New technologies for embedding intelligence in pixel front-end electronics

Valerio Re
(valerio.re@unibg.it)
Università di Bergamo and INFN Pavia, Italy

UNIVERSITÀ
DEGLI STUDI
DI BERGAMO

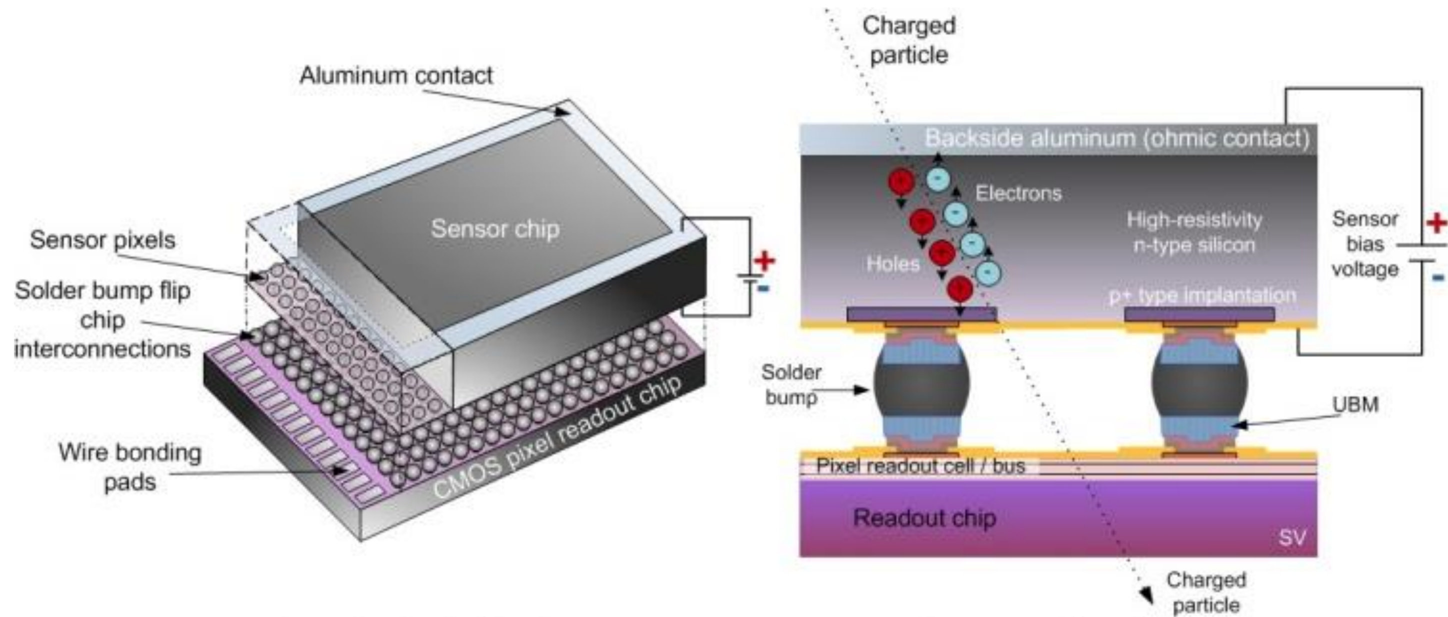
6th Summer School on Intelligent signal processing for FrontIEr Research and Industry
Aug 23rd - Sep 4th, 2021, UAM, Madrid, Spain

Front-end electronics as an integral part of particle detectors

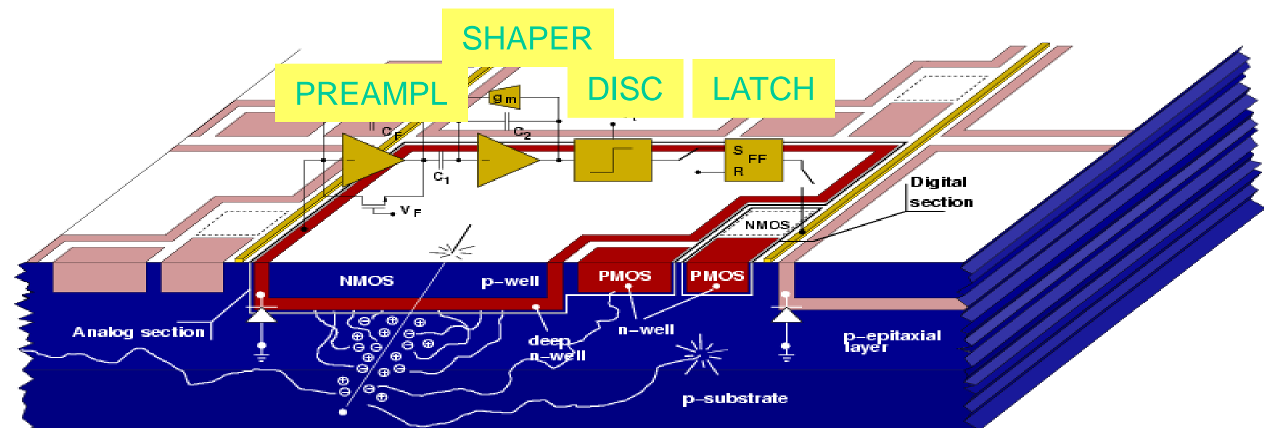
- Industrial microelectronic technologies have crossed the 10 nm frontier bringing CMOS into the nanometer world
- Digital figures of merit (speed, density, power dissipation) are driving the evolution of CMOS technologies
- For analog applications in which speed and density are important, scaling can be in principle beneficial, but what about critical performance parameters such as noise and radiation hardness?
- Nanoscale CMOS is appealing for the design of very compact front-end electronics systems with advanced integrated functionalities, such as required by semiconductor pixel sensors with low pitch for particle detection at extremely high rates and radiation levels
- 3D integration can be a key technology to enhance the performance of sensors and electronics; 3-dimensional techniques are allowing microelectronics to go beyond conventional scaling limits

Hybrid and monolithic pixel sensors and their front-end electronics

Hybrid



Monolithic



Particle tracking at LHC- Phase II:

- Very high hit rates (3 GHz/cm^2), need of an intelligent pixel-level data processing
- Small detector signals require operating at low threshold (< 1000 electrons)
- 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)

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

FCC-hh:

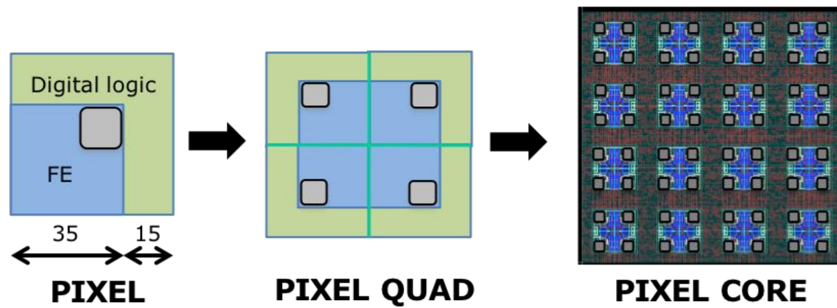
- **Radiation levels expected to increase in inner layers (25 mm):**
  up to 30 Grad and 10^{18} neutrons/cm^2
- **Smaller pixels (avoid in-pixel pileup)**
  (~25x50 μm^2) the need for higher logic density is not a function of pixel size, but of hit rate per unit area.
- **Huge data rates:**
  - Max hit rate 20 Gb/s/cm^2, will need 50-100 Gbps low-power, low-material data links

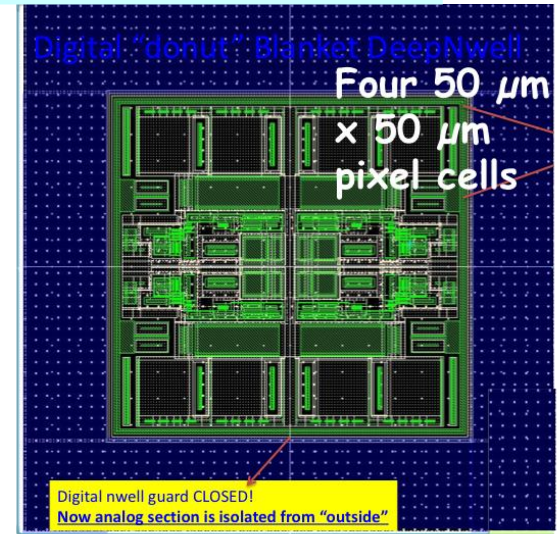
Advanced pixel detectors and readout microelectronics: photon science

X-ray imaging at free electron laser facilities (next generation):

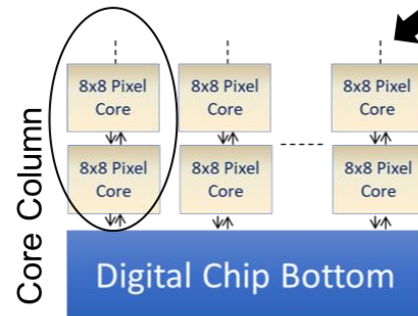
- Reduction of pixel size (100x100 μm^2 or even less), presently limited by the need of complex electronic functions in the pixel cell:
  - Large memory capacity to store images (at XFEL, ideally, 2700 frames at 4.5 MHz every 100 ms, mean frame rate ~27 kHz, maybe going to continuous wave mode of operation at 1 MHz), high data throughput
  - Maximum frame rate allowing for power cycling (needs to be done on-chip?)
  - Sharing of readout cells among small pixels
  - Advanced pixel-level processing (1 – 10000 photons dynamic range, 10-bit ADC at 5 MHz operation)

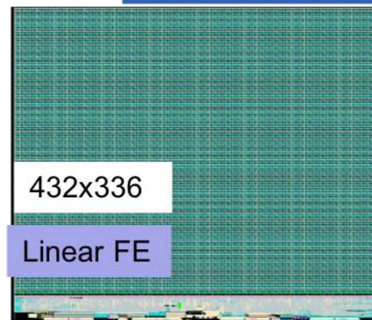
A pixel readout chip for HL-LHC (ATLAS, CMS) with 50 μm x 50 μm analog and digital cells





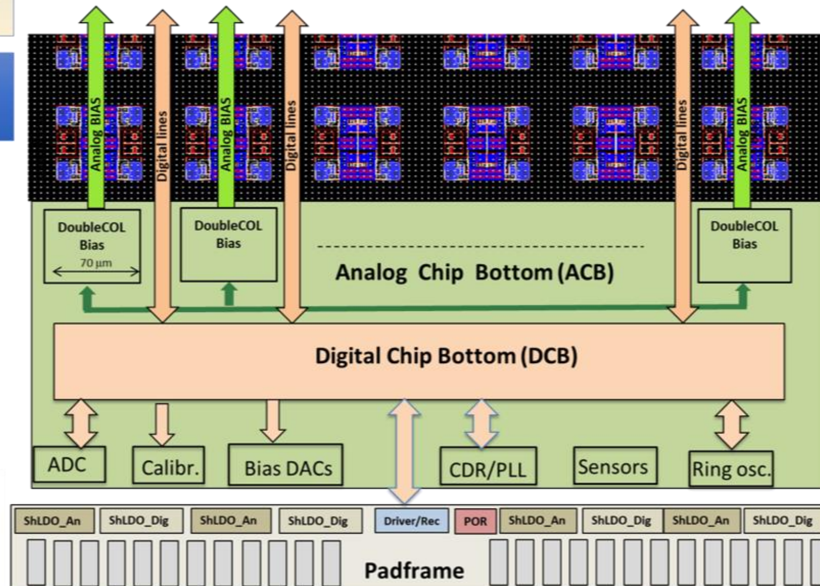
Isolation of analog front-end from the mostly digital environment makes it possible to operate at low threshold (≲ 1000 electrons) at extreme hit rates and radiation levels





RD53B-CMS (CROC_V1)

- submitted in May 2021
- size: 21.6 mm x 18.6 mm



Trends of pixel detectors and readout microelectronics

- Finer granularity (smaller pixels, more channels), challenge to reduce power dissipation
- Ultra-high radiation hardness (> 1 Grad total ionizing dose) required by some applications
- Precise timing information for triggering and reconstruction
- Increasing amount of transmitted data requires more and faster links, using the smallest possible amount of material.
- Increased and improved detector information used by intelligent processors on or close to the front-end readout, to reduce the amount of data transmitted off the detector, and hence power dissipation, through data selection and reduction
- **New microelectronic technologies will be the crucial tool to achieve the performance required by future detector systems**

Trends of pixel detectors and readout microelectronics

The microelectronic industry has reached nanoscale CMOS nodes, based on new transistor structures (FinFETs, gate-all-around devices, etc.) to improve the speed and power performance of digital systems.

Future detector readout ASICs will be mixed signal systems, including analog interfaces with sensors: achieving the required analog performance with advanced CMOS nodes will be a critical target

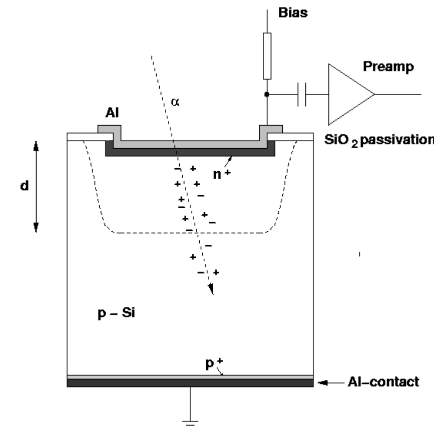
High-performance commercial CMOS image sensors are based on stacking two or three active layers (sensor, logic, memory) to reduce pixel pitch and increase functionality.

These industrial advances can provide the means to enhance or possibly revolutionize the performance of future detector readout ASICs and their interconnection to high granularity detectors

Sensor design is nowadays closely intertwined with its readout electronics and the optimization of the performance is a joint effort.

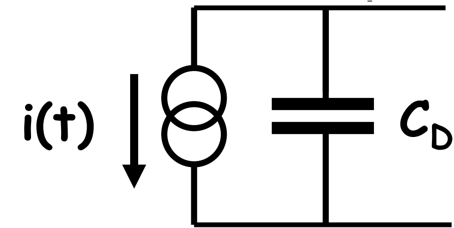
The fundamental analog signal processing functions

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)

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

filtering, signal shaping

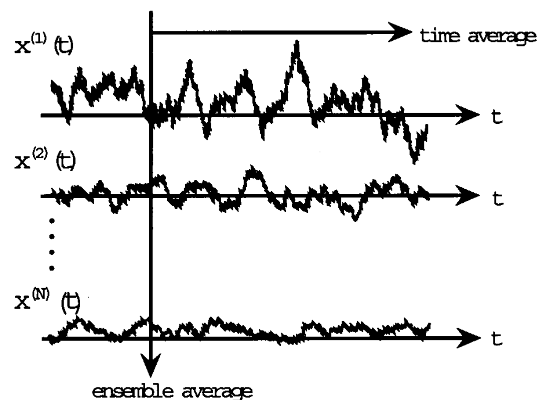
optimize the measurement of a desired quantity such as signal amplitude as a measure of the energy loss of the particle, with the best possible accuracy compatible with noise intrinsically present in the amplifying system, and with the constraints set by the different applications (power, speed,…)

Effect of electronic noise on charge measurements

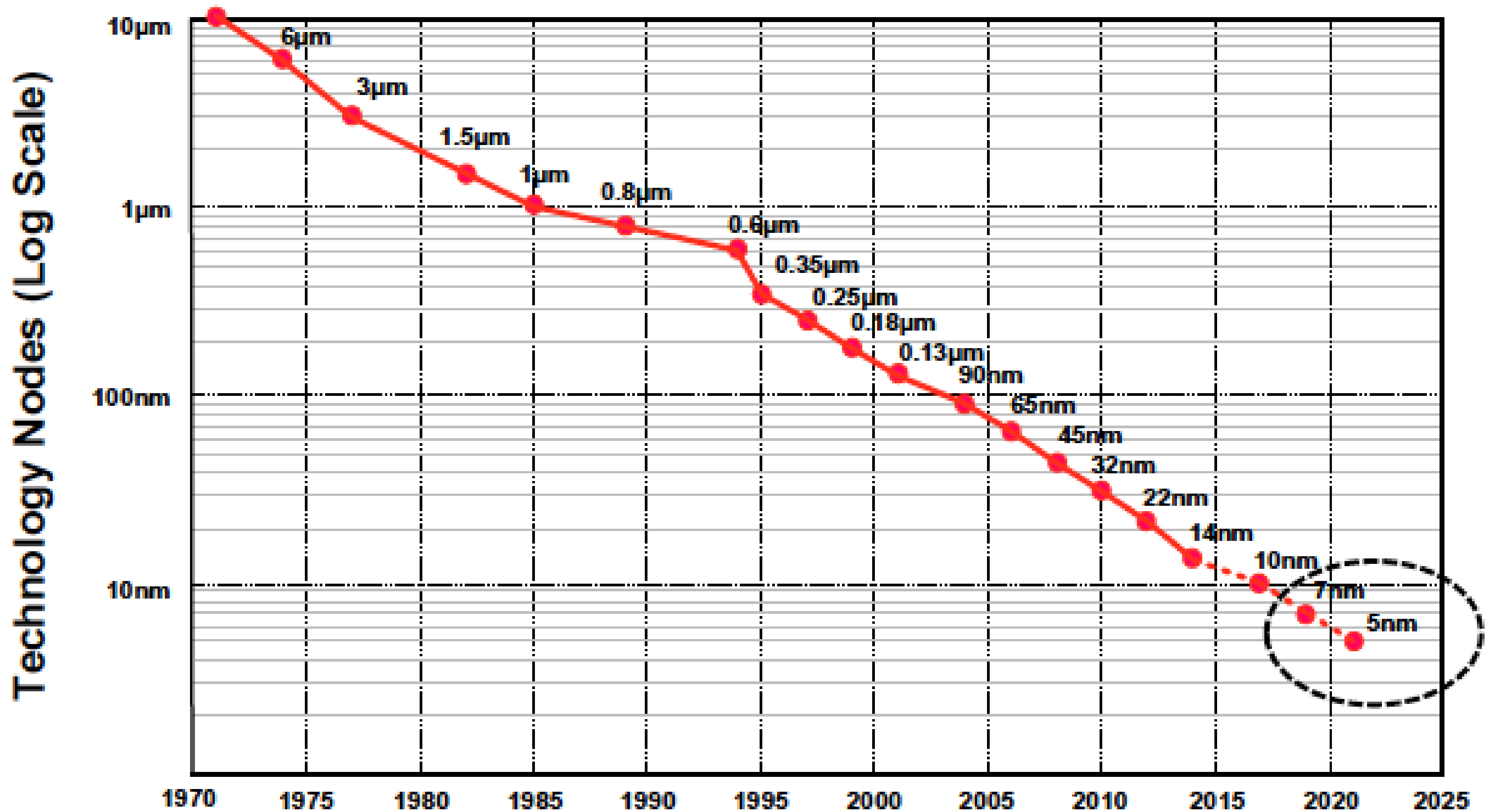
Inherent to the conduction of current in an amplifying device is a random component, depending on the principle of operation of the device.

