Vertexing and tracking with high radiation environment and high pile-up

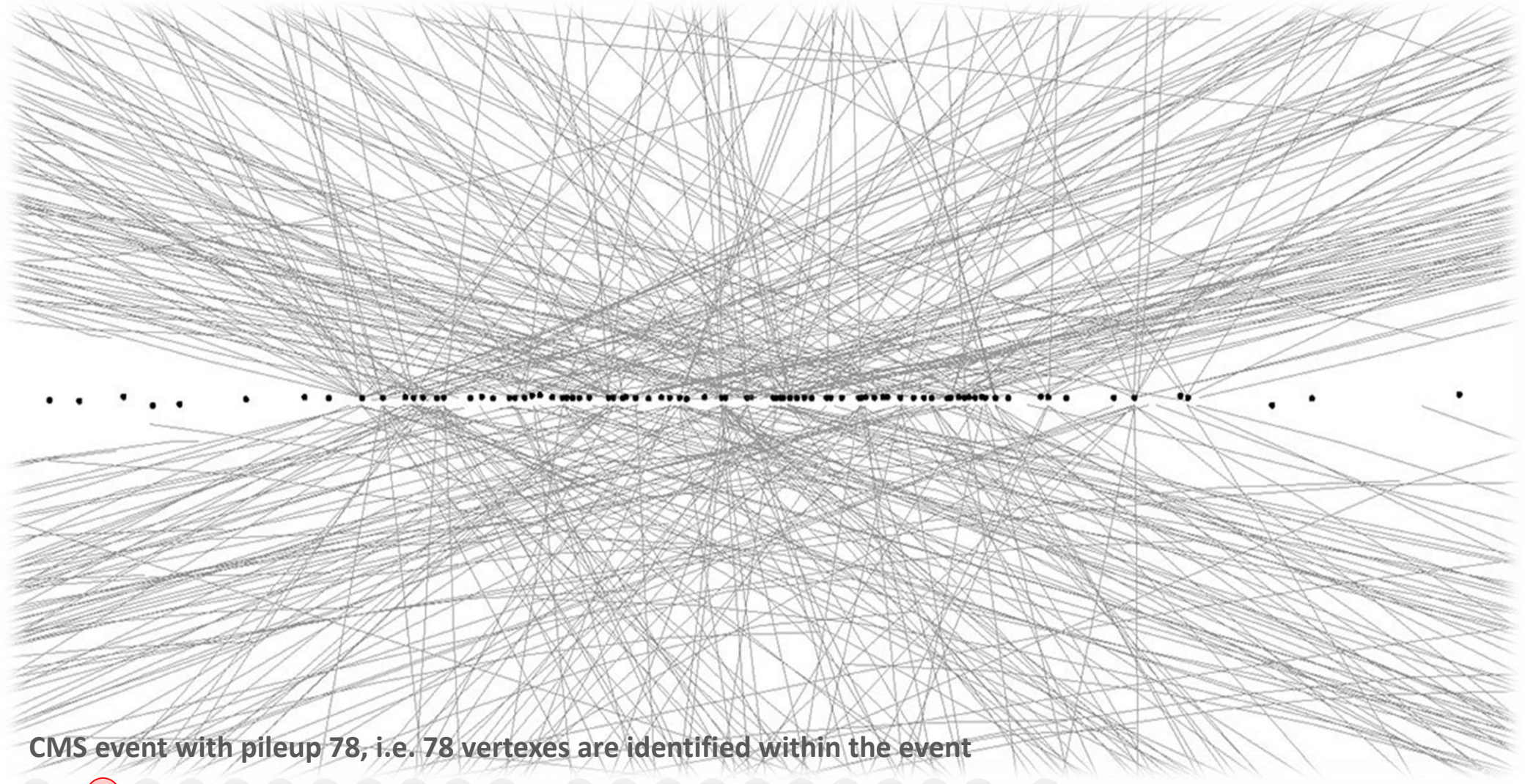
Piero Giubilato
Padova University & INFN

Signal characteristics

also known as hits density and cluster properties,

High pileup – what it is

- Every time these bunches cross one another, more than one proton-proton collision (interaction) takes place: this is called **pile-up** in the HEP community.



CMS event with pileup 78, i.e. 78 vertexes are identified within the event

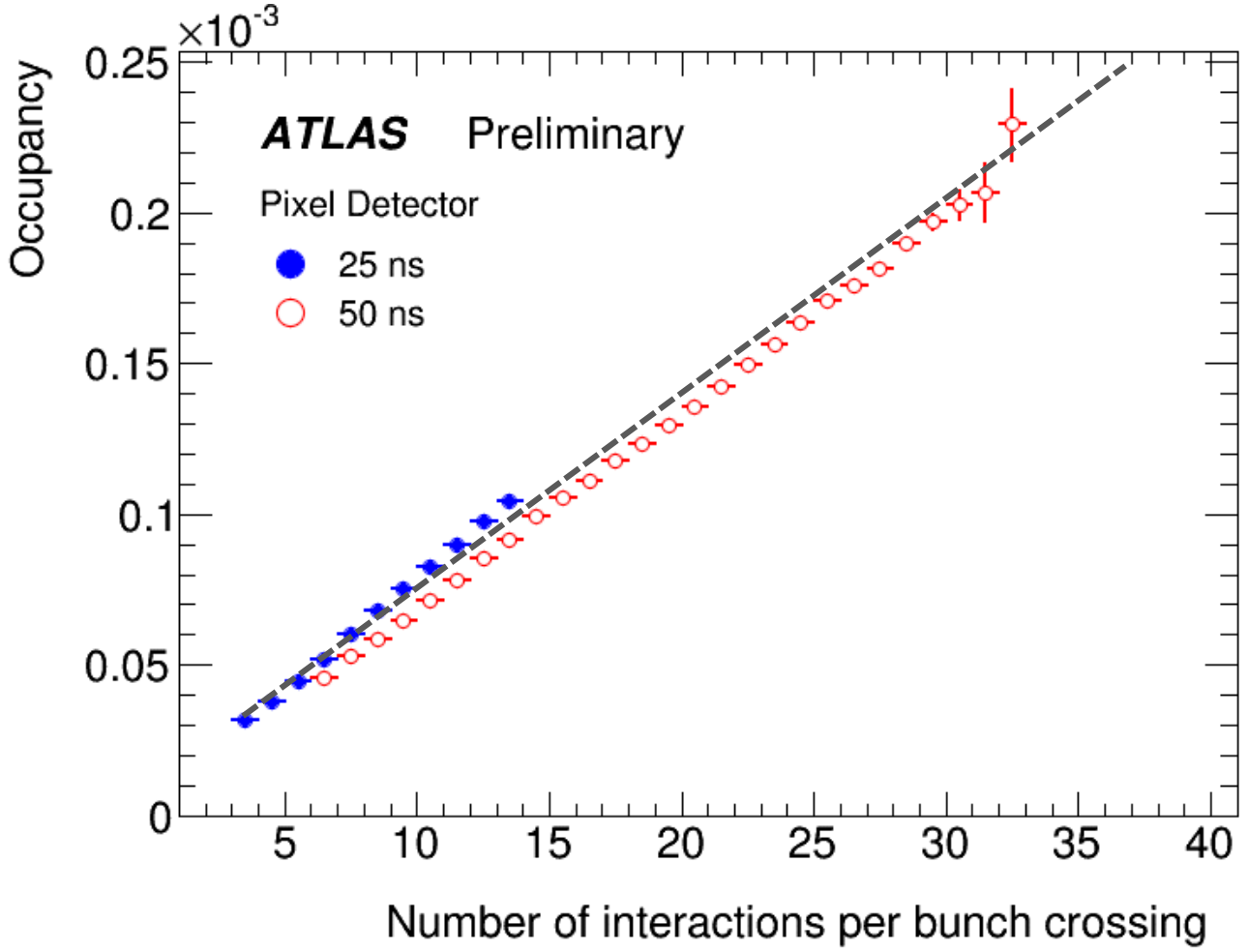
High pileup – what to expect

- When CMS/ATLAS were designed, the expected average pileup was **25**, with a bunch crossing rate of 40 MHz (25 ns) for a total instantaneous luminosity of 1×10^{34} $cm^{-2}s^{-1}$.
- Presently LHC is running at 0.7×10^{34} $cm^{-2}s^{-1}$ instantaneous luminosity, but with 50ns bunch separation, implying a higher average pileup. Both experiments (ATLAS, CMS) are coping well with such operating conditions.
- After 2024, LHC is supposed to run at 5×10^{34} $cm^{-2}s^{-1}$ instantaneous luminosity, with an expected average pileup of **140**.
- More than the pileup (cause), what matters is the **occupancy** (effect), i.e. the fraction of pixels per event with signals.

