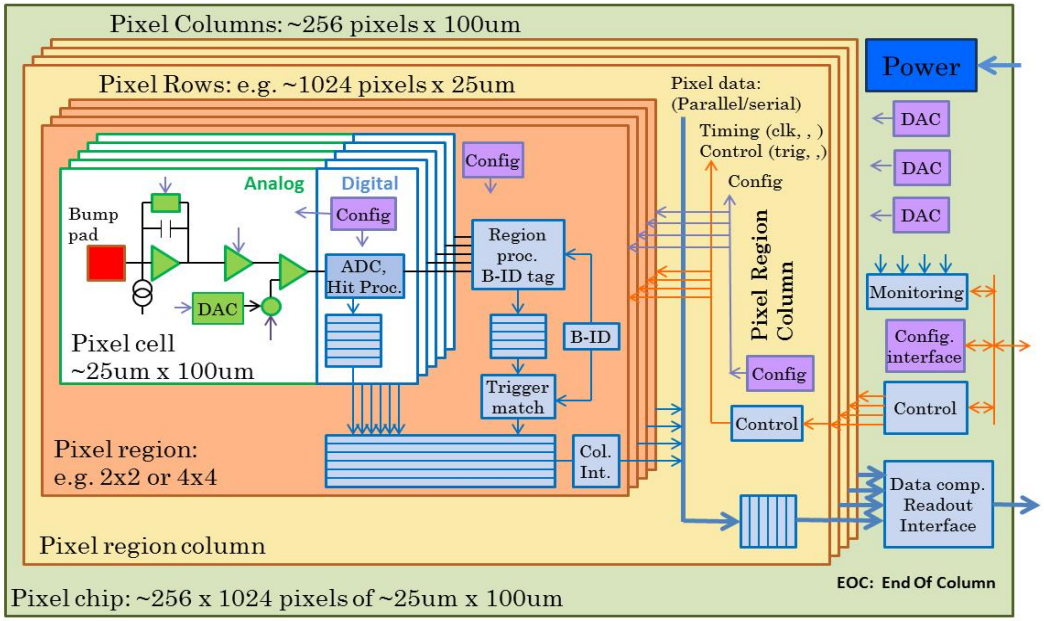
- RD53 was organized to tackle the extreme and diverse challenges associated with the design of pixel readout chips in nanoscale CMOS for the innermost layers of particle trackers at future high energy physics experiments (LHC – phase II upgrade of ATLAS and CMS, CLIC)

- It relies fully on significantly improved performance from next generation pixel chips

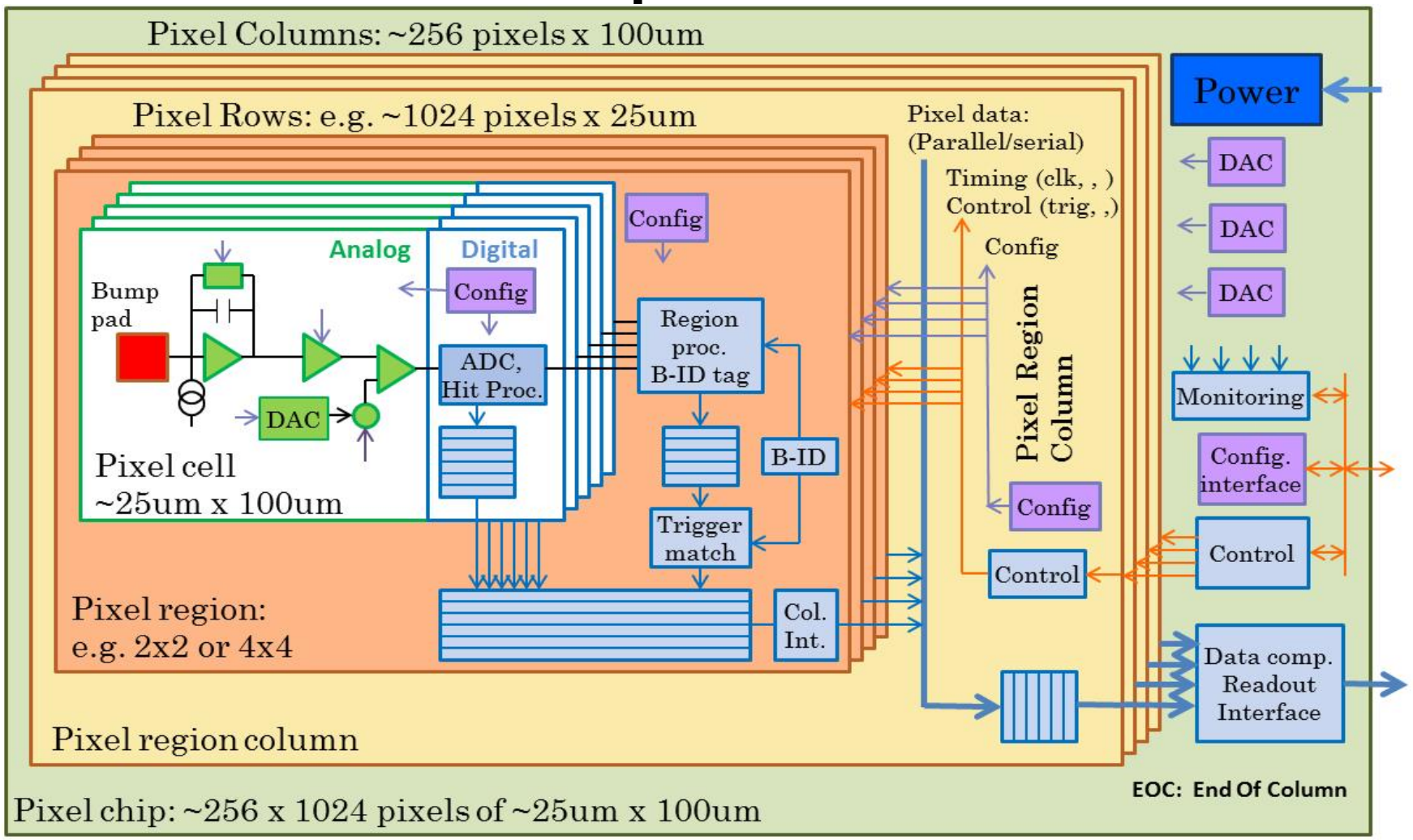
65 nm CMOS is the candidate technology to address the requirements of these applications.
It has to be fully studied and qualified in view of the design of these chips.

- ATLAS: CERN, Bonn, CPPM, LBNL, LPNHE Paris, Milano, NIKHEF, New Mexico, Prague, RAL, UC Santa Cruz.
- CMS: Bari, Bergamo-Pavia, CERN, Fermilab, Padova, Perugia, Pisa, PSI, RAL, Torino.
- Collaborators: ~100, ~50% chip designers

A mixed-signal pixel readout chip for the phase II upgrade of LHC in 65 nm CMOS



RD53 chip architecture



- 95% digital (as for FEI4)
- Charge digitization (TOT or ADC)
- ~256k pixel channels per chip
- Pixel regions with buffering
- Data compression in End Of Column
- Chip size: >20 x 20 mm^2

Analog front-end design for detector charge measurements

Radiation detectors

A measure of the information appears in the form of an electric charge, induced on a set of two electrodes, for which ultimately only one parameter (capacitance) is important.

Front-end electronics

amplifying device

(charge-sensitive preamplifier)

filtering, signal shaping

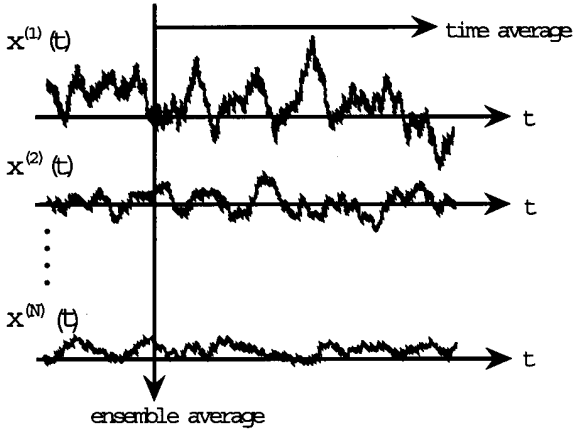
optimize the measurement of a desired quantity such as signal amplitude as a measure of the energy loss of the particle

Effect of electronic noise on charge measurements

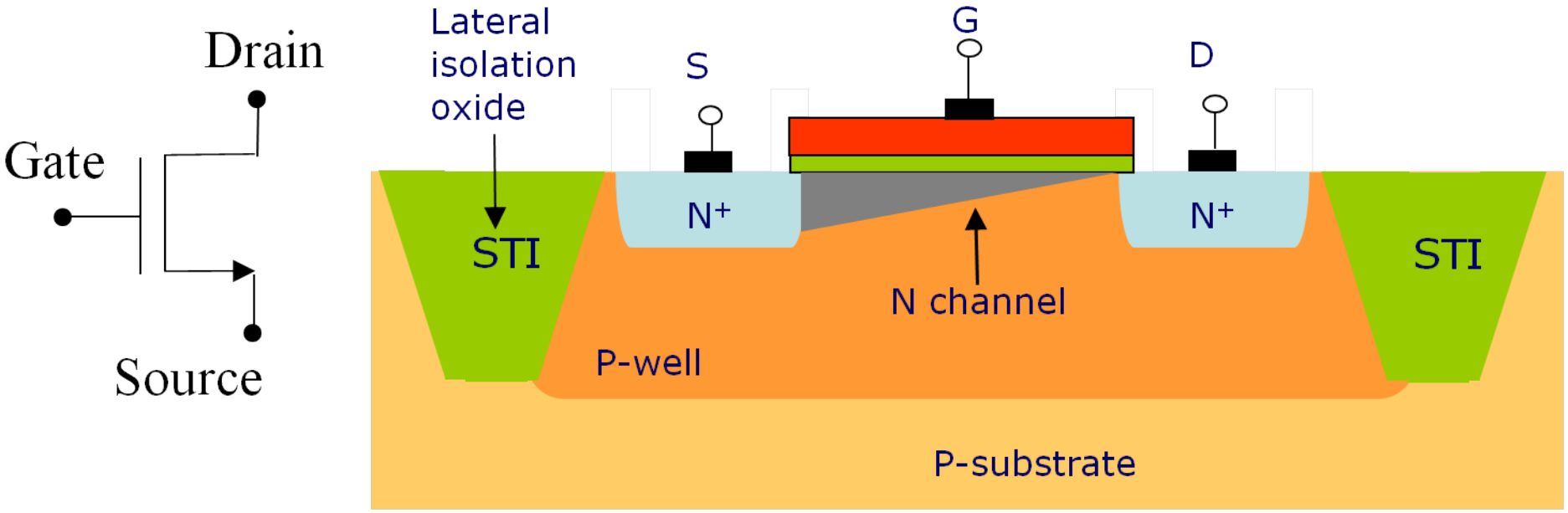
Inherent to the conduction of current in an amplifying device is a random component, depending on the principle of operation of the device.

This random component (noise) associated with amplification gives an uncertainty in the measurement of the charge delivered by the detector or of other parameters such as the position of particle incidence on the detector.

Compromises must be made in very large and complex detector systems such as modern silicon trackers.

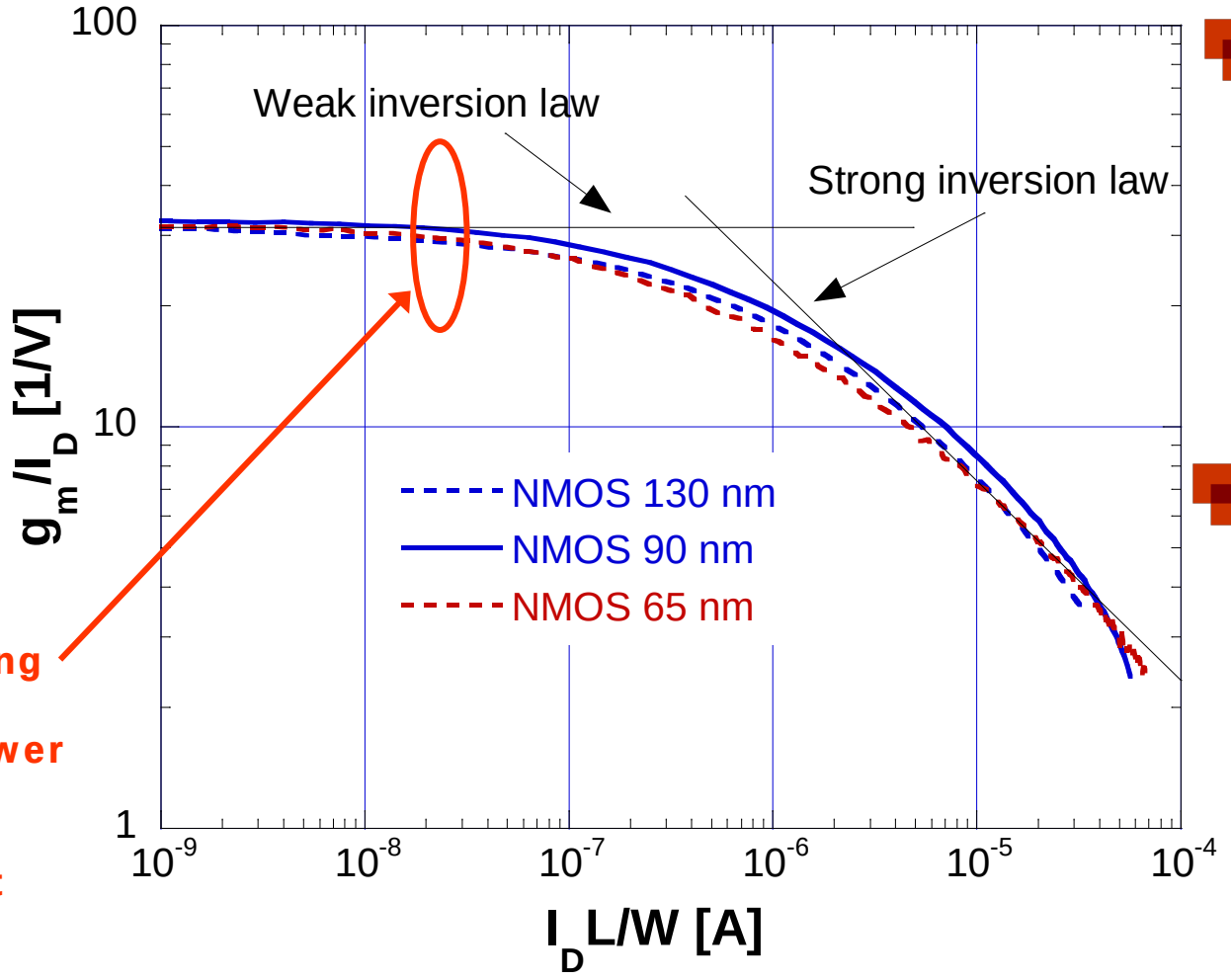


Basic element of modern electronics: the MOSFET



- Three-terminal device: an electrode controls the current flow between two electrodes at the end of a conductive channel.
- The transconductance $g_m = dI_D/dV_{GS}$ is the ratio of change in the output (drain) current and of the change in the potential of the control (gate) electrode

MOSFET essential parameters: the transconductance g_m



Typical operating point in low-power pixel sensor readout

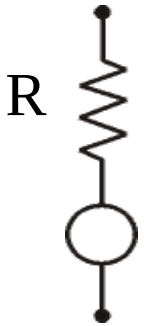
- Under reasonable power dissipation constraints, devices in nanoscale CMOS operate in the weak inversion region
- In weak inversion:

$$g_m = \frac{I_D}{nV_T}$$

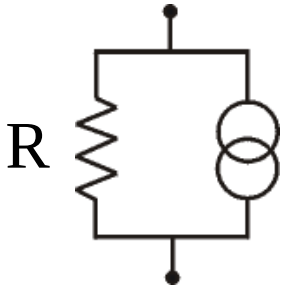
(n =1.2 in 100-nm scale CMOS)

MOSFET essential parameters: channel thermal noise

- Thermal noise arises from random velocity fluctuations of charge carriers due to thermal excitation. The spectral density (noise power per unit frequency bandwidth) is white, i.e. frequency independent. In a resistor, this can be modelled in terms of a fluctuating voltage across the resistor, or of a fluctuating current through the resistor.



$$\frac{\overline{de_R^2}}{df} = 4kTR$$

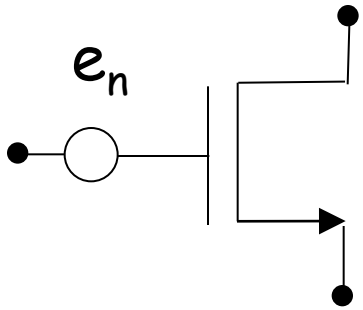


$$\frac{\overline{di_R^2}}{df} = \frac{4kT}{R}$$

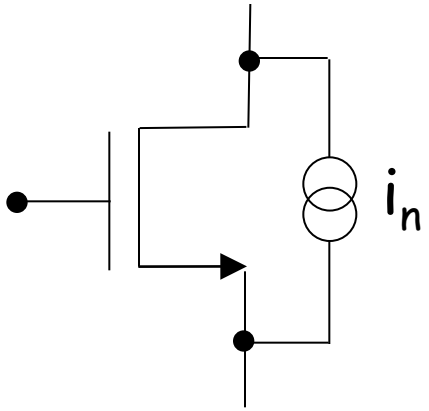
- The channel of a MOSFET can be treated as a variable conductance. Thermal noise is generated by random fluctuations of charge carriers in the channel and can be expressed in terms of the transconductance g_m.

MOSFET essential parameters: channel thermal noise

- Thermal noise in a MOSFET can be represented by a current generator in parallel to the device, or by a voltage generator in series with the gate (fluctuation of the drain current can be seen as due to fluctuations of the gate voltage).



$$\frac{\overline{de_n^2}}{df} = \frac{4kT\,\Gamma}{g_m}$$



$$\frac{\overline{di_n^2}}{df} = 4kT\,\Gamma g_m$$

k = Boltzmann's constant, T = absolute temperature

Γ = coefficient ($\cong$ 1) dependent on device operating region, short channel effects…

Statement of the problem of front-end electronics

Measurement of a charge delivered by a capacitive source with the best possible accuracy compatible with noise intrinsically present in the amplifying system, and with the constraints set by the different applications.

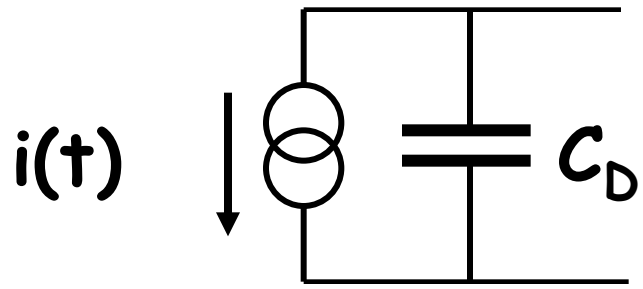
(noise - power - speed)

The discussion will be based on the nuclear electronics noise theory.

(basic equations recalled for discussion purposes)

Acquiring the signal from the sensor: the charge-sensitive preamplifier

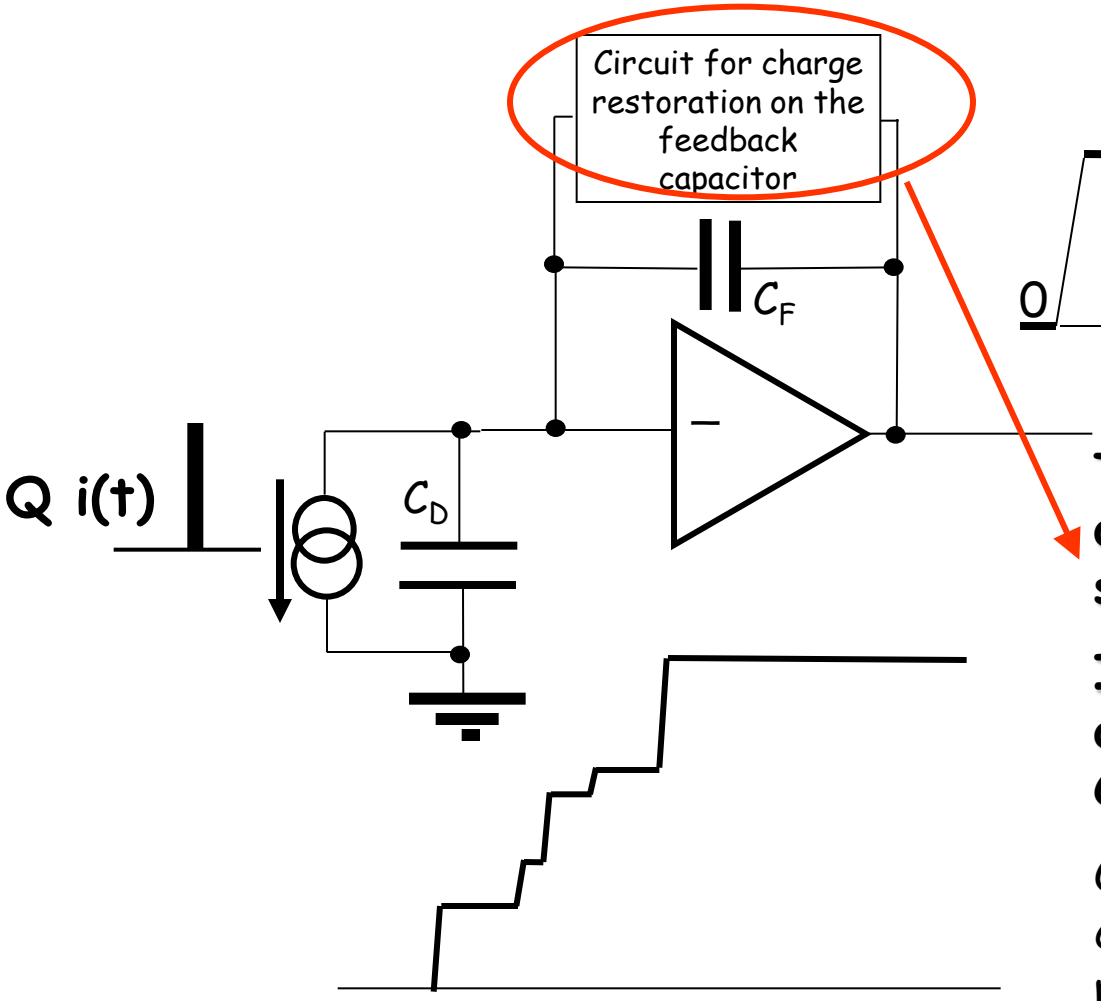
- The detector signal is a current pulse i(t) of short duration
- The physical quantity of interest is the deposited energy, so one has to integrate the sensor signal



$$E_S \div Q_S = \int i(t)dt$$

- The detector capacitance C_D is dependent on geometry (e.g. strip length or pixel size), biasing conditions (full or partial depletion), aging (irradiation)
- Use an integrating preamplifier (charge-sensitive preamplifier), so that charge sensitivity ("gain") is independent of sensor parameters

Acquiring the signal from the sensor: the charge-sensitive preamplifier

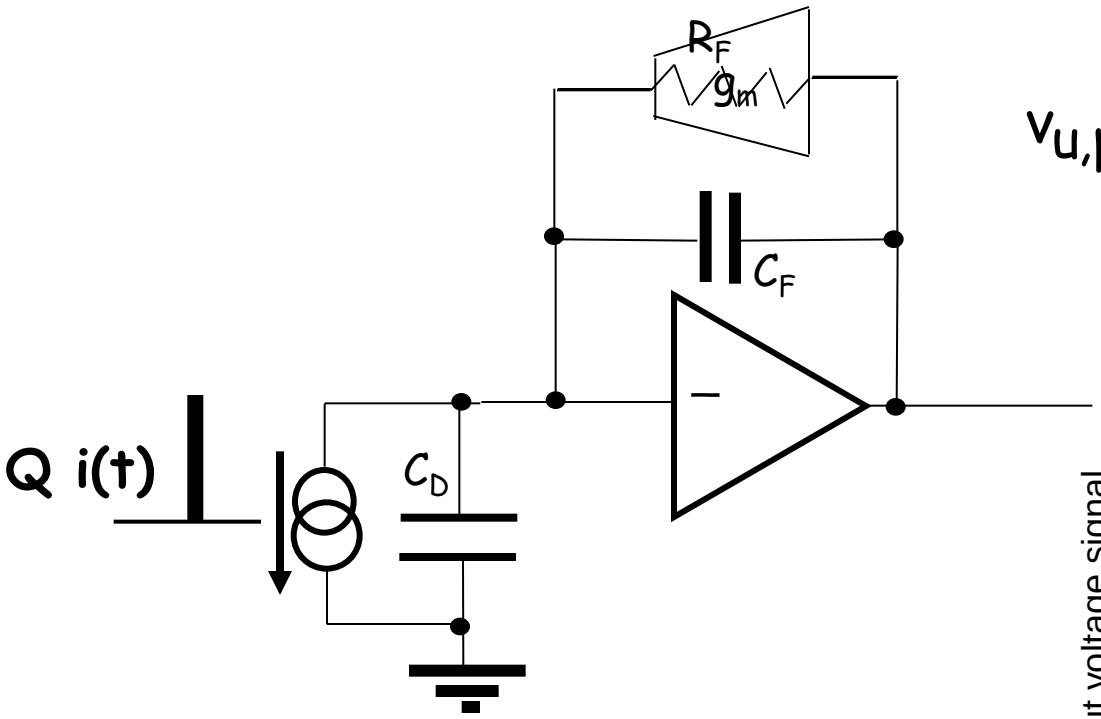


This guarantees a return to baseline of the preamplifier output, avoiding saturation.

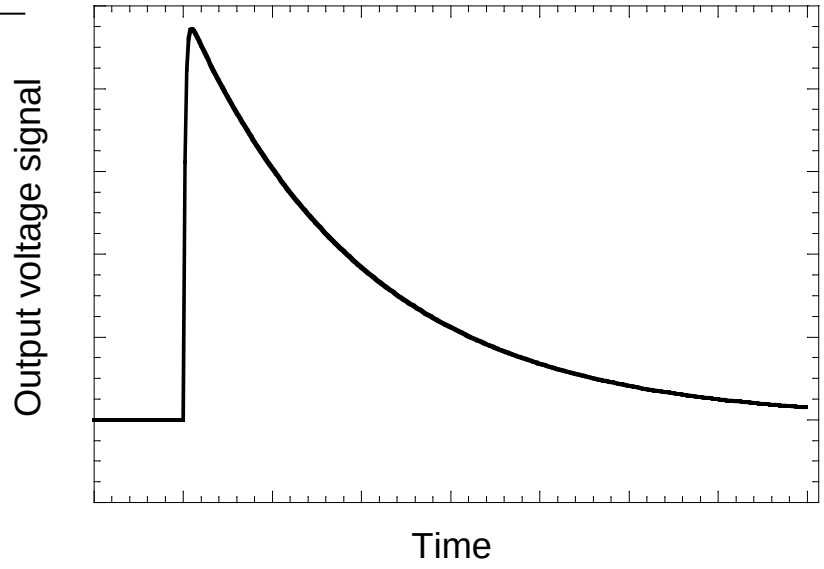
It can be achieved with a resistor R_F or, in an integrated circuit, with a CMOS circuit (transconductor).

Compensation of detector leakage current can also be performed in the preamplifier feedback (dc coupling)

Acquiring the signal from the sensor: the charge-sensitive preamplifier

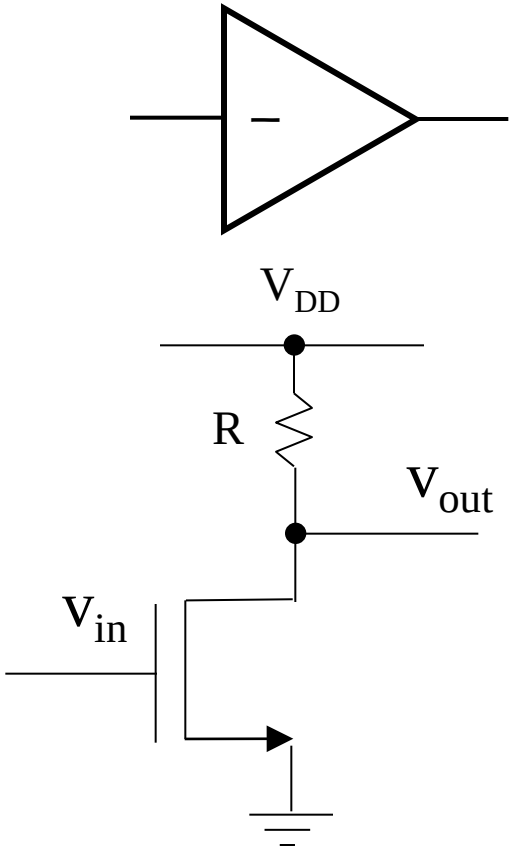


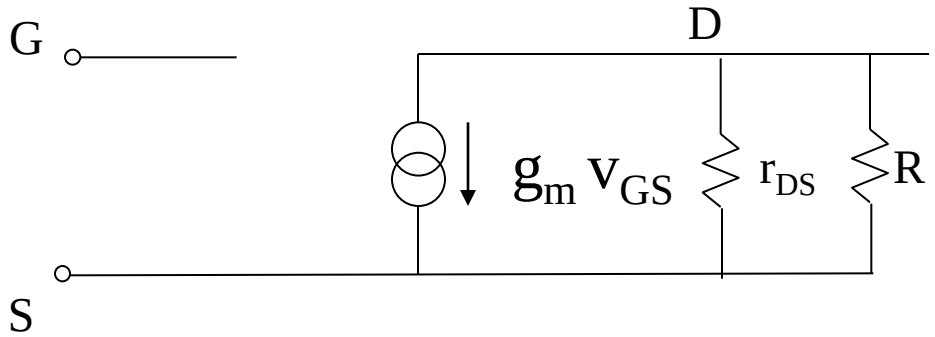
$$v_{u,pre}(t) = \frac{Q}{C_F} e^{-\frac{1}{R_F C_F} t} - \frac{g_m}{C_F} t$$



Forward gain stage: CMOS version

- The forward gain stage is an inverting amplifier which can be based on the common source configuration





$$\frac{v_{out}}{v_{in}} = -g_m\left(r_{DS} \,//\, R\right)$$

$r_{DS} \div (I_D)^{-1}$

$r_{DS} \div L$ (device gate length)

Intrinsic gain $g_m r_{DS}$

- It represents the maximum gain that can be achieved with a single transistor. It gives an indication concerning the design complexity that is necessary to implement a high gain amplifier.

$$g_m r_{DS} \propto \alpha L$$

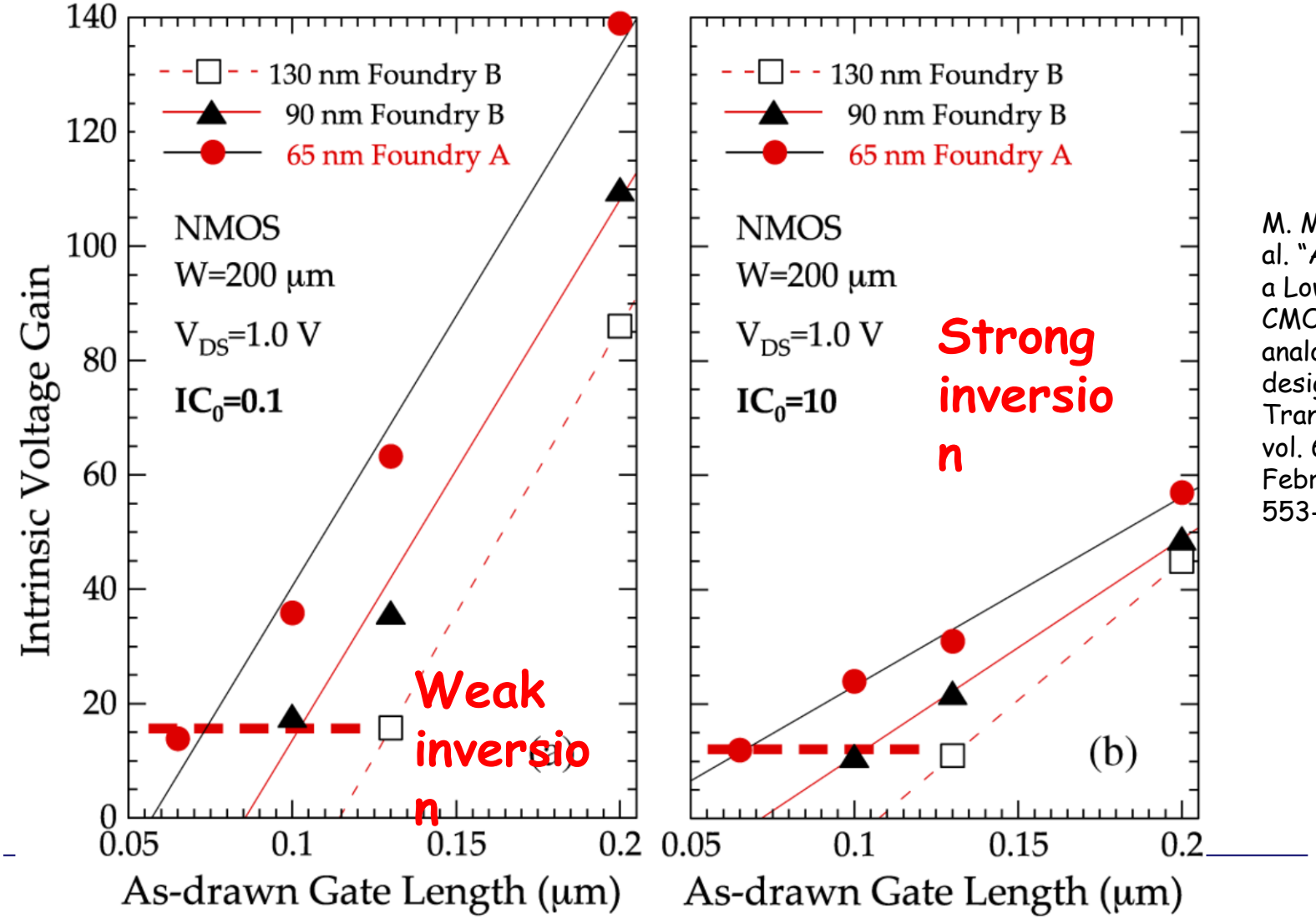
- L is the transistor gate length, α is the CMOS process scaling factor. If L_{min} scales by $1/\alpha$ in agreement with scaling rules, $g_m r_{DS}$ stays constant.

