ECFA

European Committee for Future Accelerators

Front-end electronics for future trackers in the next 20 years:

the third dimension

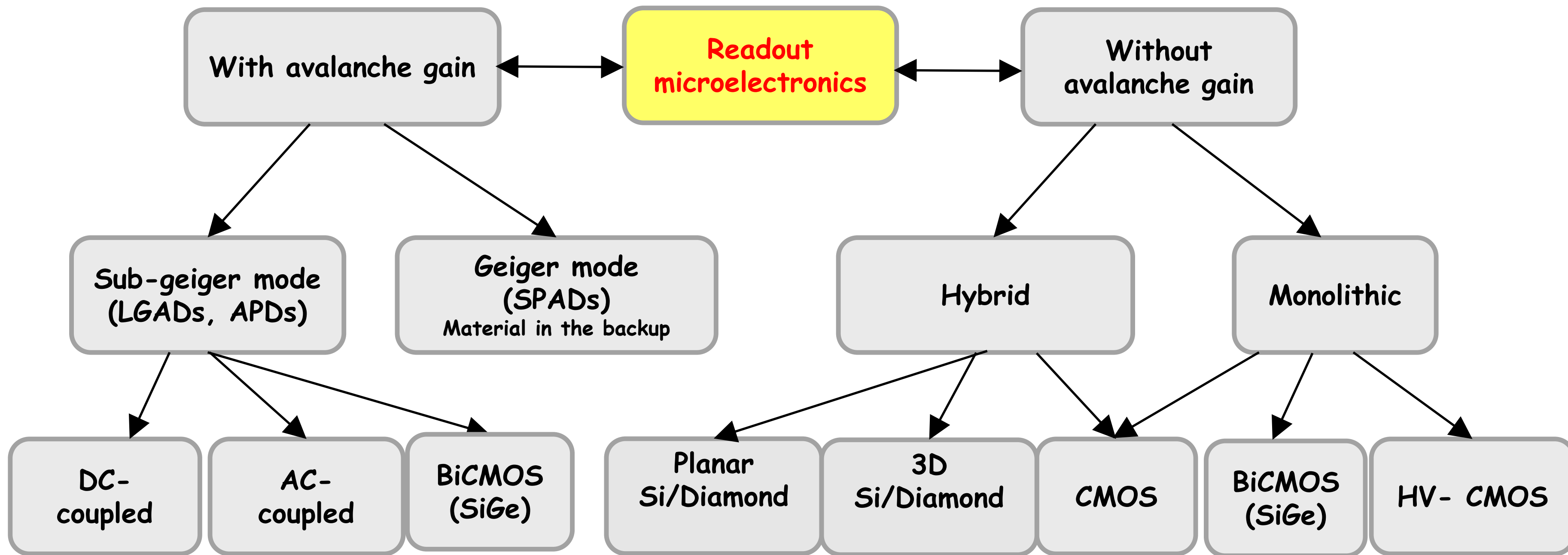
Valerio Re

UNIVERSITÀ DEGLI STUDI DI BERGAMO | Dipartimento di Ingegneria e Scienze Applicate

INFN

Sezione di Pavia

Solid state detectors for future (4D) trackers



Front-end electronics as an integral part of particle detectors

- Industrial microelectronic technologies are today crossing the 10 nm frontier bringing CMOS into the nanometer world
- Digital performances (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 technique are allowing microelectronics to go beyond conventional scaling limits

Advanced pixel detectors and readout microelectronics

Particle tracking at LHC- Phase II:

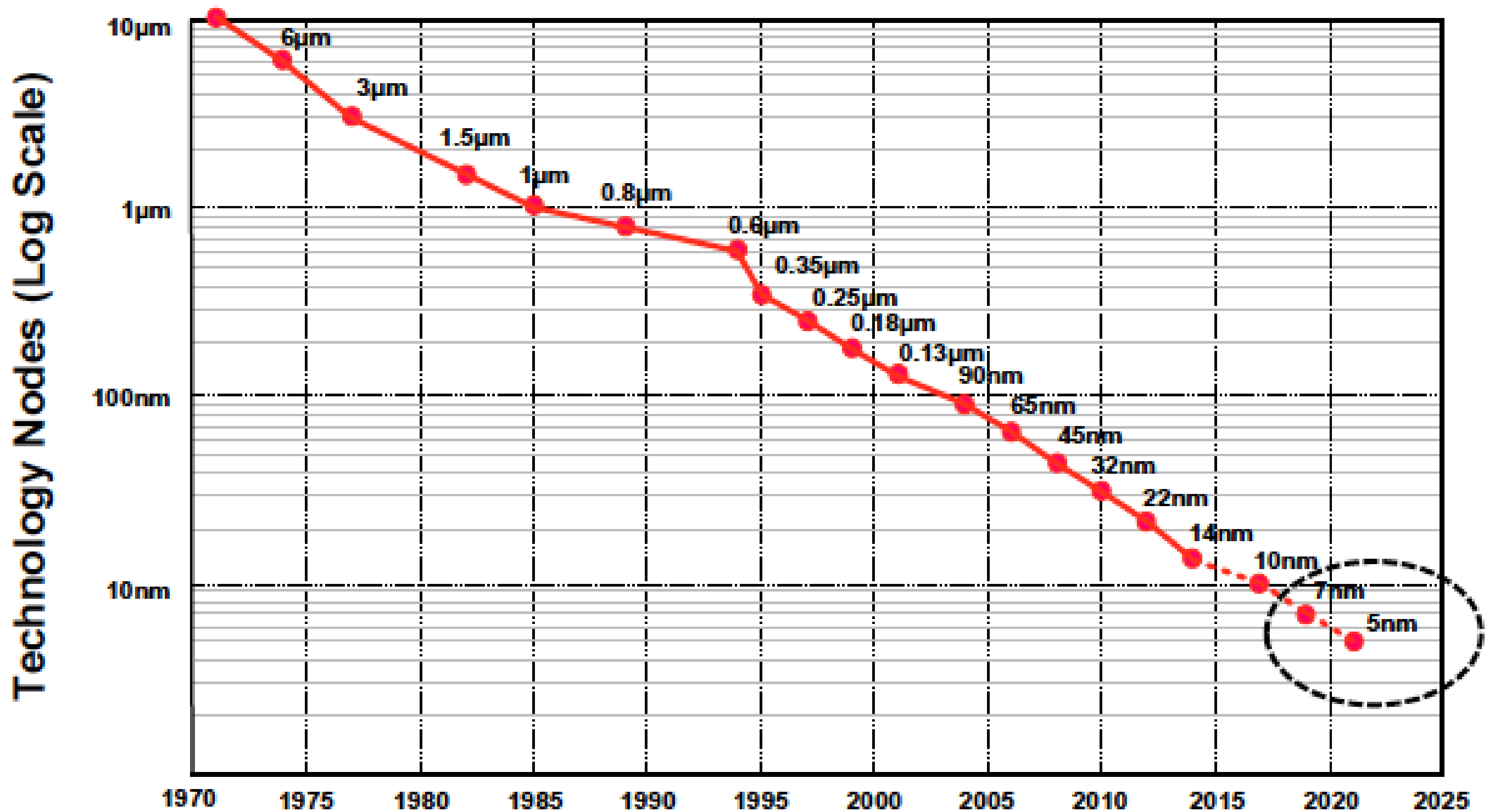
- Very high hit rates (3 GHz/cm^2), need of an intelligent pixel-level data processing
- Small detector signals require operating at low threshold (< 1000 electrons)
- Very high radiation levels (1 Grad Total Ionizing Dose, 10^{16} neutrons/cm^2)
- Small pixel cells to increase resolution and reduce 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

Scaling of microelectronic processes

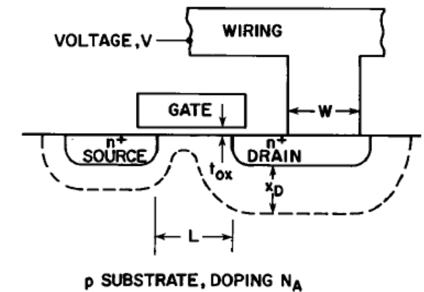


Source: Semiconductor device fabrication, https://en.wikipedia.org/wiki/Semiconductor_device_fabrication

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 *tox, L, W*	$1/\kappa$
Doping concentration *Na*	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

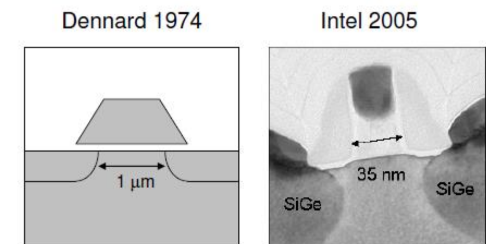


R. Dennard, IEEE JSSC, 1974

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**.

