First prototype of a tracking system with "artificial retina"

Nicola Neri
INFN, Sezione di Milano

TIPP 2014
Amsterdam, The Netherlands
2-6 June 2014

Contributors

A. Abba[1], F. Bedeschi[2], F. Caponio[1], M. Citterio[1], S. Coelli[1], J. Fu[1], A. Geraci[1], P. Marino[2], M. Monti[1], M. J. Morello[2], N. Neri[1], D. Ninci[2], A. Piucci[2], M. Petruzzo[1], G. Punzi[2], L. Ristori[2,4], F. Spinella[2], S. Stracka[2], D. Tonelli[3], J. Walsh[2]

[1] INFN-Milano/Politecnico, [2] INFN-Pisa/University/SNS, [3] CERN, [4] Fermilab

Outline

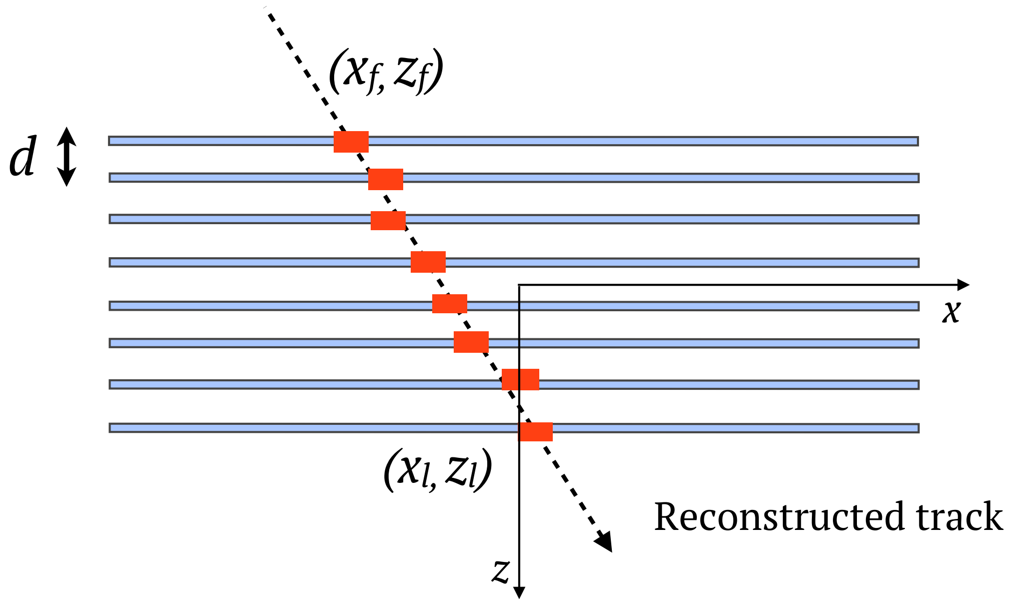
- Retina algorithm
- Silicon sensor telescope
- Retina architecture
- Tracking performance
- Perspectives for future applications

“Artificial retina” prototype

- The *“artificial retina”* algorithm for fast track finding was originally proposed by *Luciano Ristori* in *NIM A 453 (2000) 425-429.* It takes inspiration from neurobiology[1] and applies to positions sensitive detectors (*e.g.* pixel, strip detectors).
- Here we present *the first prototype of a tracking system with “artificial retina”* based on 2D tracks, for simplicity sake, implemented using commercial FPGAs of TEL62 boards
- The prototype maximal rate for track reconstruction is about 1 MHz. We plan to perform functionality tests using cosmic rays
- The “artificial retina” is modular system that can be designed to work for HEP applications, *i.e.* high rates and large detectors, providing offline-like track quality results with a latency <1μs. See LHCb-PUB-2014-026

[1] *See also M. M. Del Viva talk at TIPP14*

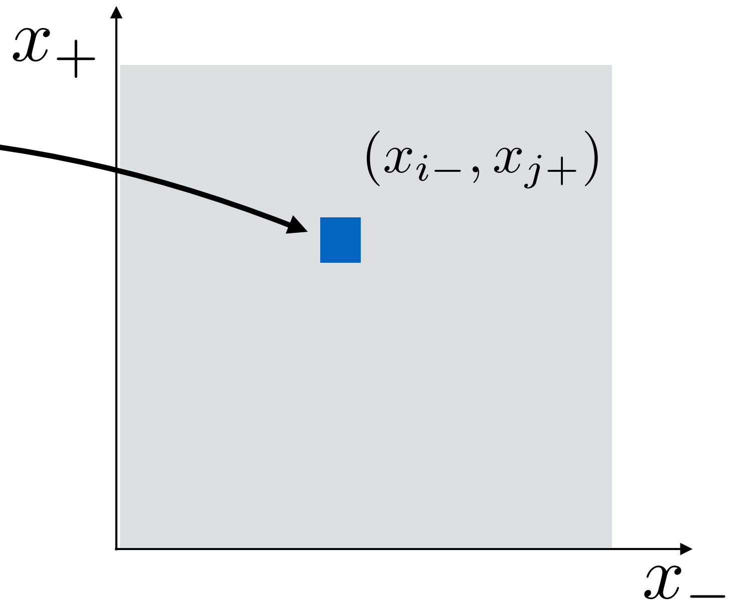
Track parameter definition



Track parameters

$$x_{\pm} = \frac{x_l \pm x_f}{2}$$

Grid of track parameters

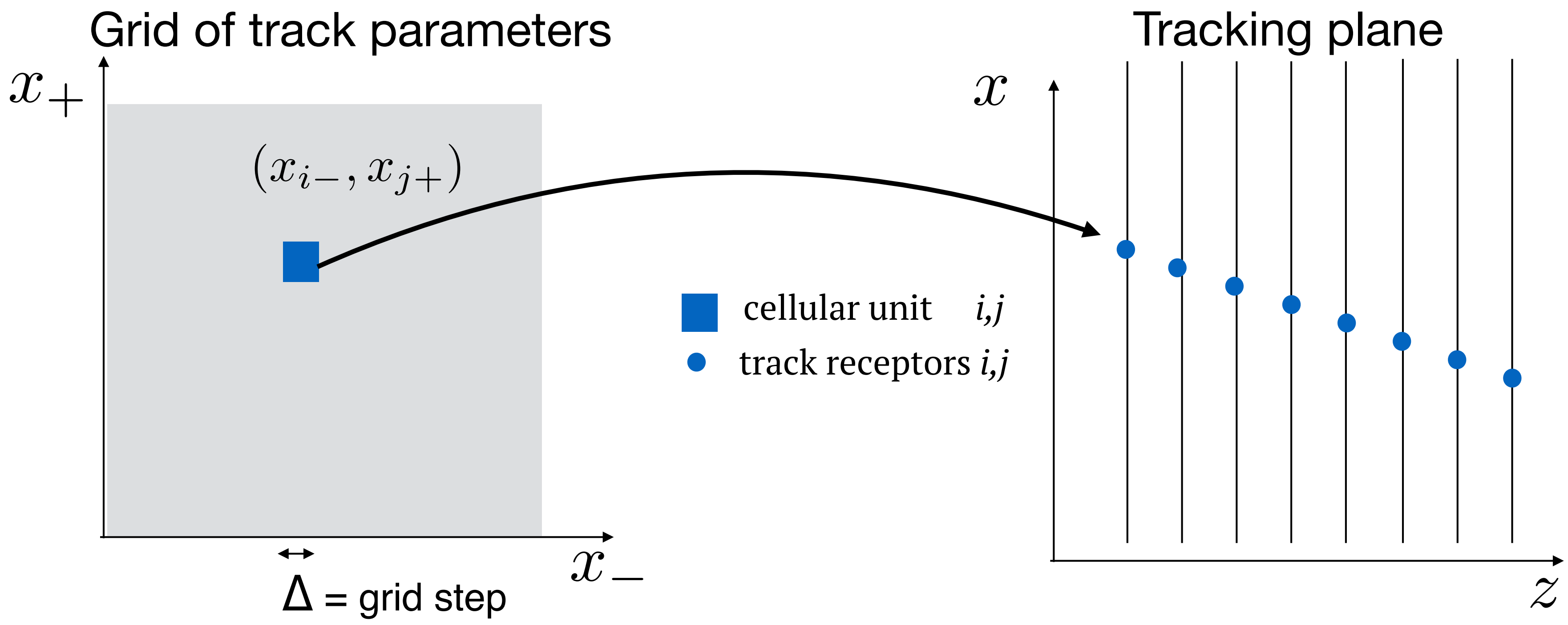


A 2-dim track is described as

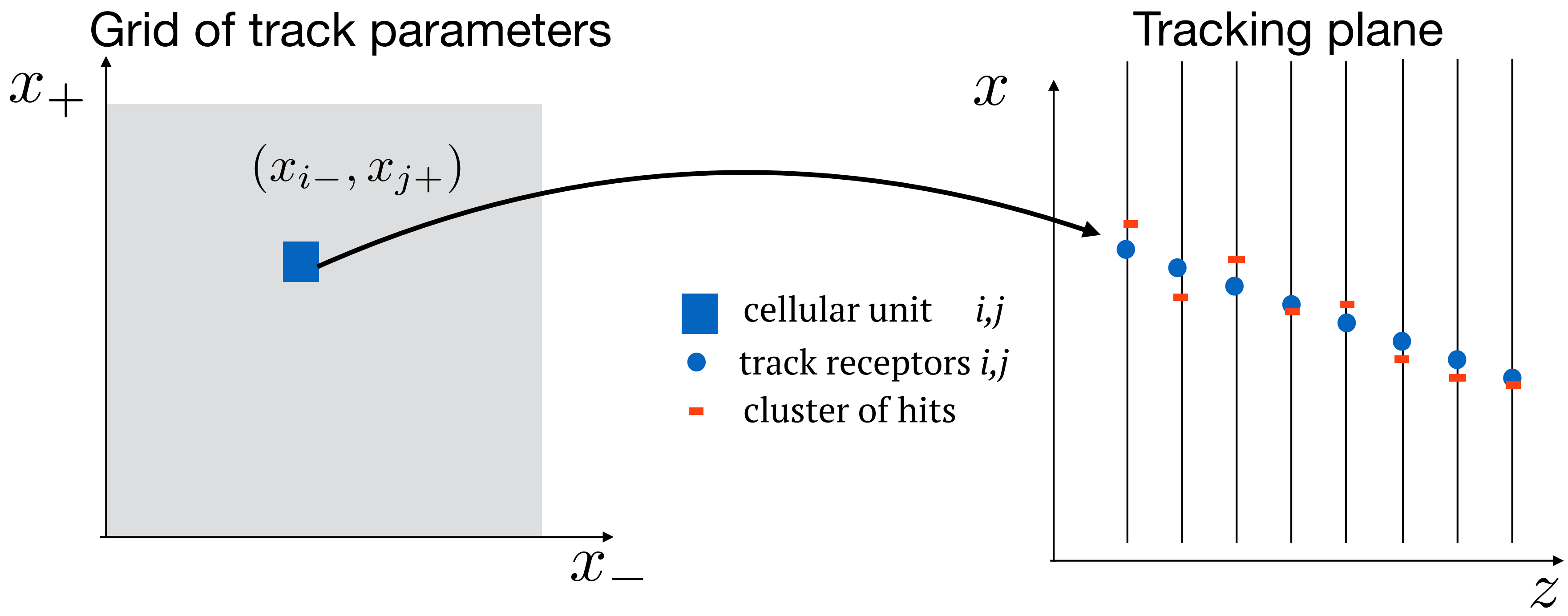
$$x(z) = x_+ + x_- \frac{z - z_+}{z_-}$$

