Front-end electronics for radiation sensors in ultra-deep submicron CMOS technologies

Angelo Rivetti
INFN-Sezione di Torino, Italy

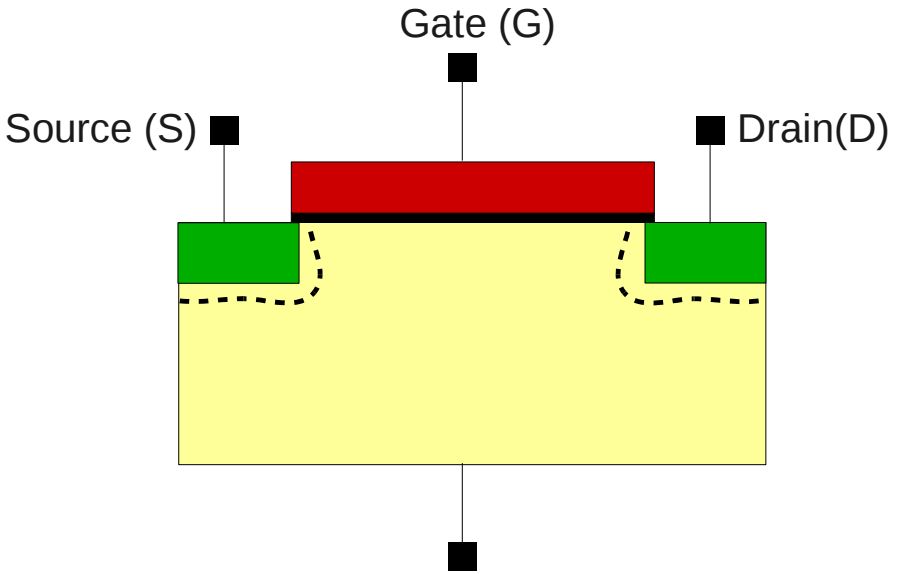
Why CMOS

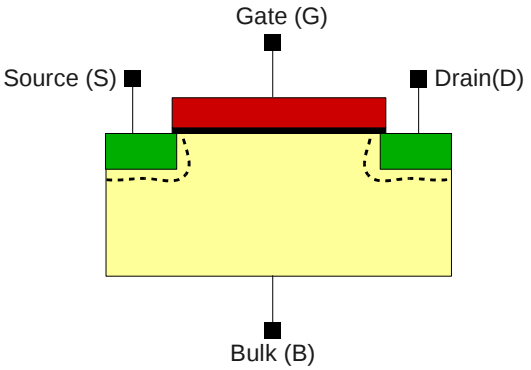
- **CMOS technologies dominate today the market of integrated circuits:**
  - **Ability to integrated on the same substrate PMOS and NMOS devices.**
  - **CMOS transistors are fairly easy to squeeze.**
- **Integration density doubles roughly every 24 months (Moore's Law).**
- **Scaling makes transistors faster: CMOS has progressively replaced bipolar transistor in a growing number of applications, including RF.**

Why CMOS in front-end electronics

- **CMOS was always the preferred choice to implement front-end for radiation detectors:**
  - **Easily available and cheap.**
  - **Allows the fabrication of good sampling circuits.**
- **CMOS became the solution for implementing front-end electronics in radiation-sensitive environments:**
- **Reason: thin gate oxide makes transistors less sensitive to radiation damage.**
- **Deep sub-micron front-end:**
  - **First generation: CMOS 0.25 µm. Now in the LHC detectors. Radiation hard with enclosed layout.**
  - **Second generation: CMOS 130 nm. Radiation hard even with standard layout. Increased functionality, smaller area.**
  - **Next step: 65 nm. Much increased functionality.**

Scaling CMOS transistors





α>1

Device width and length	→	1/α
Depletion region	→	1/α
Oxide thickness	→	1/α
Threshold voltage	→	1/α
Power supply	→	1/α
Cox	→	α
Device density	→	α^2

$$x_d = \sqrt{\frac{2\epsilon_s}{qN_A}(\phi_B + V)}$$

$$C_{OX} = \frac{\epsilon_{OX}}{t_{OX}}$$

Consequences of scaling

- For **digital circuits**: things can only **get better**!
  - Scaling is optimized for digital circuits.
  - Digital gates becomes **smaller**, **faster** and lower power.
  - Note: while **power/gate decreases, power/chip increases** to the increased number of gate and **augmented functionality.**
- For **analog** circuits situation is **more complex**.
- Analog circuits sensitive to a number of effects arising in the deep sub-micron regime.
- For analog circuits, key parameters are:
  - Transconductance g_m: $\Delta I_{DS} = \frac{\partial I_{DS}}{\partial V_{GS}} \Delta V_{GS} = g_m \Delta V_{GS}$
  - Output conductance g_{ds}: $g_{ds} = \frac{\partial I_{DS}}{\partial V_{DS}} \approx \lambda I_{DS}$
  - Noise and matching...
- The g_m/g_{ds} ratio is a metric of a gain that can be achieved by CMOS amplifiers.

Transconductance

Strong inversion

$$I_{DS} = \frac{1}{2}\mu_n C_{OX}\frac{W}{L}(V_{GS} - V_{TH})^2(1 + \lambda V_{DS}) \qquad g_m = \sqrt{2\mu C_{ox}\frac{W}{L}I_{DS}}$$

Weak inversion

$$I_{DS} = I_0\frac{W}{L}e^{\frac{V_{GS}}{nU_T}}(1 - e^{-\frac{V_{DS}}{nU_T}}) \qquad U_T = \frac{kT}{q} \qquad g_m = \frac{\partial I_{DS}}{\partial VGS} = \frac{I_{DS}}{nU_T}$$

$$I_C = \frac{I_{DS}}{2n\mu C_{ox}\frac{W}{L}U_T^2}$$

I_C < 0.1: W.I.

In between: moderate inversion!

I_C> 10: S.I.

Gm in weak inversion

$$I_C = \frac{I_{DS}}{2n\mu C_{ox}\frac{W}{L}U_T^2}$$

- Scaling the technology, C_{ox} increases.
- For the same bias current and aspect ratio, I_C decreases
- Transistors work more and more in weak inversion.
- g_m/I_d is maximized
- MOS resembles a bipolar transistor

Continuity across level of inversion

- **The EKV model was developed in an attempt to have a physical based model providing continuity of MOS characteristics across different level of inversion:**

MOS equation in saturation valid in all region of inversion

$$I_{DS} = 2n\mu C_{ox}\phi_T^2 \frac{W}{L}\left[\ln\left(1+e^{\frac{V_{GS}-V_{TH}}{2n\phi_T}}\right)\right]^2$$