30 Years of MOSFET Scaling

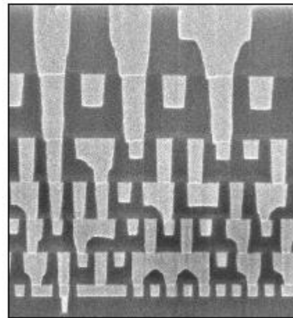


	Dennard 1974	Intel 2005
Gate Length:	1.0 µm	35 nm
Gate Oxide Thickness:	35 nm	1.2 nm
Operating Voltage:	4.0 V	1.2 V

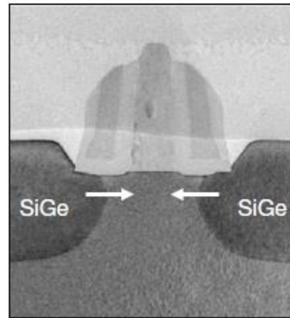
- Classical scaling ended because of gate oxide thickness limits: in very thin oxides, direct tunneling of carriers leads to a large gate leakage current.

The New Era of Device Scaling

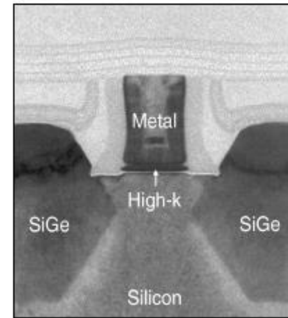
M. Bohr (Intel), "The new era of scaling in a SoC world", ISSCC 2009.



Copper + Low-k

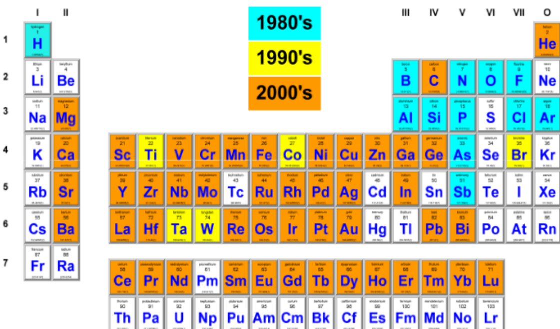


Strained Silicon



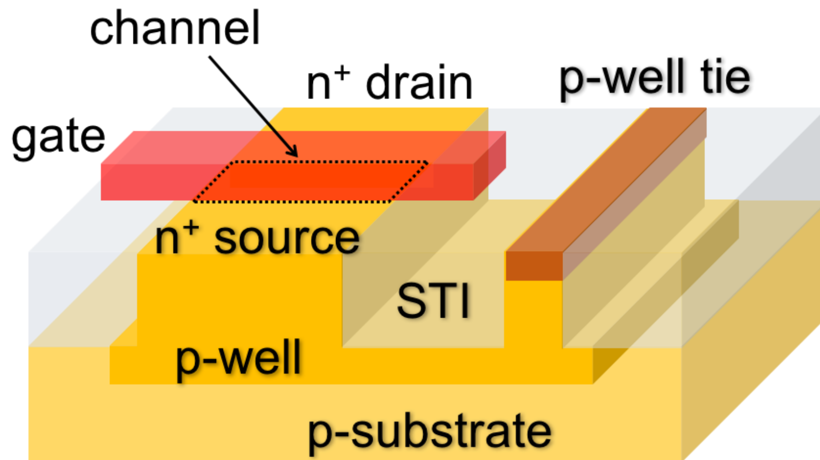
High-k + Metal Gate

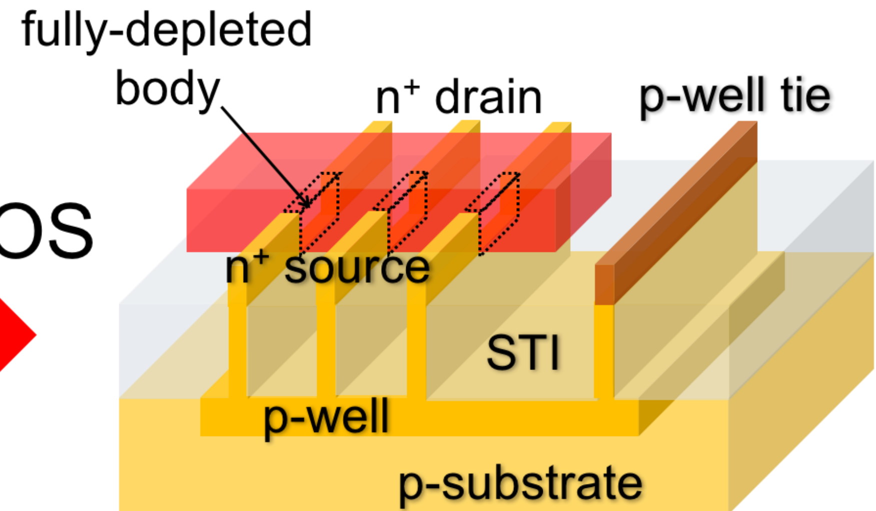
Modern CMOS scaling is as much about material and structure innovation as dimensional scaling



- 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.

