# RD53 pixel chips for the ATLAS and CMS Phase-2 upgrades at HL-LHC

Flavio Loddo - I.N.F.N. Bari

on behalf of the RD53 Collaboration

Outline

- Introduction to the RD53 Collaboration: mandate and deliverables
- Requirements
- Floorplan
- I/O interface
- Multi-chip data merging
- Serial powering
- Analog Front-end
- Pixel digital architecture
- Data flow
- Radiation hardening
- SEE strategy
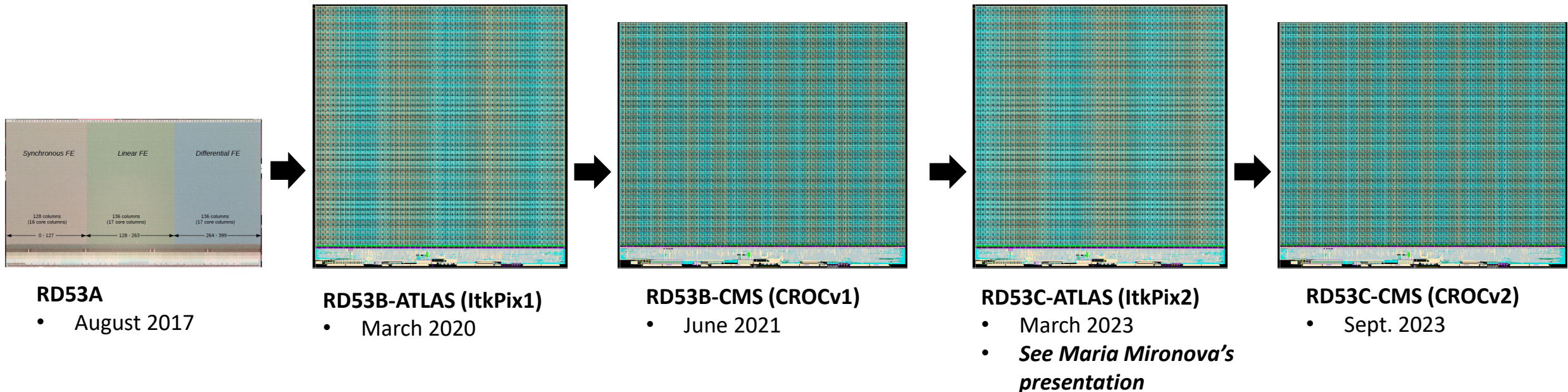- Verification plan
- Conclusions
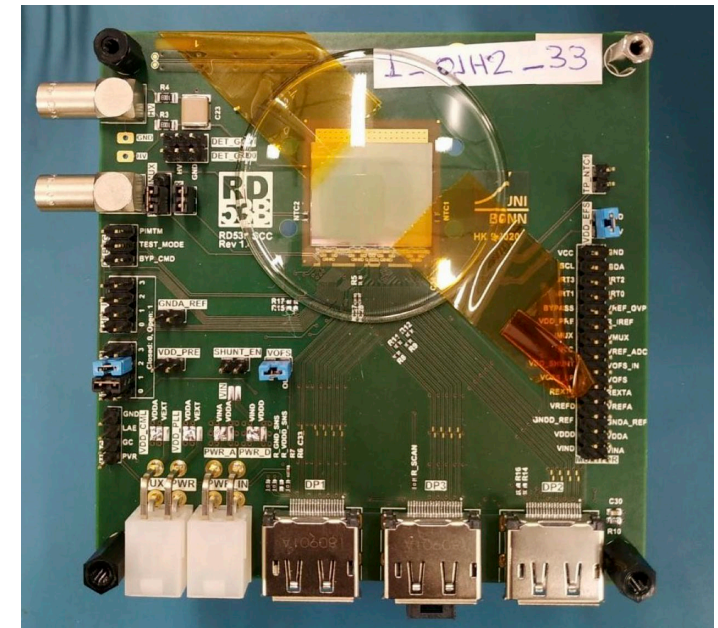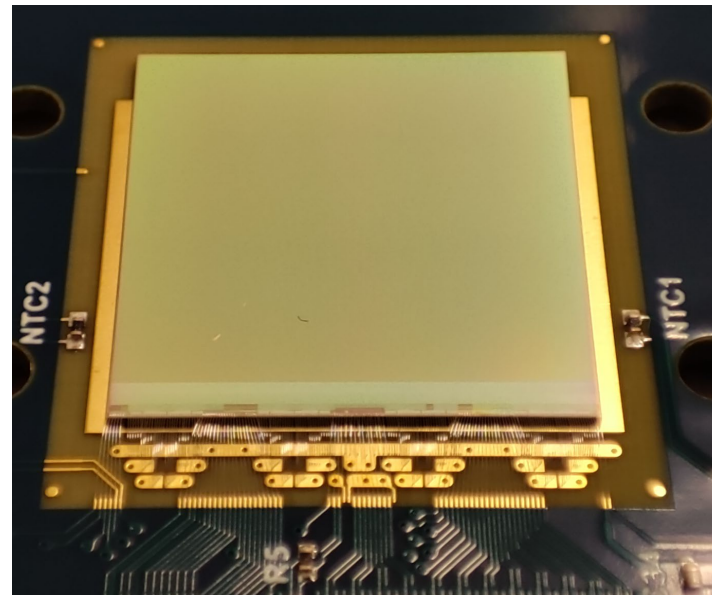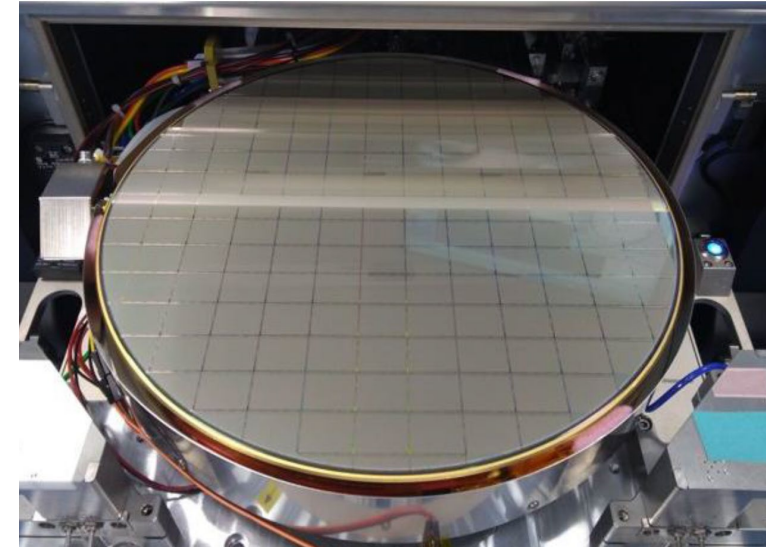
RD53 Collaboration

RD53 Collaboration is a joint effort from ATLAS and CMS Institutes to develop readout chips for the HL-LHC pixel detectors
24 institutes, started in 2013

1. Characterization of **TSMC 65nm CMOS** technology in radiation environment
2. Design of a **rad-hard IP library** (Analog front-ends, DACs, ADCs, CDR/PLL, high-speed serializers, RX/TX, ShuntLDO, …) qualified through a series of test chips
3. Design and characterization of **half-size pixel chip demonstrator (RD53A)** with design variations (3 Analog Front-Ends, 2 pixel readout architectures)
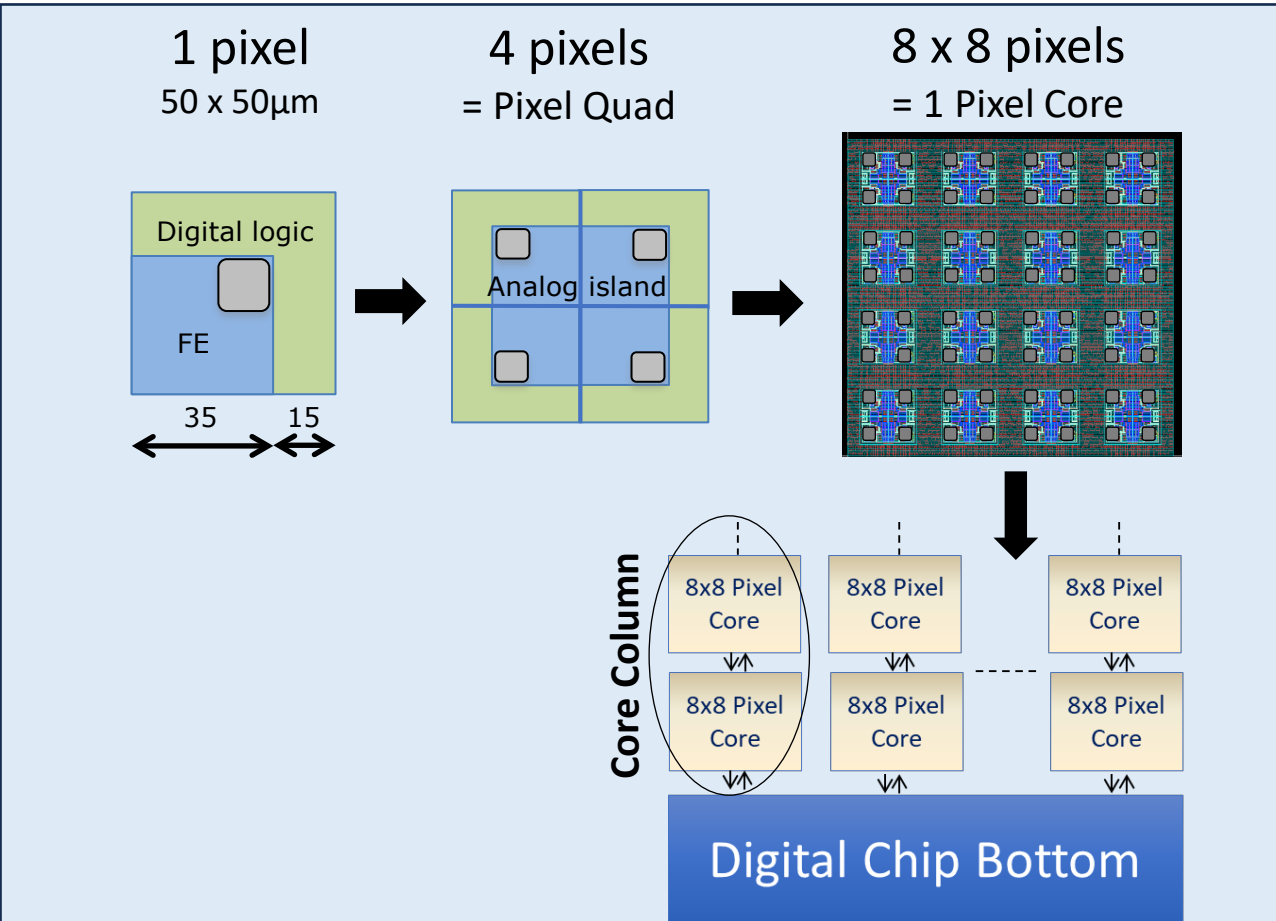4. Design of pre-production (**RD53B**) and production (**RD53C**) pixel readout chips
   ATLAS and CMS chips are two instances of the same common design, having different size and Analog Front-End, according to specific requirements of the experiments


RD53A
- August 2017

RD53B-ATLAS (ItkPix1)
- March 2020

RD53B-CMS (CROCv1)
- June 2021

RD53C-ATLAS (ItkPix2)
- March 2023
- *See Maria Mironova's presentation*

RD53C-CMS (CROCv2)
- Sept. 2023

Design requirements

Parameter	Value (CMS/ATLAS)
Technology	65 nm CMOS
Max. hit rate	3.5 GHz/cm²
Trigger rate	750 kHz / 1 MHz
Trigger latency	12.5 µs
Pixel size (chip)	50 x 50 µm²
Pixel size (sensor)	50 x 50 µm² or 25 x 100 µm²
Pixel array	432 x 336 pixels / 400 x 384 pixels
Chip dimensions	21.6 x 18.6 mm² / 20 x 21 mm²
Detector capacitance	< 100 fF (200fF for edge pixels)
Detector leakage	< 10 nA (20nA for edge pixels)
Min. threshold	< 1000 e-
Radiation tolerance	1 Grad over 10 years at -15°C
SEE tolerance	SEU rate, innermost: ~100Hz/chip
Power	< 1W/cm², Serial powering
Readout data rate	1-4 links @ 1.28Gbits/s = max 5.12 Gbits/s
Temperature range	-40°C ÷ 40°C

- RD53B Design manual and user guides: https://cds.cern.ch/record/2665301
- RD53B requirements: https://cds.cern.ch/record/2663161

Floorplan

1 pixel
50 x 50μm

4 pixels
= Pixel Quad

8 x 8 pixels
= 1 Pixel Core

Digital logic

FE

35   15

Analog island

Core Column

| 8x8 Pixel Core | 8x8 Pixel Core | 8x8 Pixel Core |
| 8x8 Pixel Core | 8x8 Pixel Core | 8x8 Pixel Core |