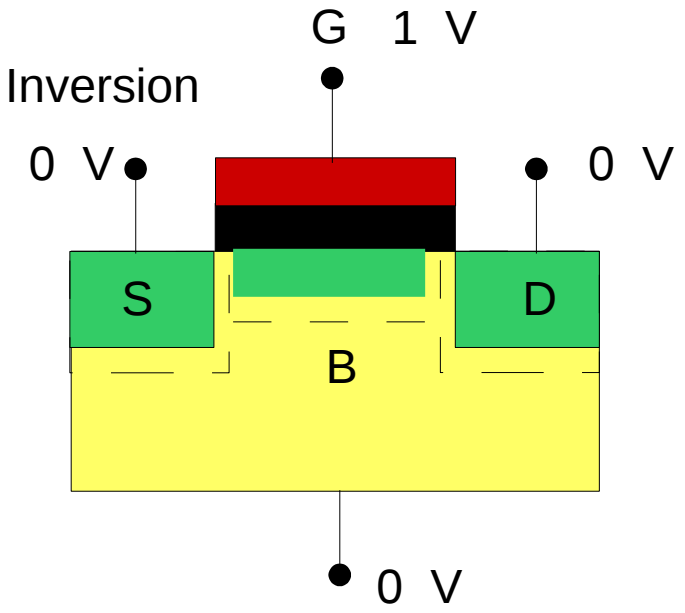
Trasconductance

$$g_m = \frac{I_D}{n\phi_T}\frac{1}{\sqrt{I_C+0.5\sqrt{I_C}+1}}$$

Key short channel effects

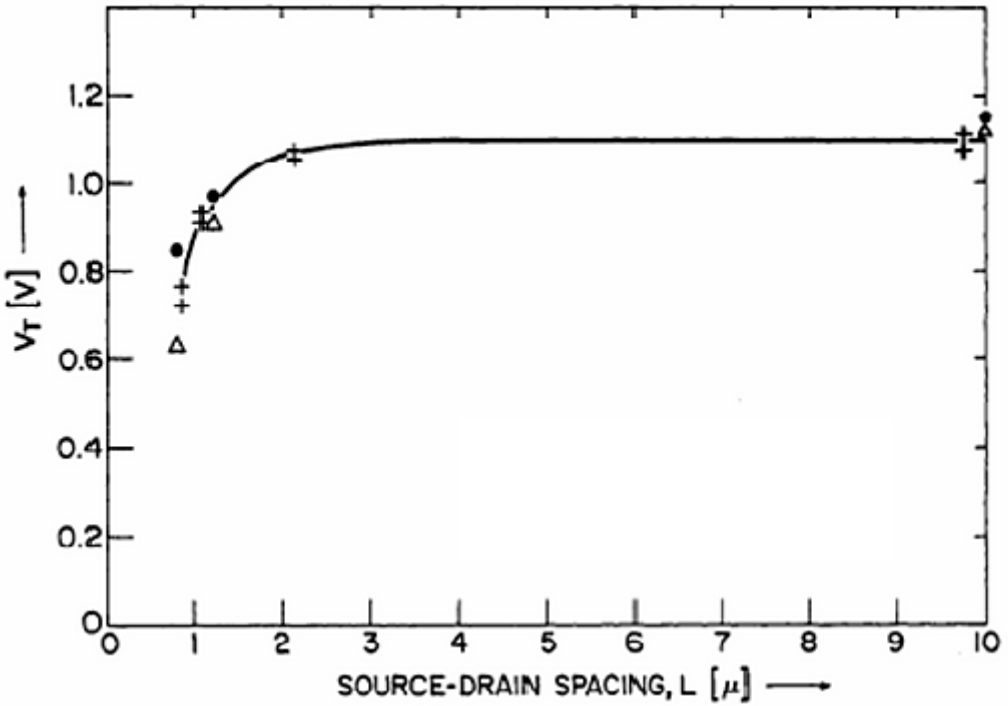
1. **Drain-induced barrier lowering: reduces g_{ds}**
2. **Surface scattering: reduces mobility/g_m**
3. **Velocity saturation: reduces g_m**
4. **Impact ionization:substrate current, reduces g_{ds}**
5. **Hot electrons: long term reliability**

Example: reverse short channel (1)

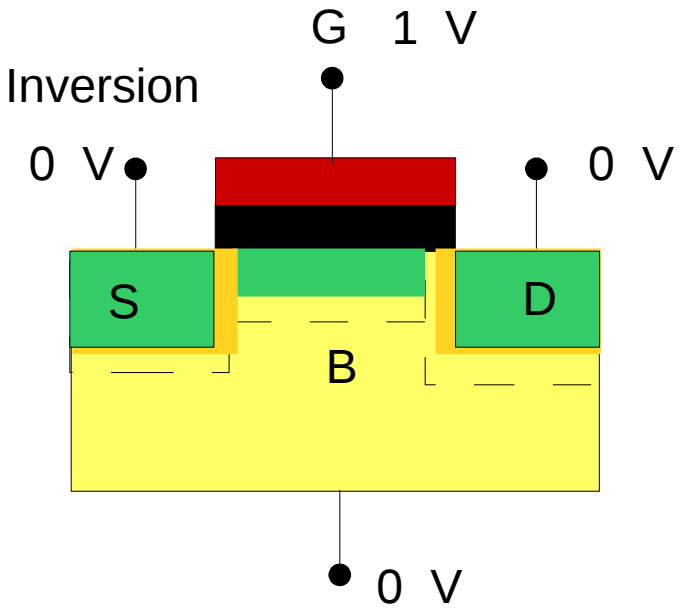


Example of traditional short channel effect: threshold is lower as channel length is decreased.

- The gate voltage needs to maintain also the depletion region.
- For short channel device source and drain depletion regions protrude significantly in the channel.
- For the same charge store on the gate, more carriers can be attracted in the channel, since the depletion region is partially supported by source/drain

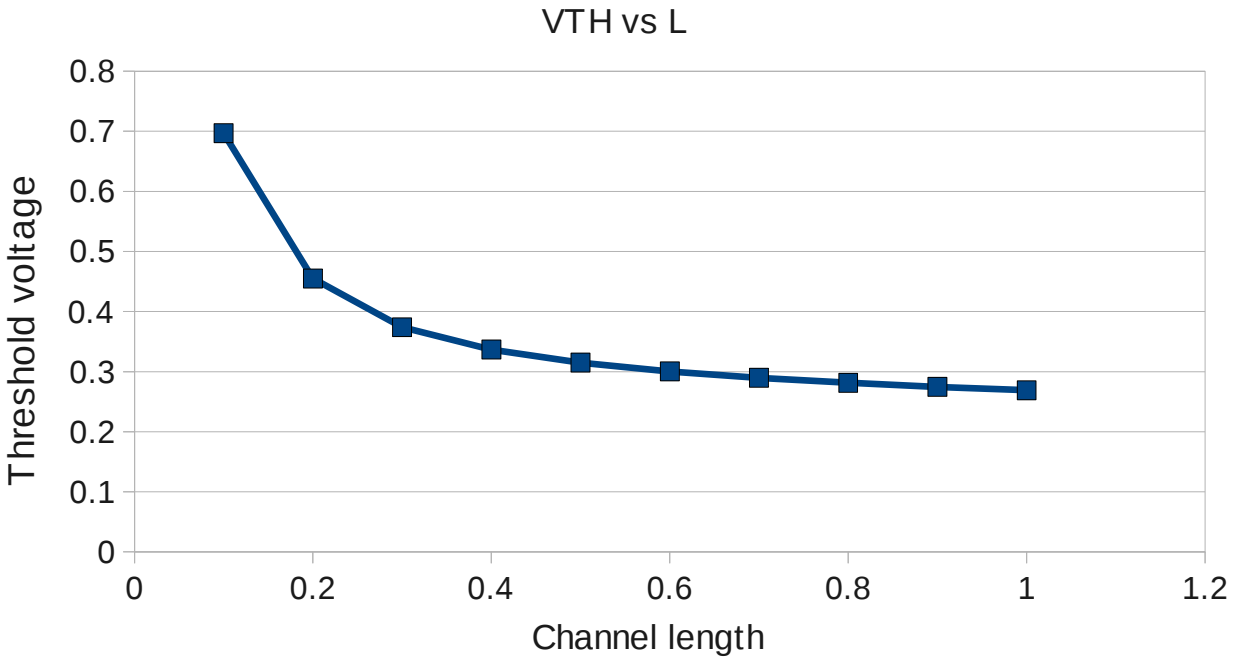


Example: reverse short channel (2)

