Progress towards a 4D fast tracking pixel detector

M. Petruzzo[1],
M.Citterio[1], J.Fu[1], P.Gandini[1],
A. Lai[3], M.Luchi[2], V. Liberali[1,2],
N.Neri[1,2], S.Riboldi[1,2]

[1]INFN, Sezione di Milano
[2]Università degli Studi, Milano
[3]INFN, Sezione di Cagliari

Connecting the Dots
April 20-30, 2020

for the **TIMESPOT** team, and in the framework of the **LHCb-RTA** project

- **TIMEPOST R&D** project
  - **TIME** & **SP**ace real-time **O**perating **T**racker
  - three years project, from 2018, funded by **INFN**
  - development of a **silicon and diamond 3D tracker with timing facilities:**
    - **demonstrated 30 ps** hit resolution at testbeam at PSI[https://pandora.infn.it/public/52d779]
  - **construction of a demonstrator** integrating sensors, electronics, **real-time** processors

https://web.infn.it/timespot/

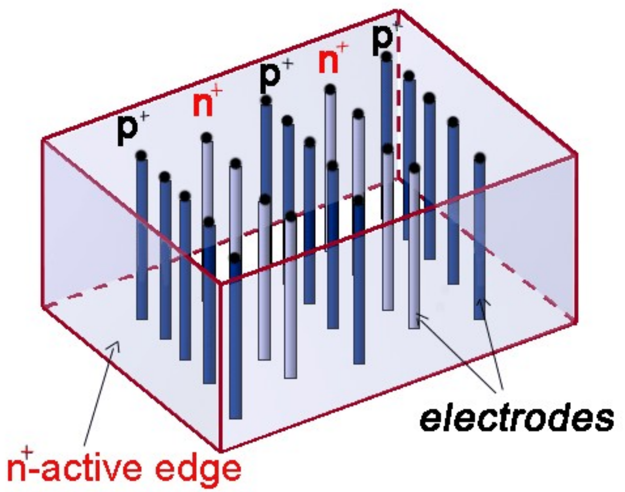
- **Work Packages and coordinators:** – Principal investigator: A. Lai
  - 3D silicon sensors: development and characterization – G.F. Dalla Betta
  - 3D diamond sensors: development and characterization – S. Sciortino
  - design and test of pixel front-end – V. Liberali
  - design and implementation of fast tracking devices – N. Neri
  - design and implementation of high speed readout boards – A. Gabrielli
  - system integrations and tests – A. Cardini

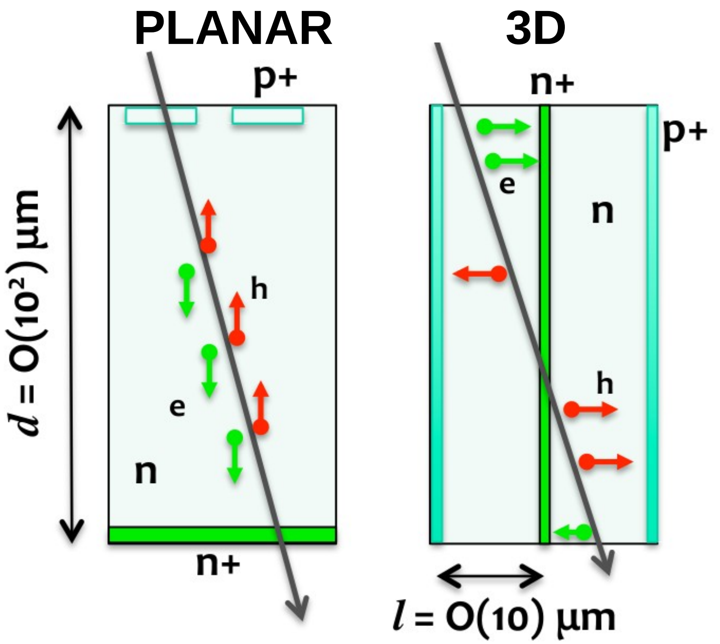
3D silicon sensors

- 3D sensors are a well established technology in which **electrodes are realized as vertical columns** through the silicon substrate
  - already used for vertex detectors (i.e. ATLAS IBL)
- Higher fabrication complexity and cost, w.r.t. to planar technology
- Geometric inefficiency (~blind electrodes)

  ...but...

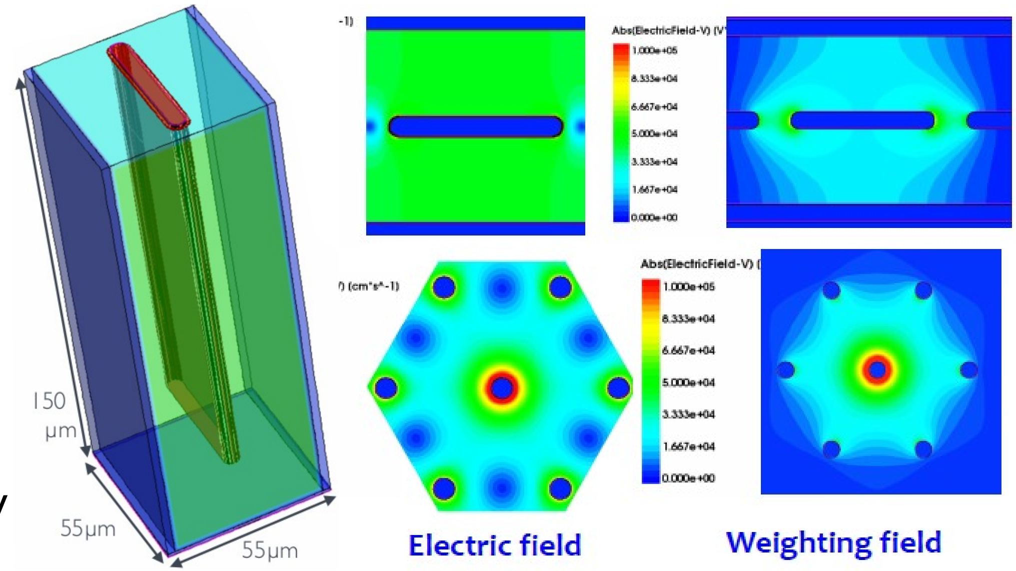
- **Radiation hardness:** > $3x10^{16}\ n_{eq}/cm^2$
- By design, **extremely fast signal**:
  - charge deposition decoupled from electrode distance
  - **potentiality for timing**, not yet exploited

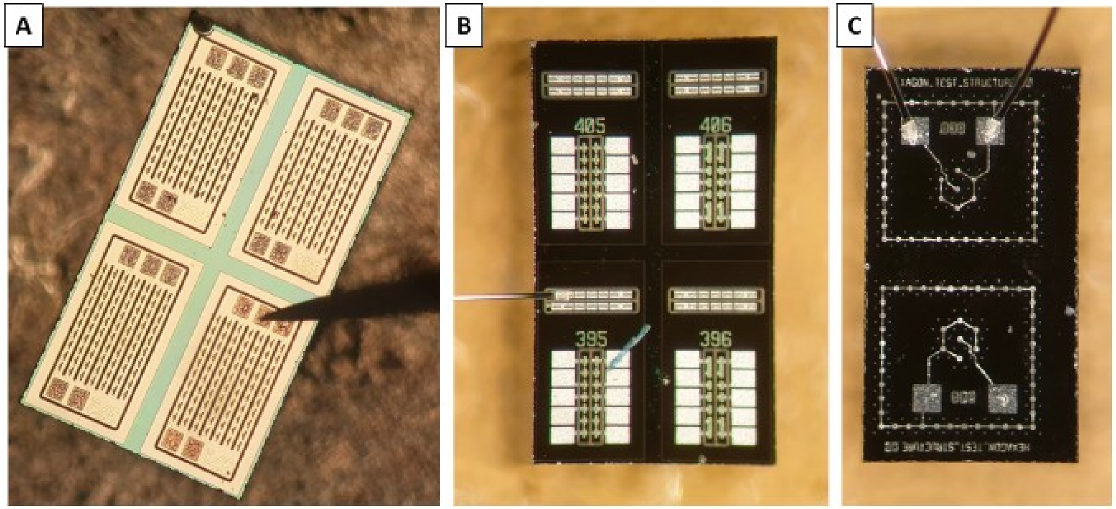




3D Silicon Trench geometry

- 3D Silicon **Trench geometry**:
  - best solution for **electric field uniformity**
  - **charge collection time < 400 ps**
  - strong mitigation of Landau fluctuations
  - most probable value of charge deposit ~2 fC
  - full depletion ~20 V – velocity saturation ~40 V