Digital Chip Bottom

Pixel array
- Digital logic synthesized for 8 x 8 pixels to form a **Pixel Core**
- All Cores are identical → efficient hierarchical implementation and verification
- **The number of Cores is a parameter of the common RTL netlist**
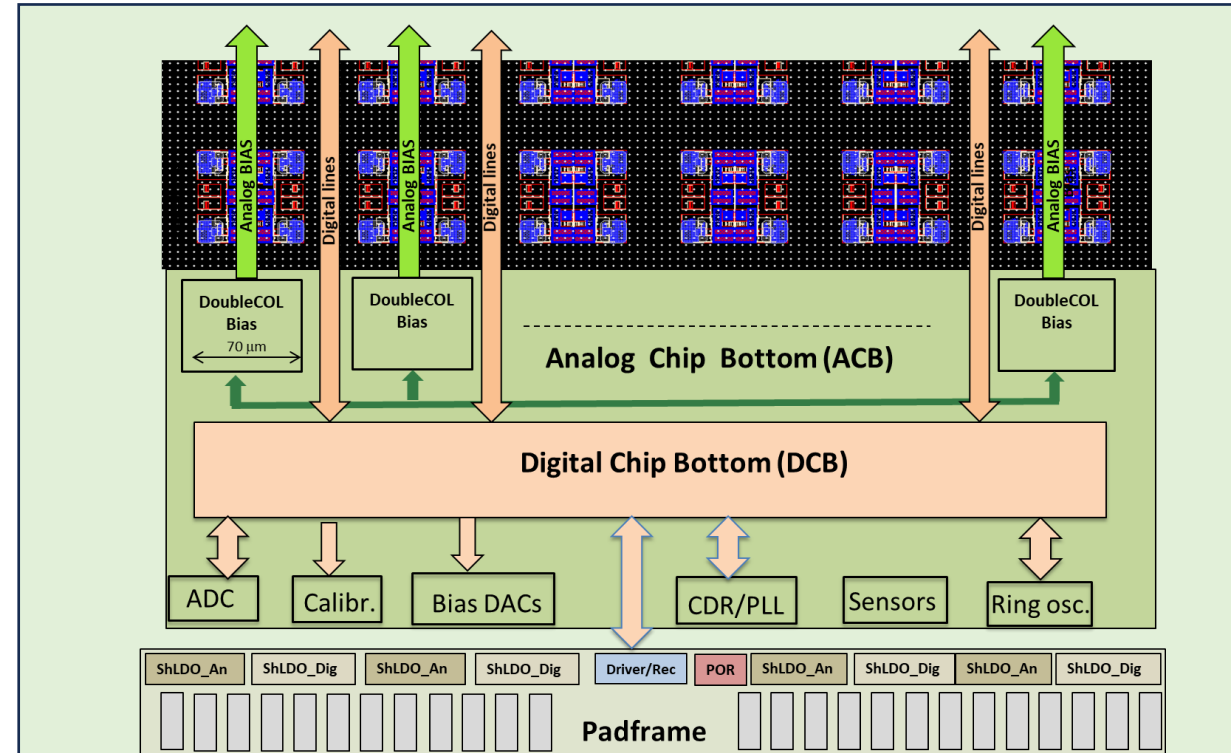
Chip periphery
- **Analog Chip Bottom (ACB)**: contains all analog and mixed/signals building block for Calibration, Bias, Monitoring and Clock/Data recovery
- **Digital Chip Bottom (DCB):** synthesized logic for communication to/from chip, readout and configuration
- **Padframe (common to ATLAS/CMS)**: complex macro containing all I/O blocks with ESD protections and distributed ShuntLDO regulator for serial powering

I/O interface
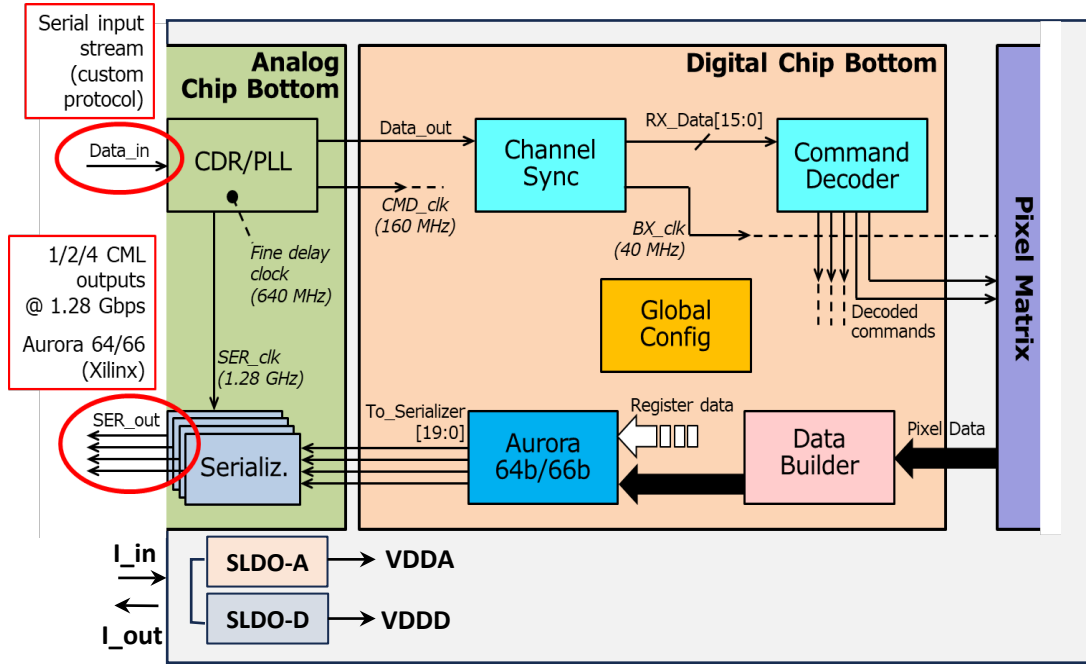
In short, extremely complex pixel readout ASIC designed to have a minimal I/O interface:

❑ **Input: command, control and timing**
  ➢ **One single 160 Mb/s differential serial link**, driving up to 15 chips (4 bit addressing + broadcast)
  ➢ CDR/PLL recovers Data and Clock
❑ **Output:**
  ➢ **Up to 4 x 1.28 Gb/s CML serial links** for data readout (hit data + service data) compatible with LpGBT and using Aurora 64b/66b encoding
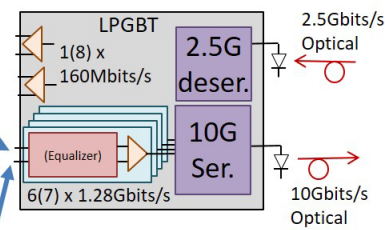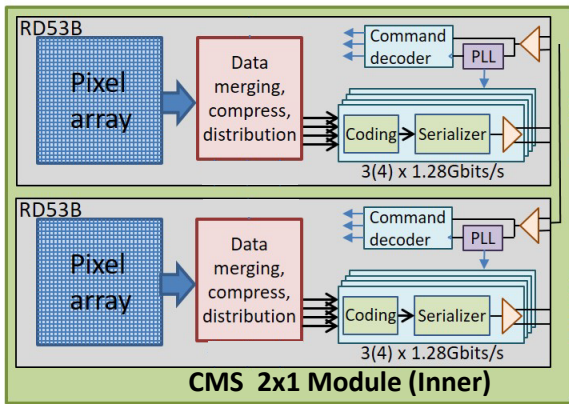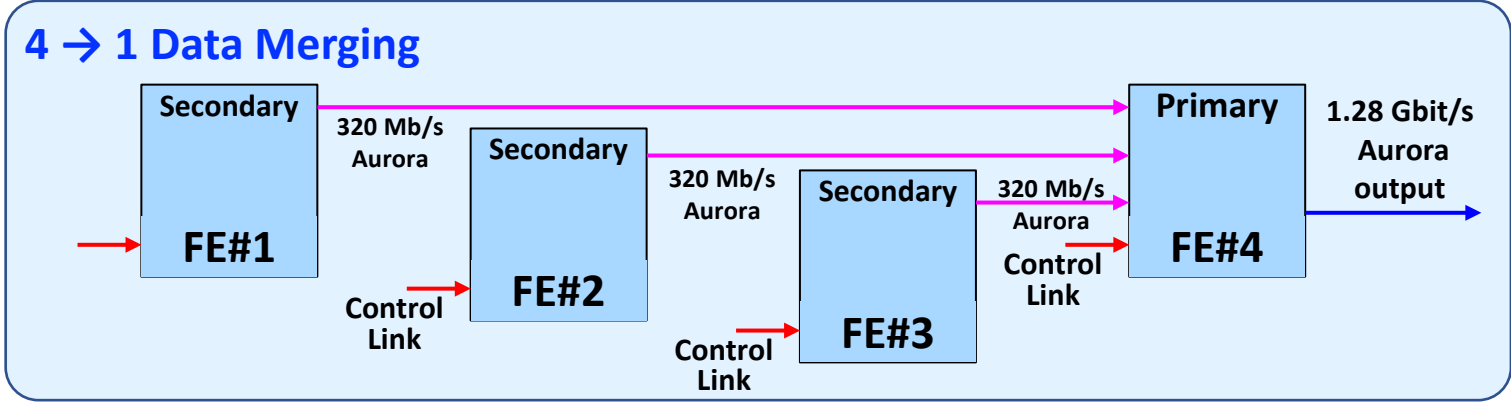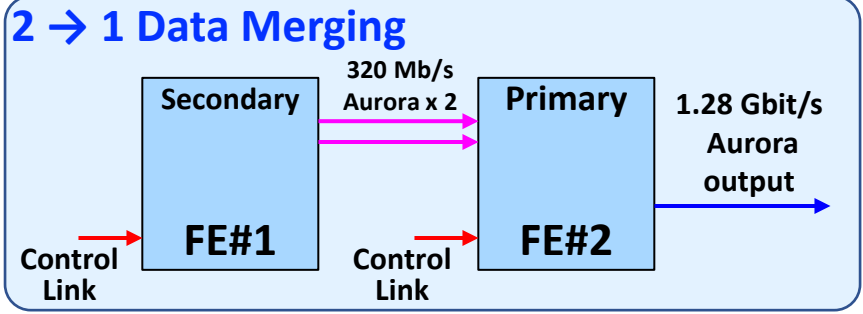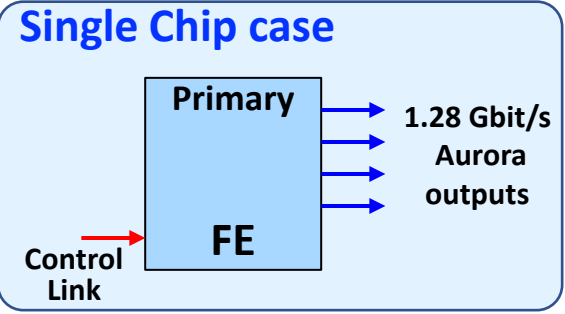❑ **Power:**
  ➢ Input/output currents for **serial powering**

Multi-chip data merging

Data merging is used to optimise the number of e-links sent off a module.

In the outer layers, one chip of the module can be configured as "primary" to aggregate serial data from one or more other "secondary" chips and merge them with its own output

Data from the secondary chip(s) is merged with data from the primary chip in a simple **round robin**

Single Chip case

Primary FE → 1.28 Gbit/s Aurora outputs
Control Link

2 → 1 Data Merging

Secondary FE#1 → 320 Mb/s Aurora x 2 → Primary FE#2 → 1.28 Gbit/s Aurora output
Control Link / Control Link

4 → 1 Data Merging

Secondary FE#1 → 320 Mb/s Aurora
Secondary FE#2 → 320 Mb/s Aurora
Secondary FE#3 → 320 Mb/s Aurora → Primary FE#4 → 1.28 Gbit/s Aurora output
Control Link

1 pixel module for 1 LPGBT
3 Readout links per chip
1 Control link per module
2 pixel chips per module

CMS 2x1 Module (Inner)
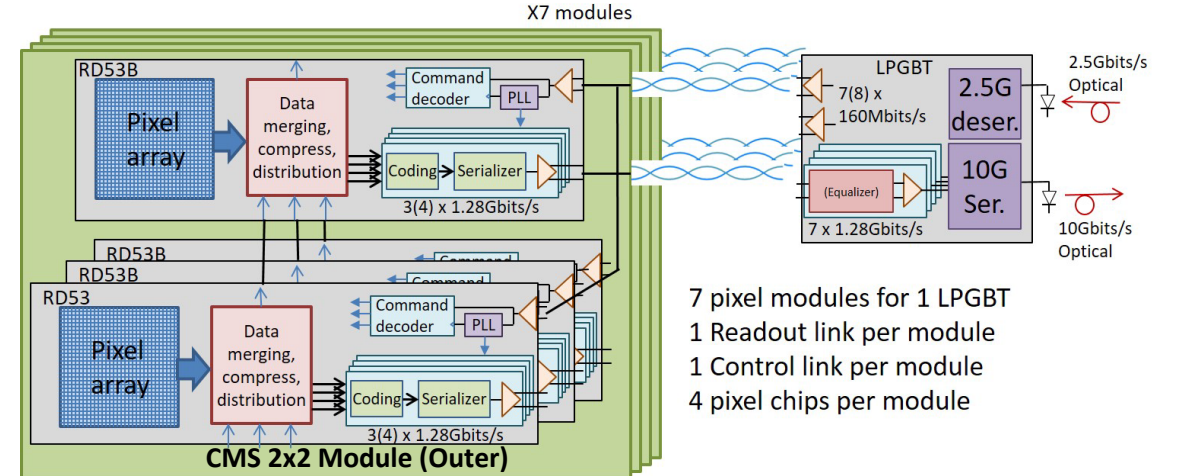
7 pixel modules for 1 LPGBT
1 Readout link per module
1 Control link per module
4 pixel chips per module
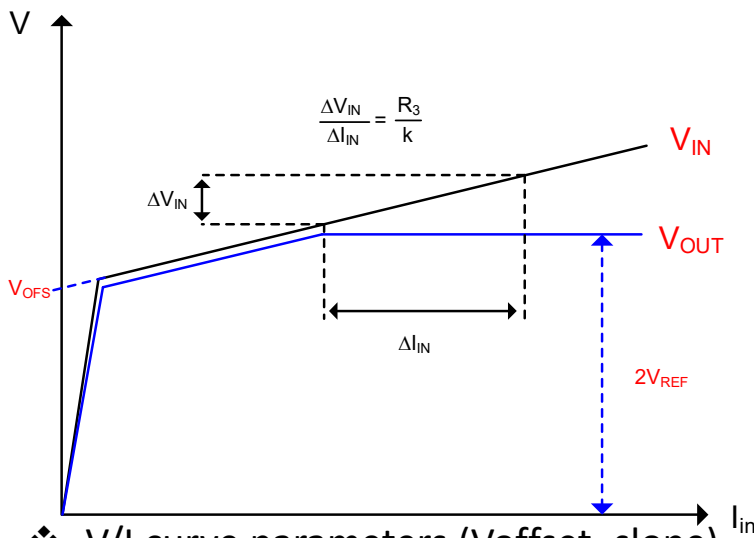
CMS 2x2 Module (Outer)