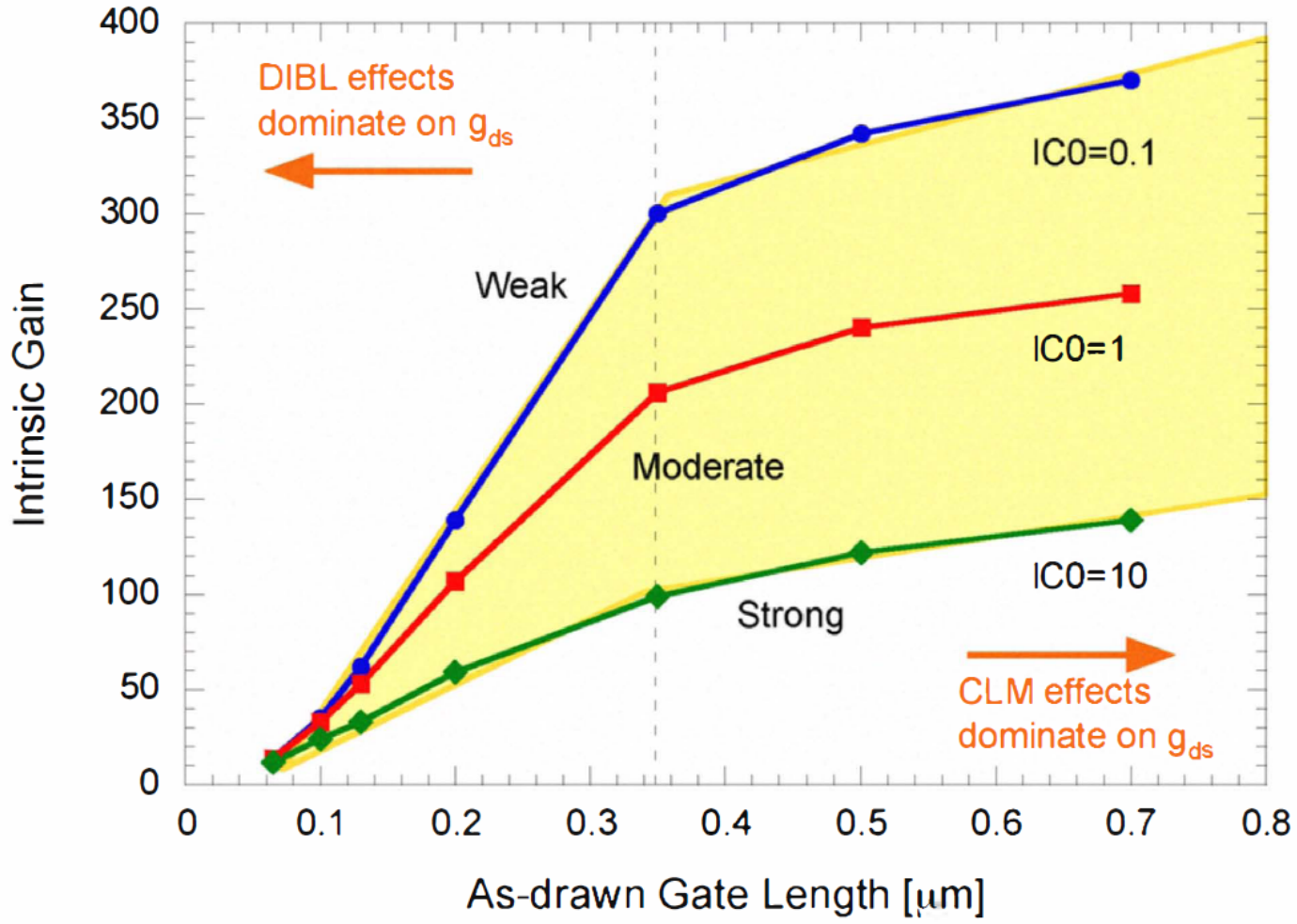


Reverse short channel effect: simulated threshold variation as a function of channel length in a 90 nm CMOS process.

- To prevent excessive extension of the source-drain depletion region into the regions around the electrodes receive a stronger substrate doping (halo doping).
- When the channel is very short the two regions tend to overlap, the local substrate doping in the channel region is increased and the threshold voltage increases.
- Another reason to keep away from minimum length devices in analog design!



Gain vs IC and channel length



M. Manghisoni et al, “Analog Design Criteria for High-granularity Detector Readout in the 65 nm CMOS Technology”, IEEE NSS-MIC Conference Records, N40-2, 2011.

Noise

- **Key noise sources in MOS transistors:**
  - **Channel thermal noise**
  - **Channel flicker (1/f) noise**
- **No big surprises concerning thermal noise.**
- **Flicker noise:**

$$e_{n1/f}^2 = \frac{K_f(I_C, L)}{C_{ox} W L f^{\alpha_f}}$$

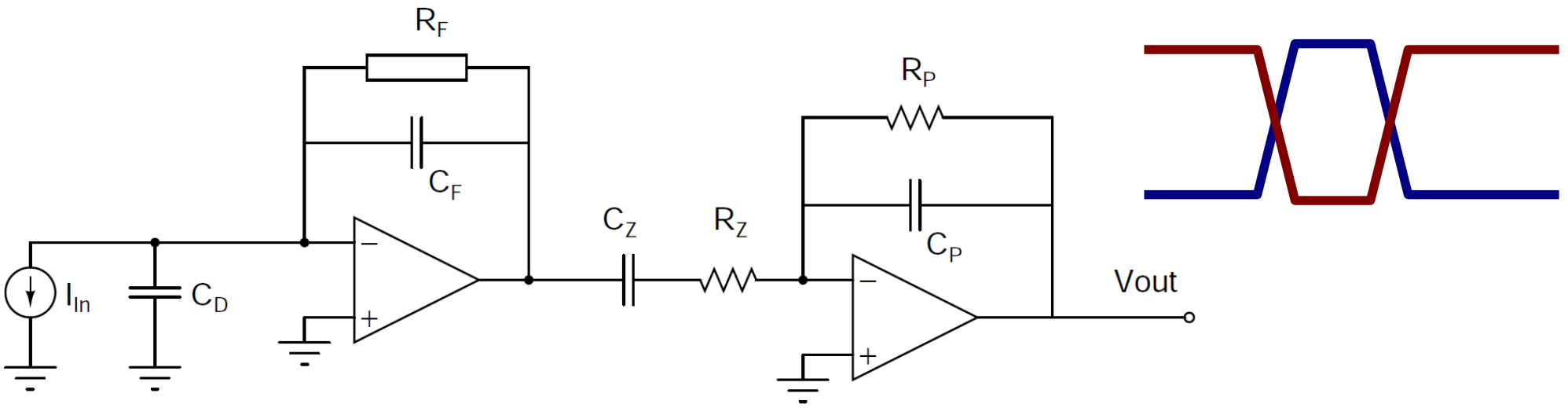
- **Kf may depend on biasing condition and channel length**
- **Deviation from pure 1/f behavior.**

Noise performance compatible with low noise design

Consequences on front-end design

- **Three fundamental blocks in front-end electronics:**
  - **Front-end amplifiers (preamplifier/shapers).**
  - **Analog to Digital Converters.**
  - **Time to Digital Converters.**

Front-end amplifiers



- **In principle, suffer from the reduced voltage swing (power supply at 1 V), but....**
  - **Use of inverting configurations mitigate the problem**
  - **Use of fully differential signal can increase the swing to 2 V p-p**
  - **Weak inversion reduces the headroom required by transistors**
  - **For very high dynamic range, thicker oxide transistor can be used.**
- **Front-end design more challenging, but not that grim!**

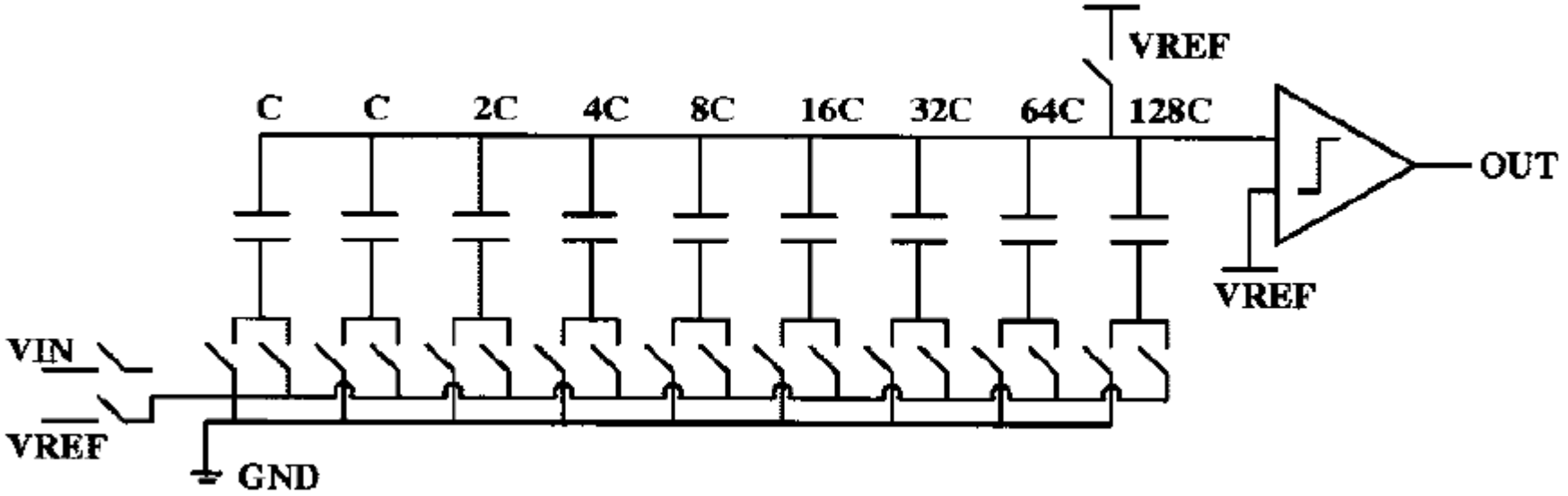
Low power ADC: some examples

Ref	Technology	Architecture	N of bit	Sampling rate (MS/s)	ENOB	Power (mW)	FOM (Fj/step)
1	90 nm	SAR	9	40	8.23	0.82	68
2	130 nm	SAR	10	50	9.11	0.82	30
3	65 nm	SAR	10	100	9.01	1.13	22
4	90 nm	FLASH+SAR	9	100-200	8.44-8.31	0.75/1.33	34.7
5	90 nm	SAR	10	50	9.5	0.32	9

$$FoM = P/(2^{ENOB} F_s)$$

- ADC with 50 MS/s, 9-10 bit resolution and < 1mW of power are now common.
- The renaissance of the SAR ADC.