Higher than 10^{-3} px^{-1} occupancy (pileup > 140 for ATLAS)

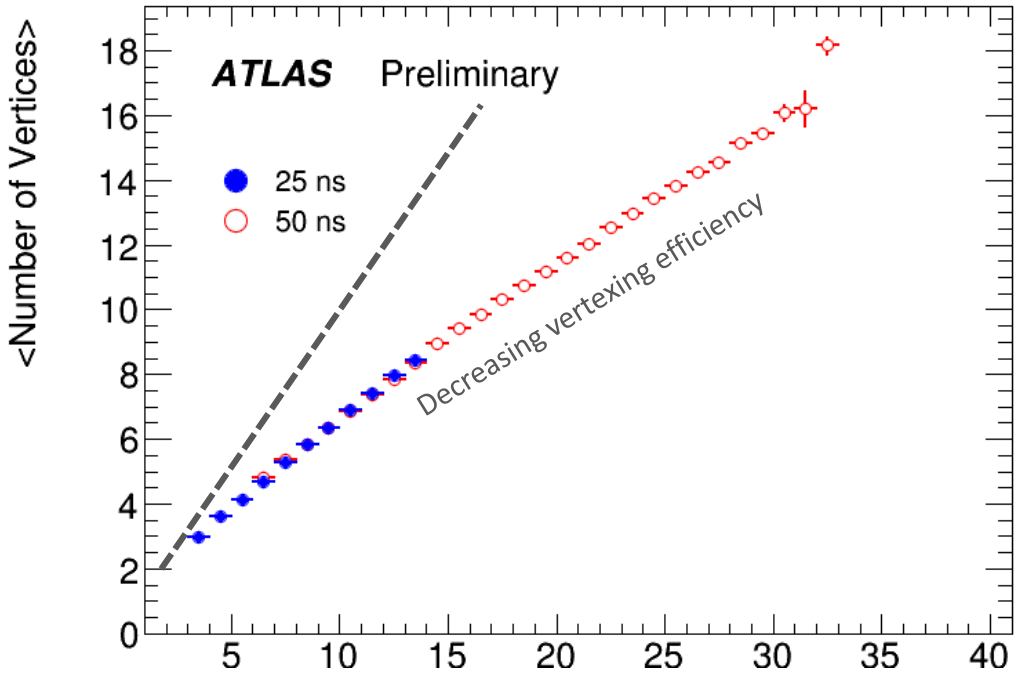
High pileup – actually better to look at occupancy

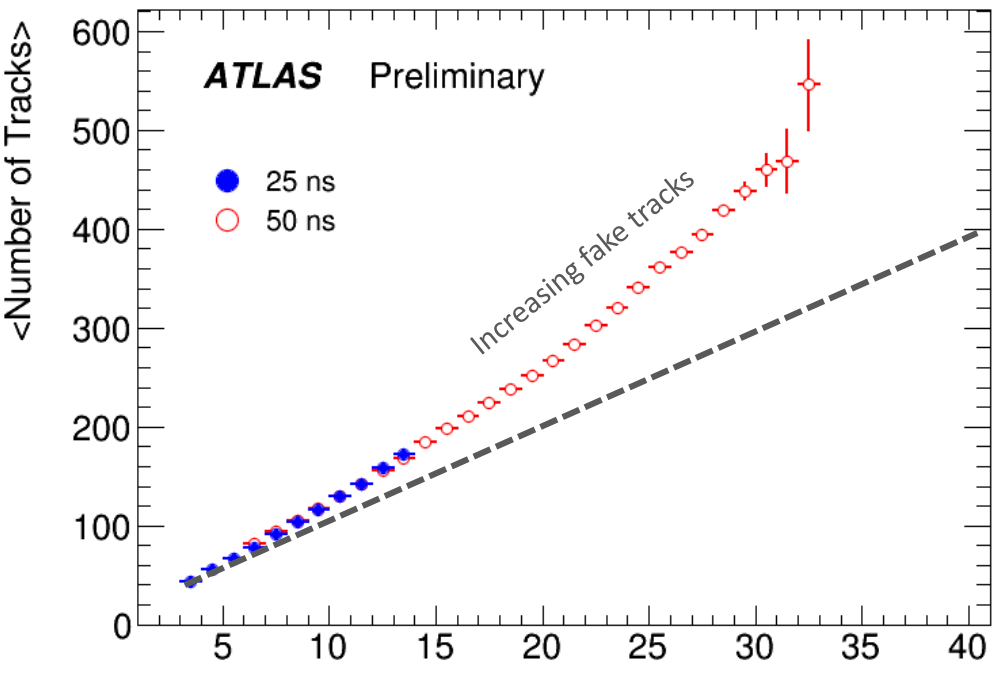
- Occupancy is what to look for. Typically, less than **10^{-3} px^{-1}** for tracking applications. In first approximation, the hits distribution can be considered uniform in HEP tracking conditions.



High pileup – challenge for tracking efficiency

- **At higher occupancies** (> 10^{-2} px^{-1}), other compression methods are more effective respect to the traditional sparsification schemes employe din HEP pixel sensors.
- Given a design occupancy, higher occupancy levels usually compromise both vertexing and tracking efficiency, mostly due to the increased fake tracks number.





High pileup – occupancy as data load

For charged particle tracking in HEP and similar applications, some assumption can be made about the data we are interested to:

- A (recorded) particle passing through a sensor generates an **hit**.
- Every hit is associated to one or more pixel (**cluster**) containing the released charge. A single pixel hit is anyway a cluster.
- For every hit, we may or may not want to know:
  - the **shape** of the cluster
  - the **total charge** of the cluster
  - the **fractional charge** (per pixel) within the cluster
- We consider timing a property of the event, not of the hit (but this is not general).

- From some basic information entropy considerations, we can estimate the minimum (best) data load necessary to transmit the hits information from the sensor to the outside world.

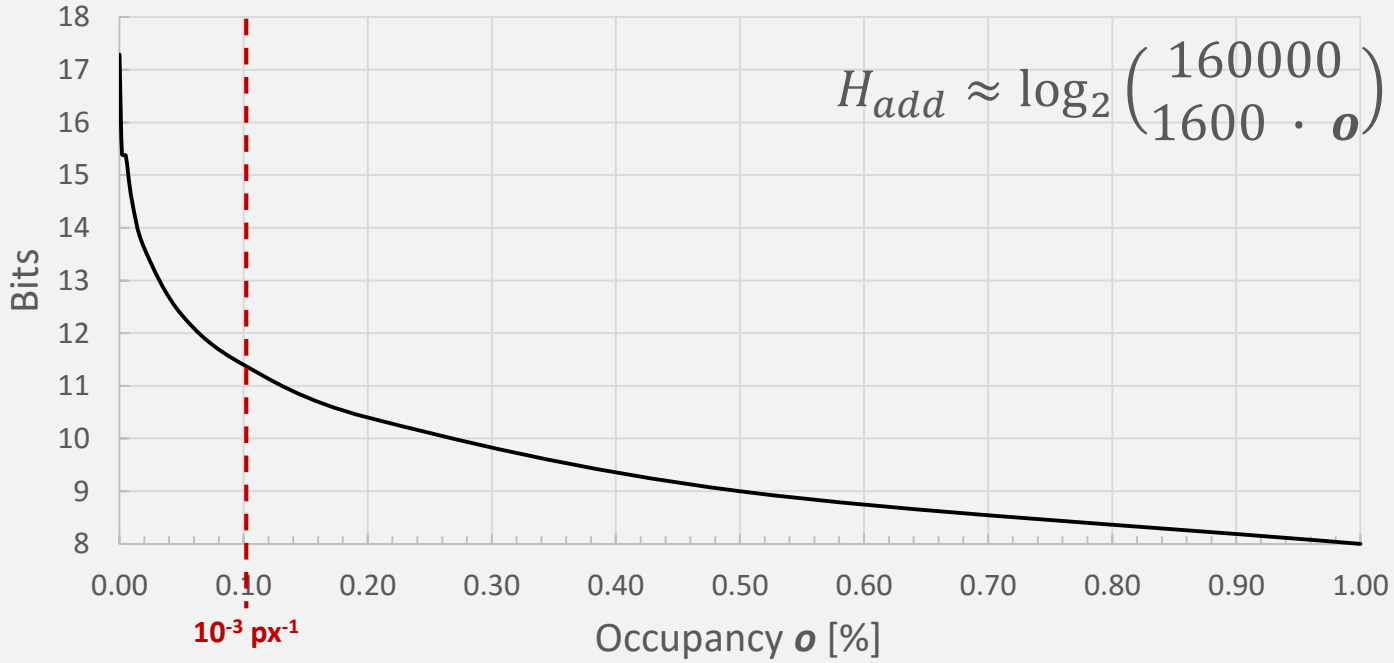
High pileup – expected data entropy 1

Cluster address entropy

Number of bits (entropy) necessary to **encode the address of the hit within the sensor**, assuming an ***n*** pixel sensor with ***k*** hits on it:

$$H_{address} \approx \log_2 \binom{n}{k}$$

Where we consider an occupancy low enough (< 1%) so the address entropy of real cluster is well approximated by that of single-pixel hits. For an example sensor of 2 × 2 cm^2 active area with 50 µm pitch pixel and an occupancy ***o*** [%] it results:



High pileup – expected data entropy 2

Cluster shape entropy

Number of bits (entropy) necessary to **encode the cluster shape**, where N is the number of possible shape, and P a "small" number of the most likely shapes (P<N), and x the fraction of clusters of a given shape.

$$H_{shape} \approx (1-x)\log_2\left(\frac{1-x}{N-P}\right)$$

Operating experience from current detectors shows how **4 bits** cover the vast majority of actual cluster shape entropy distribution found on a single sensor (not the whole detector!).

Cluster charge entropy

Number of bits necessary to **encode the cluster charge**, where D is the number of bits to digitize the charge value and p_i is the probability of each possible value (convolution of Landau charge release + Gaussian noise distribution):

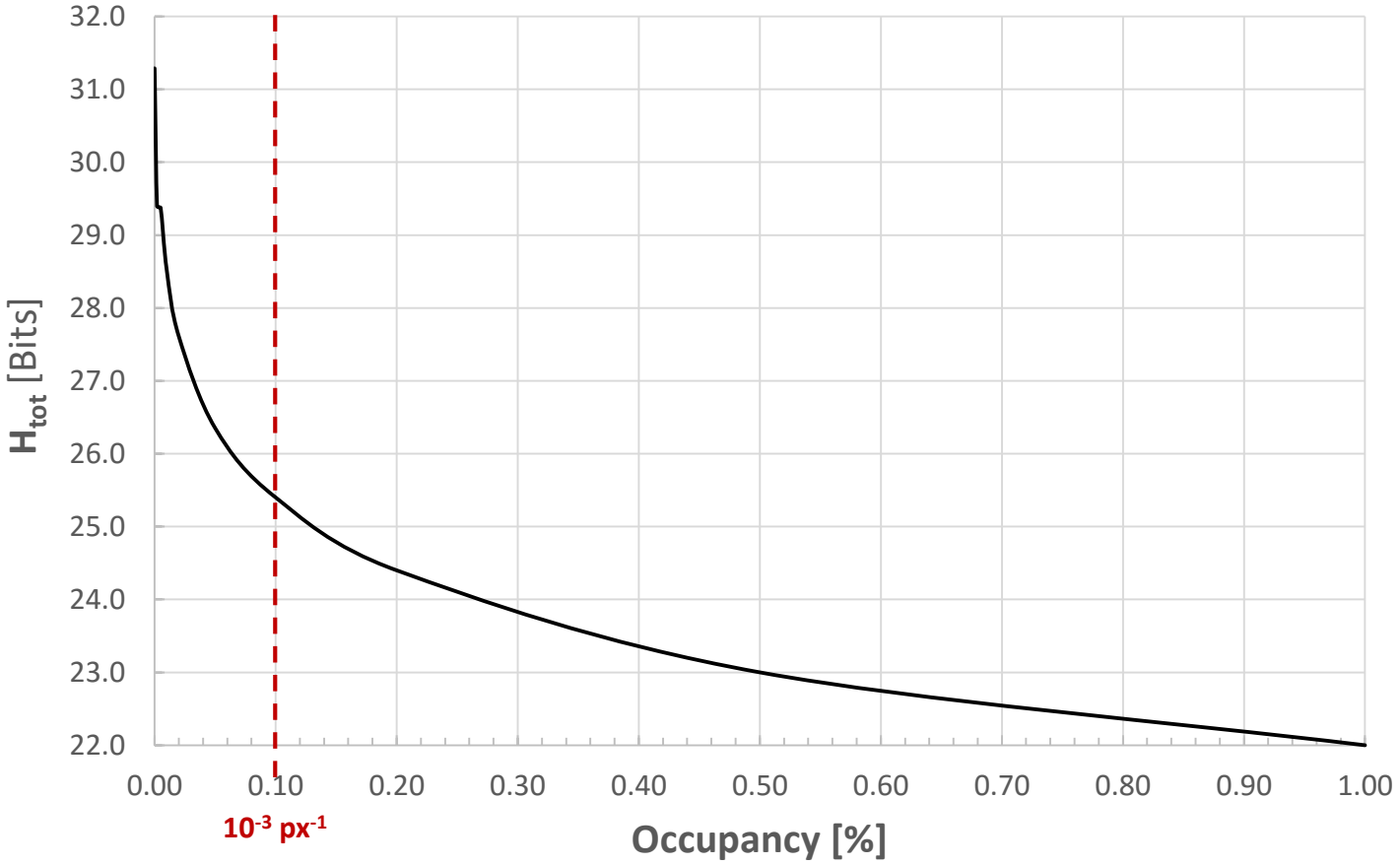
$$H_{charge} \approx -\sum_{i=0}^{2^D} p_i \log_2(p_i)$$

With a S/N of at least 40 and 1 bit error (ideal case), **5 bit** suffice to describe the expected charge spectrum. Other **4 bits** must be added if per-pixel charge knowledge is required.

High pileup – expected total data entropy

Summing all the contributions seen in the previous slides, we get the minimum number of bits necessary to describe a hit. Staying with our toy sensor of 2 × 2 cm^2 active area and 50 µm pixel pitch (200 × 200 pixels):

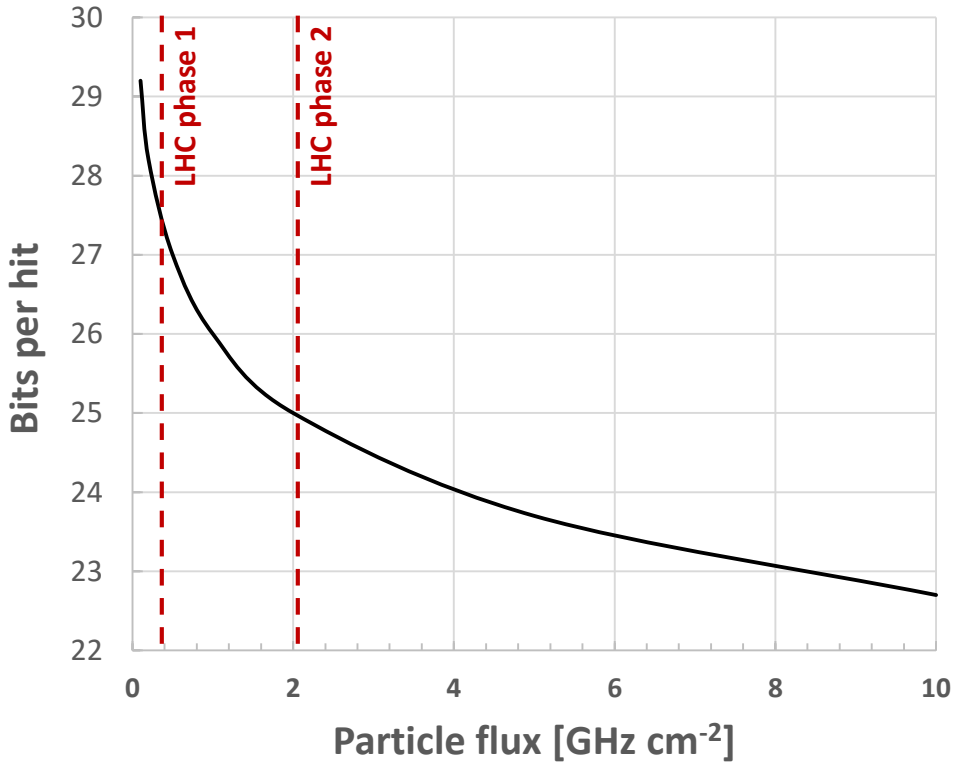
$$H_{tot} \approx H_{address} + 4 + 5 + 5 > 8 + 14 = 22\ bit$$