ShuntLDO for Serial powering

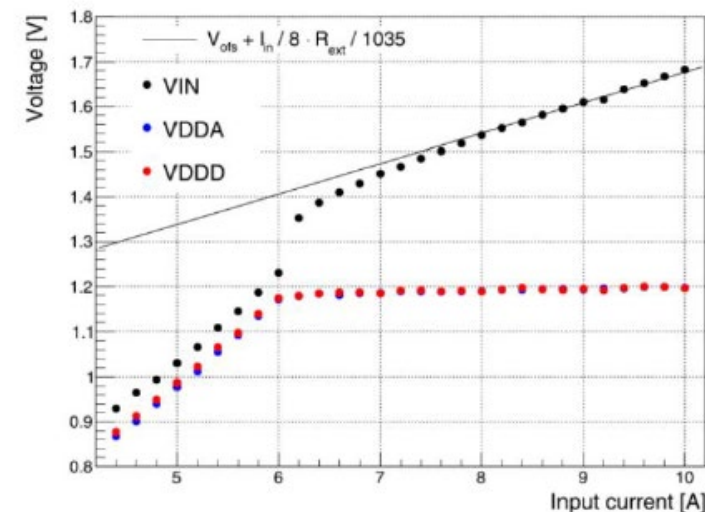Both ATLAS and CMS will adopt the innovative **serial-powering scheme:**

- Based on **ShuntLDO** regulators in the readout chips (1 x Analog, 1 x Digital)
- **Constant input current** shared between chips (2÷4) on the same module
- Modules are in serial chains: "recycle" current from one module to another
- **Input current** dimensioned to satisfy the highest load, with ~20-25% headroom for stable operation to be absorbed by the **Shunt** device
- In case of chip failure, its current can be absorbed by the other chips of the module
- Not sensitive to voltage drops (low mass & compact power routing)
- On-chip regulated supply voltages, low noise
- Radiation hardness (1 Grad)
- Current sharing defined by the Voffset and Slope, **all set by external resistors (no SEU)**

4(2) chips per module

Up to 14 modules per chain
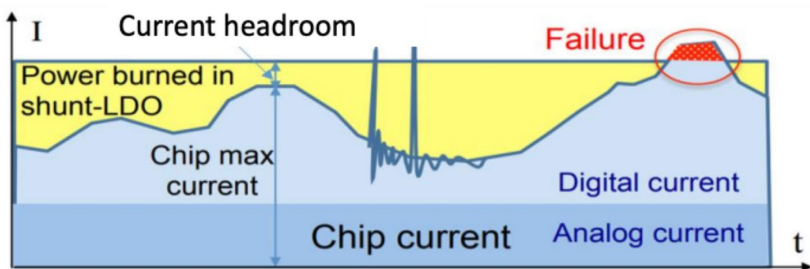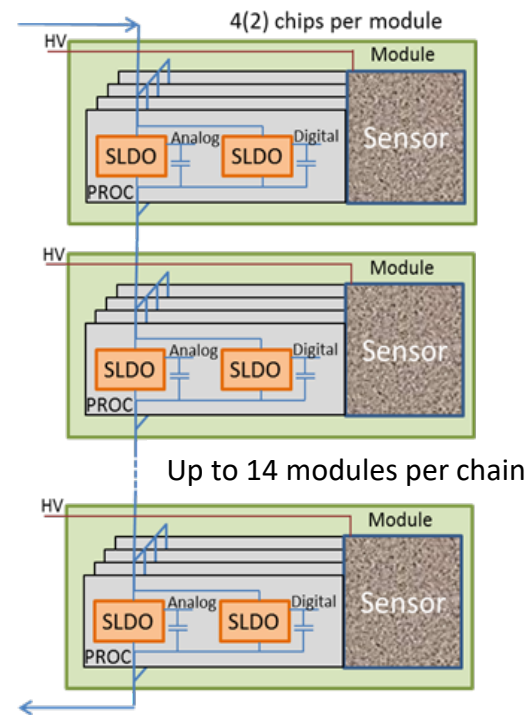
Example current consumption of one readout chip

$$\frac{\Delta V_{IN}}{\Delta I_{IN}} = \frac{R_3}{k}$$

Protections:
- *Over-voltage protection*: V_{IN} clamped to 2 V
- *Under-shunt protection*: V_{OUT} decreased in case shunt current goes below a certain threshold (due to excess load current)

❖ V/I curve parameters (Voffset, slope) defined by external resistors
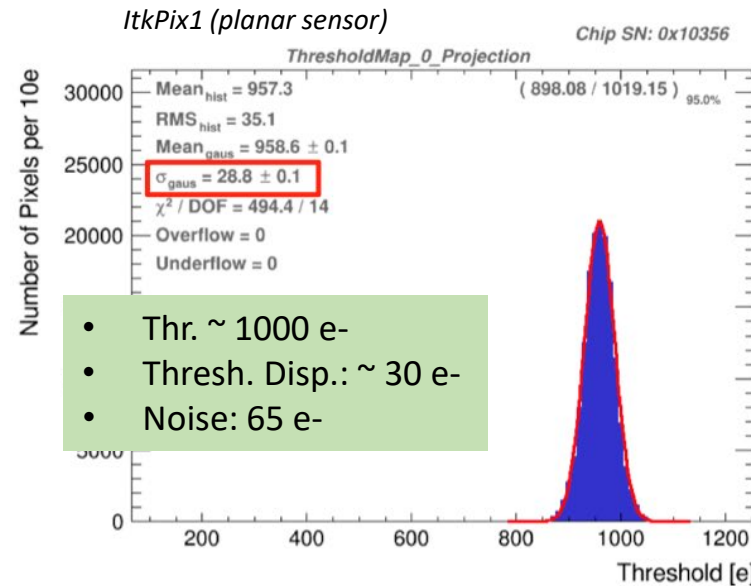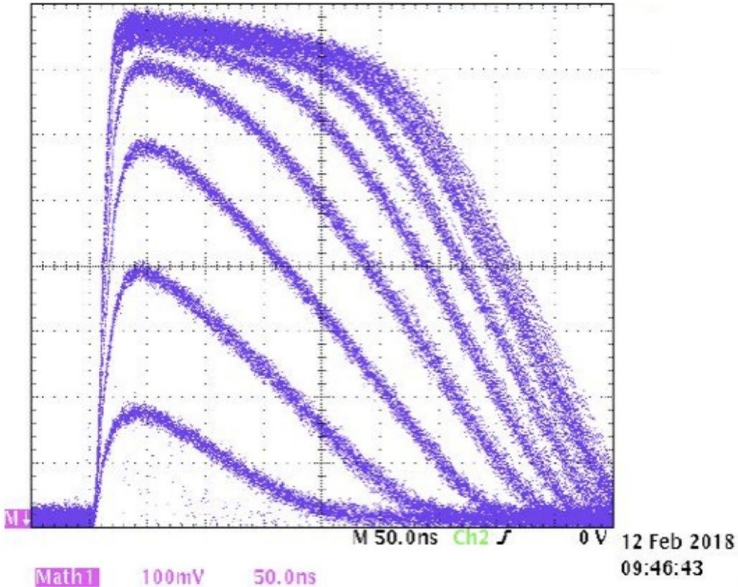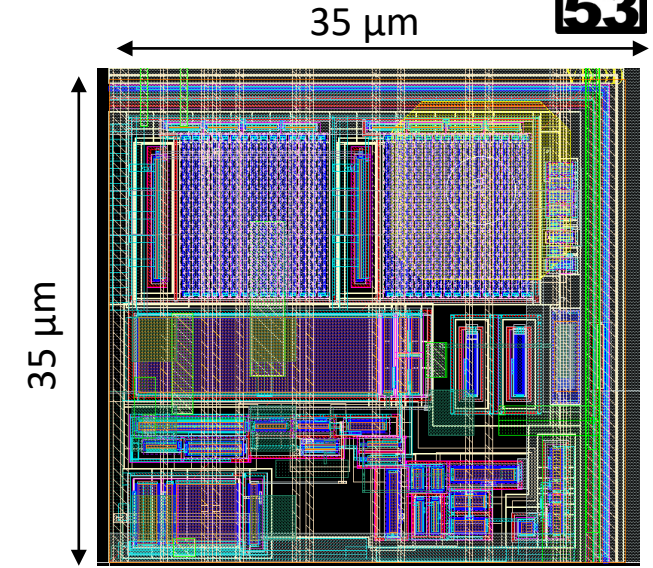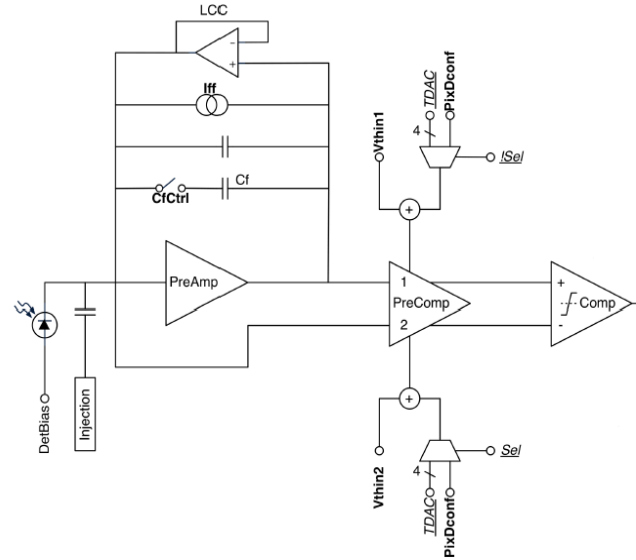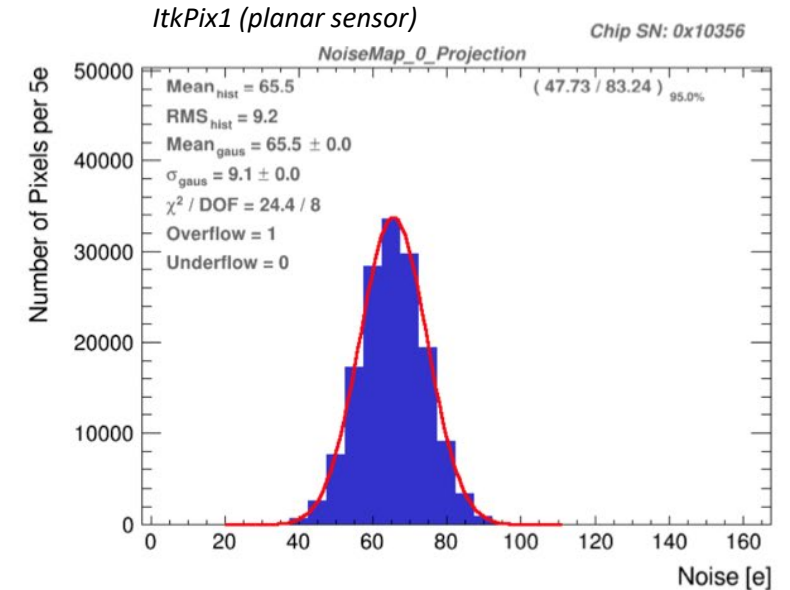❖ V_{OUT} tunable by chip configuration

$V_{ofs} + I_{in} / 8 \cdot R_{ext} / 1035$

- VIN
- VDDA
- VDDD

Differential Front-End (ATLAS)

- Continuous reset integrator first stage with DC-coupled pre-comparator stage

- Leakage current compensation circuit (LCC) for additional optional feedback in case of $I_{leakage}$ > 2 nA