Traditional SAR ADC (1)



- In the sampling phase the top plate is connected to VREF and the bottom plate to VIN.
- After sampling the top plate is left floating and all the bottom plate are switched to GND. Voltage of the top plate is VREF-VIN.
- The capacitance of the MSB (128C) is switched to VREF. The voltage of the top plate is VREF-VIN+VREF/2. If the voltage is > VREF, the MSB is set to zero and the bottom plate of the MSB capacitors is switched back to GND.
- If VREF-VIN+VREF/2 is smaller than VREF, the bottom plate of the MSB is kept at VREF and the corresponding bit is set to 1.
- The procedure continues with the other bits.

Traditional SAR ADC (2)

- In conventional SAR ADC:
- Capacitors can be switched back and forth between VREF and GND: unnecessary power consumption in the DAC.
- DAC capacitors are usually over-sized with respect to the minimum required by noise boundaries.
- Logic is synchronous with a master clock: 40 MHz clock leads to less that 4 MS/sec.

Performance improvements stem from:

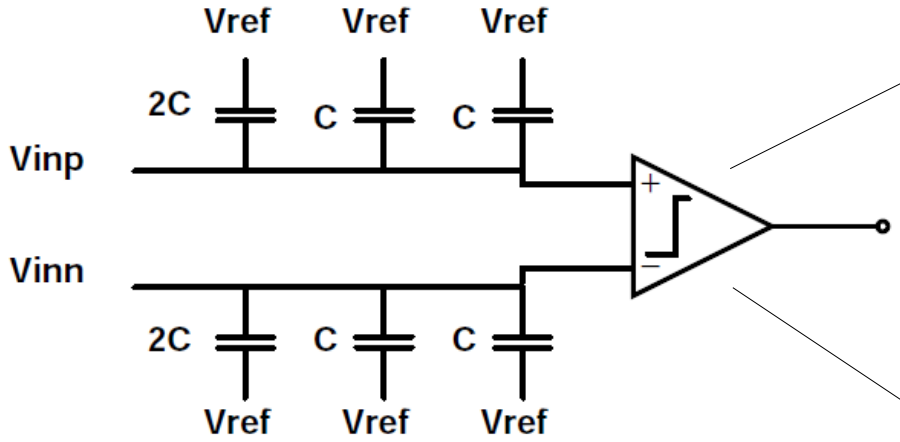
- Reduction of capacitor size
- Modified switching schemes
- Use of asynchronous logic

Modified switching algorithms

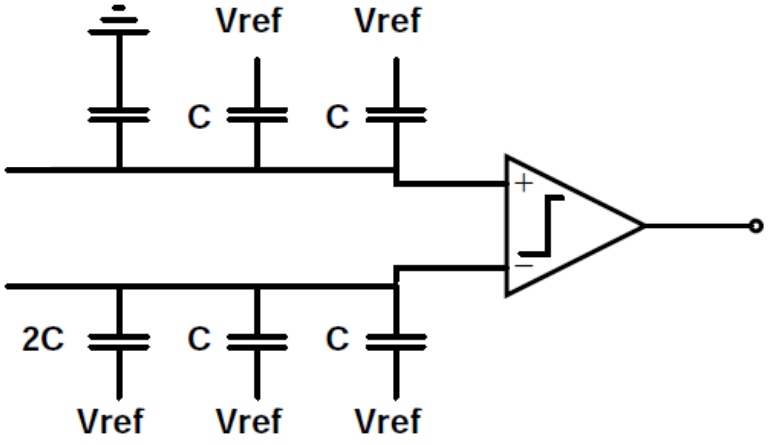
A 10-bit 50-MS/s SAR ADC With a Monotonic Capacitor Switching Procedure

Chun-Cheng Liu, *Student Member, IEEE*, Soon-Jyh Chang, *Member, IEEE*, Guan-Ying Huang, *Student Member, IEEE*, and Ying-Zu Lin, *Student Member, IEEE*

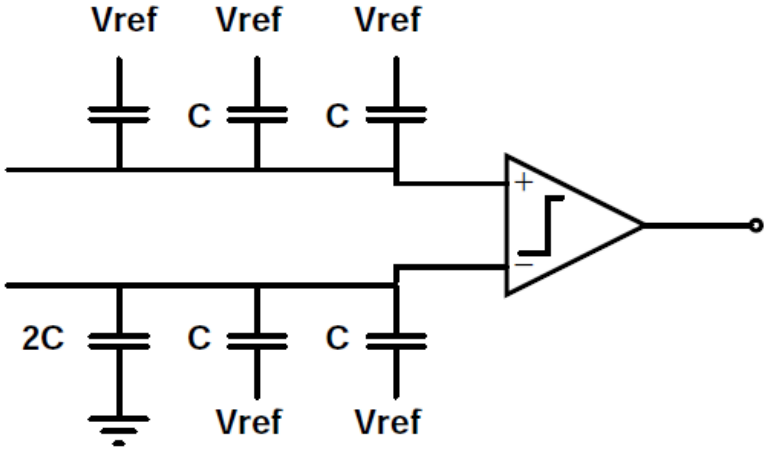
Sampling & first comparison

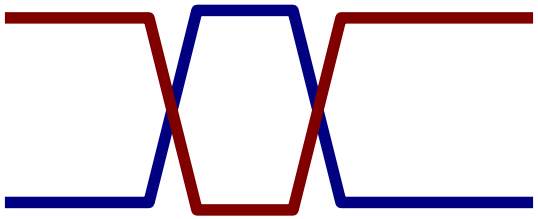


If Vinp > Vinn



If Vinp < Vinn





- **Modern ADC are always fully differential.**
- **A front-end with fully differential outputs is necessary.**
- **Alternative switching schemes can save up to 98% of the DAC power**

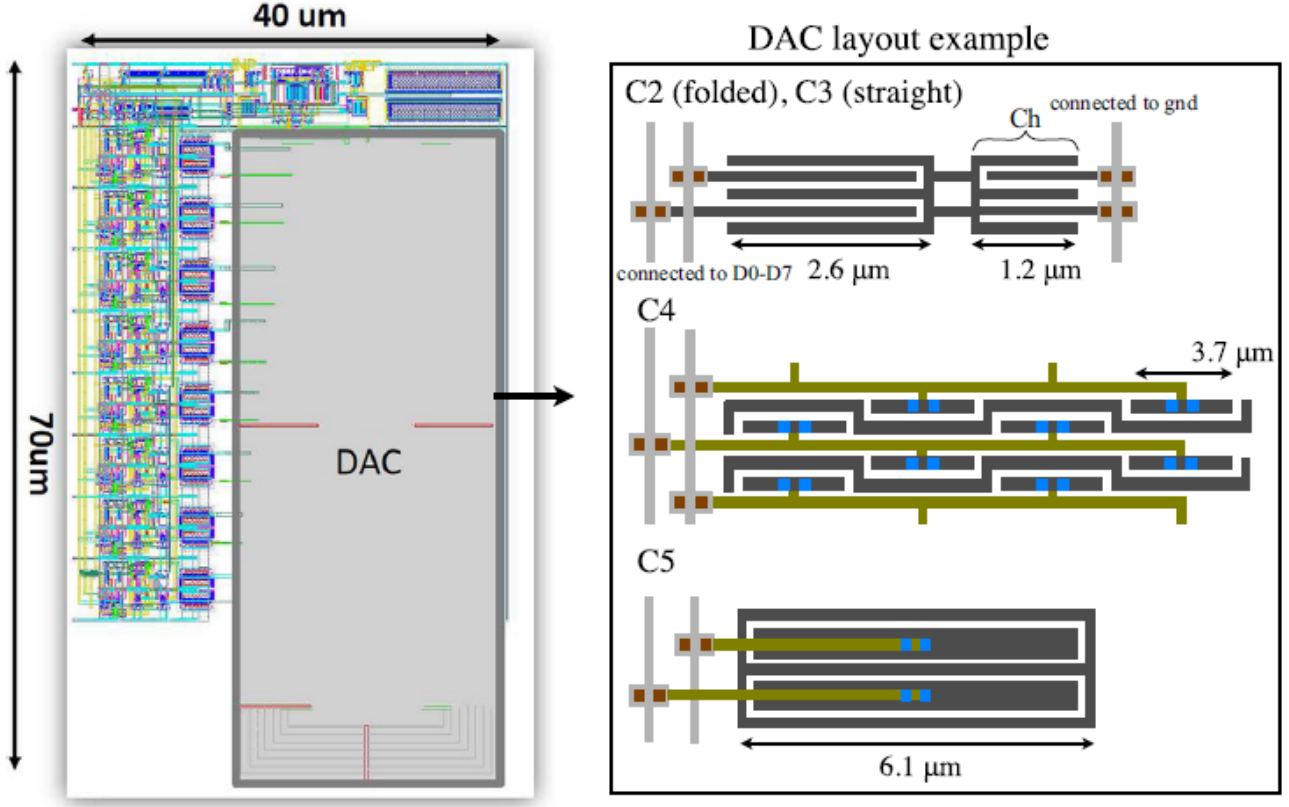
Nuclear Instruments and Methods in Physics Research A ▮ (▮▮▮▮) ▮▮▮–▮▮▮