- Keeping the intrinsic gain constant with scaling (when you depart from classical scaling rules in advanced technologies) is considered as a major challenge which affects the design of analog circuits in nanoscale CMOS

- But Low Power CMOS processes do not follow aggressive scaling trends closely…

Intrinsic gain $g_m r_{DS}$ in LP 65 nm CMOS

The value of the intrinsic gain is maintained across different CMOS technology nodes from 130 nm to Low Power 65 nm if inversion conditions are the same.



M. Manghisoni, et al. "Assessment of a Low-Power 65 nm CMOS process for analog front-end design", IEEE Trans. Nucl. Sci., vol. 61, no. 1, February 2014, pp. 553-560.

Intrinsic gain from 130 nm to LP 65 nm CMOS

Inversion coefficient $$I_{C0} = \frac{I_D}{I_Z^*}\frac{L}{W}$$

I_{C0} = 1 is the boundary between weak (I_{C0} << 1) and strong inversion (I_{C0} >> 1). For I_{C0} = 1 , $I_D = I_Z^*$. I_Z^* is a characteristic parameter of a CMOS process. It has a similar value of about 0.5 μA in NMOSFETs from the general purpose 130 nm process and the LP 65 nm technology whose parameters are shown in previous plots.

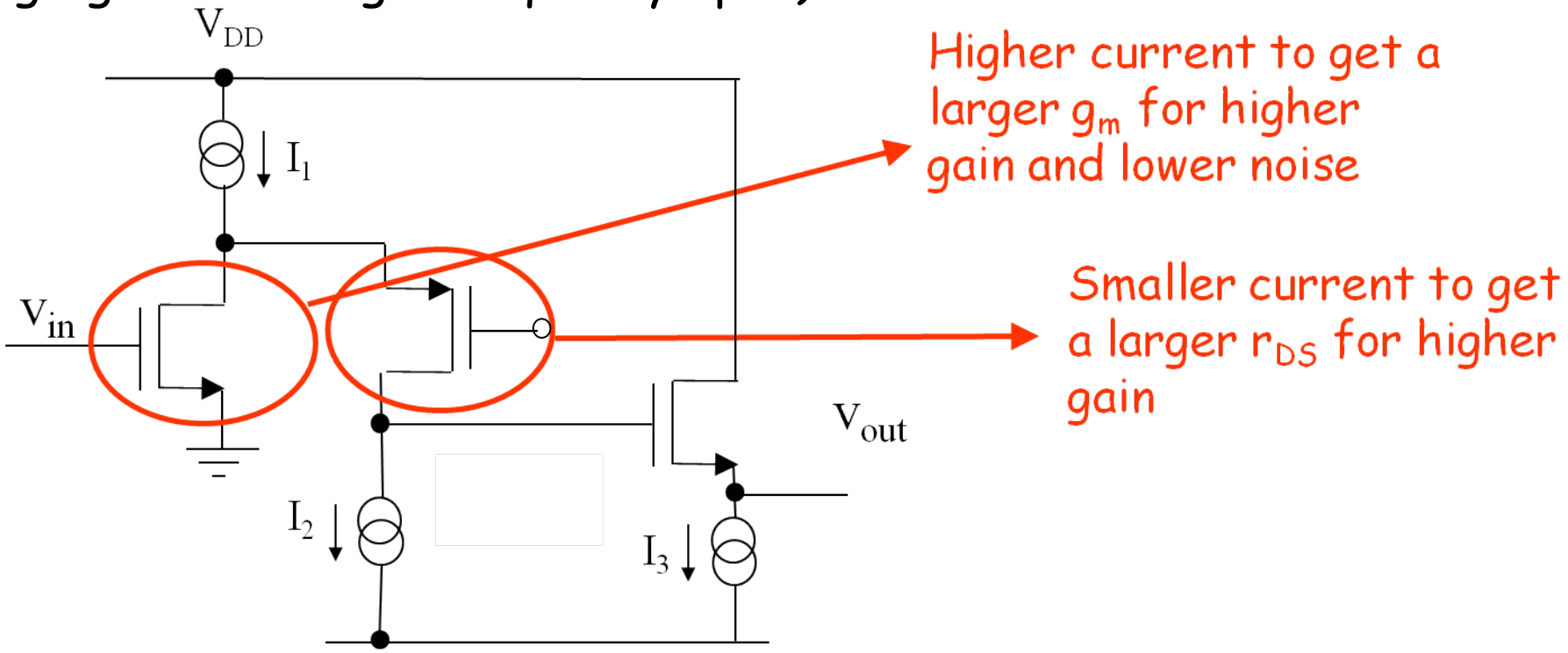
When the gate length L is reduced, the transistor stays in the same inversion conditions (equal value of I_{C0}) if the drain current I_D is correspondingly increased (at constant W).

This means that, for the minimum gate length of the process, the same value of the intrinsic gain (see previous slide) is achieved at the expense of an increased I_D. If this is not allowed because of power dissipation constraints, a value of L larger than the minimum one has to be used.

This is an example of the fact that **in analog circuits it is often impossible to take full benefit from CMOS scaling in terms of a reduced silicon area.** This is true also for other analog parameters (1/f noise, threshold dispersion,…)

Forward gain stage: CMOS version

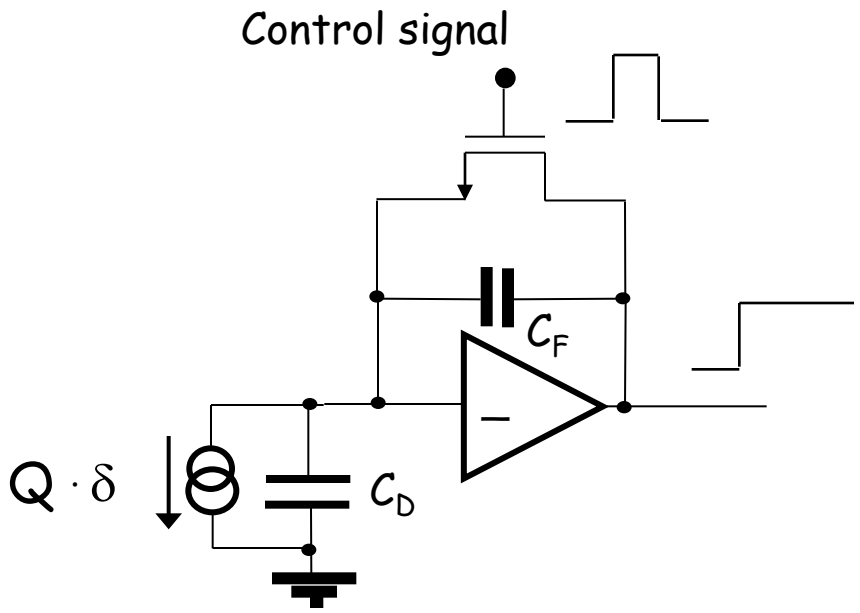
- A higher forward gain can be achieved with a folded cascode configuration. A smaller current in the cascode branch makes it possible to achieve a high output impedance.
- An output source follower can be used to reduce capacitive loading on the high impedance node and increase the frequency bandwidth (high gain in a large frequency span)



CMOS feedback network

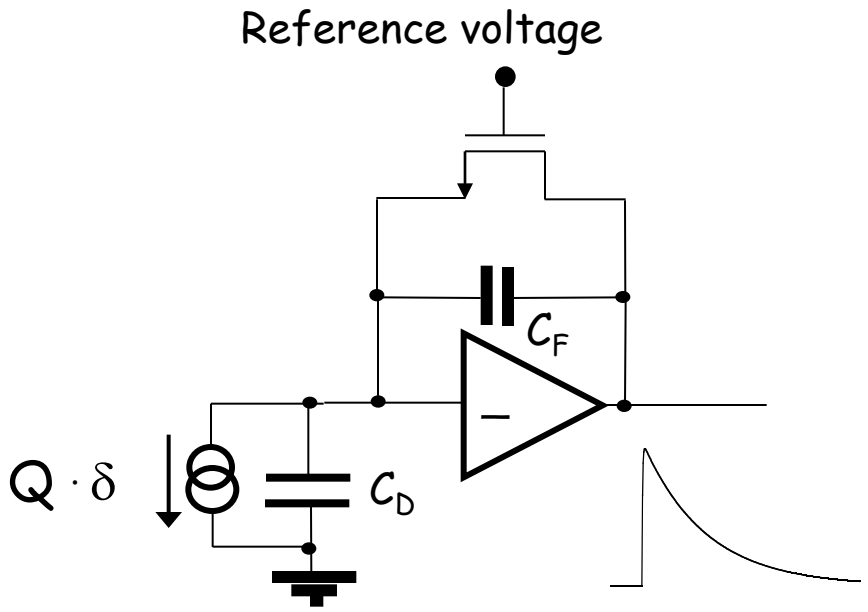
- Single feedback MOSFET

Reset switch



Can be used when you can reset the preamp at fixed times

Linear resistor

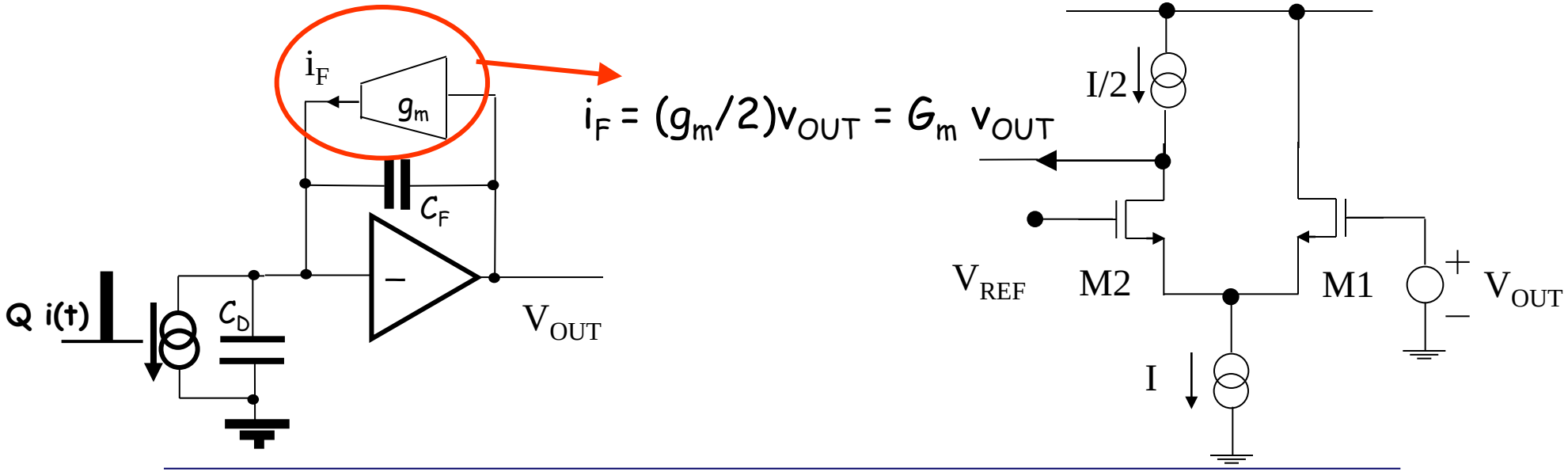


$$R_{ON} = \frac{L}{W} \frac{1}{\mu_N C_{OX} (V_{GS} - V_T)}$$

CMOS feedback network

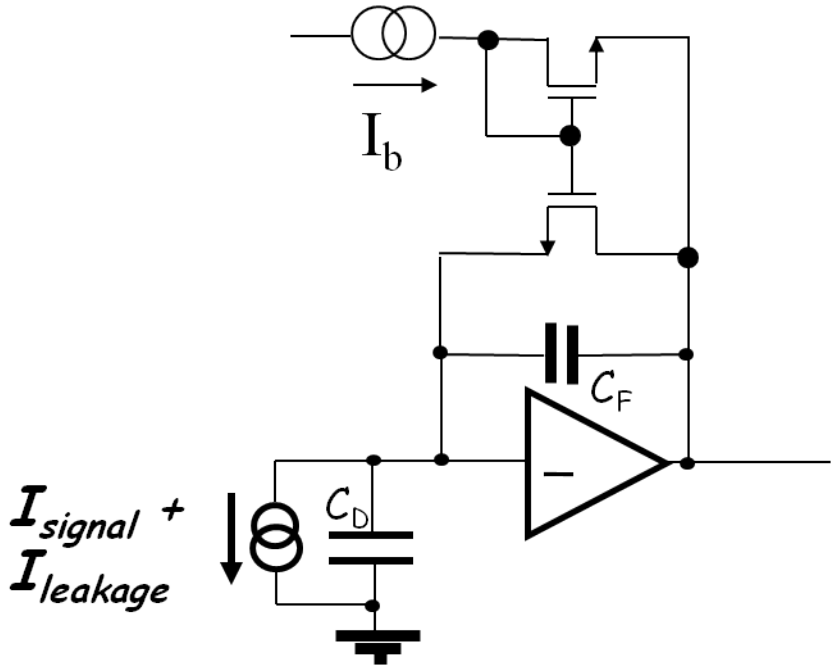
- A large feedback resistor is needed for low noise, since $\frac{\overline{di_R^2}}{df} = \frac{4kT}{R}$

- It is difficult to fabricate a large physical resistor in monolithic form, or to effectively control the resistance of a MOSFET biased in the linear region
- A large resistor can be simulated by a CMOS circuit, such as a transconductor, which can be considered to be equivalent to a resistor $R = 1/G_m$

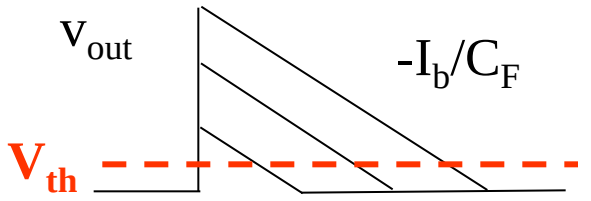


Compensation of detector leakage current

- Irradiated, dc-coupled pixel sensors may have a considerable leakage current, which may saturate the feedback transconductor or, flowing in the feedback resistor, considerably affect the dc voltage at the preamplifier output.
- A CMOS circuit can be designed to accomodate for this leakage current. A popular solution is the following:

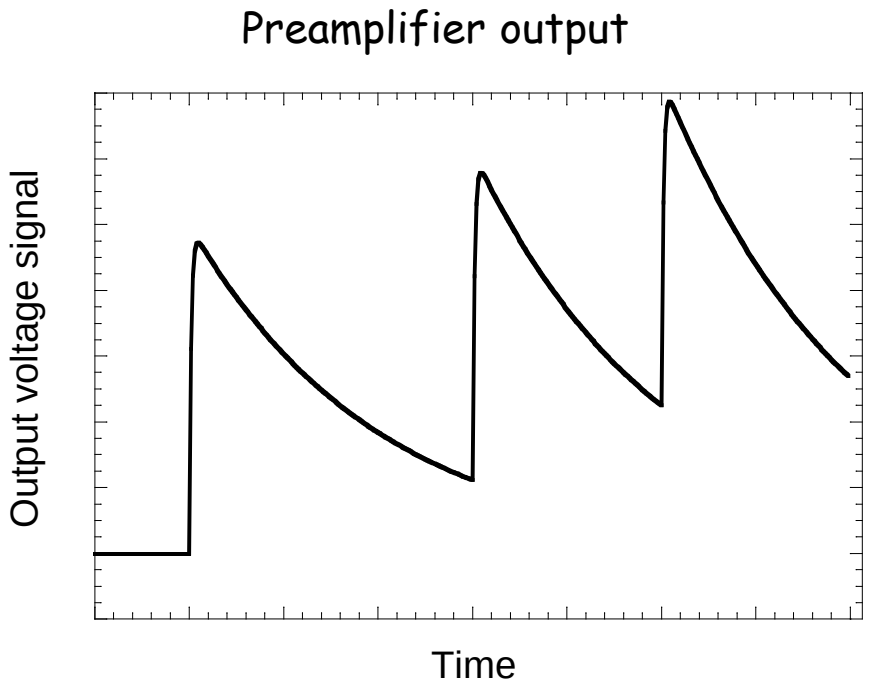


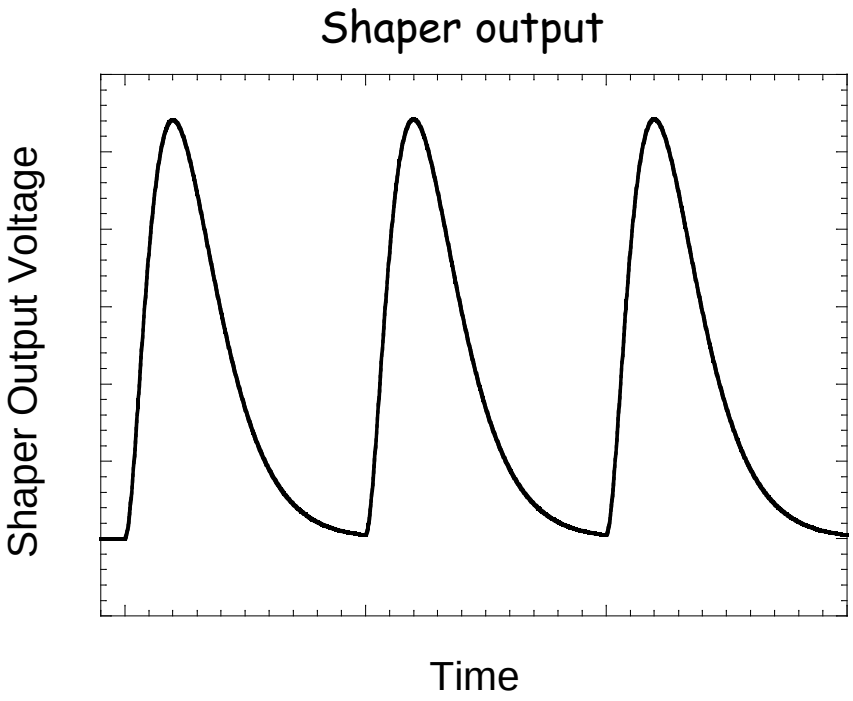
The feedback capacitor is discharged linearly by a constant current. The output signal lends itself to an amplitude-to-time conversion (time-over threshold measurement).



Processing the signal from the sensor: the shaper/filter

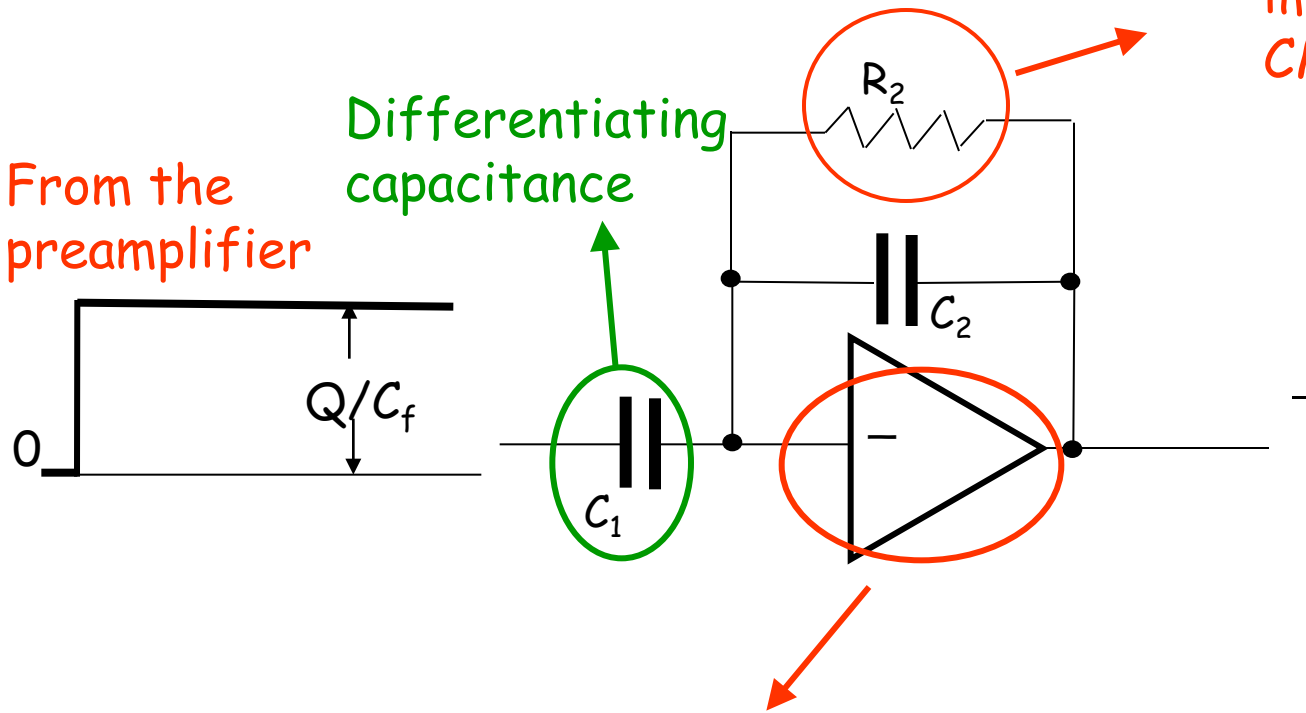
- Signal shaping: the voltage step at the preamplifier output has to be constrained to a finite duration to avoid pileup of successive signals

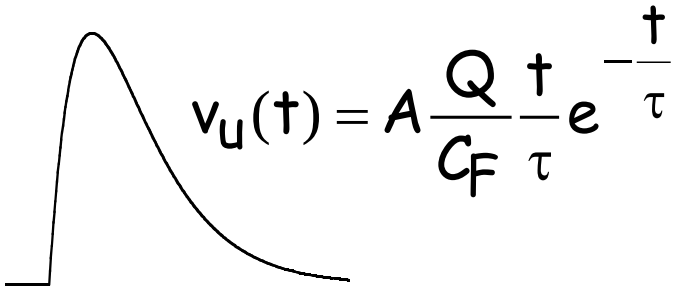




Processing the signal from the sensor: the shaper/filter

- A unipolar "semigaussian" shaper can be built with 1 differentiator (high pass) and n integrators (low-pass).
- This is a compact (n=1) implementation:





For correct values of the time constants associated to the feedback network and to the gain stage, the transfer function has two coincident poles

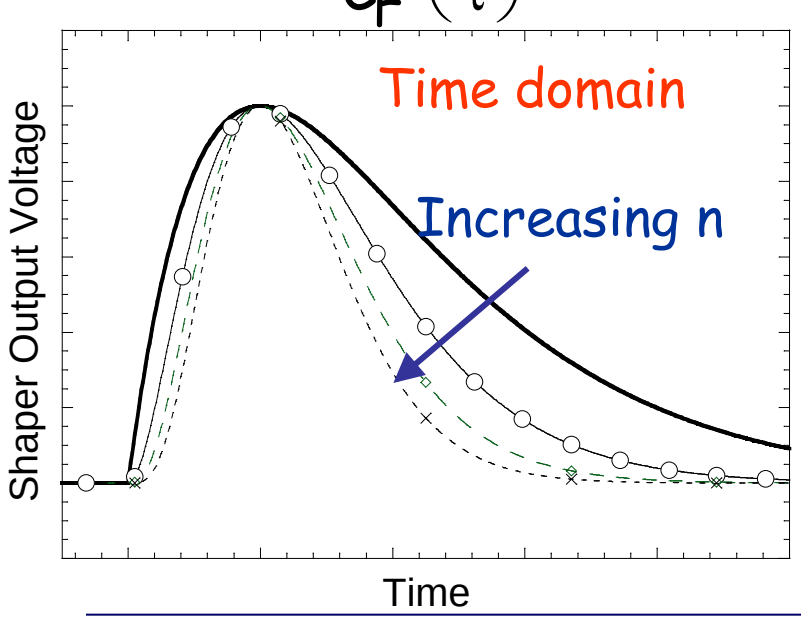
$$T(s) = \frac{s\tau}{(1+s\tau)^2}$$

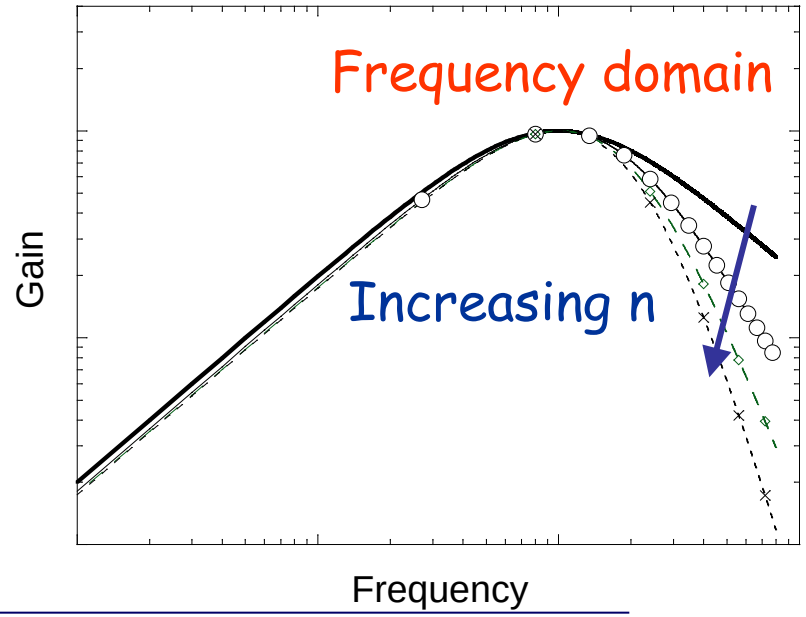
Processing the signal from the sensor: the shaper/filter