- Two-stage open loop, fully differential input comparator

❖ 10-bit DAC for global threshold

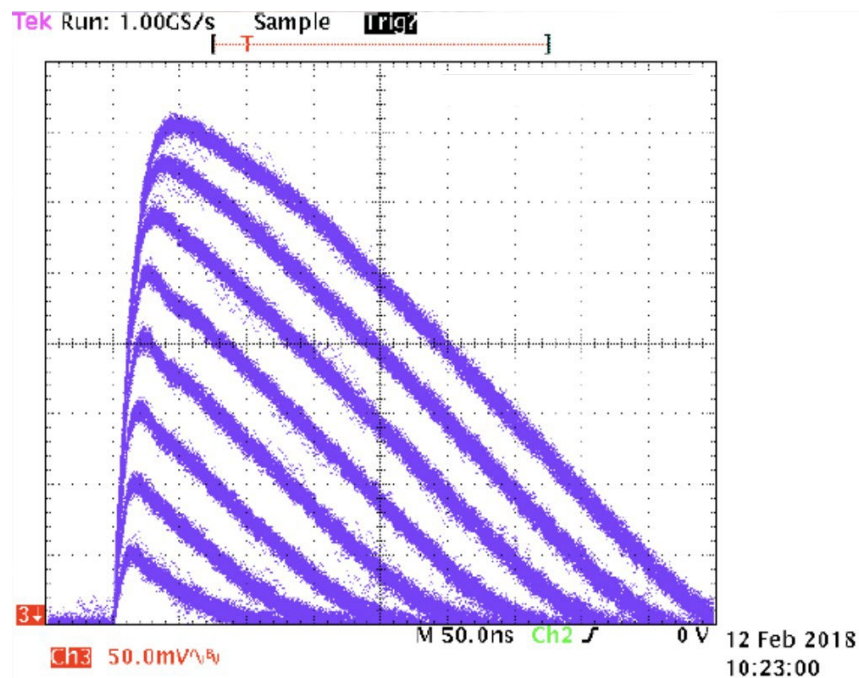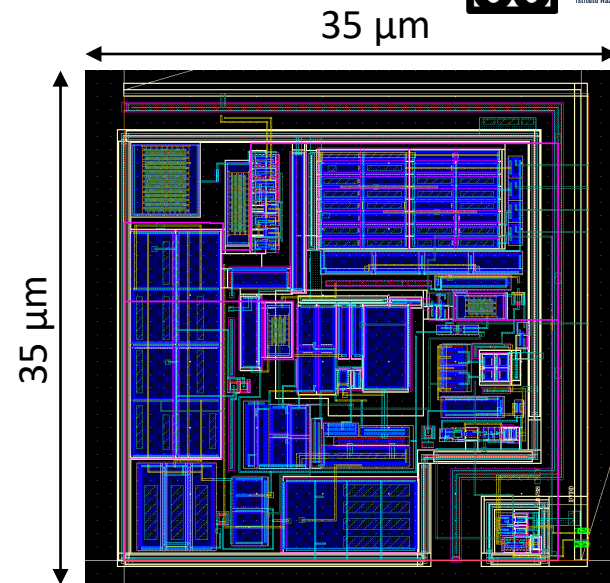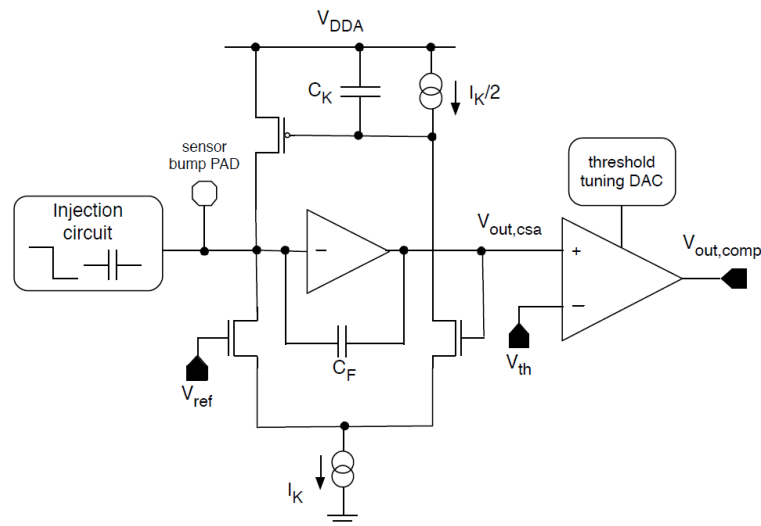❖ 4+1 bit local trimming DAC (TDAC) for threshold tuning

35 μm

35 μm

ItkPix1 (planar sensor)

ThresholdMap_0_Projection — Chip SN: 0x10356

Mean hist = 957.3
RMS hist = 35.1
Mean gaus = 958.6 ± 0.1
σ gaus = 28.8 ± 0.1
χ² / DOF = 494.4 / 14
Overflow = 0
Underflow = 0
(898.08 / 1019.15) 95.0%

- Thr. ~ 1000 e-
- Thresh. Disp.: ~ 30 e-
- Noise: 65 e-

ItkPix1 (planar sensor)

NoiseMap_0_Projection — Chip SN: 0x10356

Mean hist = 65.5
RMS hist = 9.2
Mean gaus = 65.5 ± 0.0
σ gaus = 9.1 ± 0.0
χ² / DOF = 24.4 / 8
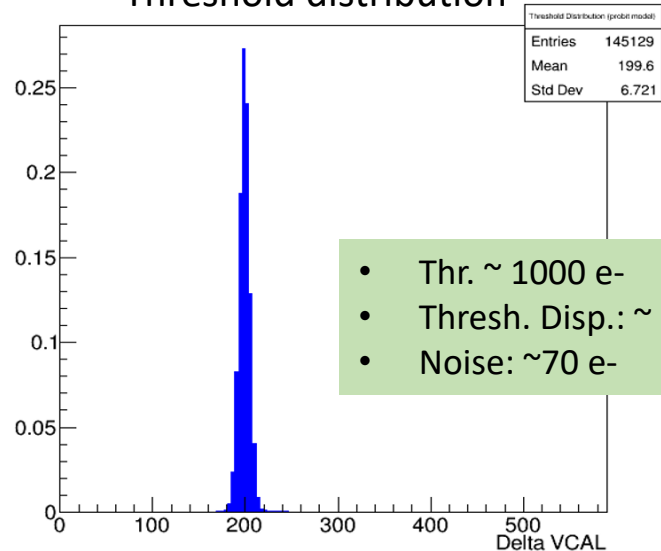Overflow = 1
Underflow = 0
(47.73 / 83.24) 95.0%

Linear Front-End (CMS)

- Charge sensitive amplifier

- Krummenacher feedback for return to baseline and leakage current compensation

- Comparator

❖ 10-bit DAC for global threshold
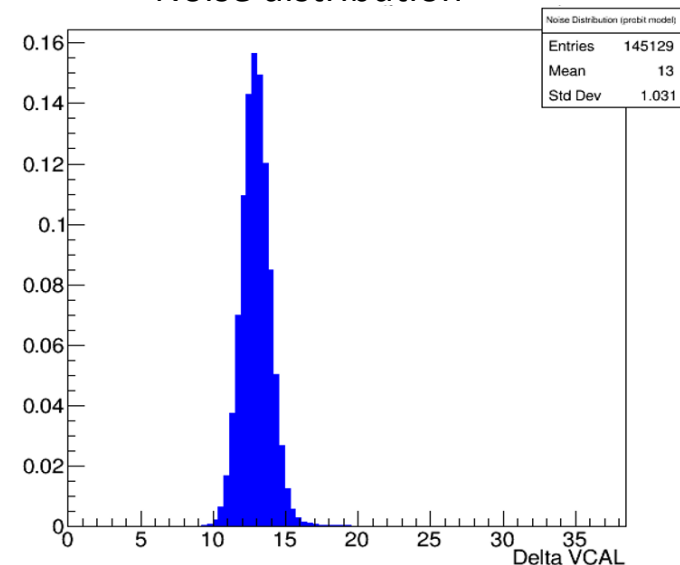
❖ 5-bit local trimming DAC (TDAC) for threshold tuning


35 μm

35 μm

Threshold distribution

Threshold Distribution (probit model)	
Entries	145129
Mean	199.6
Std Dev	6.721

Noise distribution

Noise Distribution (probit model)	
Entries	145129
Mean	13
Std Dev	1.031

- Thr. ~ 1000 e-
- Thresh. Disp.: ~ 50 e-
- Noise: ~70 e-

Tek Run: 1.00GS/s   Sample   Trig?

M 50.0ns Ch2 ∫   0 V   12 Feb 2018
Ch3 50.0mV   10:23:00

Pixel digital architecture

- Hits are stored as **Time-over-Threshold**, associated to a time stamp
- 6-bit ToT counter, but only 4 bits are stored and read-out
  - ➤ Selectable counting clock: 40 MHz or 80 MHz (dual-edge)
  - ➤ Selectable ToT dual-slope 6-to-4 mapping for charge compression
- **Each pixel has 8x4-bit ToT memories**
- The time stamp memory is shared among 4 pixels of the same 4x1 **Pixel Region**
- Trigger-matching performed with programmable 9-bit trigger latency (max. 12.5 µs)
- Hit-OR network for:
  - Programmable self-triggering
  - 11-bit Precision ToT (PTOT) + 5-bit ToA counters at 640 MHz in the chip periphery for precise AFE measurements and sensors characterization
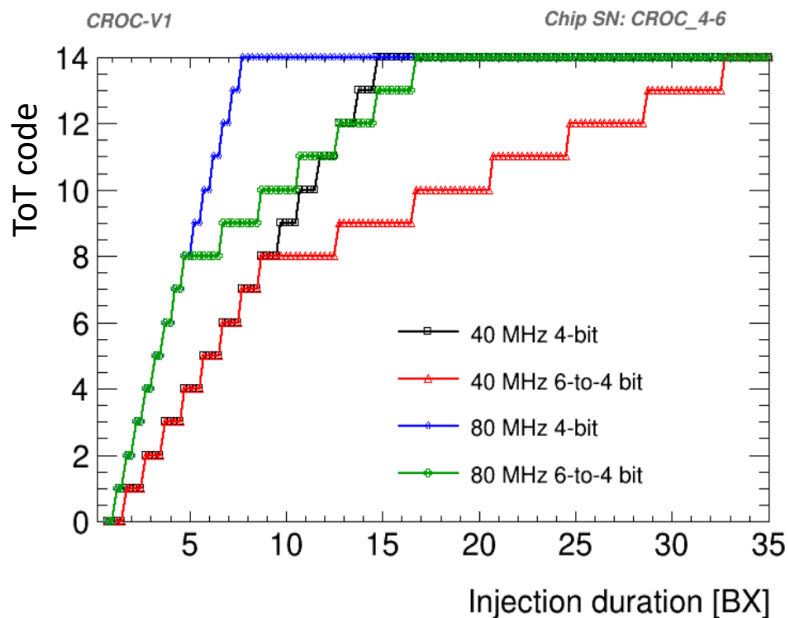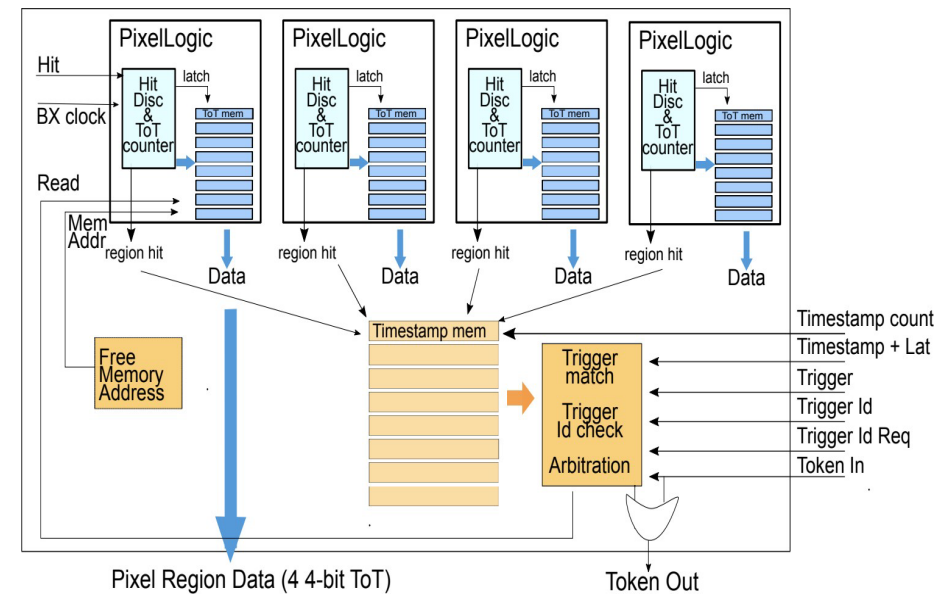
| Output 4-bit code	True ToT bin (low edge) [BX]			
	40 MHz speed		80 MHz speed	
	4-bit (DEF)	6-to-4 bit	4-bit	6-to-4 bit
0	0	0	0	0
1	1	1	0.5	0.5
2	2	2	1	1
3	3	3	1.5	1.5
4	4	4	2	2
5	5	5	2.5	2.5
6	6	6	3	3
7	7	7	3.5	3.5
8	8	8	4	4
9	9	11	4.5	5.5
10	10	15	5	7.5
11	11	19	5.5	9.5
12	12	23	6	11.5
13	13	27	6.5	13.5
14	≥14	≥31	≥7	≥15.5

Digital data flow

- **Token based readout** of hits **in parallel** for each Core Column as soon as a valid Trigger is received

- Data are stored in a FIFO at the end of each column and compressed using **Binary Tree Encoder** (up to factor of 2)

- **Multiple levels** of data processing, event building, data buffering and formatting

- Registers are read on command and periodically **("Service data")** interleaved with Physics data in a set ratio (default: 1 in 50)