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.

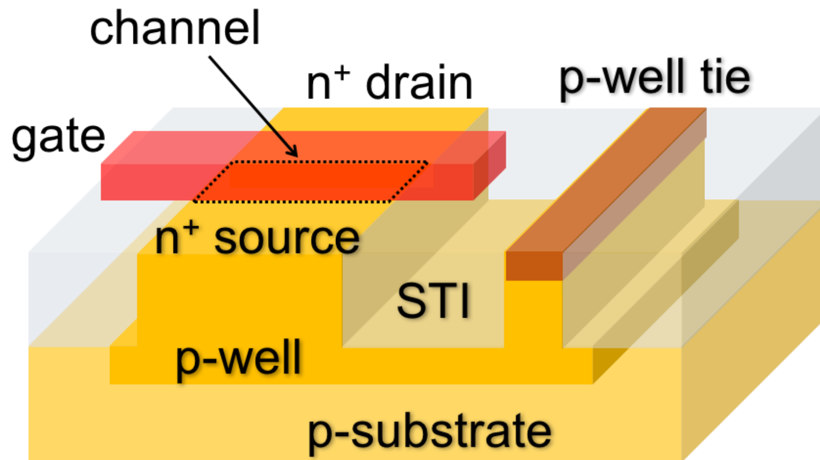


Scaling of microelectronic processes

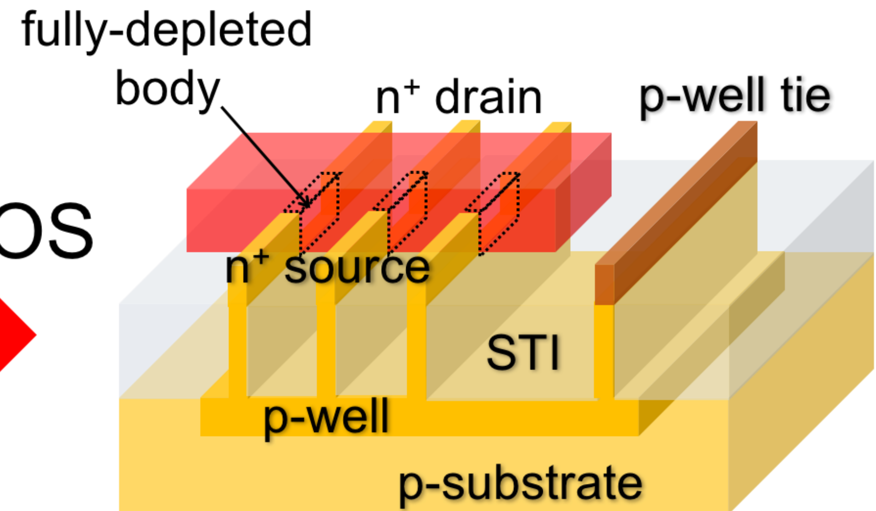


Source: Semiconductor device fabrication, https://en.wikipedia.org/wiki/Semiconductor_device_fabrication

Below 28 nm, FinFET



NMOS

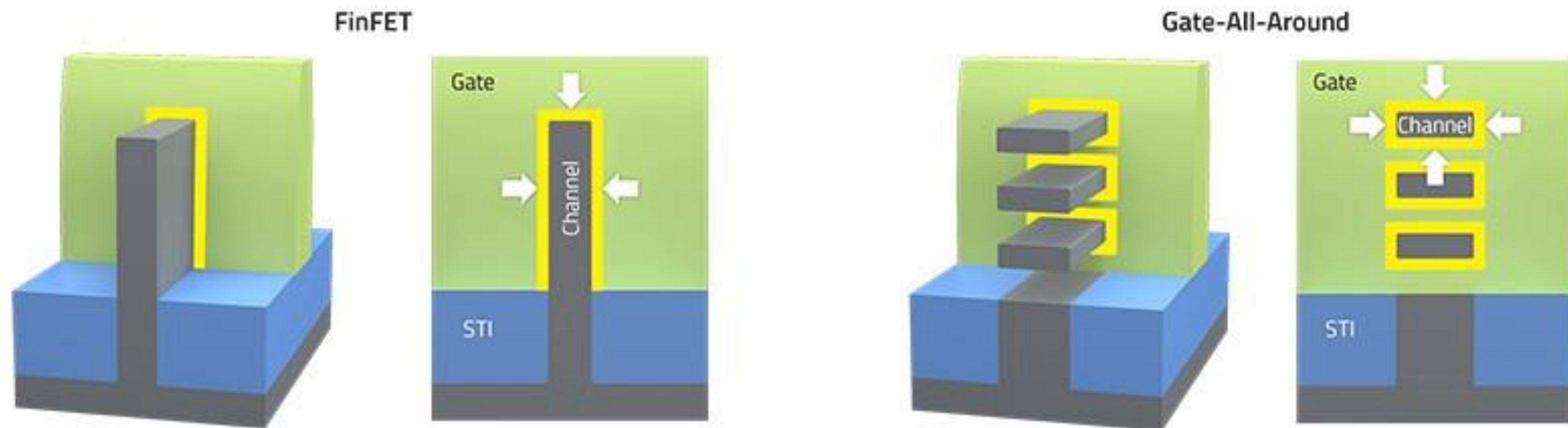


Three-dimensional, Gate-All-Around vertically stacked transistors

At reduced gate length, even the FinFET fails to provide enough electrostatic control of the channel. The scaling of the size of standard cells requires using single-fin devices, which cannot provide enough drive current

In GAA transistors, the channel is divided into separate horizontal sheets. As the gate now fully wraps around the channels, superior channel control is obtained compared to FinFET

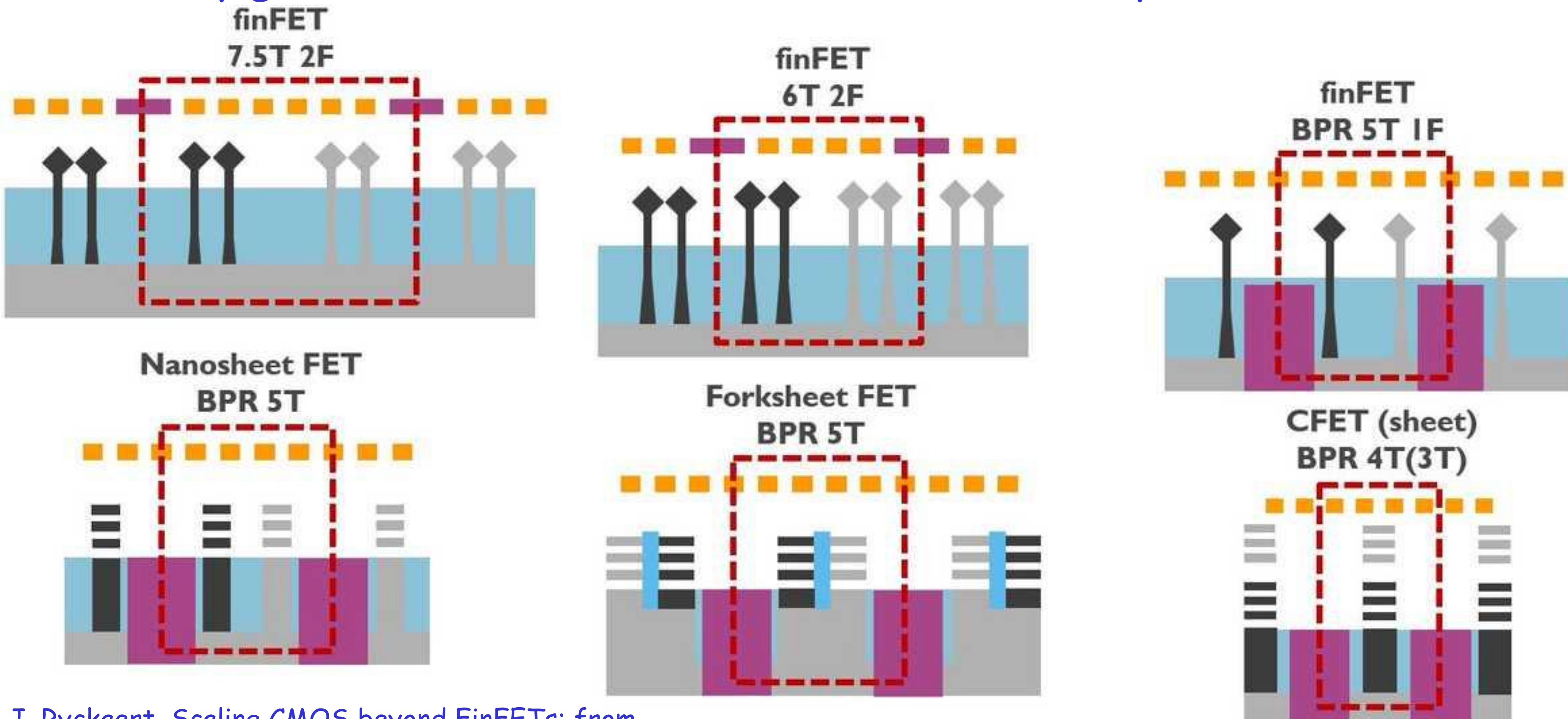
The sheet-to-sheet spacing, analogous to fin pitch, is determined not by lithography but by tightly controlled epitaxial processes



The meaning of scaling

Scaling is about density (not, or not only about the gate length of transistors) ⇒ more speed, more power/energy efficiency

For recent CMOS nodes, "7 nm", "5 nm" are not related to a feature size: they give an indication of the achievable density of transistors

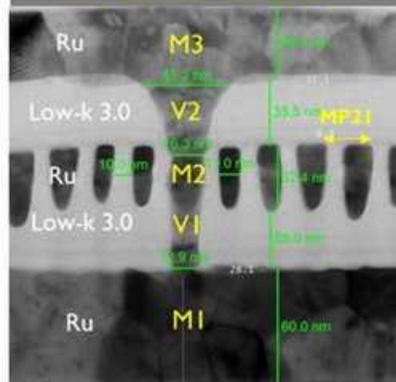


J. Ryckaert, Scaling CMOS beyond FinFETs: from nanosheets and forksheets to CFET, IMEC, 2020

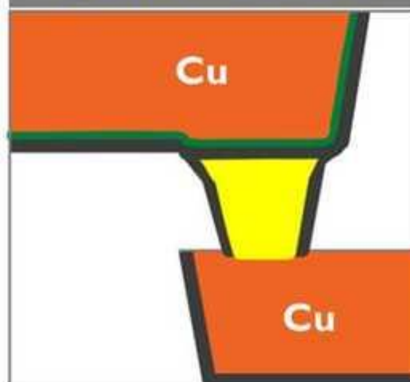
The next 20 years (?)

State of the art	Next innovation options		Further out options	
FinFET	Nanosheet	Forksheet	CFET	Sequential 3D

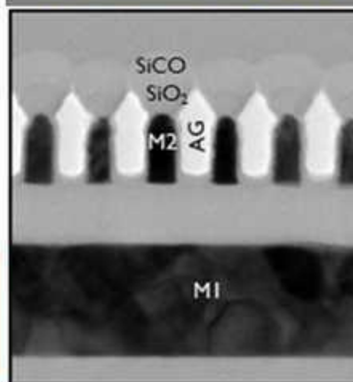
Dual damascene MP21



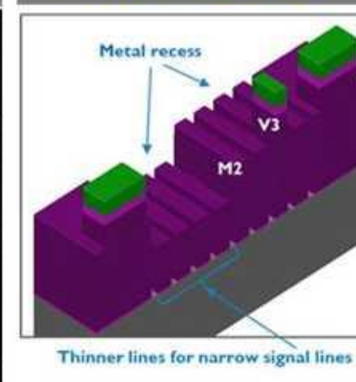
Hybrid via metallization



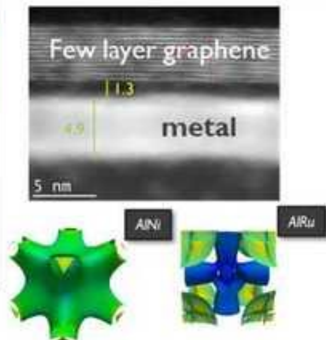
Semi-damascene w AG



H^2 with Zero Via



Hybrid graphene-metal
Or New conductor



Z. Tokei, N. Horiguchi: A view on the logic technology roadmap, IMEC, 2020

Very high carrier mobility can be achieved with 2D layered materials (or 1D materials) and nanometer-thin transistors

The adoption of advanced CMOS in the High Energy Physics community

- For various reasons (cost, accessibility, know-how,…), in the HEP community we are lagging behind the frontier of industrial technology (also CMOS image sensor processes are following the scaling of technology nodes with some delay)
- In a 20-year timescale, it is certainly possible that we'll have access to the most advanced CMOS nodes that are available today
- Can we make some predictions about the achievable performance that we can attain in the future?
- For example, how much can we shrink the size of pixel cells (considering that we'd like to integrate more and more complex functions)? I'll try to base the answer on our experience

CMOS scaling applied to pixel readout cells: from 130 nm to 65 nm for photon science (X-ray imaging)

DSSC

236μm

204μm

SRAM

CSA FE

DEPFET FE

Slow Control

ADC

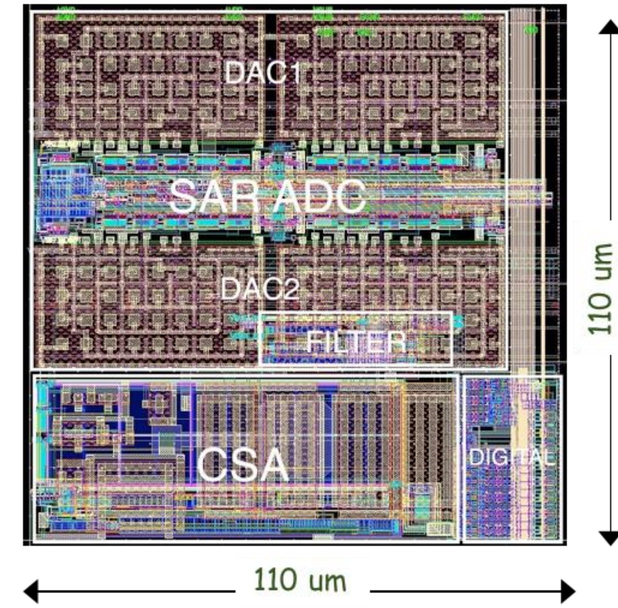
Filter

References

Injection

CSA + filter: 6000 μm^2

PixFEL

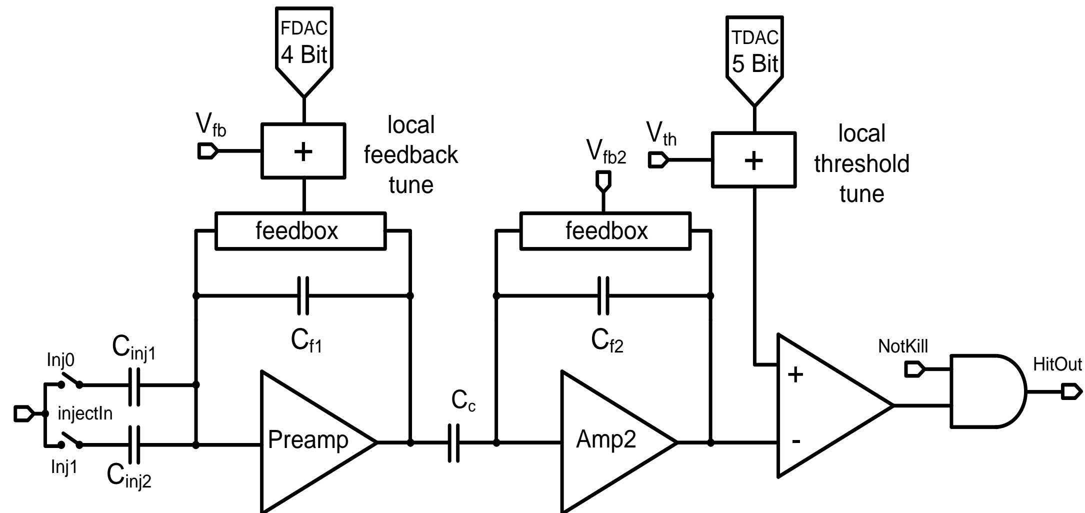


Analog section: 5000 μm^2

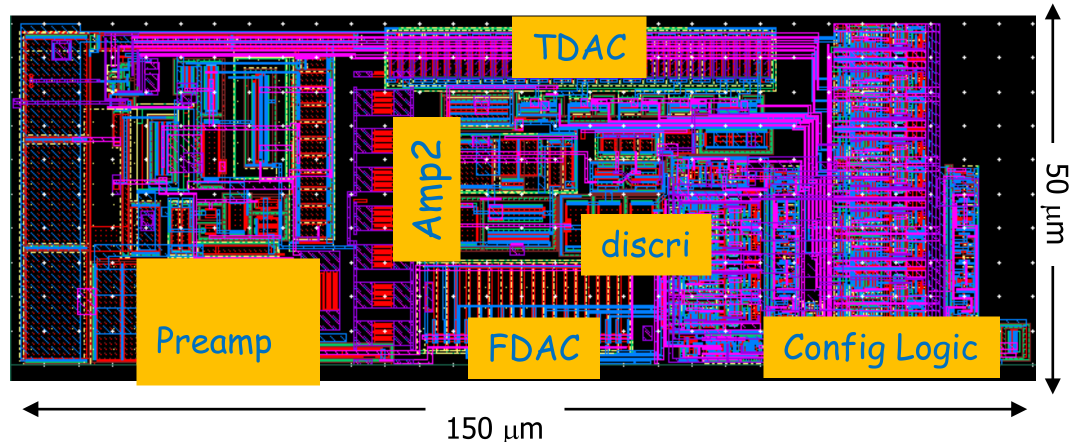
CMOS scaling applied to pixel readout cells: from 130 nm to 65 nm for HEP