- In the front-end chips for silicon microstrip detectors, a second order (n=2) shaper is often used with an additional integrator before the shaper.
- For an n^{th}-order unipolar shaper (higher n: more symmetrical pulse, higher signal rates for the same peaking time):

$$v_u(t) = A\frac{Q}{C_F}\left(\frac{t}{\tau}\right)^n e^{-\frac{t}{\tau}}$$

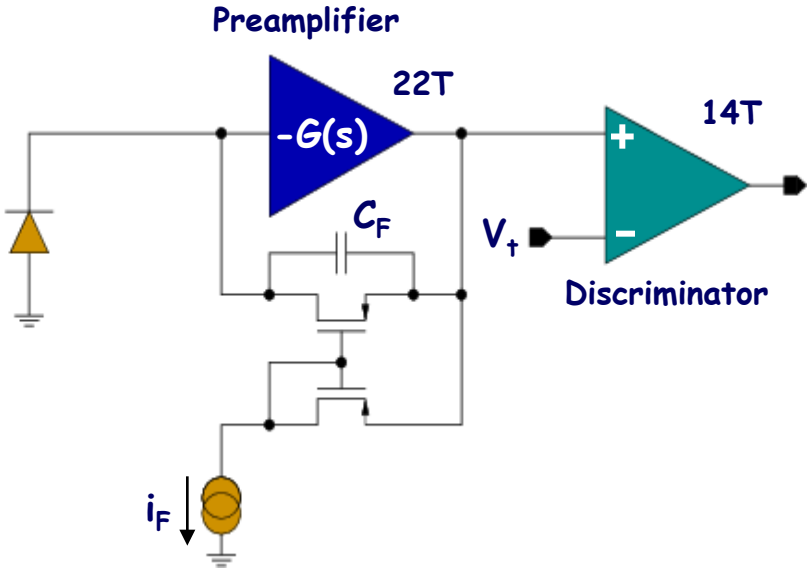
$$T(s) \div \frac{s\tau}{(1+s\tau)^n}$$

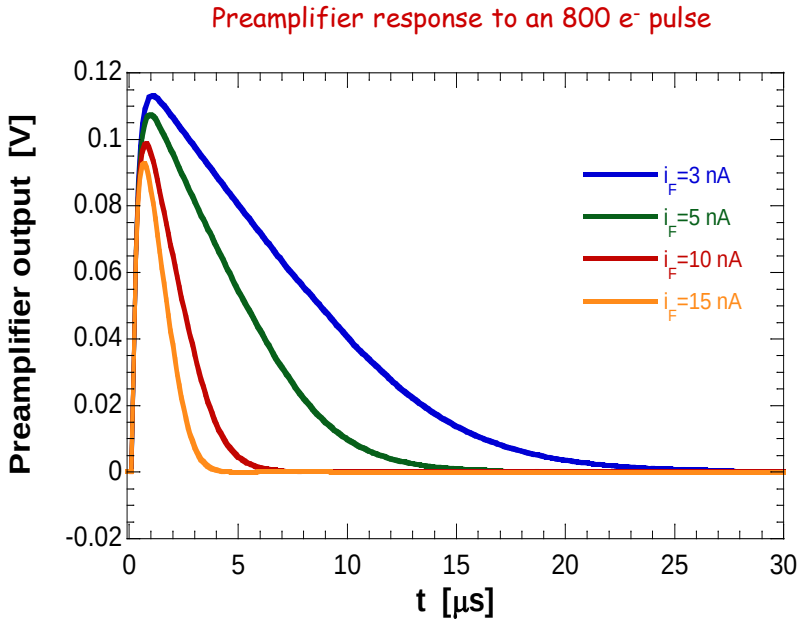




"Shaperless" analog channel

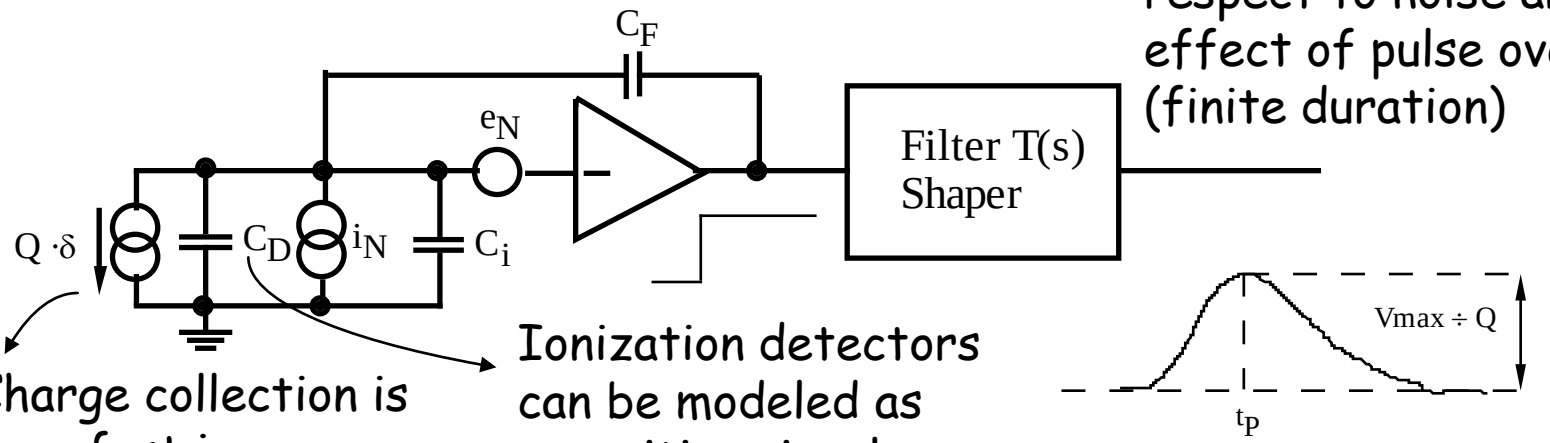
- In future applications for imaging and vertexing detectors, very small pixels may be needed with no room in the pixel for a shaper
- Under these constraints, a viable solution consists in artificially reducing the preamplifier bandwidth





Charge measuring system and the effect of noise

Filter minimizes the measurement error with respect to noise and the effect of pulse overlap (finite duration)



Charge collection is very fast in semiconductor detectors

Ionization detectors can be modeled as capacitive signal sources

Noise arises from two uncorrelated sources at the input (series and parallel noise):

$$S_{e_N}(\omega) = A_W + \frac{A_f}{f} \qquad S_{I_N}(\omega) = B_W$$

Noise sources

White series noise

$$A_W = 4kT\frac{\Gamma}{g_m}$$

White noise in the main current (drain, collector) of the input device

other components in the input stage

stray resistances in series with the input

1/f series noise

$$A_{1/f} = \frac{A_f}{f}$$

1/f component in the drain current

White parallel noise

$$B_W = 2qI_{\det} + 2qI_{G(B)} + \frac{4kT}{R}$$

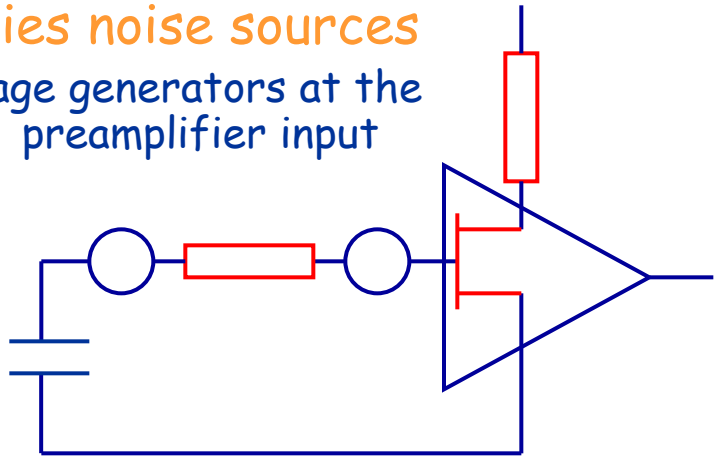
Shot noise in detector leakage current

shot noise in input device gate (base) current

thermal noise in feedback resistor

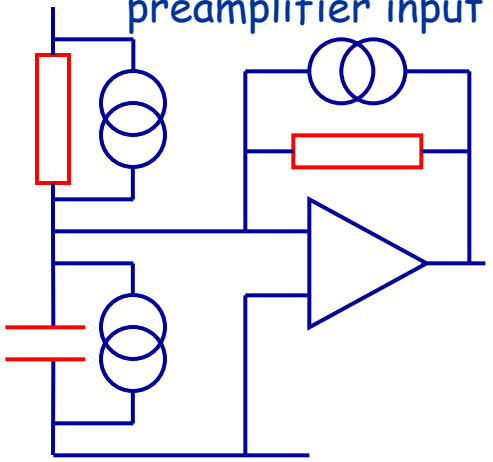
Series noise sources

Voltage generators at the preamplifier input



Parallel noise sources

Current generators at the preamplifier input



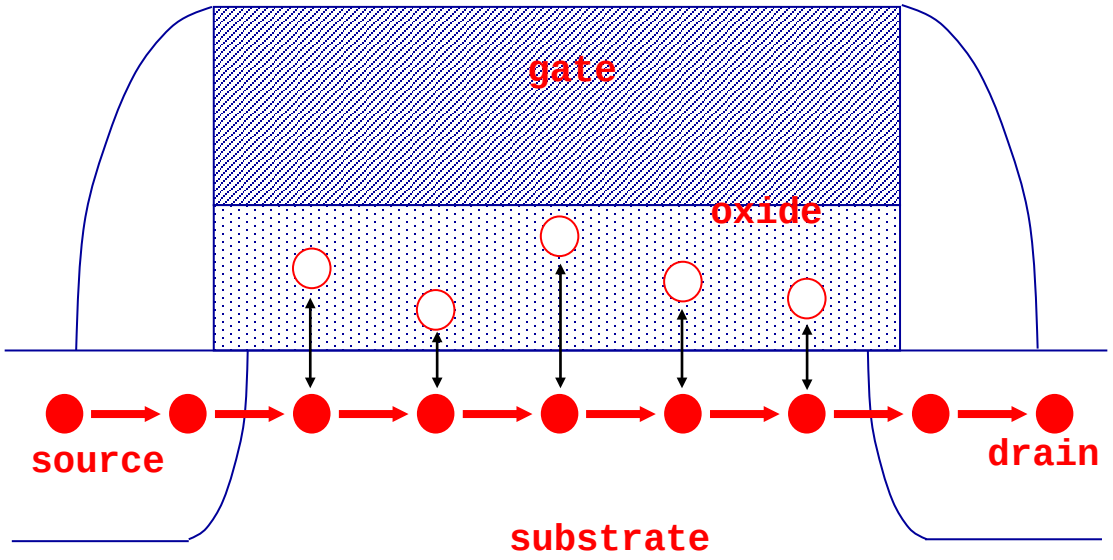
Shot noise

Shot noise is associated to device currents when charge carriers have to cross a potential barrier (P-N junctions in diodes and bipolar transistor)

$$S_I(\omega) = 2qI$$

In irradiated silicon detectors, leakage current and the associated shot noise may strongly increase

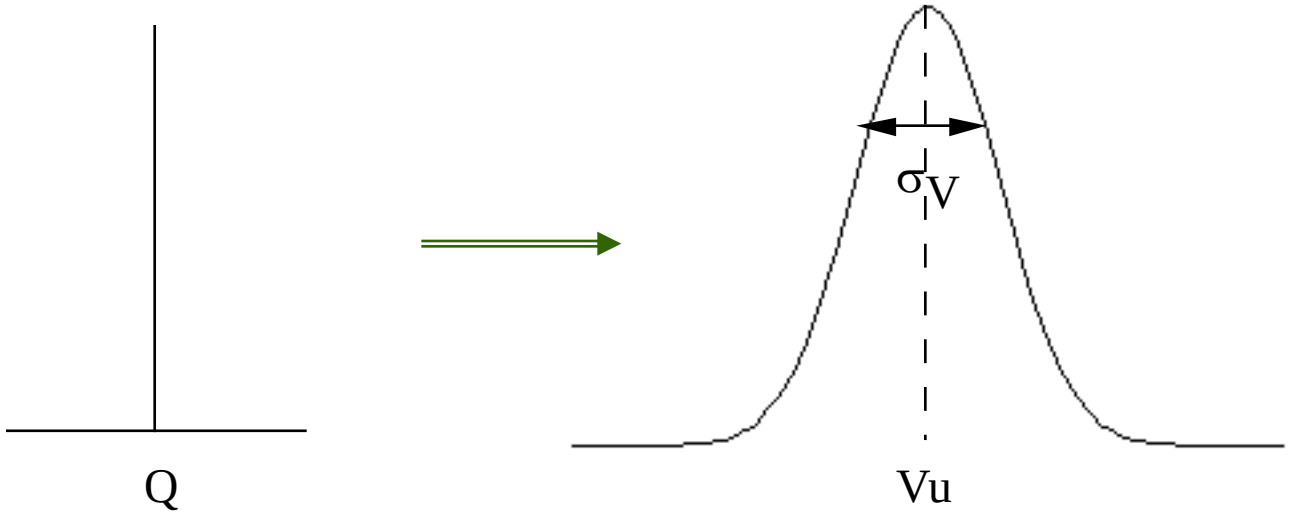
1/f noise



Interaction between charge carriers in the MOSFET channel and traps close to the Si-SiO_2 interface leads to fluctuations in the drain current.

This can be modeled with a noise voltage generator in series with the device gate, with a 1/f spectral density.

Effect of electronic noise on charge measurements

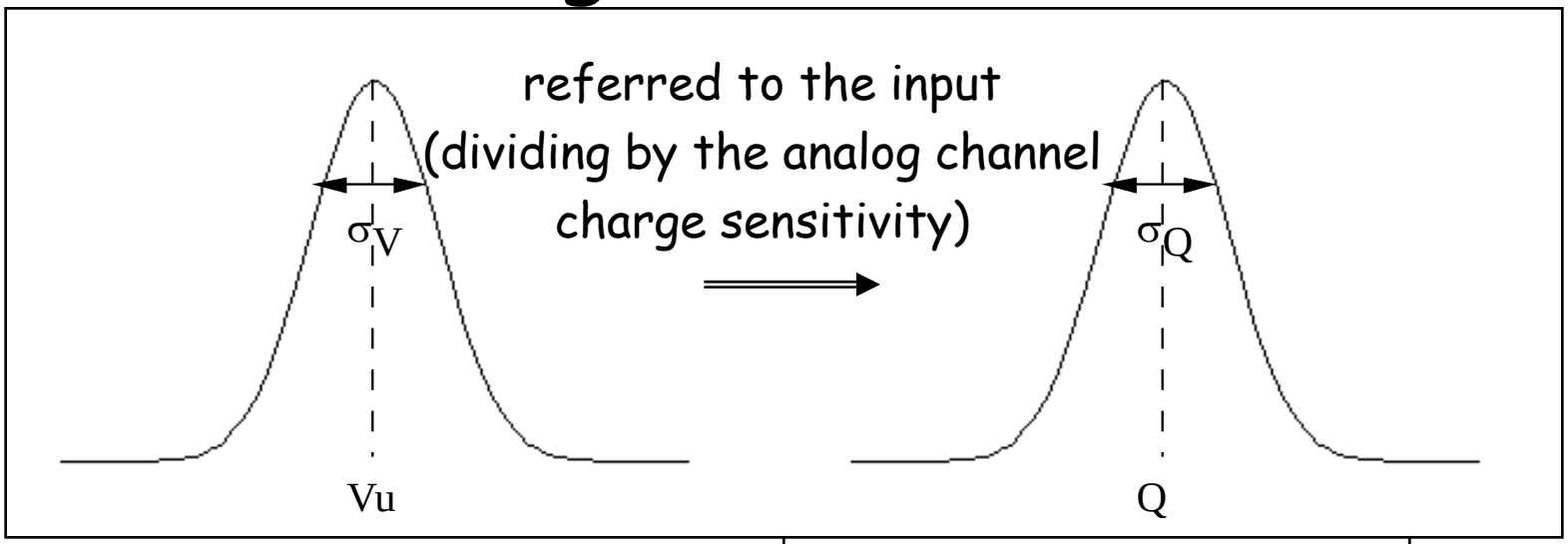


Ideally indefinitely narrow distribution of detector charge (neglecting statistics in energy deposition and charge creation)

Broadening of pulse amplitude distribution at the shaper output due to electronic noise

Because of electronic noise, the signal amplitude at the shaper output has a Gaussian probability density function

Effect of electronic noise on charge measurements



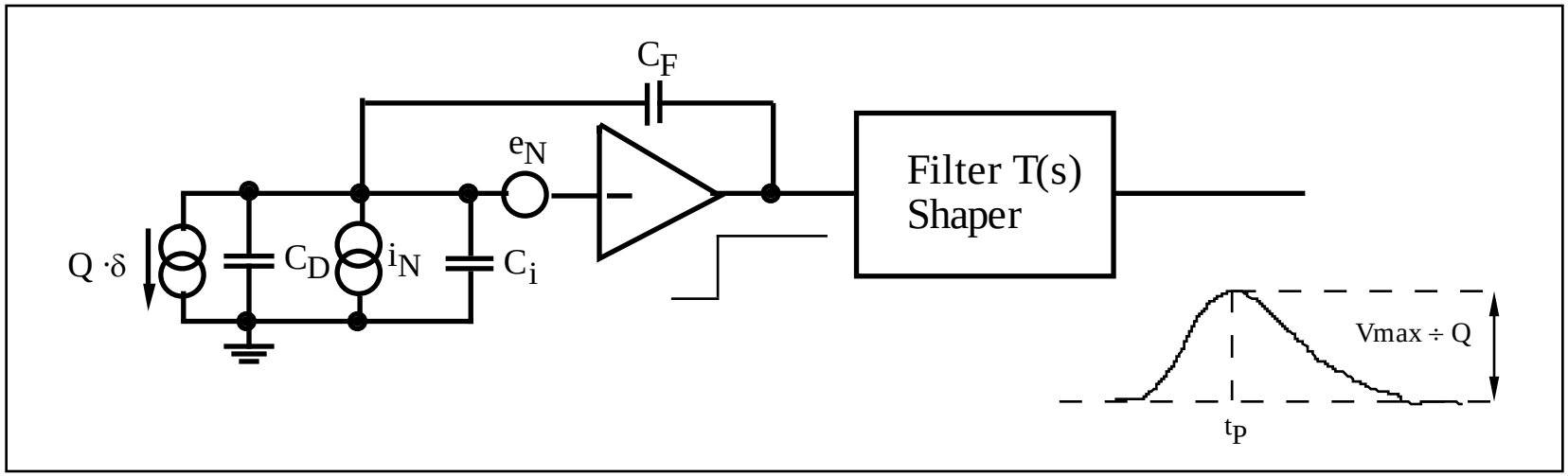
The signal amplitude at the output of the linear analog channel is characterized by a Gaussian probability density function

$$S/N = \frac{V_u}{\sigma_V} = \frac{Q}{\sigma_Q} = \frac{Q}{ENC} = \eta_Q$$

Equivalent Noise Charge = standard deviation in the charge measurement

charge injected at the input producing at the output of the linear processor a signal whose amplitude equals the root mean square output noise

Equivalent Noise Charge (ENC)



The mean square value of the noise voltage at the shaper output can be calculated as follows:

$$\overline{v_{u,N}^2} = \int_0^\infty S_u(\omega)df = \int_0^\infty \left[\left|T_{e_N}(j\omega)\right|^2 \cdot S_{e_N}(\omega) + \left|T_{I_N}(j\omega)\right|^2 \cdot S_{I_N}(\omega) \right] df$$

$$= \int_0^\infty \left[\left|T(j\omega)\right|^2 \cdot \frac{(C_D + C_i + C_F)^2}{C_F^2} \left(A_W + \frac{A_f}{f} \right) + \left|T(j\omega)\right|^2 \frac{1}{\omega^2 C_F^2} \cdot B_W \right] df =$$

Equivalent Noise Charge (ENC)

$$= A_W \frac{(C_D + C_i + C_F)^2}{C_F^2} \frac{1}{2\pi} \int_0^\infty |T(j\omega)|^2 d\omega +$$

$$+ A_f \frac{(C_D + C_i + C_F)^2}{C_F^2} \int_0^\infty \frac{|T(j\omega)|^2}{\omega} d\omega + B_W \frac{1}{C_F^2} \frac{1}{2\pi} \int_0^\infty \frac{|T(j\omega)|^2}{\omega^2} d\omega$$

$$\frac{1}{2\pi} \int_0^\infty |T(j\omega)|^2 d\omega = \frac{A_1}{t_P}$$

$$\int_0^\infty \frac{|T(j\omega)|^2}{\omega} d\omega = A_2$$

$$\frac{1}{2\pi} \int_0^\infty \frac{|T(j\omega)|^2}{\omega^2} d\omega = A_3 t_P$$

t_P = peaking time of the signal at the shaper output

A_1, A_2, A_3 = filter-dependent coefficients

Equivalent Noise Charge (ENC)

$$ENC = \frac{\sqrt{\overline{v_{u,N}^2}}}{Charge \quad sensitivity}$$

$$ENC^2 = \overline{v_{u,N}^2} \cdot C_F^2 = A_W (C_D + C_i + C_F)^2 \frac{A_1}{t_P} + A_f (C_D + C_i + C_F)^2 A_2 + B_W A_3 t_P$$

$C_T = C_D + C_i + C_F$

= total capacitance at the preamplifier input

In a well designed preamplifier, the noise is determined by the input device.

Equivalent Noise Charge (ENC)

$$ENC^2 = A_W C_T^2 \frac{A_1}{t_P} + A_f C_T^2 A_2 + B_W A_3 t_P$$

White series noise:

White parallel noise:

Neglecting noise in parasitic resistors:

$$A_W = 4kT \frac{\Gamma}{g_m}$$

$\Gamma = 0.5$ (BJT)

$\Gamma = 2/3$ (Long channel FETs)

$\Gamma \approx 1$ (Short-channel FETs)

$$B_W = 2qI$$

$I = I_B$ (BJT)

$I = I_G$ (gate tunneling current in nanoscale CMOS)

$I = I_{leak}$ Detector leakage current

Equivalent Noise Charge (ENC)

$$ENC^2 = A_W C_T^2 \frac{A_1}{t_P} + A_f C_T^2 A_2 + B_W A_3 t_P$$

Series 1/f noise (MOSFET):

$$A_f = \frac{K_f}{C_{OX} WL}$$

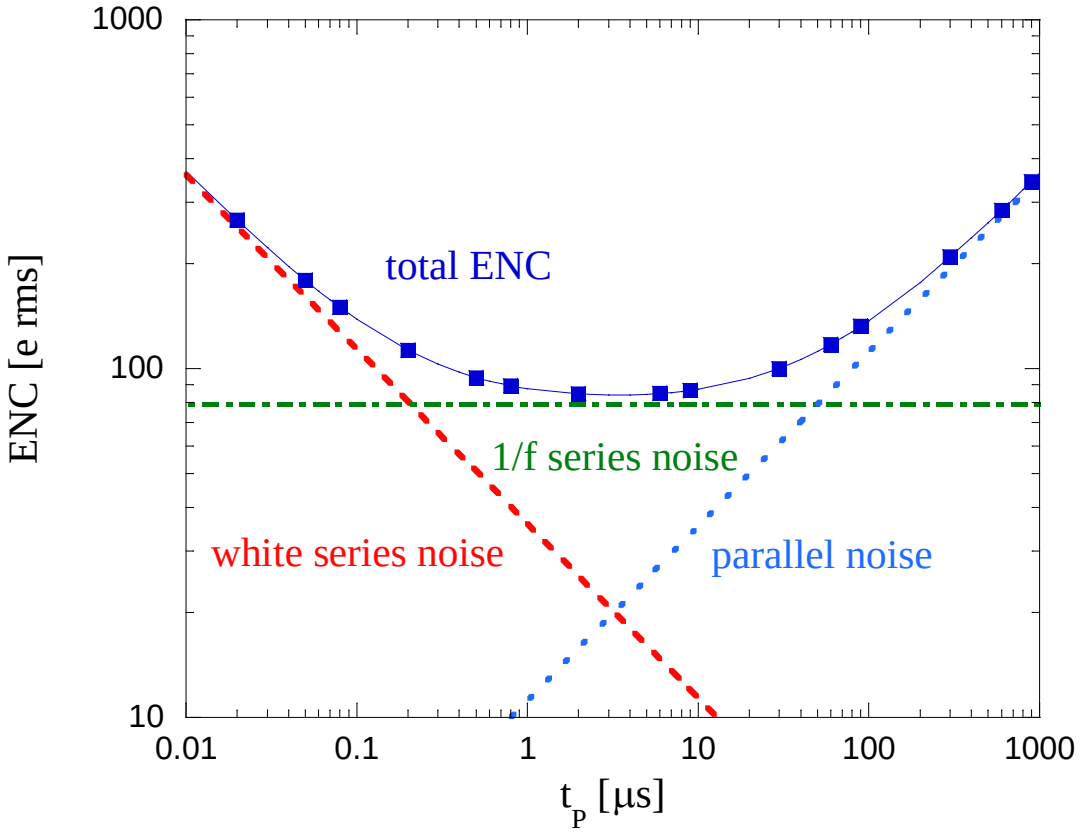
K_f: 1/f noise parameter; depends on the gate oxide quality

C_{OX}: Oxide capacitance per unit gate area

WL: Transistor geometry (gate Width and Length)

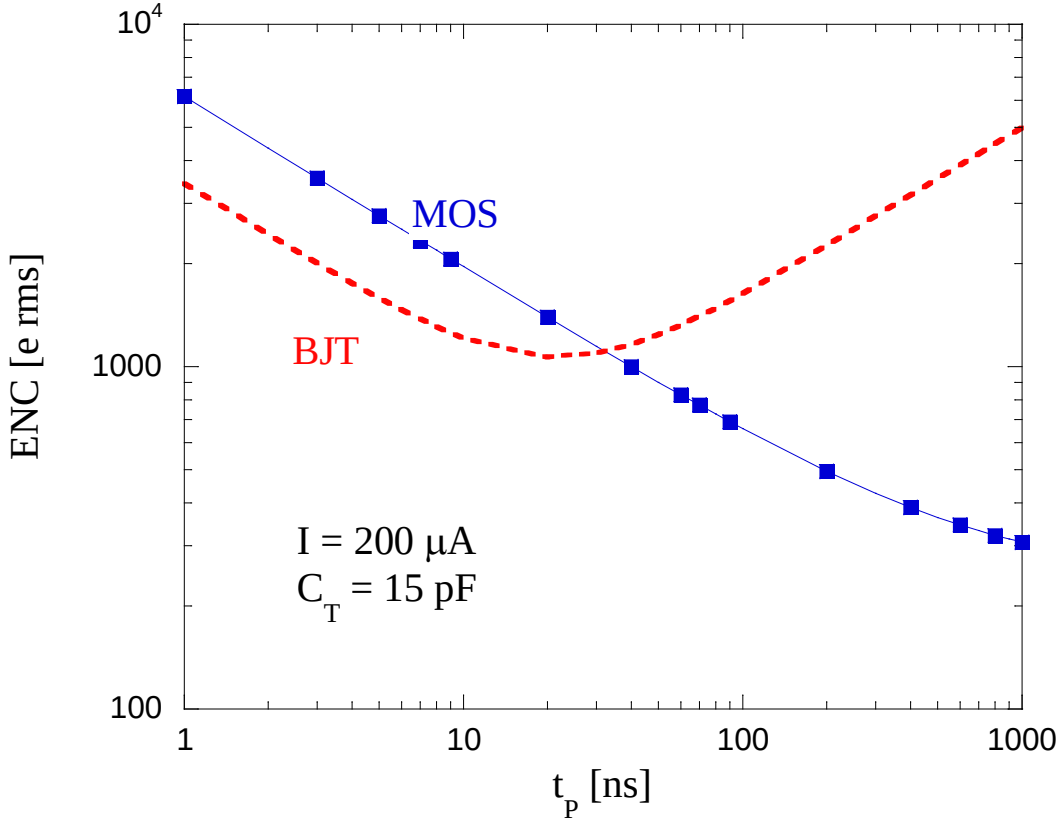
The ENC contribution from 1/f noise is independent of the peaking time of the signal at the shaper output; it is weakly dependent on the shape of the transfer function of the shaper.

- In trackers for high luminosity colliders, event rate is very high, and the peaking time has to be short (< 25 ns).
- White series noise is usually dominant here, except with irradiated sensors, where leakage current (and the associated shot noise) may increase to a very large extent.



ENC: BJT vs MOSFET

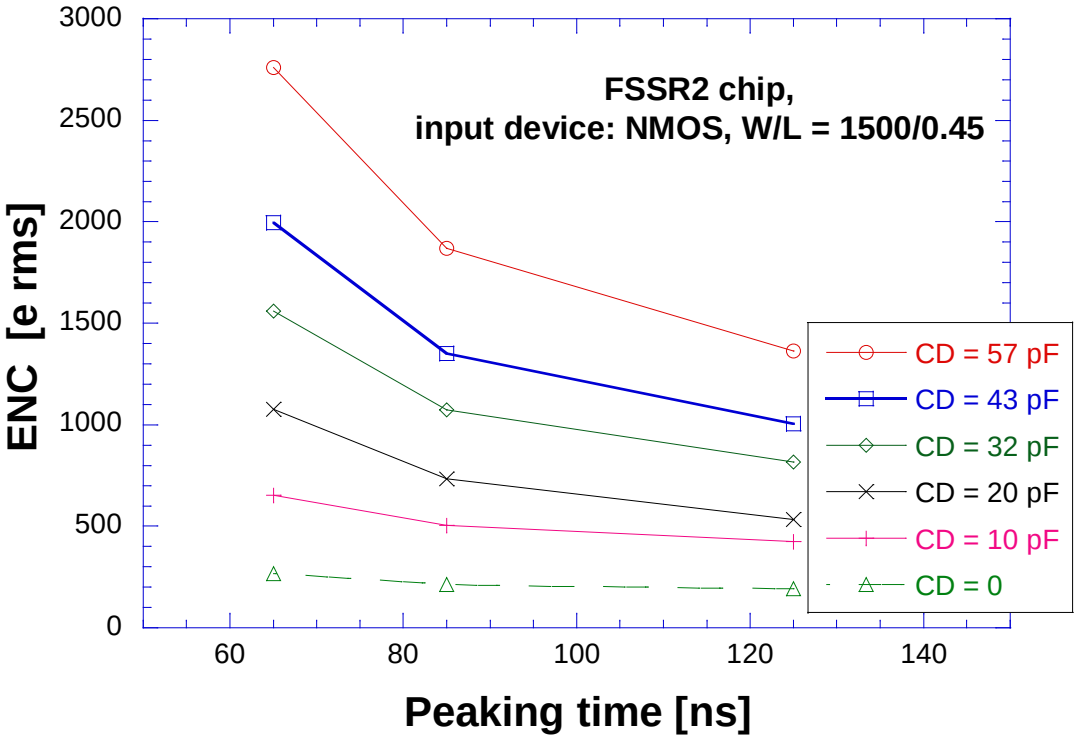
- Bipolar transistors have a larger g_m/I ratio with respect to MOSFET, which means a lower series white noise for a same current
- BiCMOS (SiGe) technologies are an appealing alternative for fast readout systems; since they are less dense (and more expensive) than CMOS, their use is mostly limited to strip front-end chips

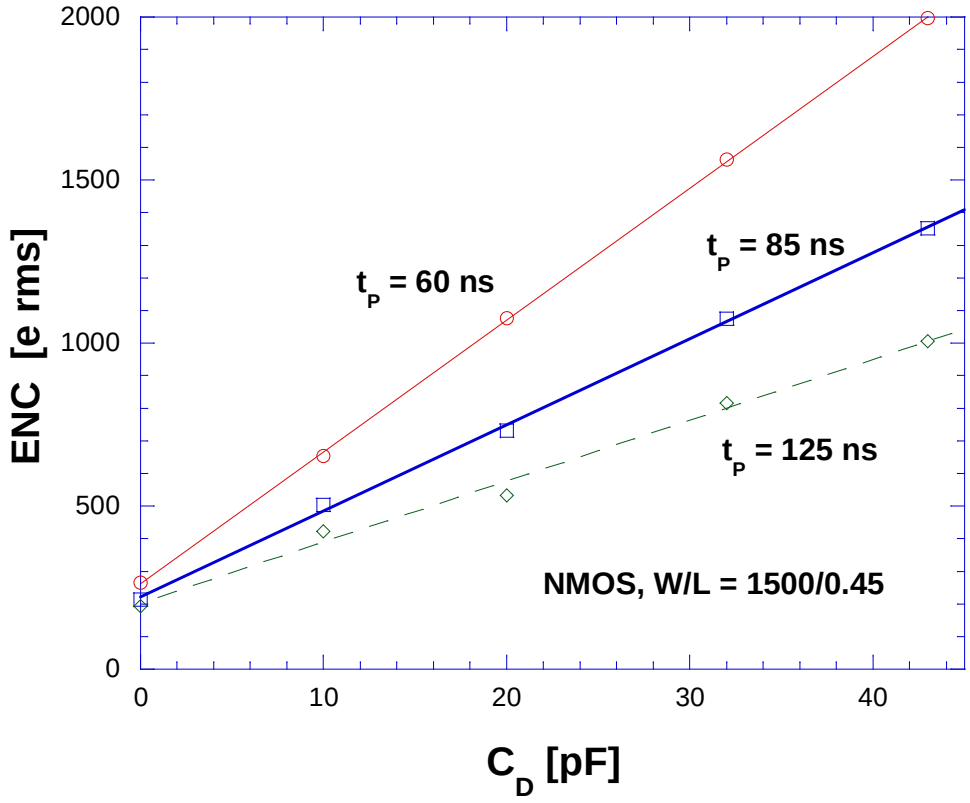


Noise and detector capacitance

- White and 1/f series noise terms (dominant in CMOS) give a contribution to ENC linearly increasing with the detector capacitance ($C_T = C_D + C_{IN} + C_F$).

$$ENC^2 = \left(A_W \frac{A_1}{t_P} + \frac{K_f}{C_{OX} W L} A_2 \right) C_T^2$$





CMOS 65 nm: noise and radiation tolerance

- Critical aspects for analog design in nanoscale CMOS **(with focus on LP 65 nm)** for detector readout integrated circuits:

Gate leakage current 1/f noise	Interaction of charge carriers with the gate oxide; tools for evaluating the quality of the gate dielectric
Thermal noise	Charge carriers in the device channel (short channel effects, strained silicon)
Radiation hardness	Radiation-induced positive charge in the gate oxide and in lateral isolation oxides (see presentation by Lodovico Ratti)

Thermal noise and CMOS scaling

- The origin of thermal noise can be traced to the random thermal motion of carriers in the device channel.
- When the MOSFET is biased in saturation ($V_{DS} > V_{DS,sat}$), the following equation can be used for the power spectral density of thermal noise in all inversion conditions:

$$S_W^2 = 4k_B T\Gamma \frac{1}{g_m}$$

$$\Gamma = \alpha_W n \gamma$$

- k_B = Boltzmann's constant
- T = absolute temperature
- $n = 1 + \frac{g_{mb}}{g_m}$

  (n = 1 – 1.5; proportional to the inverse of the slope of the I_D-V_{GS} curve in the subthreshold region)

- γ = channel thermal noise coefficient (depends on inversion region; varies with the inversion layer charge: = 1/2 in weak inversion, = 2/3 in strong inversion)
- α_W = excess noise coefficient

Excess thermal noise coefficient

$$S_W^2 = 4k_B T n \gamma \frac{1}{g_m} \alpha_W$$

α_W = 1 for long-channel devices. Short-channel devices can be noisier (α_W > 1) mainly because of two effects related to high longitudinal electric fields (E = V_{DS}/L) in the channel.