$$z_{\pm} = \frac{z_l \pm z_f}{2} \qquad \text{(constant terms)}$$

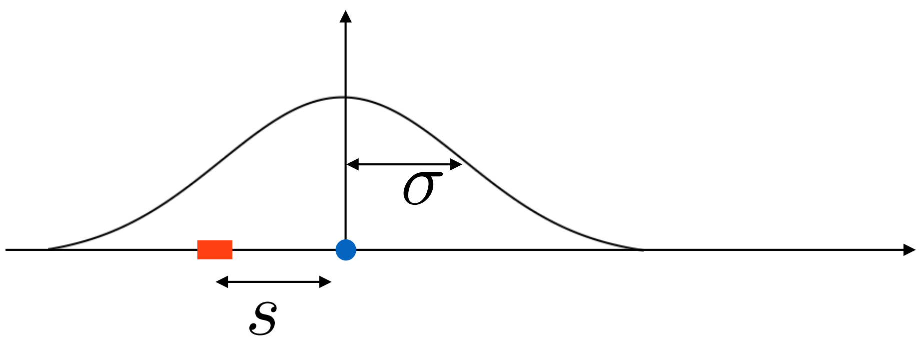
Retina receptors



Retina receptors



Receptor response

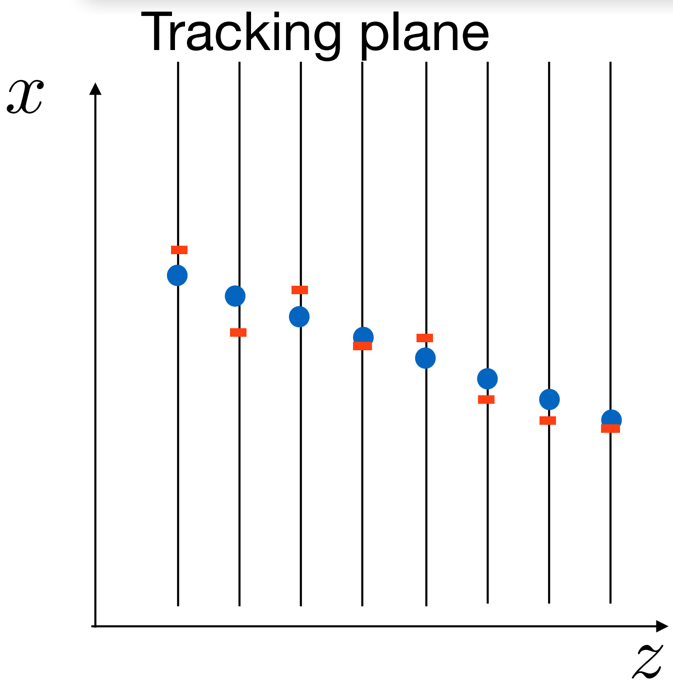


σ = width of the receptor response field.

$\sigma \simeq \Delta$ grid step

It is much larger than the obtainable resolution on track parameters and has to be tuned.

Retina algorithm



s_{ijk} distance between cluster in layer k and track receptor i,j

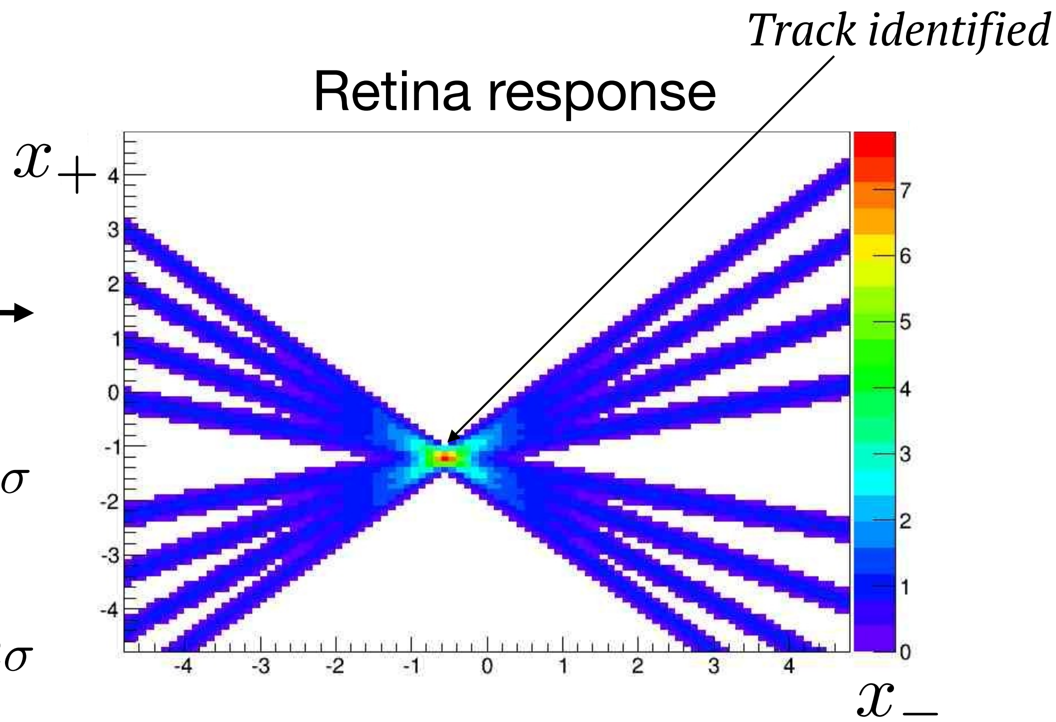
$$s_{ijk} = \left| x_k - x_{j+} - x_{i-} \frac{z_k - z_+}{z_-} \right|$$

- cluster of hits
- track receptors i,j

Excitation of the cellular unit i,j

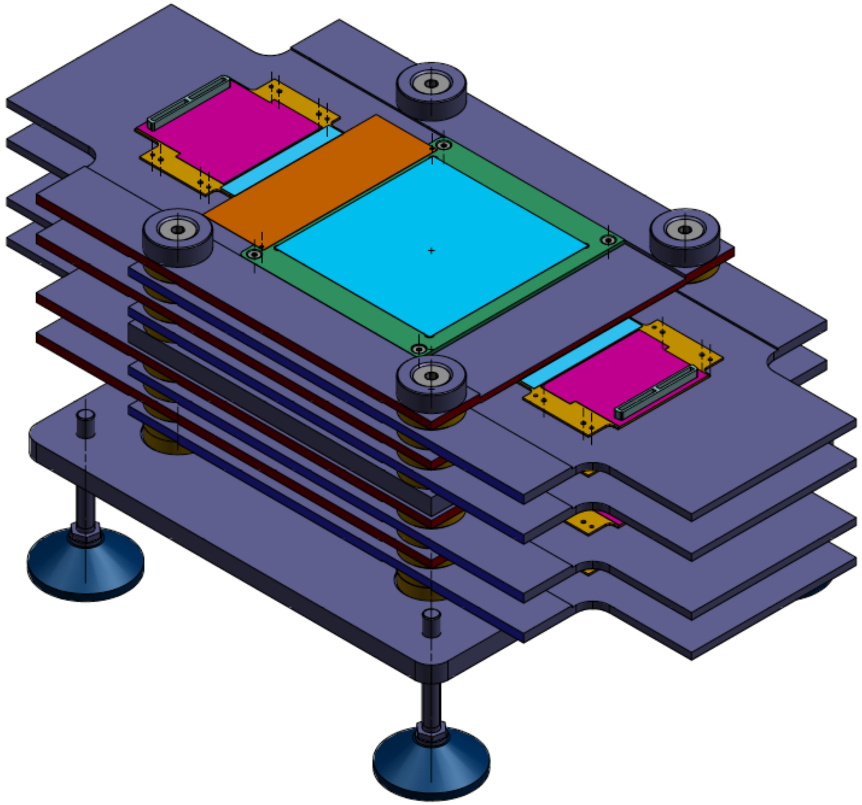
$$W_{ij} = \sum_k \exp\left(-\frac{s_{ijk}^2}{2\sigma^2}\right) \quad \text{if } s_{ijk} < 2\sigma$$

$$W_{ij} = 0 \quad \text{if } s_{ijk} > 2\sigma$$



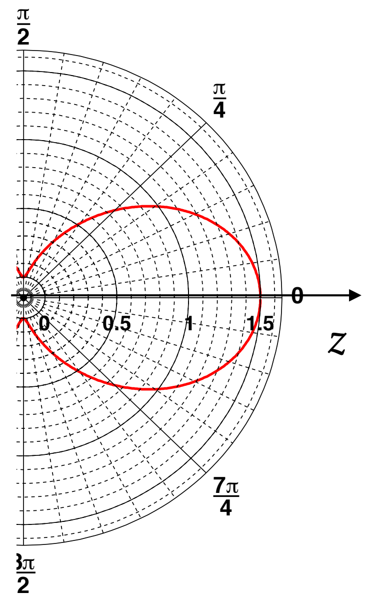
Hardware implementation

Silicon strip telescope

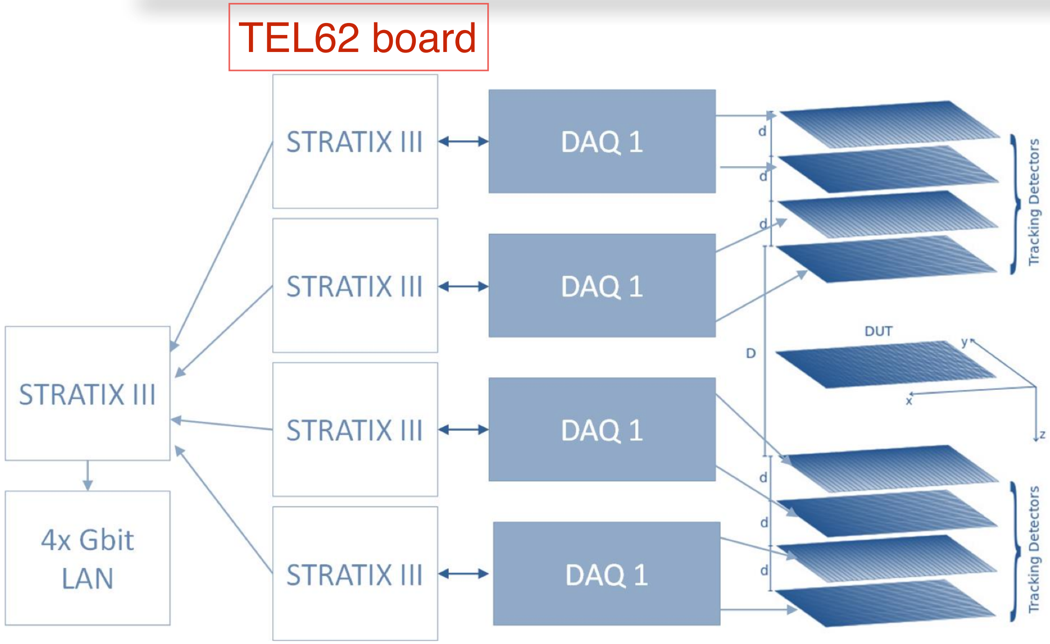


- 8 plane telescope made of single-sided silicon sensors
- STM OB2 sensors: ~10 cm x 10 cm active area, 512 strips with 183 μm pitch and 500 μm thickness
- 0.8 cm distance between planes
- expected rate from cosmic rays is of the order of 1 Hz assuming I ≈ 1 cm^{-2} min^{-1} for horizontal detectors