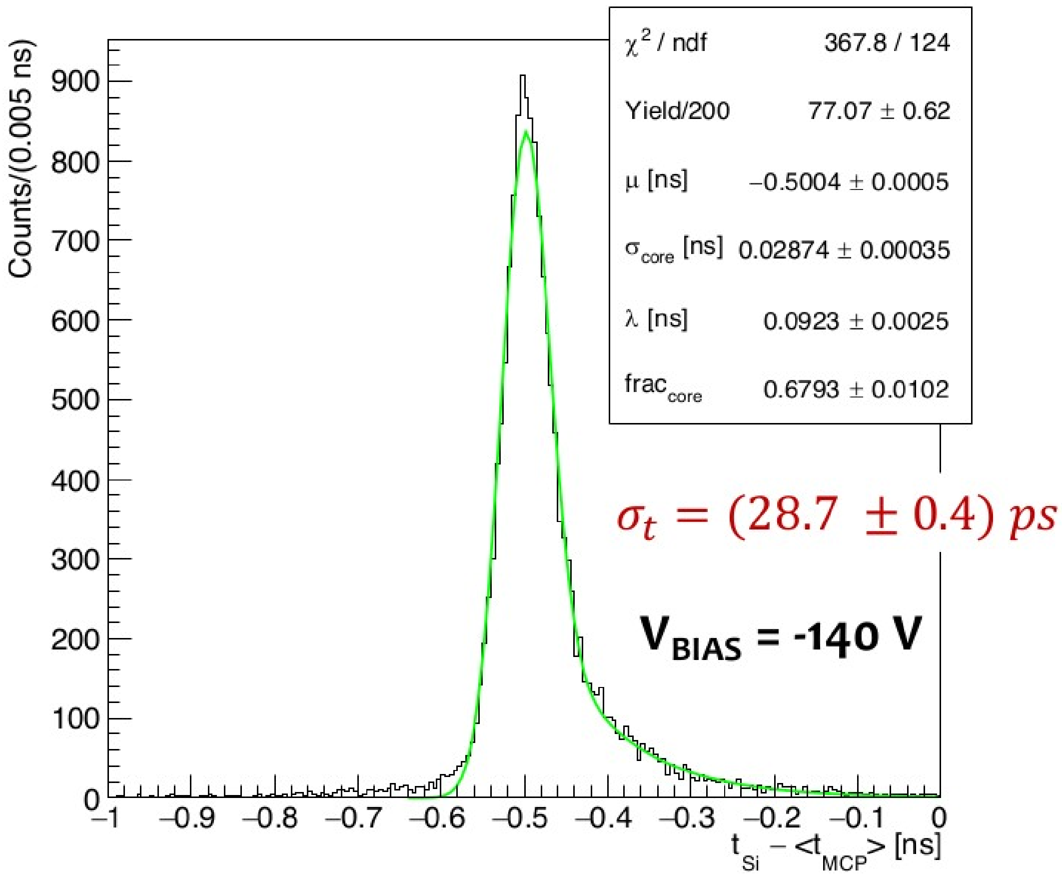


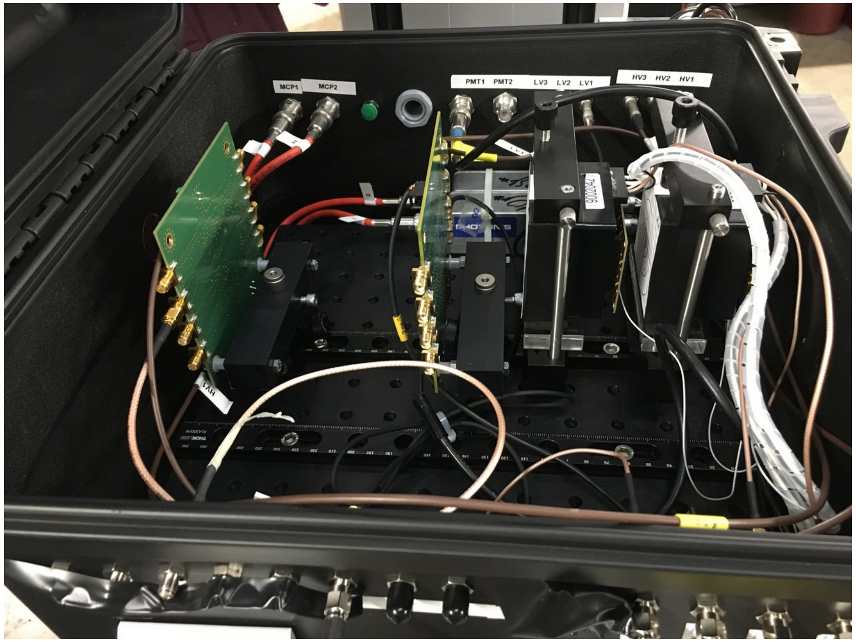


- 1st 3D trench batch produced by FBK in June 2019
  - test structures:
    pixel strip r/o (10 pixels connected on the same read-out pad) and single pixel r/o
  - TimePix geometry with 256x256 55x55 μm^2 pixels

Testbeam at PSI

- **Test structures tested at PSI with a 270 MeV/c π+**
- Timing is measured from the difference with two reference MCPs (~18 ps resolution)
- DUT time is obtained from 35% CFD on the signal





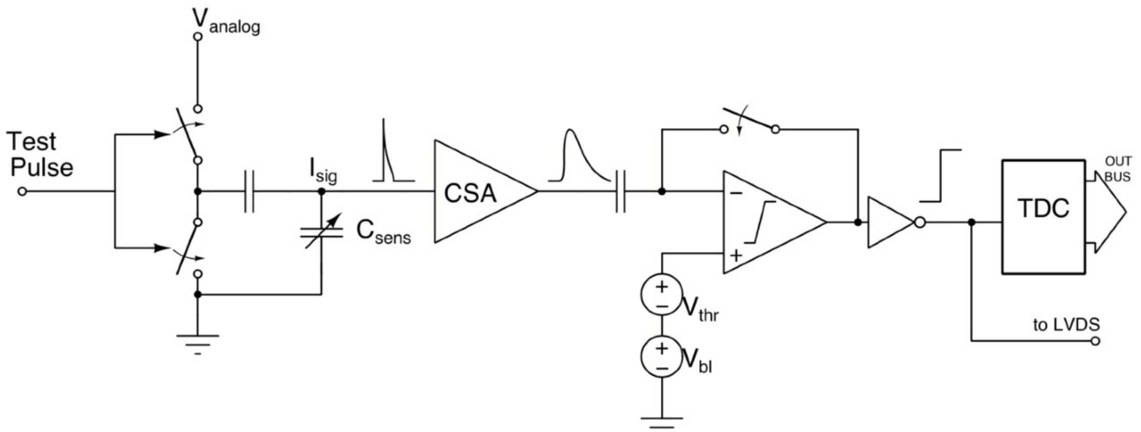
$$\sigma_t^2 = \sigma_{t,\,Jitter}^2 + \sigma_{t,\,Landau}^2 + \sigma_{t,\,TimeWalk}^2 + \sigma_{t,\,MCP}^2$$

Time resolution mainly dominated by DUT jitter

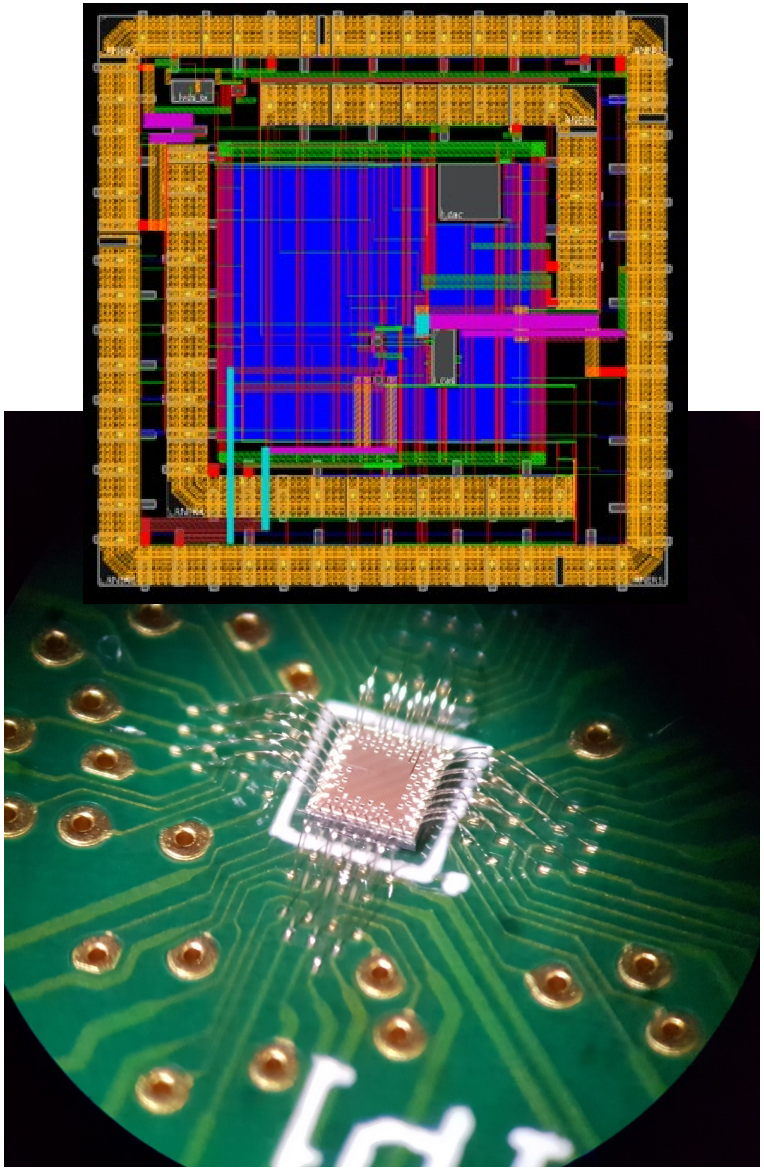
- σ_t = ~45 ps (no TOT correction)
- **σ_t = < 30 ps** (numerical filters to reduce high frequency noise)

Front-end ASIC

- **First front-end ASIC prototype on 28 nm CMOS** technology with full pixel readout chain:
  - charge injection circuit with prog. capacitance
  - charge sensitive amplifier (CSA) with sensor leakage current compensation
  - leading-edge discriminator with discrete-time offset compensation
  - 8 independent channels
  - 3 TDC technologies
- **Configurable power between 4.1 μW and 7.2 μW**



1.5x1.5 mm² mini@sic



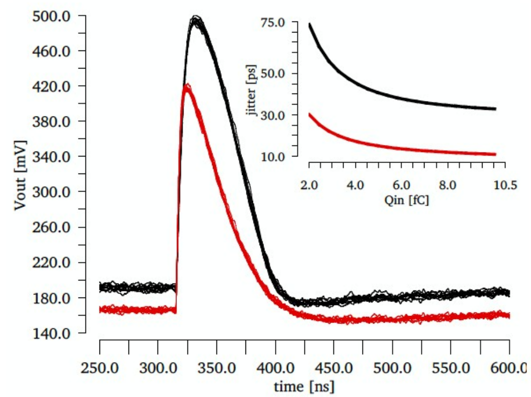
Prototype first results

Prototype zero, to gain confidence with 28nm technology

CSA performance ~x2 worse than simulation (under estimated parasitic in the feedback loop)

Overall measured time resolution was in the range of 150 ps (LP) to 70 ps (HP)

Input Signal	Delta		Sensor	
Power [μW]	***4.1***	***7.2***	***4.1***	***7.2***
G [mV fC^{-1}]	190	168	150	124
σ_n [mV]	2.8	2.0	2.8	2.0
ENC [e]	94	77	120	103
t_{pk} [ns]	16.4	7.7	18.2	10.2
t_A [ns]	2.1	2.1	4.2	3.5
TOT [ns]	100	98	79	78
SR [mV ns^{-1}]	53	98	39	68
σ_j [ps]	54	21	74	30
σ_p [ps]	66	65	67	66
σ_{mm} [ps]	33	26	40	29



	First Scheme	Second scheme	Third scheme
Size (μm²)	23 x 22	27 x 22	23 x 21
LSB (ps)	190	50	22
RMS (ps)	47	15	37
Power Active (μW)	1200	1200	65
Power Standby (μW)	10	10	34