- Reduction of charge carrier mobility:
  at increasing field strength the carrier velocity is saturated at $v_{sat}=\mu_0 E_C$ (μ_0 low-field mobility, E_C critical field strength)

$$\mu = \frac{\mu_0}{1+\frac{E}{E_C}}$$

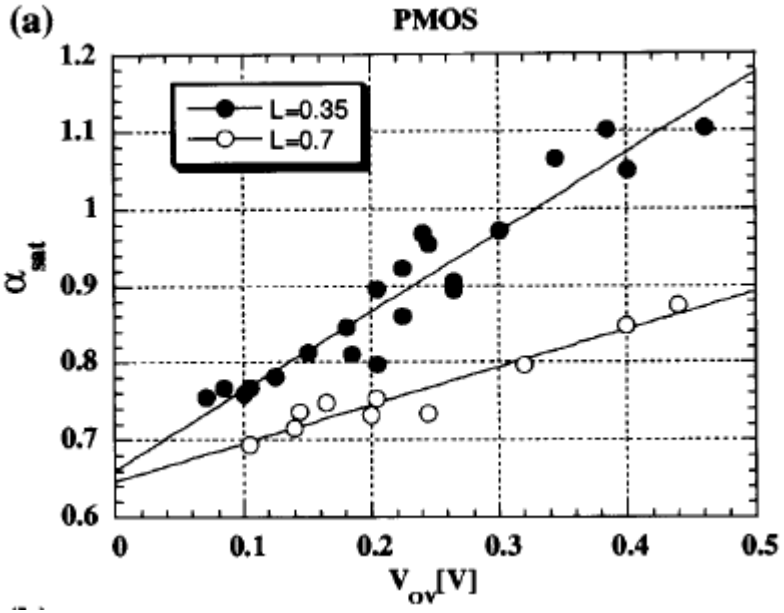
- Increase of charge carrier temperature:
  at increasing field strength the temperature T_e of carriers in the channel increases with respect to the temperature T of the device lattice

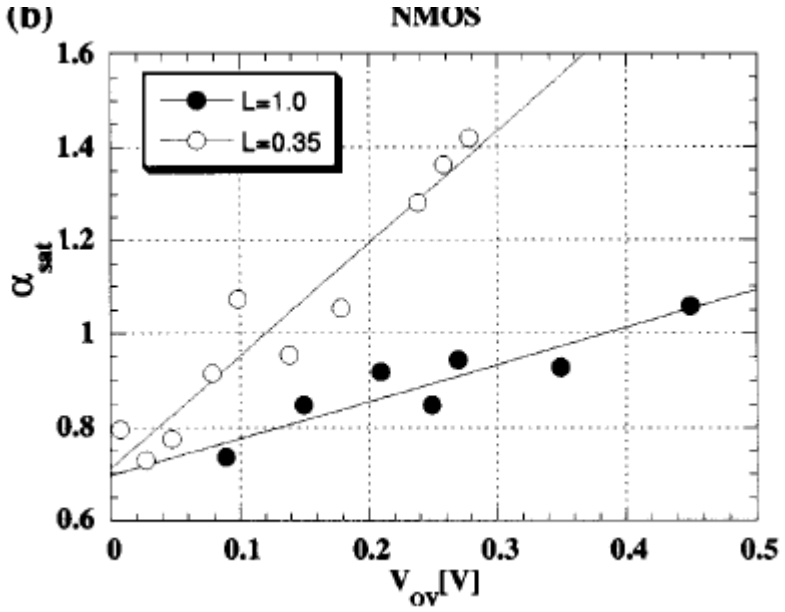
$$\frac{T_e}{T} = \left(1+\frac{E}{E_C}\right)^b$$

Excess thermal noise coefficient

In saturation, the longitudinal electric field is $E = (V_{GS} - V_{TH})/L$.

It can be shown that short-channel phenomena affect thermal noise only if the value of the ratio E/E_C is not negligible, which does not happen if the device is biased in weak/moderate inversion.





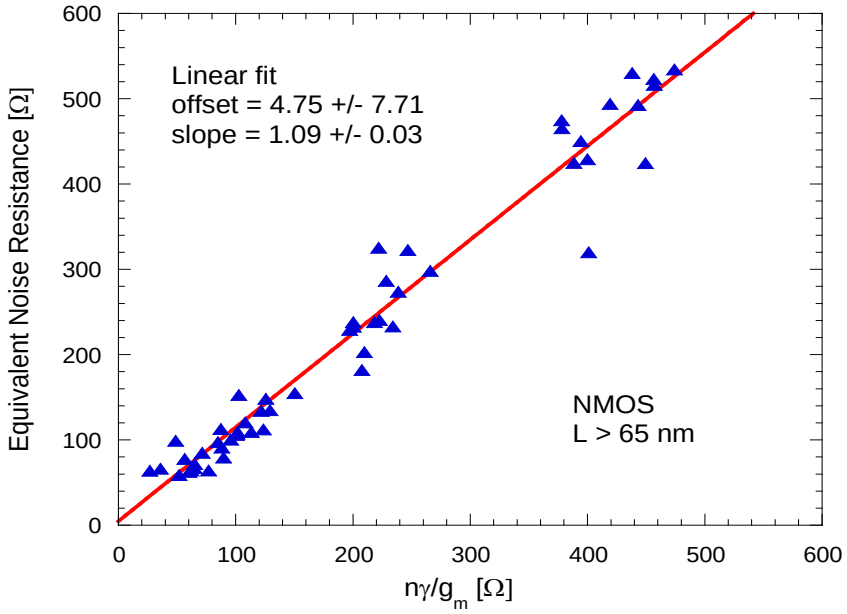
350 nm CMOS

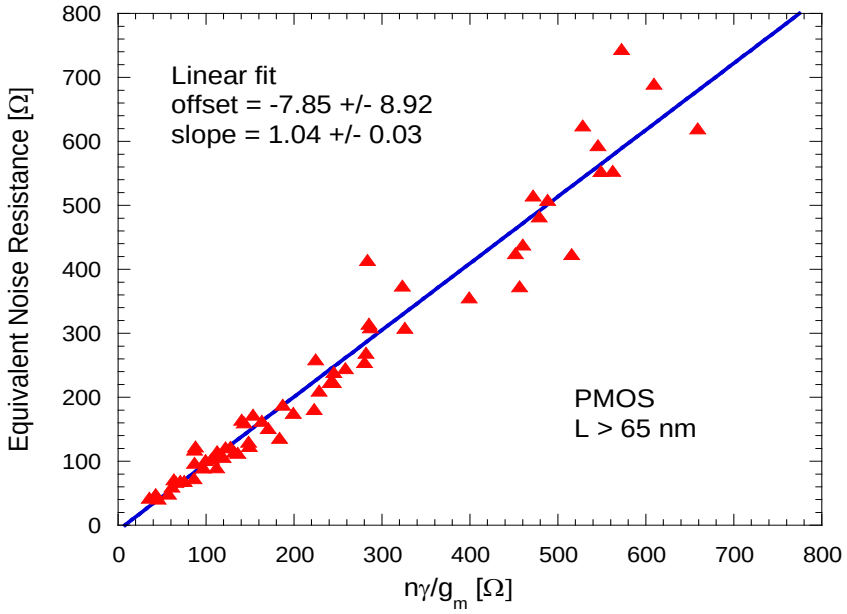
White noise in 65 nm CMOS

Evaluated in terms of the **equivalent channel thermal noise resistance**:

$$R_{eq} = \frac{S_W^2}{4k_B T} = \alpha_W \frac{n\gamma}{g_m}$$

- α_w excess noise coefficient
- n proportional to $I_D(V_{GS})$ subthreshold characteristic
- γ channel thermal noise coeff.





α_w close to unity for NMOS and PMOS with L > 65 nm → **no sizeable short channel effects** in the considered operating regions (except for 65 nm devices with $\alpha_w \approx 1.3$)

Negligible contributions from **parasitic resistances**

White noise in 65 nm CMOS

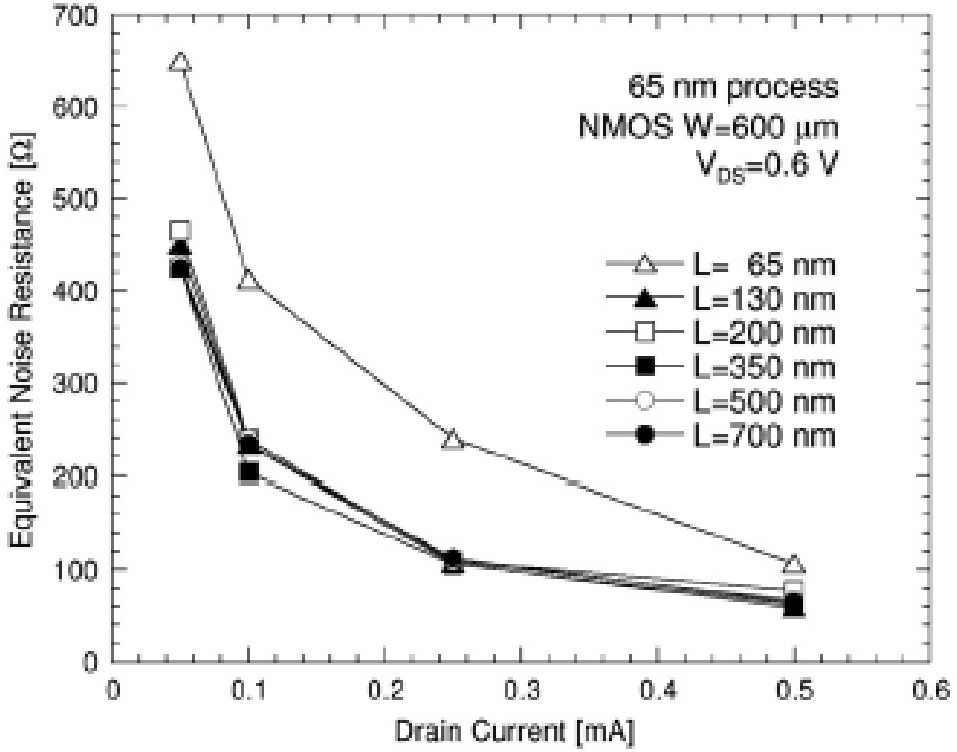
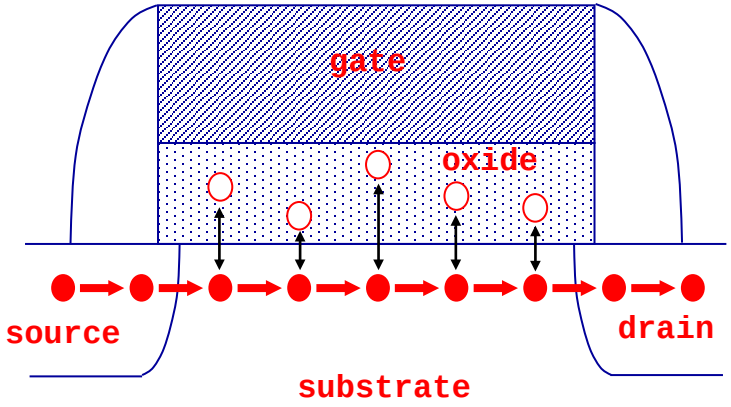


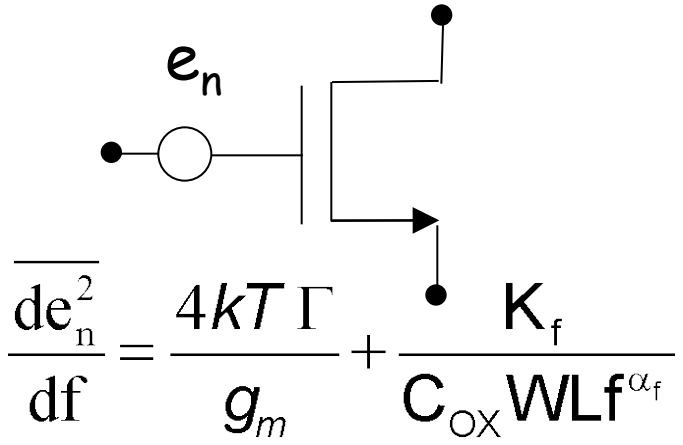
Fig. 7. Equivalent channel thermal noise resistance R_{eq} for NMOS devices belonging to the 65nm process (V_{DS}=0.6V).

1/f noise: gate stack fabrication process

- Interaction between charge carriers in the MOSFET channel and traps close to the Si-SiO_2 interface leads to fluctuations in the drain current. This can be modeled with a 1/f term in the spectral density of the noise voltage generator in series with the gate.



- The process recipe for the gate stack (gate electrode and dielectric) may affect the density of oxide traps and their interaction with charge carriers in the channel, impacting on the 1/f noise spectral density.



- For a physical oxide thickness < 2 nm (same order of the tunnelling distance) the traps at the interface between the gate dielectric and the gate electrode (fully silicided poly gates) can play a major role.

- 1/f noise may be affected by mechanical stress in the silicon channel (enhanced carrier mobility and drive current).

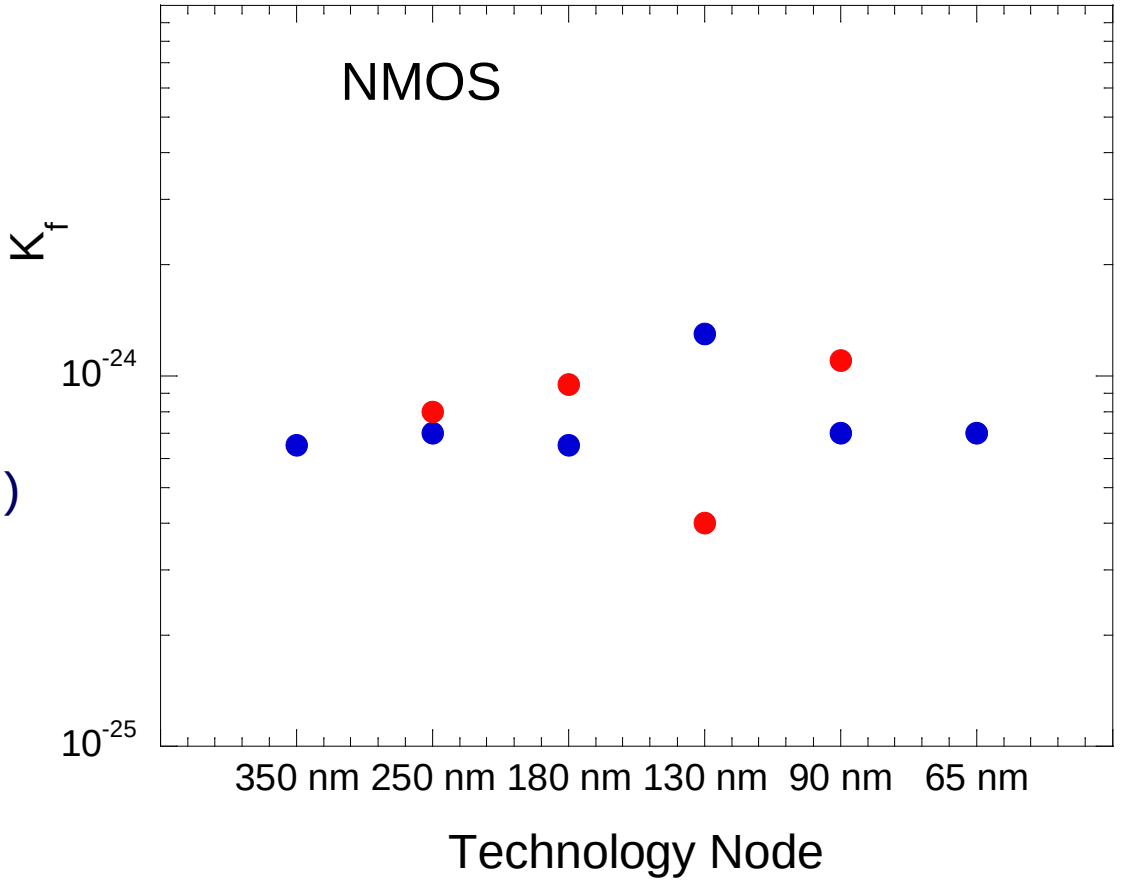
1/f noise from 350 nm to 65 nm NMOS

The 1/f noise parameter K_f does not show dramatic variations across different CMOS generations and foundries in NMOS.

$$S^2_{1/f}(f) = \frac{K_f}{C_{OX} W L f^{\alpha_f}}$$

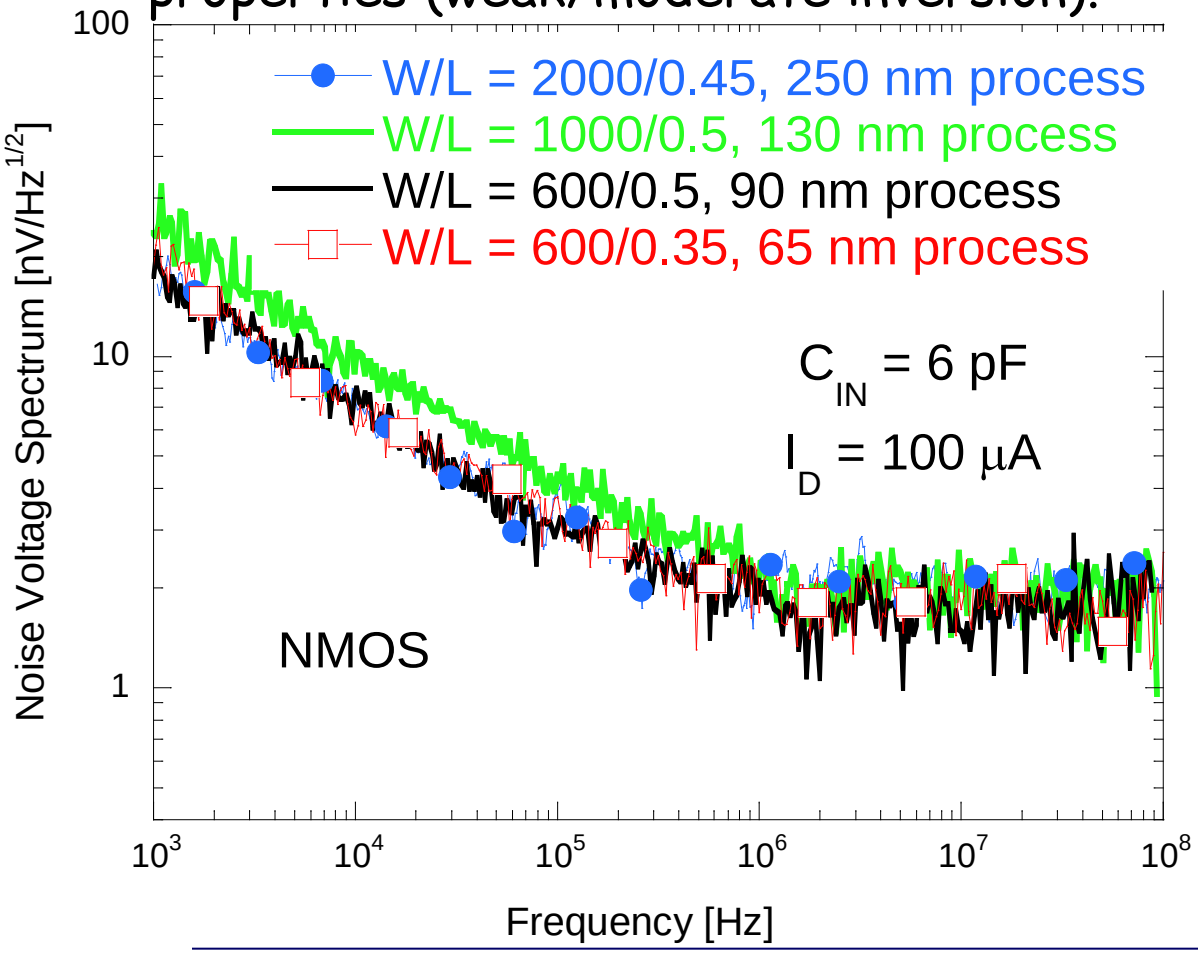
- k_f 1/f noise parameter
- α_f 1/f noise slope-related coefficient

($\cong$ 0.85 in NMOS, $\cong$ 1 – 1.1 in PMOS)



Noise in NMOS: CMOS generations from 250 nm to 65 nm

- 1/f noise has approximately the same magnitude (for a same WLC_{OX}) across different CMOS generations. White noise has also very similar properties (weak/moderate inversion).



1/f noise

$$S_{1/f}^2(f) = \frac{K_f}{C_{OX} W L f^{\alpha_f}}$$

- k_f 1/f noise parameter
- α_f 1/f noise slope-related coefficient

Channel thermal noise

$$S_W^2 = \frac{4k_B T \Gamma}{g_m}, \qquad \Gamma = \alpha_W n \gamma$$

- k_B Boltzmann's constant
- T absolute temperature
- α_w excess noise coefficient
- γ channel thermal noise coefficient

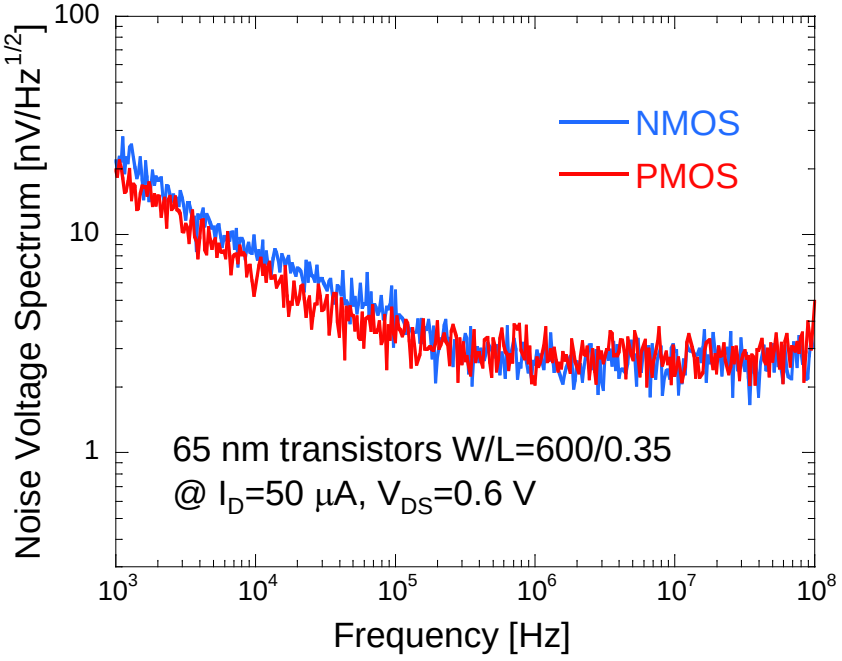
In weak inversion: $g_m = \frac{I_D}{nV_T}$

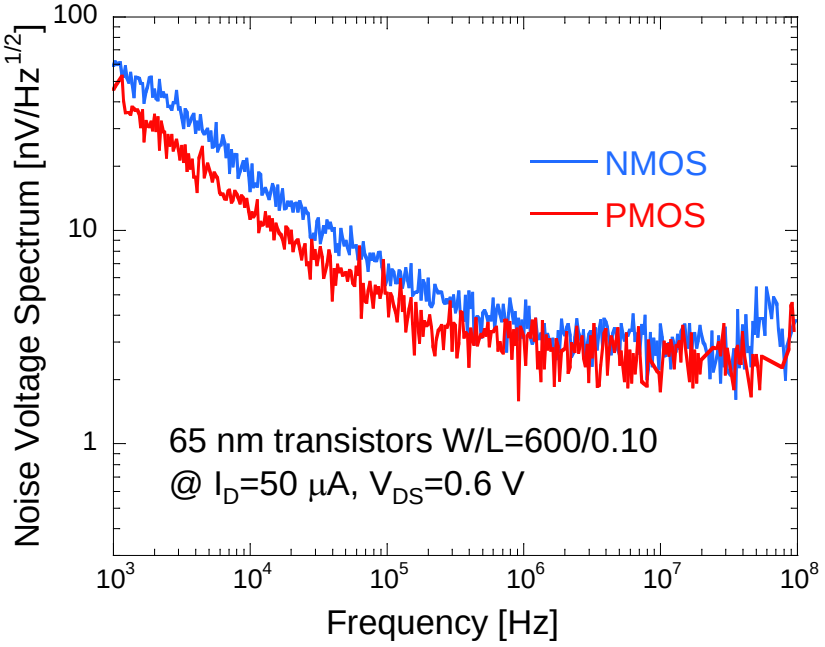
65 nm LP process: 1/f noise in PMOS

- In older CMOS times, PMOS transistors had a considerably lower 1/f noise with respect to NMOSFETs. This difference tends to decrease with newer CMOS generations.

- In the 65 nm LP process, **NMOS and PMOS have similar 1/f noise**.

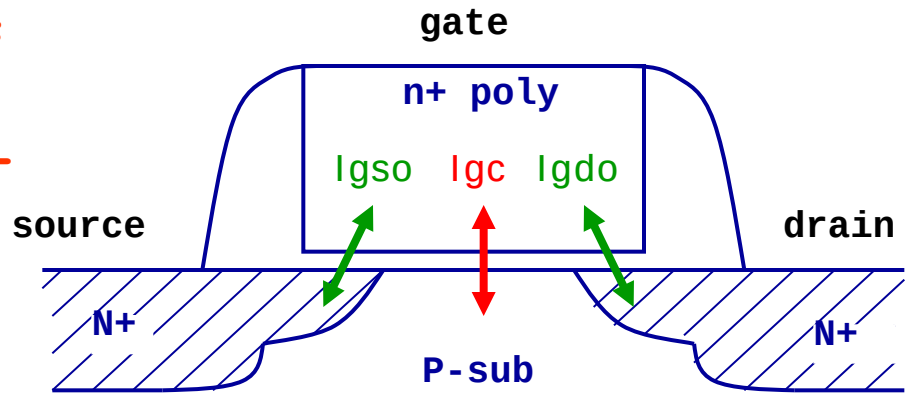
This could be explained by a "surface channel" behavior for both devices, and/or by the fact that the gate dielectric nitridation decreases the barrier energy experienced by holes across the silicon-dielectric interface. This would make it easier for the PMOS channel to exchange charges with oxide traps.