$$\frac{dN}{d\cos\theta} \propto \cos^2\theta$$



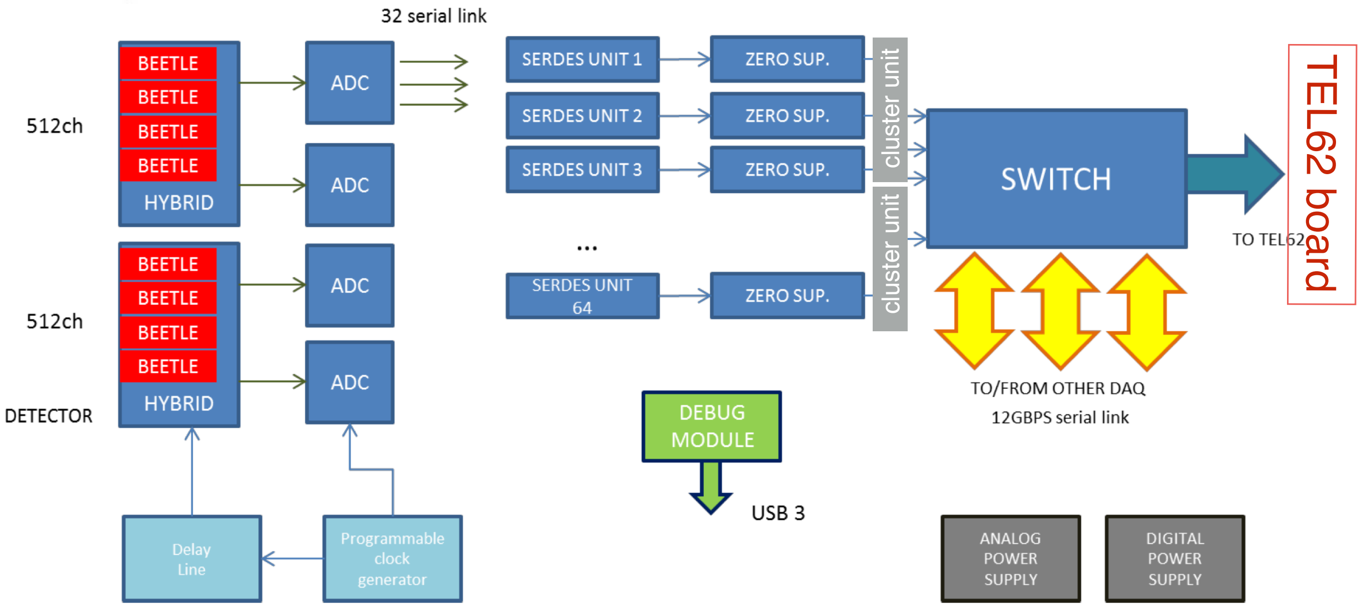
Data acquisition system



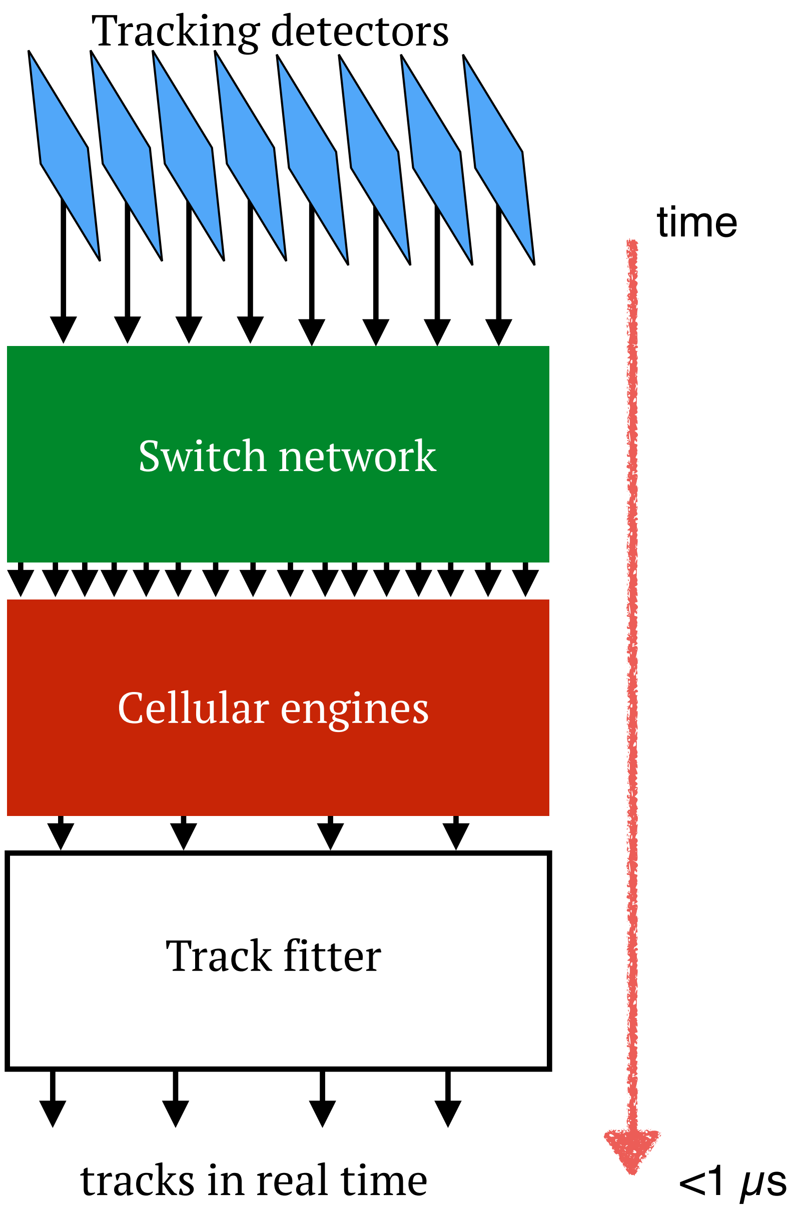
See F. Caponio, RT2014 presentation "The Readout Architecture for the Retina-Based Cosmic Ray Telescope"

- Custom DAQ board based on Xilinx Kintex 7 FPGA: 1 board for 2 detectors readout.
- Maximal readout rate 1.25 MHz
- Data delivered to 4 Altera Stratix III FPGAs of the TEL62 board for data processing and retina response

- 4 Beetle chips for each detector (128x4=512 channels).
- Digitalization with multichannel 12-bit ADC and zero suppression (threshold comparator).
- Clustering of adjacent strip hits
- Data routed to appropriate processing units by the switch



Retina architecture



The retina architecture has been implemented on commercial FPGAs: Altera Stratix III

Two level switch: 1st level in custom DAQ board and 2nd level in TEL62 board. Deliver data according to coordinate information to appropriate processing units, "engines"

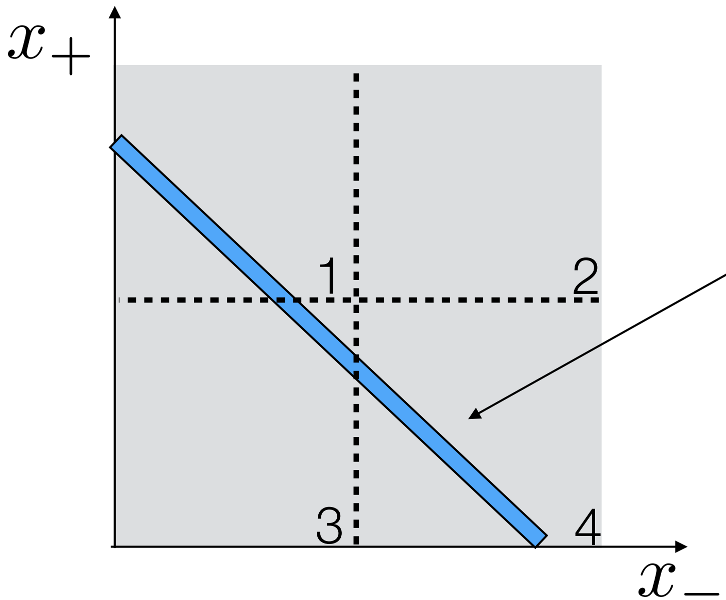
Engines are devoted to the the calculation of the cell excitation level, "weight"

Track parameters determined by interpolation of the response of the retina between adjacent cells of the local maximum

Delivering the data

- Engines receives data, through the switch, from all the tracking detectors
- Divide the grid in 4 regions corresponding to the number of available FPGAs (4 Altera Stratix III) for the processing engines.

A cluster seen by the retina



A cluster *(x,z)* corresponds to a line in the grid of parameters (x_+, x_-)