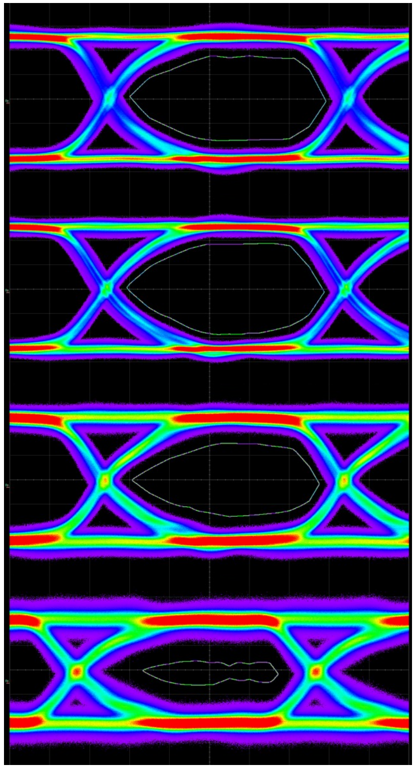
- **LVDS Tx/Rx** test:
  - loopback configuration
  - **BER < 10^{-15}** with low-power mode @ **1.5 Gpbs**
  - “silicon proof” link in all power modes

High Power (1.6 mA)

Typical Power (4.1 mA)

Low Power (2.7 mA)

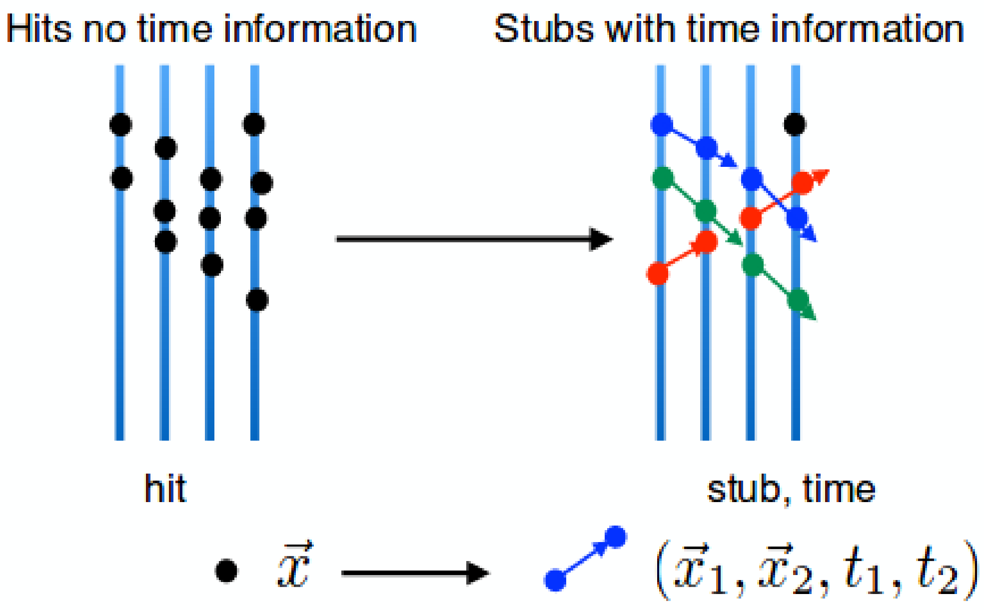
Ultra-Low Power (1.6 mA)



4D stub based tracking

[N. Neri et al., JINST 11 (2016) no.11, C11040]

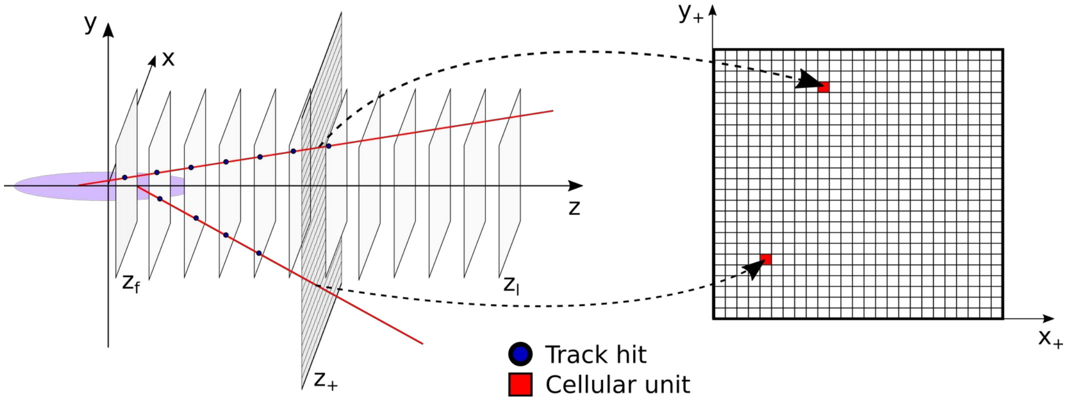
- **Stub approach:**
  - stubs are identified as **doublets of hits** in adjacent planes
  - a stub provides a **"track hint"** with **no assumptions on the particle origin**
  - **geomtrical cuts** are applied to filter stubs not compatible with tracks from the luminous region
  - **tracks identified from clusters of stubs** with similar parameters

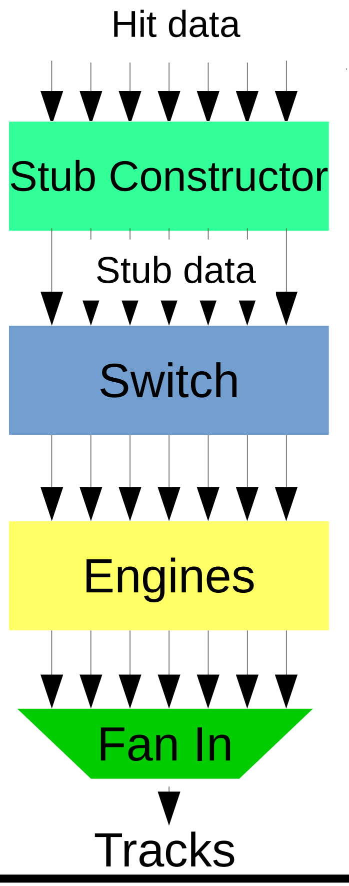


- **Stubs + Timing:**
  - stub creation is an important step: need to maintain low fake stub rate
  - particle velocity is required to be compatible with the speed of light
  - **time** information allows **further combinatoric suppression** and fake stub rate reduction

4D track finding device

- **Parallel track finding algorithm**
  - identification of stubs (hit doublets) + geom+time filtering
  - projection of stubs to a 2D reference plane
  - tracks identified as clusters of stubs with similar parameters





Evaluates and filter the combinations of hits in adjacent detector

Implemented as full-mesh network, delivers the stub data to the engines

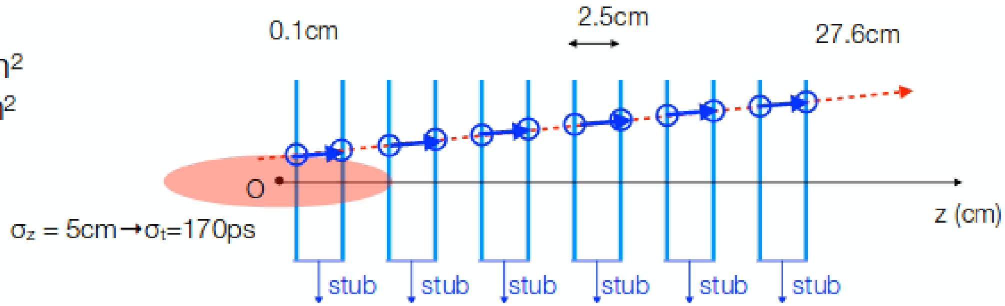
Organized in “regions”, receive the stubs, identify and reconstruct the tracks

Collect the track results from the Engines

- The system has implemented in **FPGA**:
  - **highly parallelized** architecture
  - **pipelined** architecture
  - **low latency** budget **<1 μs**
- **FPGAs** allows for reprogramming flexibility while keeping high performance in dedicated task
- **Particular focus to the latency:**
  - data transfer between the modules is **self managed**, through the implementation of a **custom hold logic**
  - reduced data serialization
  - **guaranteed minimum latency**

Algorithm simulation

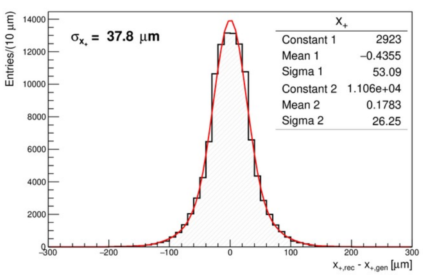
Sensor area = 6x6cm²
pixel size = 55x55μm²
thickness = 200 μm
time res σ_t=30 ps

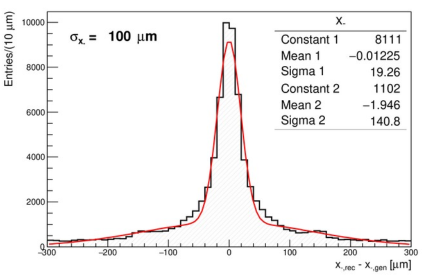


- **LHCb VELO-like detector:**
  - 12 planes in the forward region
  - pile-up ~40
  - ~1200 tracks/event

- **Luminous region:** σ_z=5cm, σ_t=167ps . **90'000 engines** in [-2,2]x[-2,2]cm² reference plane

$\sigma_{x+/y+}$= 37.8 μm
$\sigma_{x-/y-}$= 100.0 μm





First row: resolution on track parameters (x+,x+-) **without** using the time information of the stubs
Second row: resolution on track parameters (x+,y+, t) **using the time information** of the stubs