A 10 MS/s 8-bit charge-redistribution ADC for hybrid pixel applications in 65 m CMOS ☆

Tetsuichi Kishishita*, Tomasz Hemperek, Hans Krüger, Manuel Koch, Leonard Germic, Norbert Wermes

University of Bonn, Physikalisches Institut, Nussallee 12, 53115 Bonn, Germany

70 um x 40 um, 8 bit, 40 uW@10MS/s in 65 nm



More on fast ADCs

26.4 A 3.1mW 8b 1.2GS/s Single-Channel Asynchronous SAR ADC with Alternate Comparators for Enhanced Speed in 32nm Digital SOI CMOS

Lukas Kull[1,2], Thomas Toifl[1], Martin Schmatz[1], Pier Andrea Francese[1], Christian Menolfi[1], Matthias Braendli[1], Marcel Kossel[1], Thomas Morf[1], Toke Meyer Andersen[1], Yusuf Leblebici[2]

[1]IBM Research, Rüschlikon, Switzerland, [2]EPFL, Lausanne, Switzerland

Specifications	[1]	[2]	[3]	[4]	[5]	This work		
Architecture	SAR	Ti-SAR	Ti-SAR	SAR	SAR	SAR		
CMOS Technology (nm)	65	65	65	28	40	32		
Resolution (bits)	8	6	8	8	6	8		
Supply Voltage (V)	1.2	1.2	1.0	1.0	1.0	1.0	1.1	0.9
SNDR near Nyquist (dB)	44.5	31.5	42.75	43.3	30.5	39.3	39.3	38.8
Sampling Speed (GHz)	0.4	1	1	0.75	1.25	1.2	1.3	1.0
Speed per Channel (GHz)	0.4	0.5	0.5	0.75	1.25	1.2	1.3	1.0
Power (mW)	4.0	6.7	3.8	4.5	6.08	3.1	4.2	2.0
FOM (fJ/conf step)	73	210	24	41	178	34	43	28
Area (mm^2)	0.024	0.11	0.013	0.004	0.013	0.0015		
Area for 64GS/s (mm^2)	3.8	7.0	8.3	0.26	0.67	0.080	0.074	0.096

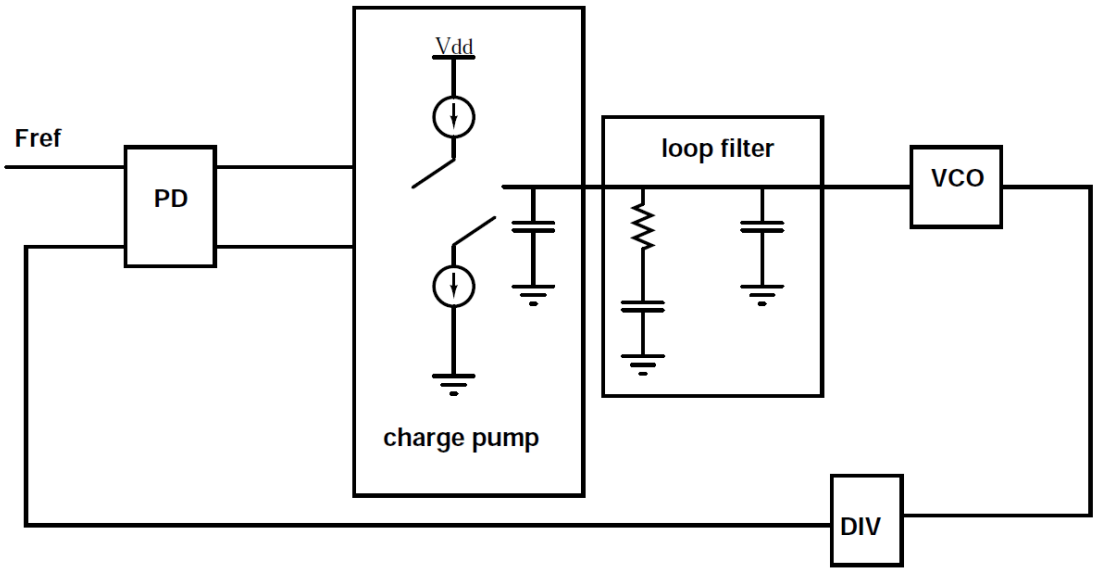
A 9-bit 50MS/s Asynchronous SAR ADC in 28nm CMOS

Tuan-Vu Cao, Snorre Aunet, Trond Ytterdal
Circuits and Systems group, Department of Electronics and Telecommunications
Norwegian University of Science and Technology, Norway
Emails: {vu.tuan.cao, snorre.aunet, trond.ytterdal}@iet.ntnu.no

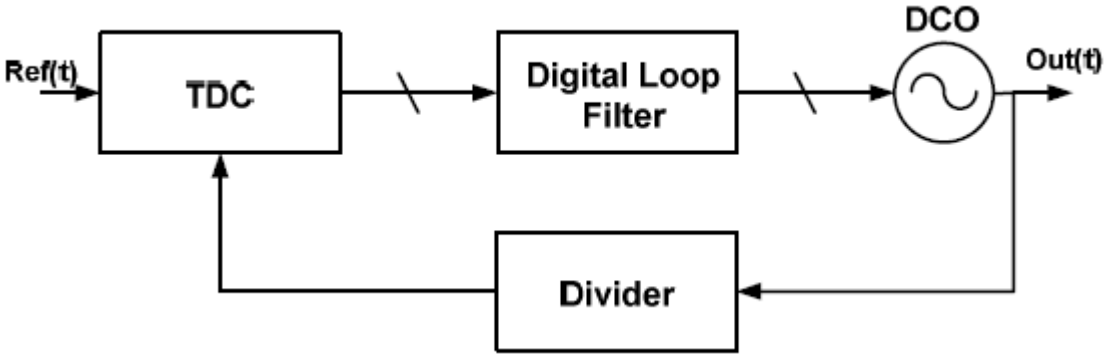
- Simulations indicate that 9 bit and 50 MS/s are achievable with 45 uW of power!
- Digitization and basic DSP functions at 50 MS/s could be achieved with 150 uW of power per channel (excluding the very front-end part).

TDC growth

Due to the interest in ALL Digital PLL (ADPLL), TDC left the niche market of instrumentation (HEP, range finding, testing) to become more general purpose components, with an significant increase on the number of published papers in the recent years.