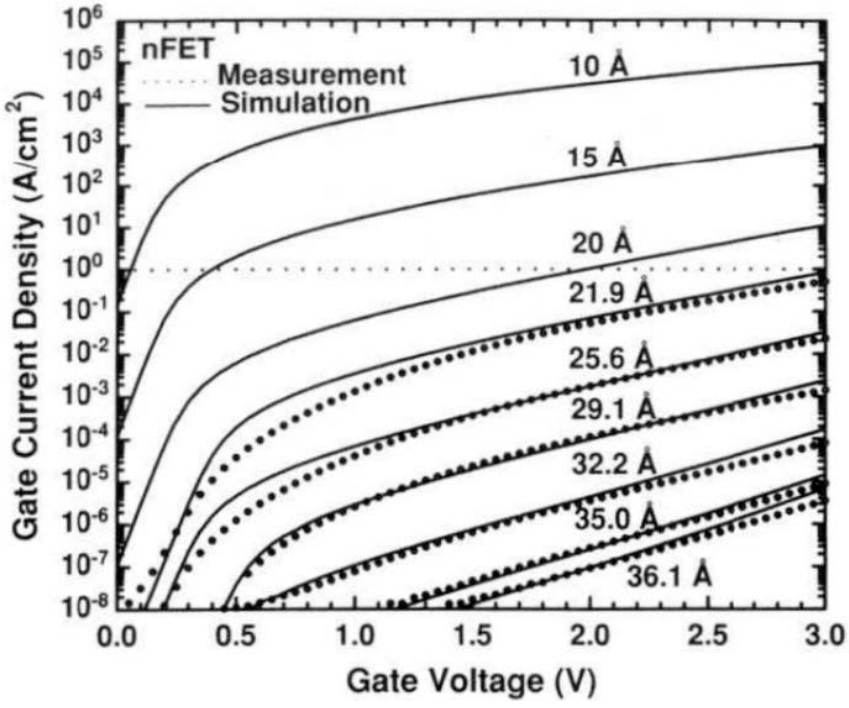


Gate current

- **Charge carriers have a nonzero probability** (larger for electrons with respect to holes) **of directly tunneling through a silicon dioxide layer with a physical thickness < 2 nm (100-nm scale CMOS).**

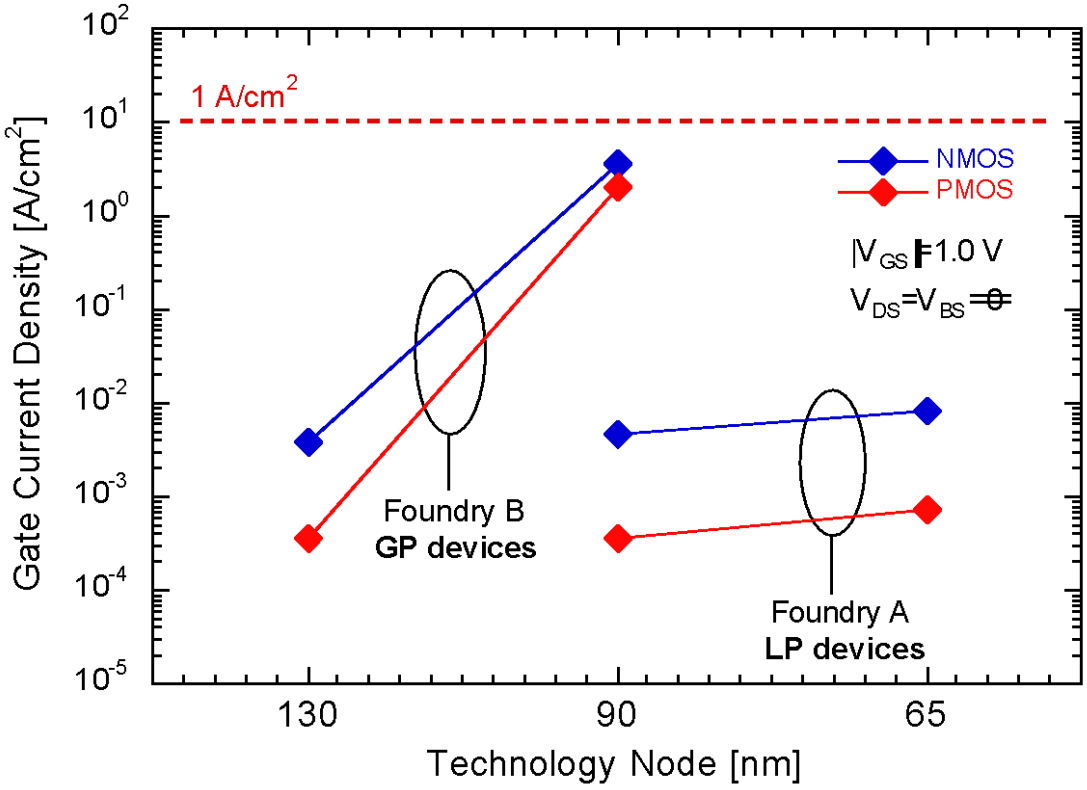


- ***A reduction of physical oxide thickness of a few Å may give several orders of magnitude increase in the gate current.***
- This current results in an **increase of the static power consumption** (manageable limit of gate leakage current density = 1 A/cm^2) for digital circuits and might **degrade analog performance** (shot noise in the gate current, discharge of storing capacitors, current load on global voltage references,…)
- **Gate dielectric nitridation** increases the dielectric constant, allowing for films with a **larger physical thickness** as compared with SiO_2 ($C_{OX} = \varepsilon_{OX}/t_{OX}$). This **mitigates the gate leakage current**; however, its value can sizably change in devices from different foundries.

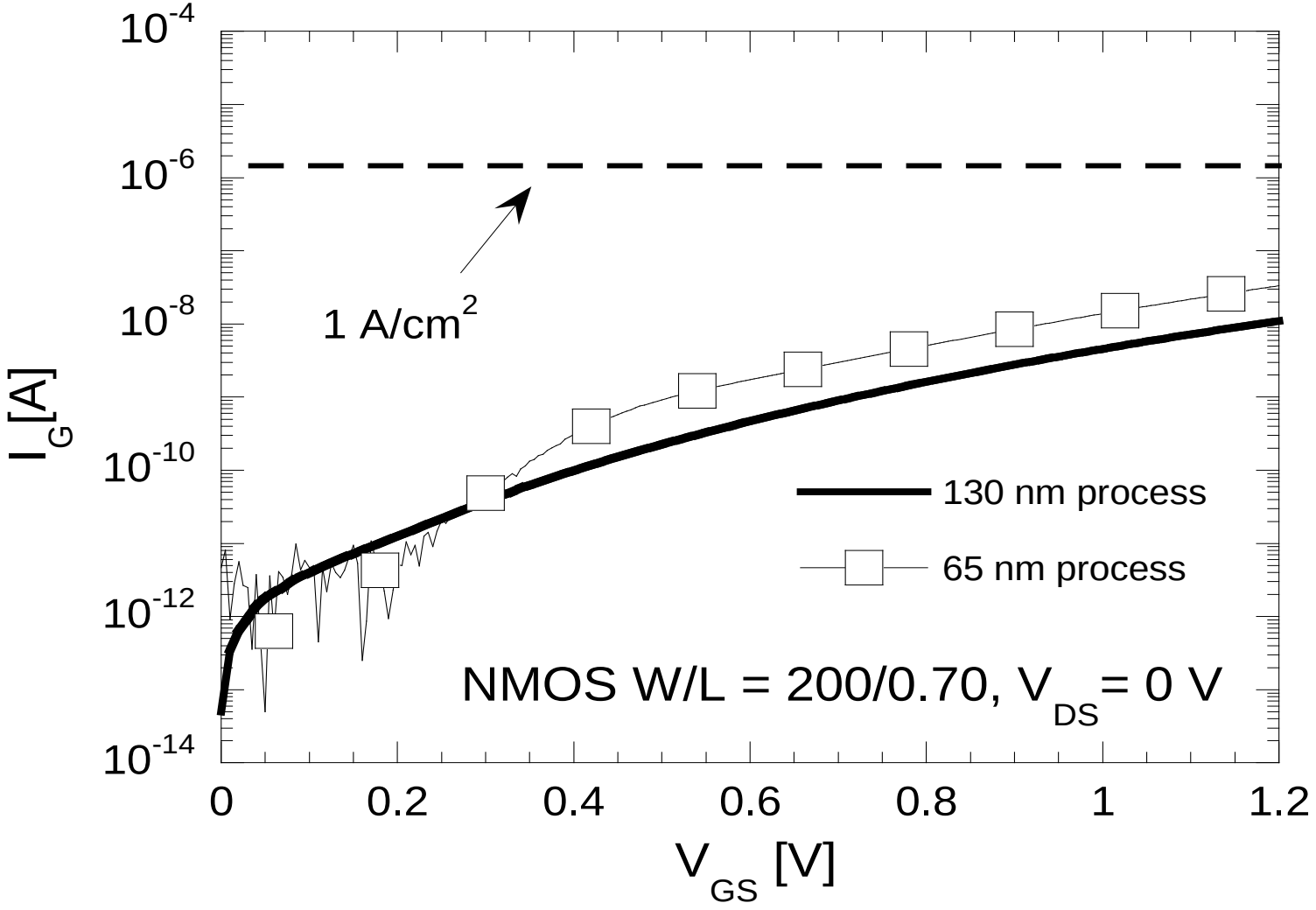


Gate leakage current in nanoscale CMOS flavours and generations

- Interestingly, the **gate leakage current is observed to decrease for the types of stresses adopted by the industry in advanced CMOS** (tensile and compressive stress for NMOS and PMOS, respectively)
- These types of stresses increase electron and hole populations in energy bands where they have low tunneling probability through SiO_2 (in addition to enhancing their mobility in the channel)
- We made tests on **130 nm and 90 nm GP (General Purpose) transistors** and on **65 nm LP (Low Power) transistors (V_{DD} = 1.2 V).** These LP devices were optimized for a reduced leakage (larger equivalent oxide thickness, different level of nitridation with respect to other flavours, different silicon stress).



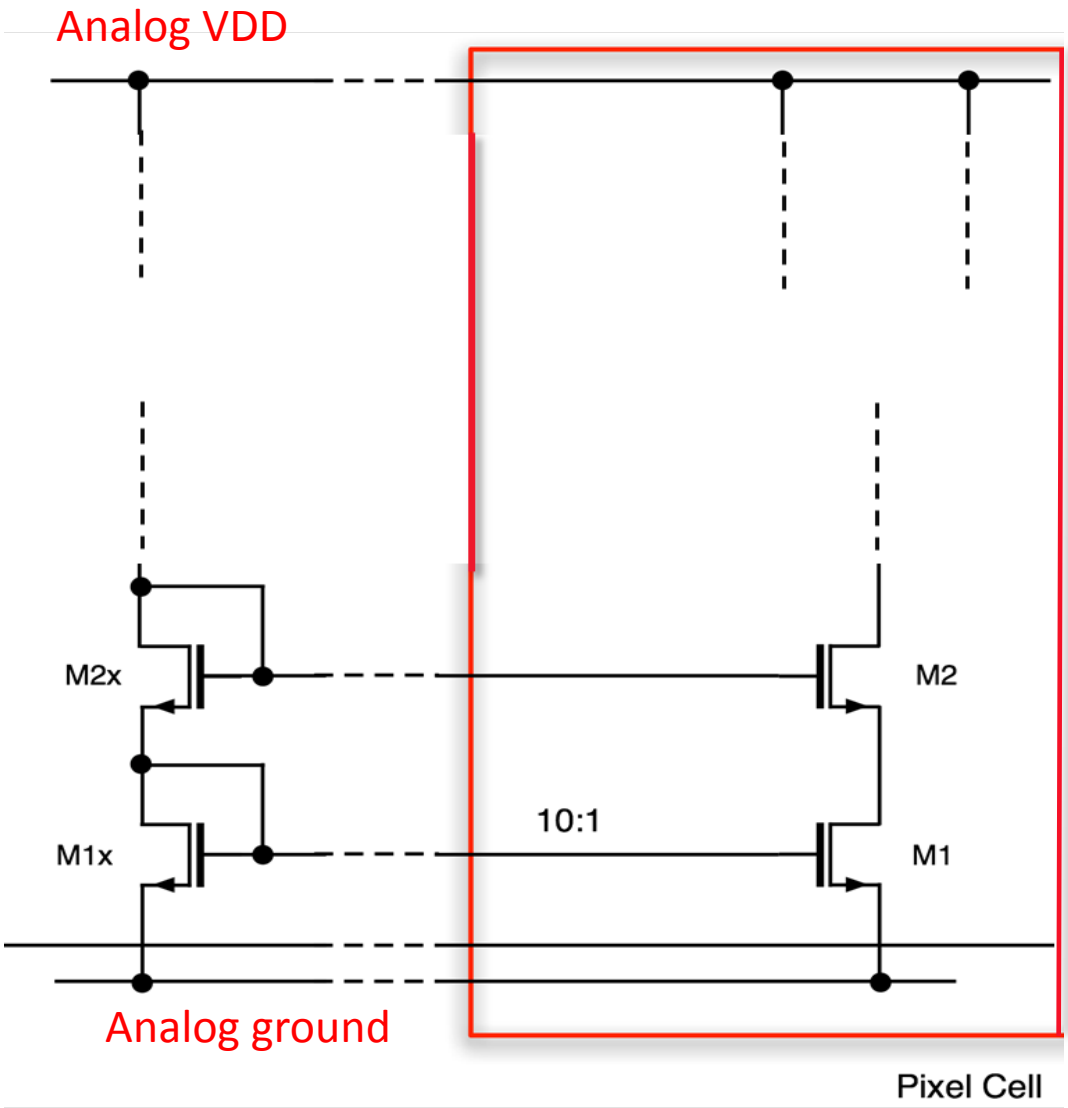
Gate current in 130 nm and LP 65 nm



Gate current and global design of a CMOS readout chip

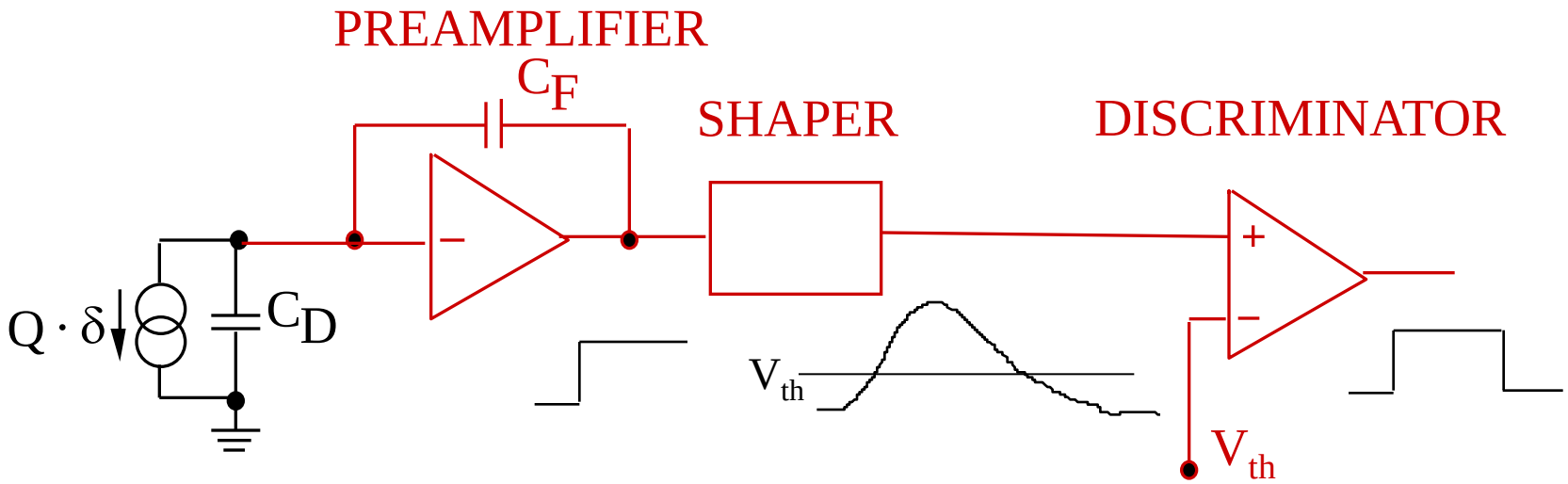
In a single transistor in the LP 65 nm process, the gate current has a very small value (of the order of 100 pA/μm^2 for the NMOS, 10 pA/μm^2 for the PMOS).

However, when bias voltages for analog front-end circuits have to be distributed across a large chip (e.g., a 4 cm^2 pixel readout integrated circuit), the total gate leakage current has to be taken into account in the design of reference voltage and current generators (typically, DACs) and current mirrors.



Extracting a hit information from the sensor signal: the discriminator

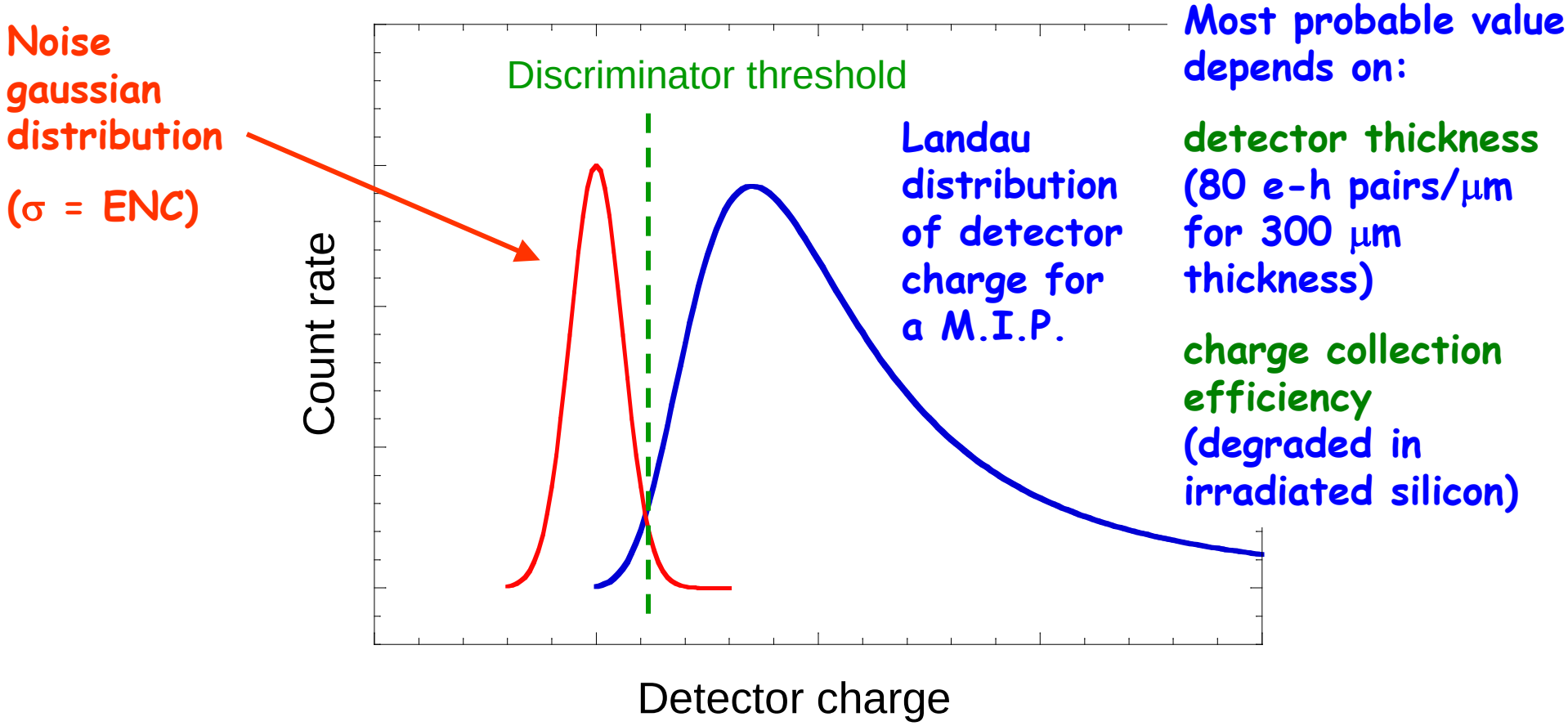
- Binary readout: hit/no hit information from a discriminator
- This can also be associated to an ADC system, providing an information about the charge delivered by the detector



- In a multichannel readout chip, channel-to-channel threshold variations due to device mismatch may degrade detection efficiency and spurious hit rate

Efficiency and noise occupancy

- An excessive threshold dispersion can lead to channels with high noise hit rate or reduced efficiency in signal detection.



- As for the noise, the discriminator threshold and its dispersion (divided by the analog channel charge sensitivity) can be treated in term of input-referred charges, Q_{th} and σ_{qth} respectively.

- For a second-order semigaussian shaper, and series white noise as the dominant contribution to ENC, the frequency of noise hits can be calculated as:

$$f_n = \frac{\sqrt{3}}{\pi t_p} e^{-\frac{Q_{th}^2}{2ENC^2}}$$

- In practical conditions, the number of noise hits can be kept at acceptably low values by satisfying this condition:

$$Q_{th}^{sig} > 4\left(ENC + \sigma_{qth}\right)$$

- To maintain an adequate efficiency, a channel-by-channel threshold adjustment may be necessary (threshold DAC in the pixel cell)

Threshold dispersion

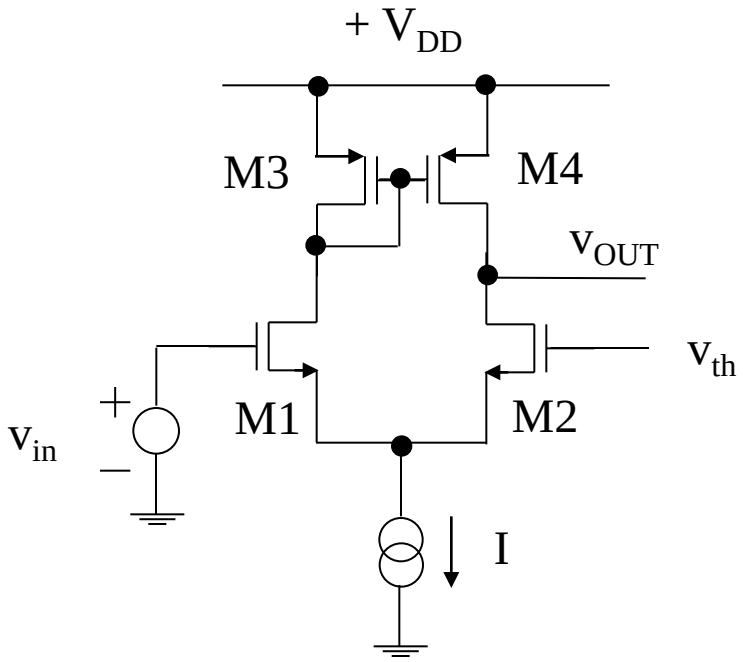
- Discriminator threshold dispersion is given by statistical variations of the threshold voltage of MOSFETs in the differential pairs used in the discriminator input stage:

$$\sigma^2(\Delta V_{th}) = \frac{A_{vth}^2}{WL}$$

$$A_{vth} \div t_{OX}$$

(from stochastic dopant fluctuations model)

=



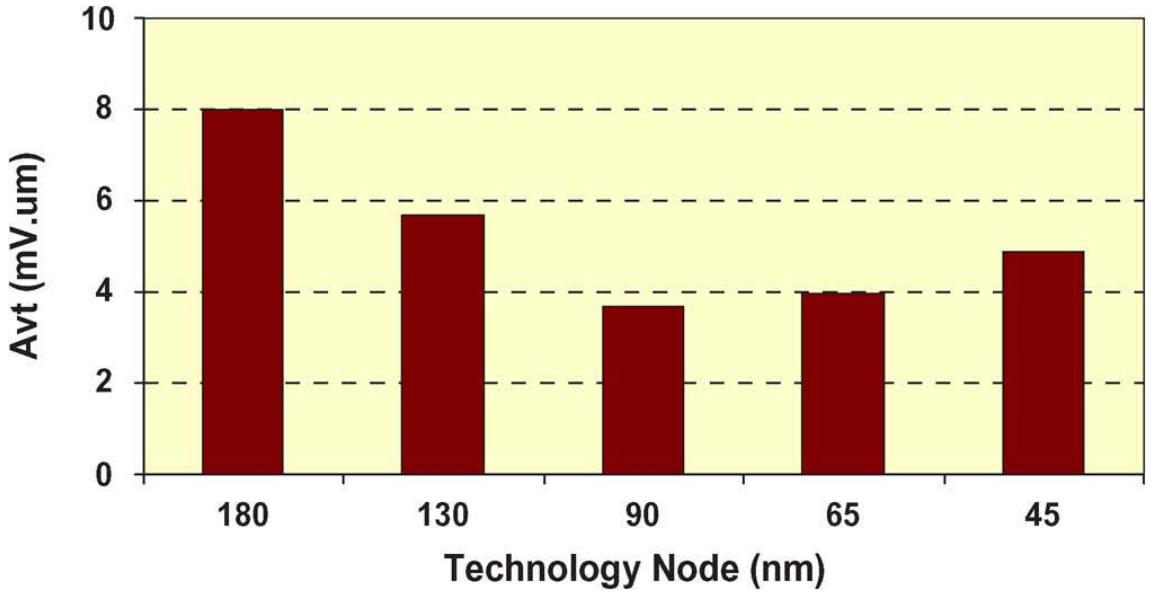
Large area transistors help reduce the effect of threshold mismatch

To maintain an adequate efficiency, a channel-by-channel threshold adjustment is often necessary (threshold DAC in a pixel readout cell)

Threshold dispersion in nanoscale CMOS

Small feature size and low supply voltage increase the impact of variation of transistor properties on chip performance.

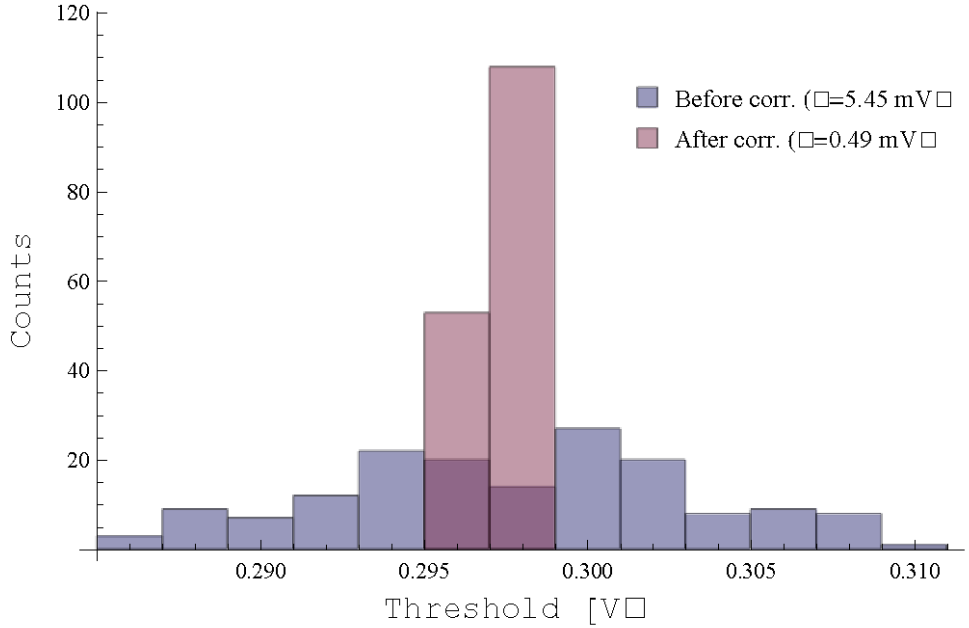
While A_{vth} has the expected reduction with t_{ox} scaling until about 90 nm technologies, in more recent technology generations it has not been decreasing. This could be due to the reduction in t_{ox} scaling, the increase in channel doping required to reduce short channel effects, and the contribution of additional process steps.

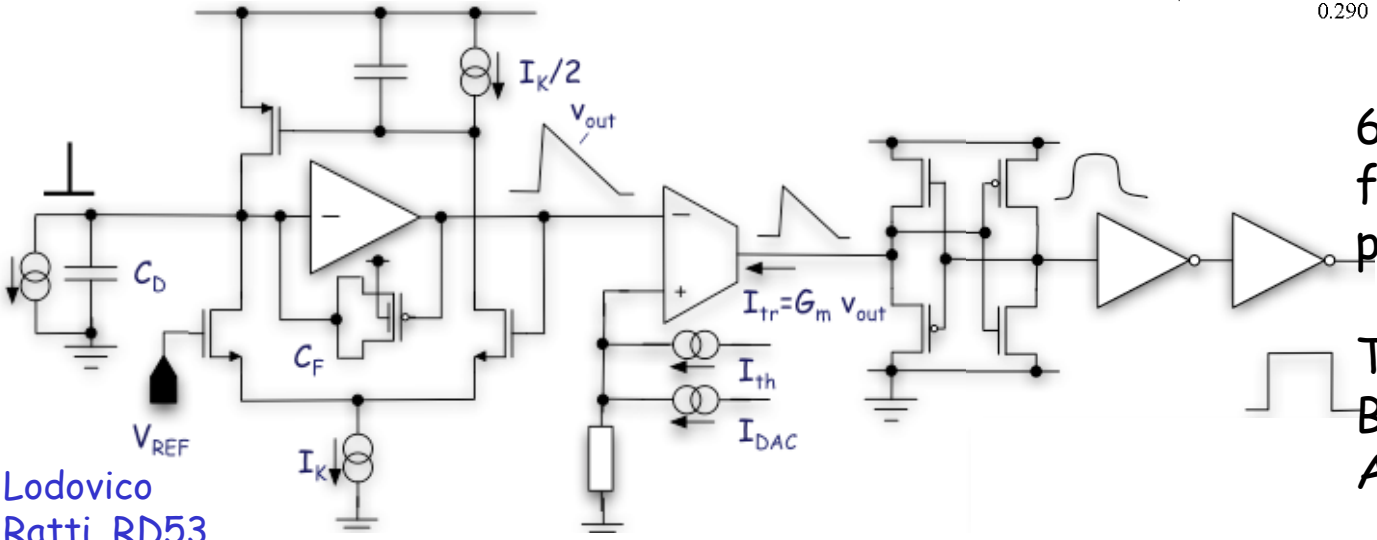


S. Saxena: "Variation in transistor performance and leakage in nanometer-scale technologies", **IEEE Trans. El. Dev**, vol. 55, no. 1, January 2008, pp. 131-144.

Digital calibration of the discriminator threshold

Nonidealities of nanoscale CMOS transistors may be addressed by appropriate design techniques. This includes the digital calibration of analog circuits. In an analog front-end channel, the discriminator threshold may be finely adjusted by a local digital-to-analog converter (DAC).