Below 28 nm, FinFET

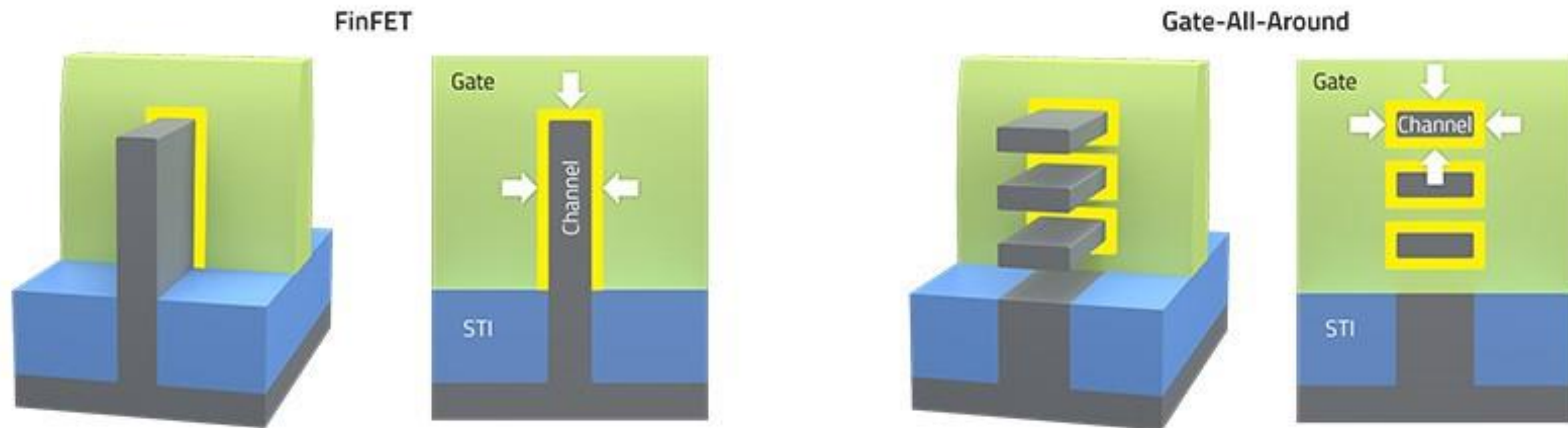




Three-dimensional, Gate-All-Around vertically stacked transistors

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

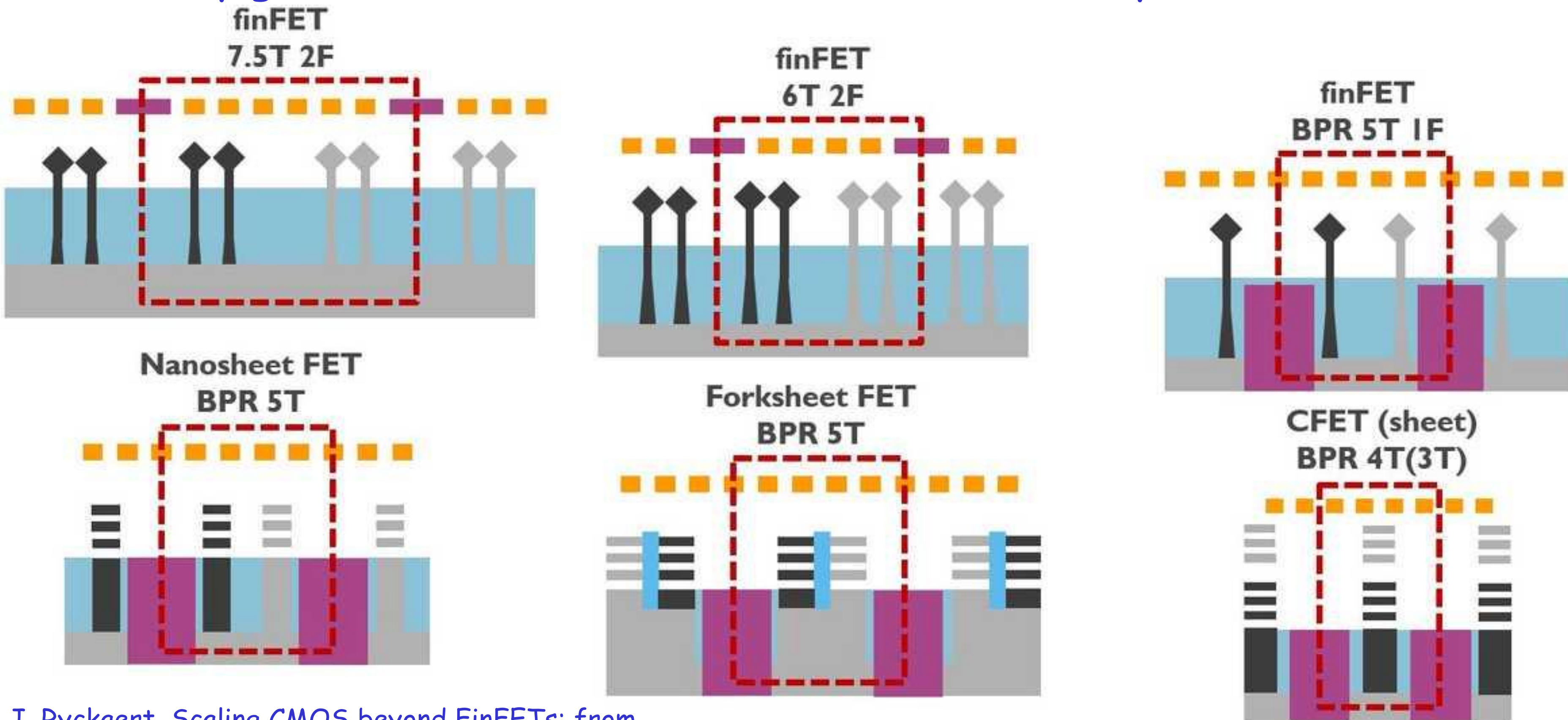
In GAA transistors, the channel is divided into separate horizontal sheets. As the gate now fully wraps around the channels, superior channel control is obtained compared to FinFET
The sheet-to-sheet spacing, analogous to fin pitch, is determined not by lithography but tightly controlled epitaxial processes.



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



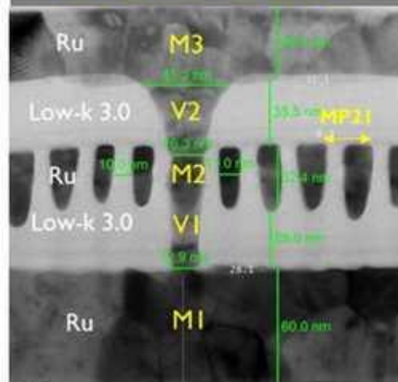
J. Ryckaert, Scaling CMOS beyond FinFETs: from nanosheets and forksheets to CFET, IMEC, 2020

The next 20 years (?)

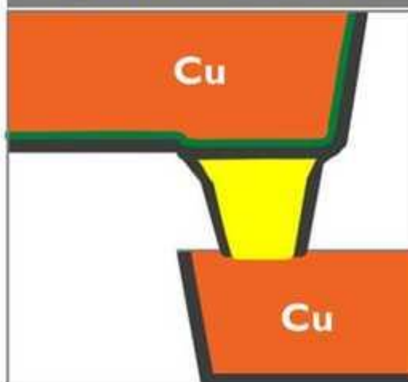
State of the art	Next innovation options		Further out options	
FinFET	Nanosheet	Forksheet	CFET	Sequential 3D

NMOS PMOS

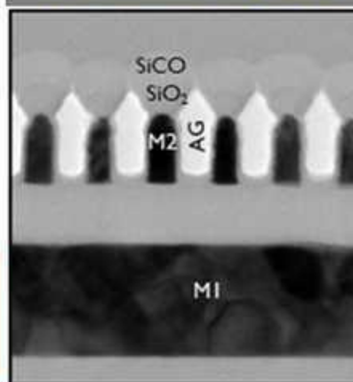
Dual damascene MP21



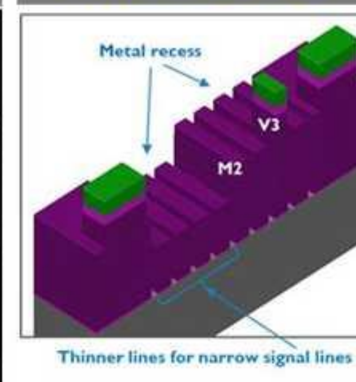
Hybrid via metallization



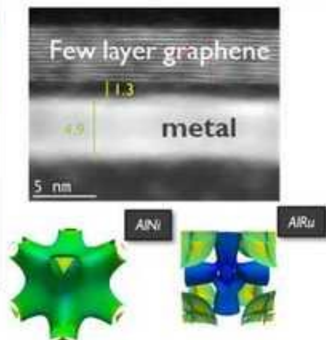
Semi-damascene w AG



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 our community

- For various reasons (cost, accessibility, know-how,…) 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 cell? 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