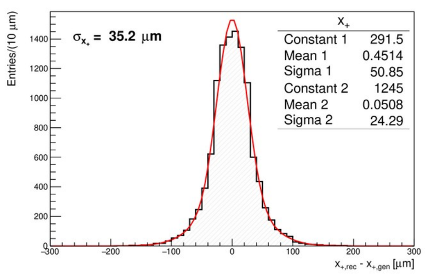
Efficiency = 99%

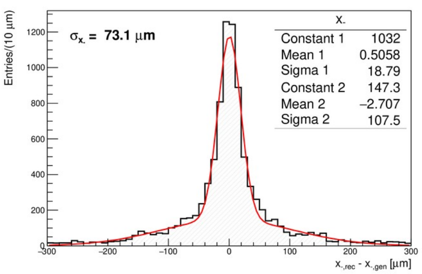
Purity = 64%

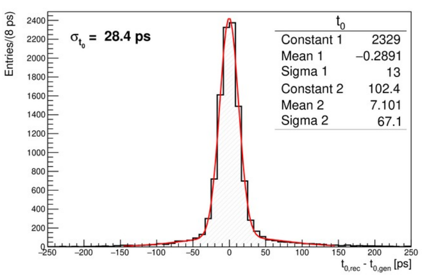
Time

Time

$\sigma_{x+/y+}$= 35.2 μm
$\sigma_{x-/y-}$= 73.1 μm
σ_t = 28.4 ps







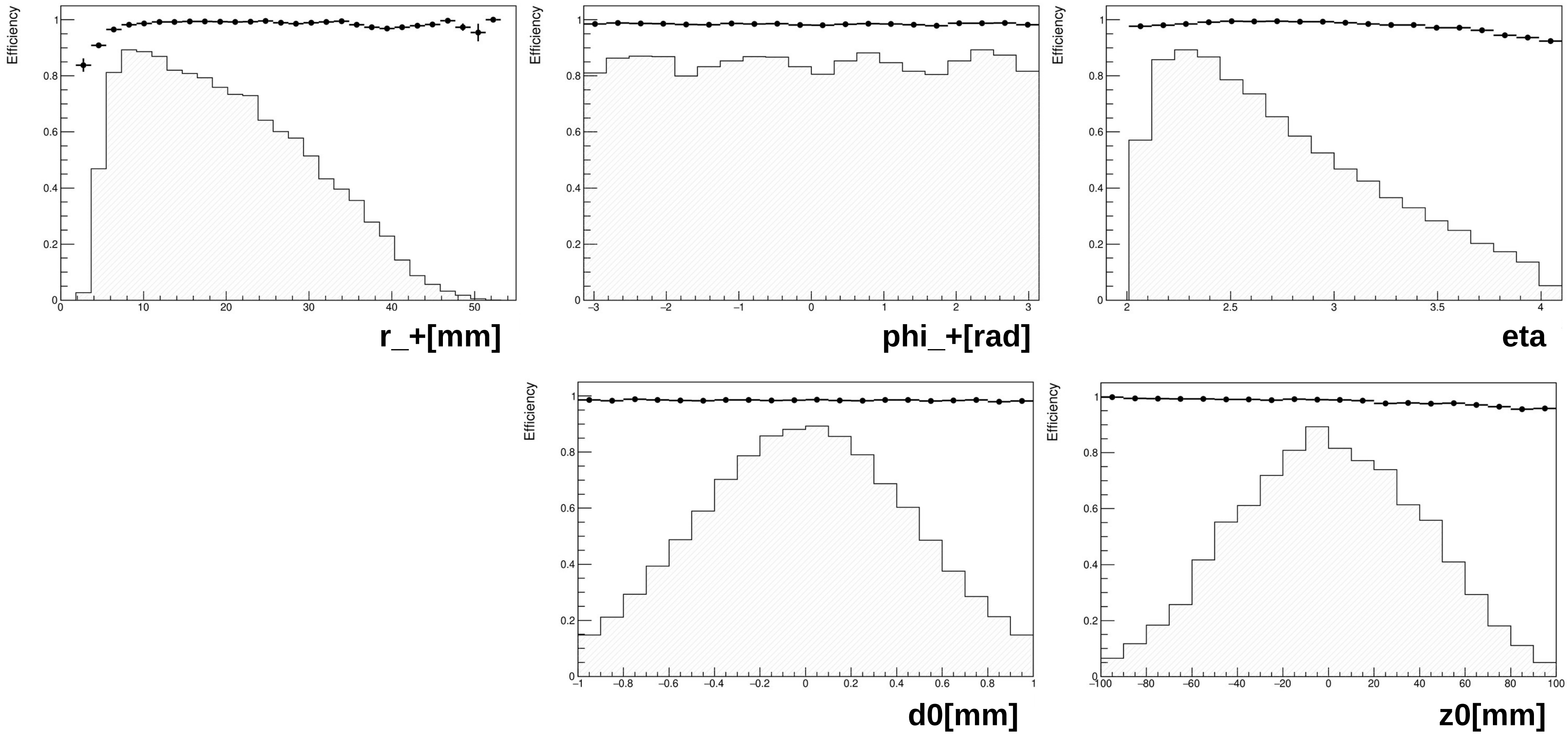
Efficiency= 99%

Purity = 85%

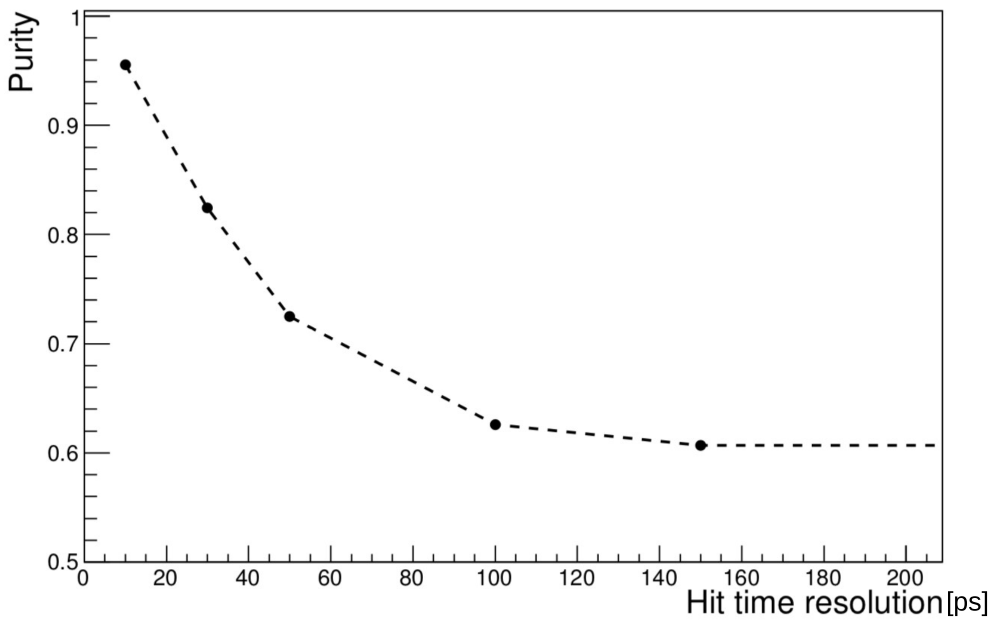
- The **track parameters resolution improves** when including the time information
- The reconstruction **efficiency is stable**
- The tracks **purity improves**

Efficiency

- **Reconstruction** (per track) **efficiency** vs track parameters r_+, φ, d_0, z_0, η :
  - track efficiency: ~98%
  - track purity: > 80% with 1200 tracks per event, using timing information
  - track purity: ~ 60 % without timing



Timing benefits

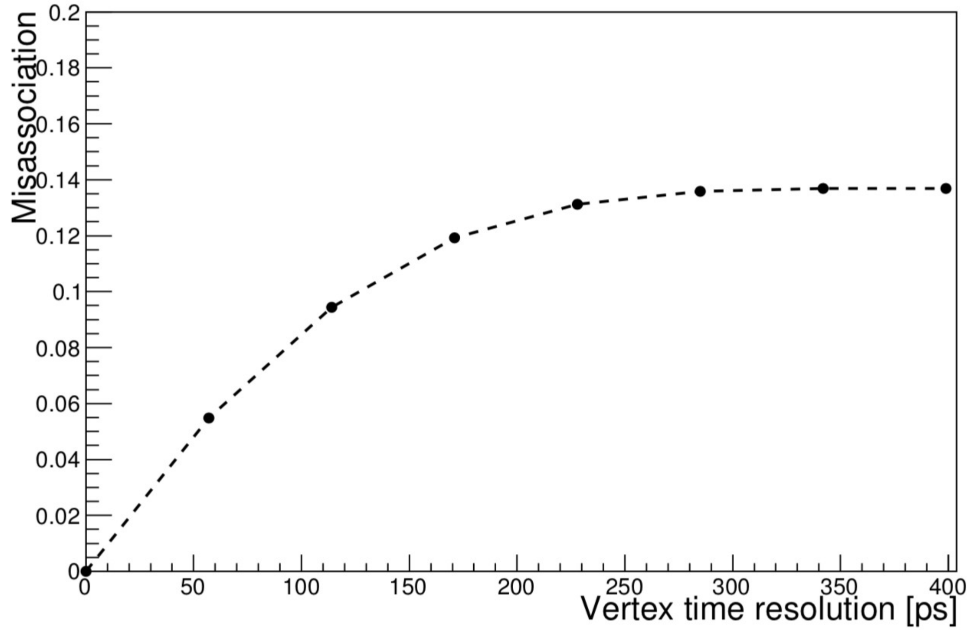


- The **track parameters resolution improves** when including the time information
- The reconstruction **efficiency is stable**
- The tracks **purity improves**

Before		After
Efficiency = 99% Purity = 64%	Time 30 ps per hit →	Efficiency= 98.5% Purity = 82%