65 nm continuous-time analog front-end channel for HL-LHC pixels

Threshold dispersion:
Before correction 380 e rms
After DAC correction 35 e rms

Lodovico Ratti, RD53 Analog WG meeting

Specifications for threshold setting, noise, power, speed in the ATLAS/CMS RD53 pixel analog front-end

Need of **detecting charge released by MIPs in heavily damaged sensors (4000 e-) at extreme irradiation levels without efficiency degradation**

- Threshold setting for higher tolerable noise occupancy: $Q_{th,min}$ = 1000 e-
- ENC < 150 e rms at C_D = 100 fF (mostly determined by series noise, ENC is proportional to C_D)
- Threshold dispersion after local tuning: σ_{Qth} < 40 e rms

  (includes contributions by discriminator threshold mismatch and pixel-to-pixel preamplifier gain variations)

These specifications are based on the so called 4σ rule (empirical):

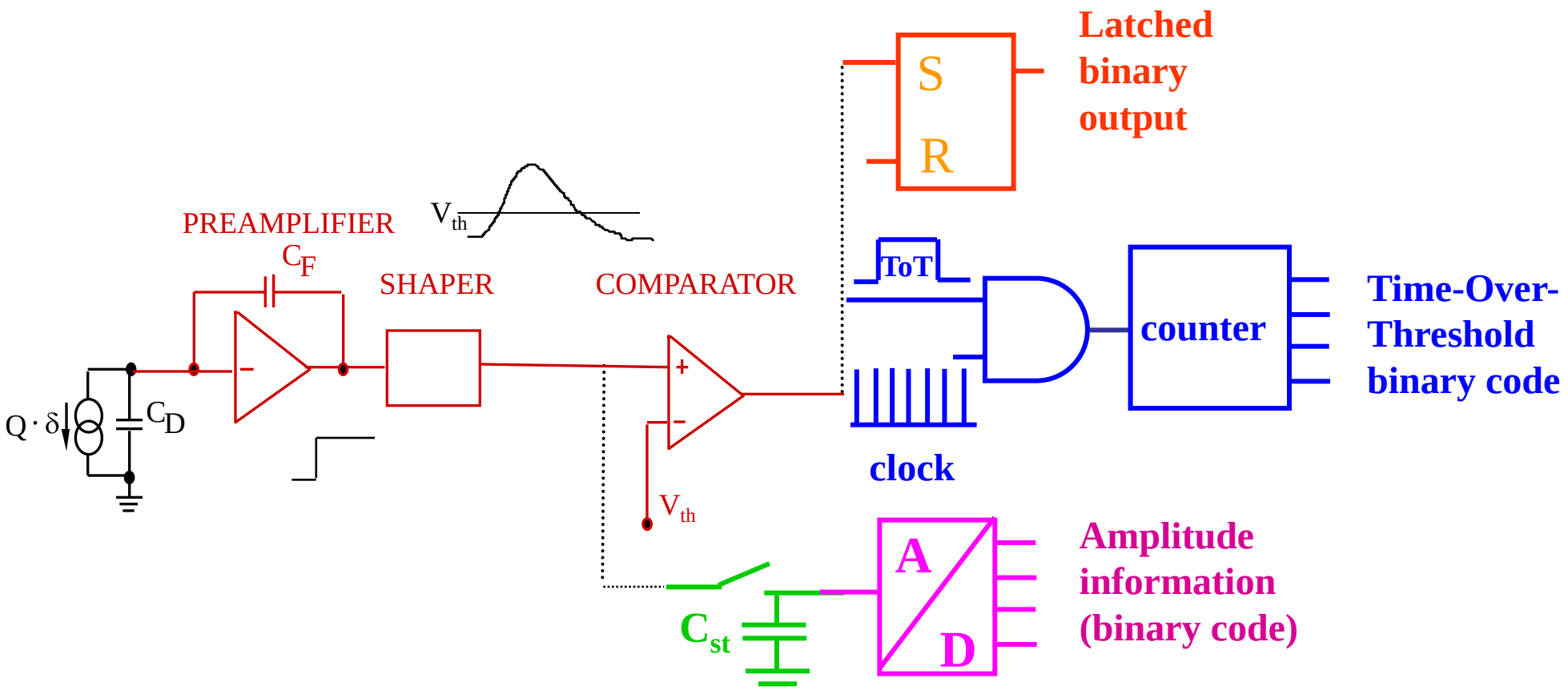
$Q_{th,\,min} > 4 \cdot ENC + 4 \cdot \sigma_{Qth}$. They have to be achieved by analog circuits integrated in **a small silicon area** and operating at **very low power**:

- Maximum power dissipation: 6 μW/pixel

(50 μm x 50 μm, or 25 μm x 100 μm)

- Maximum Hit time resolution = 25 ns, A/D conversion time < 400 ns

Analog-to-digital conversion

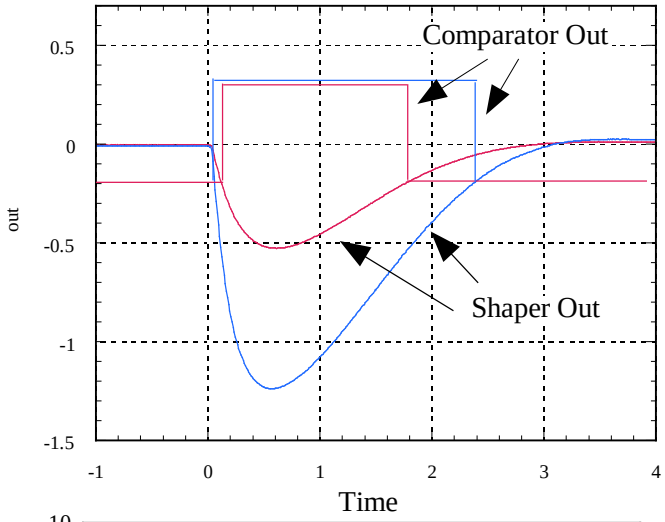


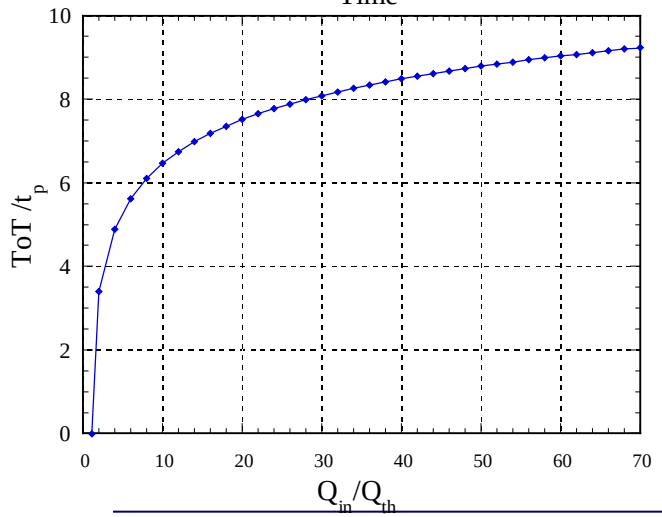
Flash, ramp, SAR,……

Time-Over-Threshold (ToT) analog-to-digital conversion

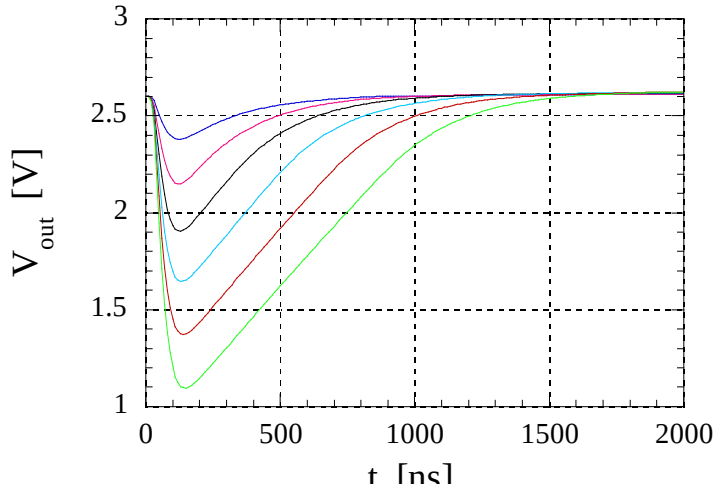
The ADC conversion of ToT is straightforward, avoiding circuit complexity in a chip with a very high functional density.

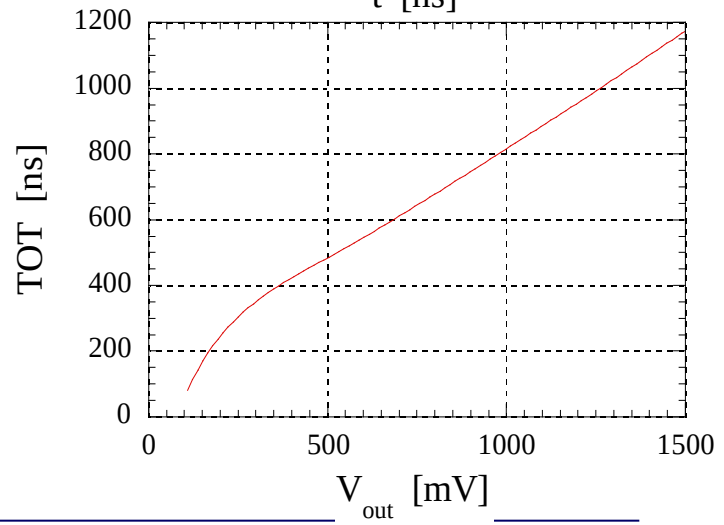
Compression type characteristic



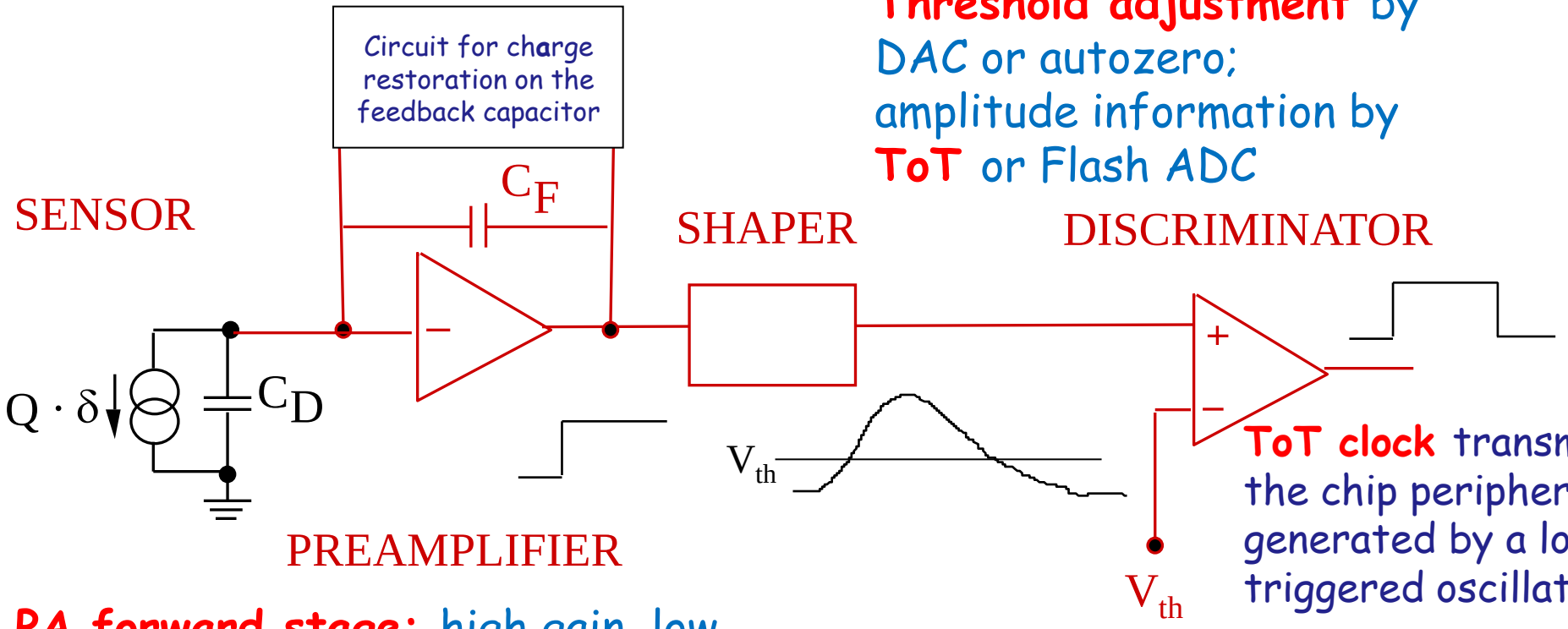


Pseudo-linear characteristic





General schematics of a pixel analog front-end in 65 nm CMOS



Threshold adjustment by DAC or autozero; amplitude information by **ToT** or Flash ADC

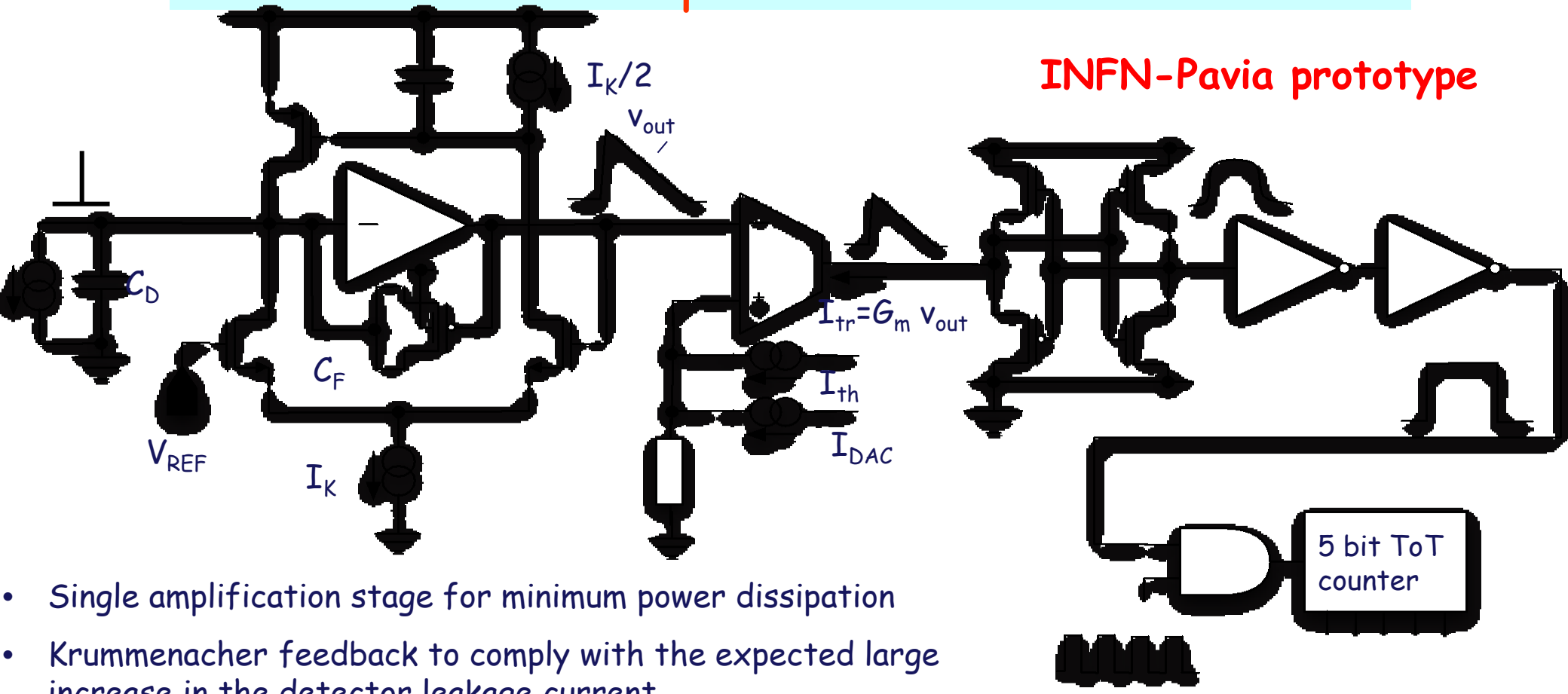
ToT clock transmitted by the chip periphery, or generated by a locally triggered oscillator

PA forward stage: high gain, low noise

Charge restoration: handle sensor leakage current after extreme irradiation levels (max. 20 nA at 2×10^{16} n/cm^2)

Trend is to **skip the shaping stage** in the RD 53 HL-LHC pixel cell, mostly because of power constraints (noise filtering in the preamplifier, or correlated double sampling)

How the analog channel may look like in a HL-LHC pixel readout



- Single amplification stage for minimum power dissipation
- Krummenacher feedback to comply with the expected large increase in the detector leakage current
- High speed, low power current comparator
- Relatively slow ToT clock – 80 MHz
- 5 bit counter – 400 ns maximum time over threshold
- 30000 electron maximum input charge, ~450 mV preamplifier output dynamic range
- Selectable gain and recovery current

Analog design in nanoscale CMOS for advanced pixel detectors

- The **Analog Working Group of RD53** is studying the best design choices for the front-end stages, that interface the silicon sensor with the large and mostly digital readout chip

- Severe constraints for noise, power, speed, silicon area, immunity to digital interferences have to be taken into account in the classical analog blocks (preamplifier, discriminator,…)

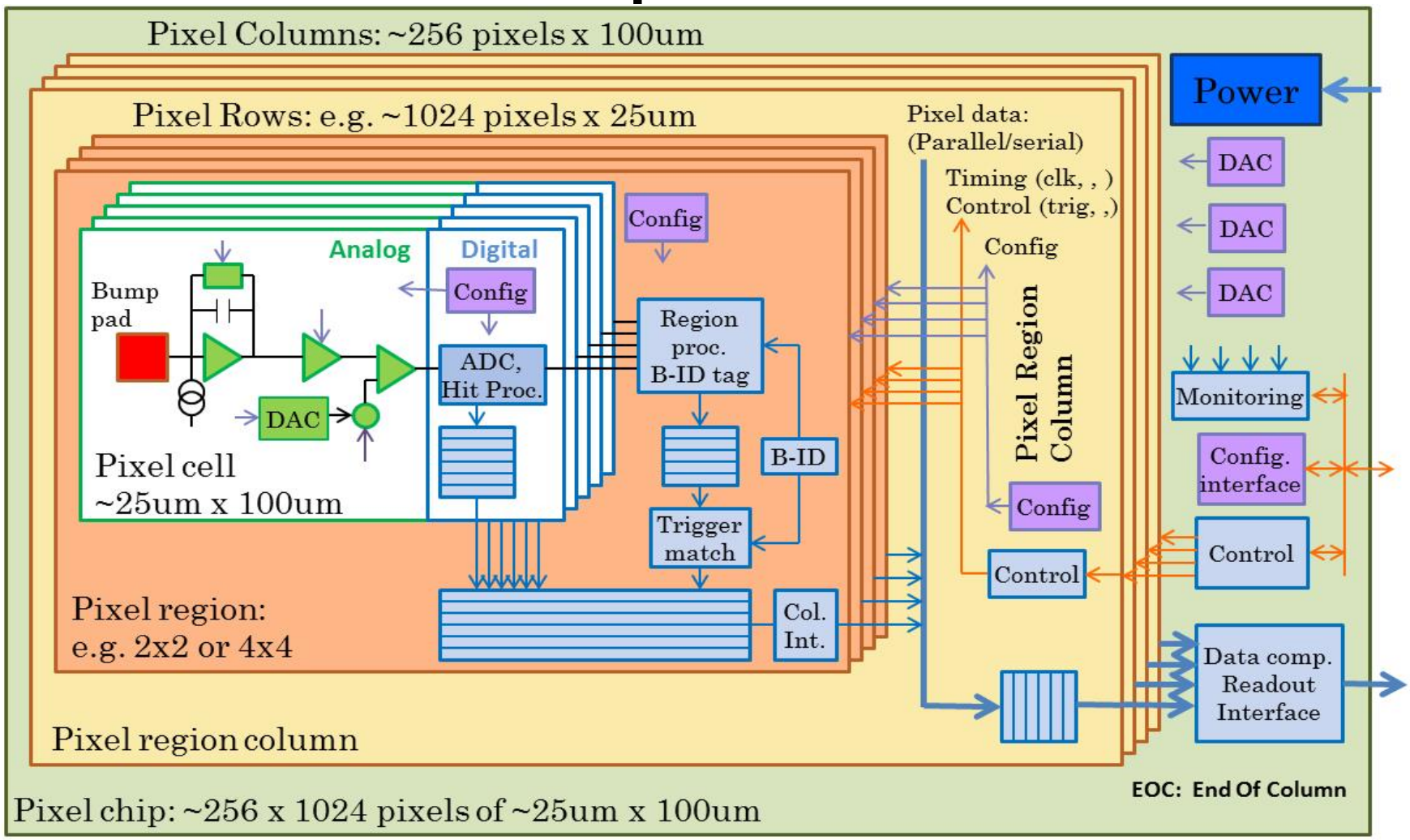
- Scaling of transistor size to the nanometer region also stimulates analog design ideas that depart from the usual schemes of pixel front-end circuits, going beyond a simple translation of old schematics into a more advanced technology. Among them:

  - ❑ Switched-capacitor techniques
    (avoid local tuning in the pixel cell and additional hardware)
  - ❑ Fully differential architectures
    (increase immunity to interferences)
  - ❑ Current-mode schemes
  - ❑ …..

Readout architecture

- Digital information of hit signals is further processed by circuitry associated to each pixel (strip) and at the chip periphery. Position (pixel or strip address), timing (time stamp) and possibly pulse amplitude (from ADC) information must be provided.
- All architectures perform data sparsification, processing only data from channels where the signal exceeds the discriminator threshold
- Often, a trigger system selects only a fraction of the events for readout, reducing the data volume sent to the DAQ. In this case, information for all hits must be buffered for some time, waiting for a trigger signal (delay of several μs).
- Triggerless (data push architectures) are also available. All hits are read out immediately (as long as the rate is not too high). This allows the tracker information to be used for Level 1 Trigger

RD53 chip architecture

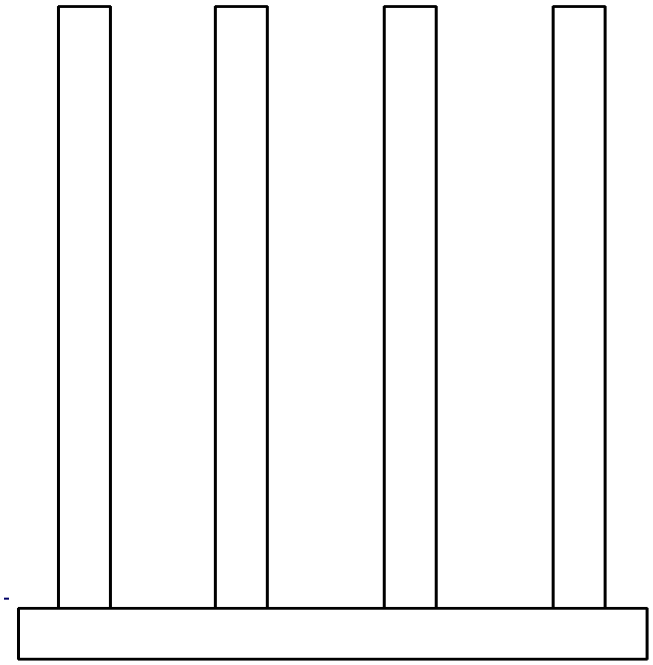


- 95% digital (as for FEI4)
- Charge digitization (TOT or ADC)
- ~256k pixel channels per chip
- Pixel regions with buffering
- Data compression in End Of Column
- Chip size: >20 x 20 mm^2

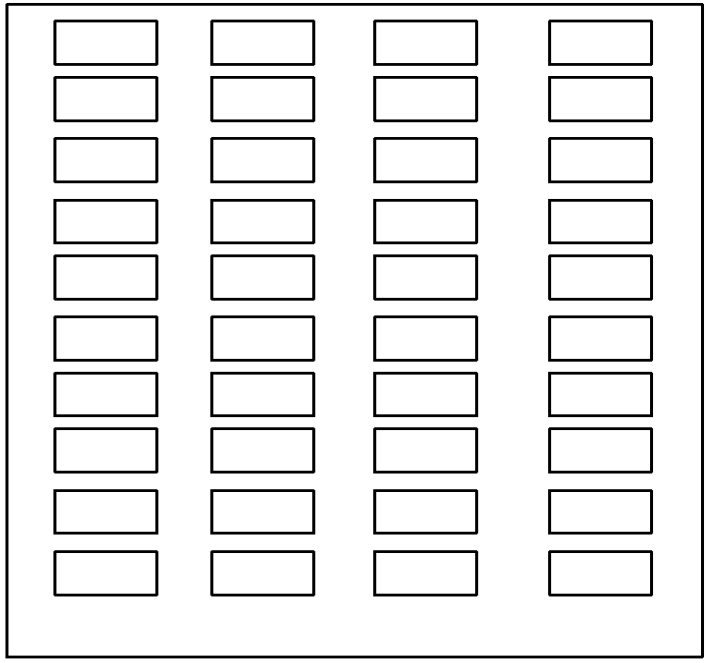
Matrix architecture: digital columns or digital cores?

- It may be best to distribute a large amount of digital logic into the pixel array and to store hits locally.
- The desire to minimize chip periphery area is likely to mean that chip design will be based on a distributed architecture as in FE-I4, which implies a tight and delicate integration with the low noise analog front-ends per pixel

Does digital array look like this?



Or like this?



Digital cores

- Just as for digital columns, digital cores can be subdivided into regions for hit and latency memory sharing.
- The core size has nothing to do with the optimum region size. The cores size has to do with
  - On-chip hit movement (bandwidth) and clock distribution
  - Power consumption
  - Layout constraints

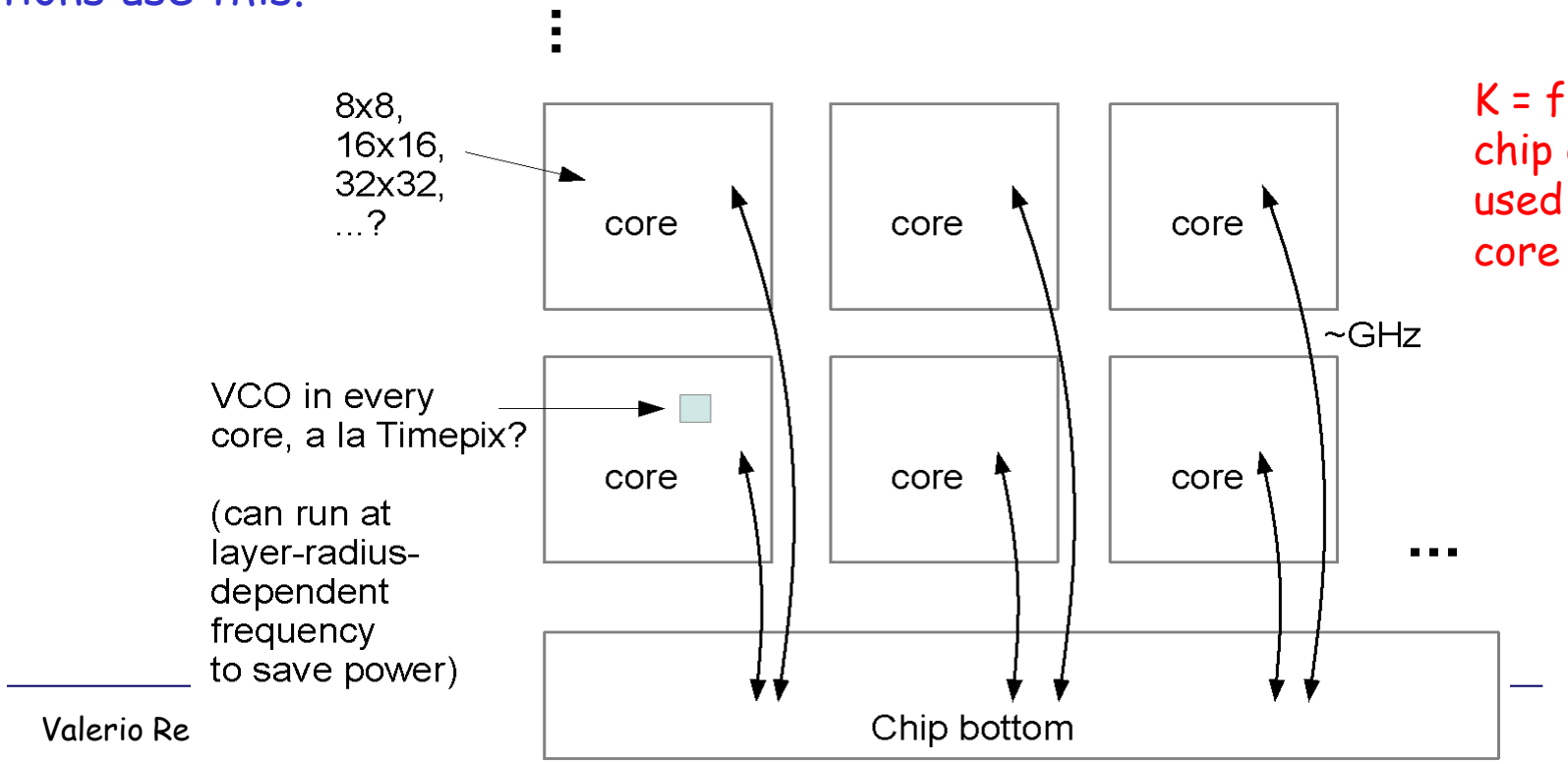
- How is each core connected to the bottom of chip?

  Point-to-point?
- What bandwidth and latency does this connection need?

  need physics simulation to understand readout efficiency vs. core BW
- Do cores talk directly to each other?
- What is the optimum core size (in pixel rows x pixel columns)?

Moving data at low power

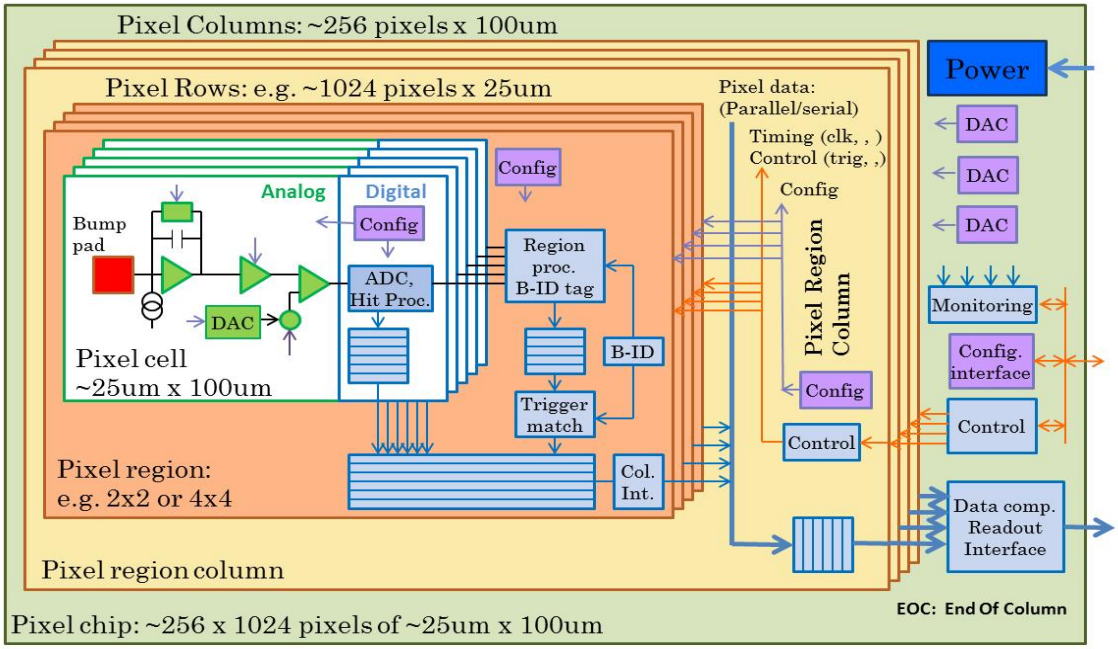
- An important way to help low threshold may be to keep power constant
- This would imply using a DC-balanced protocol for moving data out of the cores
- So the power is constant regardless of now much data there is. This results in some parameters to optimize
- Total power in core->bottom links =
  power per transition * (chip output BW * K) * number of cores
- Low power per transition wants low voltage differential. Lots of memory/CPU applications use this.