DSSC

236µm

204µm

SRAM

CSA FE

DEPFET FE

Slow Control

ADC

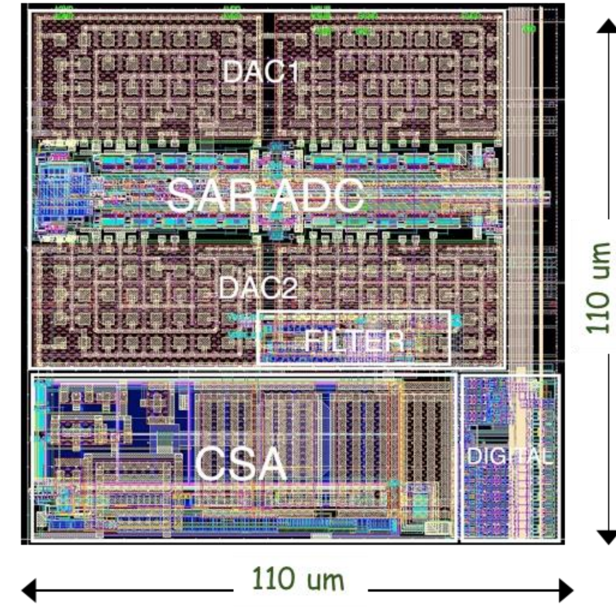
Filter

Injection

References

CSA + filter: 6000 μm^2

PixFEL

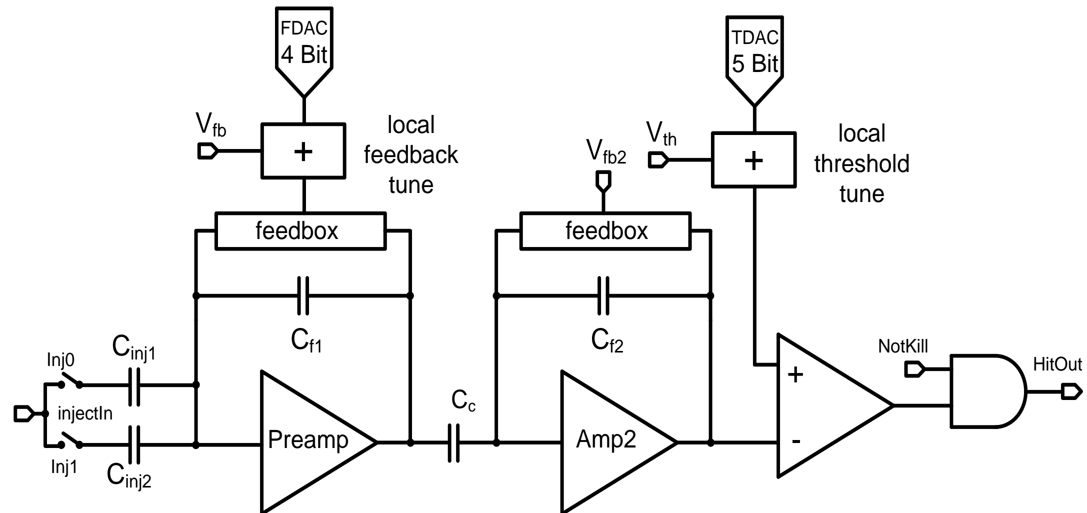


Analog section: 5000 μm^2

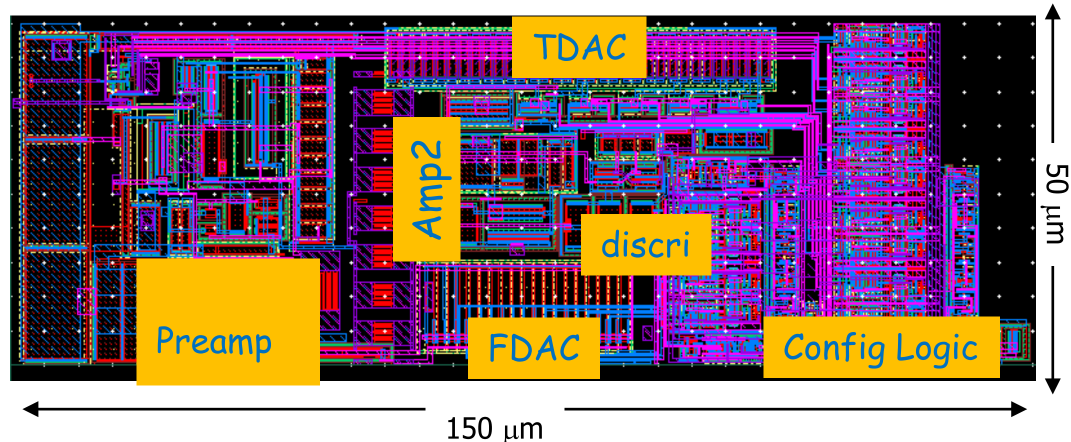
CMOS scaling applied to pixel readout cells: from 130 nm to 65 nm for HEP

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

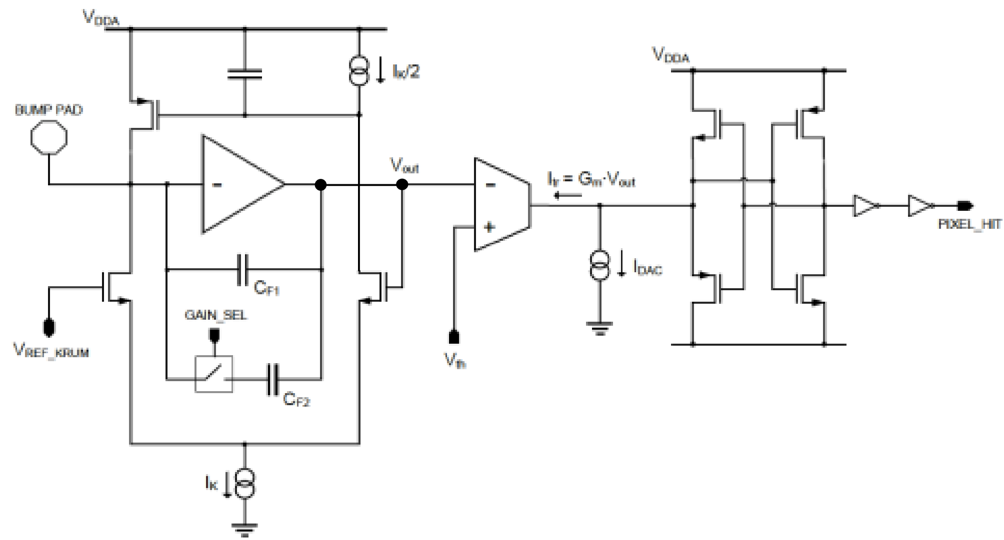
130 nm CMOS FE-I4 readout chip for pixel sensors in ATLAS IBL: the analog pixel cell



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

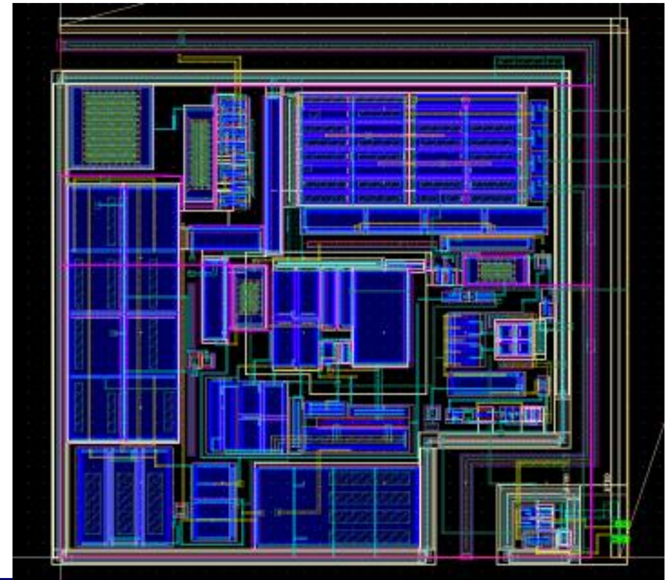


65 nm CMOS readout chip for pixel sensors in the phase II upgrade of the CMS inner tracker at HL-LHC: the Linear Front-End



In a 50 μm x 50 μm pixel cell, the area allocated to the analog front-end is about 1000 μm^2.