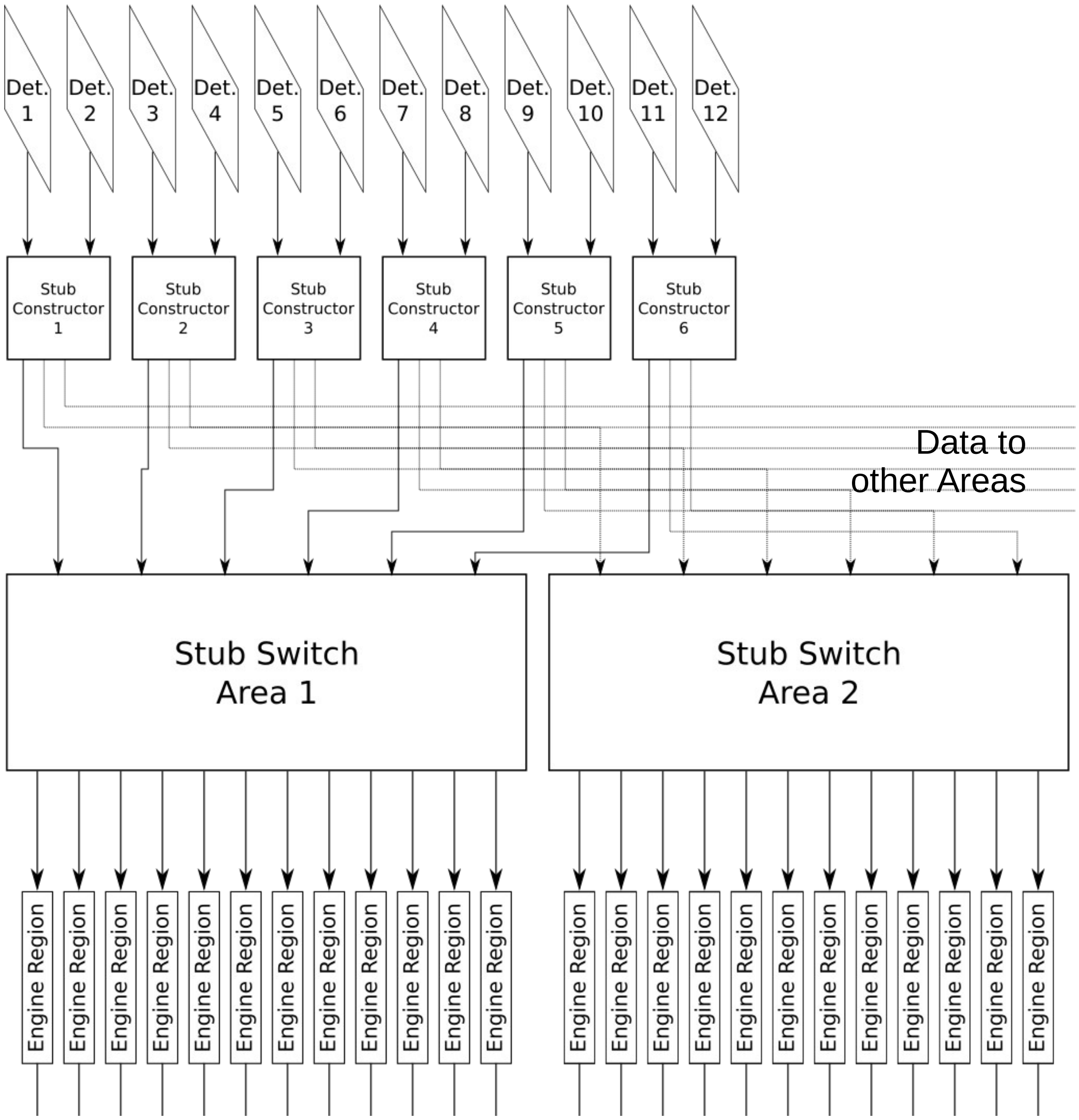
- Inclusion of timing information of the hits allows to evaluate the **time of the track**, to be used to better associate tracks to their primary interaction

- **Track mis-association** >10% (no time information)
- **<1% using precise time** information of the hit in offline reconstruction

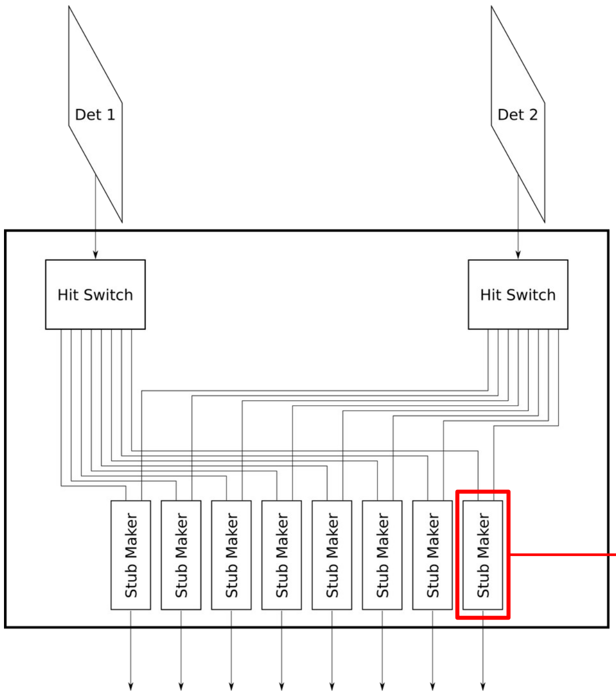


Hardware architecture

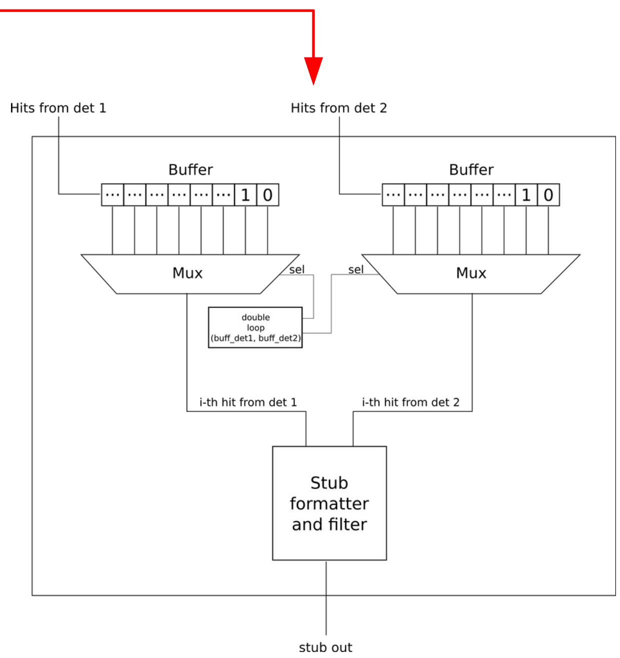
- **Independent Stub Constructors:**
  - one for each plane doublet
- **Independent Stub Switches:**
  - one for each Area of Engines
- **Engines** in the 2D tracking plane organized into multiple **independent regions**
- **Absence of "lateral" communication** between modules, make the system **modular and scalable**



Stub Constructor



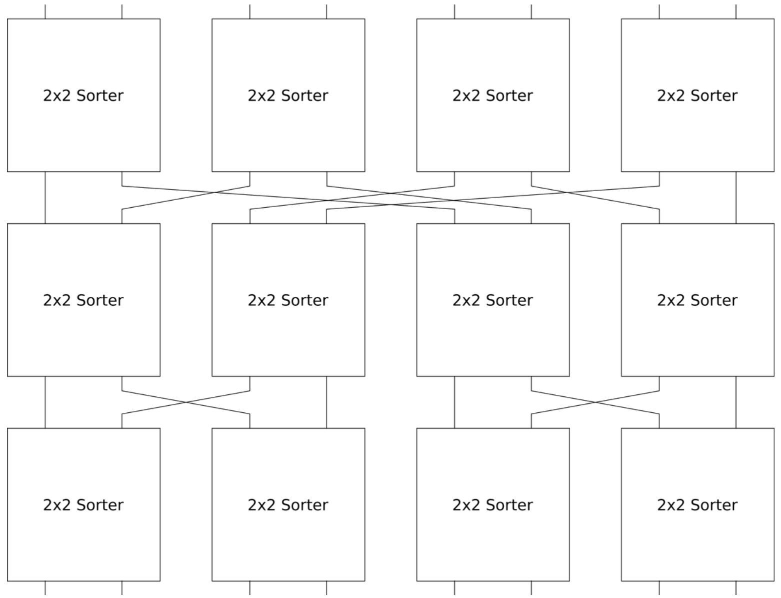
- **One module for each couple of sensors**
- **Sensors are divided in exclusive regions.** Only compatible regions on first and second sensor are processed by dedicated **Stub Makers**.
- **Sensor regions** are optimized to have **uniformly populated Stub Makers** and reduce the total latency.
- Hit data are delivered to the Stub Makers by dedicated **Hit Switches** based on the hit coordinates (r,phi) and pre-computed patterns.

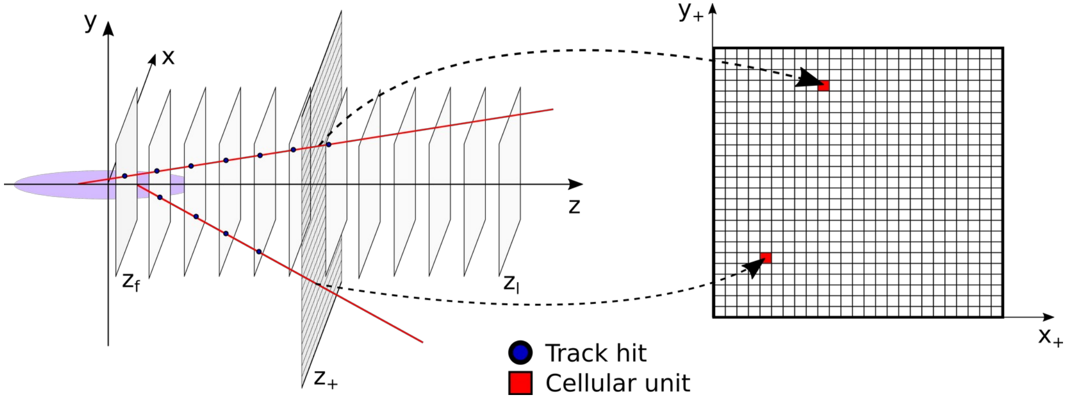


- The **Stub Maker evaluates the combination of hits** from two compatible sensors regions. It is composed of:
  - **two independent buffers** to store the hits
  - **two mulitplexers** to perform a double loop over the buffer entries
  - **a filter** to apply geom+timing cuts and suppress fake stubs
- In the **simplest case one combinatorial** unit is used

Switch + Engines

- **Switch:**
  - **full-mesh network** of Sorters (2x2 in the example) arranged in multiple layers.
  - **data are delivered based on a pre-evaluated address** (in the Stub Constructor), based on the readial and azimuthal coordinates of the stub projection to the reference plane
  - an **hold logic is implemented to pause the data flow** when the data can not be accepted by the following items in the distribution chain

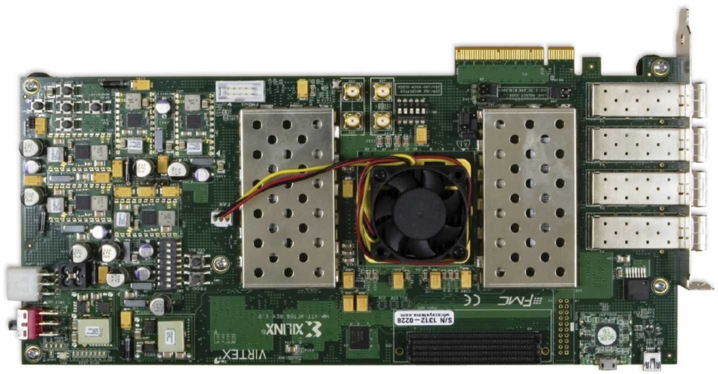




- **Engine:**
  - it corresponds to **1 cell (+8 lateral) in the tracking plane,** receives Stub data from the Switch
  - the Stub coords are used to identify the cell. A counter in the found cell is increased
  - **a track is identified if the central cell is a local maximum**, w.r.t. the the lateral cell's counter
  - the candidate track parameters are evaluated as the **average of the incoming Stub** parameters