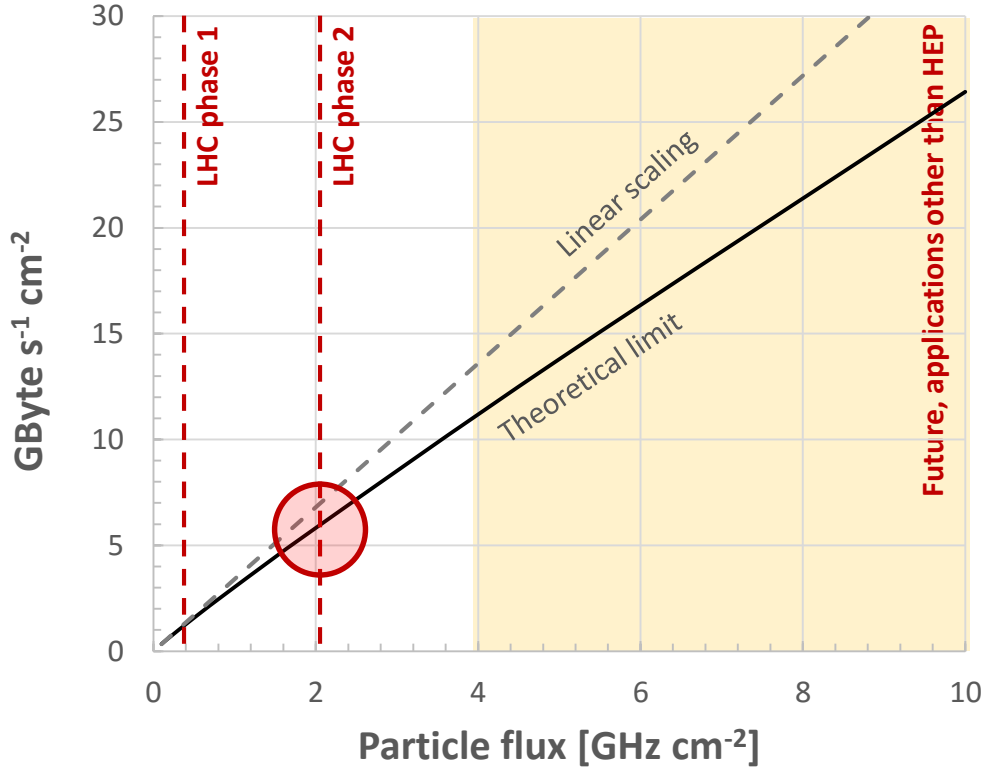
High pileup – expected total data load

Summing all the contributions seen in the previous slides, we get the minimum number of bits necessary to describe a hit. Staying with our toy sensor of 2 × 2 cm^2 active area and **50 µm pixel pitch** (200 × 200 pixels):

Hit encoding vs. particle flux

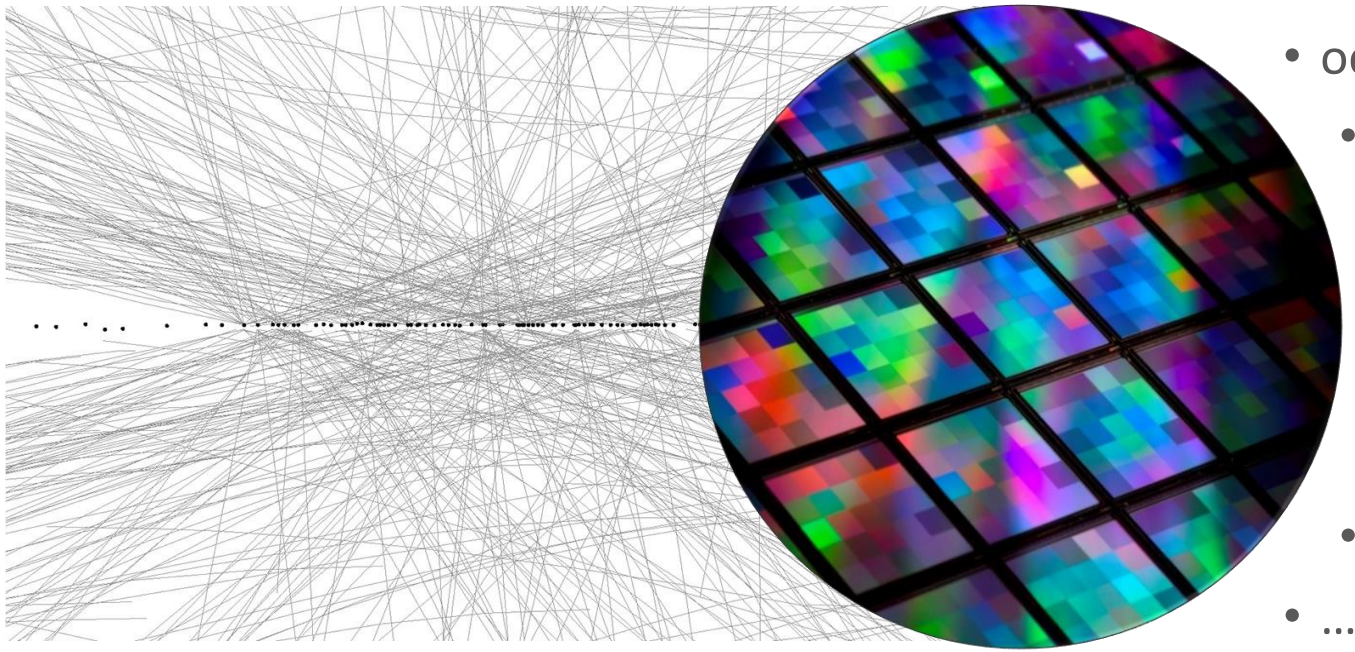


Raw data rate vs. particle flux



High pileup – the sensor electronic as an encoding machine

We can look at the sensor electronic (whichever embedded or connected) as an encoding machine tuned to optimize the data perflow given some (many) boundary conditions.



- occupancy
  - **pixel cell size**
    - **radiation tolerance**
    - power budget
    - **material budget**
    - technology node
  - financial budget
- …many others…

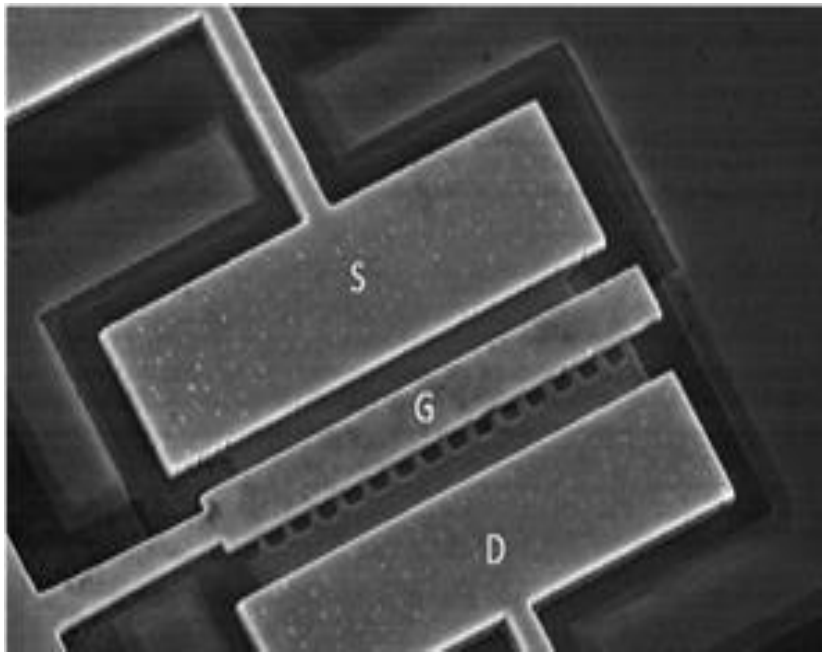
$> 5 \frac{GByte}{s}$ → *Trigger available in HEP* $< 0.5 \frac{GByte}{s}$