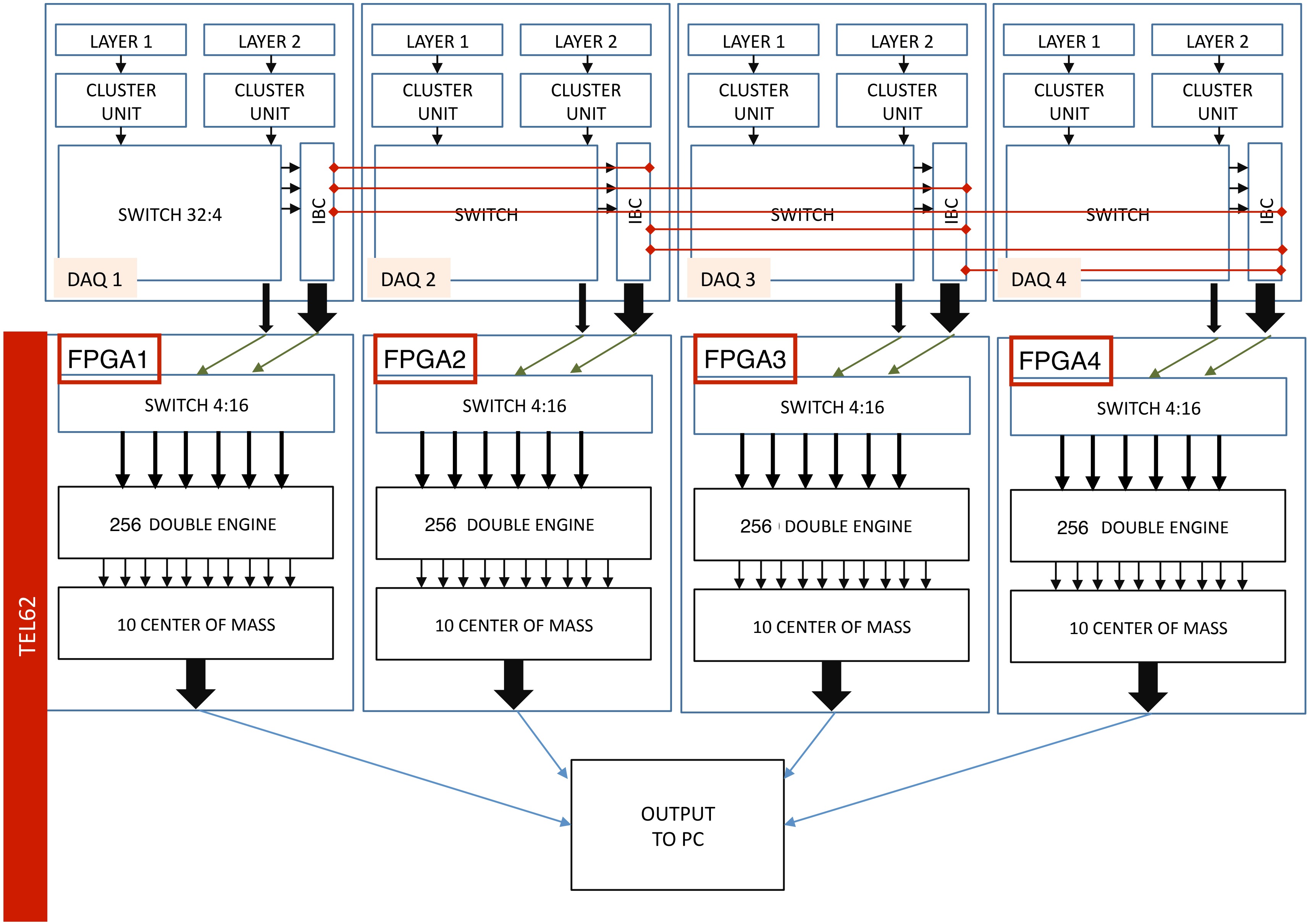
$$x_+ = -x_- \frac{z - z_+}{z_-} + x$$

- Engines with non negligible weights belong to different regions of the grid

$$|x - x_+ - x_- \frac{z - z_+}{z_-}| < 2\sigma$$

- Deliver the data to the engines (in different FPGAs) using a *full mesh switch*
- *z* determines the slope and *x* the intercept with x_+ axis of a cluster in the (x_+, x_-) plane
- Data path is determined by the cluster coordinates *(x,z)* using 8 bit information: 5 bit for *x* and 3 bit for *z*

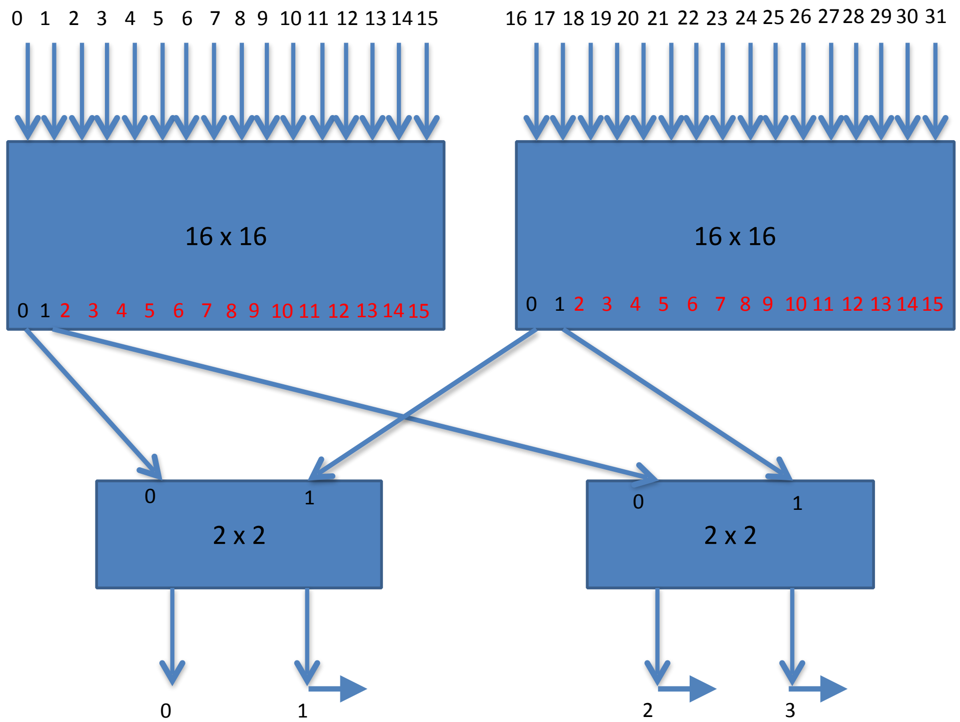
Detailed implementation



Switch modules

1st switch 32x4 in DAQ board

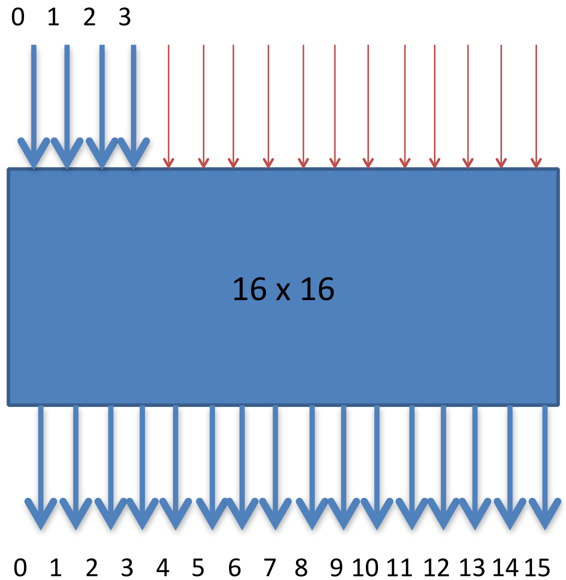
16 analog inputs from each sensor.



4 output ports: 1 output to 2nd switch level and 3 output to the other DAQ boards

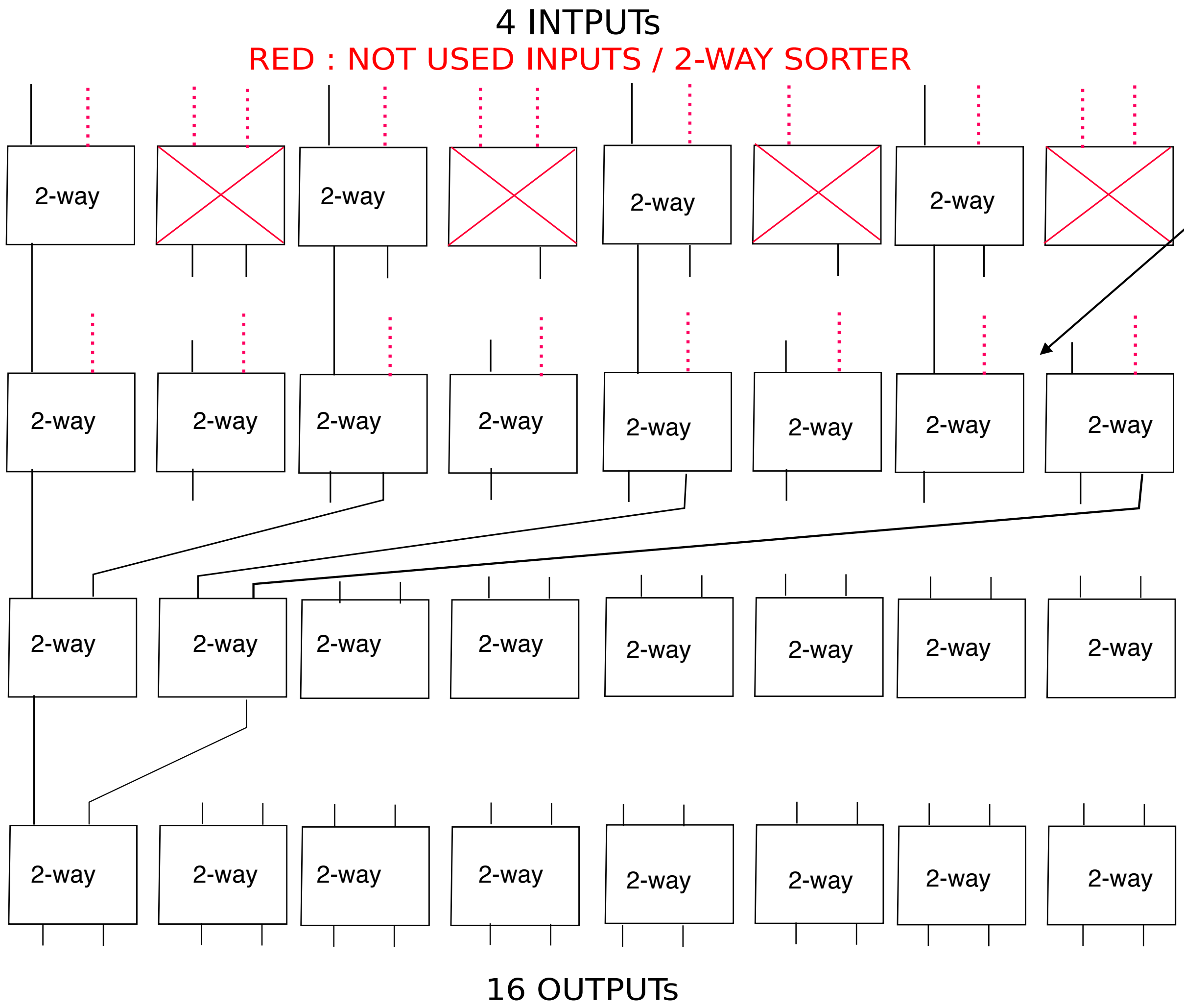
2nd switch 4x16 in TEL62 board

4 input ports: 1 from each DAQ board

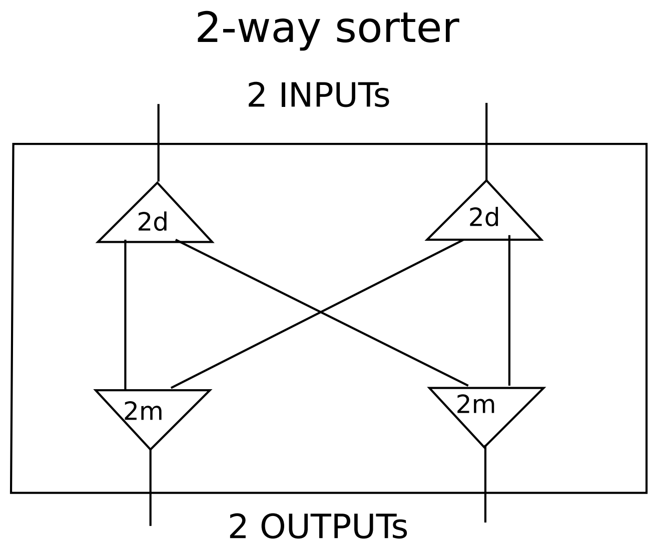


16 output ports, each connected to 16 engines in parallel

Details of the 4x16 way dispatcher

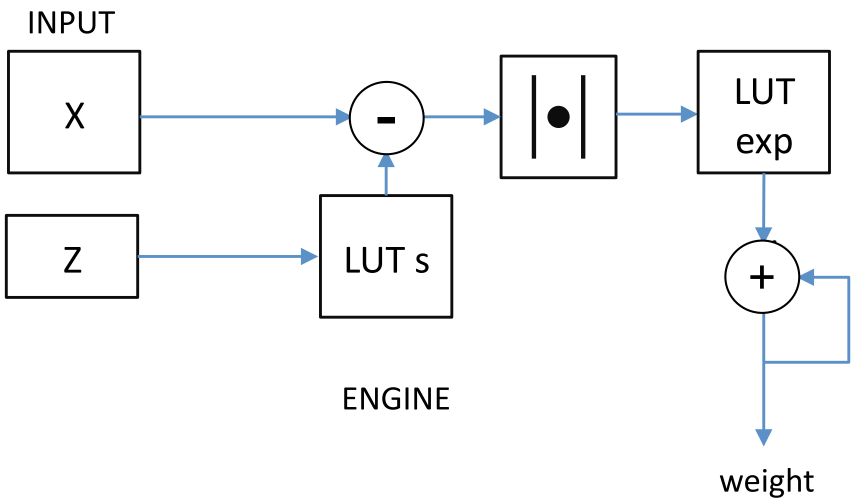


- Each box is a programmable two way sorter.
- Each input can be delivered to the left, to the right or both output ports according to LUT information.