Hardware implementation

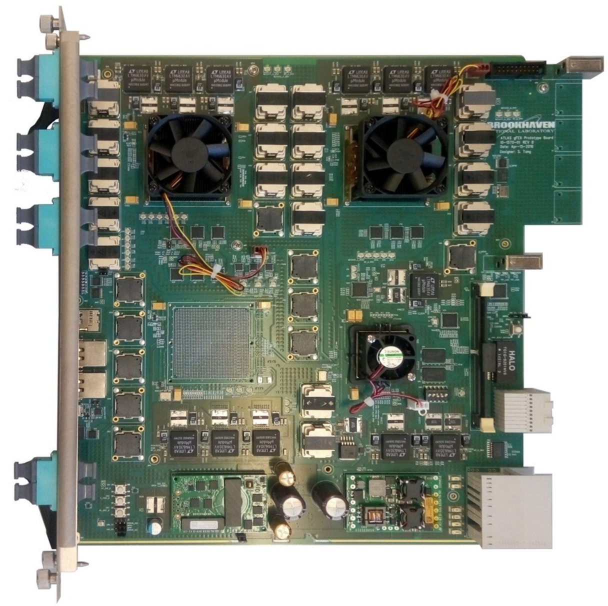
Xilinx VC709 Evalation board:



- **PCIe DMA**, implemented using Nikhef **WUPPER** : up to 60 Gbps data transfer rate [https://redmine.nikhef.nl/et/projects/wupper]
- **optical links** based on GTH transceivers: 4 (up to 12) bi-directional links at **12.8 Gbps**
- **DDR3 RAM**: 2x 4 GB banks, 100 Gbps max.read/write rate (per bank)

Stub Contructor implementation **on VC709** ongoing

gFEX board, developed at BNL for ATLAS calorimeter:

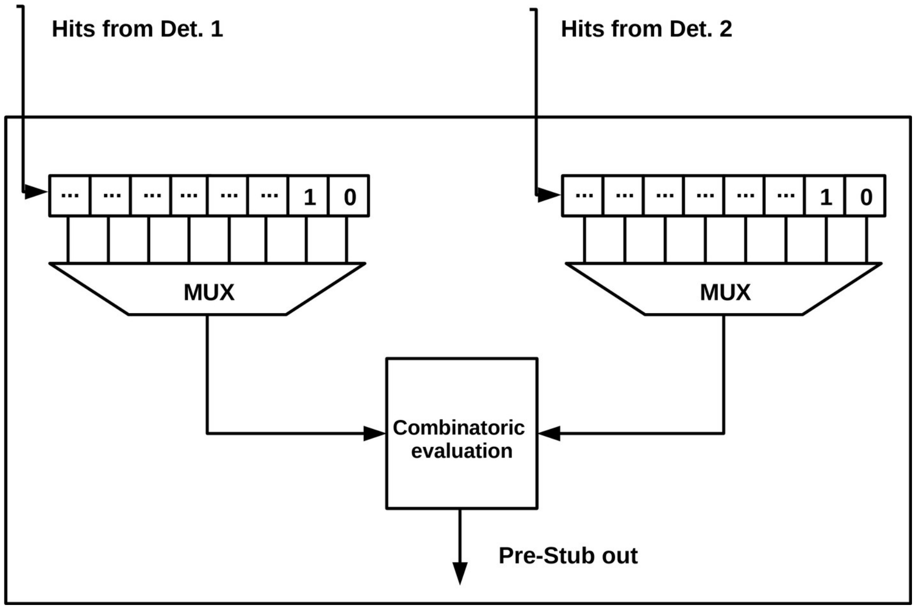


- two **Xilinx Virtex Ultrascale** FPGAs
- high-speed optical transceivers → ~**1.6 Tbps input data rate**
- one Xilinx Zynq FPGA

Switch + Engines + Fan Ins fully implemented

Stub Maker Test

- **The combinatorial** logic of the Stub Maker has been tested in hardware on the VC709:
  - no stub filtering (to be tested)
  - **clock: 200 MHz**
  - **buffer depth: 8.** Data in excess are lost
- **Test:**
  - “Events” formed by N hits + 1 End of Event data: N (+1 E.o.E) expected combinations
  - groups of M “Event” hits provided to the Stub Maker
  - during the processing of one event, an **hold signal is automatically generated** and back propagated, queue the data
- **Result**: **test passed,** providing both the correct combinations and proving the hold logic behavior in this module



- The implementation of the full Stub Constructor is ongoing:
  - the **Switch** has **already** been **tested** (see next slide)
  - the Stub filter has to be finalized and tested

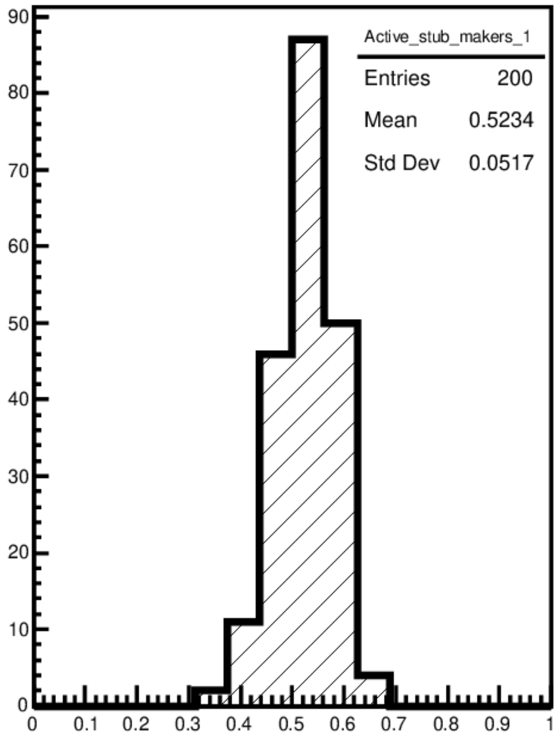
Number of Stub Makers

- The number and the occupancy of the Stub Makers depend on how the sensors are divided:
  - **dividing each sensor in 8*32 (r,phi)-regions,** the number of correlated regions in first and second sensor of each couple is evaluated as the number of correlated r-regions and phi-regions
- **14815 Stub Makers for 8 detector couples** in a VELO-like configuartion
  - it does not linearly scale with the number of (r,phi)-regions
  - **815 Stub Makers produces Stubs that points to 1/64 of the full track parameter space**, covered by the Engines implemented in the gFEX

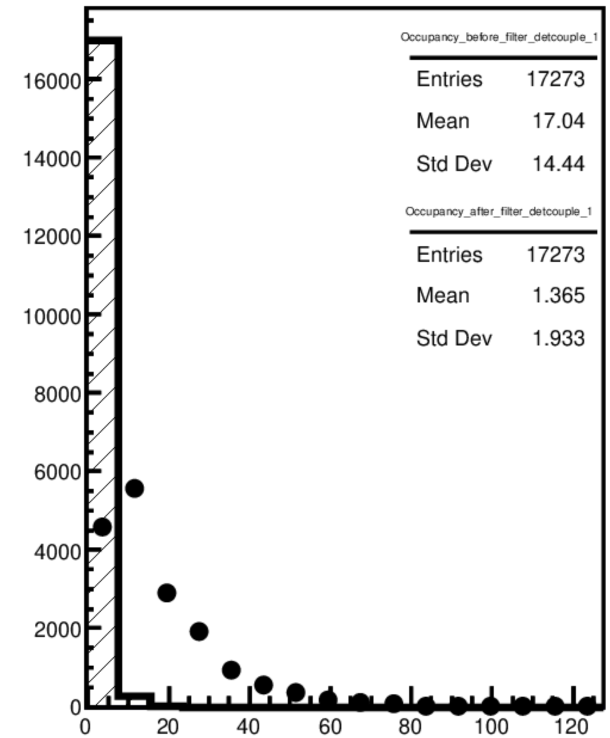
Det. couple	#corr. r-regions	#corr. phi-regions	#stub makers	#stub makers pointing to 1/64 of track space
1	26	96	2496	120
2	26	96	2496	165
3	23	96	2208	180
4	20	94	1880	72
5	15	95	1425	90
6	17	92	1564	75
7	17	90	1530	65
8	16	76	1216	48
Total			**14815**	**815**

Occupancy of the Stub Makers

- In a software simulation, it has been estimated that **about 50% of the Stub Makers is processing data for each event, which have to process ~40 combinations** (in the worst case)



Detector couple 1
(not the worst case)



- **Distribution of the number of "active" Stub Makers**, processing at least one data during an event
- Distribution of the **number of candidate Stubs processed** by each Stub Maker (**dots**) and distribution of **filtered stub (shaded/filled)** per event

Stub Constructor throughput

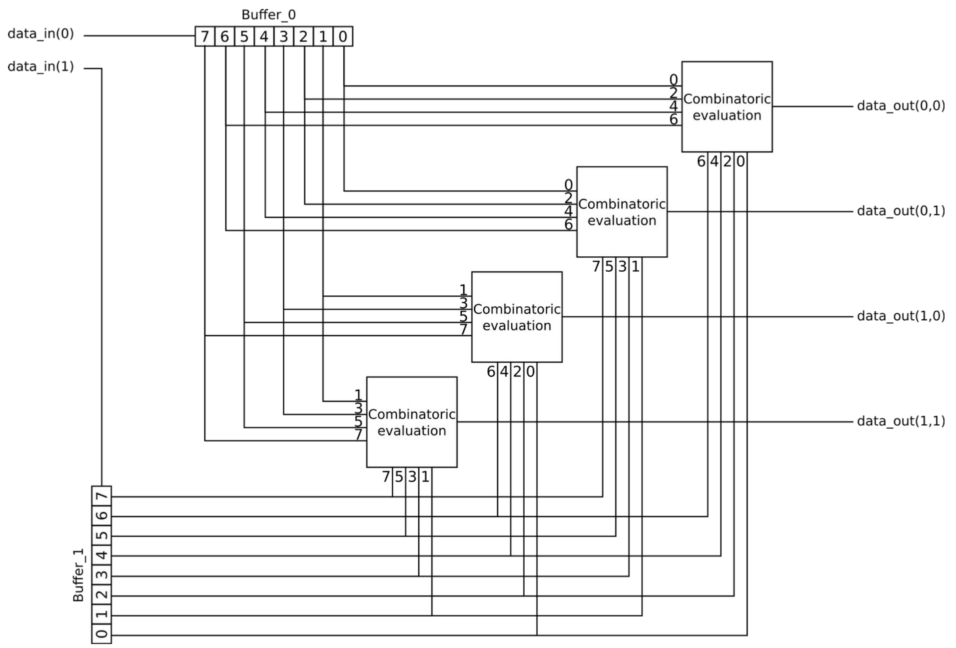
- In order to process events at 40 MHz rate in a VELO Upgrade II -like configuration, the following **equation need to be satisfied:**

$$\frac{1}{f} * Occ * N < \frac{1}{40\ MHz}$$

where:

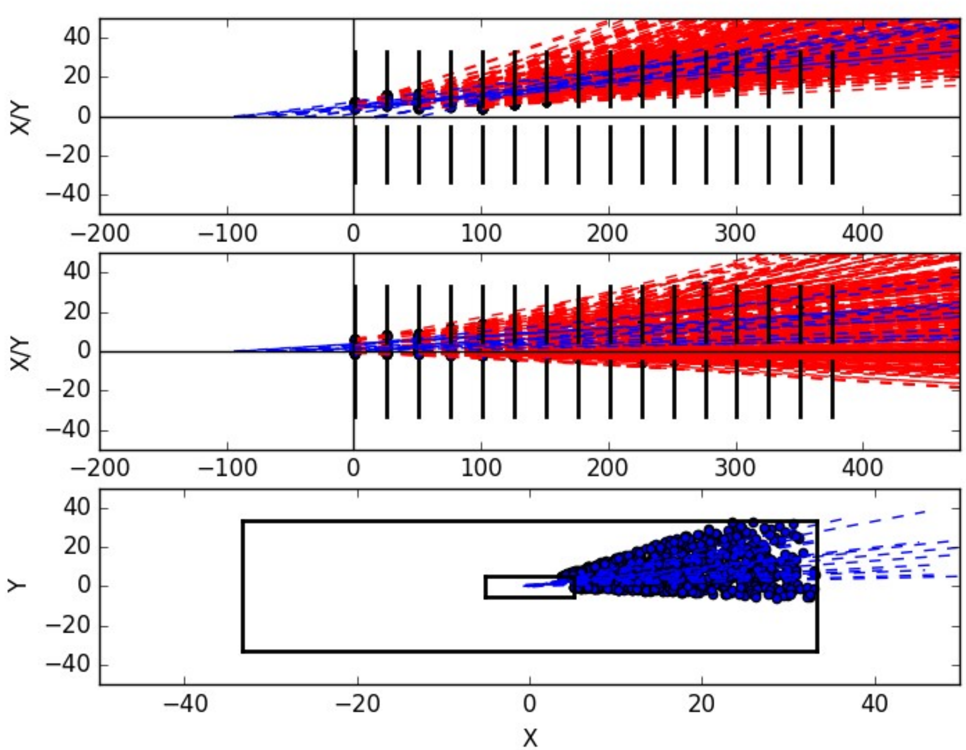
  - **f = 200 MHz** is the processing clock,
  - **<Occ.> = 50%** is the fraction of Stub Makers processing data,
  - **<N> = 40** is the number of identified pre-stubs from each Stub Maker

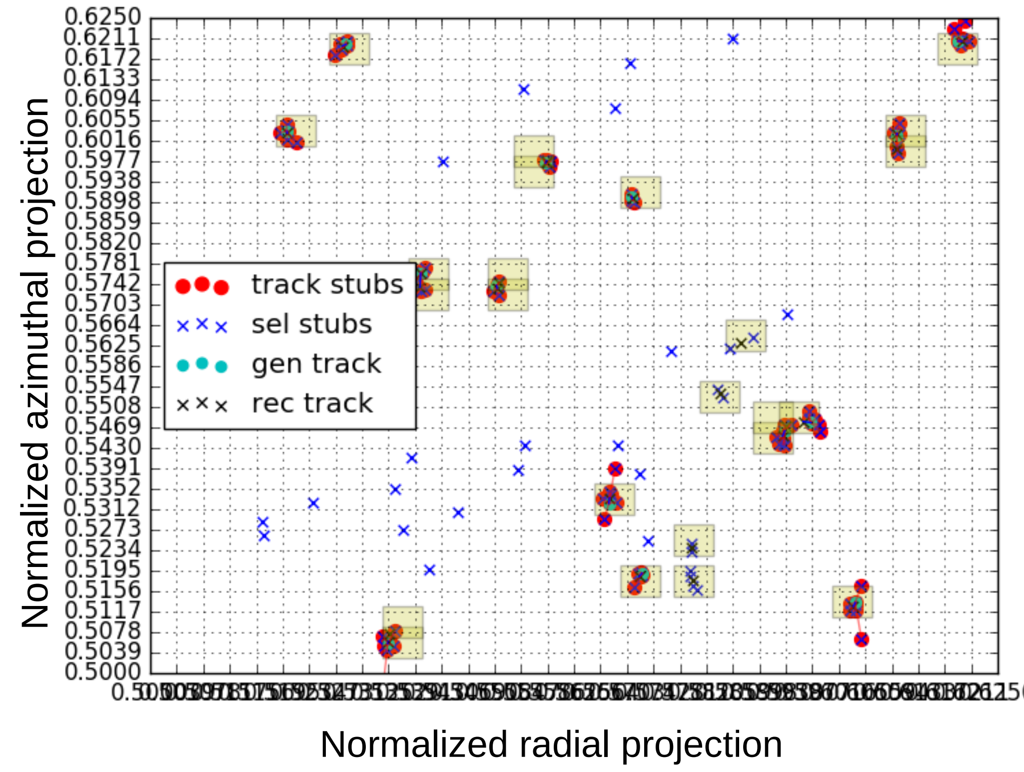
- With this number the relation is not satisfied, and **we obtain a processing rate of 10 MHz**
  - a simple way to satisfy the relation is to **increase the number of combinatoric processes** within the Stub Makers **by a factor 4** (see fig.)



- Comment:
  - the estimation is based on a processing clock f=200MHz, this value could also be increased to enhance the acceptable event rate

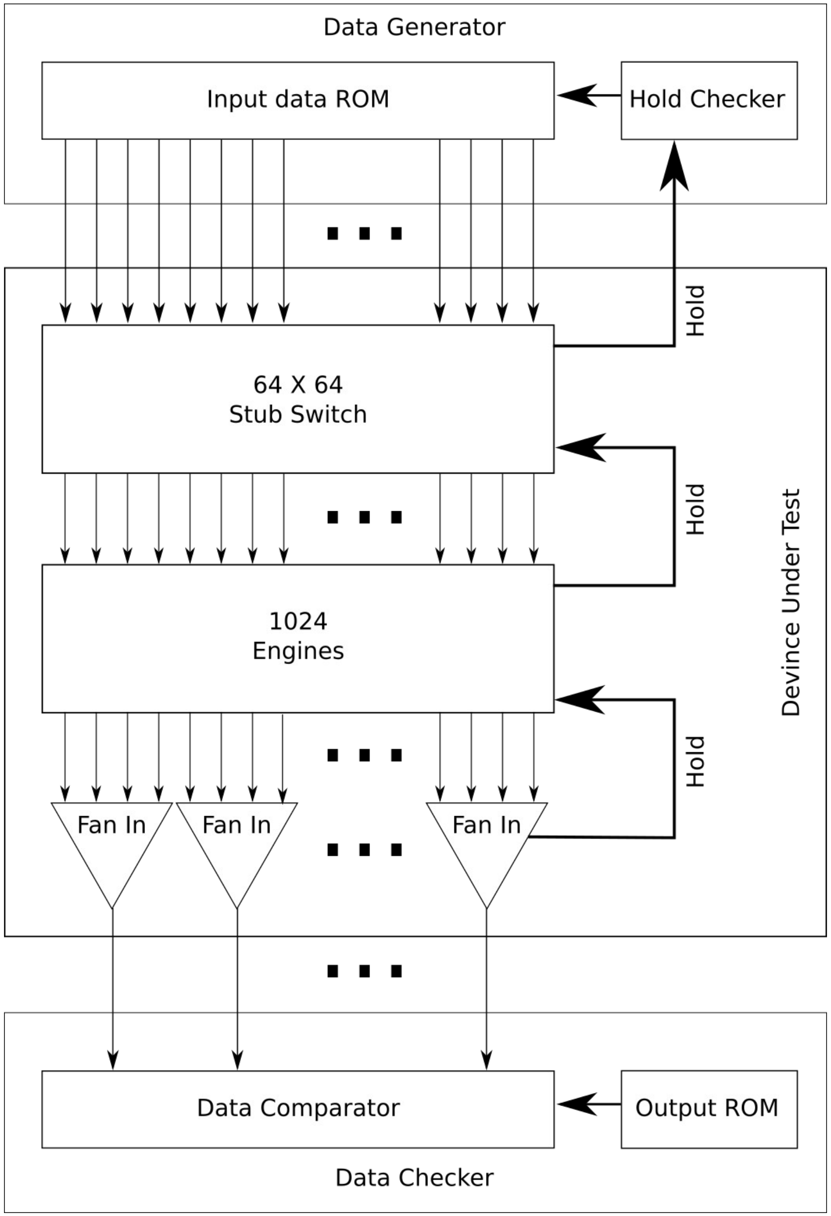
Switch + Engine test





- **The main space** of track parameters (radial and azimuthal coords of stub projection) has been divided in **8x8 (r,phi)-sectors.**
  - **already filtered stubs**, evaluated in software simulation, have been used to perform the following test
- The **system is modular and scalable** thanks to the absence of “lateral communication”
  - **results from one sector test can be extended to the full system**

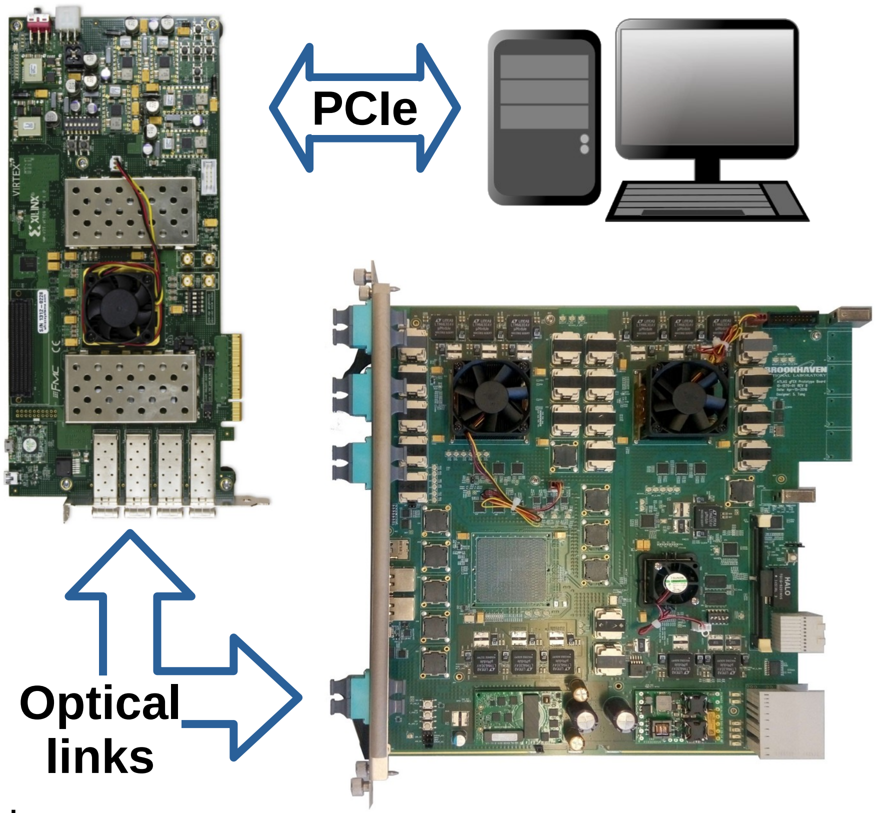
Switch + Engine test (2)