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



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

SOC Radiation Tolerant Ecosystem

Proposed Scheme

Draft 1.0

Analog functions

ADC

DAC

Slow Control Interfaces

JTAG

GPIO

SPI

I²C

Interconnect Bus

FE channel

FE analog blocks

eFPGA fabric

micro-code SRAM

RISC-V

SRAM

data

e-Link

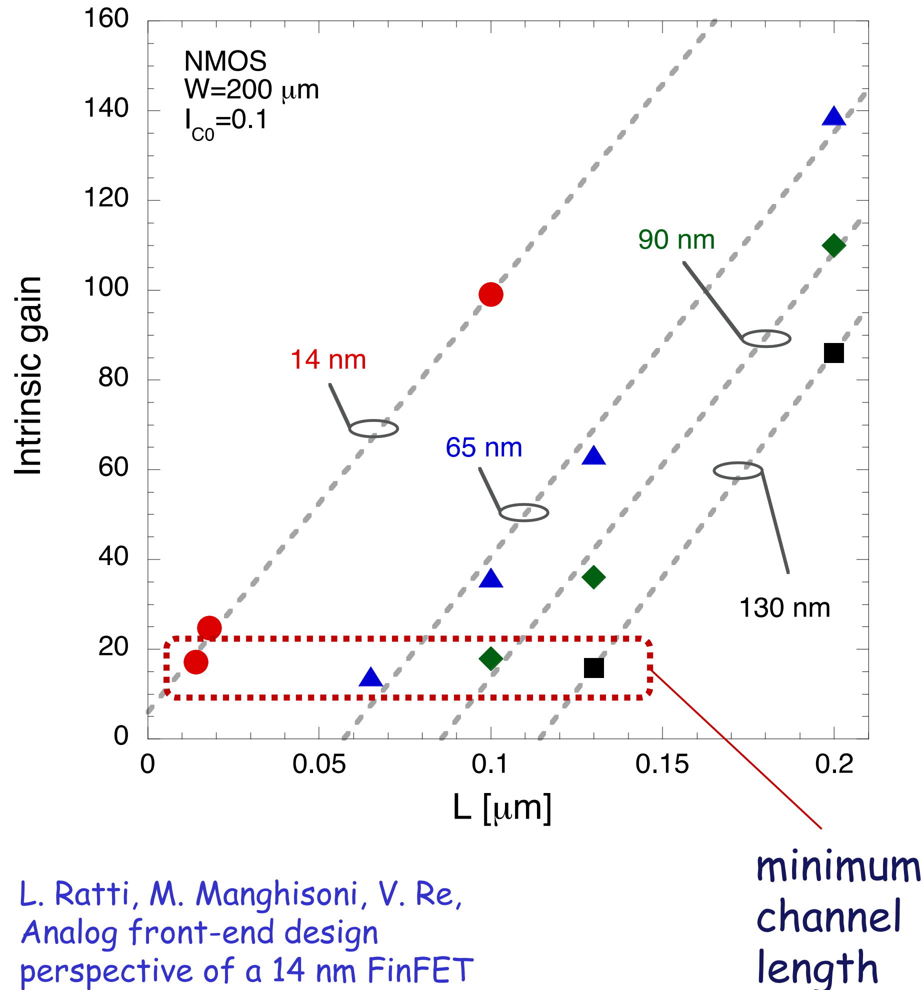
any-Link

High Speed Interfaces

Programmable Logic

Critical parameters for analog front-end design in nanoscale CMOS

Intrinsic gain in weak inversion



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

A complete picture of analog parameters is not available yet (at least to me), but a hint of the behavior of planar 28 nm CMOS and of 14 nm FinFETs can be extracted from preliminary experimental results and from the general features of recent CMOS nodes

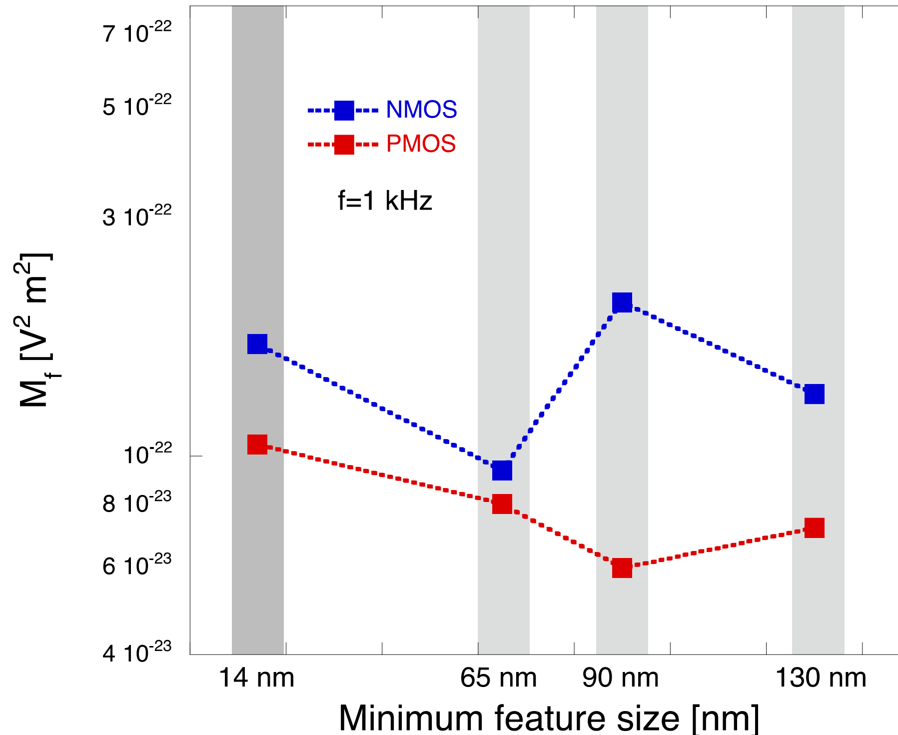
Improved gate control was developed in FinFET to decrease leakage current, reduce short channel effects (which also may lead to higher gain) and process-induced variability

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

Preliminary FinFET studies

Low frequency noise power parameter

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



- 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

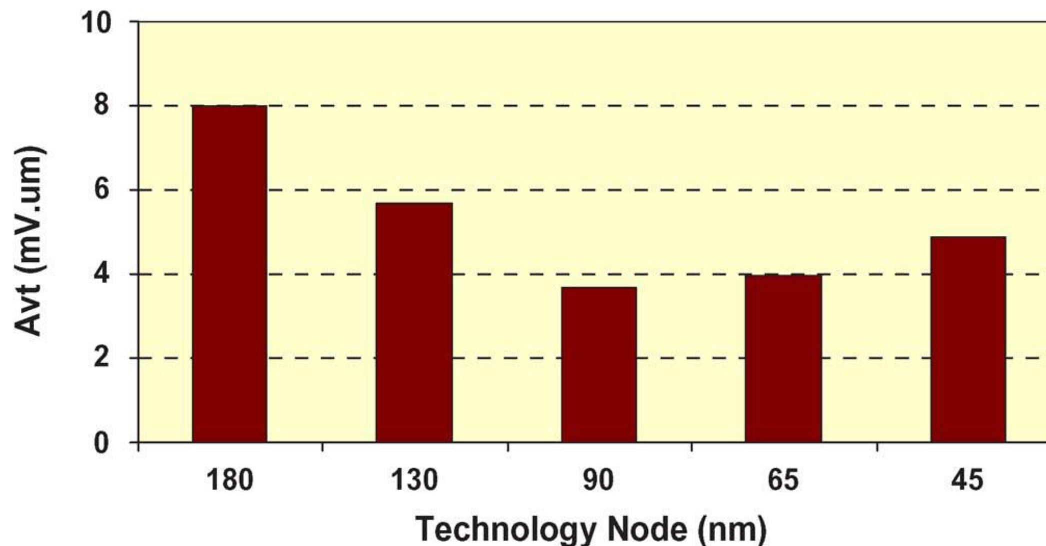
(S. Yang et al, *28nm metal-gate high-K CMOS SoC technology for high-performance mobile applications*, 2011 IEEE Custom Integrated Circuits Conference)

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

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
($W_p \geq 300$nm $L_p \geq 120$nm $L_n \geq 120$nm)

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)

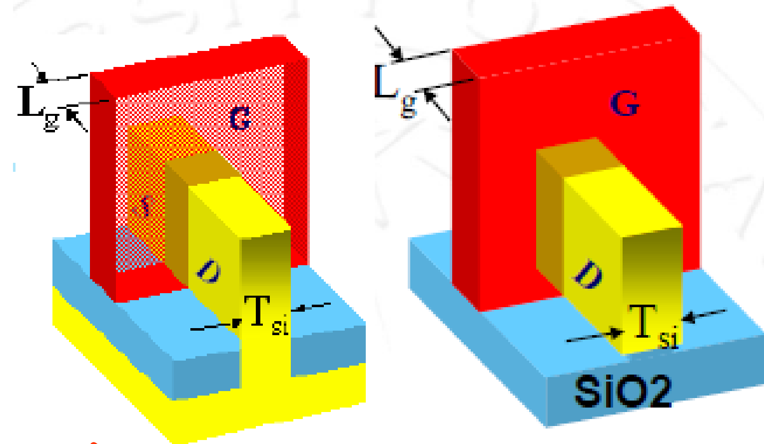
Evolution of scaling and radiation hardness

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)

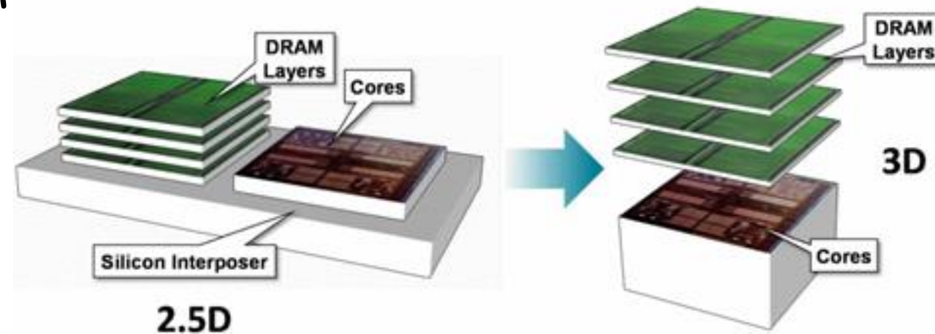
- Carbon nanotubes and graphene have generated much interest: not yet clear if they will be a replacement for Si CMOS. Because of their extremely low volumes (few atomic layers), their radiation response may mostly depend on interfaces and surrounding materials.