16 engines each output = 16x16=256 engines

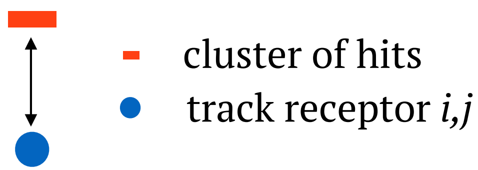
Engine module



Each engine determines the excitation level of 2 adjacent cells in the grid of tracking parameters x_+, x_-.
N cells = 2 N engines

Fit 256 engines x FPGA (using 50% Altera Stratix III)
4x256=1024 engines = 2048 cellular units for the retina

- A 10x16 bit LUT for the calculation of the distance s_{ijk}



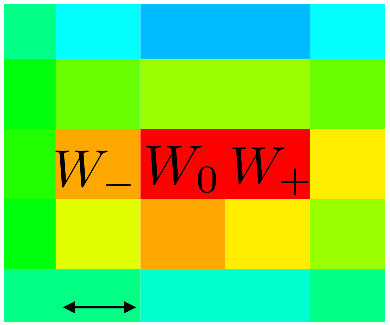
$$s_{ijk} = \left| x_k - x_{j+} - x_{i-} \frac{z_k - z_+}{z_-} \right|$$

- A 10x16 bit LUT for the calculation of the exponential

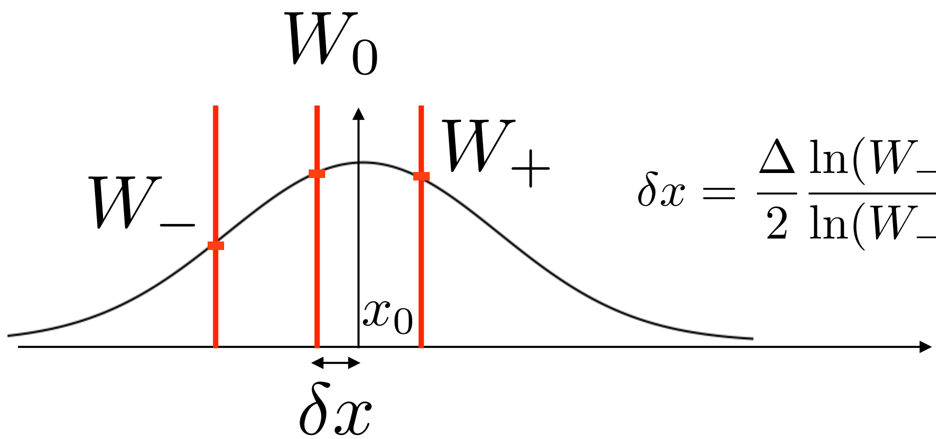
$$W_{ij} = \sum_k \exp\left(-\frac{s_{ijk}^2}{2\sigma^2}\right) \qquad \text{if } s_{ijk} < 2\sigma$$

Track parameter determination

- Identify the engine with the max local weight
- Weights of the maximum and adjacent cells to the Track Fitter unit for interpolation
- Gaussian interpolation for track parameters: ln(x) stored in a 10x16 bit LUT



Δ = grid step



$$\delta x = \frac{\Delta}{2} \frac{\ln(W_-/W_0) - \ln(W_+/W_0)}{\ln(W_-/W_0) + \ln(W_+/W_0)}$$

- Similar results have been obtained with Center of Mass calculation and then applying a correction factor

$$\bar{x} = \frac{\sum_i W_i x_i}{\sum_i W_i}$$

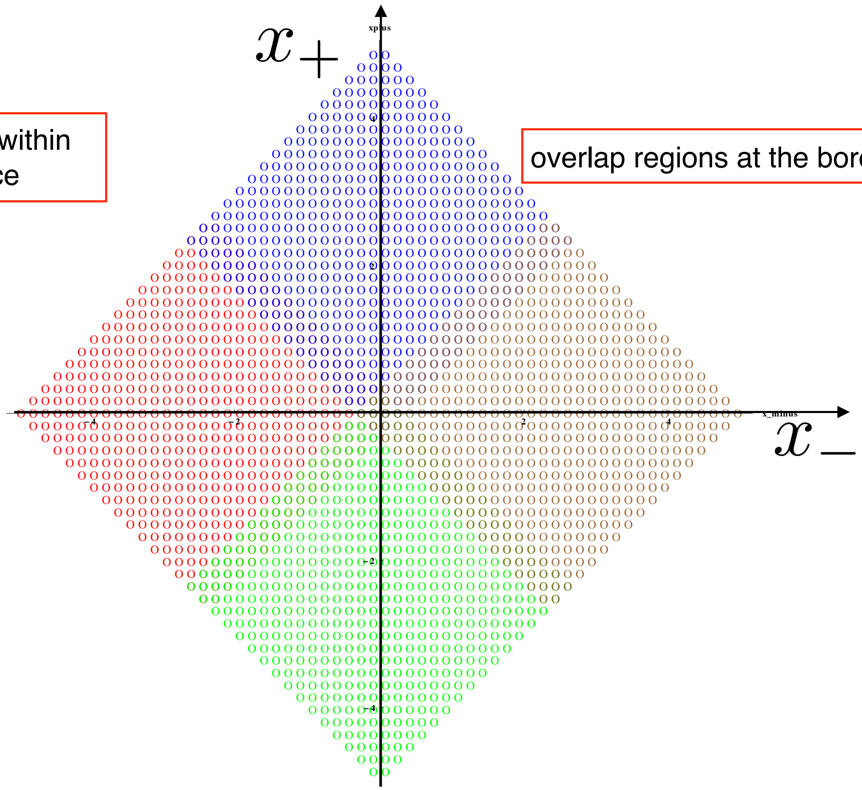
$$\delta x = \alpha(\bar{x} - x_0)$$

Resources and latency

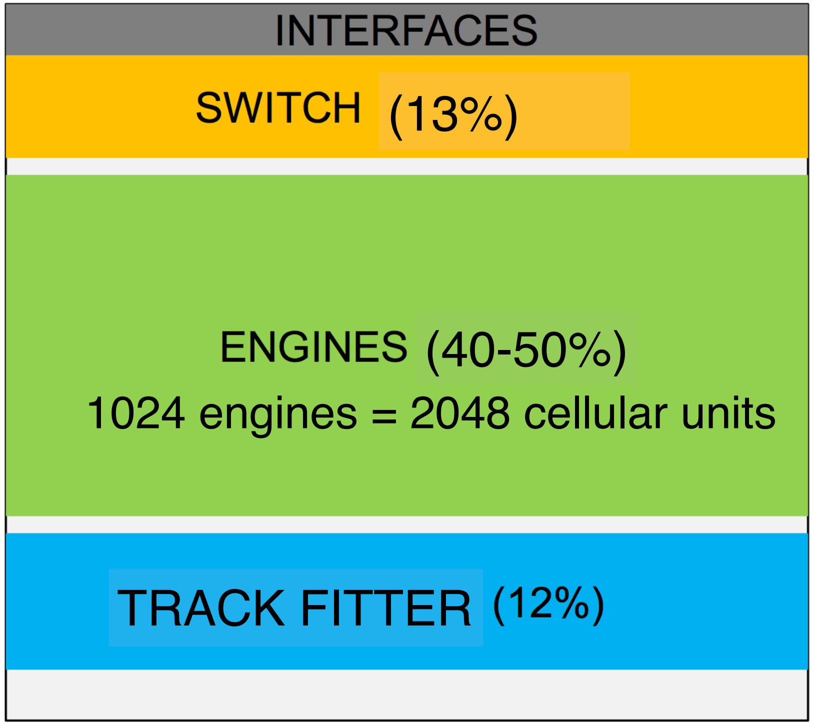
- Here we present a solution based on 4 Altera Stratix III FPGAs