Analog section in the readout cell in pixel front-end chips for particle tracking at LHC

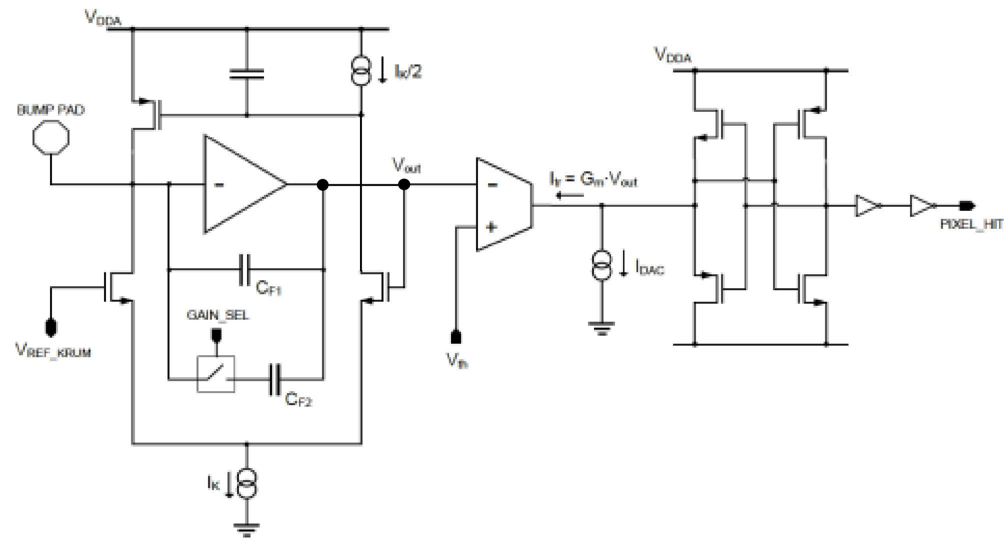
130 nm CMOS FE-I4 readout chip for pixel sensors in ATLAS IBL: the analog pixel cell



About 6000 μm^2 for the analog front-end in a 250 x 50 μm^2 pixel

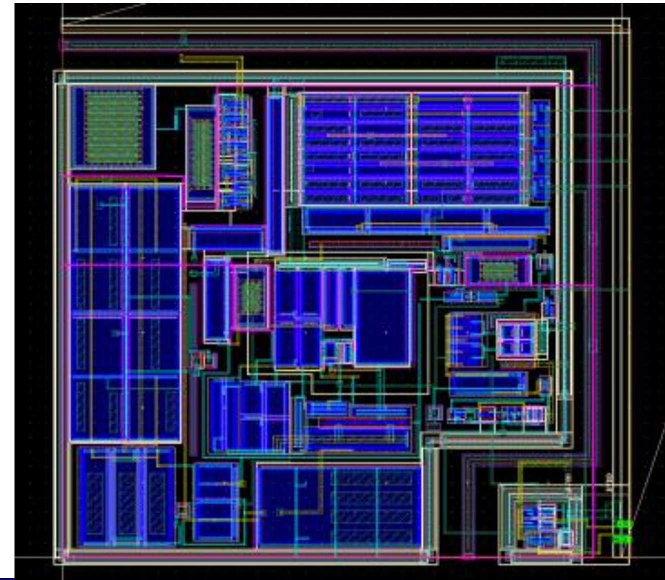


65 nm CMOS readout chip for pixel sensors in the phase II upgrade of the CMS inner tracker at HL-LHC: the Linear Front-End

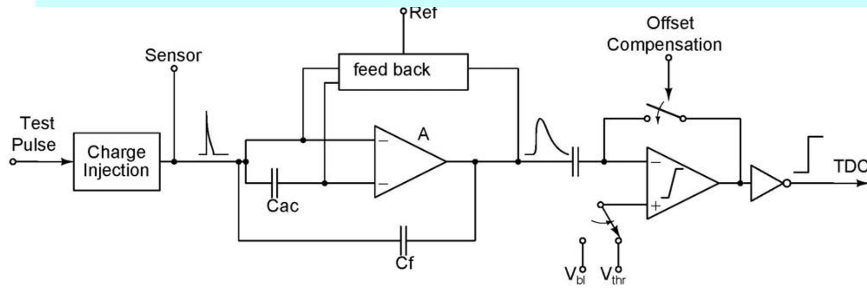


In a 50 μm x 50 μm pixel cell, the area allocated to the analog front-end is about 1000 μm^2
(A/D conversion based on Time-over-Threshold)

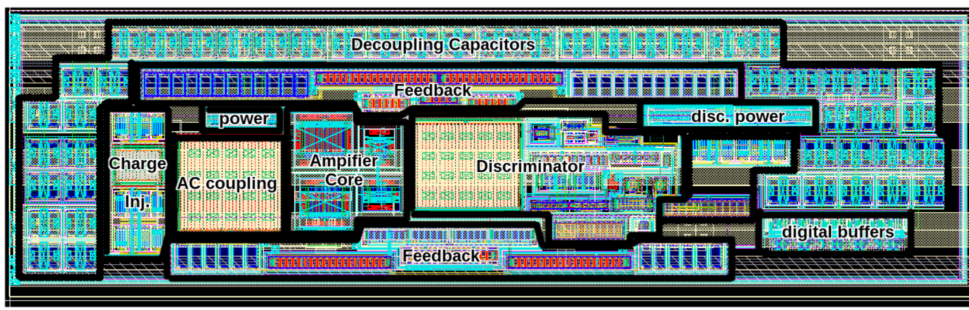
With respect to FE-I4, area reduction is also achieved by changing the design (e.g., no shaper)



TIMESPOT: a 28 nm CMOS chip for the readout of pixel sensors with high performance timing



Pixel pitch: ≤ 55 μm
Time resolution: ≤ 50 ps per pixel
(target = 30 ps or better)

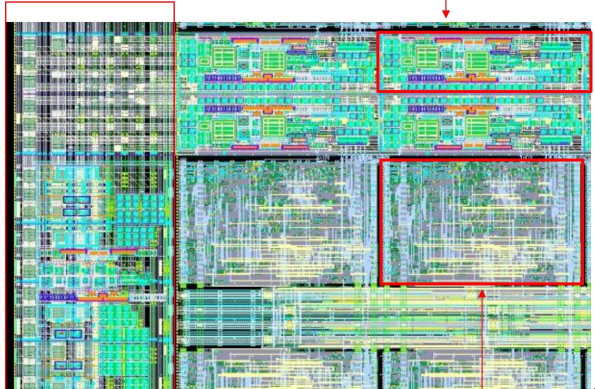


To maximize sustainable rate, **1 TDC per pixel channel** has been integrated

Maximum input signal TDC rate = 3 MHz
24 bits output word (ToA + ToT) serial @160 MHz

50 μm

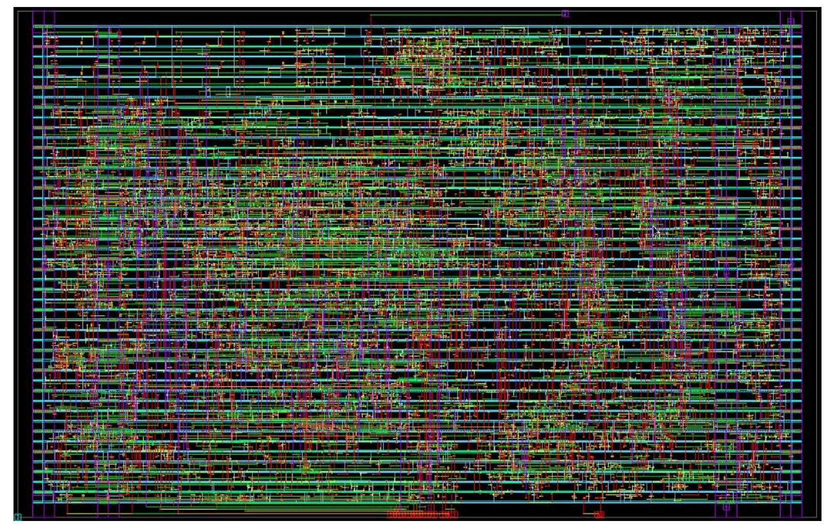
Analogue cell
(CSA amplifier and Leading Edge discriminator):
size 50 x 15 μm²



Analogue services

TDC: 50 x 31.5 μm²

nal Sumn



INFN CSN5 R&D project
P.I. Adriano Lai, INFN Cagliari

The design of future readout chips

- Previous examples shows that the size of analog circuits is not shrinking by simple CMOS scaling: it may require changes in the pixel cell design

- Digital signal processing can (and probably must) be greatly enhanced in future chips design, extracting high quality data to be sent off chip. This will also allow designers to fully exploit CMOS scaling to the nanometer level

- Still, detection of small signals requires analog circuits (typically preamp + discriminator) capable of operating at low charge threshold, while avoiding spurious hits (low noise and threshold dispersion) and surviving in a hostile digital environment

Critical parameters for analog front-end design in nanoscale CMOS

- An analog designer cares about a set of crucial parameters when she/he has to adopt nanoscale CMOS for detector front-end integrated circuits:
  - Intrinsic gain
  - Thermal noise

  } Charge carriers in the device channel (short channel effects, strained silicon)

  - Gate leakage current
  - 1/f noise

  } Interaction of charge carriers with the gate oxide; tools for evaluating the quality of the gate dielectric

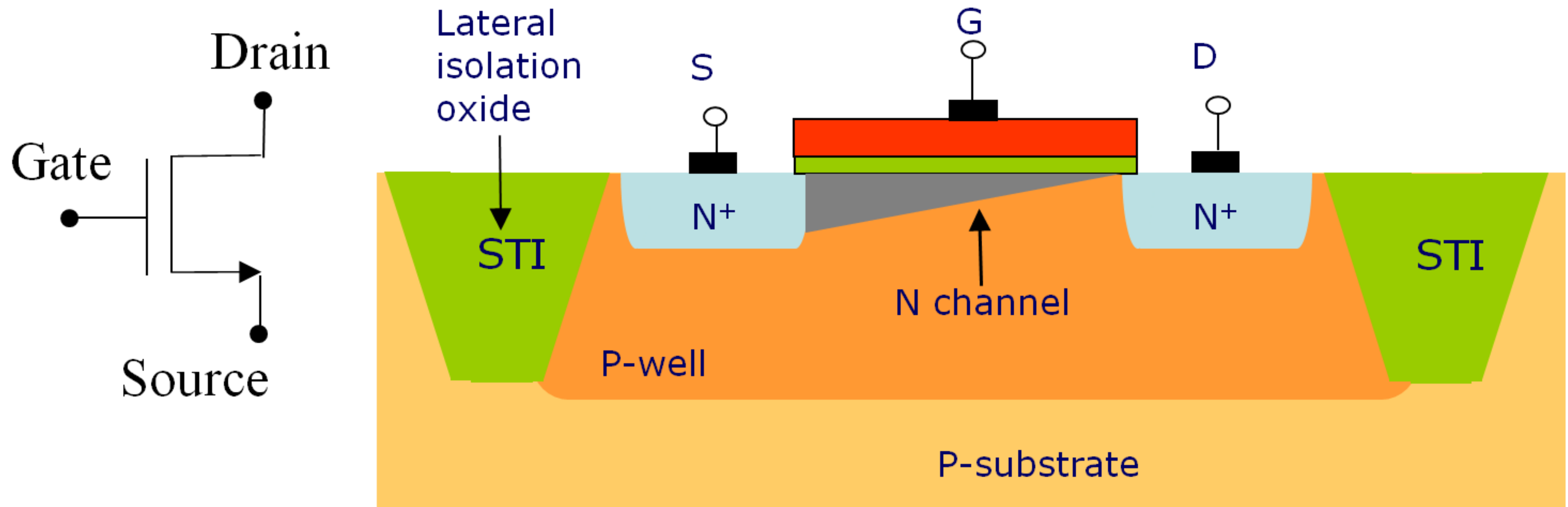
  - Radiation hardness

  } Radiation-induced positive charge in the gate oxide and in lateral isolation oxides

  - Device matching

  } Random components of local process parameters dependent on device size

Basic element of modern electronics: the MOSFET

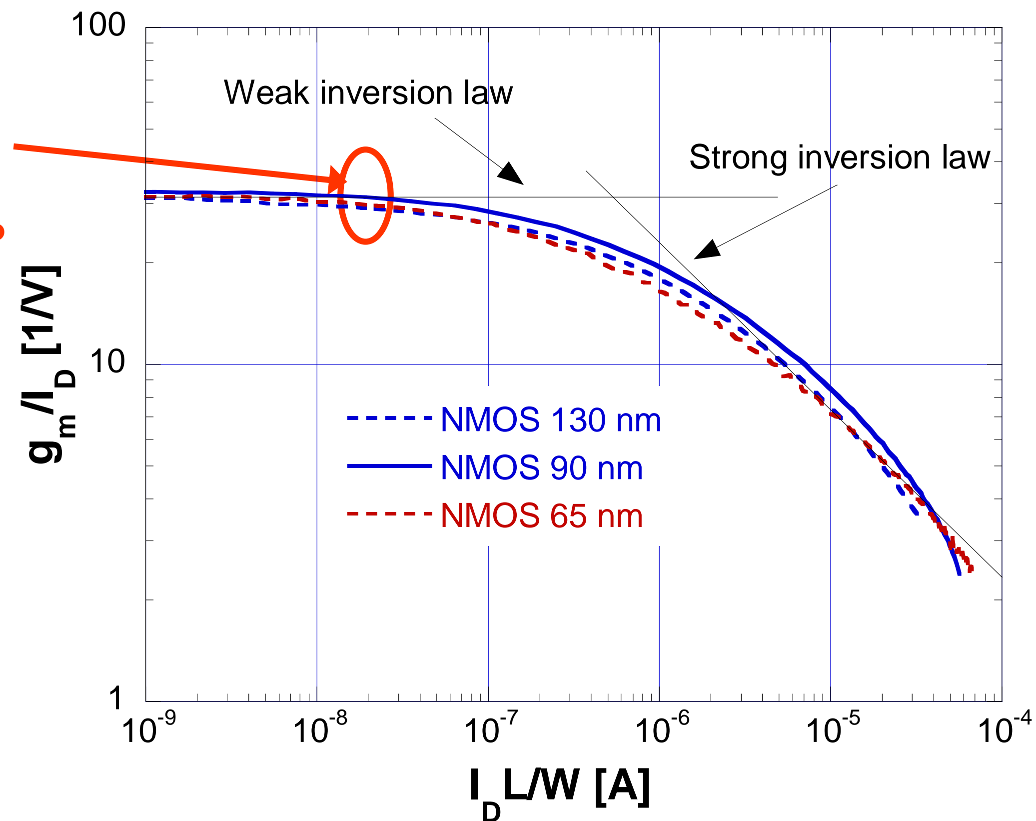


- Three-terminal device: an electrode controls the current flow between two electrodes at the end of a conductive channel.
- The transconductance $g_m = dI_D/dV_{GS}$ is the ratio of change in the output (drain) current and of the change in the potential of the control (gate) electrode

MOSFET essential parameters: the transconductance g_m

Operating point for W/L =400/0.2 (strips), I_D = 40 μA

W/L =20/0.1 (pixels), I_D = 4 μA



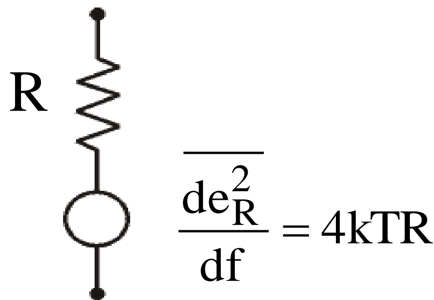
- Under reasonable power dissipation constraints, devices in deep submicron CMOS operate in the weak inversion region
- In weak inversion:

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

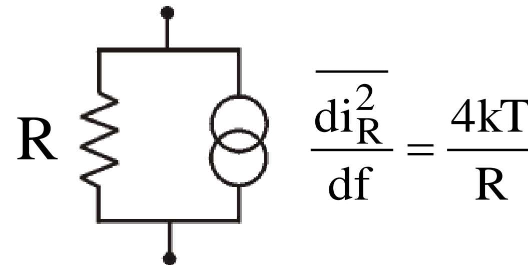
(n =1.2 in 100-nm scale CMOS)

MOSFET essential parameters: channel thermal noise

- Thermal noise arises from random velocity fluctuation 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.