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 `



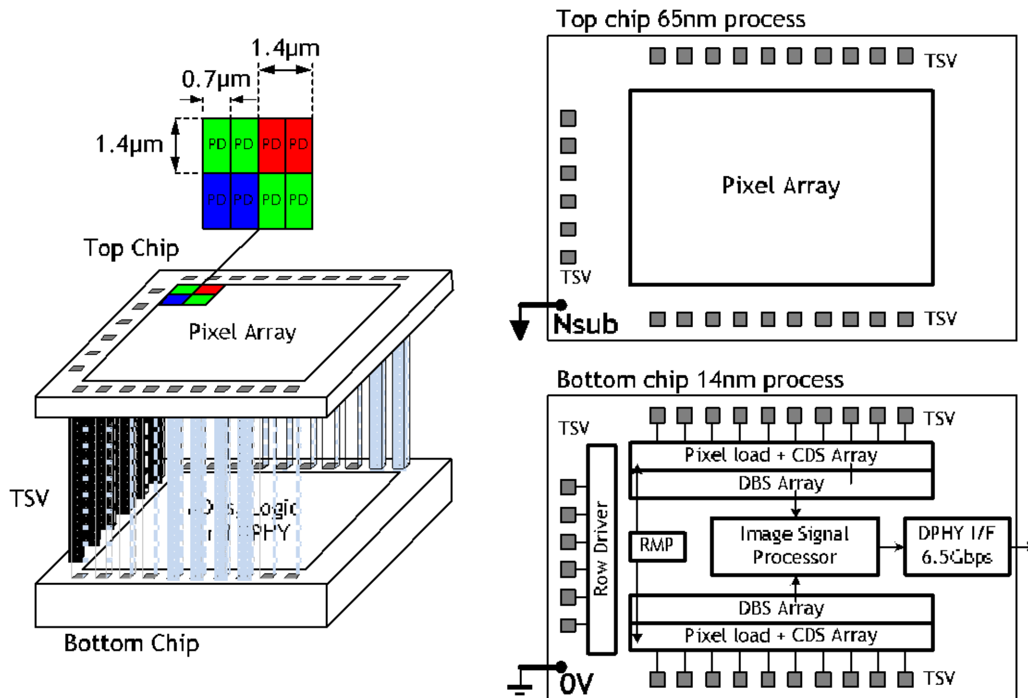
- For our pixel sensors, 3D integration could be leveraged to shrink pixel size and pitch, increase pixel-level electronic functions, reduce dead areas, decrease amount of material by aggressive thinning

3D 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

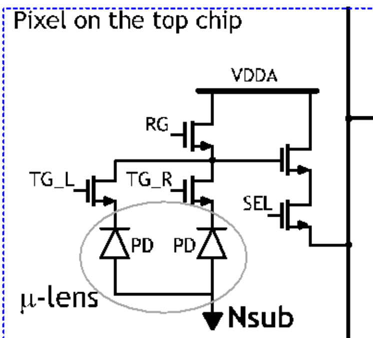
Image sensor 3D-stacked with a FinFET readout

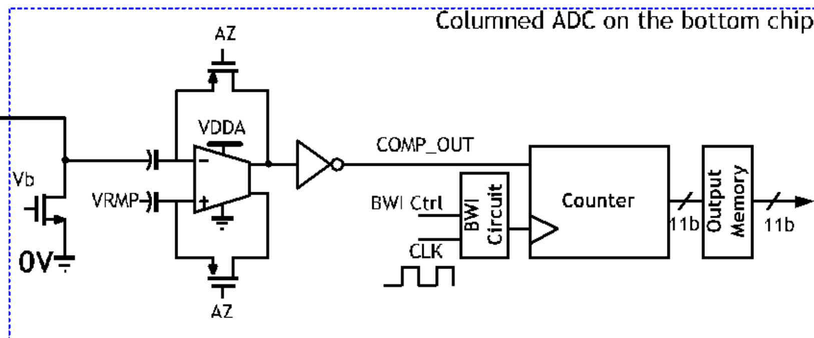
M. Kwon et al., A Low-Power 65/14nm Stacked CMOS Image Sensor", Samsung, ISCAS 2020



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



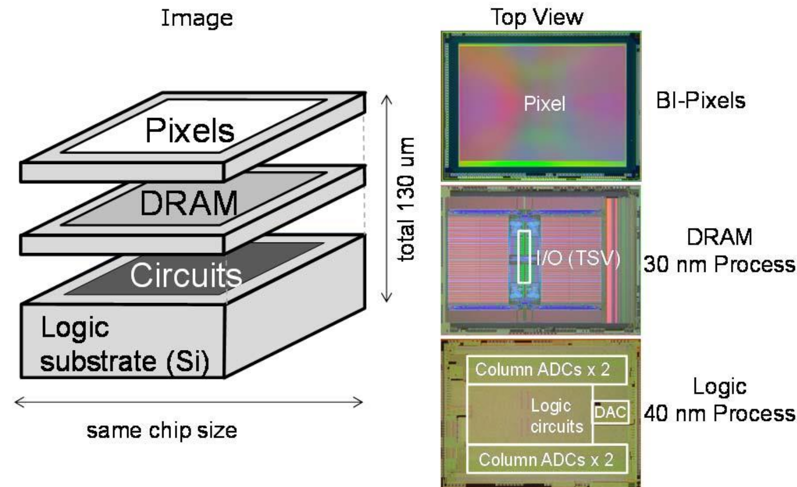


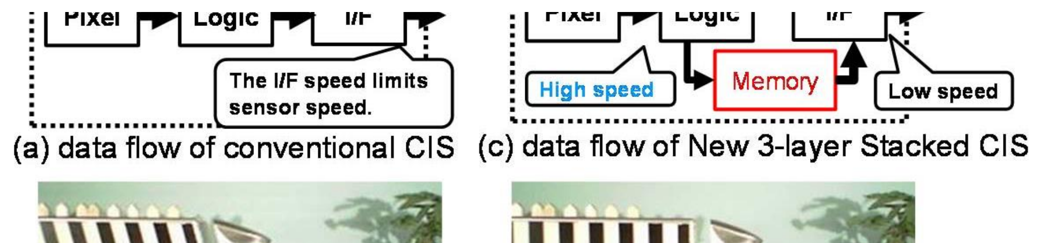
Wafer level stacking
pixel pitch 1.4 µm,
12 Mpixel
11-bit column parallel ADC

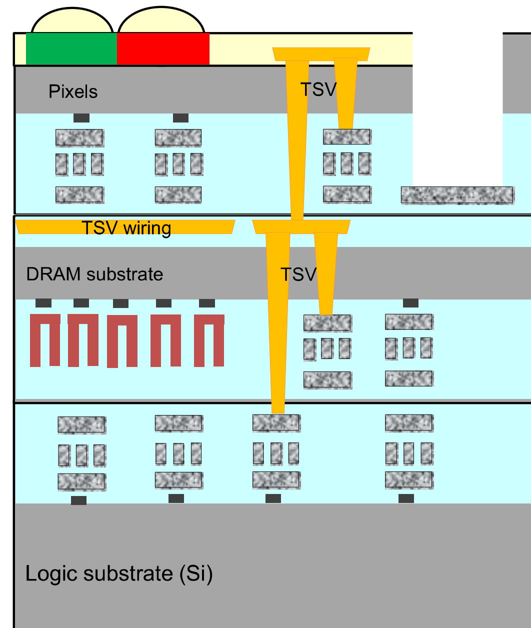
CMOS image sensors with 3 device layers: pixels, DRAM, logic

"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







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

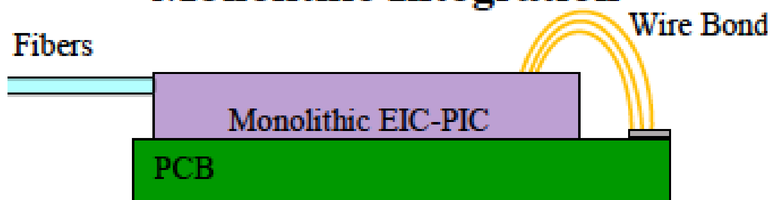
Samsung version with 28 nm logic and 20 nm DRAM