grid of cellular units within telescope acceptance

overlap regions at the borders

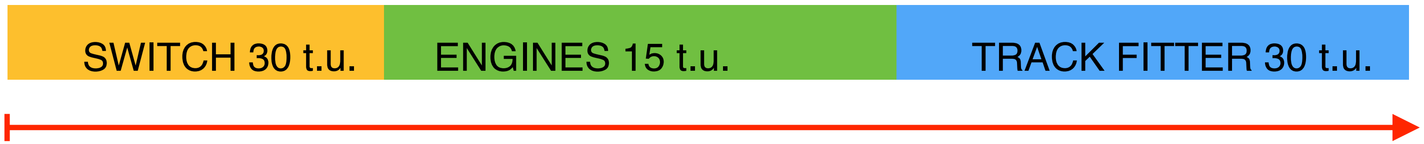


Altera Stratix III



25-35% BACKUP

Latency of Retina response

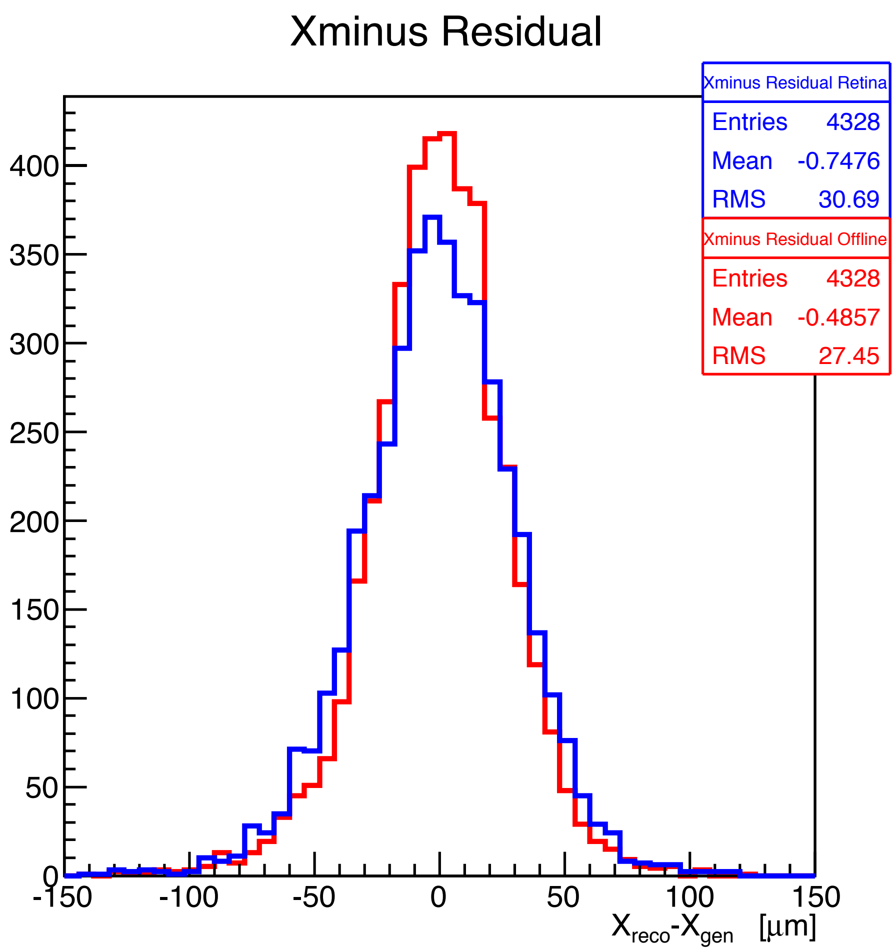


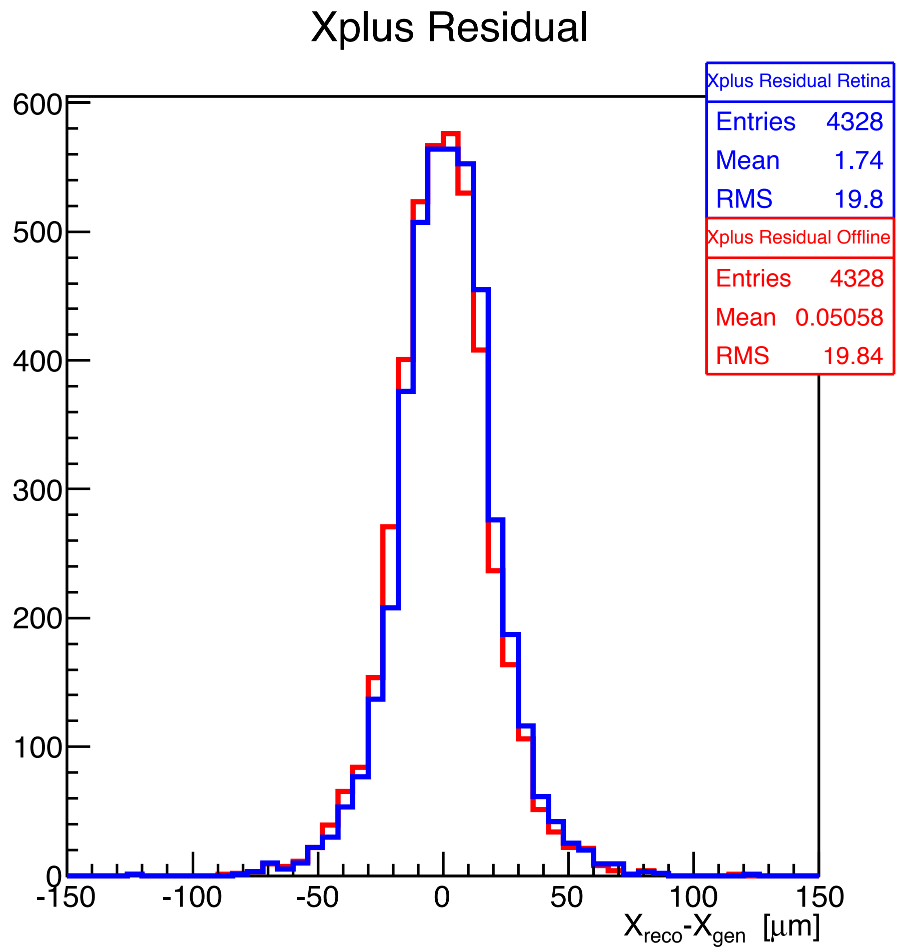
<100 t.u.=0.5 μs

1 t.u. at 200 MHz clock = 5 ns

Tracking performance

- Track parameter resolution with retina (blue) is comparable with offline results (red)





- We have completed our design simulation and we are ready to implement everything in hardware in the next months
- Let’s briefly talk about the perspectives for the future

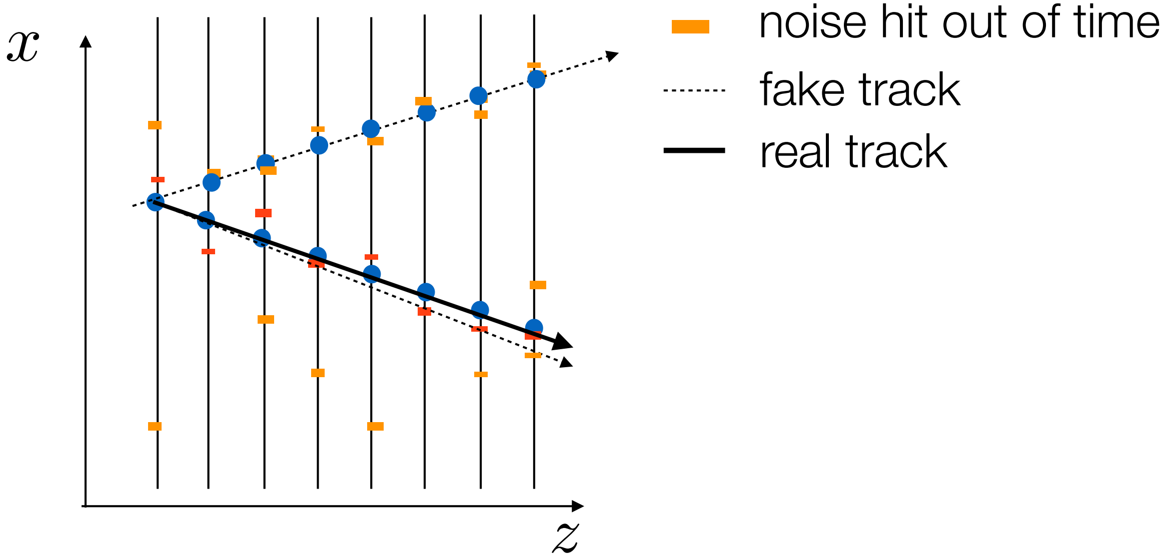
Using precise time information of the hit

- Time information could be used to further suppress noise hits
- R&D on ultrafast silicon pixel detectors aims to achieve ~10-20 ps time resolution. **JINST 9 (2014) C02001**

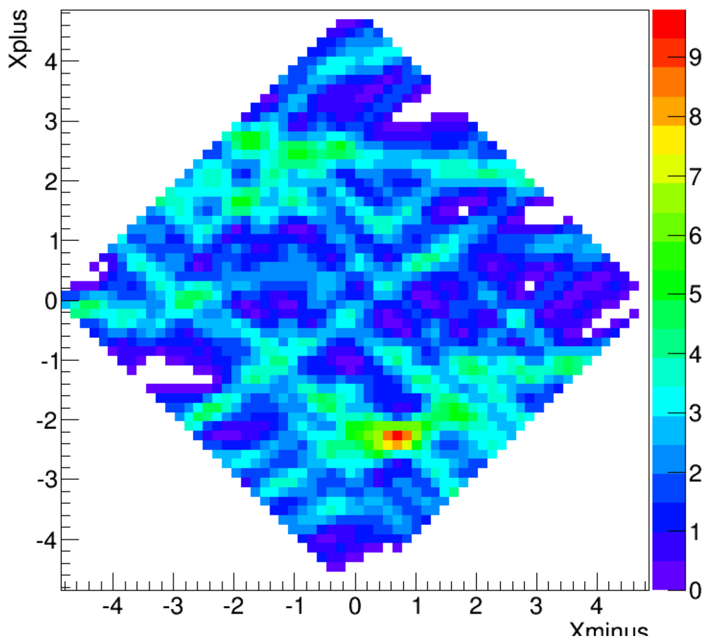
$$W_{ij} = \sum_k \exp\left(-\frac{s_{ijk}^2}{2\sigma^2}\right) \exp\left(-\frac{t_{ijk}^2}{2\sigma_t^2}\right)$$

$$t_{ijk} = (t_{k,meas} - t_{ijk,exp})$$

Retina with spatial information



No time information



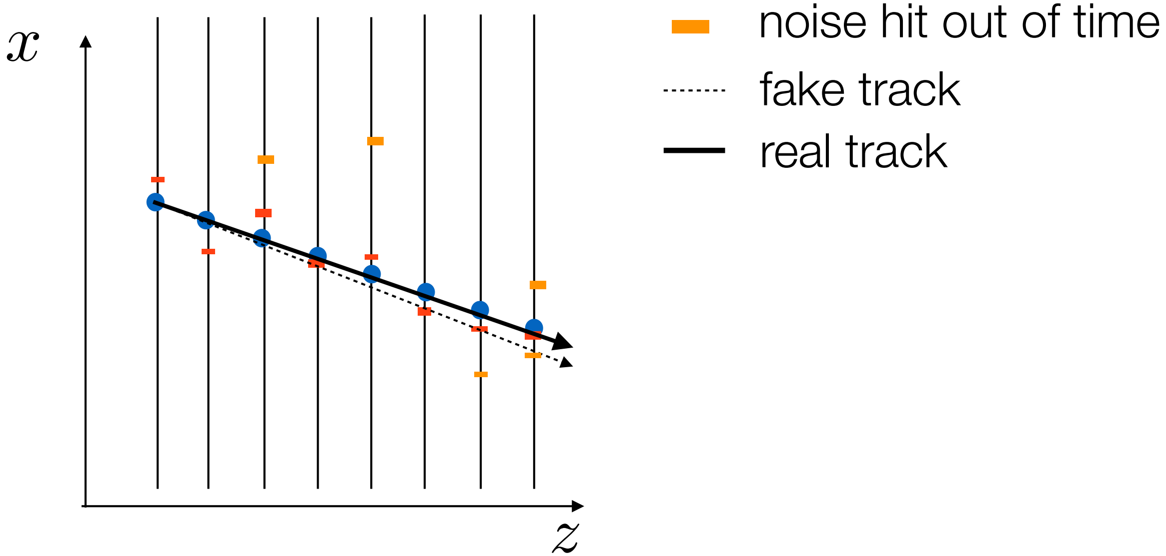
Using precise time information of the hit

- Time information could be used to further suppress noise hits
- R&D on ultrafast silicon pixel detectors aims to achieve ~10-20 ps time resolution. This is an interesting possibility for the future.