For a toy chip of 4 cm^2, 50 µm pixel pitch and 2 GHz cm^{-2} particle flux, **minimum ideal data rate**

1m electrical channel @ 4 Gbit s^{-1}, about 2-3 times better than present best, **so this is optimistic**

Radiation tolerance

Just about the most frequently used transistors (CMOS)

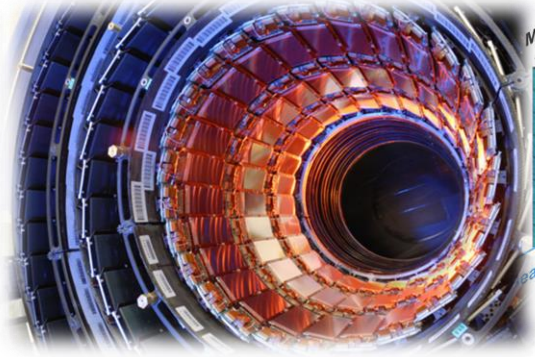


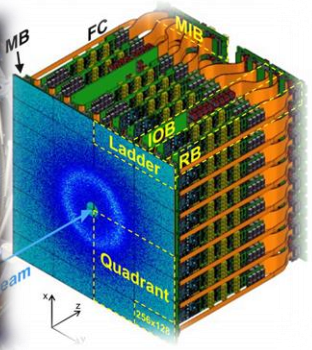
Radiation – framing the problem

Typical space levels (NASA)

changing the sensor is less a concern here, plus the sensor partially screens the electronic

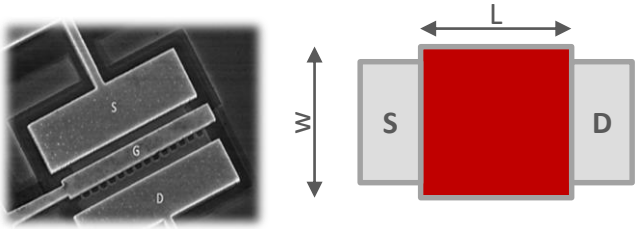
		COTS	Rad Tolerant	Rad Hard	LHC Phase2	HL LHC	FEL (3yrs)
TID	$[Gy]$ (rad)	100 (10krad)	500 (50krad)	1×10^4 (1Mrad)	1×10^6 (100Mrad)	**10×10^6 (1Grad)**	up to 10^9 (100 Grad)
NIEL	$\left[\frac{1\ Mev\ n_{eq}}{cm^2}\right]$	-	-	-	2×10^{15}	2×10^{16}	≈ 0 (< 300 keV γ)
SEU LET_{thr}	$\left[\frac{Mev}{mg\ cm^2}\right]$	5 Mev	20	80 ÷ 150	40	40	-

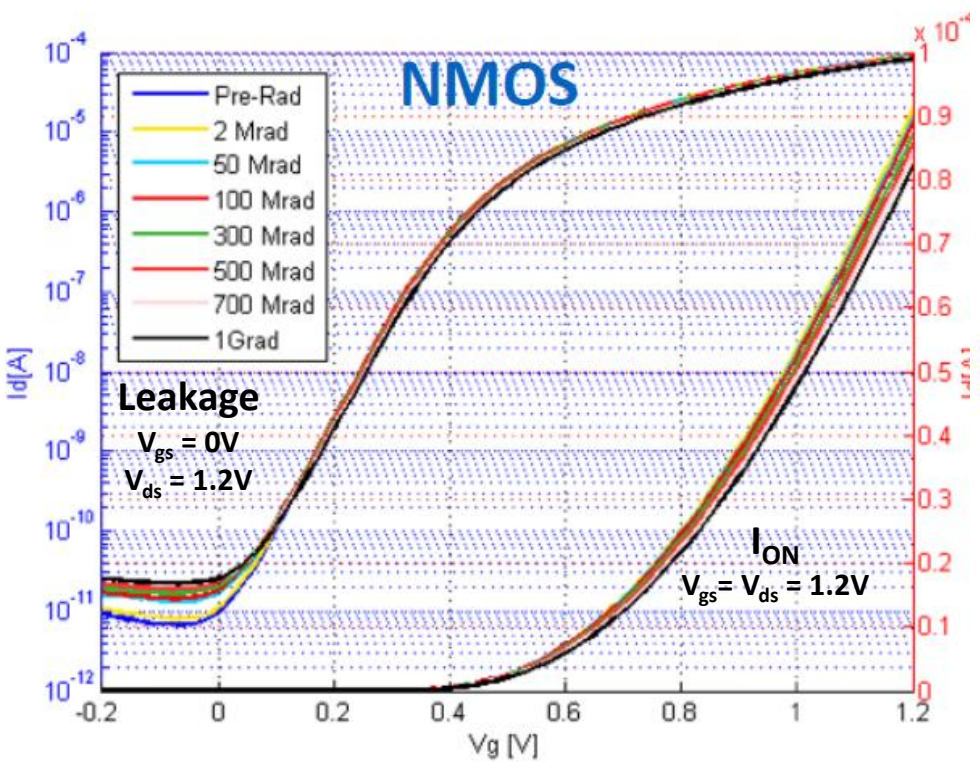


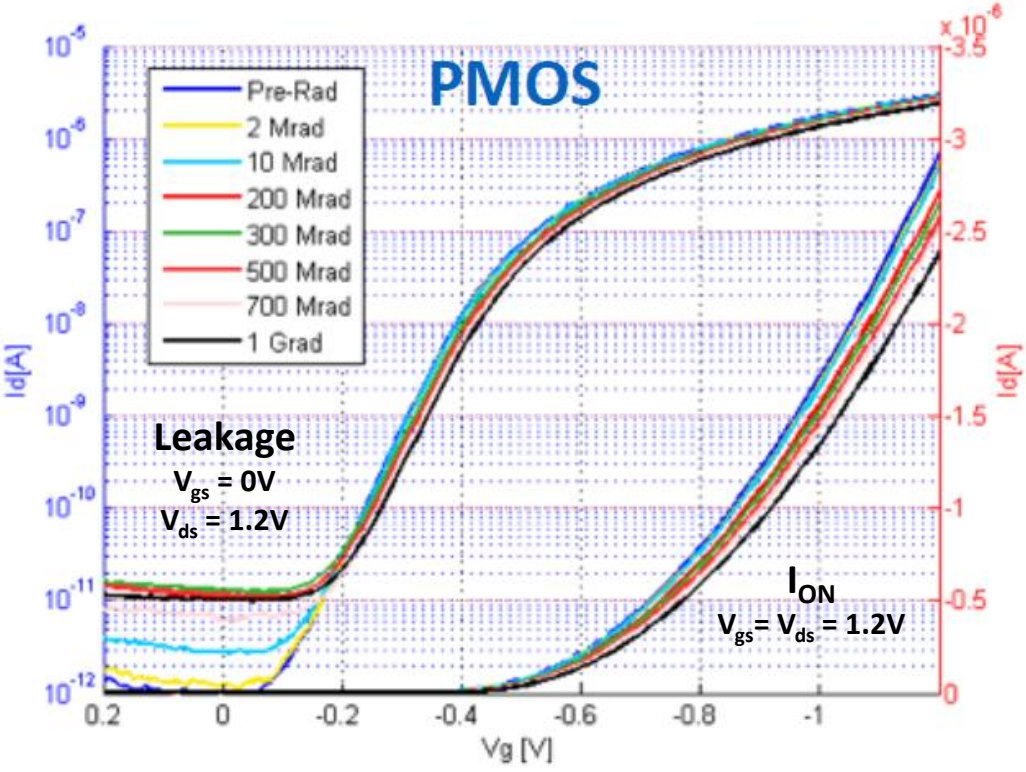


Radiation – 65nm “large” transistors

Large transistors (long and wide channel) characteristics undergo small changes up to 10 MG (1 Grad). Even if increase, the leakage current is well below 40 nA nm, therefore not a problem for usual configurations. **Thin gate oxide is rad-hard!**