3D 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

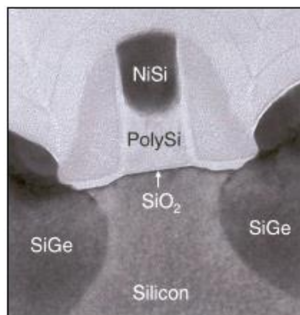
Conclusions

- 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
Vd N+ D
Vb P+ TIE
N LDD
Fz
N LDD
STI
Vsb
Δy
HALO
Id
Vdb
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.*

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)

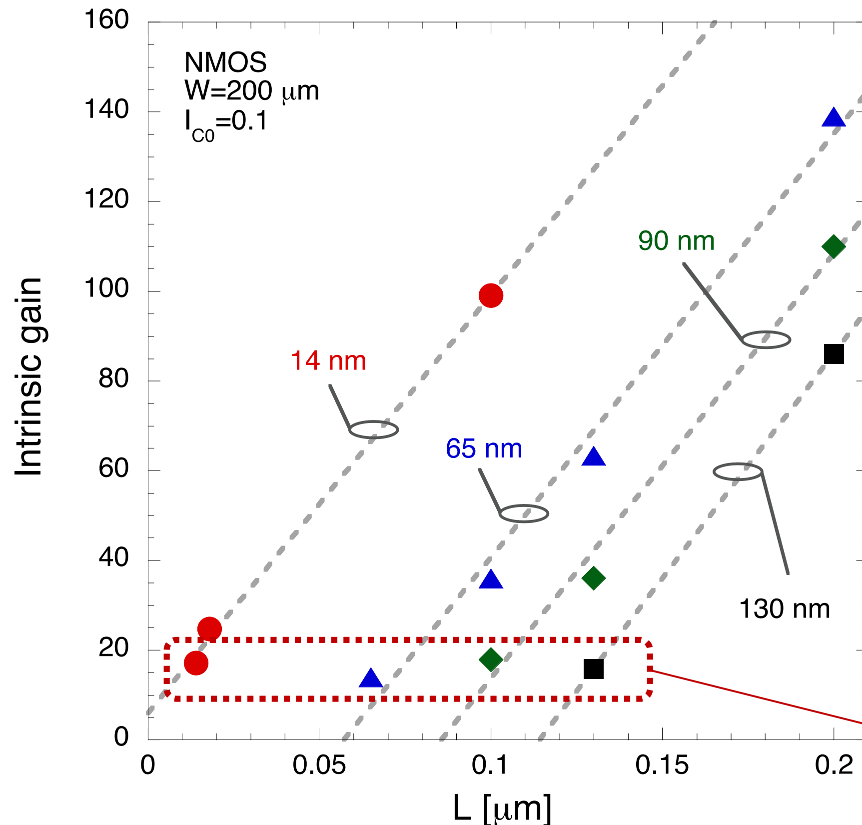
- In these data-driven devices, each pixel continuously monitors for a change of sufficient magnitude from a memorized value. When the change exceeds a threshold, the camera sends an event, transmitted from the chip with x-y location, time-stamp, and polarity of the change

- 3D stacking allows for higher complexity of pixel electronics at small pitch without degrading fill factor

- Architectures moving closer to pixel sensors for particle detection

Preliminary finFET studies

Intrinsic gain in weak inversion



- In all the examined technology nodes, the intrinsic voltage gain at the minimum channel length features very similar values, in agreement with constant field scaling rules

$$A_{vi} = \frac{g_m}{g_{ds}} \propto \alpha L$$

with α the scaling factor

minimum channel length

Critical parameters for analog 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:
  - Gain
  - Thermal noise

  } Charge carriers in the device channel (short channel effects)

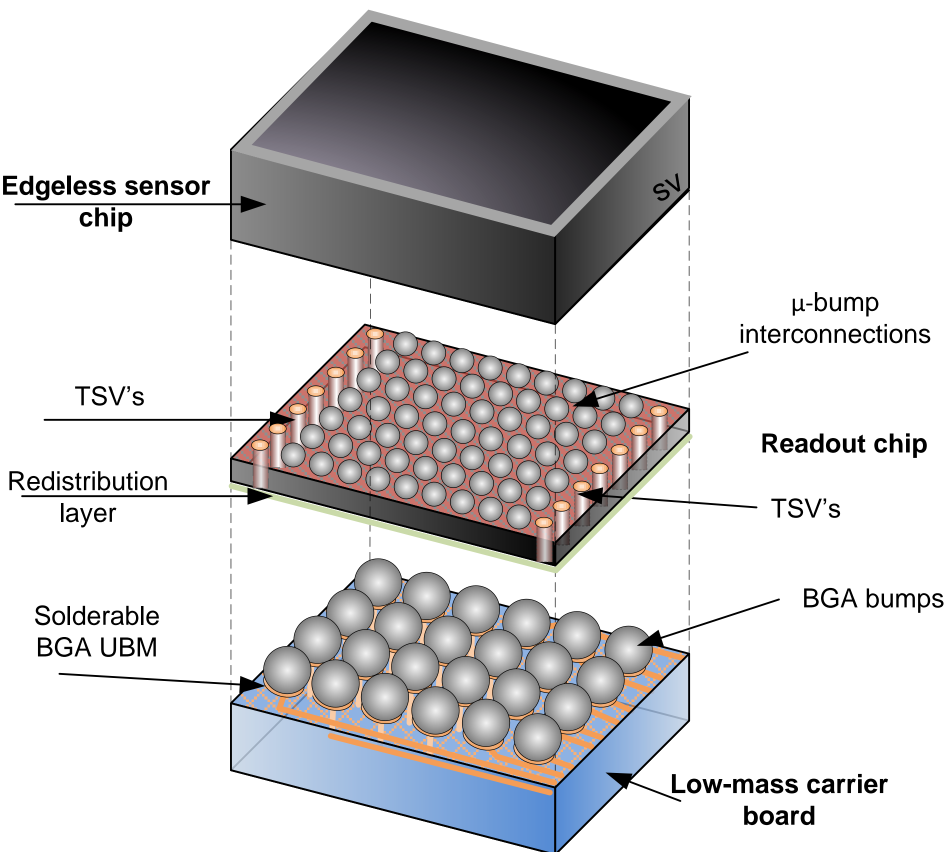
  - 1/f noise
  - Gate leakage current

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

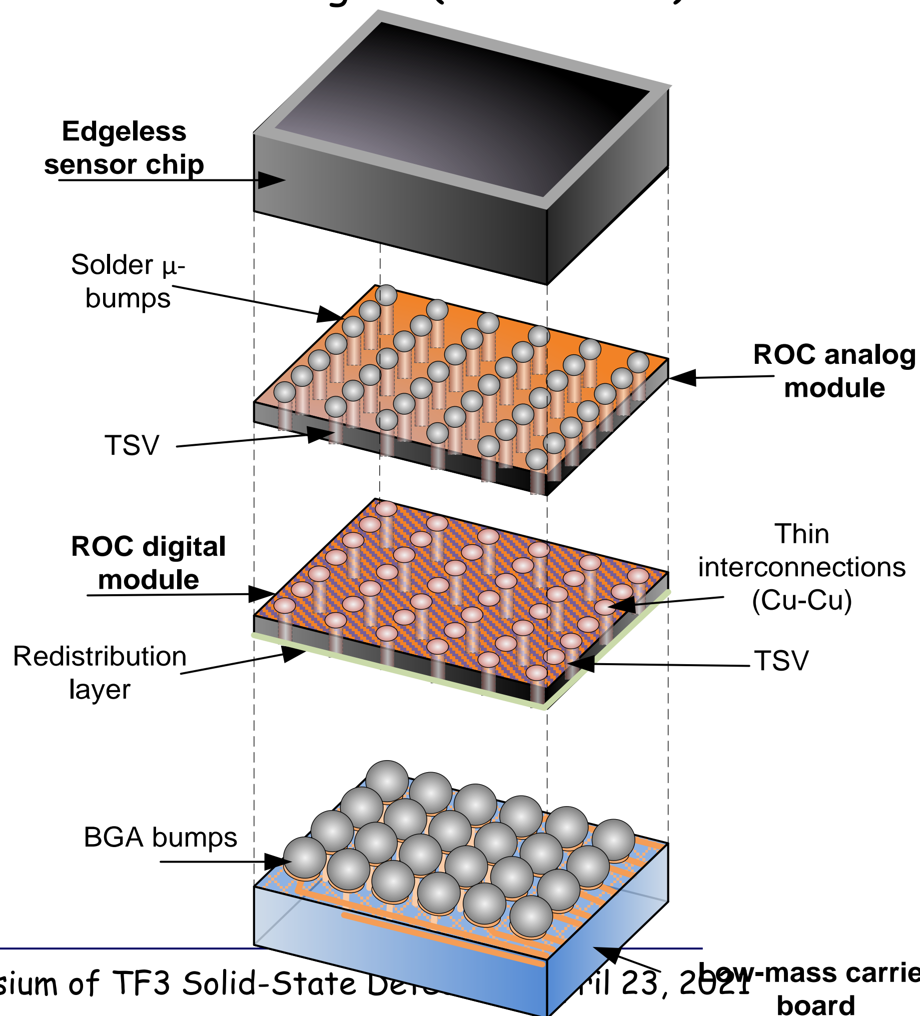
  - Threshold dispersion } Dopant fluctuations in the device channel
  - Radiation hardness } Radiation-induced positive charge in the gate oxide and in lateral isolation oxides