- **Aurora frames are built** and sent to the high speed serializers
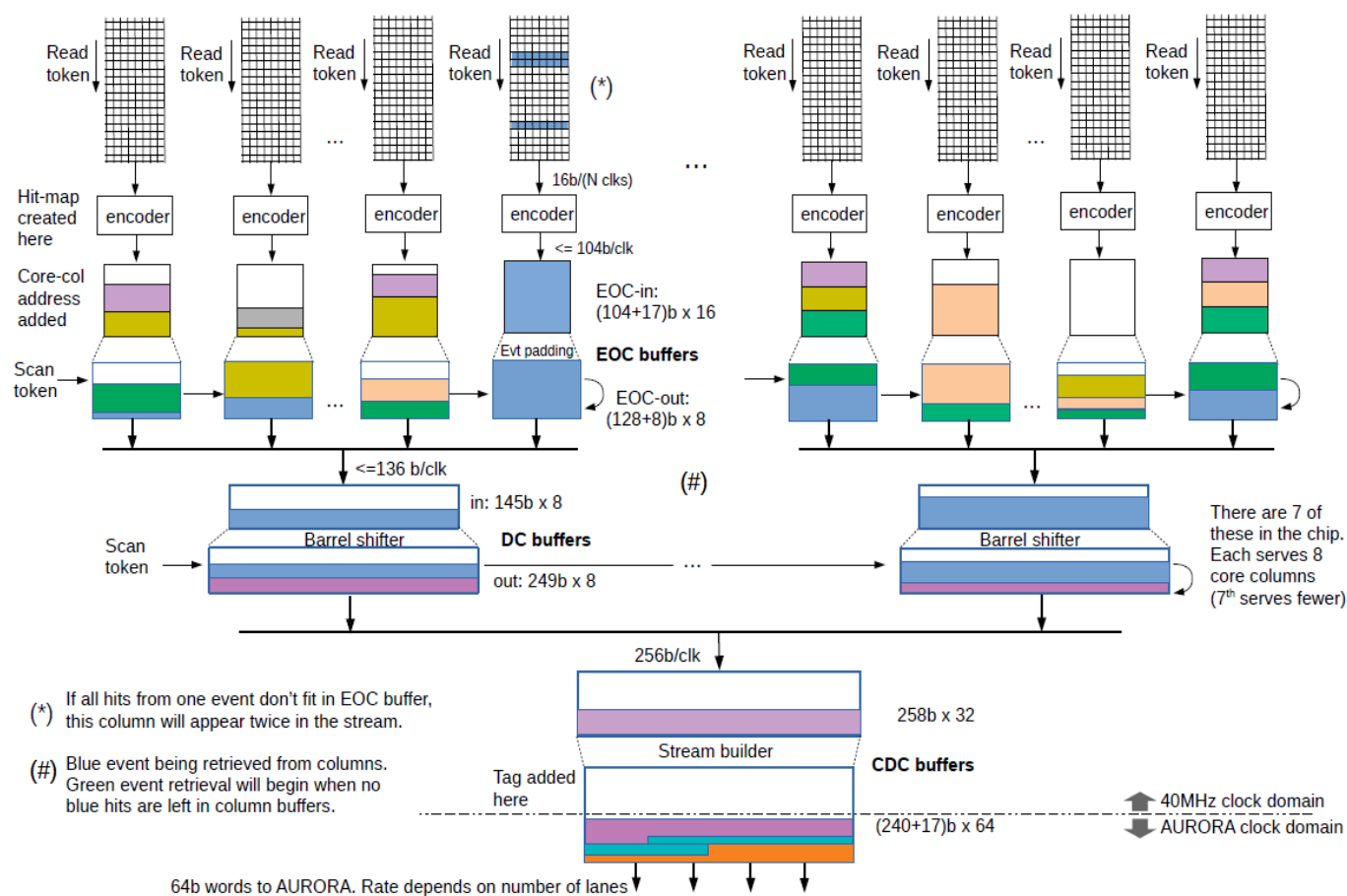
Radiation hardness

- Extensive TID X-ray tests on RD53A/B and small prototypes to qualify IPs, analog FEs and digital standard cells
- **The TID damage concerns essentially the digital design because of the small area devices (min L, small W)** → **more severe for "strength 0"**
- From measurements on Ring Oscillators: **gate delay degradation larger at low dose-rate (LDR): more severe for "strength 0"**
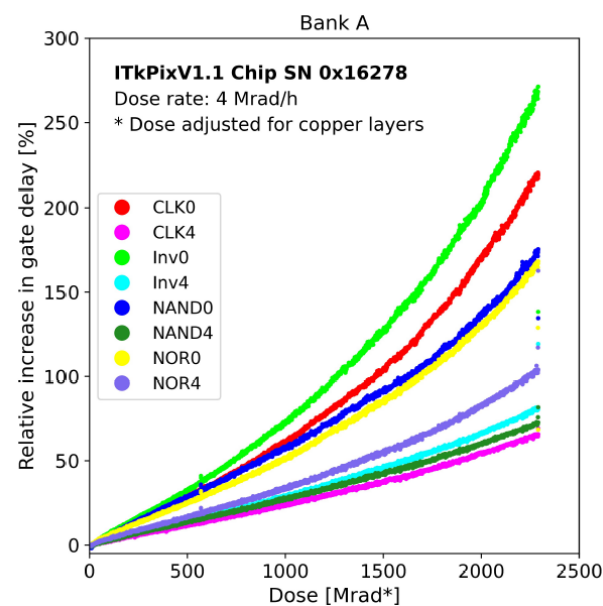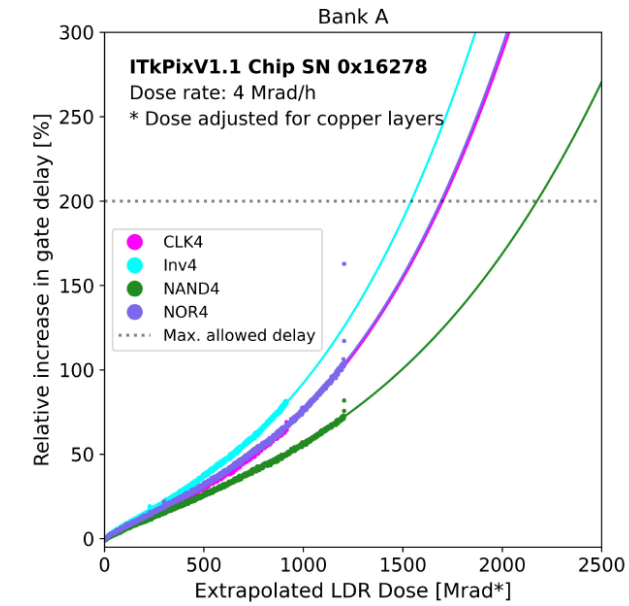
Delay change at 5 Mrad/h and 25 krad/h
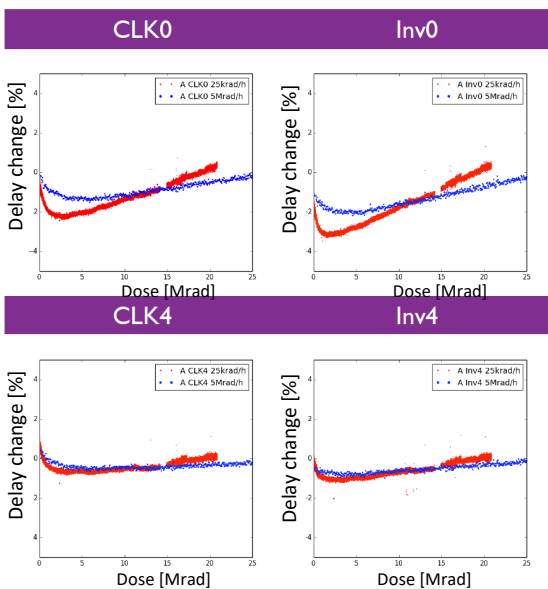
Ratios of damage at low and high dose rate

Gate	LDR/HDR damage
CLK 0	4.8
CLK 4	2.4
Inv 0	4.6
Inv 4	2.5
NAND 0	3.2
NAND 4	1.7
NOR 0	2.7
NOR 4	1.8

Gate delay at HDR X-ray to 2.3 Grad

Estimated behaviour at LDR of strength 4 gates

- **No minimum size digital cells (strength 0)**
- **Irradiation corner models and extreme corners** from the foundry used to predict the TID effect and design to guarantee good timing for the digital design
- **Tests done at low temperature (-20°C) and high dose-rate, extrapolating the behaviour using the LDR/HDR correction factor**, give confidence that RD53 chips will meet TID specifications (at least 1 Grad) if **operated at cold temperature**.
- Irradiated chips must be cooled while under power, **since room or high temperature annealing is detrimental**

Single Event Effect (SEE) protection

The expected rate per chip of Single Event Upsets (SEUs) is ~ 100 Hz in the innermost layer

Target behaviour:

- Stable operation: biasing and the Clock and Data Recovery (CDR) as critical for the overall chip functionality should be very SET robust

- No power cycling: need to recover without power-cycling (detrimental for a serially powered system)

- Occasional hit/event is allowed to be lost

Principle: protect with triplication vital circuitry (configuration registers, state machines, memory pointers) as far as feasible within space limits (but cannot protect everything!)

Note: at these SEE densities, no protection is perfect, so the command protocol includes:

- a **CLEAR** command for fast recovery from residual SEE or operational issues

- provision for **continuous reprogramming** of global registers and pixel configuration

Design strategy:

- **Intensive SEE simulations** of both digital logic (RTL, Gate-level) and critical analog IPs (bandgap, drivers, PLL, biasing blocks)

- **Test campaigns** with Heavy Ions, protons, laser

  - ➤ **Both simulations and tests are needed to identify sensitive nodes of IPs and critical digital circuits** needing improvements for the production chips

  - ➤ As an example, a critical fault condition was found with SEU simulations that could never been found with testing, because it requires high trigger rate, high hit rate, and high SEU rate all at once

Single Event Effect (SEE) protection

Pixel Configuration

- 8 bits → 1.28 Mbit/chip
- No space to protect all bits
- 3 less critical bits unprotected
- 5 more relevant bits protected with TMR without self-correction
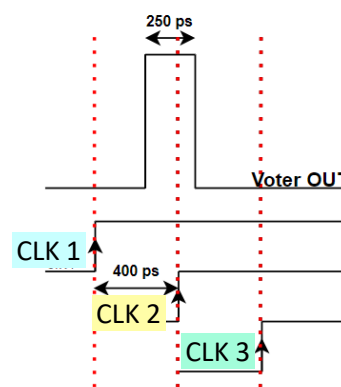 - ➢ **With protons: 100 x more SEE tolerant than unprotected**
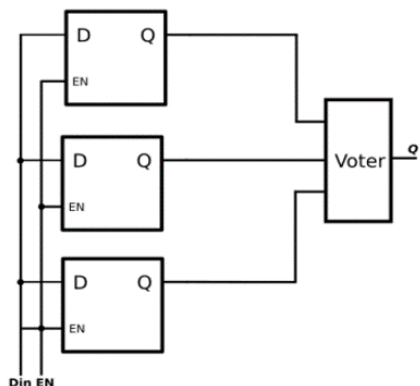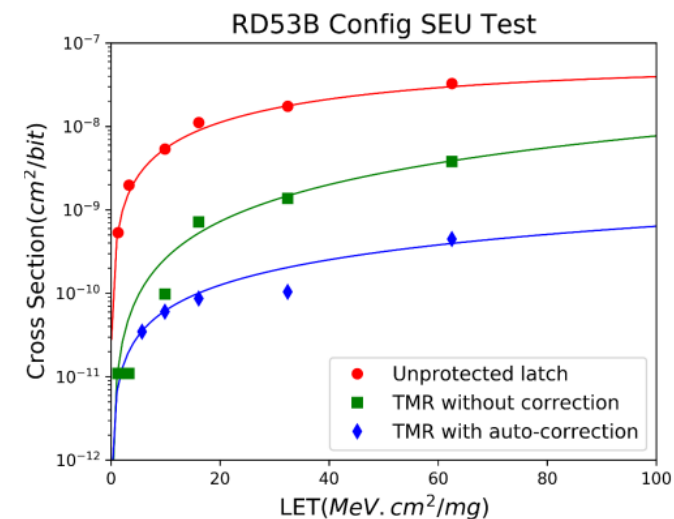

Global Configuration and state machines

- ~90 kbit/chip
- TMR with self-correction and triplicated clock with skew ($\Delta T \sim 400ps$) for **Single Event Transient** filtering
  - ➢ **With protons: 400 x more SEE tolerant than unprotected**
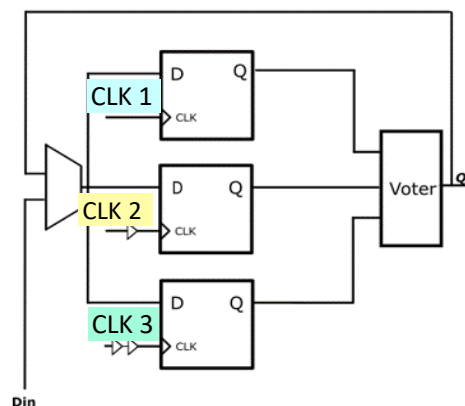
SET filtering



SEU cross-section comparison from heavy-ion testing:

RD53 verification framework

Very complex digital architecture:

- **U**niversal **V**erification **M**ethodology (IEEE Std. 1800.2-2020)
- industry-like approach, well-supported by tool vendors

RD53 verification framework

- verify that chip outputs match outputs predicted by the framework, for randomised configurations and inputs
- Initial framework version used for architecture studies and RD53A-RD53B verification
- Improved version implemented for final functional and SEU verification of production chips
- <u>Metric-driven verification</u>: **simulate until the desired coverage is reached (goal is 100%)**
  - **Functional coverage**: the features of the chip
  - **Code coverage**: blocks of exercised, toggling of signals
- Use **constrained randomization** to test all feasible combinations of configuration and inputs plus some directed tests for specific block verification
- Regressions with **> 1700 tests** to reach coverage goals
- Also used for SEU injection simulations