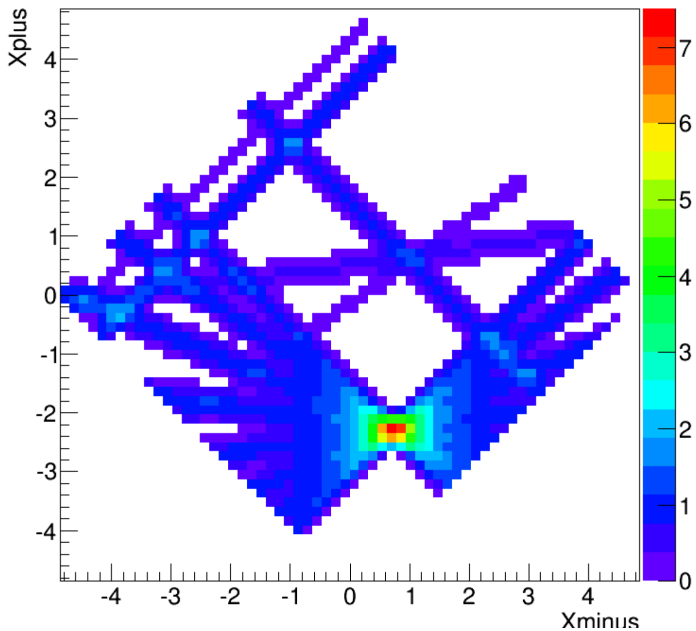
$$W_{ij} = \sum_k \exp\left(-\frac{s^2_{ijk}}{2\sigma^2}\right)\exp\left(-\frac{t^2_{ijk}}{2\sigma_t^2}\right)$$

retina can fit time of the track

Retina with spatial information and time information



time resolution 100 ps



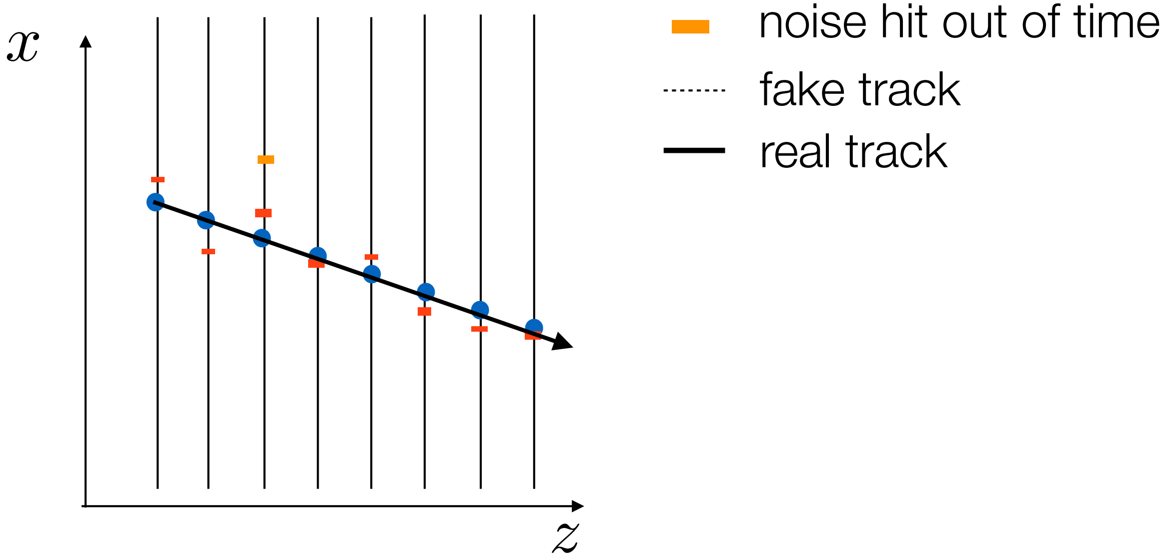
Using precise time information of the hit

- Time information could be used to further suppress noise hits
- R&D on ultrafast silicon pixel detectors aims to achieve ~10-20 ps time resolution. This is an interesting possibility for the future.

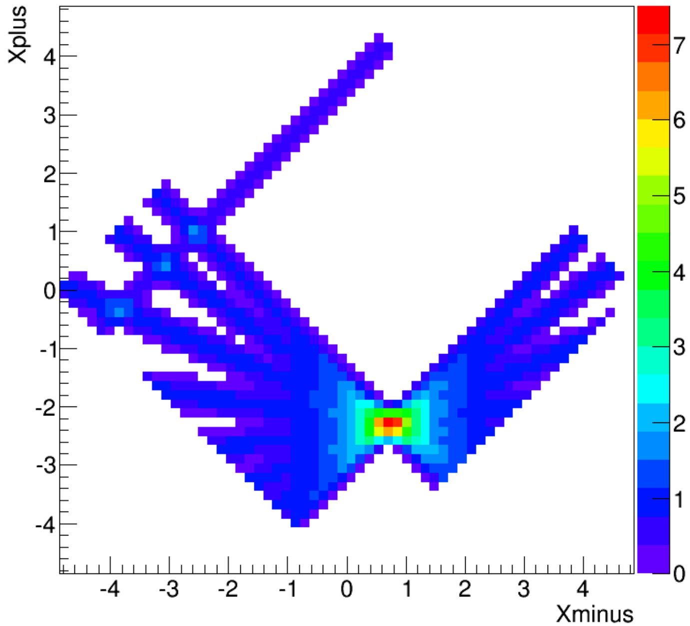
$$W_{ij} = \sum_k \exp\left(-\frac{s_{ijk}^2}{2\sigma^2}\right)\exp\left(-\frac{t_{ijk}^2}{2\sigma_t^2}\right)$$

retina can fit time of the track

Retina with spatial information and time information



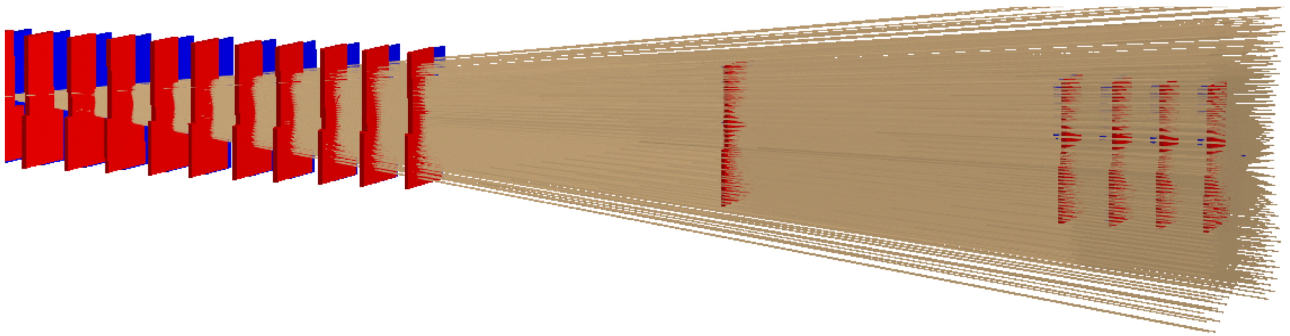
time resolution 10 ps



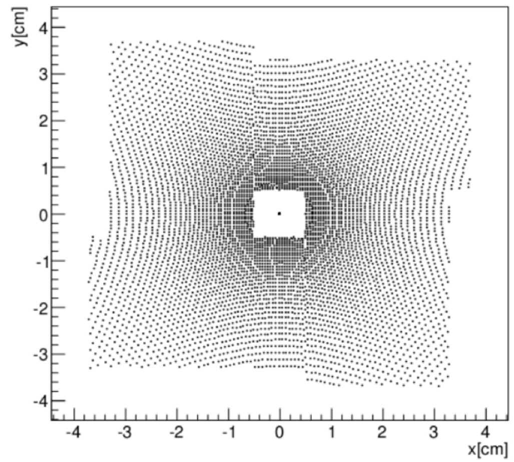
Possible applications in HEP

- The retina architecture is modular and more FPGA resources can be included to cope with high particle rates and large detectors, e.g. 40 MHz event rate and 300 tracks/event.
- An example of the application of the retina to HEP experiments has been documented here: LHCb-PUB-2014-026. *See also D. Tonelli's talk and G. Punzi talk at INSTR14.*
- Such a device can deliver tracks with offline-like quality with <1μs latency at 40 MHz