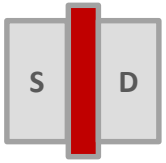


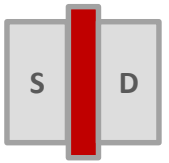


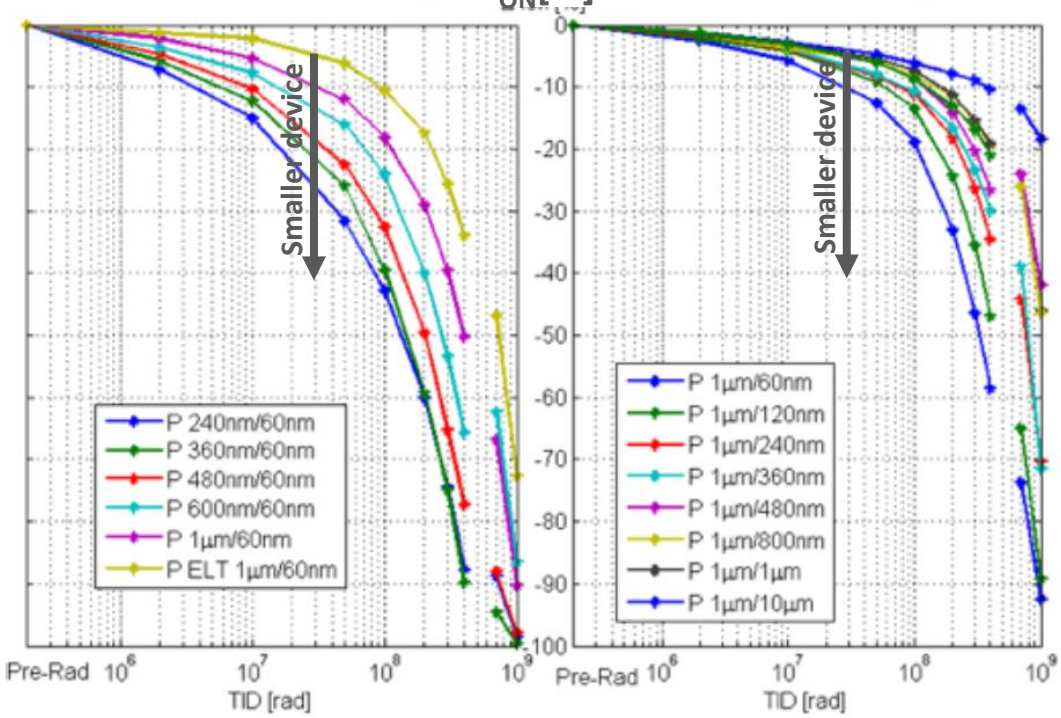
Radiation – 65nm transistors aspect ratio influence

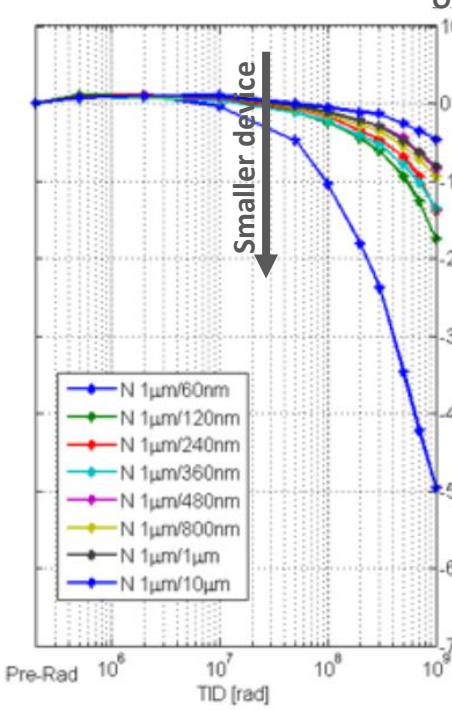
Faccio et al.
TWEPP 2015

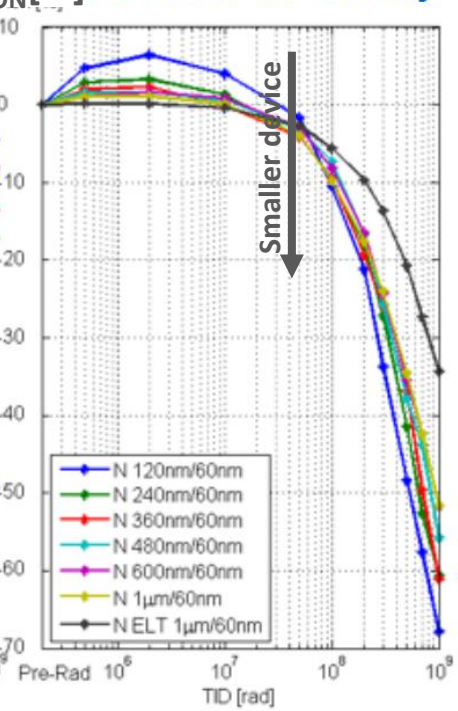
Severe radiation damage in "narrow" and "short" channel transistors. The effect has been shown to depend on bias and temperature applied during and after irradiation. The narrow and short configuration (not shown here) is affected as well!







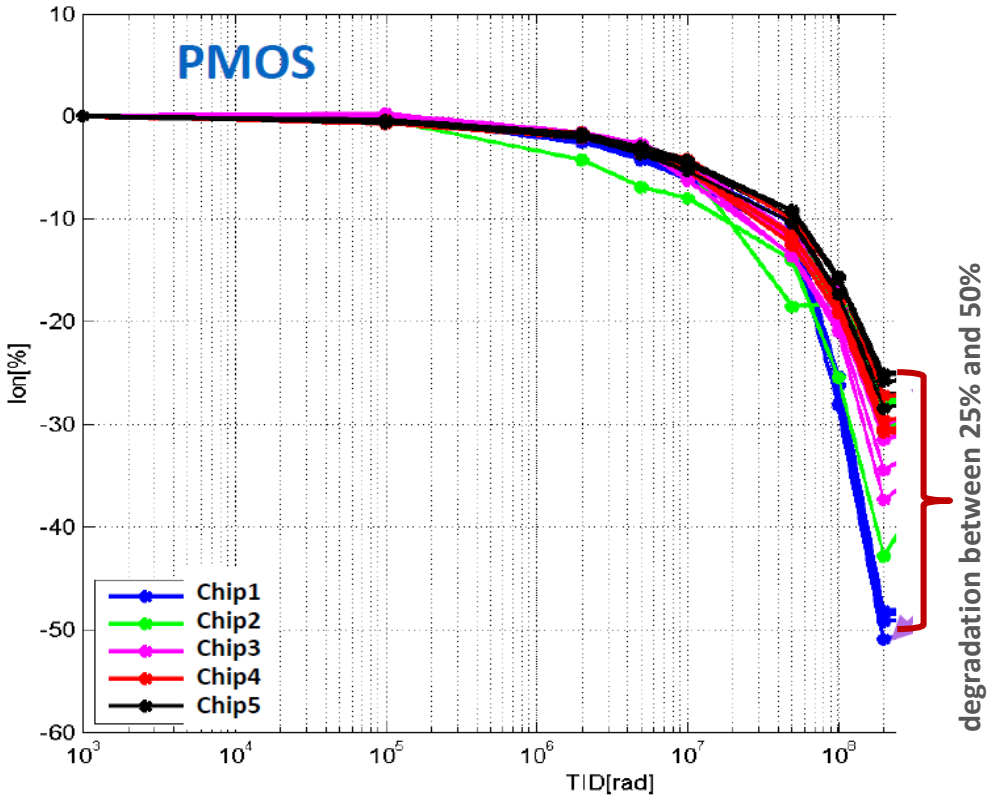




Radiation – 65nm transistors other issues

Faccio et al.
TWEPP 2015

"small" transistor response to radiation also shows **large differences** between different chips.