Conclusions
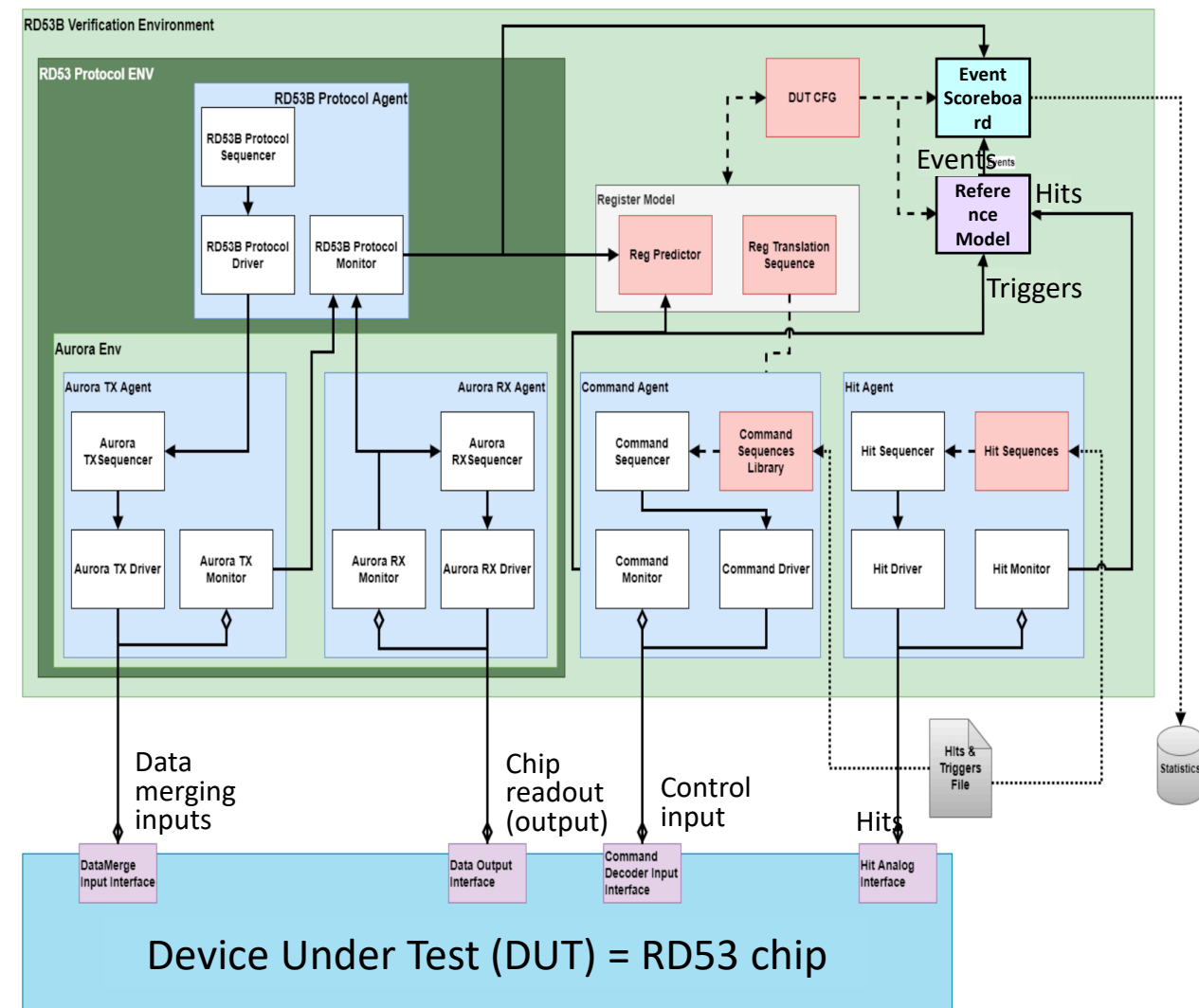
- The readout chips for the ATLAS and CMS HL-LHC pixel detectors are being developed by the **RD53 Collaboration** on 65 nm CMOS technology since 2013

- **Collaborative design** across ~24 institutes around the world

- The first half-size prototype **RD53A** has been crucial to initially explore design variations and to understand how to cope with the complexity of implementing such large and complex pixel chips operating in such harsh environment
  - ➢ RD53A was extensively used by the two experiments to characterize sensors

- The final chips are two different "instances" of a common design having different sizes and different Analog Front-End

- Full-size **pre-production** chips **RD53B-ATLAS (ITkPix1)** and **RD53B-CMS (CROCv1)** implementing all requested functionalities requested have been produced in 2020 and 2021

- **Verification is crucial in all its different aspects**: functional, SEE, analog and M/S, power

- **The ATLAS final production chip ITkPixV2 has been recently received and is currently under test**
- **The submission of the final CMS chip CROCv2 is estimated to take place in few weeks**

Thank you for your attention!

Backup slides

Analog and M/S blocks summary

| Block	Description
Analogue front end	The ATLAS chip versions use a differential front end. The CMS chip versions use a linear front end.
Shunt LDO	Enables start-up and serial powering. Constant input current shared between chips, modules on serial chains. 1 LDO for digital power, 1 LDO for analogue power.
Clock & Data Recovery (CDR)/PLL	Recovers a 160MHz clock and command/trigger stream. The PLL generates internal clocks: 160 MHz, 64 MHz, 640 MHz and 1.28 GHz.
Bias circuit	Provides biases to the pixel array. Based on bandgap references.
Calibration circuit	Injects hits into the pixel array, to calibrate its response.
Monitoring block	Digitises analogue quantities using a voltage mux, current mux and 12-bit ADC
Temperature and Radiation sensors	Temperature sensors: polysilicon resistors. Radiation sensors: based on PMOS devices with a linear variation in voltage in the dose range 10 - 1000 Mrad.
LVDS pads/drivers	Pads and drivers for differential inputs/outputs

Bias circuit

- BIAS network is based on Bandgap reference circuits, to provide a reference voltage/current with low sensitivity to temperature variations
- Tuning by means of 4 wire-bond trimming pads (<u>no risk of SEU bit flips</u>), whose optimal value is found during **wafer probing**
- The tuned current I_{ref} is replicated and used as reference to 23 Digital-to-Analog converters to bias the analog Front-end, the CDR and other IPs
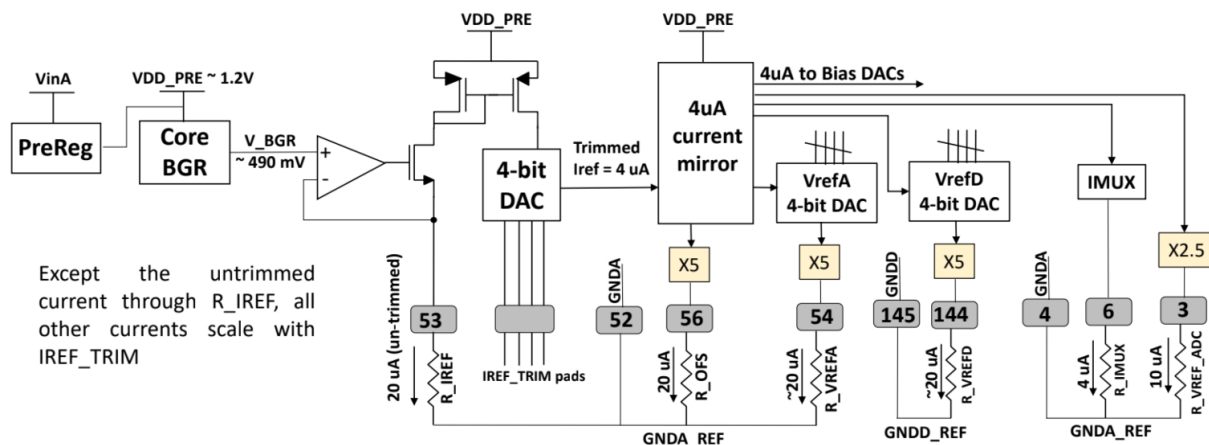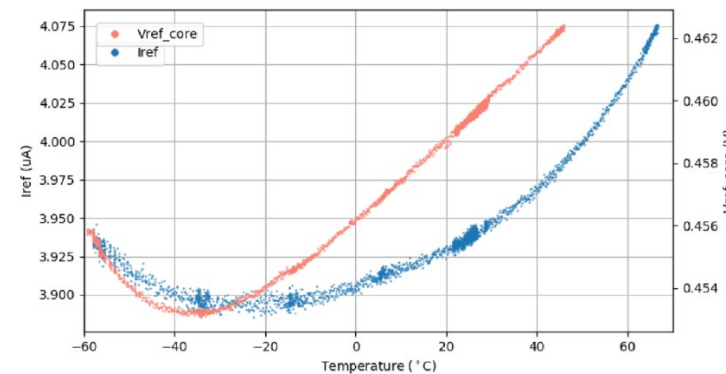
4% difference over 120°C temp. range
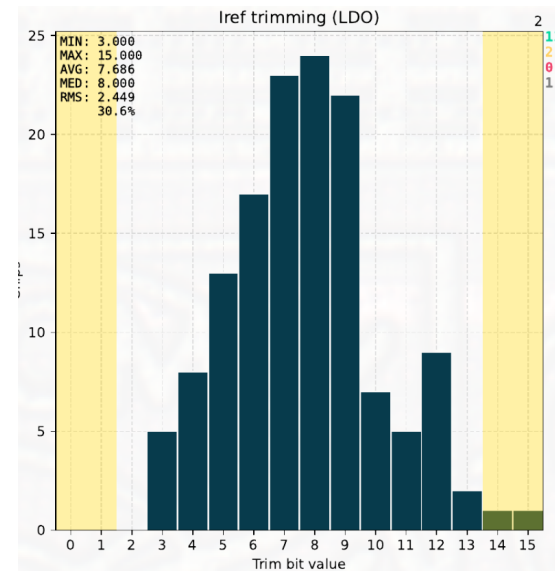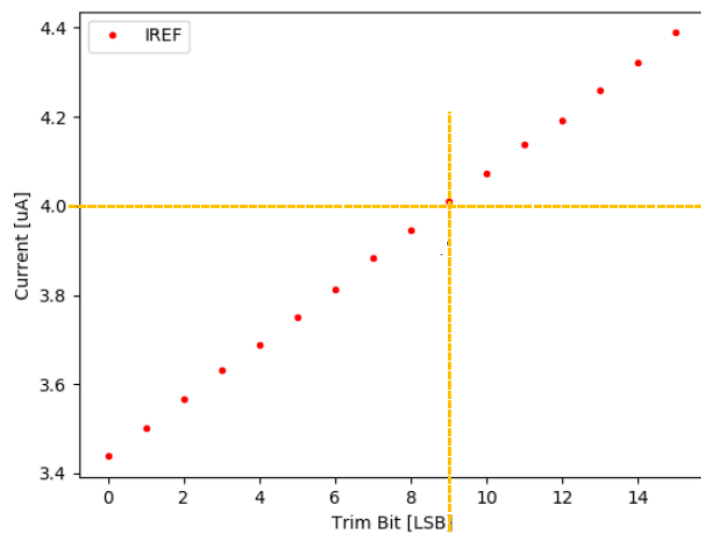
Calibration circuit

- **Each pixel** is equipped with a calibration injection circuit for test and calibration purpose
- The analog injection uses two distributed voltages, provided by two **12-bit voltage DACs**, to generate a precise voltage step fed to an injection capacitor
- Two selectable ranges
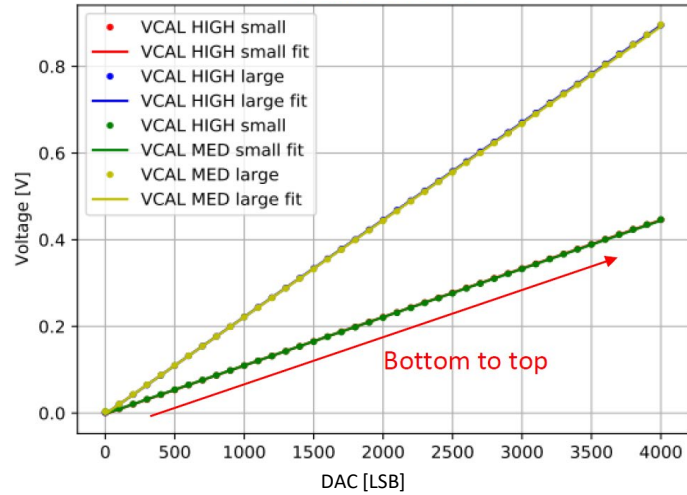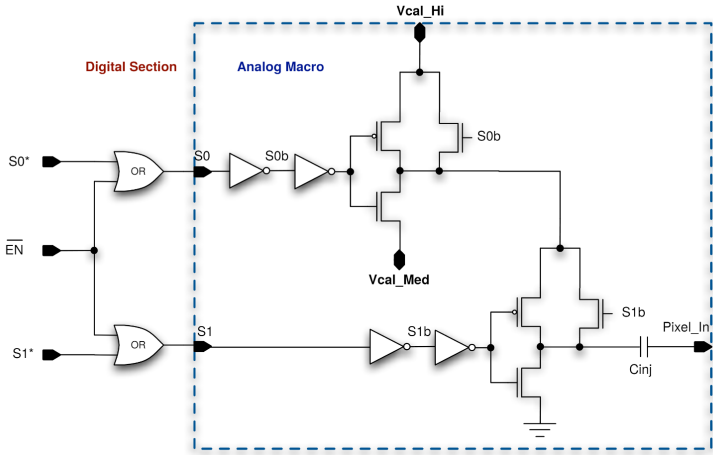
- Possibility to measure the value of injection capacitor using a dedicated circuit to define precisely the injection step
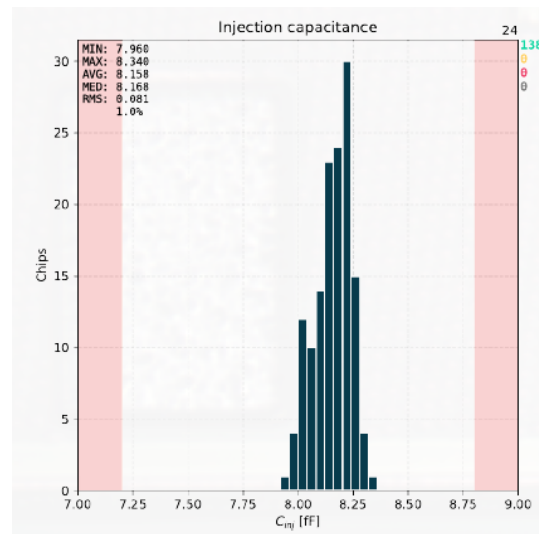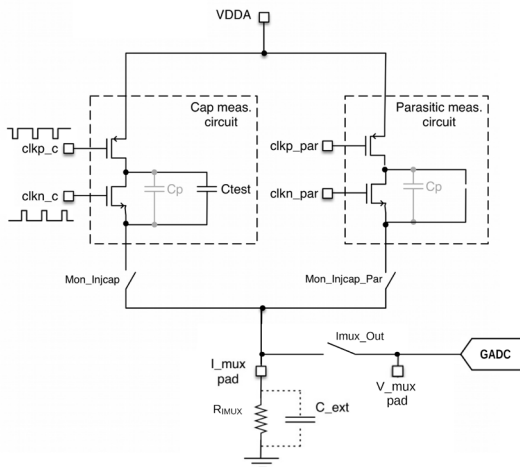
Monitoring block