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

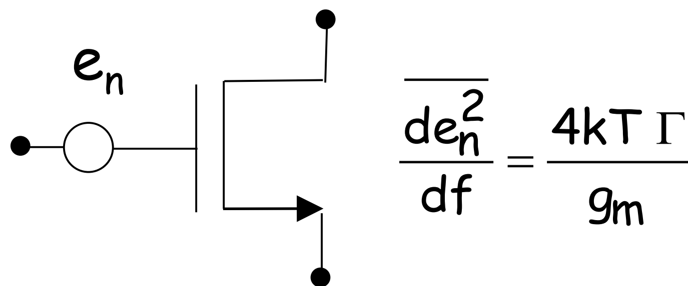


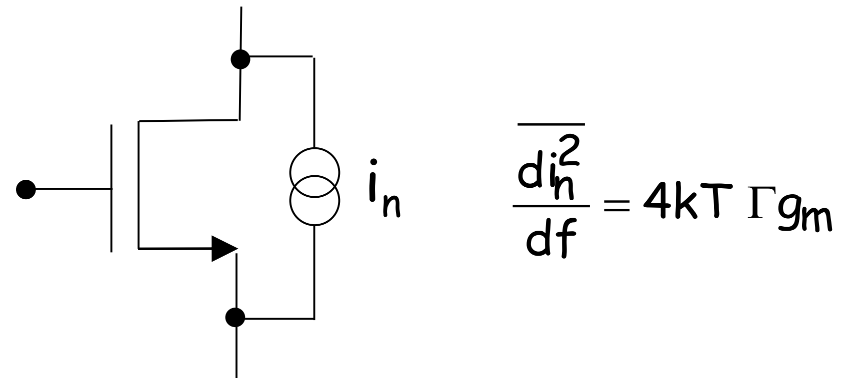
$$\frac{\overline{di_R^2}}{df} = \frac{4kT}{R}$$

- 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 (fluctuations of the drain current can be seen as due to fluctuations of the gate voltage).



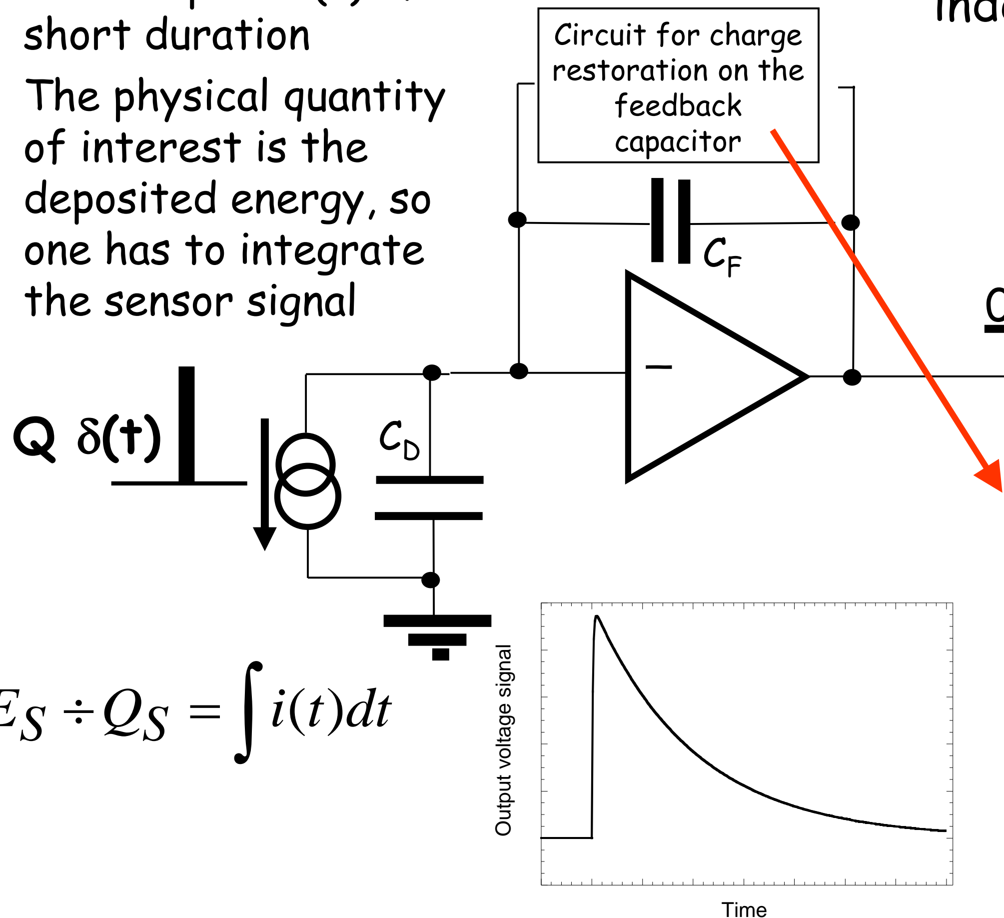


k = Boltzmann's constant, T = absolute temperature

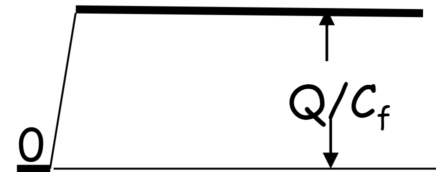
Γ = coefficient ($\cong$ 1) dependent on device operating region, short channel effects…

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



- Use an integrating preamplifier so that charge sensitivity ("gain") is independent of sensor parameters (C_D)



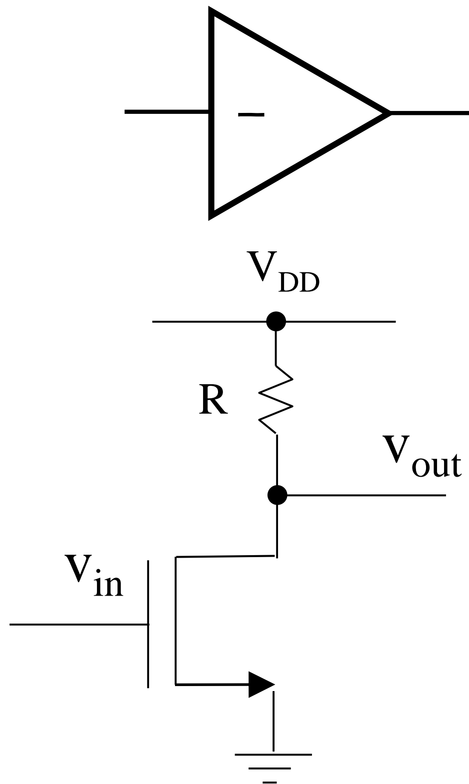
This guarantees a return to baseline of the preamplifier output, avoiding saturation.

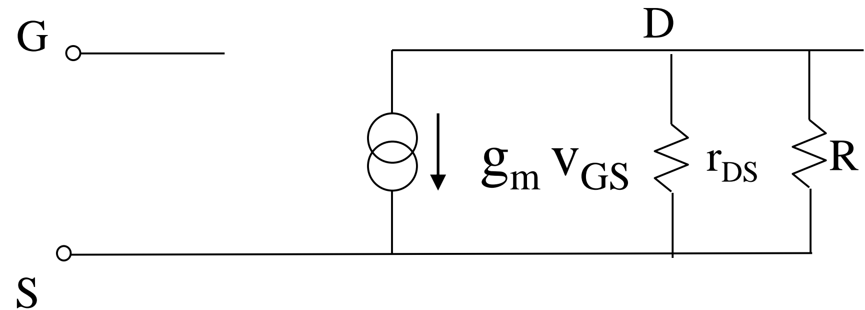
It can be achieved with a resistor R_F or, in an integrated circuit, with a CMOS circuit (transconductor).

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

Forward gain stage: basic single transistor amplifier

- The forward gain stage is an inverting amplifier. Its most basic version can be in principle based on the common source configuration.





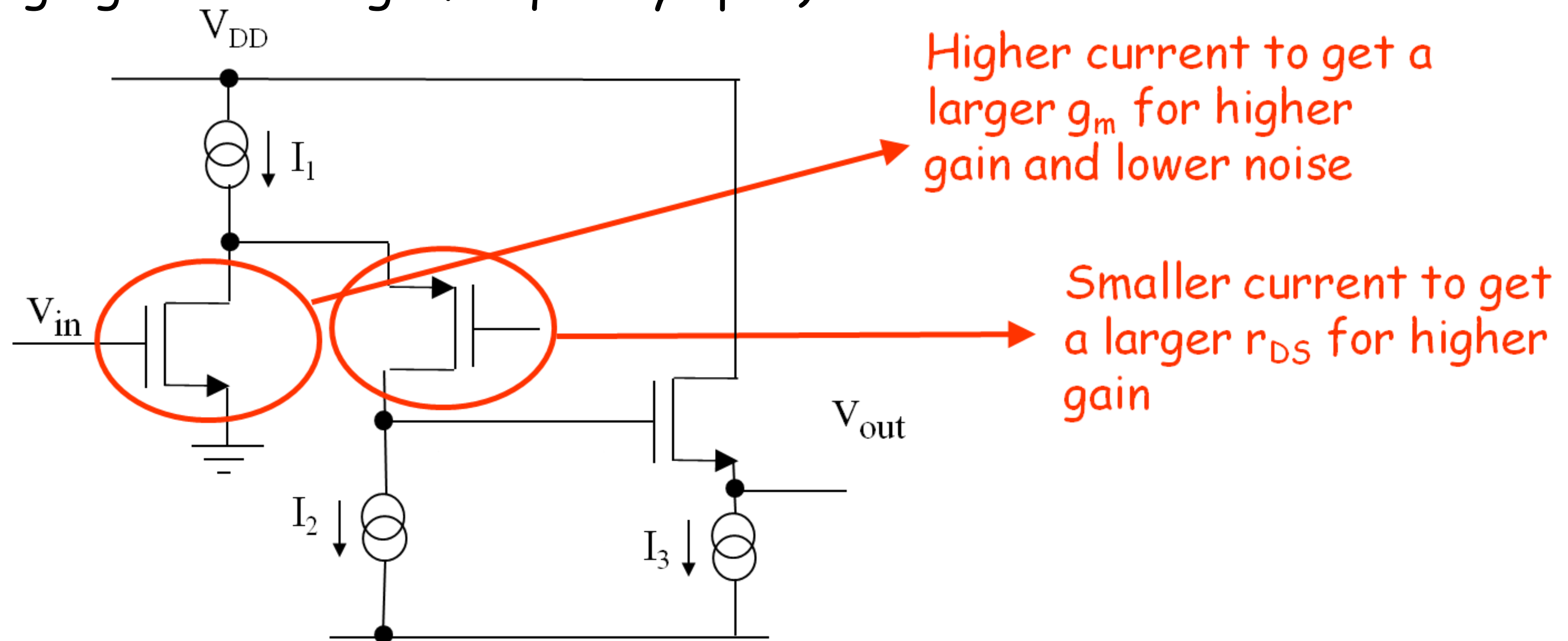
$$\frac{v_{out}}{v_{in}} = -g_m (r_{DS} // R)$$

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

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

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)



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.

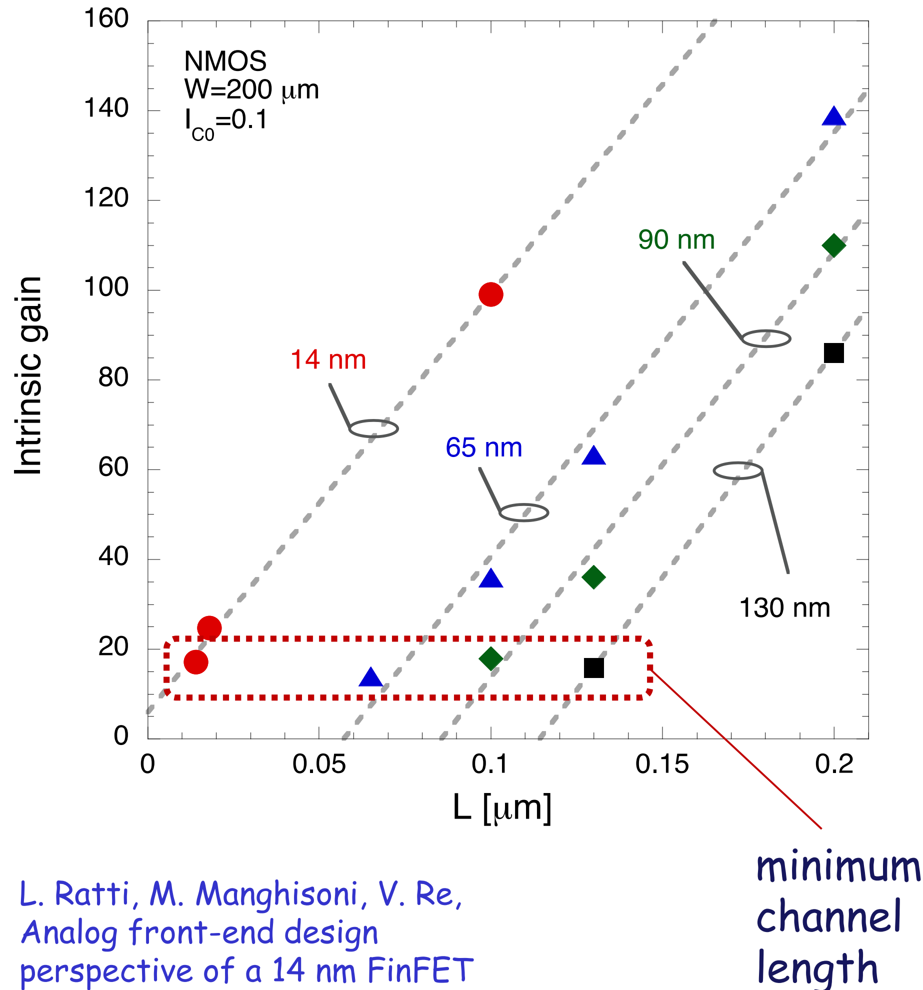
$$g_m r_{DS} \mu a L$$

- 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

Intrinsic gain across CMOS generations

Intrinsic gain in weak inversion



minimum channel length

L. Ratti, M. Manghisoni, V. Re, Analog front-end design perspective of a 14 nm FinFET technology, 2019 IEEE NSS

The value of the intrinsic gain is maintained across different CMOS technology nodes if inversion conditions are the same

Improved gate control was developed in FinFET to decrease leakage current, reduce short channel effects (which also may lead to higher gain) and process-induced variability

The improved electrostatics of GAA transistors can help in increasing the gain

Intrinsic gain, CMOS scaling and design contraints

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 whose parameters are shown in previous plots.

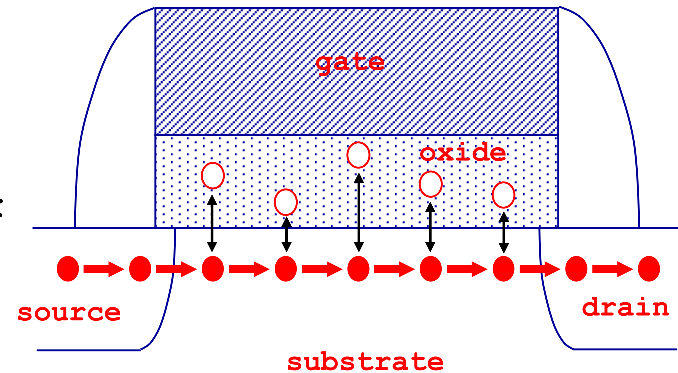
When the gate length L is reduced, the transistor stays in the same inversion conditions (equal value of I_{C0}) if the drain current I_D is correspondingly increased (at constant W).

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,…)

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.

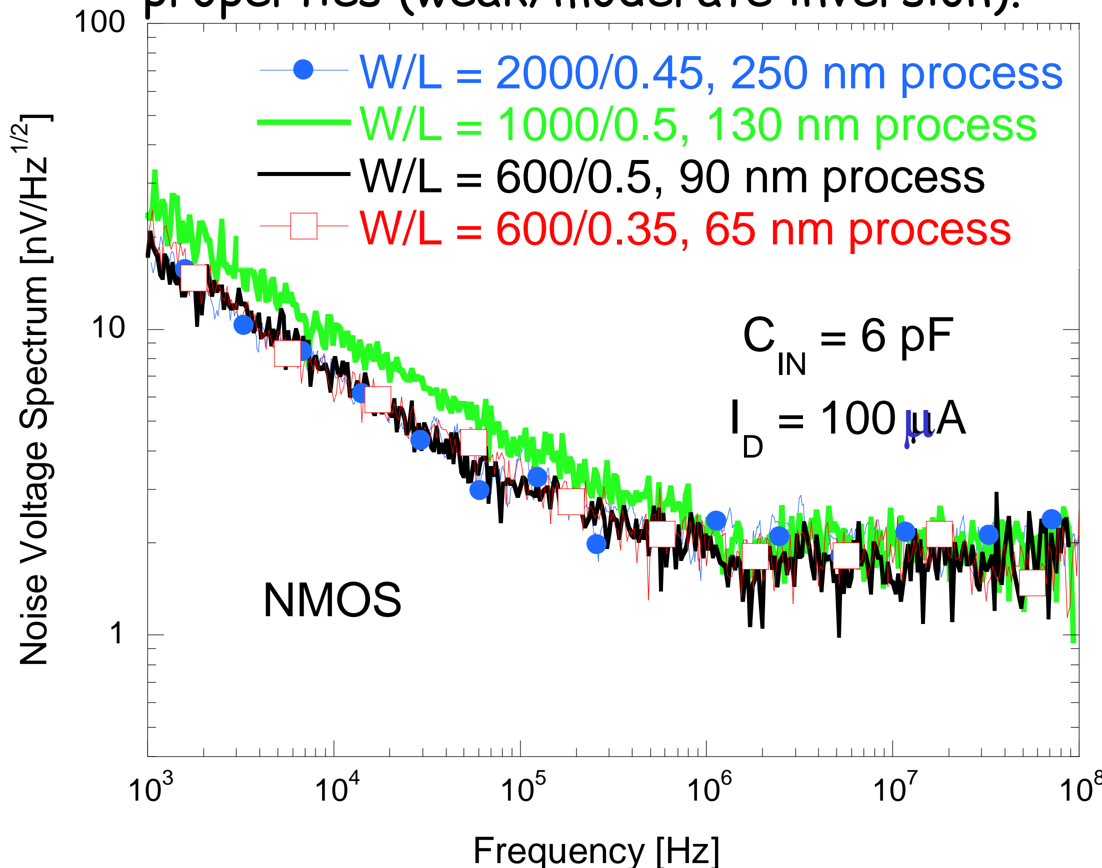
e_n

$$\frac{\overline{de_n^2}}{df} = \frac{4kT\,G}{g_m} + \frac{K_f}{C_{OX} W L f^{a_f}}$$

- 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).