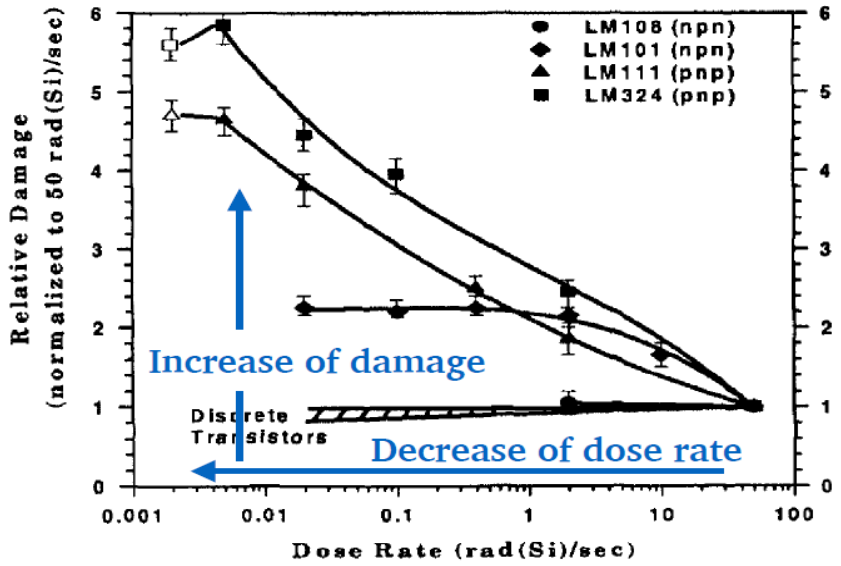


5 identical transistor in each chip (W = 1µm, L = 60nm)
5 chip irradiated in same condition
Irradiation at T = 25C, $V_{gs} = V_{ds} = 0V$

Many other effects found investigating 65nm in different collaboration, **look at the recent literature**. In particular:

- strong dependence on temperature
- non-monotonic behaviour of the degradation.

Plus, similarities with the behaviour of bipolar devices, which COULD imply:



Enhanced Low Dose Rate Sensitivity (ELDRs): TID degradation increases at lower dose rates

Radiation – where we are

1 Grad (10 MGy) seems a difficult frontier: gate oxide and leakage almost ok (and in case enclosed layout transistors), but:

- Sidewalls, spacers, shallow trenches source of charge trapping, less quality structures than gates and source/drain implants (similar to poor quality oxides in bipolars).
- The previous HEP experience (250 nm) hit a sweet spot between gate oxide thickness and parasitic structures influence (which is much more relevant for smaller nodes).
- Many other non well understood mechanism: traps transport mechanism, interfaces, holes mobility degradation, non SiO_2 materials, etc.
- Smaller the node, more difficult to get process detail from the "big" foundry providing it!

- **We DO NOT fully understand what is going on!**
- **Difficult we have the resources to investigate it!**

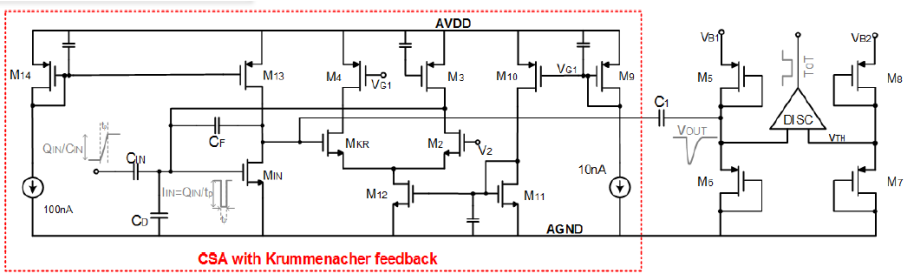
Plus (no room here to discuss them, equally important)

- CMOS normally not affected by **NIEL**: must anyway be checked for the 10^{16} n_{eq} level and smaller technology nodes, no room to discuss this here.
- **Single Event Effect** more relevant due to the gate density, appropriate design mandatory in every part of the design, therefore even more complexity and challenges for the design team.

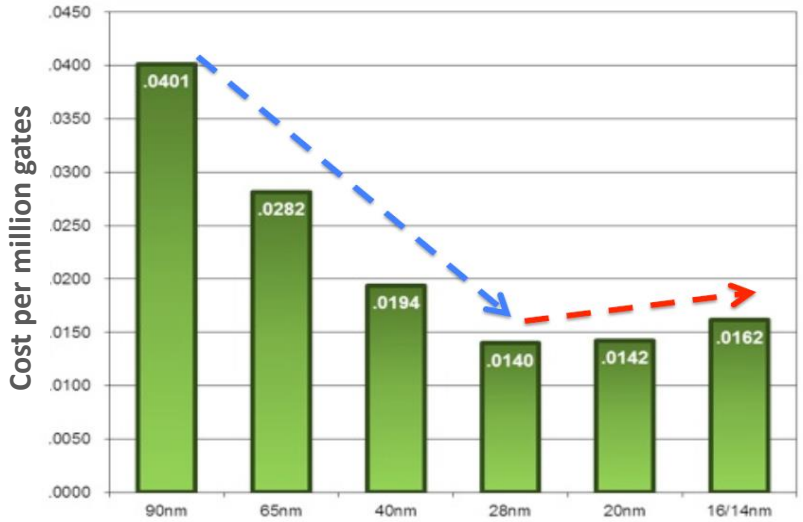
Radiation – possible planar solutions?

Using even smaller nodes could alleviate the problem… or worsen it?

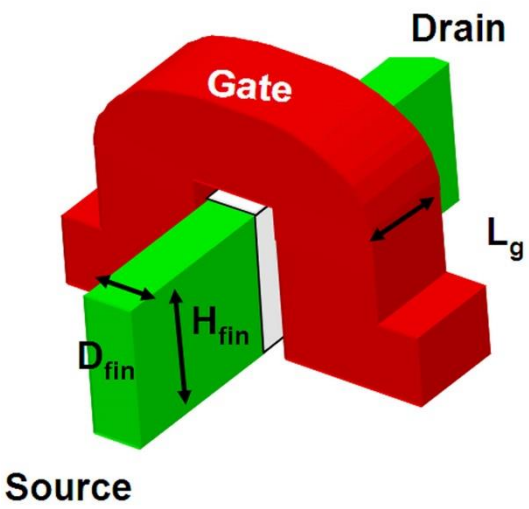
Scaltech 28 project (A. Baschirotto) investigates the 28nm node characteristics for pixel and services (high speed links) applications in harsh environments.

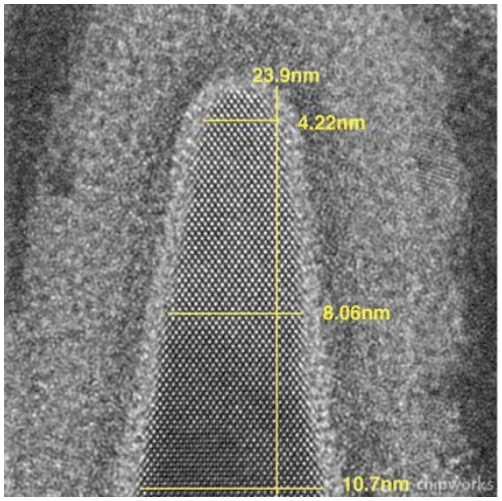


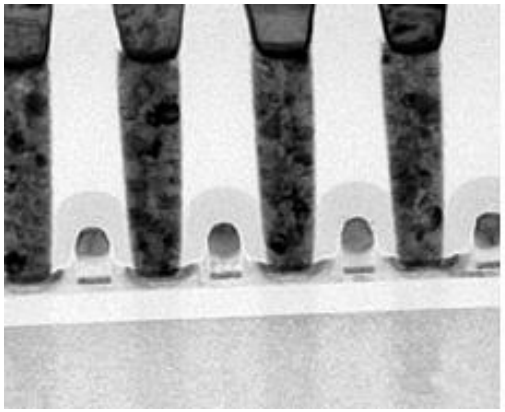
Cost becomes competitive for LARGE applications (few 100000 chips production, i.e. ≥200m^2 surface)

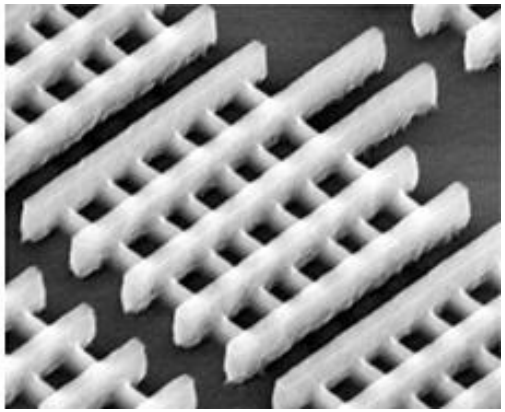


Even **more advanced nodes** could address the problem? Difficult (and expensive) to investigate.