- The Monitoring block enables digitization and readout of internal parameters (T, voltages and currents from different parts of the chip)
- Consists of a current mux, a voltage mux and a **12-bit Analog to Digital Converter (ADC)**
- Monitoring can be performed at any time, also during data-taking, via the normal data output links
- 5 temperature sensors in different positions (near the ShLDOs, in the centre of periphery, top/bottom of the matrix)
- Ring oscillators → measurements of digital cells speed degradation with TID
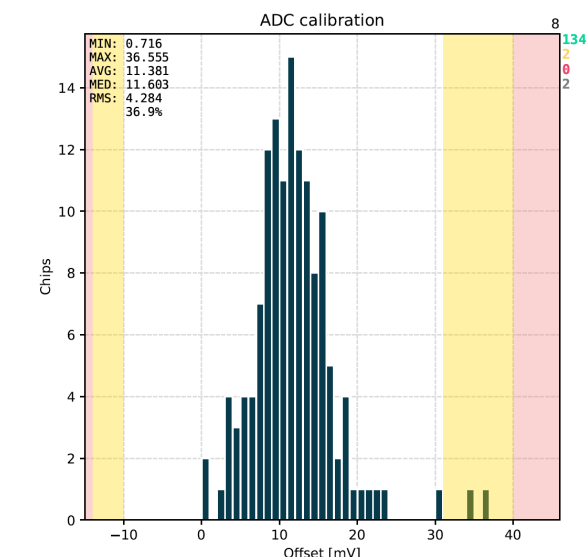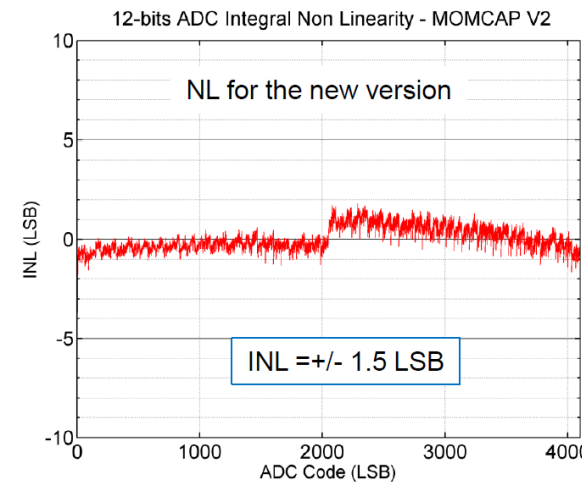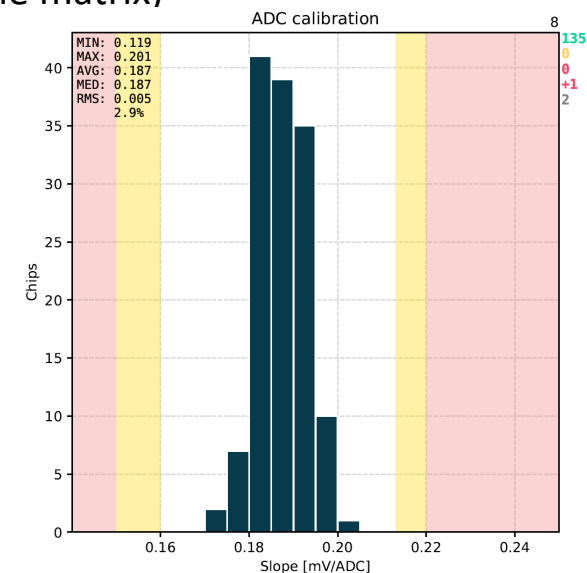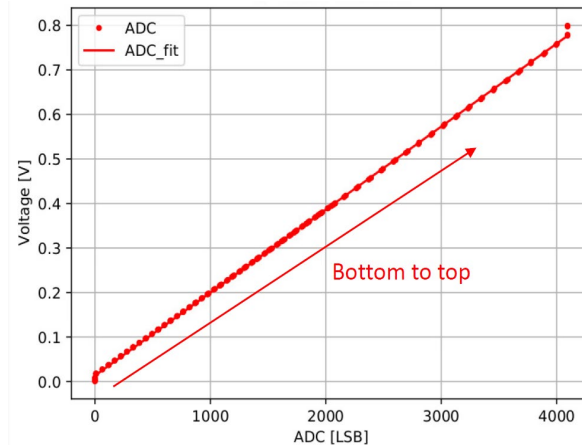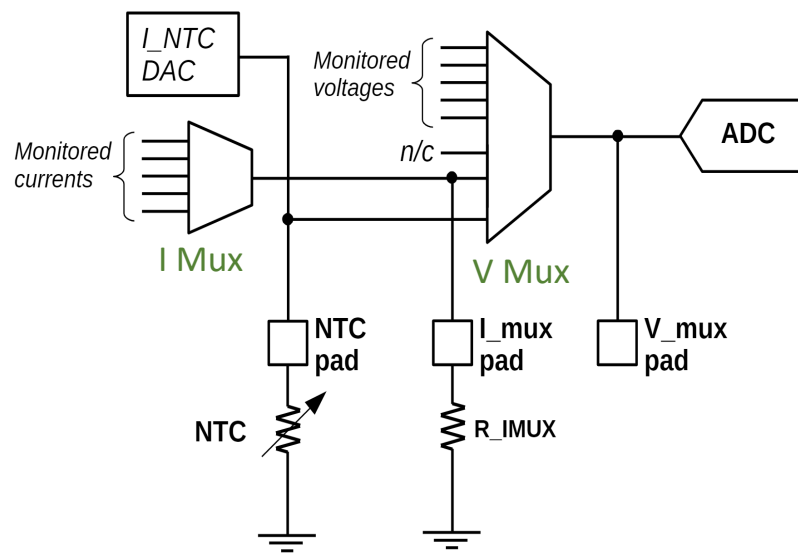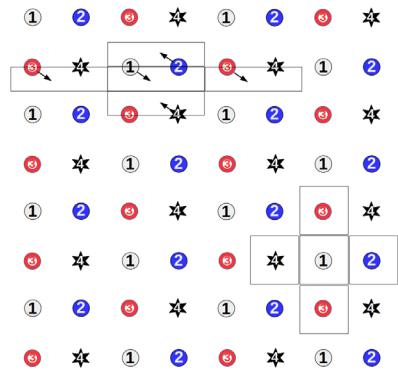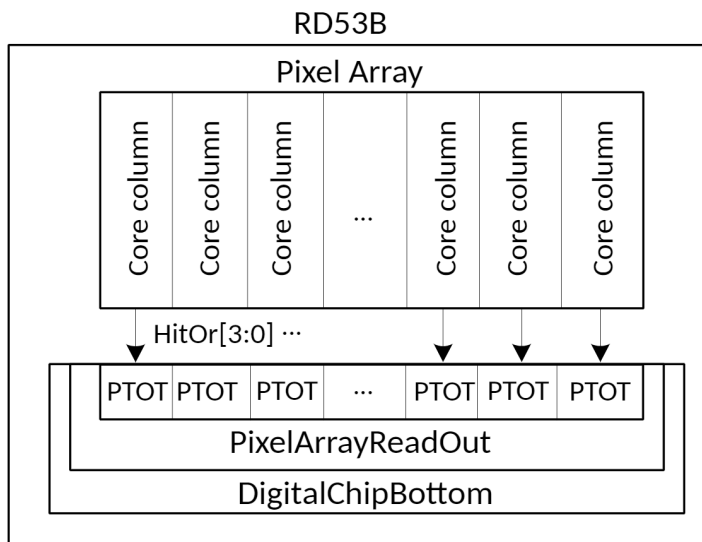
- Flexible auto trigger function, based on a Hit-OR network from the pixel array
- Hit-OR network consists of 4 OR lanes per Core Column, with a mapping such that neighbor pixels are mapped on different lines
- At the end of Core Column, the 4 lanes are combined to build the global Hit-OR with programmable patterns
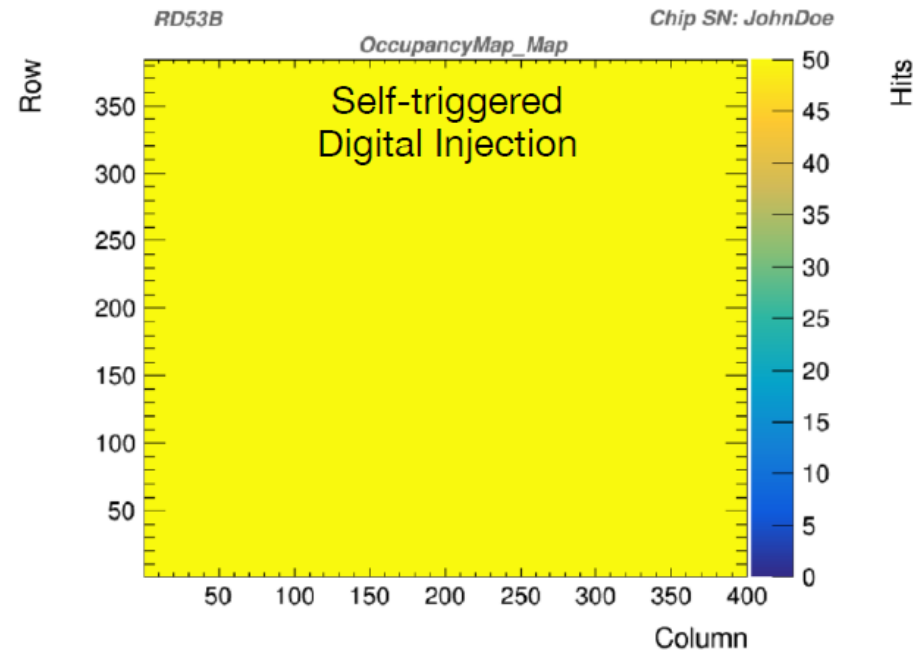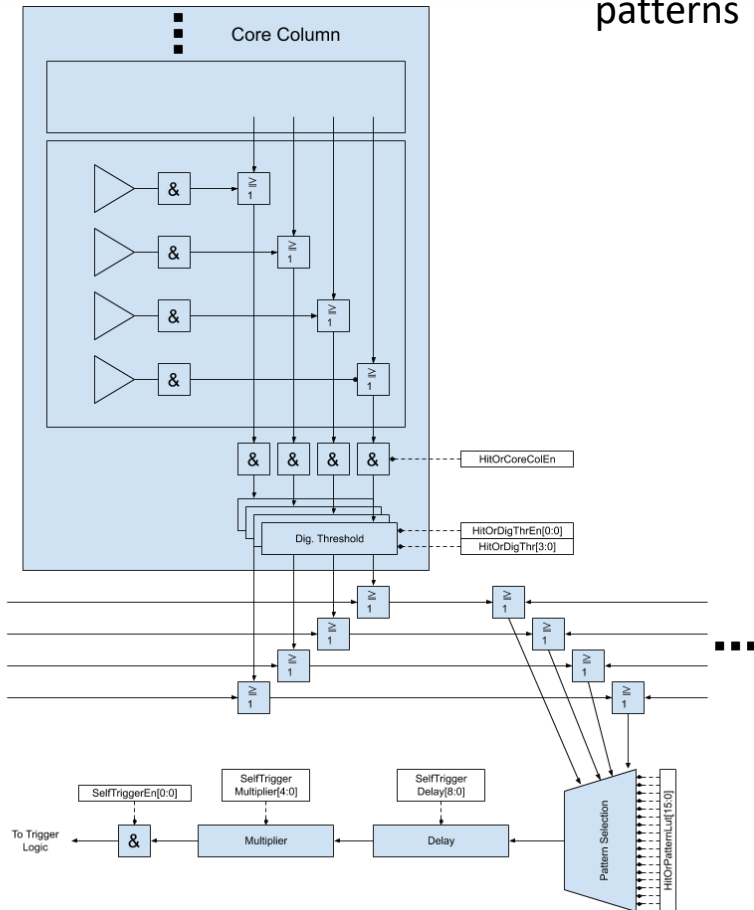
Precision ToT and ToA