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

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

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

Channel thermal noise

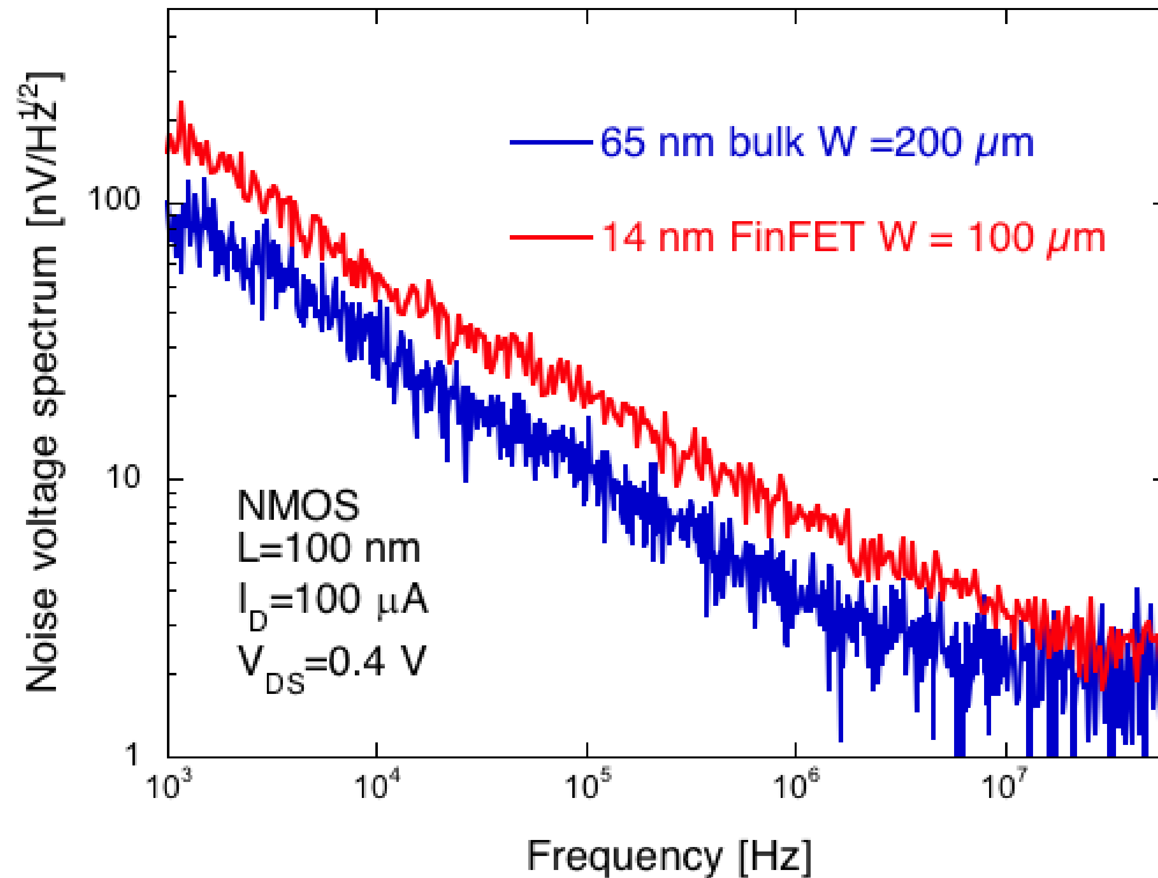
$$S^2_W = 4 k_B T \frac{\Gamma}{g_m}$$

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

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

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

1/f noise in FinFET



In FinFETs (as in 65 nm CMOS) in weak/moderate inversion, thermal noise is not degraded by short channel effects

For a similar gate capacitance, in the tested 14 nm N-type FinFETs 1/f noise is higher (even if not dramatically) than in 65 nm CMOS

From available data, **FinFET transistors appear to be in a similar ballpark as previous CMOS nodes (including 28 nm) with 1/f and thermal noise**

L. Ratti, M. Manghisoni, V. Re, Analog front-end design perspective of a 14 nm FinFET technology, 2019 IEEE NSS

(S. Yang et al, 28nm metal-gate high-K CMOS SoC technology for high-performance mobile applications, 2011 IEEE Custom Integrated Circuits Conference)

Gate leakage current in nanoscale CMOS flavors and generations

- **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)

- The increased thickness of gate oxide made possible by high-k materials takes care of the gate leakage currents in advanced nodes (from need to reduce static power in digital circuits)

- Also from past design experience, gate leakage current does not appear to be a problem for analog circuits

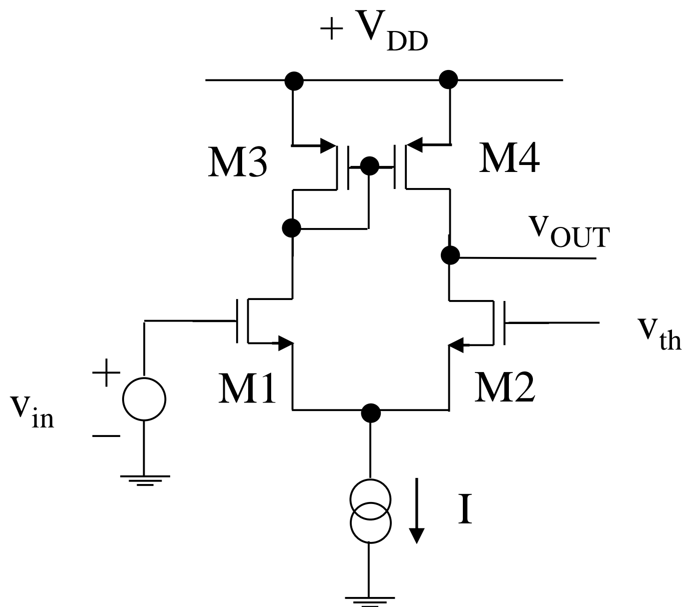
Threshold dispersion

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

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

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

(from stochastic dopant fluctuations model)



Large area transistors help reduce the effect of threshold mismatch

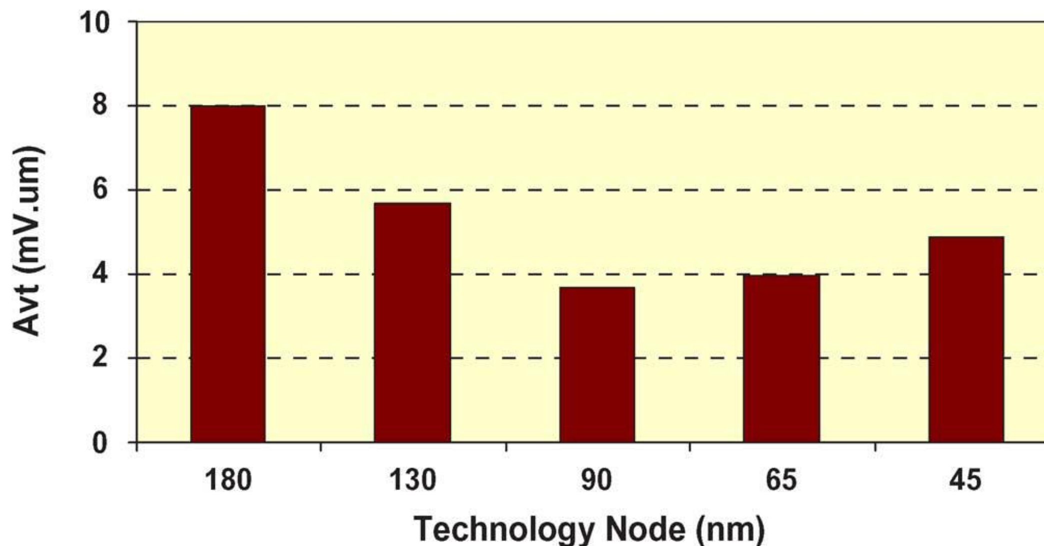
To maintain an adequate efficiency with a small number of noise hits, a channel-by-channel threshold adjustment is often necessary (threshold DAC in a pixel readout cell)

Threshold dispersion in nanoscale CMOS

While the threshold mismatch has the expected reduction with t_{ox} scaling until about 90 nm technologies, in more recent technology generations it has not been decreasing that much. 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.

However, threshold mismatch could be smaller in FinFET, thanks to the better electrostatic control of the gate

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.

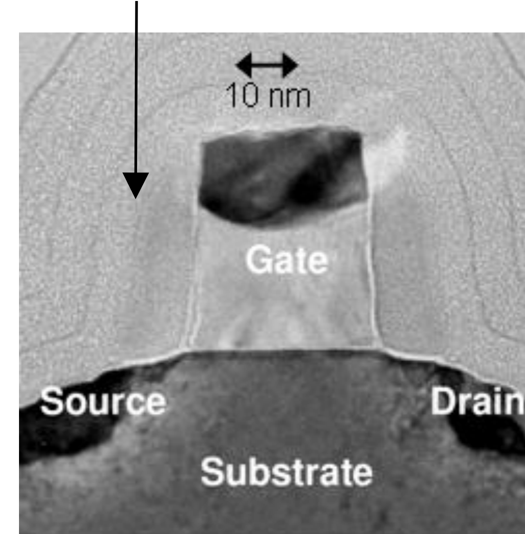


S.-Y- Wu et al, A highly manufacturable 28nm CMOS low power platform technology
2009 IEEE symposium on VLSI Technology

≅ 3 mV.μm for 28 nm, may go even lower with FinFET

Radiation effects in planar CMOS

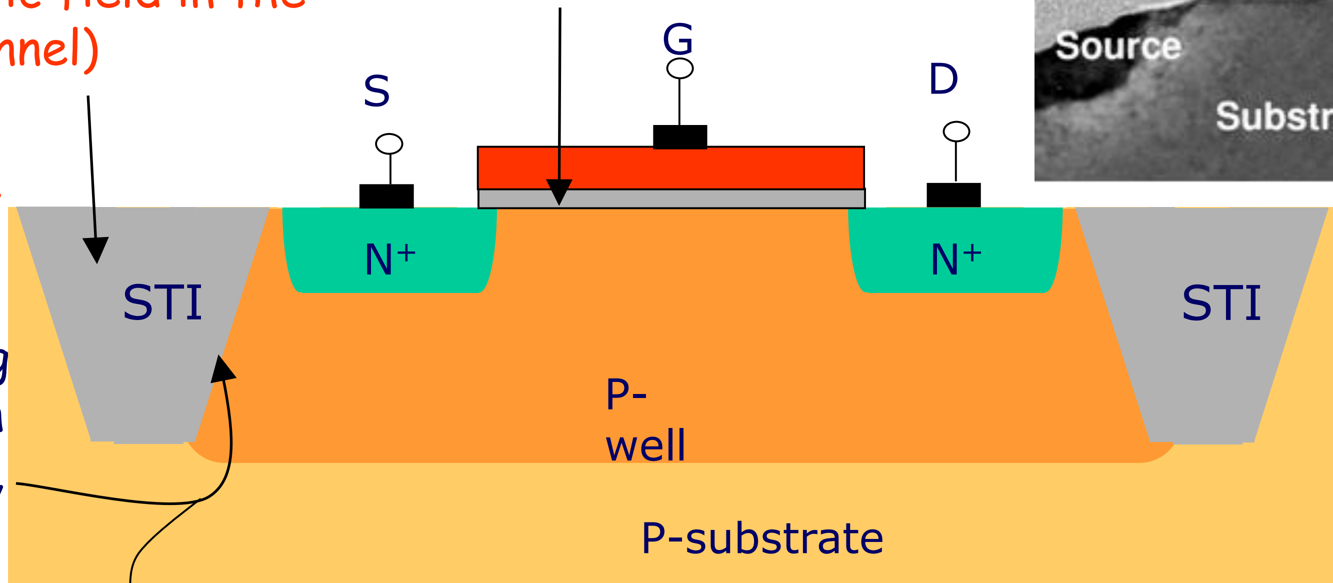
Spacer dielectrics may be radiation-sensitive



Thick Shallow Trench Isolation Oxide (~ 300 nm); radiation-induced charge-buildup may turn on lateral parasitic transistors and affect electric field in the channel)

Thin (rad-hard) gate oxide for core devices, becomes thicker (and rad-softer) for I/O transistors

Doping profile along STI sidewall is critical; doping increases with CMOS scaling, decreases in I/O devices



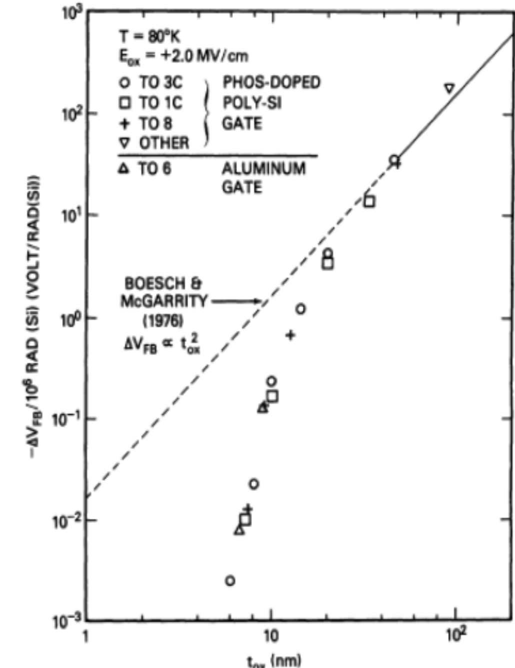
Increasing sidewall doping makes a device less sensitive to radiation (more difficult to form parasitic leakage paths)

Ionizing radiation effects in sub-100 nm CMOS

- Ionizing radiation generates electron-hole pairs in SiO_2 regions; a fraction of holes is trapped, with a **buildup of positive charge**. Holes reaching the interface with the silicon bulk may create **interface states**, which can trap electron or holes depending on the device bias and type

- **Radiation-induced positive charge is removed from thin gate oxides by tunneling** (which also prevents the formation of interface states)

- **Isolation oxides remain thick** (order of 100 nm) also in nanoscale CMOS, and they are **radiation soft**.

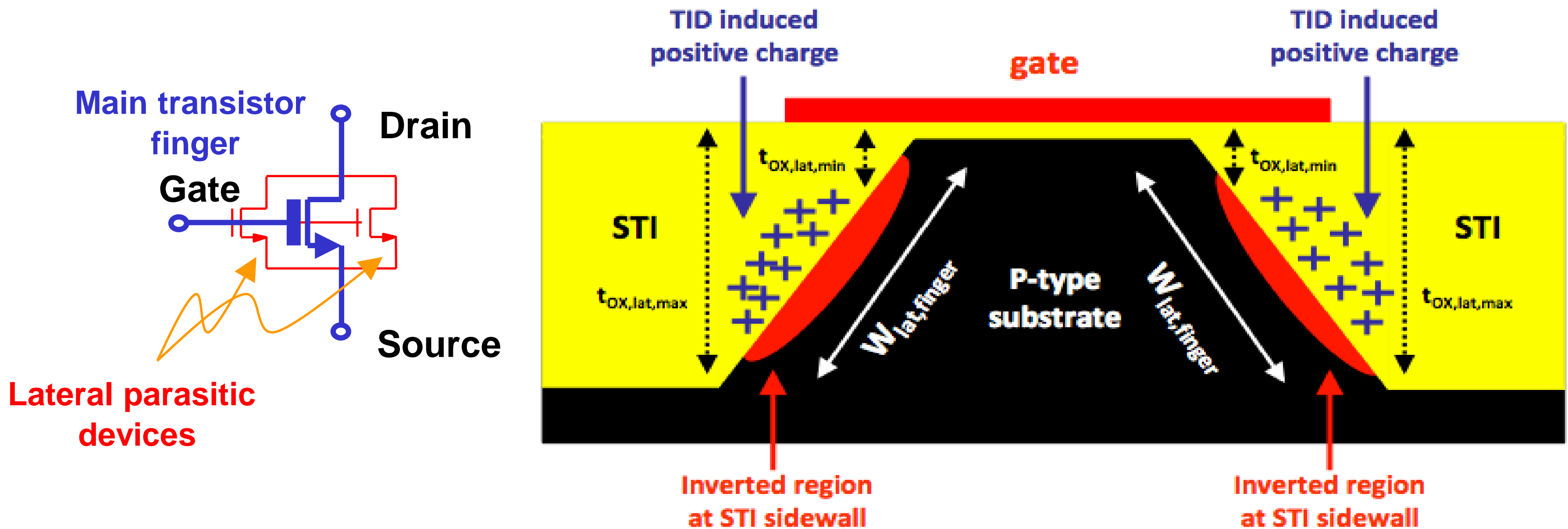


- With scaling, the effect of positive charge buildup in STI oxides appears to be **mitigated by the higher doping of the silicon bulk**.