K = fraction of chip output BW used by a single core

System aspects of mixed-signal design of large chips in 65 nm CMOS

- The local adjustment of the threshold in a pixel cell is an example of design strategies that can be adopted to overcome limitations associated with small area devices.
- Generally speaking, the design of a complex microelectronic system such as a 65 nm pixel readout chip has to devise solutions that allow analog and digital circuits to work together in a same substrate, complying with low noise and low power dissipation requirements (among others)



The devil is in the details: integration of mixed-signal functions in a multichannel chip

- Non-ideal effects may degrade the performance of low-noise analog circuits in strip or pixel readout integrated circuit:

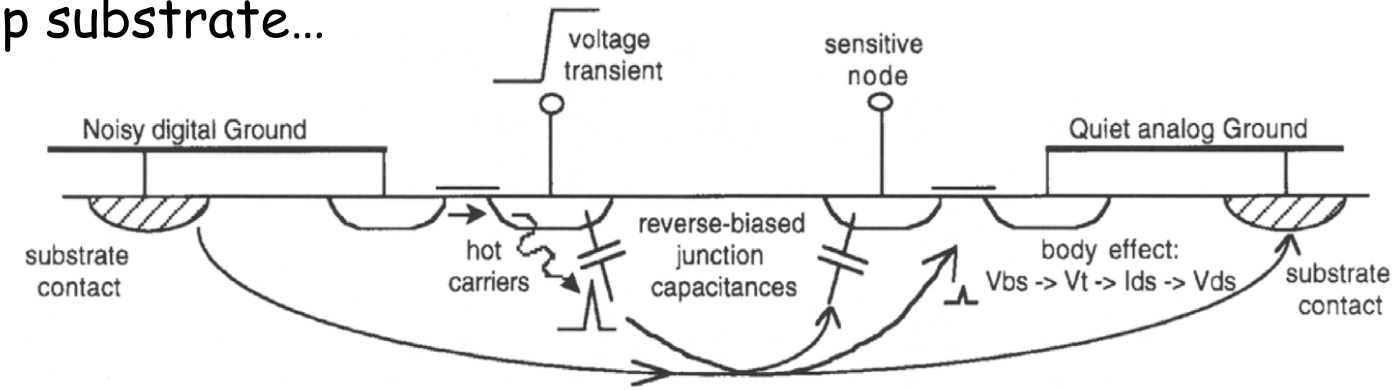
  Interferences from digital signals (clocks, commands, readout lines)

  Voltage drops on interconnections distributing power supply voltages across large area chips

  .....

Digital-to-analog interferences

- There are many ways by which interferences can propagate through the chip substrate…



- These effects can be mitigated by using differential, low-level digital signals (LVDS), by isolation techniques, by layout tricks (separated analog and digital power and signal routing) …

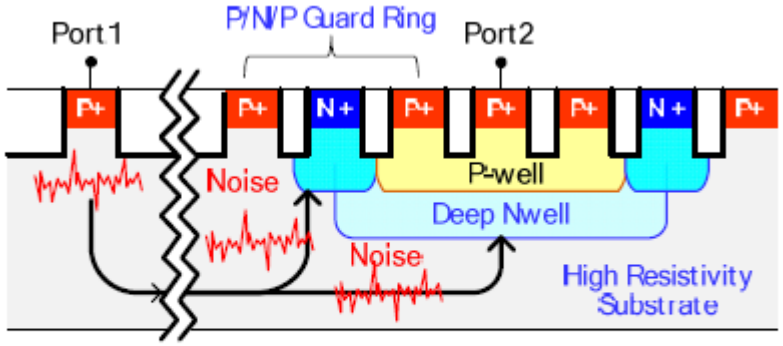


Figure 6: Illustration of the three main isolation techniques; PNP guard ring, deep NWELL and high resitivity substrate.

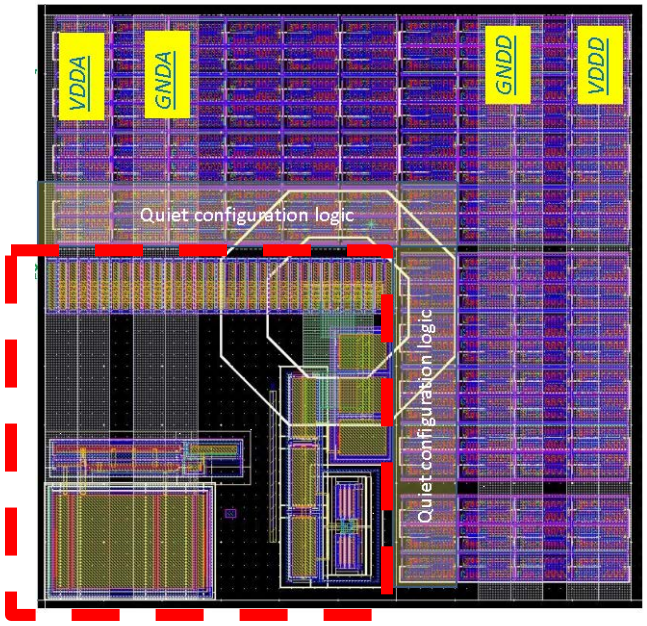
I. A. Young (Intel), "Analog mixed –signal circuits in advanced nanoscale CMOS technologies for microprocessors and SoC", 2010 ESSCIRC

Integration of the analog readout cell in a (mostly digital) pixel matrix

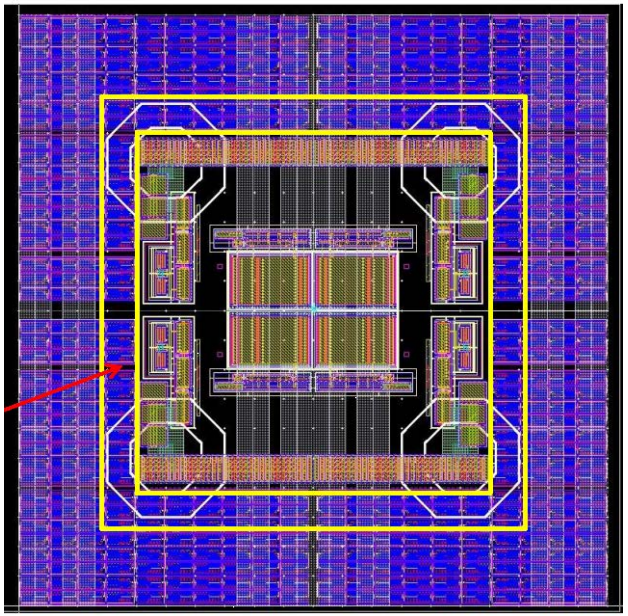
- In a 50x50 µm^2 pixel cell, a correct layout is crucial to avoid digital interferences in the low-noise analog front-end

- The "analog island" concept (A.Mekkaoui, LBNL)

50 µm x 50 µm pixel cell



Four 50 µm x 50 µm pixel cells



Analog section

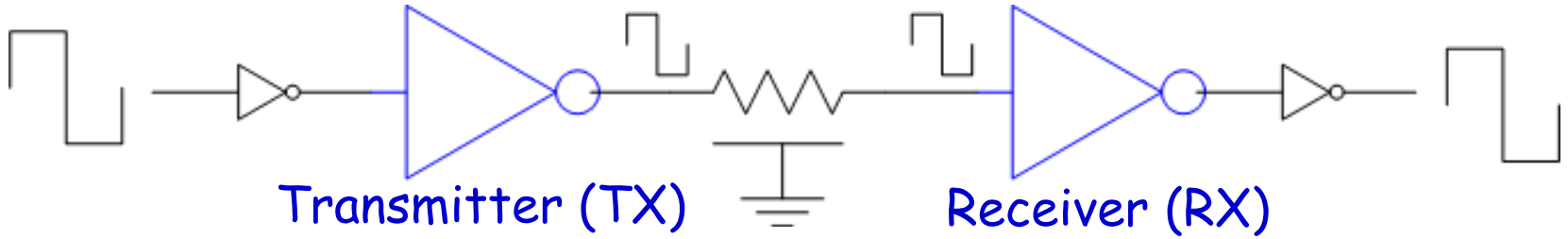
Clock distribution across a large 65 nm chip

Clock distribution is a crucial aspect of modern multi-GHz microprocessor design. A single clock source is fed through hierarchies of clock buffers to eventually drive almost the entire chip.

Increasing process and device variations in deep sub-micron semiconductor technologies further adds to timing uncertainties known as clock skews.

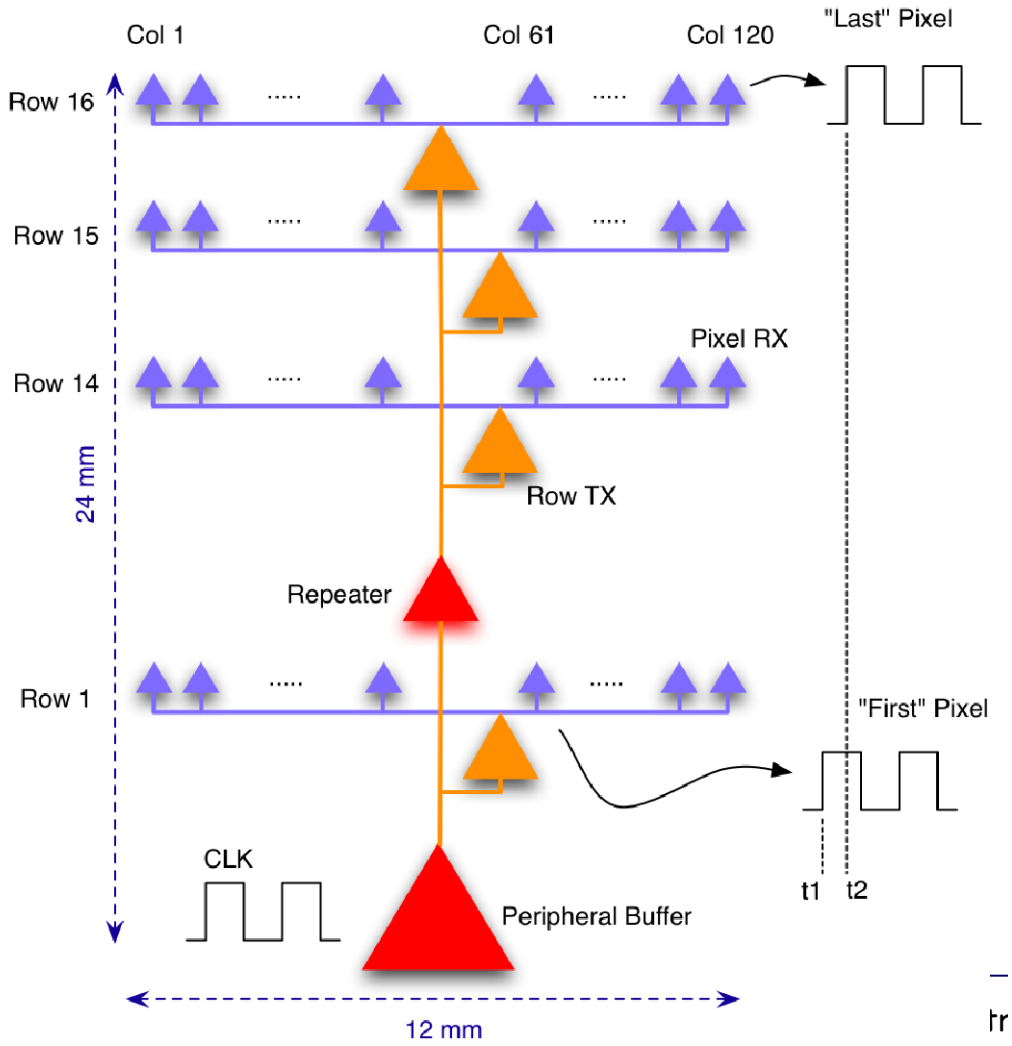
The load of the entire chip is substantial, and sending a high quality clock signal to every corner of the chip can be very expensive in terms of power dissipation.

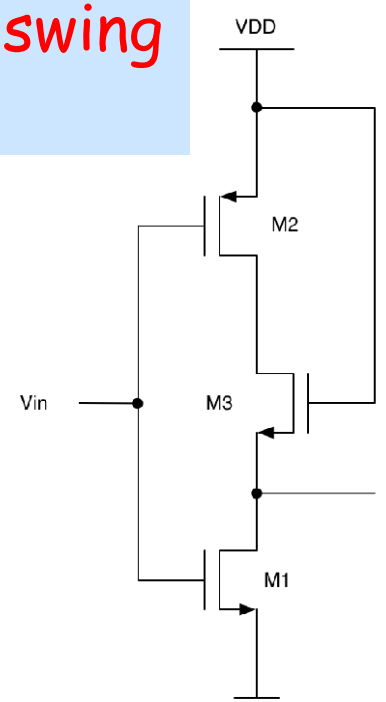
This could be seen as a typical "mixed-signal" problem, that can be tackled by distributing an "analog-like" signal (aka a reduced swing digital signal) that is then regenerated close to the pixel readout cells.



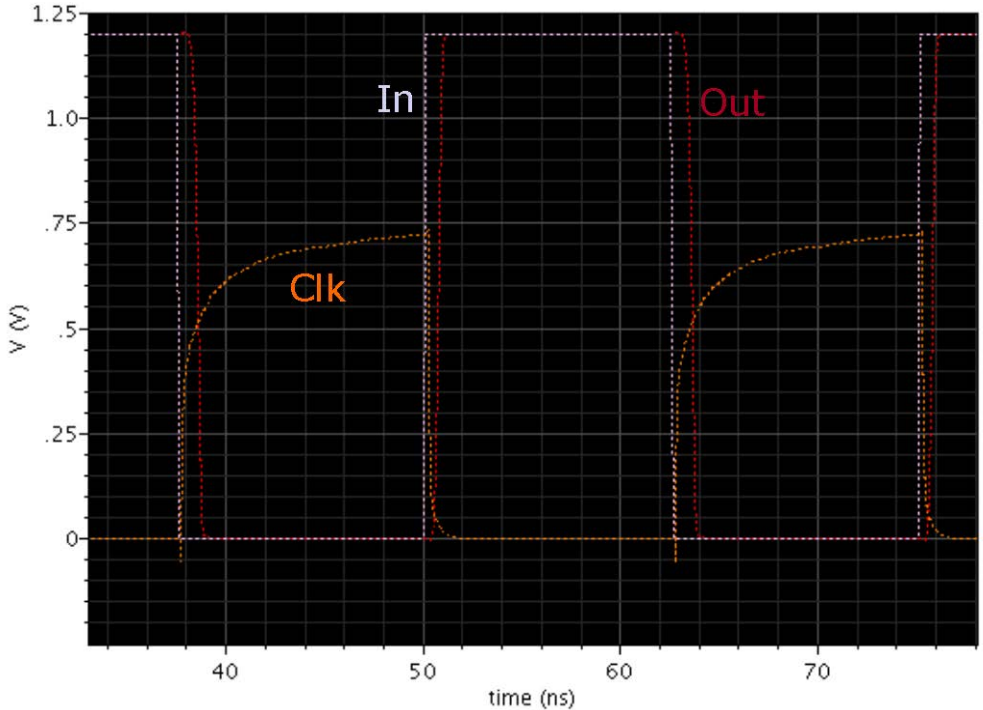
$$E_{bit} = C V_{DD} V_{swing}$$

Row distribution scheme of a reduced swing clock signal in a pixel readout matrix





Reduced swing clock transmitter (requires an asymmetric inverter in the receiver to regenerate nominal digital voltage levels)



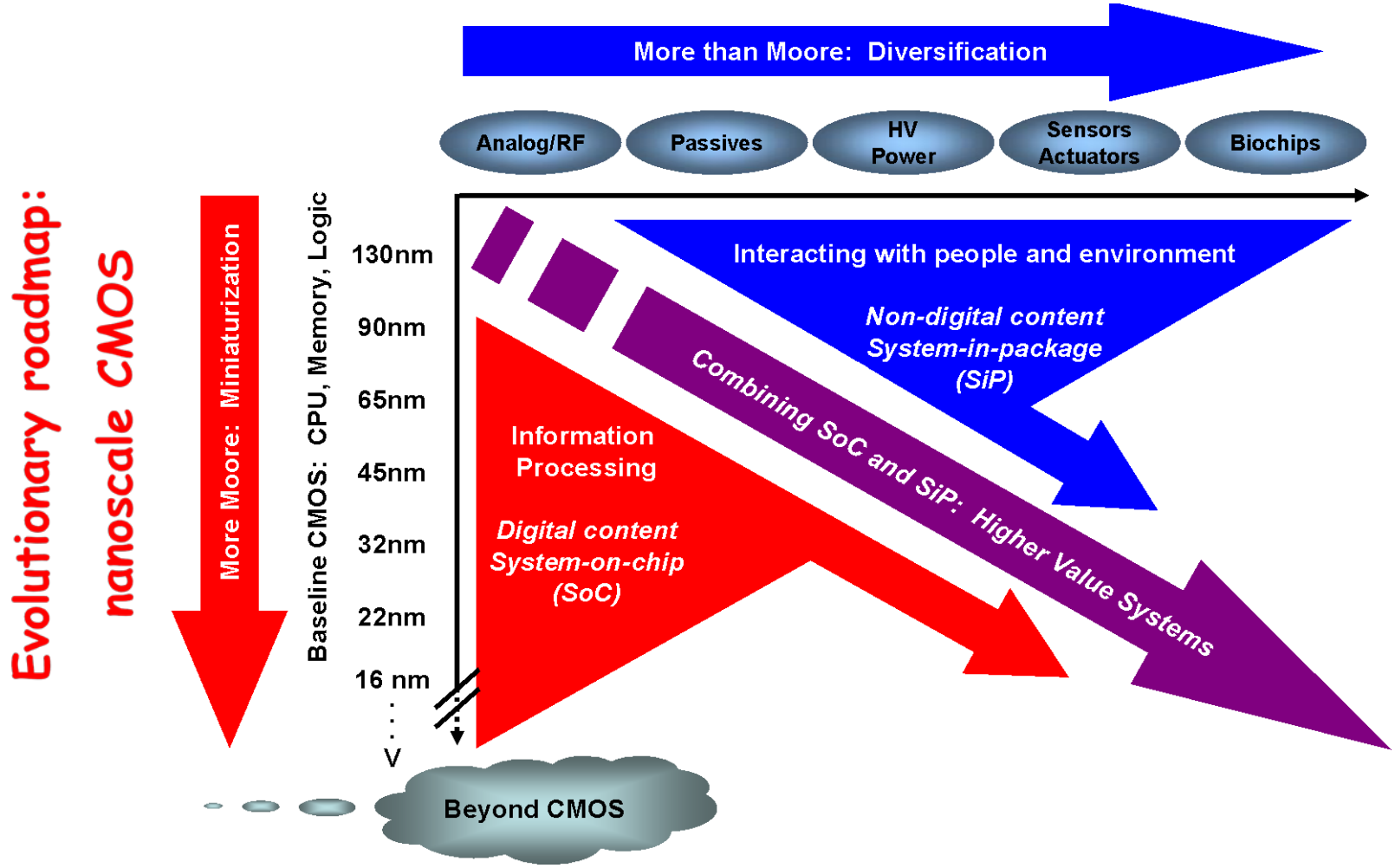
RD53 plans

- The challenge of analog design in 65 nm CMOS for pixel readout at the HL-LHC is being tackled by our community of microelectronic designers in a collaborative way (RD53, the INFN project CHIPIX 65)
- Current goal: developing/qualifying technology, tools, architecture and building blocks required to design next generation pixel chips for very high rates and radiation
- First round of submissions in 2014, first prototypes currently being characterized, other submissions in Spring 2015
- These prototypes will allow us to evaluate the best solutions for the architecture of the analog front-end

➔ **Full pixel chip prototype in 2016**

Evolution of microelectronic technologies

No roadmap, room for new ideas:
monolithic sensors, 3D integration



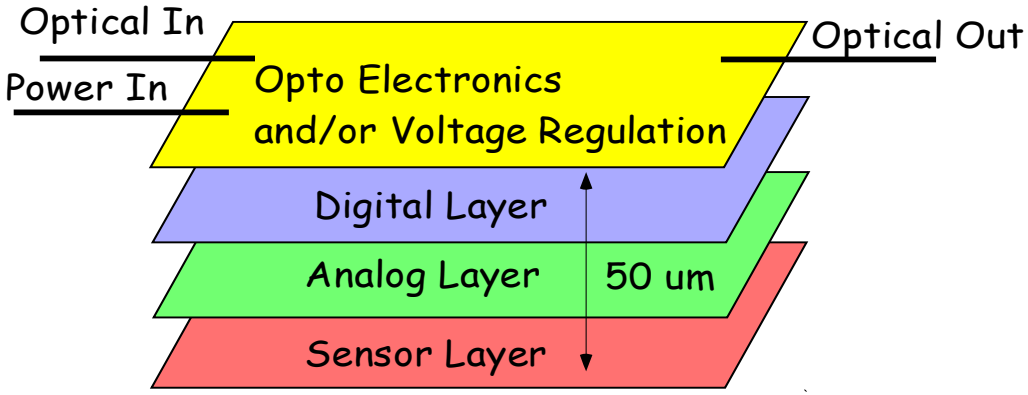
A further leap: 3D integration

- The microelectronic industry has developed 3D integration of **thinned and bonded CMOS tiers with vertical interconnections through the silicon layers** as an alternative or complementary way to device scaling, with the goal of enhancing memory capacity and microprocessor speed (by reducing length of interconnections) and of improving the performance of image sensors (by including pixel level high-speed signal processing).

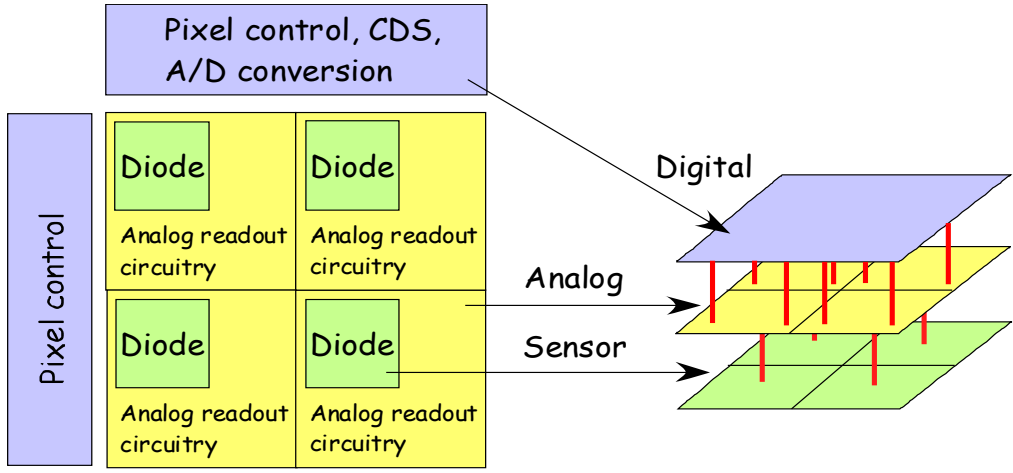
- The semiconductor detectors and front-end electronics communities in HEP and photon science plan to take benefit from 3D integration for **new pixel sensor with advanced functionalities, smaller form factor, less material and dead area, separation and optimization of sensing, analog and digital functions,…** New concepts may also be enabled by this technology.

3D vertical integration

- A "3D" chip is generally referred to as a chip comprised of 2 or more layers of active semiconductor devices that have been thinned, bonded and interconnected to form a "monolithic" circuit.
- Often the layers (sometimes called tiers) are fabricated in different processes.
- Industry is moving toward Vertical Integration to improve circuit performance.
  - Reduce R, L, C for higher speed
  - Reduce chip I/O pads
  - Provide increased functionality
  - Reduce interconnect power and crosstalk
- This is a major direction for the semiconductor industry.



Physicist's Dream



Conventional MAPS 4 Pixel Layout

3D 4 Pixel Layout

Prospects and support to 3D integration in the pixel detector community

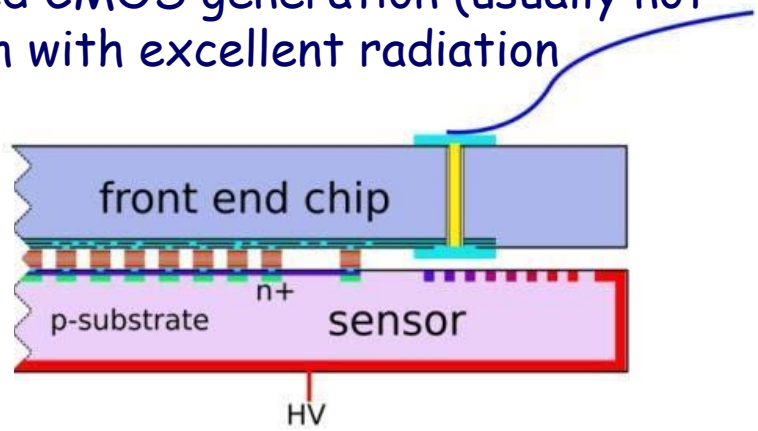
3D integrated circuits based on homogeneous layers (same CMOS technology) and high density TSVs and interconnections remain a very promising approach to advanced pixel detector readout and other applications.

The **AIDA WP3** project supported the less aggressive **"via last"** variant of 3D integration, where **low-density TSVs are etched in fully processed CMOS wafers**. It is a mature technology, presently available at various vendors.

This technique makes it possible to **use heterogeneous layers** (different technologies) for sensors and front-end electronics and to fabricate four-side buttable devices with minimal dead area.

A high-resistivity, fully depleted sensor can be combined in a low-mass assembly with a readout chip designed in an aggressively scaled CMOS generation (usually not available in the typical MAPS "Opto"processes), both with excellent radiation hardness (among other properties).

Low-density peripheral TSVs can be used to reach backside bonding pads for external connection. The interconnection technology can be chosen according to the pixel pitch.



Conclusions

- Front-end electronics for silicon trackers in future experiments is an exciting challenge for integrated circuit designers
- Classical analog problems (signal amplification and shaping, noise, threshold dispersion) will require clever solutions
- New industrial technologies (nanoscale CMOS, 3D integration, …) will be exploited to achieve increasingly demanding specifications
- 3D integration is progressing in the microelectronic industry, and we have to be ready to exploit it. Ultimately, it may allow designers to avoid using sub-50 nm processes for analog and digital circuits in very small pixel readout cells.
- R&D activities in these technologies have to be supported by our community, since they enable new concepts for detector systems. The new **AIDA-2020 project** will provide this support for 65nm CMOS (with CERN support) and for 3D integration.
- Technology watch for novel devices and processes has to continue, since the evolution of microelectronics is not going to end soon.

References

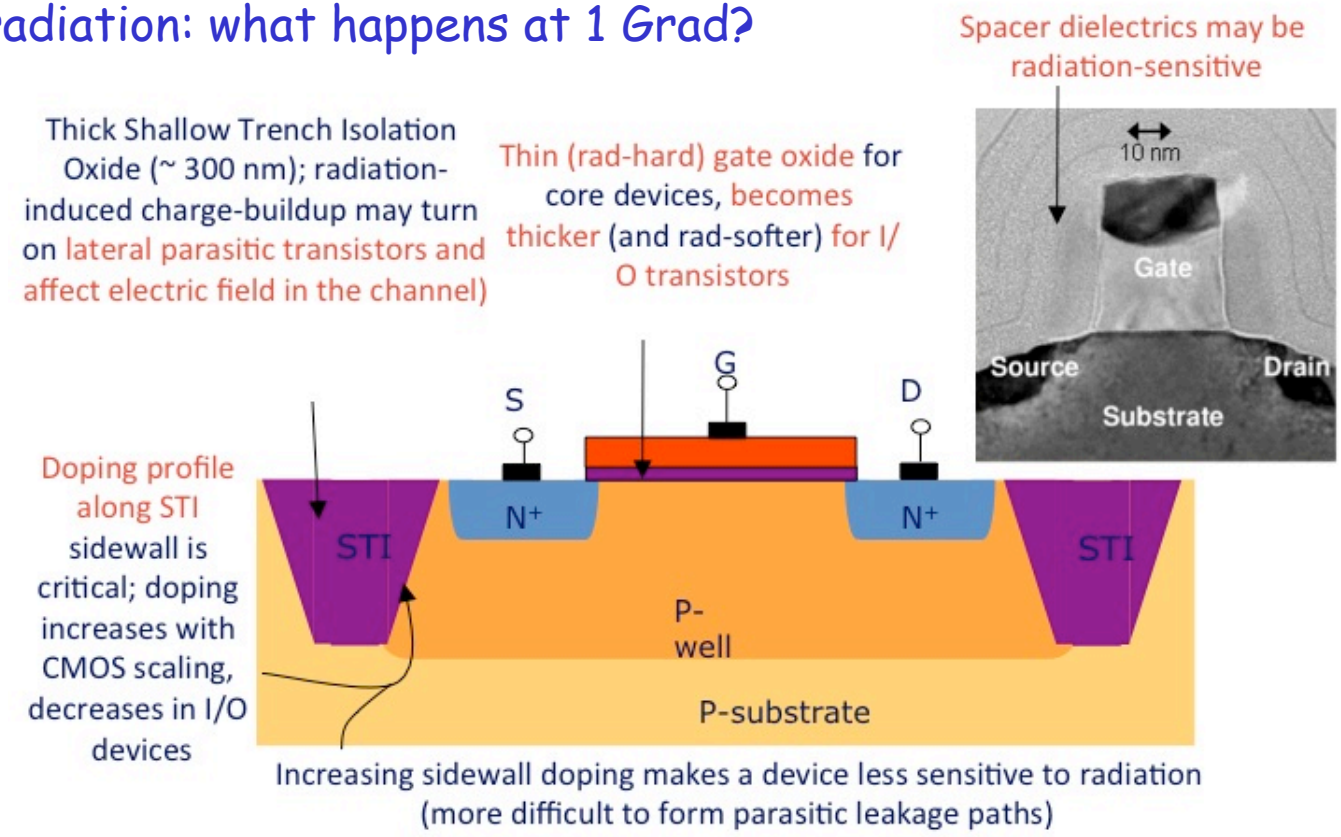
- E. Gatti, P.F. Manfredi: "Processing the signals from solid-state detectors in elementary-particle physics", La Rivista del Nuovo Cimento, 1986
- V. Radeka: "Low-noise techniques in detectors", Ann. Rev. Nucl. Part. Sci., 1988
- G. Lutz: "Semiconductor radiation detectors"
- L. Rossi, P. Fischer, T. Rohe, N. Wermes: "Pixel Detectors. From Fundamentals to Applications"
- H. Spieler: "Semiconductor detector systems"

Backup slides

65 nm CMOS at extreme radiation levels

- At the HL-LHC design luminosity, for an operational lifetime of 10 years, the innermost pixel layer will be exposed to a total ionizing dose of 1 Grad, and to an equivalent fluence of 1-MeV neutrons of 2 x 10^{16} n/cm^2.
- If unacceptable degradation, **a replacement strategy** must be applied for inner pixel layers.
- Nanoscale CMOS (with very thin gate oxide) has a large intrinsic degree of tolerance to ionizing radiation: what happens at 1 Grad?