RD53B

- PTOT module can be used for high resolution **Time over Threshold** and **Time of Arrival** measurement of the HitOR lines, using 640 MHz counting clock (1.5625ns resolution)
  - ➢ 11-bit PToT counters
  - ➢ 5-bit PToA counters, measuring the phase difference from HitOr leading edges and next BX clock rising edge
- Each Core Column is equipped with a PTOT module. Can be triggered for readout via the normal path, just like a Pixel Core
- Can be used to make precision measurements of analog front-end, like <u>time walk</u>
- Allows to reconstruct the amplifier output waveform (sort of <u>oscilloscope</u>)
- Working in ItkPixv1, bugged in CROCv1, fixed for final chips
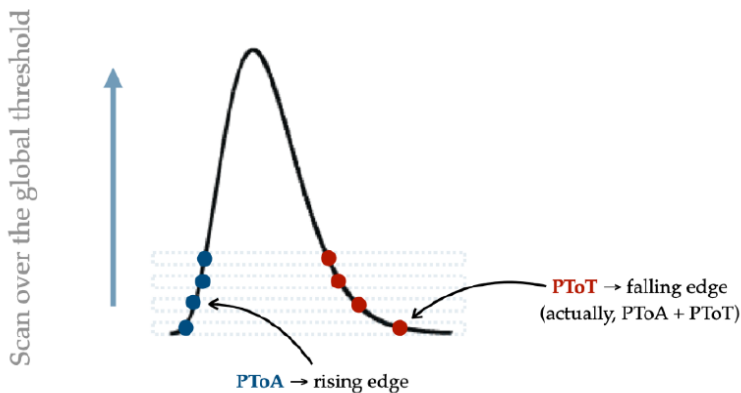
- Inject fixed calibration pulse
- Scan the threshold
- Sample PToA and PToT at each step
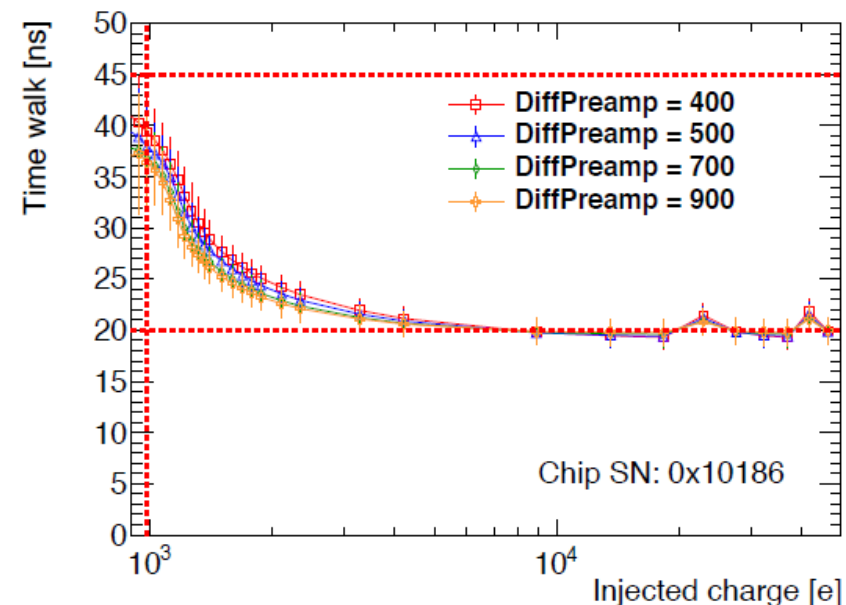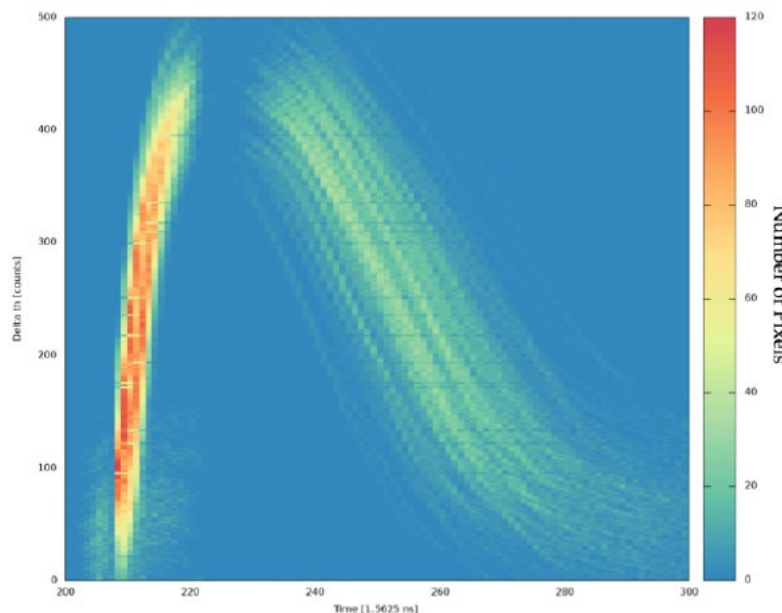
Clock and reset strategy

Clock generation and distribution

— CDR/PLL used to recover 160 MHz clock and command stream from the transitions on the input serial control link

— generated internal clocks: 160 MHz (channel-synch + DataMerge), 64 MHz (Aurora 64b/66b encoding), 640 MHz (PTOT) and 1.28 GHz (serializers + fine-delays)

— 40 MHz master clock generated by simple clock-division into ChannelSynchronizer (160MHz/4), used my most of chip-periphery and into pixel array

— 4 different clock domains overall running into chip-periphery: standard Clock Domain Crossing (CDC) techniques adopted to deal with multiple clocks (also checked with formal tools)

— automatically-inserted clock-gating used to reduce power consumption

Reset scheme

— **NO reset pin** (nor Power-On Reset) implemented in the chip to minimize the risk of false resets due to SEEs (therefore NO reset line triplication required)

— **synchronous reset** adopted everywhere (with the only exception in the per-pixel asynchronous hit-sampling logic)

— Global and Pixel configuration have **hardcoded default values** selected by a multiplexer at power-up and to switch the muxes to use register values "magic numbers" need to be written into two key registers

— individual blocks have independent reset signals that can be sent using the **GlobalPulse** command

— an additional fast-command called **Clear** can be used to fully and only resets the datapath and control state machines inside the chip

— self-reset of the CDR/PLL if no **Sync** commands are not sent for $\sim 1 \ \mu s$

$$J_{RMS} = 11.2 \ ps$$
$$J_{P-P} = 98 \ ps$$
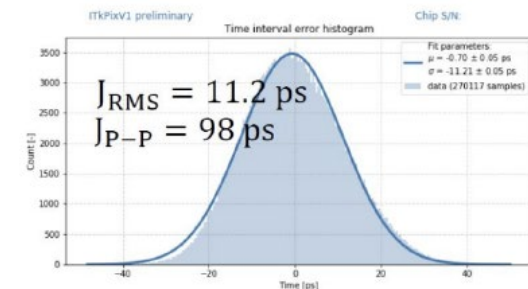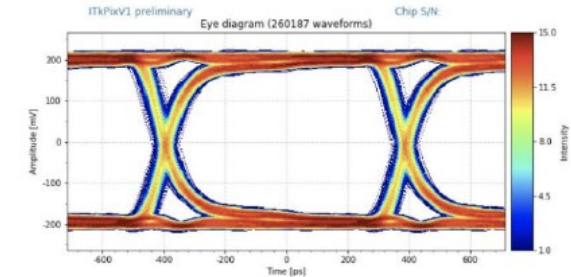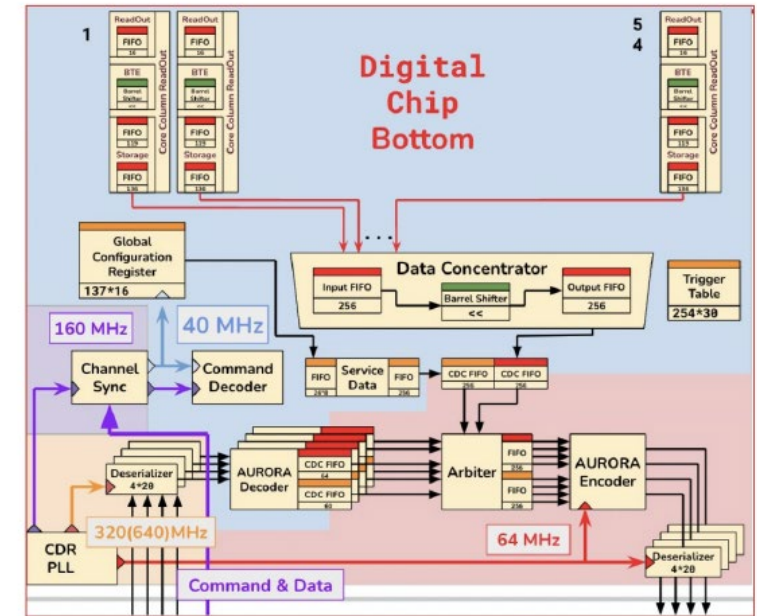
Data reduction and debug options

There are several ways to **reduce** the amount of Physics data by chip configuration:

- **Binary readout:** leave out the ToT bits, ~25% bandwidth saved
- **Event truncation:** for each Core Column, if the size of the event is over a threshold, trim it
- **Timeout truncation:** for each event, if it takes too long to read out, trim it using a timeout

There are **debug options** to add information:

- **Raw output:** output a simple hit map (do not use the Binary Tree compression)
- **CRC:** append a 32-bit Cyclic Redundancy Check to the end of a Stream
- **BCId and Level1 ID:** add 16 bits containing the Bunch Crossing ID, or Level 1 ID, or both.

Isolated hit removal
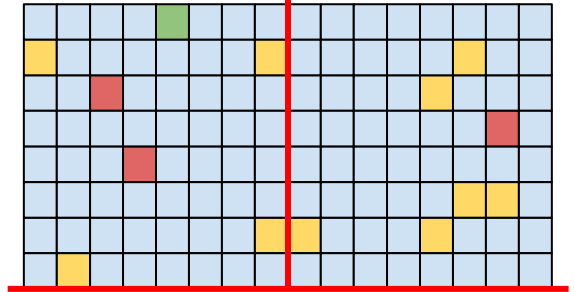
Pixels in red are considered isolated

It is only easy to look within the data for one Core Column at a time, so pixels touching vertical boundaries are not removed.

Chip interfaces: control link

> **Fast Commands** = 16 bits, 100 ns to send
> **Slow Commands** = variable length, up to some ms to send

- The RD53 control protocol was designed to meet these needs:
  - **High radiation environment:**
    - The protocol symbols are a Hamming distance of 2 bit flips apart, allowing error detection
    - To speed up recovery from errors, the **CLEAR** command, which resets the data path, is a minimum-sized **Fast Command**
    - Fast Commands can be interleaved by DAQ during **Slow Commands** such as Write Register
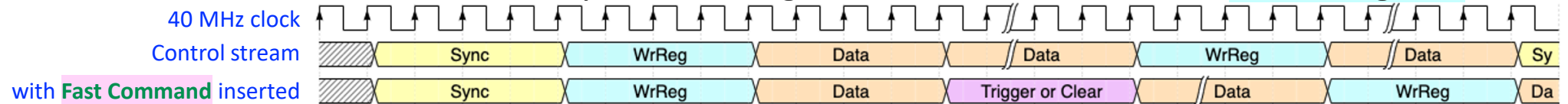
    - Write Register commands have features to auto-increment register addresses and to broadcast to up to 15 chips on a shared bus, to speed up chip configuration
    - These together mean that **the chip can be continuously reconfigured in ~100 ms**
  - **High trigger rate:**
    - Trigger commands are minimum-sized **Fast Commands**
  - **Minimise cabling and material in the detector:**
    - Both commands and the main 160MHz clock are recovered from this 1 differential pair
    - The periodic Sync command guarantees sufficient transitions to recover the clock

Chip interfaces: data output and formats

Hit data readout and compression

- Triggers cause hit data to be read out from the Pixel Array using tokens
- Data is stored in a FIFO at the end of each column
- Hit data from columns is compressed using **Binary Tree Encoder** (up to factor of 2)
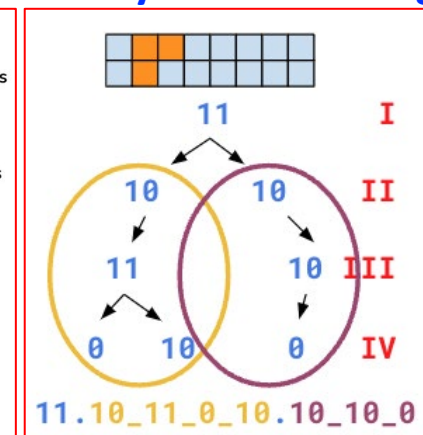
Event building and Streams

- Data is collected from all Core Columns and formatted in order to form a full event. This is done in a multi level event building stage.
- Each event is full contained and referenced by the associated Trigger Tag
- All events are built preserving trigger order (NO event mixing)
- Events are encapsulated into a variable length Stream due to Binary Tree encoding and zero suppression
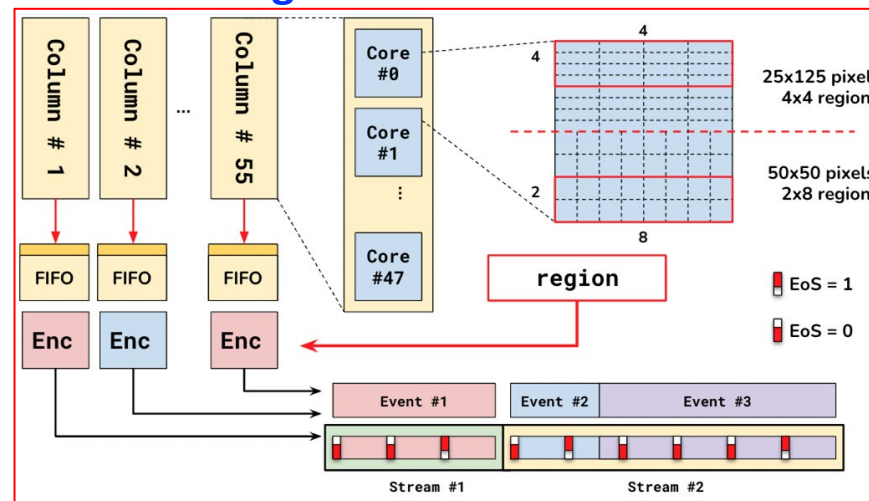
Formatting within Streams

- Streams are split into 64-bit words (Aurora data frames)
- The first bit shows if the current frame is the last of the stream (End of Stream/EoS bit)
- The next 8 bits are the Event Tag (only in the 1st frame of that event)
- Next: 6-bit address of the core column ("ccol" in the diagram)
- Next: 6-bit address of the core row ("crow")
- 2x8 region Binary Tree Encoded address ("BT" in the diagram)
- Next: the Time Over Threshold data (4 bits) for each hit pixel ("ToT")
- Once the event is finished we close the Stream padding with zeros or insert a new event separating it from the previous one with a three bit separator (111) that never can occur in event data.

Event building

Binary tree encoding

Encoding of Physics and Register ("Service") data