- However, the radiation-induced noise degradation may be sizable. This is associated to **noisy lateral parasitic transistors**. The use of enclosed devices (when possible) for low-noise functions may help.

Radiation effects in lateral isolation structures

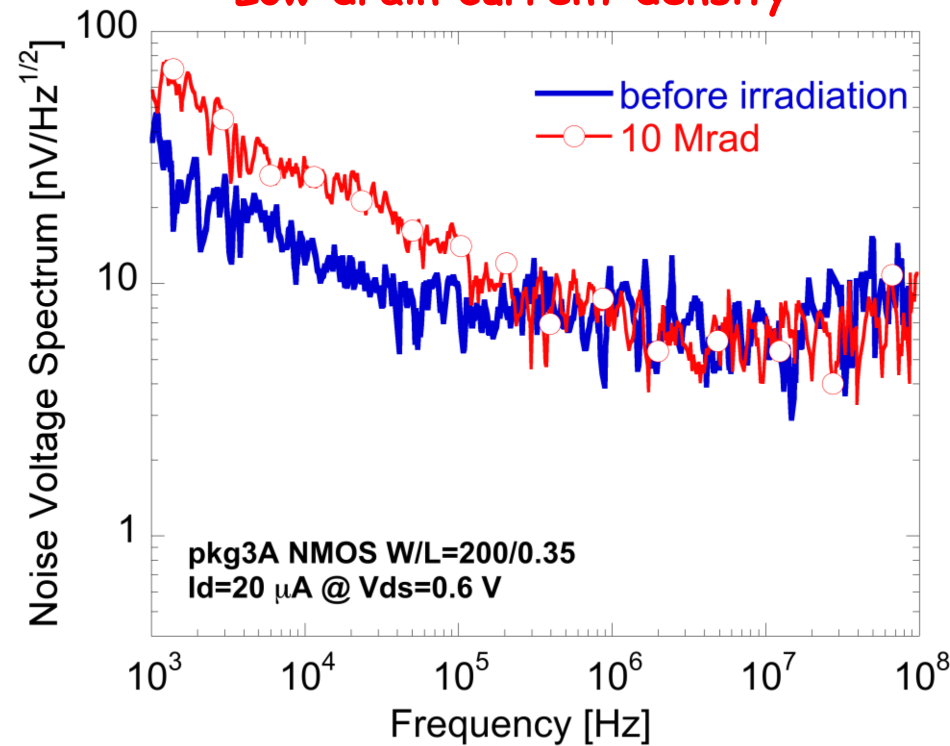
- In an interdigitated device, the **lateral leakage** can be modeled considering that **two parasitic transistors for each finger** are turned on



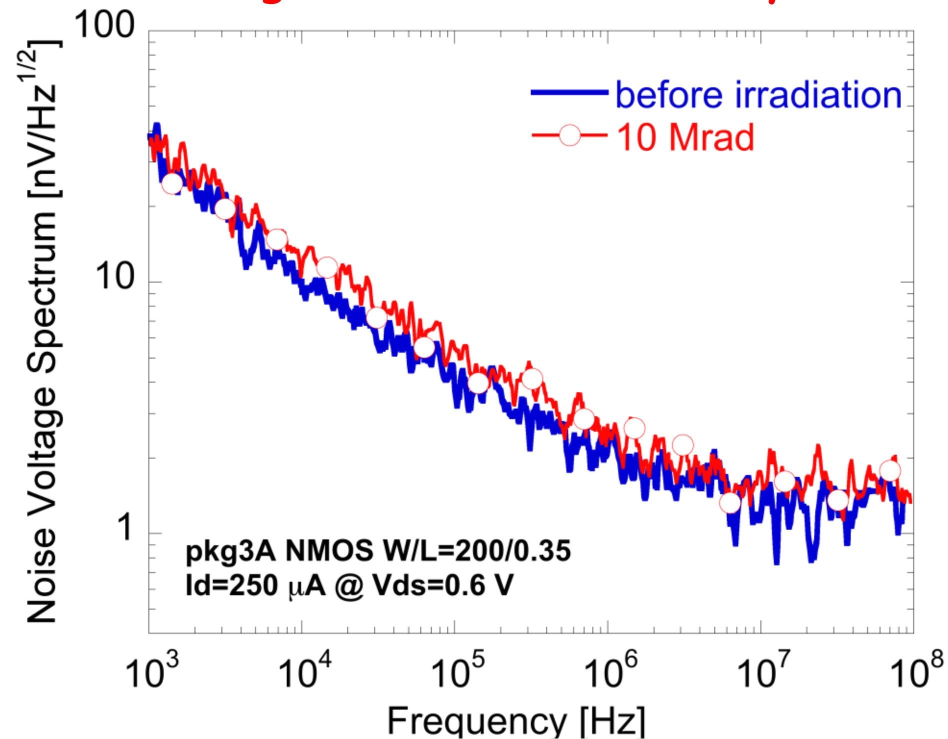
- The two lateral parasitic transistors have an effective gate width $\mathbf{W_{lat,finger}}$, the same gate length **L** as the main device and gate oxide capacitance $\mathbf{C_{OX,lat}}$ inversely proportional to an effective oxide thickness $\mathbf{t_{OX,lat}}$
- The parasitic devices **add a contribution to the total drain current and noise** of NMOSFETs

NMOSFETs
TSMC LP 65 nm technology – 10 Mrad

Low drain current density



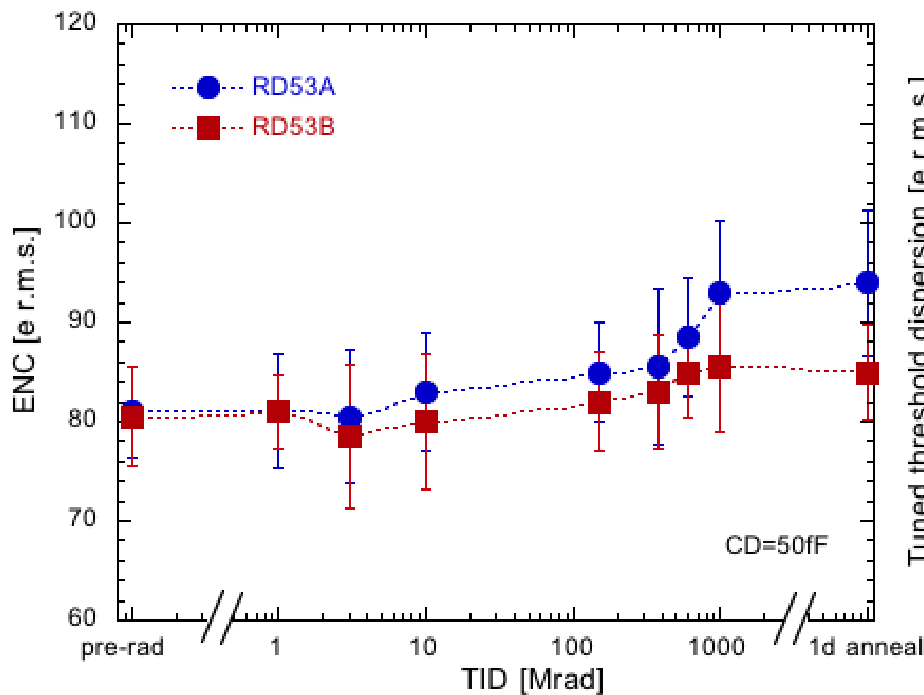
High drain current density



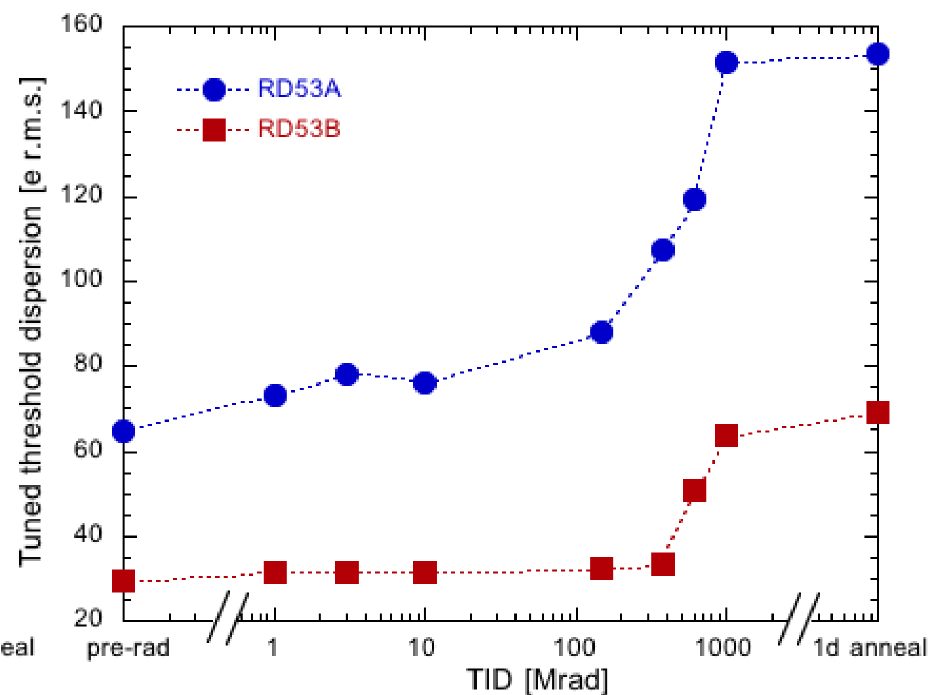
- **Moderate 1/f noise increase at low current density**, due to the contribution of lateral parasitic devices
- At **higher currents** the degradation is almost negligible because the impact of the parasitic lateral devices on the overall drain current is smaller
- **No increase** in the **white noise** region is detected

Radiation hardness of a 65 nm CMOS prototype chip with the RD53B Linear Front-End

Noise



Tuned threshold dispersion



Tolerance to high Total Ionizing Dose of nanoscale CMOS

In RD53, the extensive characterization of the LP 65 nm CMOS technology led to the definition of analog design guidelines to prevent degradation of transconductance and excessive threshold voltage shift
(Wp ≥ 300nm Lp ≥ 120nm Ln ≥ 120nm)

Can similar criteria be defined for 28 nm CMOS, for FinFET and GAA processes?
What is the noise behavior at extremely high TID?
Can a 28 nm CMOS chip (or, e.g., a 14 nm or a 5 nm one) work with acceptable performance at TID > 1 Grad?
(see the excellent and extensive work by CERN, Padova et al, to characterize radiation hardness of 28 nm CMOS at very high total ionizing dose)

A significant radiation-induced parasitic leakage current can be observed for bulk FinFETs due to charge trapping in isolation oxides, particularly for narrow-fin transistors

(D. Fleetwood, Evolution of Total Ionizing Dose Effects in MOS Devices with Moore's Law Scaling, IEEE TNS, 2017)

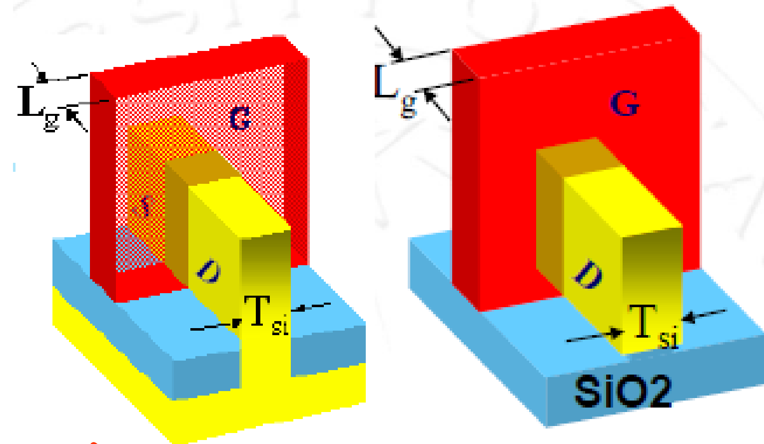
Evolution of scaling and radiation hardness

Alternate gate dielectrics

- With a high dielectric constant (high-k) material, a much thicker gate dielectric can be used, with the equivalent capacitance of much thinner SiO_2-based structures (in ≤ 45 nm CMOS).
- Thicker dielectrics are more sensitive to ionizing radiation; as always, actual behavior will depend on process details. Hafnium-based dielectrics with good radiation tolerance have been reported

Advanced multiple-gate devices

- "3D" gate structures have been devised as a way to avoid short-channel effects in aggressively scaled MOSFETs (≤ 22 nm). Control of lateral gates on silicon channel may be beneficial in terms of radiation tolerance (no lateral leakage). However, radiation effects in these advanced devices may be more complex than in bulk MOS



Carbon-based electronics (Beyond CMOS)

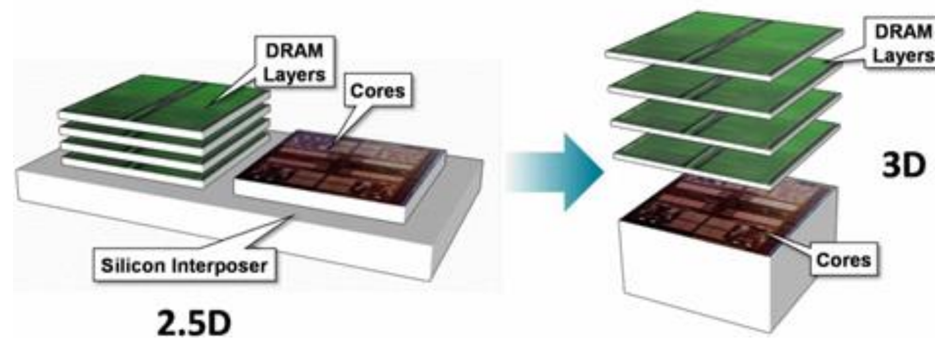
- Carbon nanotubes and graphene have generated much interest: not yet clear if they will be a replacement for Si CMOS. Because of their extremely low volumes (few atomic layers), their radiation response may mostly depend on interfaces and surrounding materials.

3D integration as a tool to advance the state of the art of pixel sensors

- **The increase of functional density can be achieved by stacking layers of electronics, vertically interconnected by Through-Silicon Vias (TSV)**

  ⇒ interconnect delays can be reduced

  ⇒ each layer can be optimized for a dedicated function (sensing, analog processing, DSP, memory, optical data transmission)



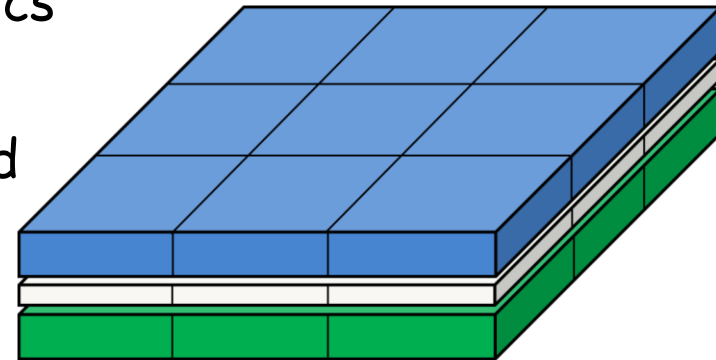
- For our pixel sensors, 3D integration could be leveraged to shrink pixel size and pitch, increase pixel-level electronic functions, reduce dead areas, decrease amount of material by aggressive thinning

3D integration as a tool to advance the state of the art of pixel sensors

- **Improve resolution ⇒ shrink pixel size and pitch**, down to about 20 μm or even less

- **Preserve or even increase pixel-level electronic functions**
  handling of high data rates, large dynamic range, high resolution analog-to digital conversion and timing, sparsification, large memory capacity, intelligent data processing, independent analog and digital substrates: the limit to the minimum size of pixel readout cells can be overcome with multiple tiers of electronics

- **4-side buttable tiles**
  large area detector with minimum or no dead area, thanks to TSV and backside metallization and patterning (RDL)