Conventional analog PLL
Needs an analog filter



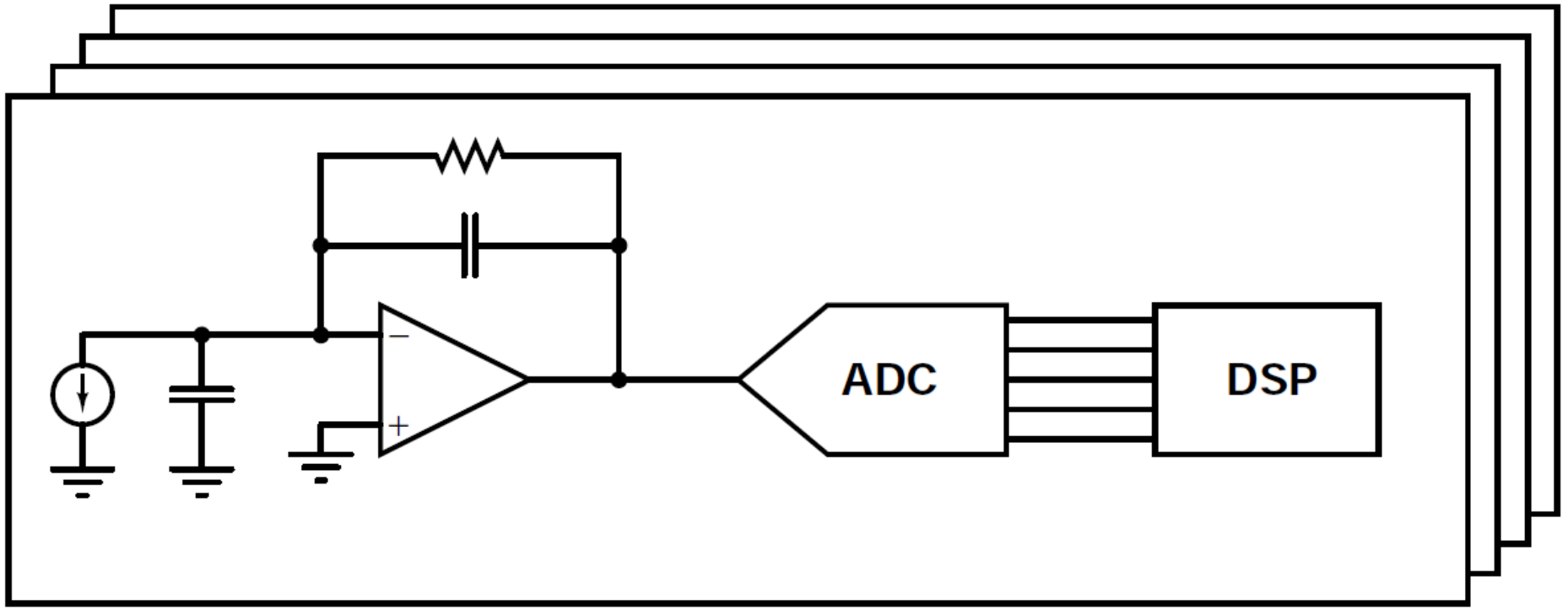
ALL Digital PLL
Phase difference measured with a TDC

Some recent TDC designs

Ref	Technology	Architecture	Resolution (ps)	Sampling rate (MS/s)	Range (ns)	Power (mW)	Area
1	130 nm	GRO	1	50	12	2.2-21	0.04
2	130 nm	Vernier-ring	8	15	32	7.5	0.26
3	90 nm	Passive inter.	4.7	180	0.6	3.6	0.02
4	90 nm	Delay line	20	26	0.64	6.9	0.01
5	65 nm	2D delay line	4.8	50	< 0.6	1.7	0.02
6	90	Time Amp.	1.25	10	0.64	3	0.6
7	90	Vernier+GRO	3.2	25-100	40	3.6-4.5	0.027

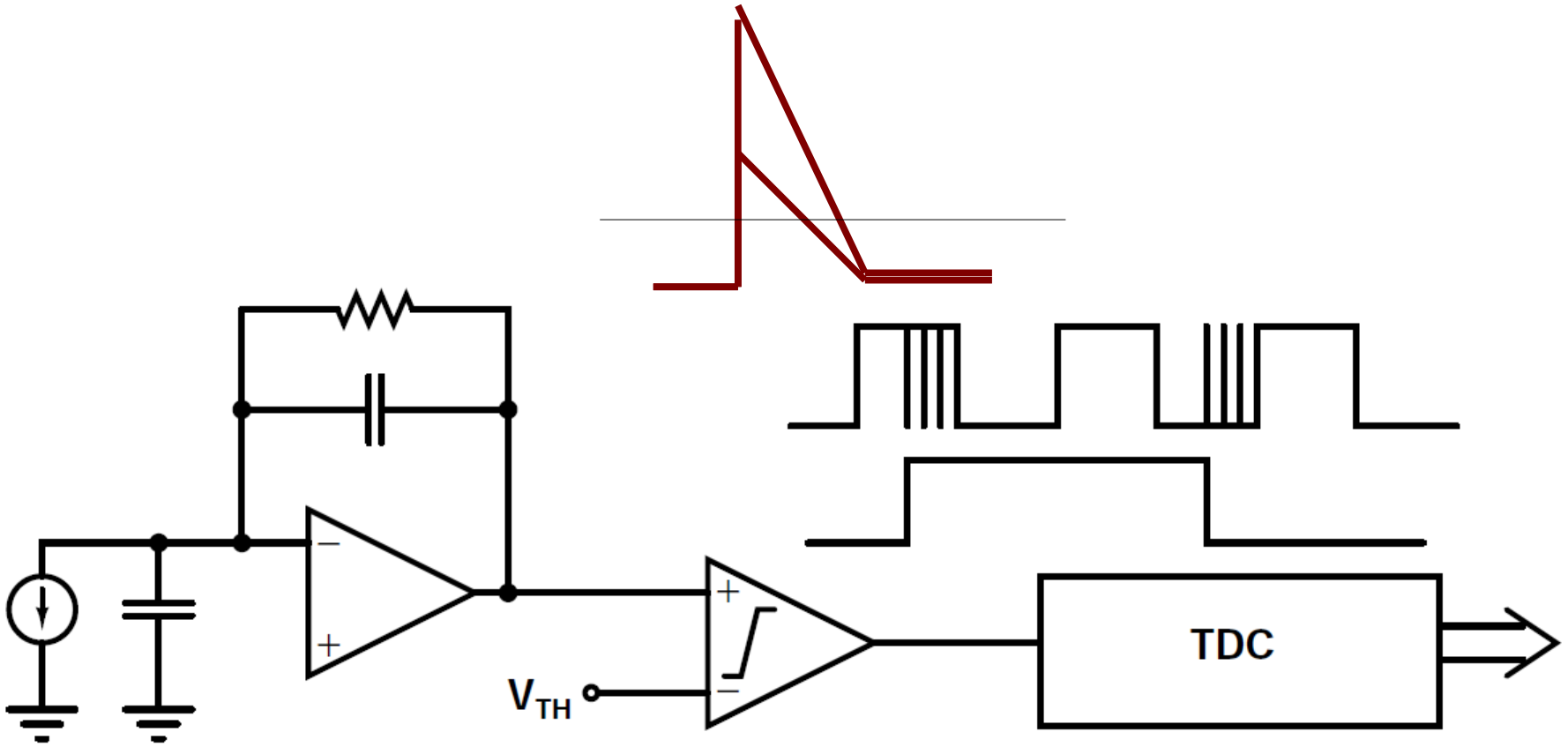
- TDCs are now reaching the sub-ps resolution
- Many different architectures
- Dynamic range low in many high resolution TDC

Future front-ends (1)