2 telescope with 10 tracking layers each



Grid of cellular units

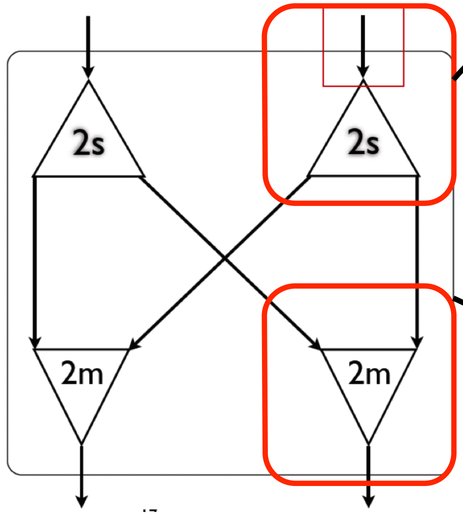


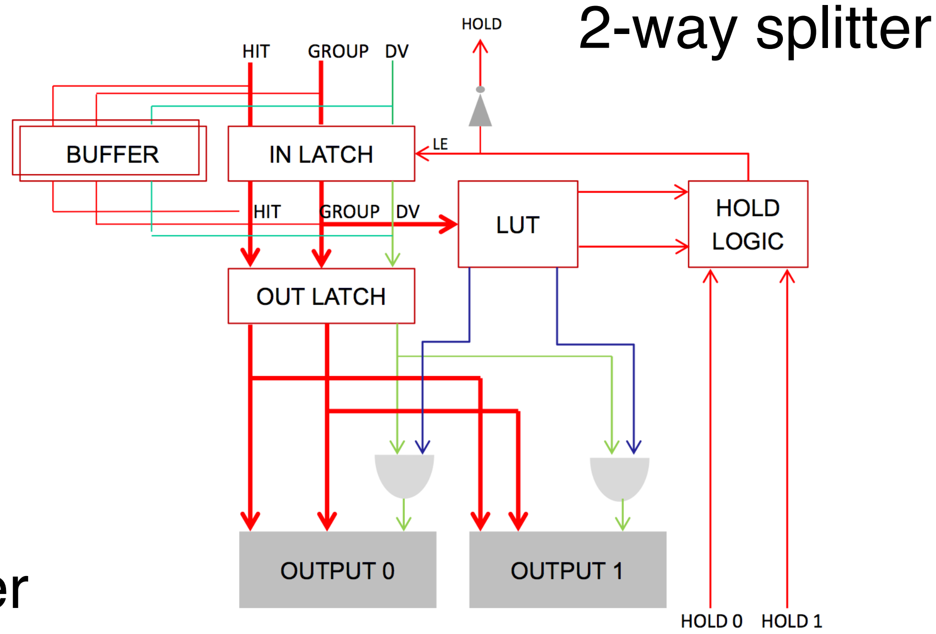
Conclusions

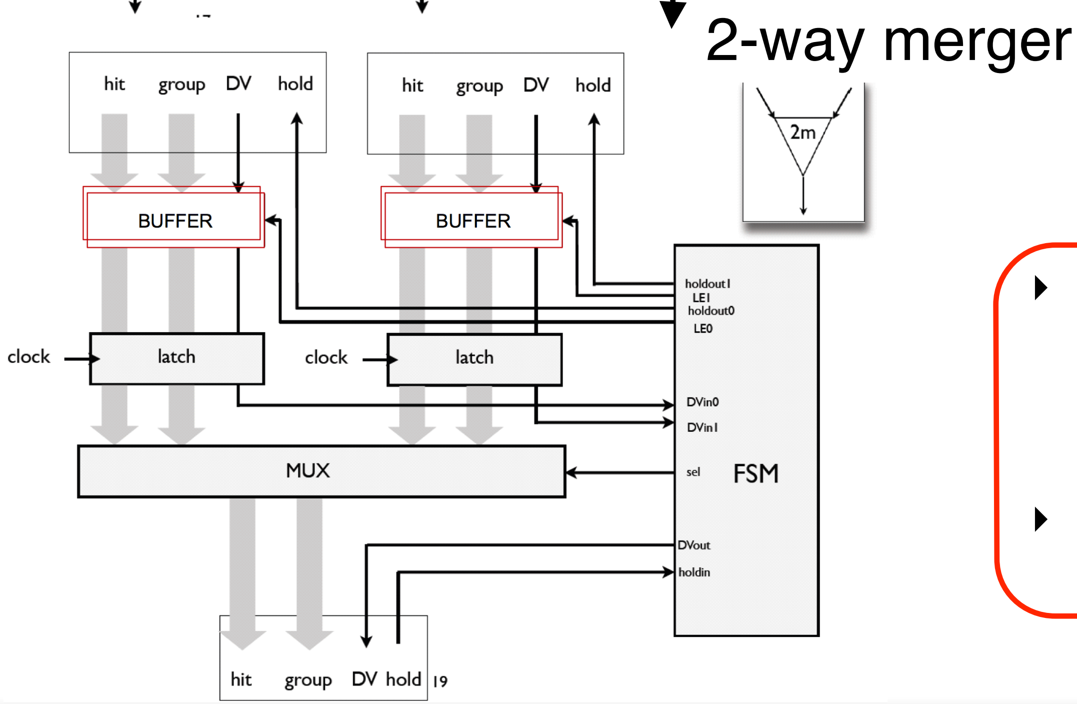
- The first prototype of a tracking system with artificial retina has been designed and it is under construction
- The artificial retina prototype has been designed to run on commercial FPGAs available on TEL62 boards
- The tracking performance are comparable with offline results with a latency of the response $< 1\mu s$
- The system is modular and suitable for HEP detectors with very high particle rates, e.g. LHC experiments. See LHCb-PUB-2014-026.

Backup slides

2-way sorter



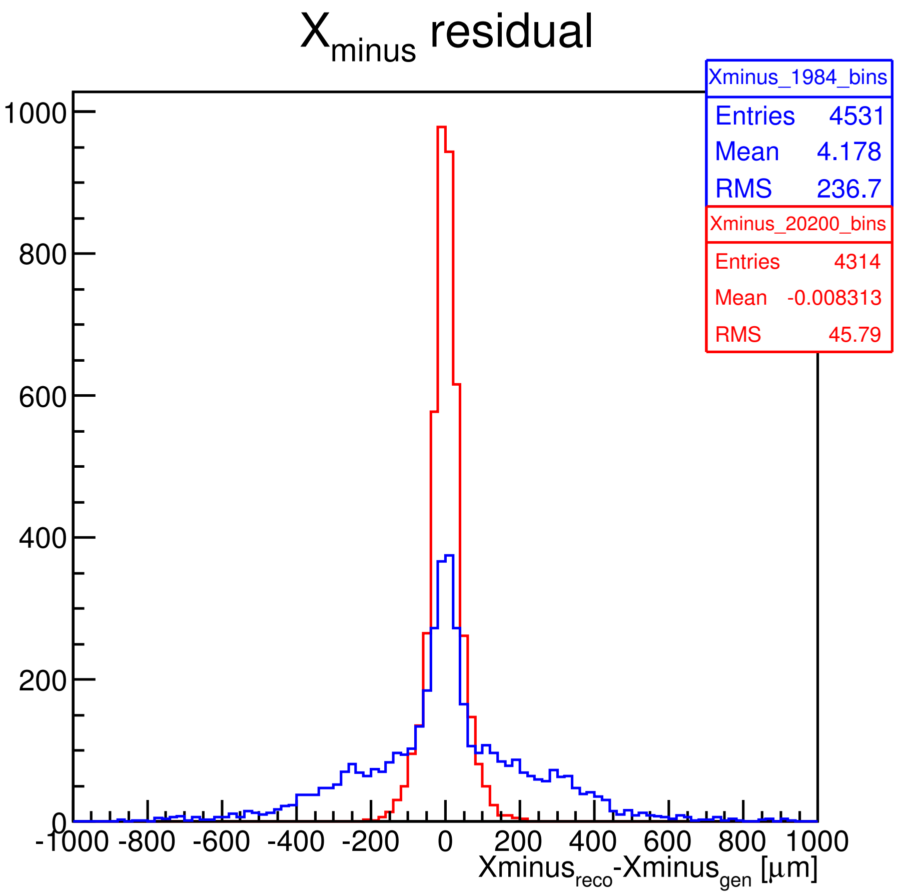


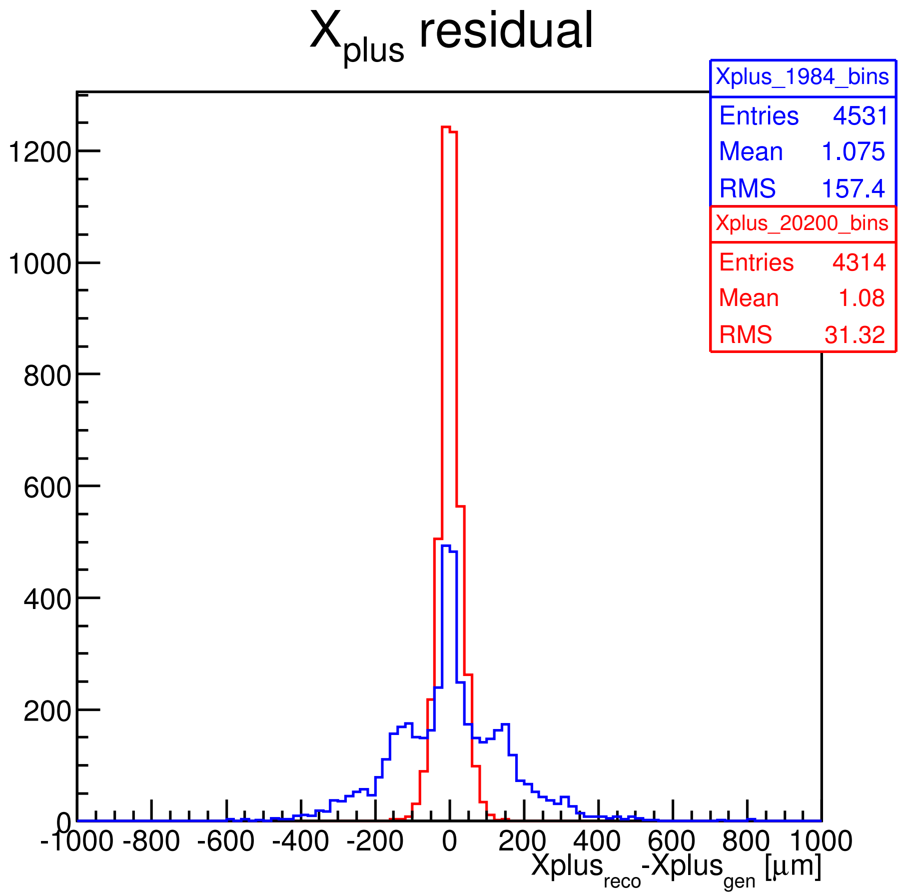


- The 2-way sorter acts also as a memory buffer in case of traffic jam. Input stream can be held
- Switch functionalities validated with VHDL simulation

Performance in presence of bkg

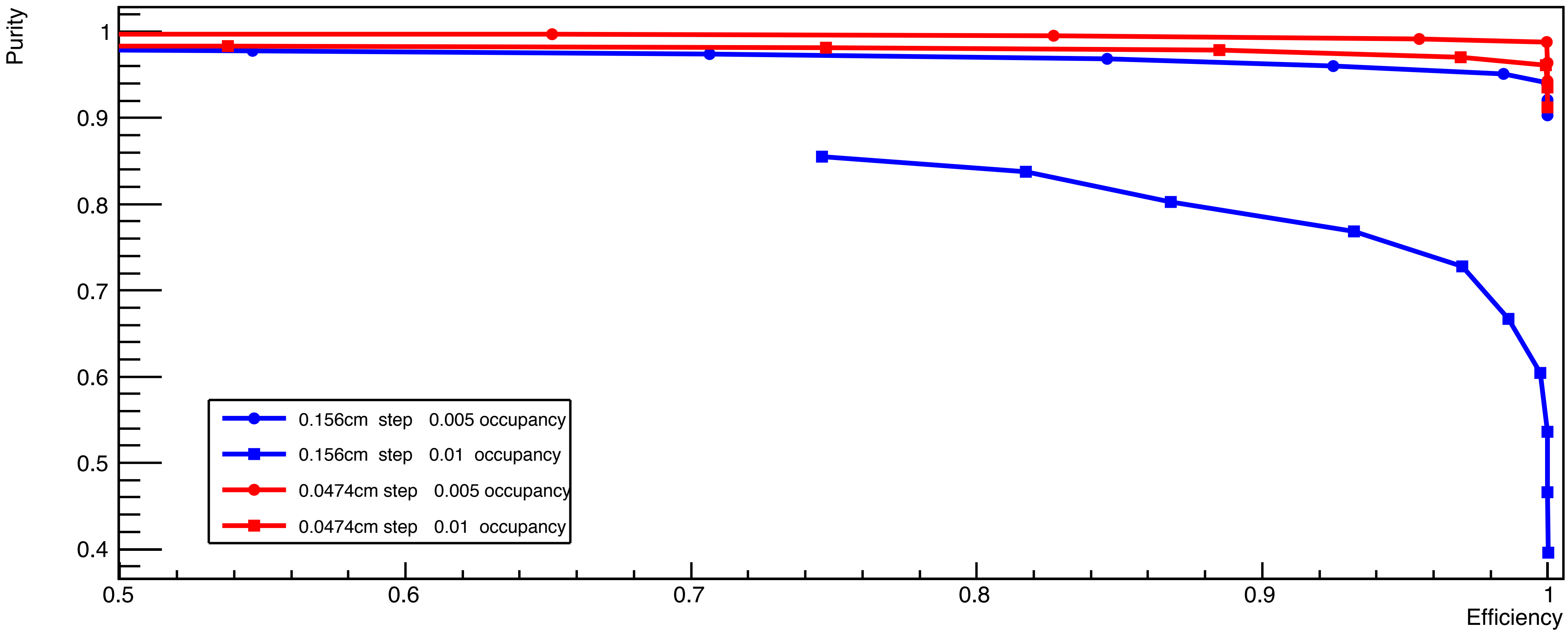
- Track parameter resolution in presence of noise hits (0.5% occupancy) with about 2k cells (blue) and 20k cells (red).





Performance in presence of bkg

- Efficiency vs purity in presence of noise hits (0.5% ● and 1% ■ occupancy) with about 2k cells (blue) and 20k cells (red).



Using hit time information

- Track residuals for 1% occupancy and 2K cells with no time (black) information, and with 100 ps (red), 10ps (blue) time resolution