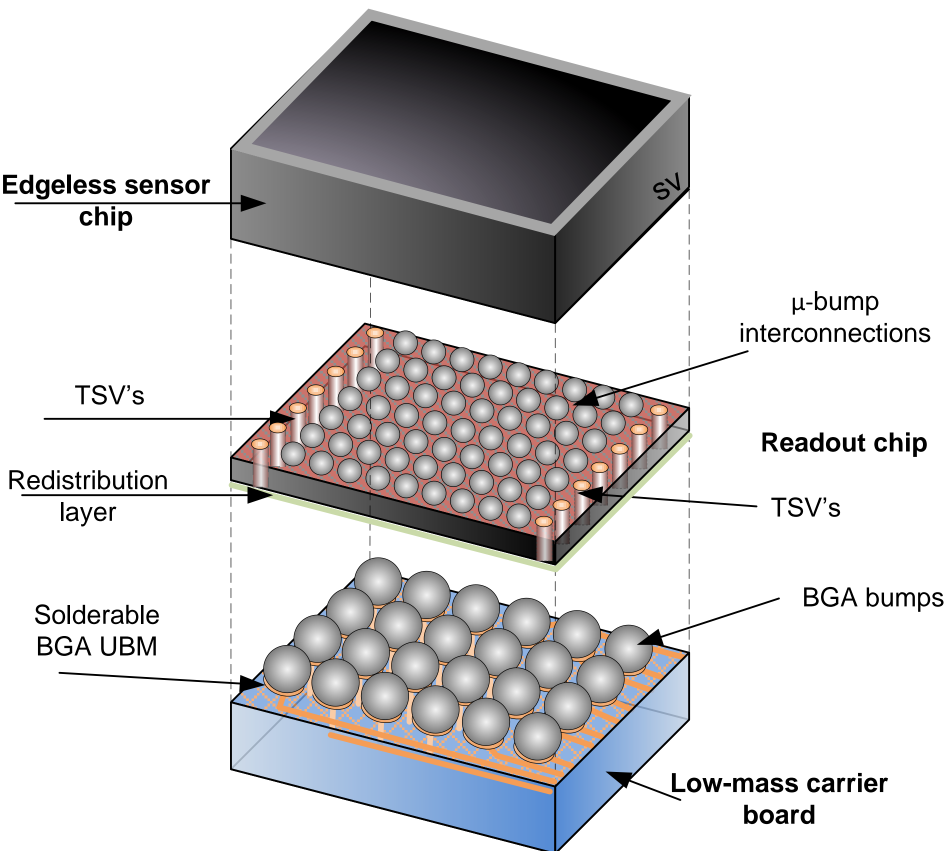


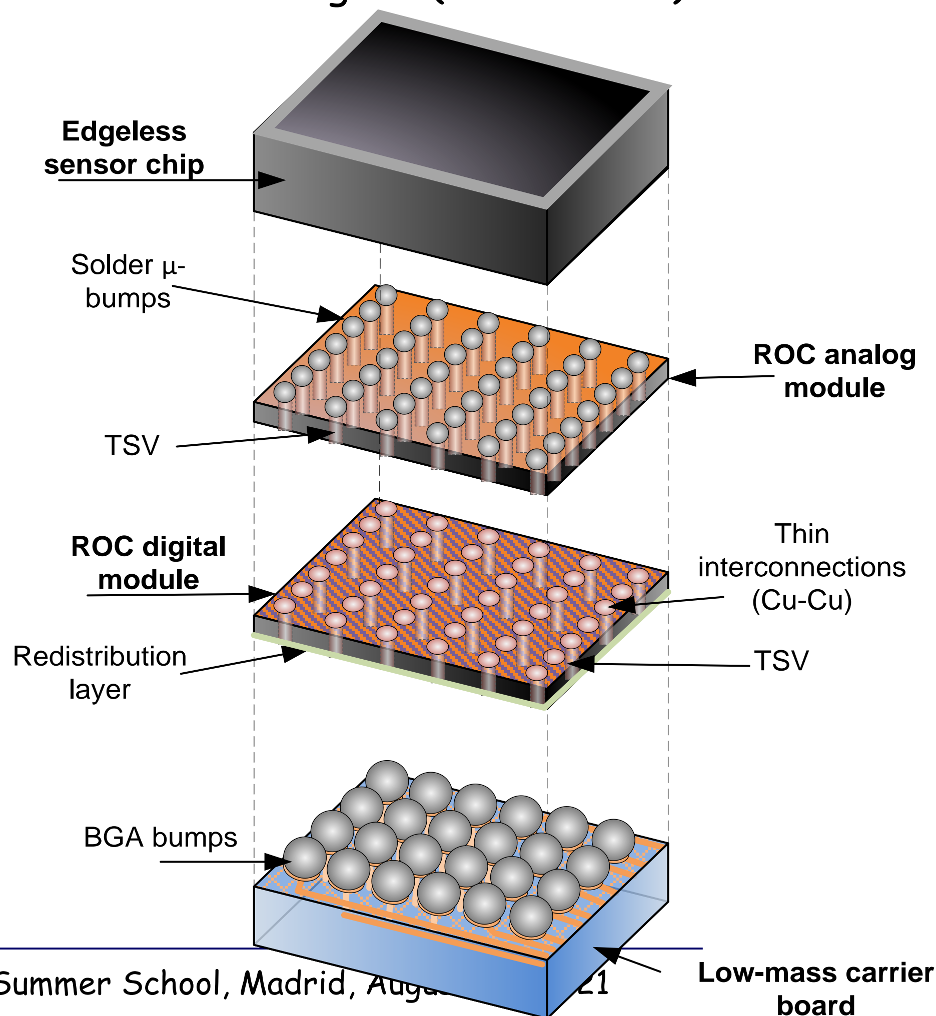
- **Decrease amount of material ⇒ thin sensor and electronics** reduce errors in track reconstruction due to multiple scattering of particles in the detector system (thinning to a 50 -100 μm total thickness, needed to optimize TSV geometry)

Pixel sensor tiles

Peripheral TSV (via last)
Single layer readout chip
TSV carry I/O to backside
Pixelated ASIC layout with I/O, control and other functions in the periphery

Small pitch TSV (via middle)
Double layer readout electronics
Signal exchange through dense inter-tier bonding interfaces
Separate functions of analog (pixelated) and digital (distributed) tiers





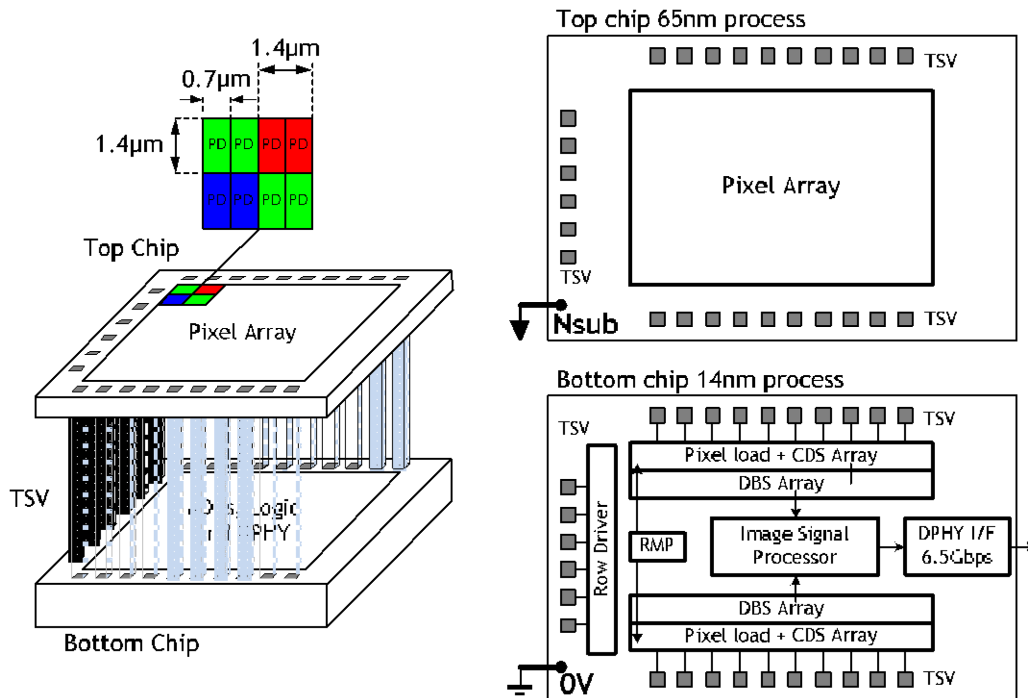
MUX 2010 WORKSHOP,
SAMI VÄHÄNEN – CERN PH-ESE

3D in commercial microelectronics: imagers

- "BSI and 3D-stacked processing continue to offer improved performance with increased on-chip functionality and new features being integrated at the pixel level" (from the ISSCC 2020 report)
- A clear industrial technology trend is based on the stacking of CMOS image sensors with a CMOS mixed-signal readout chip, both in decananometer technologies (see also ISSCC2021)
- sub-μm pixel CMOS image sensors have been fabricated thanks to small pitch bonding interfaces and wafer stacking
- 3-layer devices with image sensor, RAM, and logic are available thanks to high-density TSV, opening the way to event-based imaging, to AI processing and machine learning
- The resulting architectures may stimulate interesting ideas for particle tracking detectors

Image sensor 3D-stacked with a FinFET readout

M. Kwon et al., A Low-Power 65/14nm Stacked CMOS Image Sensor", Samsung, ISCAS 2020

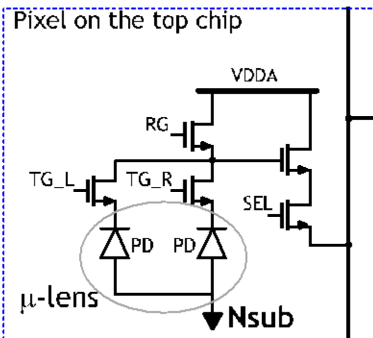


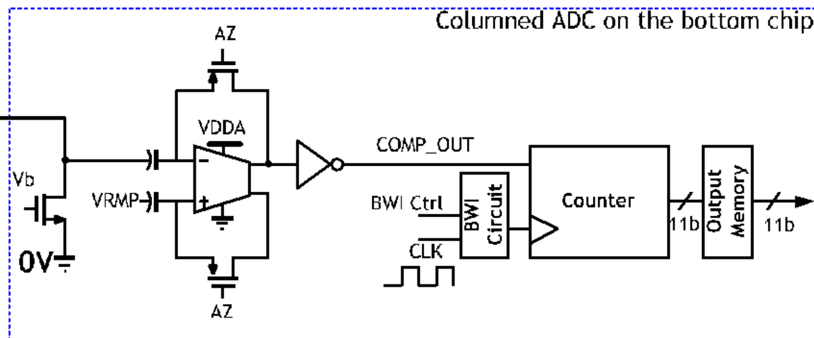
65 nm backside illuminated CMOS Image Sensor

output signal from pixels on the top chip is transferred to correlated double sampling (CDS) circuits on the bottom chip throughout TSV

14 nm CMOS readout chip (with 3D FinFET transistors)

single-slope ADCs, row driver to control pixels on the top chip, image signal processor (ISP) and mobile industry processor interface

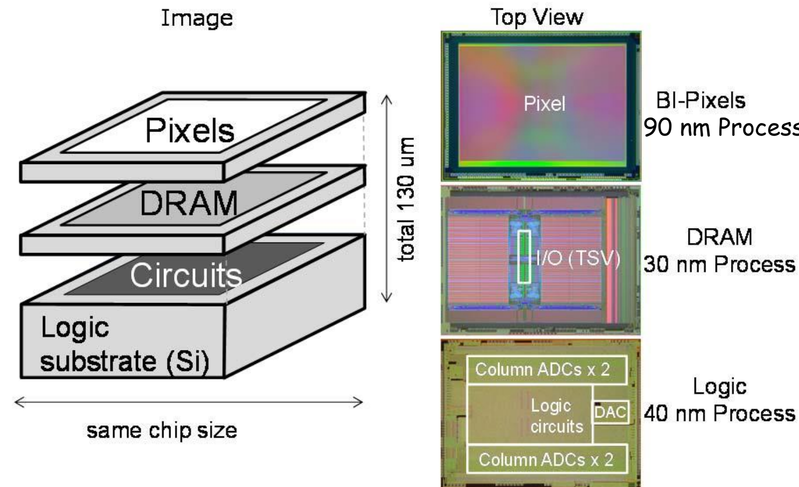




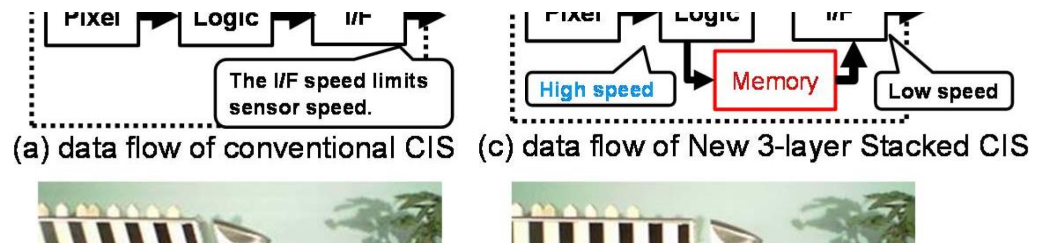
Wafer level stacking
pixel pitch 1.4 µm,
12 Mpixel
11-bit column parallel ADC

CMOS image sensors with 3 device layers: pixels, DRAM, logic

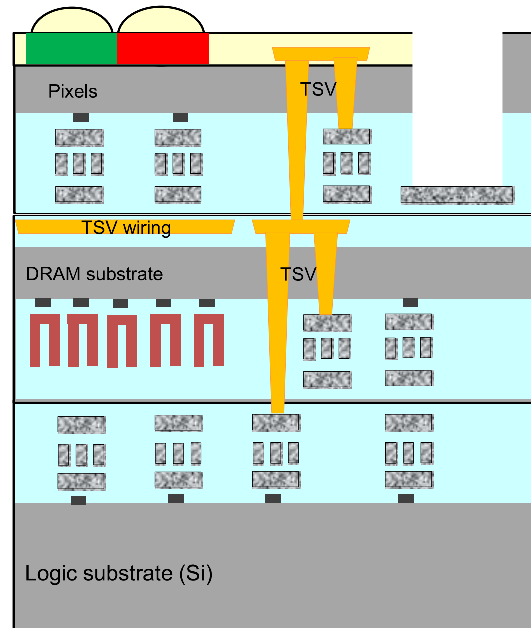
"Pixel/DRAM/logic 3-layer stacked CMOS image sensor technology" H. Tsugawa et al. Sony, 2017 IEEE IEDM



- Three bonded Si substrates, each electrically connected by TSVs through sensor or DRAM (thinned to 3 μm)
- Thanks to DRAM, readout speed from the pixel can be increased (960 fps super slow motion video) without being limited by the speed of the I/O interface



(a) data flow of conventional CIS (c) data flow of New 3-layer Stacked CIS



TSV have a minimum diameter of 2.5 μm and a pitch of 6.3 μm (35000 TSV ≅ number of row and columns)

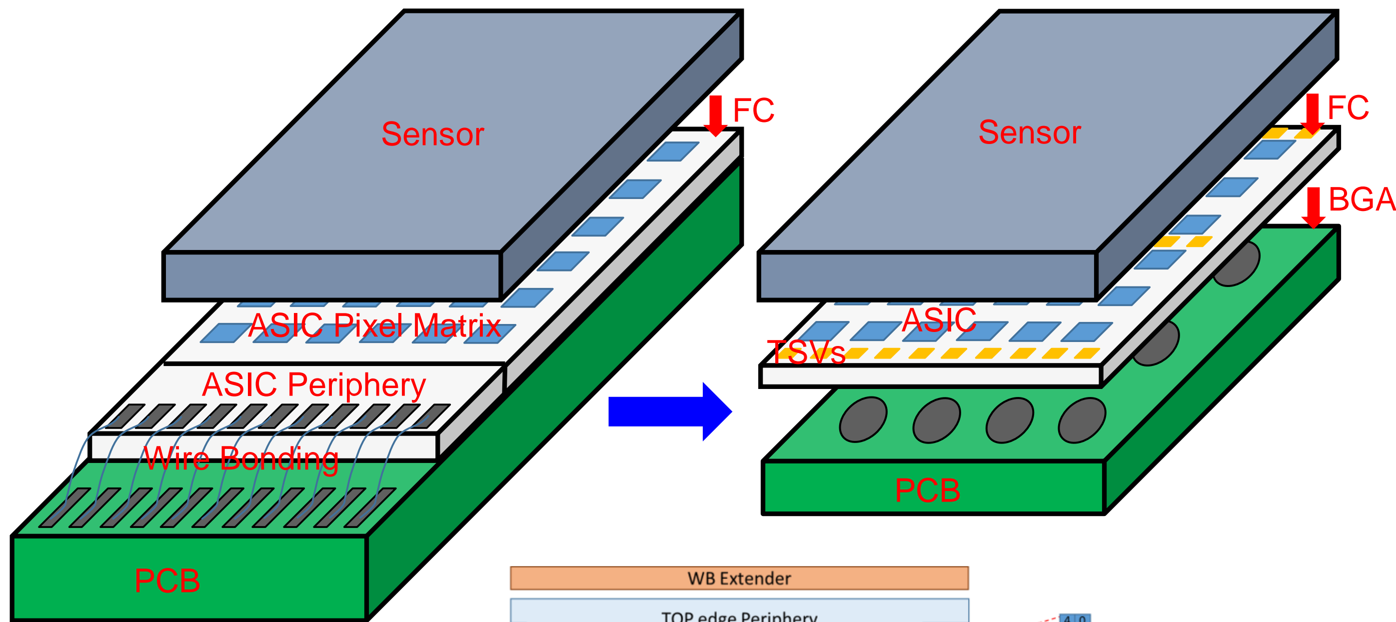
19.3 Mpixel, 1.22 μm x 1.22 μm pixels

Samsung version with 28 nm logic and 20 nm DRAM

3D stacked CMOS sensors and advanced functionalities

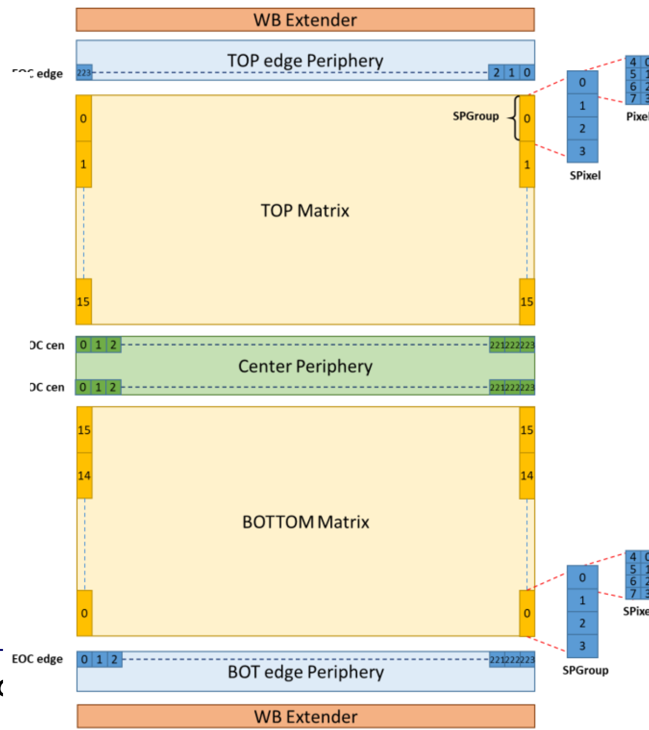
- Event-based sensors respond to brightness changes asynchronously and independently for every pixel (high temporal resolution and low latency, very high dynamic range, and low power consumption)
- In these data-driven devices, each pixel continuously monitors for a change of sufficient magnitude from a memorized value. When the change exceeds a threshold, the camera sends an event, transmitted from the chip with x-y location, time-stamp, and polarity of the change
- 3D stacking allows for higher complexity of pixel electronics at small pitch without degrading fill factor (embedding smart functions such as object recognition, movement detection, using AI and machine learning)
- Architectures moving closer to what is needed by pixel sensors for particle detection

3D integration for the next generation of TIMEPIX



TIMEPIX4 chip
large single threshold particle tracking detector chip with improved energy and time resolution (sub-ns) and with high-rate imaging capabilities

3 "hidden" peripheries with TSV:
- TOP, BOTTOM Edge: Data Readout (8x 10 Gbps Serializers)
- CENTER: SC, Analog Blocks (DACs, ADC, Band-Gaps…)

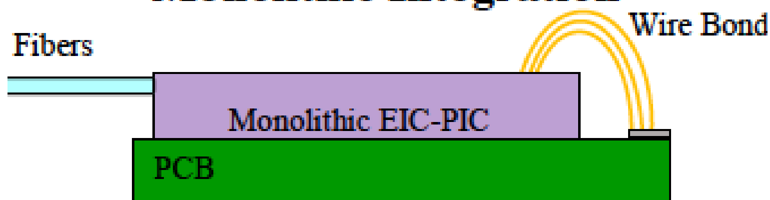


- **functions normally associated with the chip periphery located throughout the pixel matrix** (digital-on-top design)
- **more flexibility in the choice of readout architectures**.
- Chips abuttable on 4 sides.