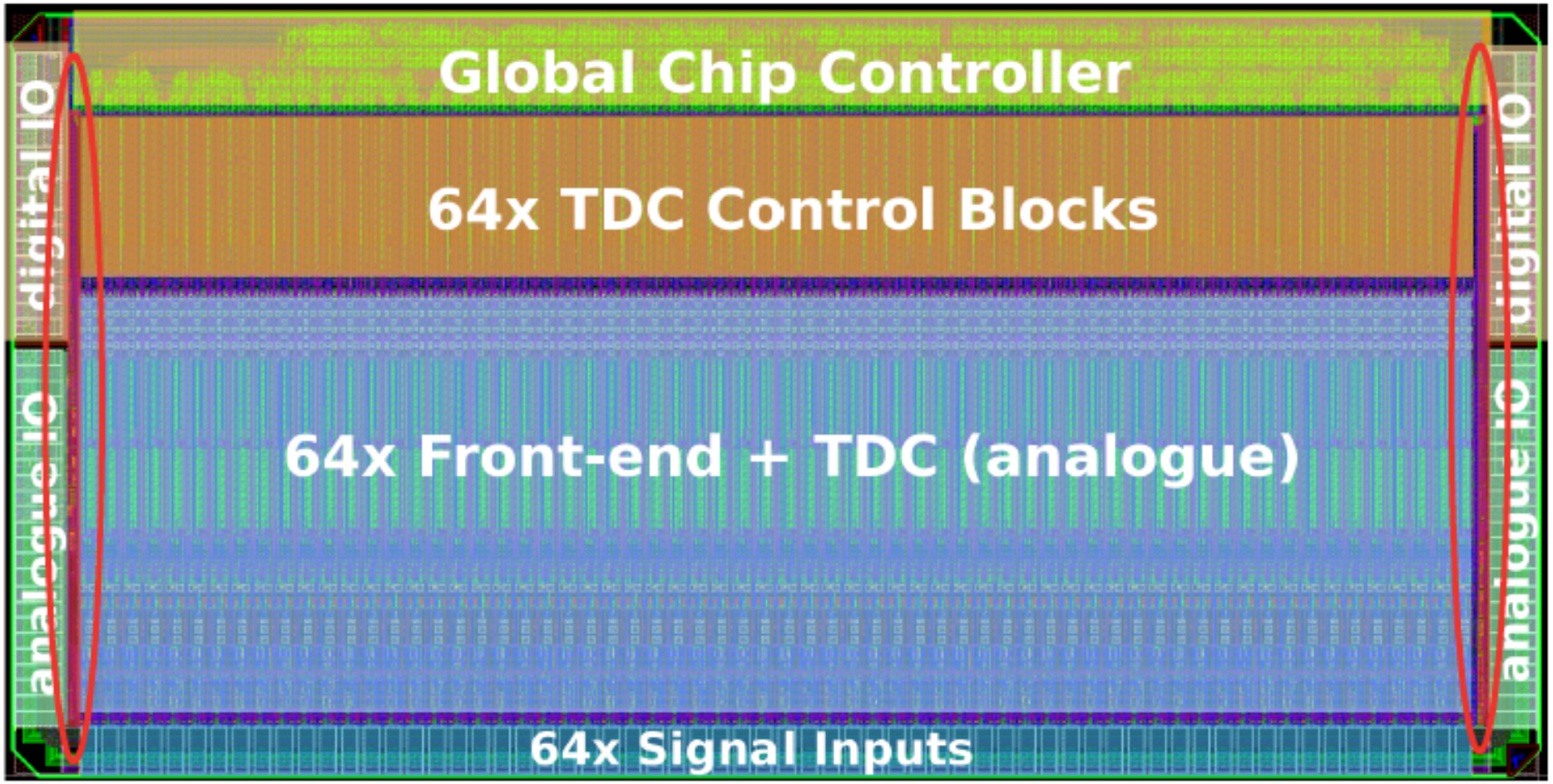
- **Future front-end will look more and more like this**
- **Embedded DSP allows for flexiblity, feature extraction, correction of analog imperfection**

Future front-ends (2)



- Energy can also be extracted from timing measurement.
- Use of TDC allows for fast time over threshold system.

A time based front-end



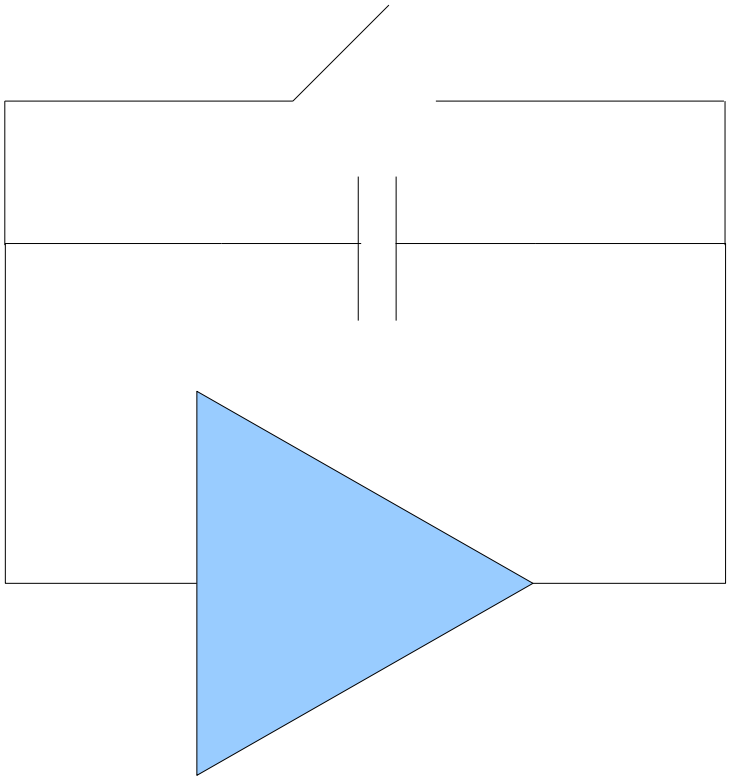
64 channels, 128 TDC, 300 kHz per channel, 25-50 ps binning

Pixel detectors

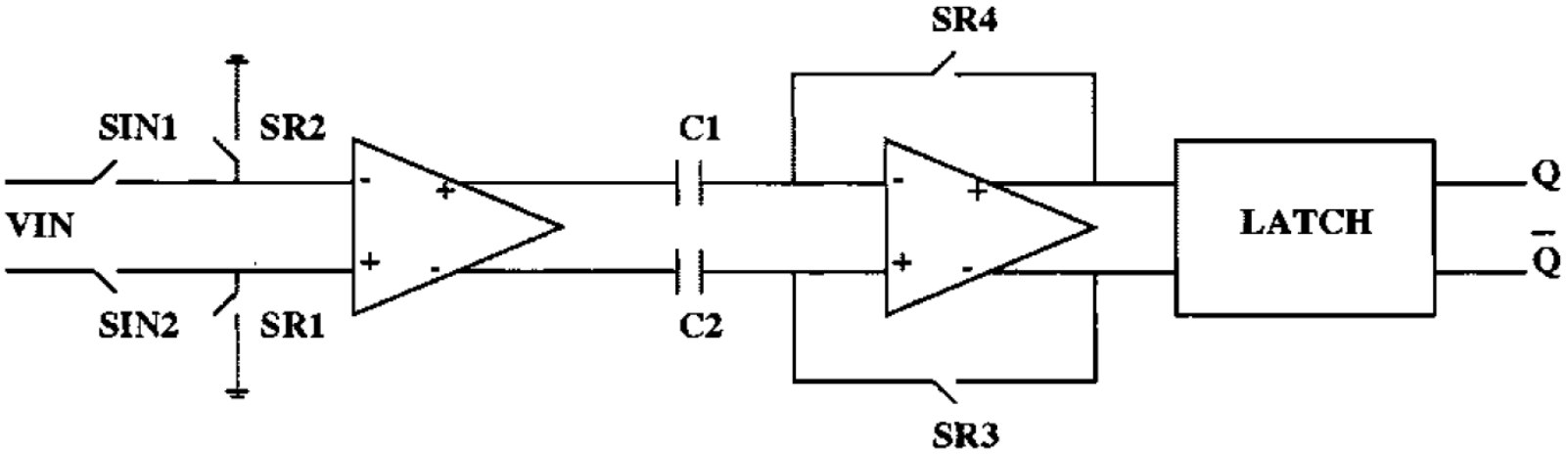
- **Hybrid pixels likely to play a key role also in next generation of experiments.**
- **Area available for the front-end electronics is one of the factors limiting the minimum pixel size.**
- **Moving to more and more scale technologies can improve significantly future vertex detectors.**
- **Future vertex detectors at the upgraded LHC will have to withstand much increased data rate O(2 Ghz/cm^2).**
- **Present generation of hybrid pixels: CMOS 130 nm**
- **Next generations CMOS 65 nm.**
- **Smaller pixels (e.g. 25 x 100)**

Discrete-time front-end

- **In 65 nm and beyond transistors become very small.**
- **When a transistors is used as an analog switched, charge injection become minimal.**
- **In many applications, discrete time front-end can be (again!) an attractive option.**



Discrete-time comparators



- Discrete time comparator allow the use of positive feed-back.
- High sensitivity and no problem with in-time threshold.
- Offset can be suitable stored on capacitors and subtracted.
- Due to the leakage of the switches, the procedure must be repeated from time to time
- Very small residual offset can be achieved.