- Radiation induced electric charge is associated with thick lateral isolation oxides
- CMOS is usually not affected by non n-ionizing radiation: what happens at 2 x 10^{16} n/cm^2?

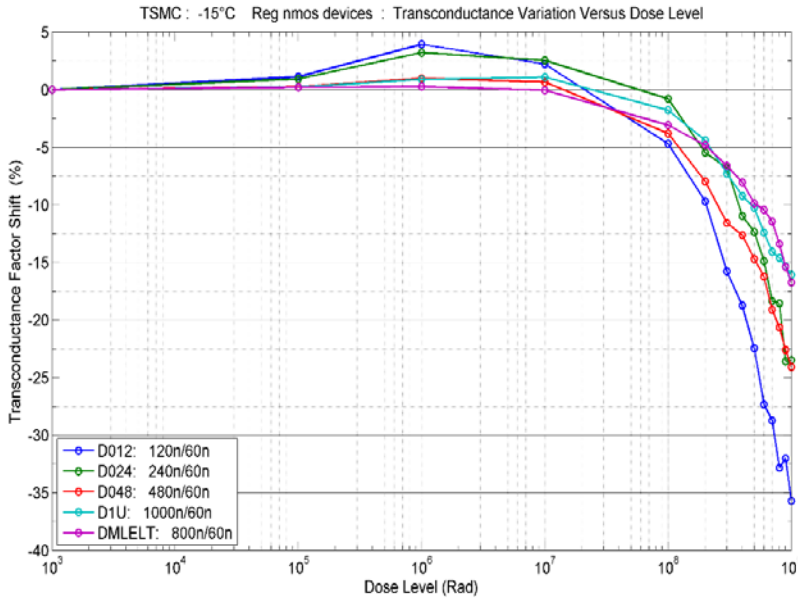


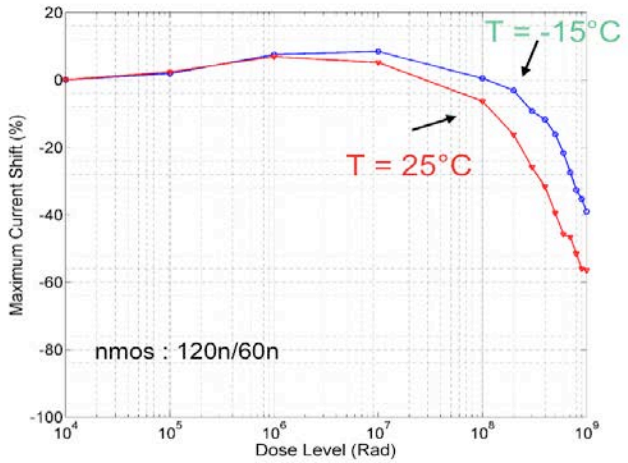
The analog front-end at extreme total ionizing dose

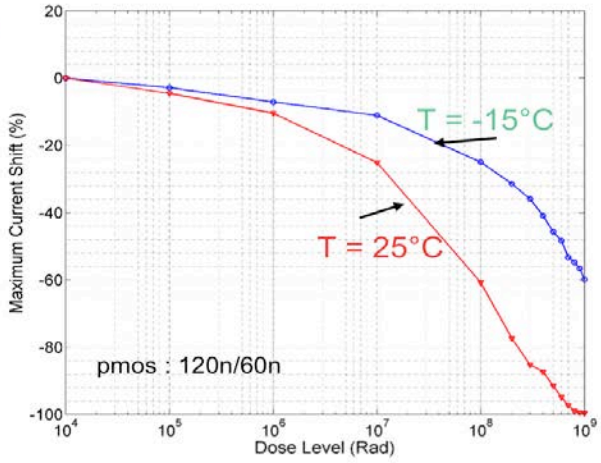
- Among other effects, PMOSFETs (especially minimum size ones) show a large transconductance degradation, which becomes very steep over 100 Mrad (partial recover after annealing)
- This is probably not so critical for the design of analog blocks, where minimum size transistors can be avoided if necessary; the study of radiation effects on noise is ongoing
- Damage mechanisms have yet to be fully understood; they appear to be **less severe at** the foreseen operating temperature of the pixel detector at HL-LHC (**about - 15 °C**)

CPPM data with X-rays

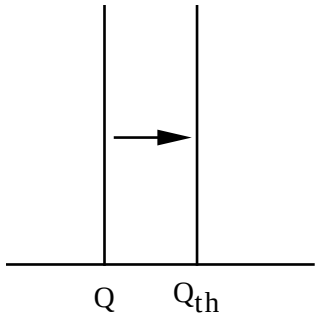
(Fermilab and Padova studying radiation effects with other sources)



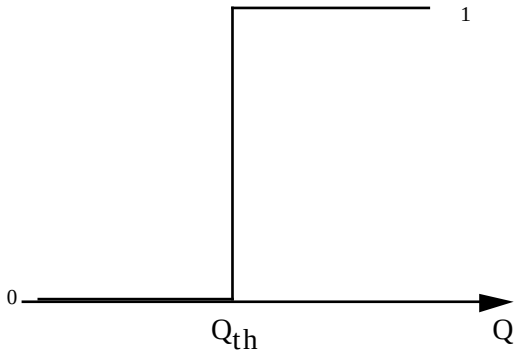


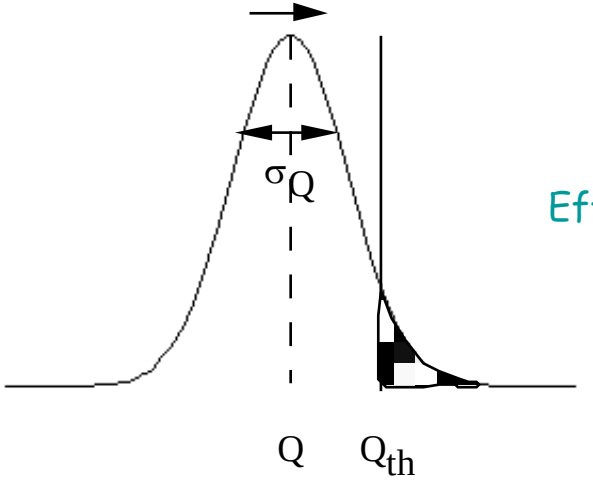


Effect of noise on discriminator firing efficiency

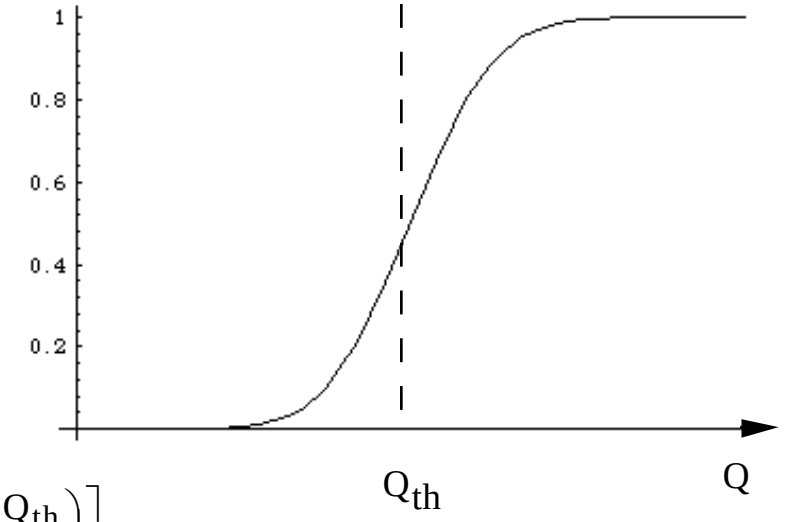


Noiseless system





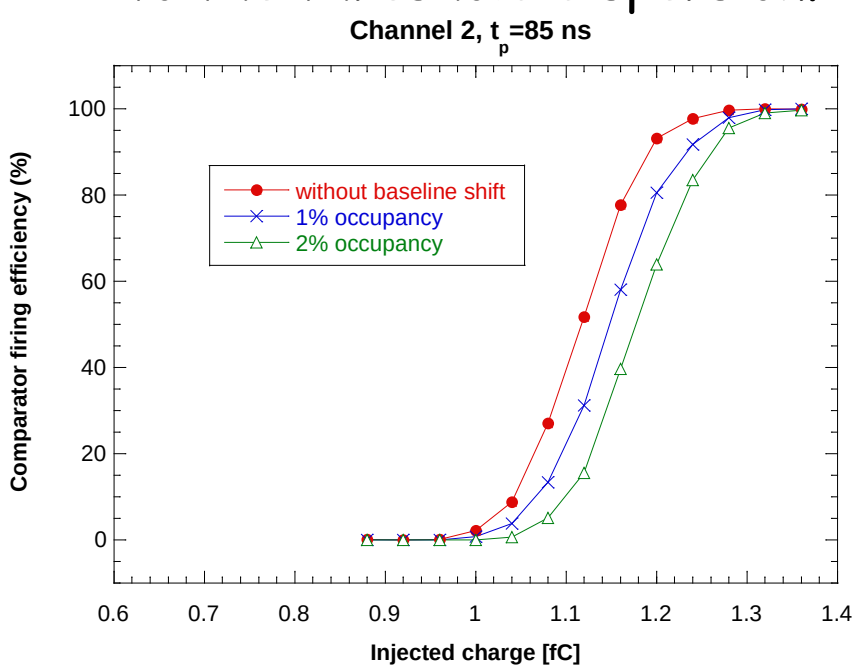
Effect of noise



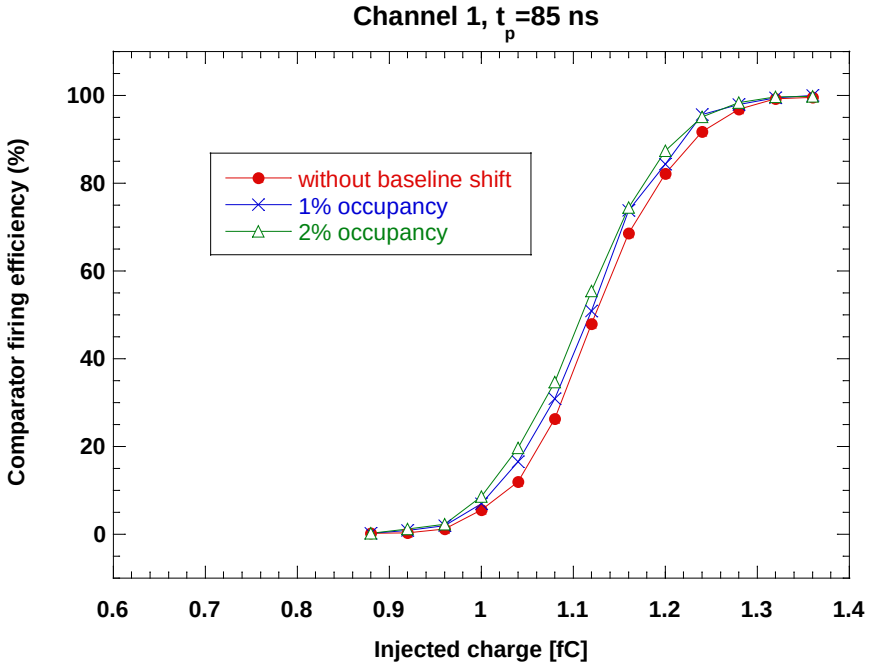
$$P(Q_{th}) = \int_{Q_{th}}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_Q} \exp\left[-\frac{(q-Q)^2}{2\sigma_Q^2}\right] dq = \frac{1}{2}\left[1 + \mathrm{Erf}\left(\frac{Q-Q_{th}}{\sqrt{2}\sigma}\right)\right]$$

Processing the signal from the sensor: the baseline restorer

- Since the signal at the preamplifier output is not an ideal voltage step, but returns to baseline with a long time constant, the signal at the shaper output has a long tail. This results in a baseline shift at the discriminator input, with related statistical fluctuations, adding to the threshold dispersion.

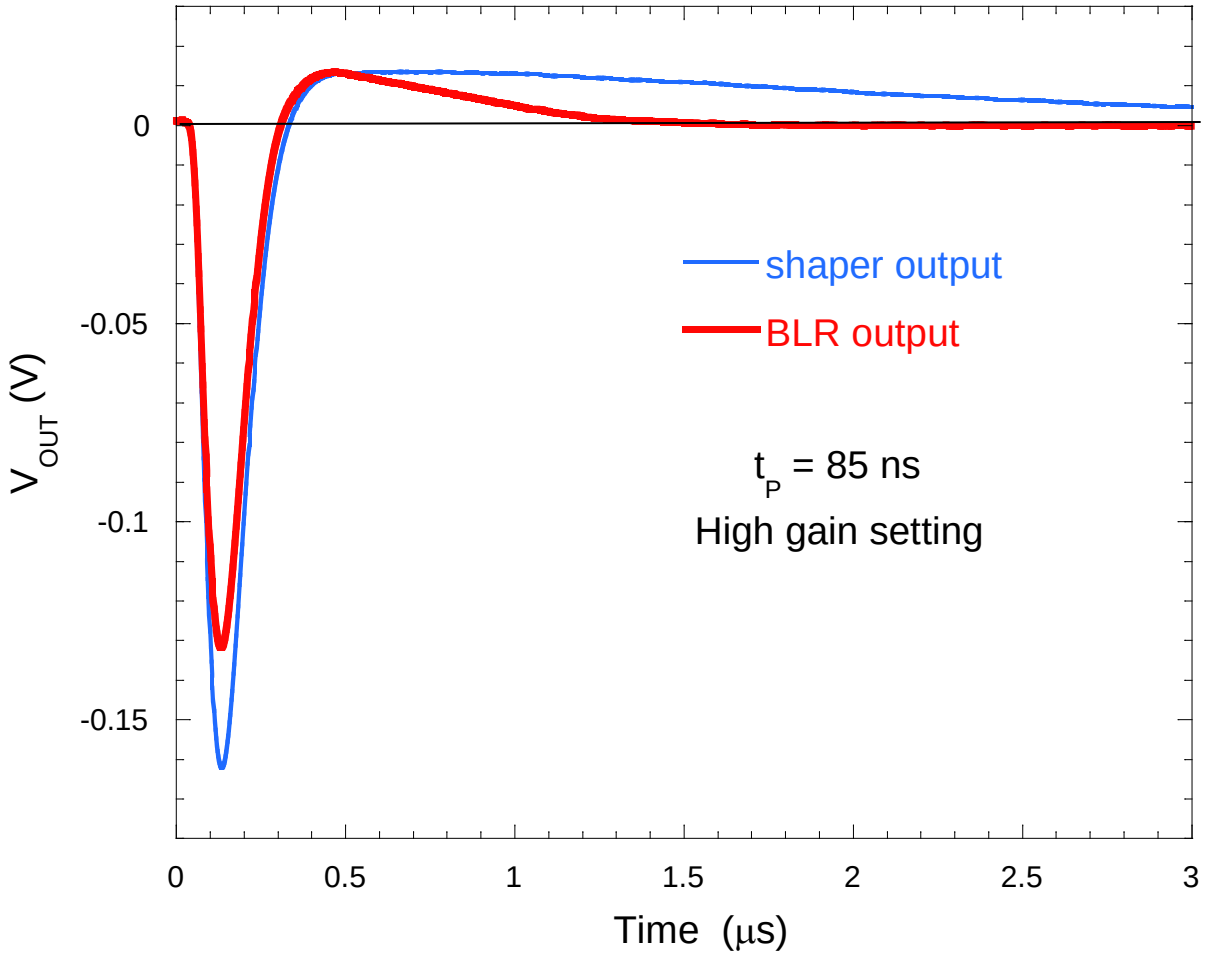


Input signal discriminator scan without BLR

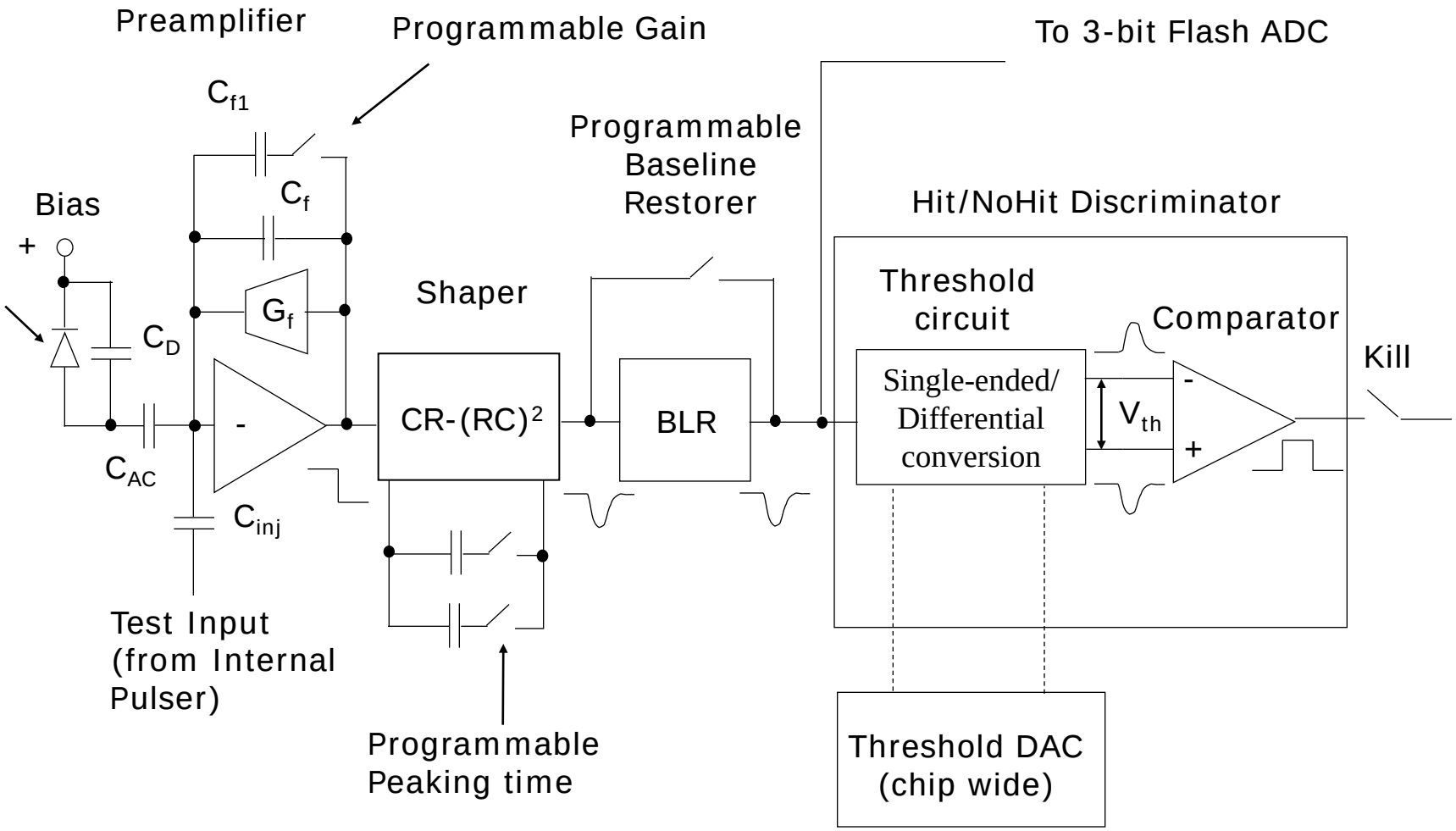


Input signal discriminator scan with BLR

Shift and fluctuations of the baseline at the discriminator input can be removed by a baseline restorer.



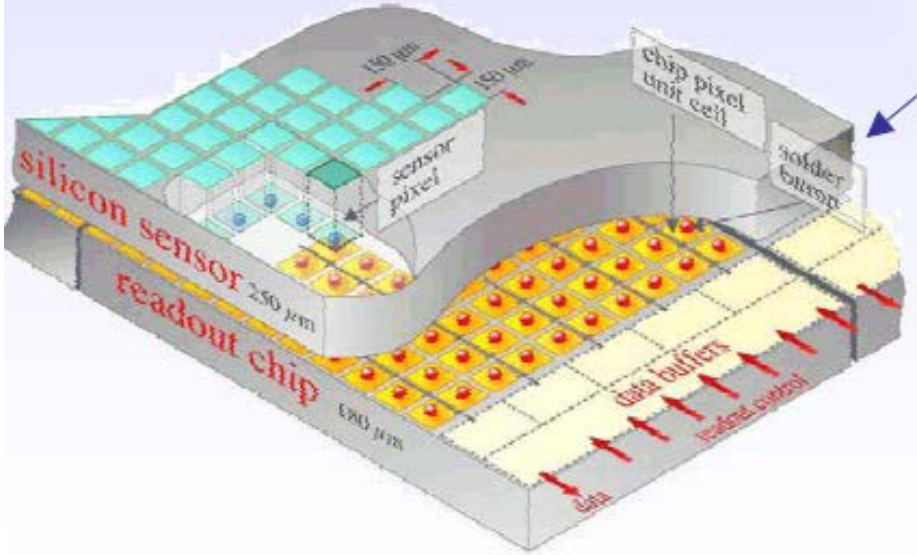
Analog channels (FSSR2 chip)



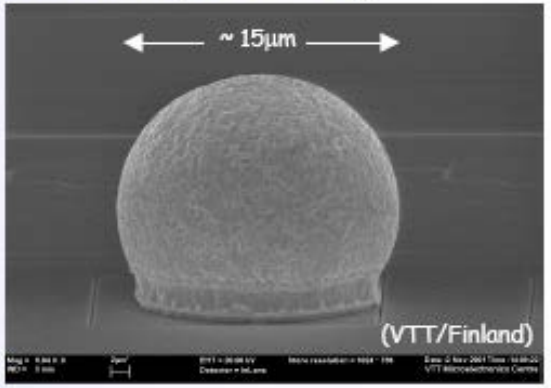
Hybrid pixel sensors

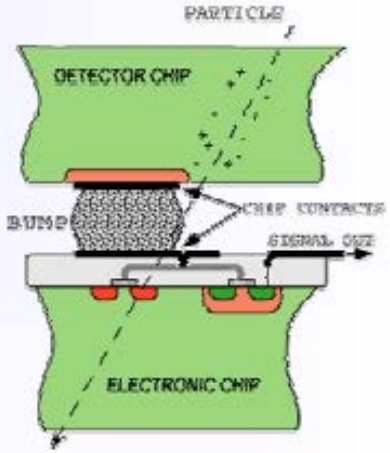
HAPS – Hybrid Active Pixel Sensors

- segment silicon to diode matrix with high granularity (⇒ true 2D, no reconstruction ambiguity)
- readout electronic with same geometry (every cell connected to its own processing electronics)
- connection by "bump bonding"
- requires sophisticated readout architecture
- Hybrid pixel detectors will be used in LHC experiments: ATLAS, ALICE, CMS and LHCb



Solder Bump: Pb-Sn

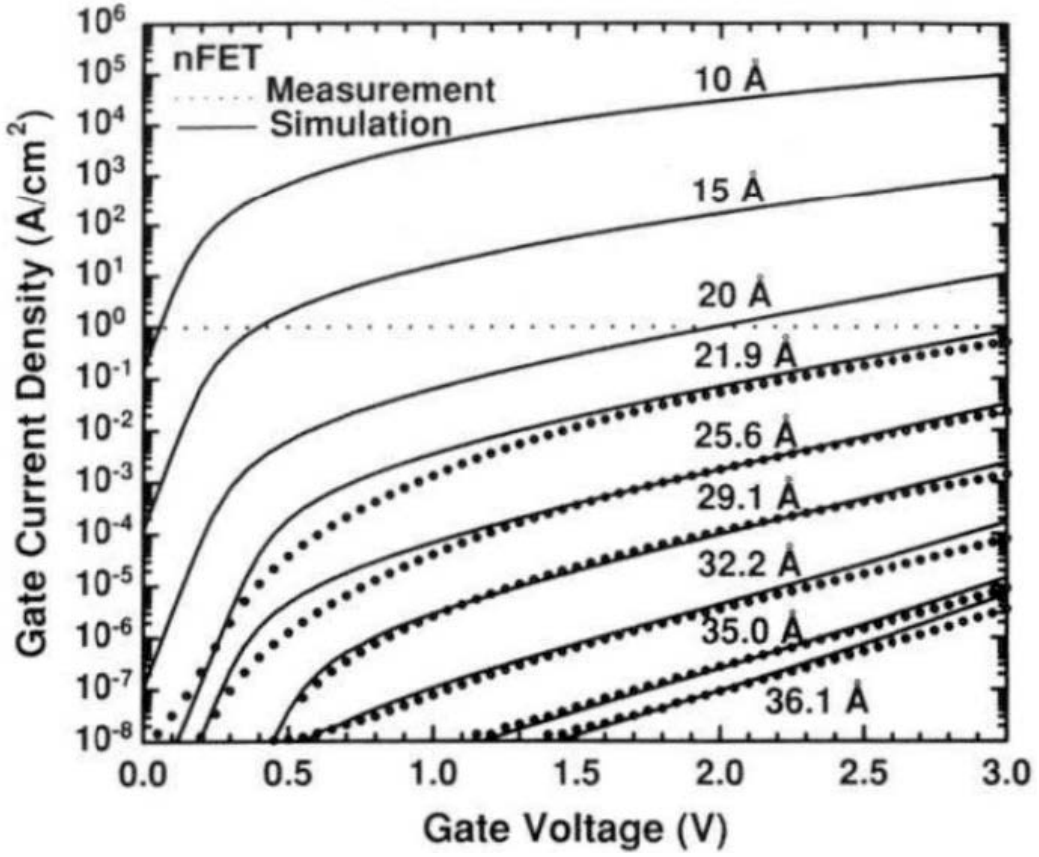




Flip-chip technique

CERN Academic Training Programme 2004/2005

Gate leakage current shot-noise in nanoscale CMOS



$$S_{IG}(f) = 2qI_G$$

90 nm CMOS process:

- Current density = 1 A/cm^2
- W = 1000 μm
- L = 0.1 μm

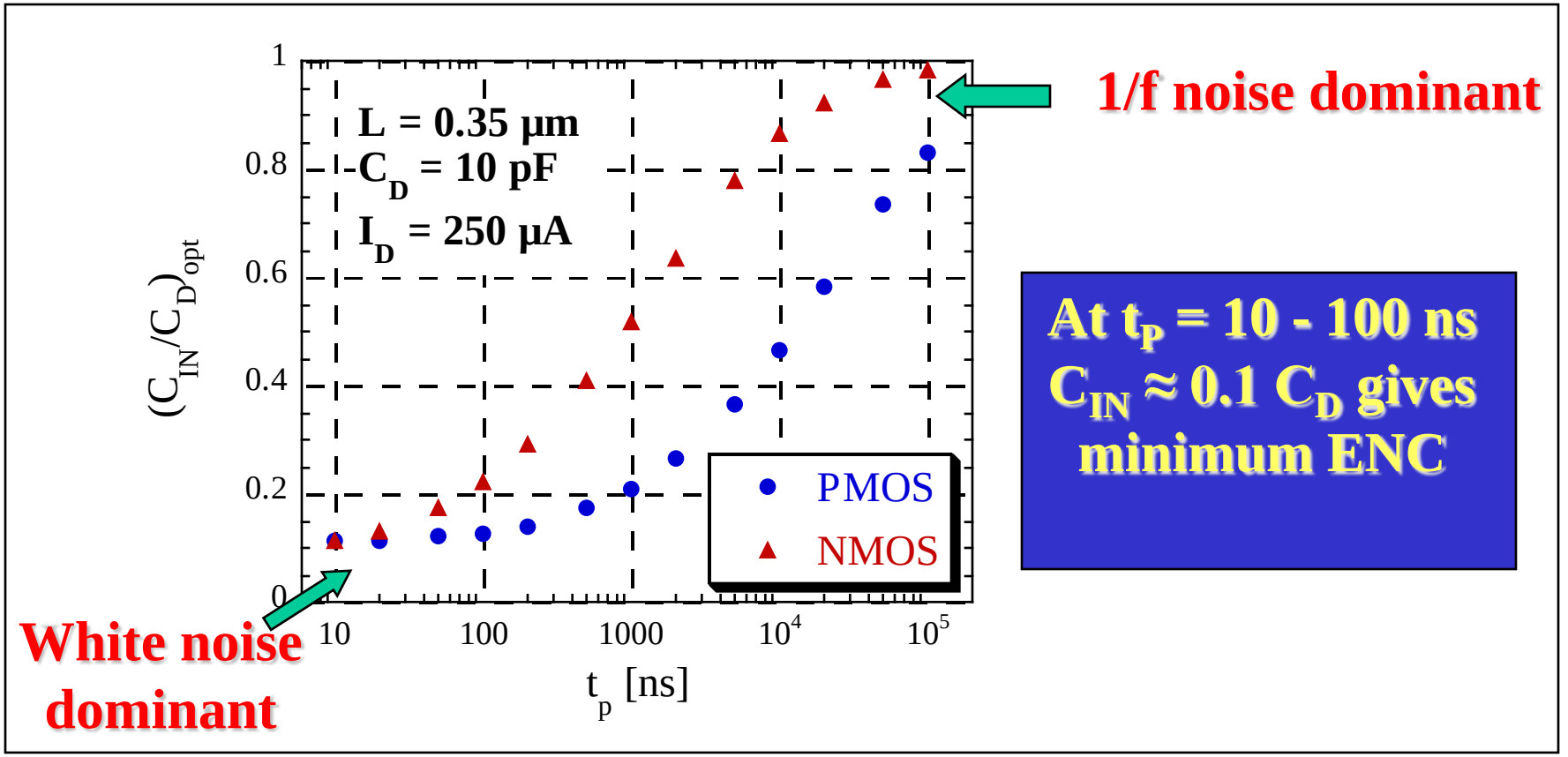
$\Rightarrow I_G = 1\mu A$
$\Rightarrow S_{IG} = 2qI_G = 0.56$ pA/√Hz

Non negligible noise contribution

Capacitive matching

- It is possible to minimize ENC by a correct choice of the dimensions of the preamplifier input device (gate width W and length L)
- Conditions for optimum matching between the preamplifier input capacitance ($C_{IN} = C_{OX}WL$) and the detector capacitance C_D depend on the input device operating region (most often, weak or moderate inversion) and on which series noise contribution is dominant (white or 1/f)
- This optimization has to comply with constraints on the power dissipation, which limit the drain current in the input device (in weak inversion, $A_W \div 1/g_m \div 1/I_D$)

Capacitive matching in a deep submicron technology



0.18 µm technology

Capacitive matching in a deep submicron technology

- Optimum ENC and input NMOS gate width in the C_D region of pixel detectors