X. Llopart, "The Timepix4 chip and its design approach ", VERTEX2019, Lopud, Croatia

3D integration developments for pixel detector in HEP and photon science

- 3D chip stacking and high density bonding are being successfully used for industrial image and TOF high performance sensors based on advanced CMOS ≤ 65 nm
- These technologies can be very interesting also for new detectors in our field, e.g. for high resistivity CMOS sensors, LGAD, SiPM,...
- 3D integration of a readout chip with a silicon photonics device may open way to high-rate data transmission (≥ 100 Gb/s)

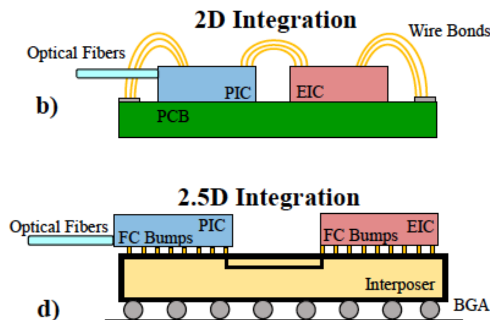


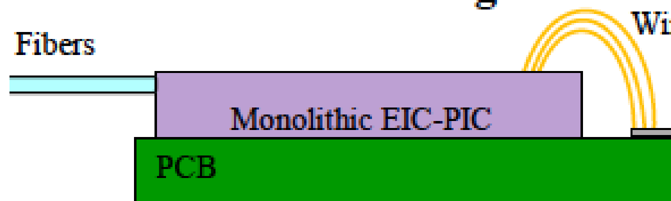
Journal of Lightwave Technology 2020
Silicon Photonic 2.5D Multi-Chip Module Transceiver
for High-Performance Data Centers, N. C. Abrams, et al

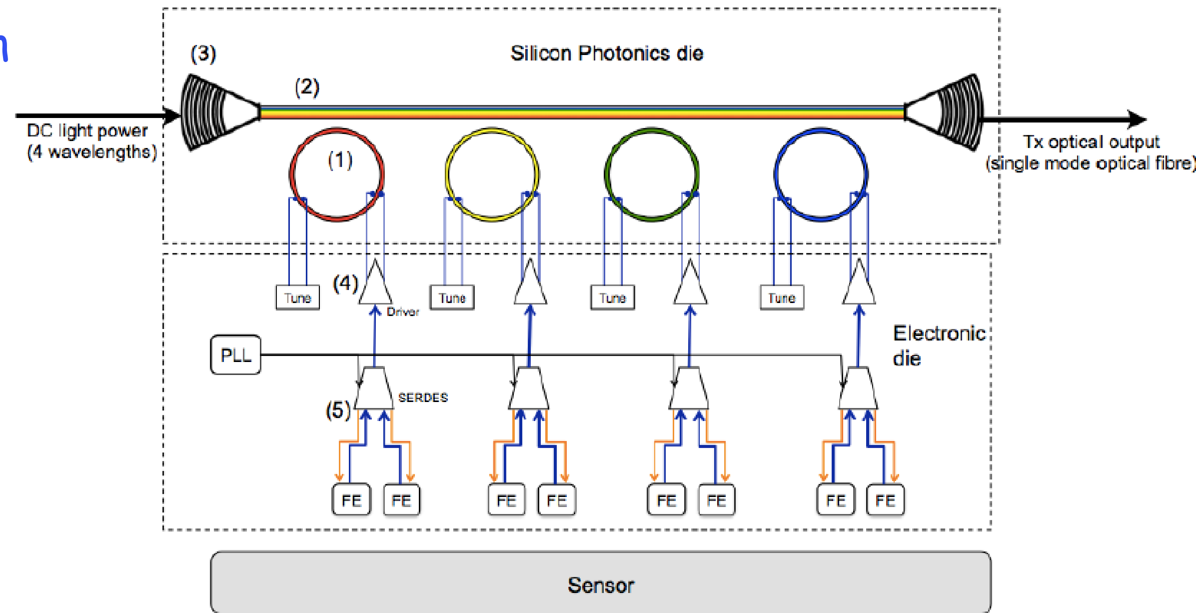
A potential application of 3D integration: the INFN FALAPHEL project

Integration of Silicon Photonics and high-speed microelectronics for high rate data transmission, operating at extremely high dose levels (≥ 1 Grad), and 100 Gb/s data rate, using wave/space division multiplexing techniques.

Hybrid (3D or 2.5D) integration of Silicon Photonics modulators with high speed radiation hard 28 nm pixel readout ASIC







In the more aggressive solution, based on 3D integration, the readout ASIC (bonded to a sensor) is flipped on top of the photonic chip

Journal of Lightwave Technology 2020
Silicon Photonic 2.5D Multi-Chip Module Transceiver for High-Performance Data Centers, N. C. Abrams, et al

(P.I. F. Palla; INFN Padova, INFN Pavia, INFN Pisa, Univ. Pisa, Univ. Bergamo, S.S. S.Anna, Pisa)

Embedding intelligence in a pixel readout chip

More intelligence can be integrated in the front end ASICs to allow for data reduction and possibly an overall system power reduction, though this is a trade-off with optimal reconstruction using off-detector electronics or software.

Such complex logic at the front-end must be adaptable to changing experimental conditions. Ideally it would be programmable, and 'FPGA-like' or 'CPU-like'.

This programmability and configurability will ultimately enable the community to develop fewer ASICs of higher complexity.

SOC Radiation Tolerant Ecosystem

Proposed Scheme

Draft 1.0

Analog functions

ADC

DAC

Interconnect Bus

Slow Control Interfaces

JTAG

GPIO

SPI

I²C

FE channel

FE analog blocks

eFPGA fabric

micro-code

SRAM

RISC-V

SRAM

data

Programmable Logic

e-Link

any-Link

High Speed Interfaces

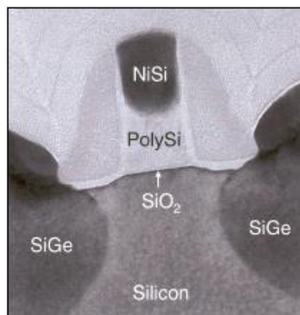
Conclusions

- Exploiting progress in microelectronic technology is essential to achieve the unprecedented performance requirements of future detector generations
- The progress of microelectronic technology is not slowing down: transistors are advancing thanks to new materials and architectures, 3D integration already allows for high connectivity between sensing, logic and memory layers
- Infrastructures at the HEP institutes for the design of complex mixed-mode CMOS ASICs have to be built up to match future challenges
- The design of the front-end electronics has to be considered as a crucial aspect in a system-level development of new solid-state sensors

Backup slides

Nanoscale MOSFETs

65 nm Transistor



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

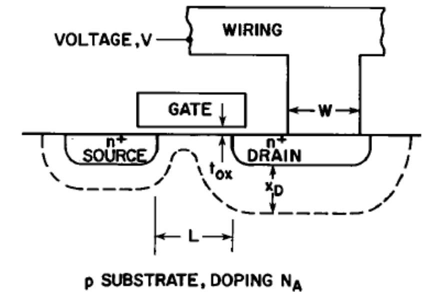
TENSILE STRESS Fz (GOOD N - GOOD P)
COMPRESSIVE STRESS Fy (GOOD N - OK P)
COMPRESSIVE STRESS Fx (GOOD P - BAD N)
IMPLANT RANGE VARIABILITY Δy
OXIDE TRAP DENSITY (Not)
INTERFACE TRAP DENSITY (Nit)
TRANSVERSE FIELD (Ey)
LATERAL FIELD (Ex)
CAP LAYER
POLY/METAL GATE G
Ig Vg
Fy Ey
Ge IMPLANT RAISED S
Ge IMPLANT RAISED D
SPACER
ON or Hi-K
SPACER
Vb P+ TIE
Vs N+ S
Fx CHANNEL Ex
N LDD
N LDD
Vd N+ D
Vb P+ TIE
Fz
Δy
HALO
Id
Vsb
Vdb
STI
STI
SA
SB
P-TUB (BULK B)
R1
R2
JUNCTION
GATE EDGE

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 ($< 2\ \mu m$) 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.*

The old ways of CMOS scaling

- **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**)

Device or Circuit Parameter	Scaling Factor
Device dimension t_{ox}, L, W	$1/\kappa$
Doping concentration N_a	κ
Voltage V	$1/\kappa$
Current I	$1/\kappa$
Capacitance $\varepsilon A/t$	$1/\kappa$
Delay time/circuit VC/I	$1/\kappa$
Power dissipation/circuit VI	$1/\kappa^2$
Power density VI/A	1



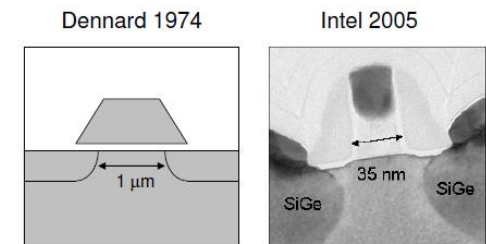
R. Dennard, IEEE JSSC, 1974

Classical MOSFET scaling was first described in 1974

- 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.

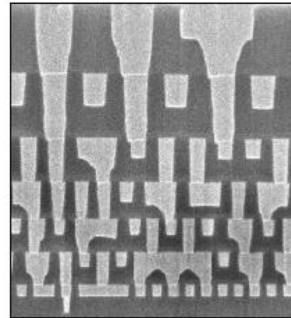
30 Years of MOSFET Scaling



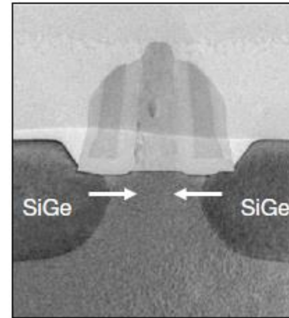
	Dennard 1974	Intel 2005
Gate Length:	1.0 µm	35 nm
Gate Oxide Thickness:	35 nm	1.2 nm
Operating Voltage:	4.0 V	1.2 V

The New Era of Device Scaling

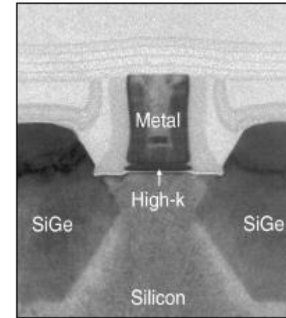
M. Bohr (Intel), "The new era of scaling in a SoC world", ISSCC 2009.



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 reducing gate capacitance) by using materials with higher dielectric constant than SiO_2.

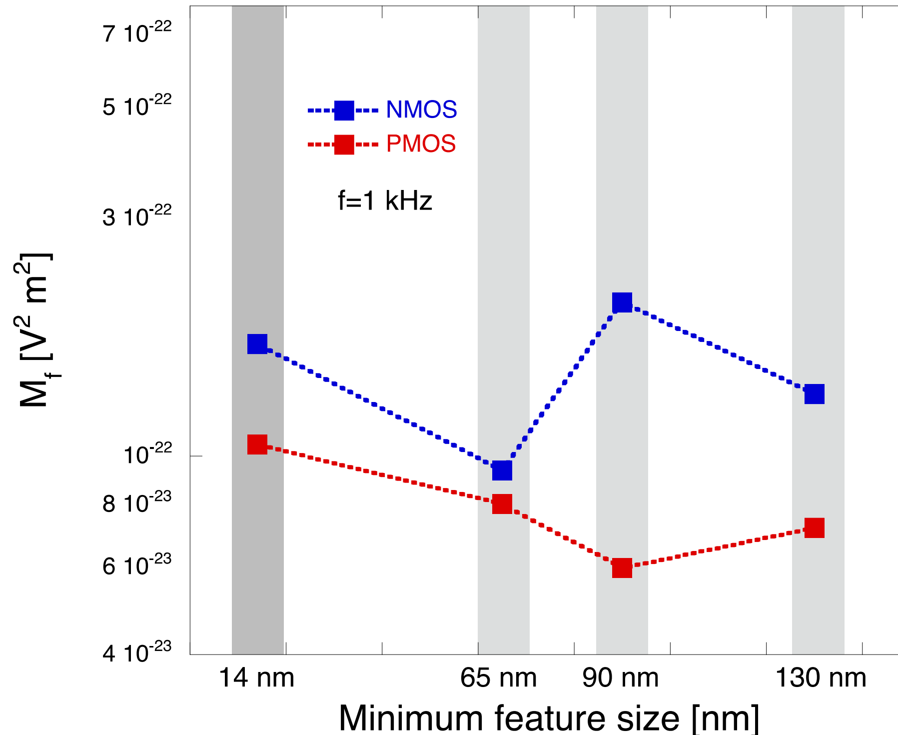
Elements used in Silicon Chip Fabrication

1980's
1990's
2000's

Courtesy: J.P.Colinge

Preliminary FinFET studies

Low frequency noise power parameter



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

- To account for the difference in α_f

$$M_f(f) = \frac{K_f}{C_{OX}} f^{\alpha_f - 1}$$

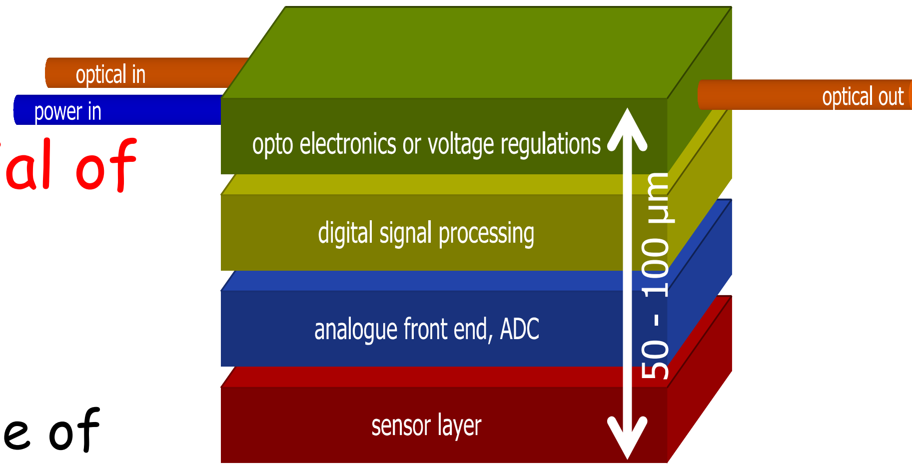
- From available data, **FinFET transistors appear to be in a similar ballpark as previous CMOS nodes (including 28 nm) with 1/f and thermal noise**

L. Ratti, M. Manghisoni, V. Re, *Analog front-end design perspective of a 14 nm FinFET technology*, 2019 IEEE NSS

(S. Yang et al, *28nm metal-gate high-K CMOS SoC technology for high-performance mobile applications*, 2011 IEEE Custom Integrated Circuits Conference)

What is 3D integration?

- 3D electronics: "integration of thinned and bonded silicon integrated circuits with vertical interconnects between the IC layers."[1]



- 3D electronics has the potential of being:
  - Denser (smaller form factor)
  - Faster (reduced delay because of shorter interconnects)
  - Lower power (smaller interconnect capacitance)
  - Lower cost (sizably less expensive than aggressive CMOS scaling)
  - Integration of dissimilar technologies (sensor, analog, digital, optical)

1) Philip Garrou, Christopher Bower, Peter Ramm, Handbook of 3D Integration Technology and Applications of 3D Integrated Circuits,Wiley-VCH, 2008.