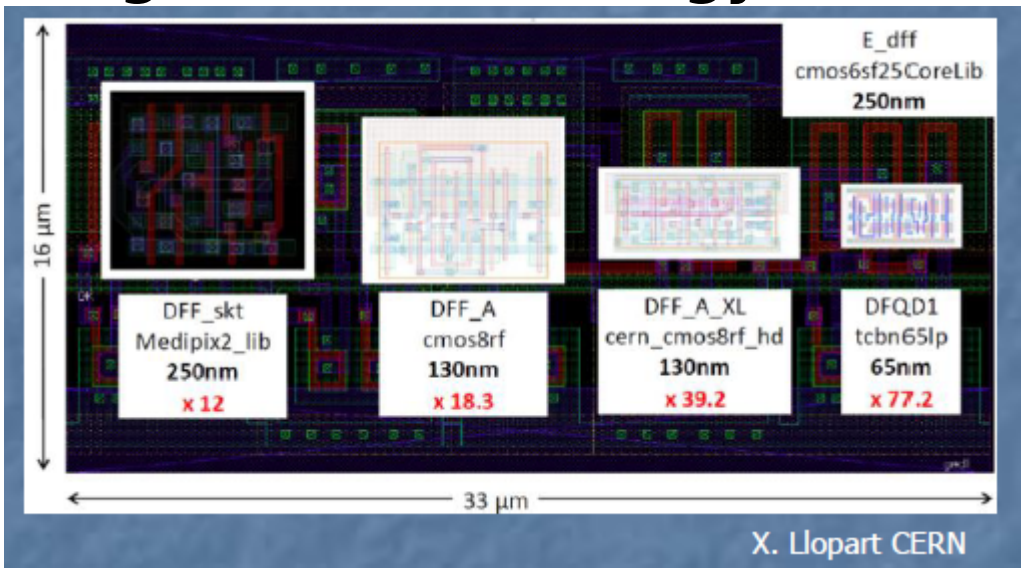
$$V_{\mathrm{OSR}} = \frac{\Delta Q}{A_{01}C} + \frac{V_{\mathrm{OSL}}}{A_{01}A_{02}} + \frac{V_{OSA2}}{A_{01}(1+A_{02})}$$

UDSM usage

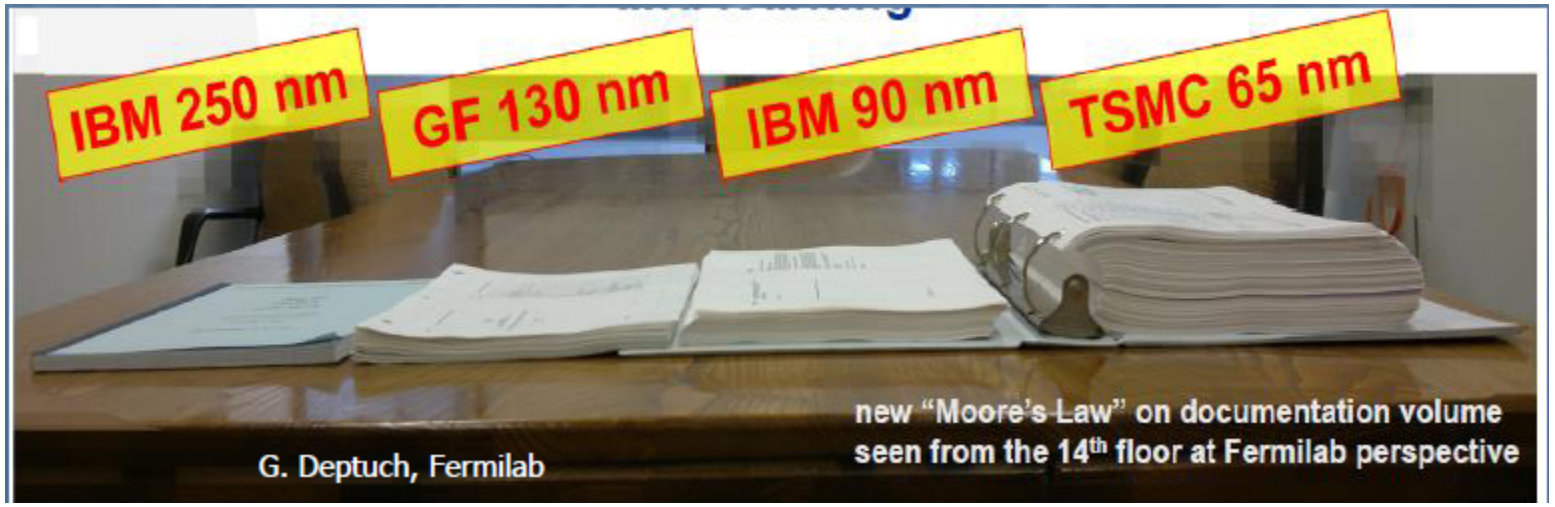
- **Deep sub-micron technologies are however very complex**
- **A lot of reliability rules, especially for obtaining good analog performance.**
- **Chip design is becoming more and more complex, needs to share expertise and common block among different design center:**
- **Cooperation and coordination of efforts is paramount**
- **CERN R&D 53 initiative**
  - **Oriented to new generation of pixel detectors for LHC upgrades**
  - **Common ATLAS-CMS effort**

UDSM usage

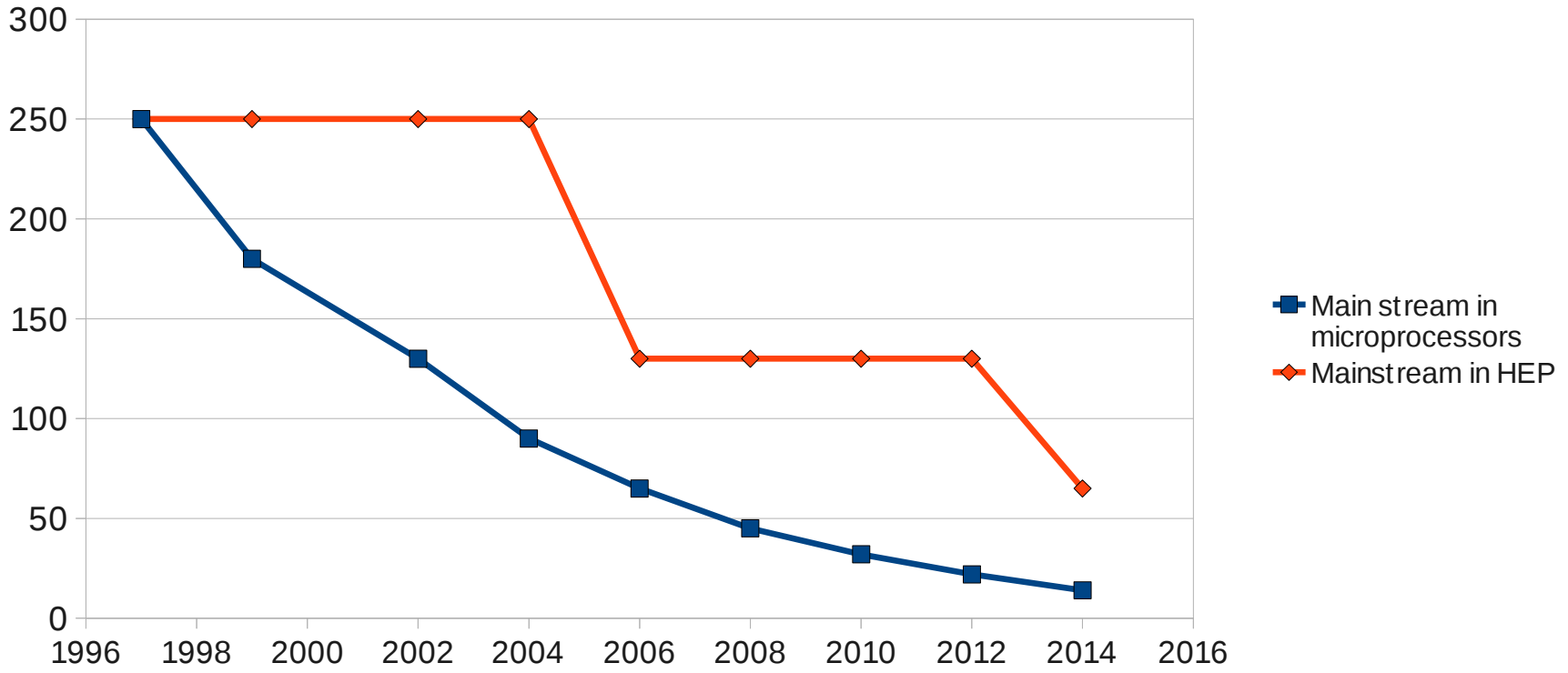
- **Direct scaling of the technology…..**



- **Reverse scaling of the design rules:**



A final thought



- **MPW: 15 mm2 in 28 nm: 156 000 (standard), 189 000 SOI**
- **1 mm2: 15 000 euros**
- **http://cmp.imag.fr/products/ic/?p=prices**

UDSM can allow a significant leap in the design of front end electronics for radiation sensors
It is really time to move forward!