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



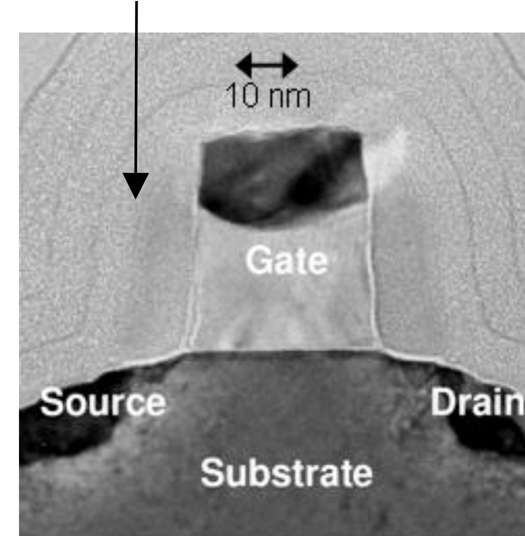
MUX 2010 WORKSHOP,
SAMI VÄHÄNEN – CERN PH-ESE

Radiation effects in planar CMOS

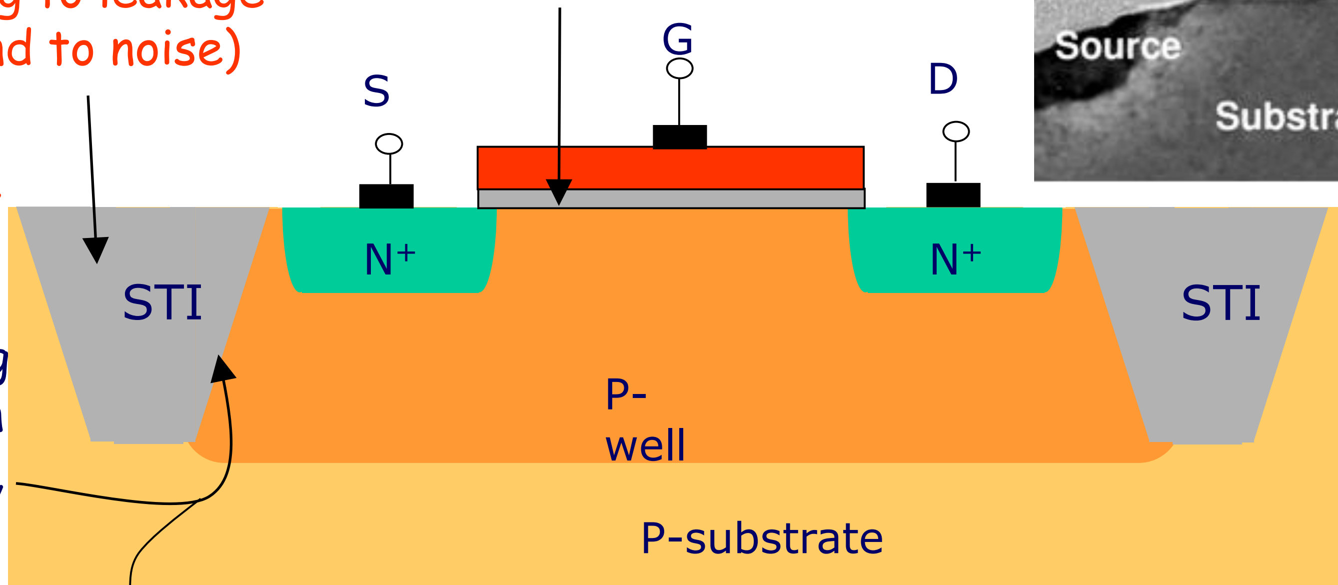
Spacer dielectrics may be radiation-sensitive

Thick Shallow Trench Isolation Oxide (~ 300 nm); radiation-induced charge-buildup may turn on lateral parasitic transistors contributing to leakage current and to noise)

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

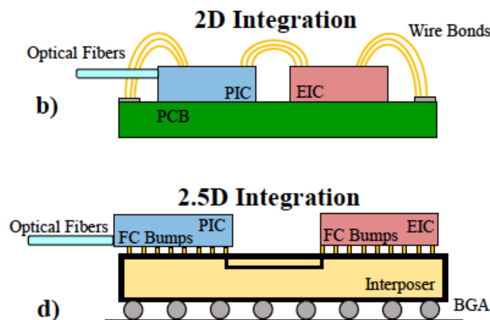


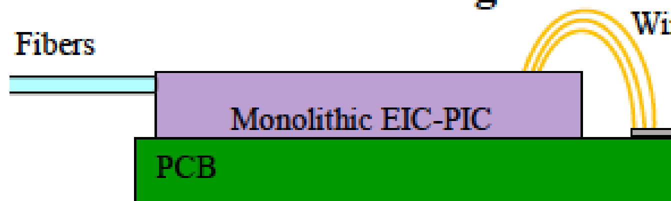
Increasing sidewall doping makes a device less sensitive to radiation (more difficult to form parasitic leakage paths)

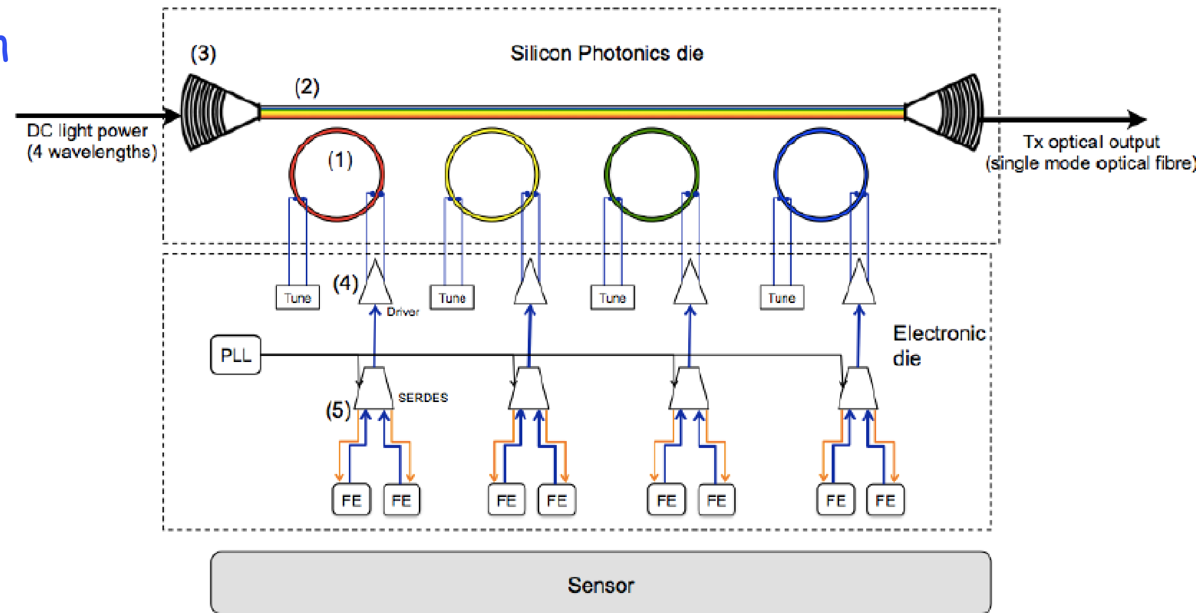
A potential application of 3D integration: the INFN FALAPHEL project

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)