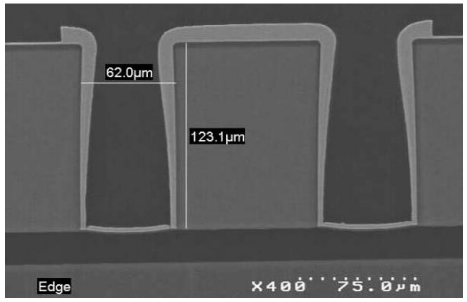


Radiation – possible non planar solutions?

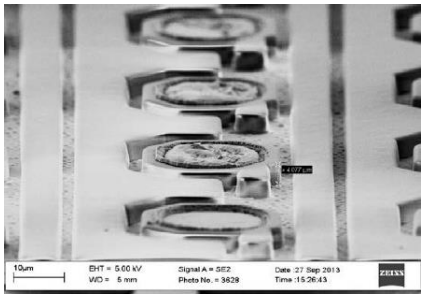
Is shrinking the transistor size proves too difficult, let stack more than one layer of transistors per unit area (also known as 3D!)

Lot of research done by many group on vertical integration in these lasts years. Here the Medipix 3 (130nm) example, but many others available.

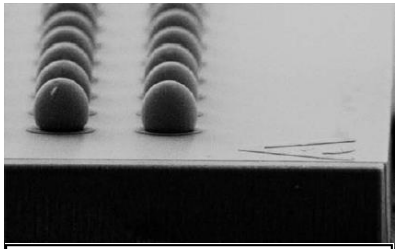
Through Silicon Vias (TSV)



ROC pads

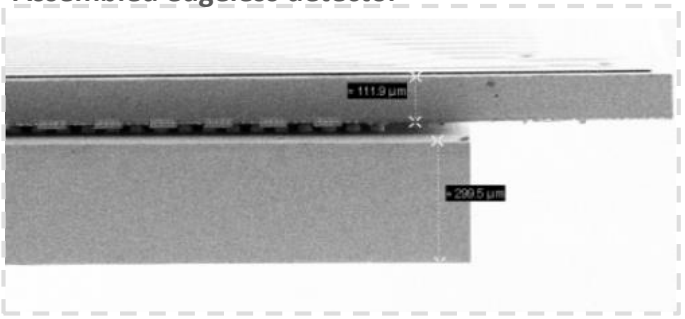


Sensor SnPb balls

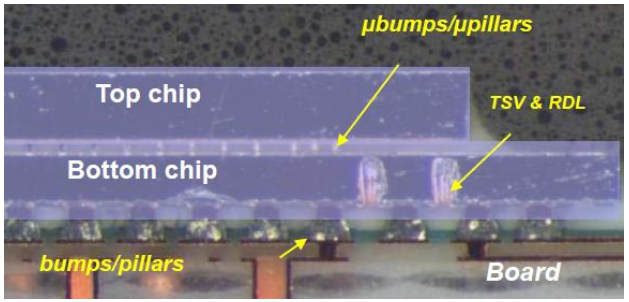


TSV: CEA-Leti, FR
Flip chip: Advacam, FI

Assembled edgeless detector



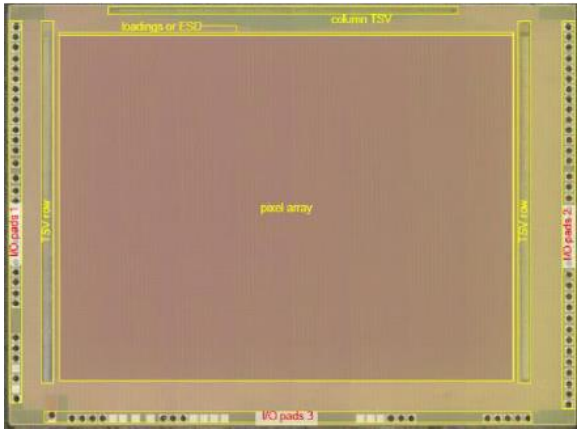
Detector mounted on board



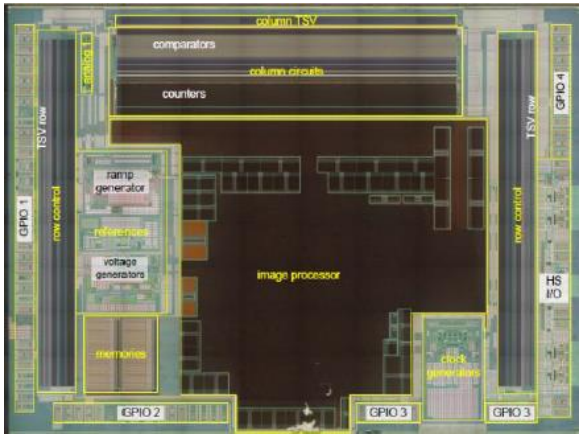
Still our pixel are “large”, i.e. we have no technology for such an integration with say 1 µm or few pixel pitch.

Sony stacked CMOS, mix different processes among the two layers.

90 nm node pixel array



65 nm node processing logic



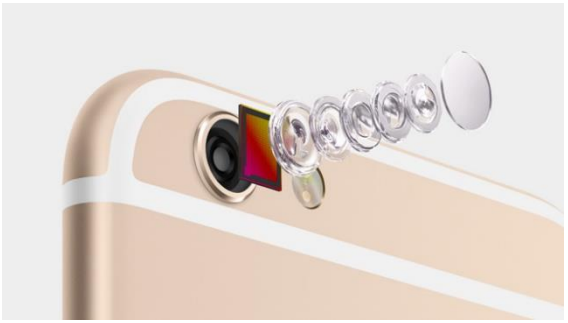
Radiation – practical solutions

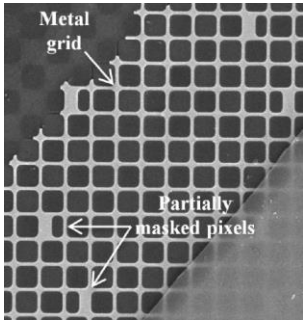
- Investigating and proving the effectiveness/reliability of advanced technological solutions is becoming more and more difficult for the scientific community.
- E.g., practical steps for using 65nm tech node in a 1 Grad environment (**R&D 53)**:
  - Consider changing the innermost layers at interval(s), i.e. design the whole system from scratch to support this possibility.
  - Research different libraries and provider implementation of a specific node within dedicated test chips test chips to find which available commercial solution offers the best performance.
  - Use small transistor only where really necessary!
  - Custom-design cells with larger W/L transistors, without compromising the overall footprint (R&D 53 working on this, up to double size transistor with similar or 30% more cell footprint).
  - Part of the design can stay healthy even with "degraded" transistors, e.g. low frequency circuits can operate successfully even if the transistors can supply only low Id currents after strong irradiation.

Maybe not fancy, but definitely feasible!

Radiatio – avoid entering a panda niche

With **1 Grad target** we must pay attention not becoming a panda within the ICs market, as there is no real market interest (at least so far) for any 1 Grad device. Differently from 20 years ago, we are NO MORE driving the technology, we must use what we can afford to buy.





- Design your system **top-down**!
- Use as much as possible available **commercial solutions** (hardware, IP cores, libraries, …)
- **Then look for what not available/affordable commercially, and focus on that**.
- Try to maximize the scientific value/outcome of what you develop in-house:
  - IP cores, libraries, entire chip could be useful and/or shared for/with other applications.
  - Look for potential commercial outcomes/partnership for what you are designing.
- Design/development team must grow in size to manage the challenge of newer technologies. We likely need to update our organization within the community.

Conclusions

High occupancy tracking

- The occupancy level (pileup) is not a concern in absolute term, what difficult is to implement a machine (chip) to handle it which can withstand the environment constraints.
- Tracking problem are, from the detector point of view, a particular subset of more general information theory problems. Computational power is affordable nowadays, be open when you try solving them!
- E.g., if you need very high density tracking, do not fear considering to embed a JPEG core inside your chip, and then focusing on how to make it working in a radiation environment.

Radiation hardness

- Radiation hardness is clearly one of the most "peculiar" constraint we have, and by itself justify many of our custom-made chips, **but is not the only one**!
- Like it or no, **the market will be the major drive on our choices and opportunities** in the microelectronic field (CCD were better, CMOS cheaper, who won the imagers battle?).
- Custom-made processes are nice, but unless the industry find them commercially viable, it will be unlikely to actually get them. **Adapt to use and/or modify what available**, or develop something with **different possible applications** in mind.