- **Architecture test, implemented in the same FPGA**, on the gFEX board with a system clock: **320 MHz**
- **3 Blocks:**
  - Data generator
  - Device Under Test (**Switch+Engines+Fanins**)
  - Data checker
- **Data Generator provides stub data stored in a ROM,** filled with data associated to 398 simulated events in a **1/64 sector of the whole detector**
- The **Data Checker compares the received track data with the values stored in a ROM**, populated with the expected results
- **Data processed in hardware are compatible with data stored in the Output ROM**, evaluated from behavioral simulation.
- The obtained event processing throughput is **40.9 MHz**

Next steps

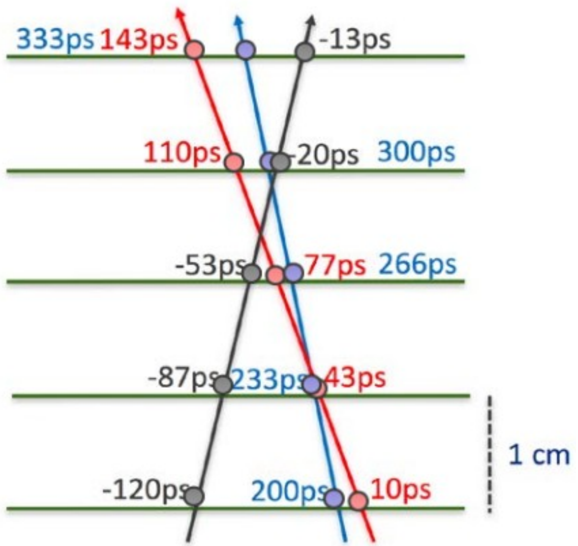
- **Full system test:**
  - software generation of **simulated data on PC**
  - data transferred to the VC709 board via PCIe interface
  - **stub generation in the VC709**
  - data transferred to the **gFEX** via optical links and **track reconstruction** onboard
  - **reconstructed tracks transferred to the PC,** via the the VC709 board

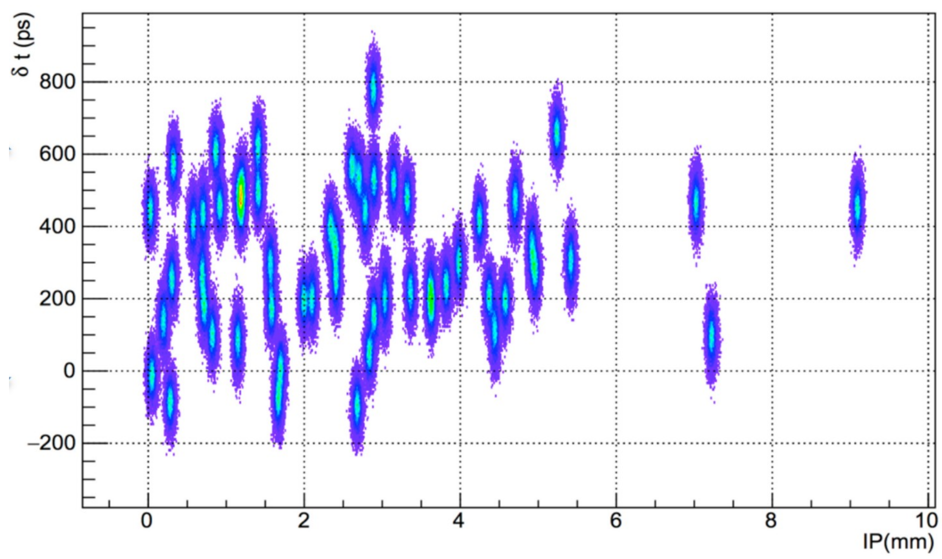


- **Collaboration with INFN Bologna**, within TIMESPOT project
  - development of a **LHCb VELO U2 simulation** including timing information of the hits
  - development of clustering algorithm in FPGA
- **Upgrade from VC709 to Xilinx VCU128:**
  - increased resource and transfer rate up to 1 Tbps

Role of timing

- **Timing in track reconstruction:**
  - **hits distributed in time** with RMS ~ 170 ps
  - need ~30 ps hit resolution to discriminate hits with similar positions, based on time
  - **use only compatible hits** in the pattern recognition

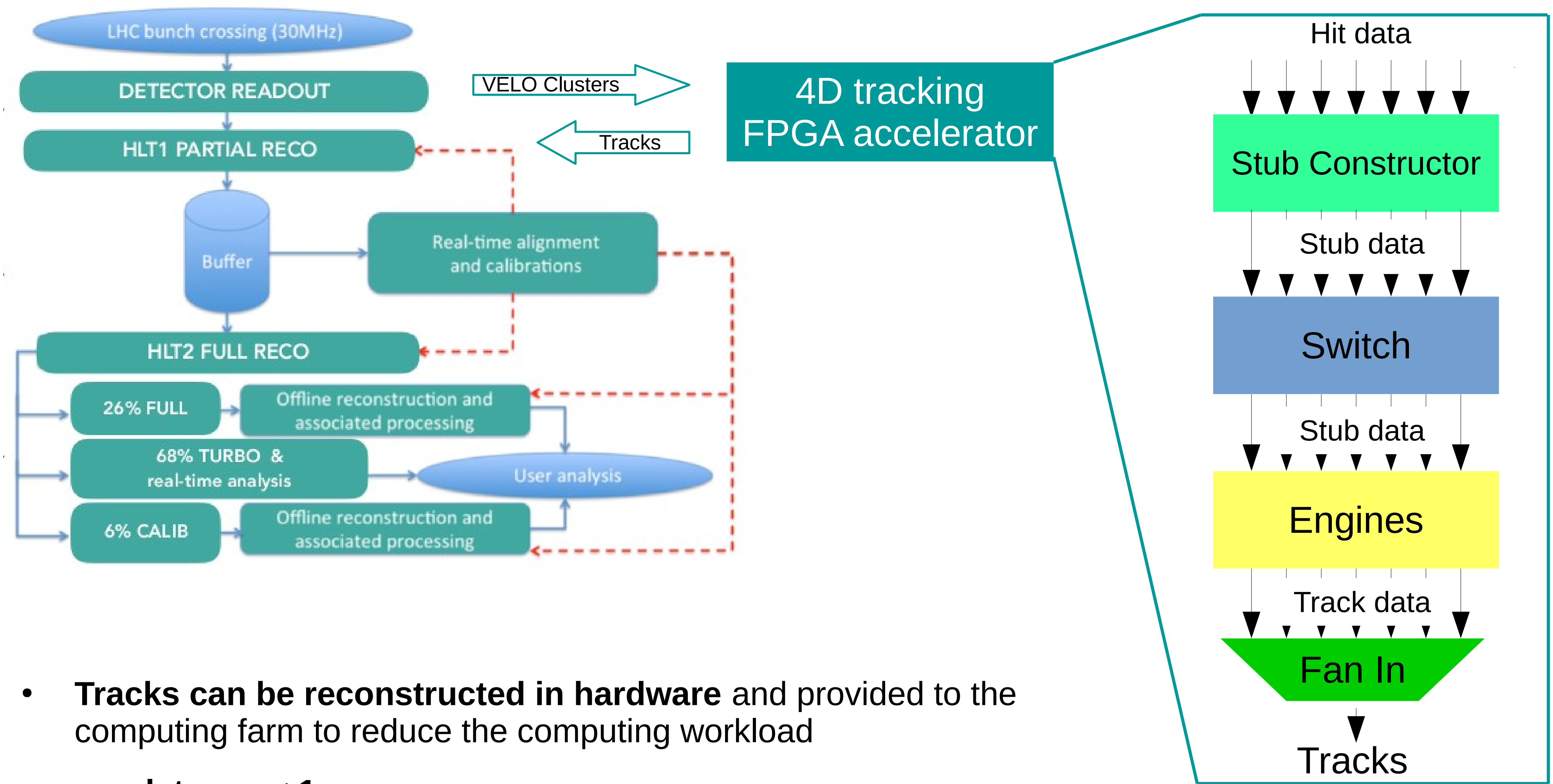




- **Timing in PV-association:**
  - additional power to select the correct primary vertex
  - hit time resolution needed: ~200 ps
  - enough to separate tracks with similar spatial parameters

- **Others:**
  - precise measurement of time of primary vertices
  - track time stamping for better association of track upstream and downstream the magnet
  - timing in PID and calorimetry
  - **reduce data to process in HLT by selecting only interesting pp interactions**

FPGA Accelerator in LHCb



- **Tracks can be reconstructed in hardware** and provided to the computing farm to reduce the computing workload
  - **latency < 1 μs**
- Tracks provided with the **collection of hits** for refined fit in software

Conclusions

- **3D silicon sensors with “Trench” pixel** geometry provided unprecedented timing performance, with **up to 30 ps timing resolution** from tests using a 270 MeV/c π+ beam at PSI.
- **28 nm CMOS front end, with high precision TDC**, is being developed. Electronics, at the moment, is the limiting factor and optimization of the design is ongoing

- **A “Stub” based fast track finding device** has been developed for a possible application to an Ugrade II LHCb VELO-like detector
- The system has been **implemented and tested in FPGA**:
  - stub identification and construction strategy tested in high-level simulation
  - main components of the Stub Constructor (Stub Makers) tested in hardware, using a **system clock of 200 MHz**. “Assembly” and test of the full Stub Constructor ongoing
  - Switch + Engines tested in low-level simulation and hardware
  - **tested at 40 MHz event rate**, using a system clock of 320 MHz