# Simulation and performance of an artificial retina algorithm for 40MHz track reconstruction

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

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

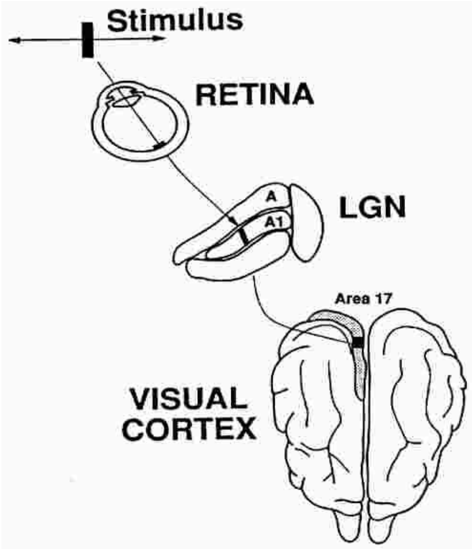
**WIT 2014, Philadephia, USA**

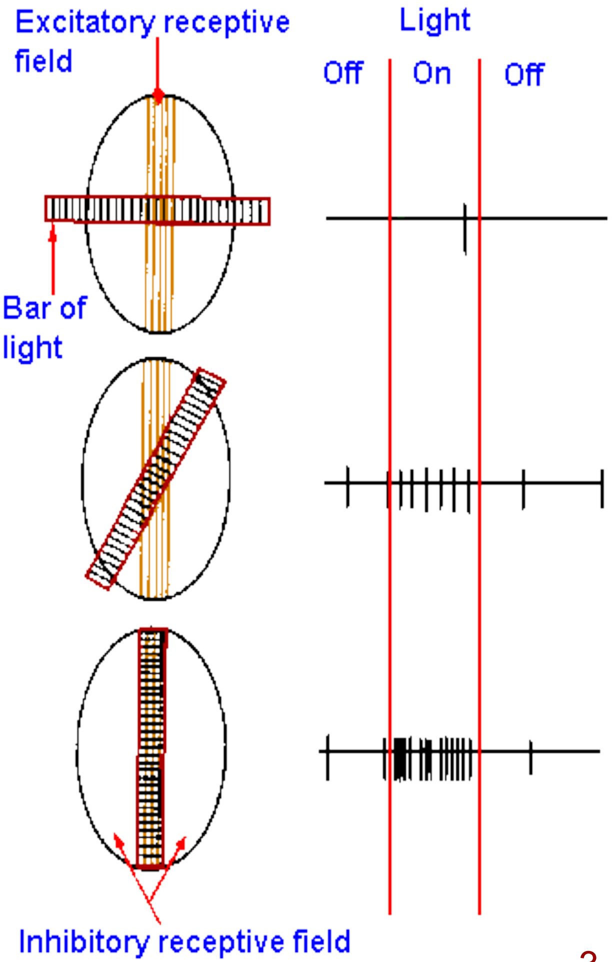
SCUOLA NORMALE SUPERIORE

# A “cellular” tracking algorithm

➜ Original idea in [L. Ristori, NIM A, 452 (2000) 425]

- Inspired by mechanism of visual receptive fields [D.H. Hubel, T.N. Wiesel, J. Physiol, 148 (1959) 574]
- Experimental evidence that V1 functionality can be modeled as a “trigger” [MM. Del Viva, G. Punzi, D. Benedetti, PloS one - DOI: 10.1371/journal.pone.0069154]



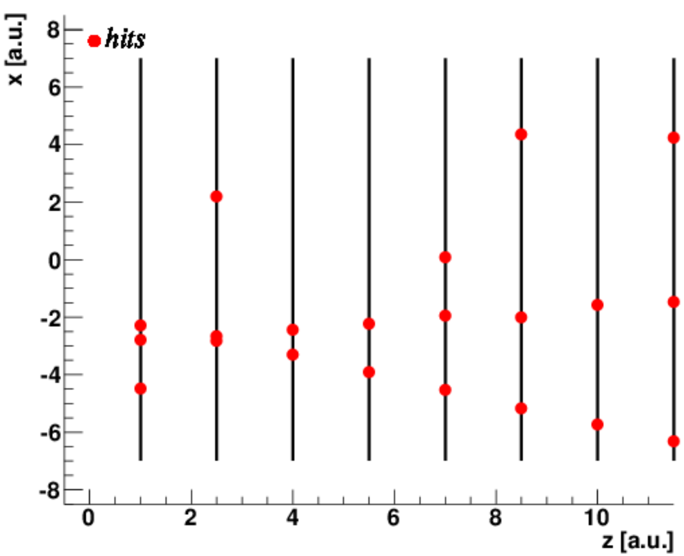



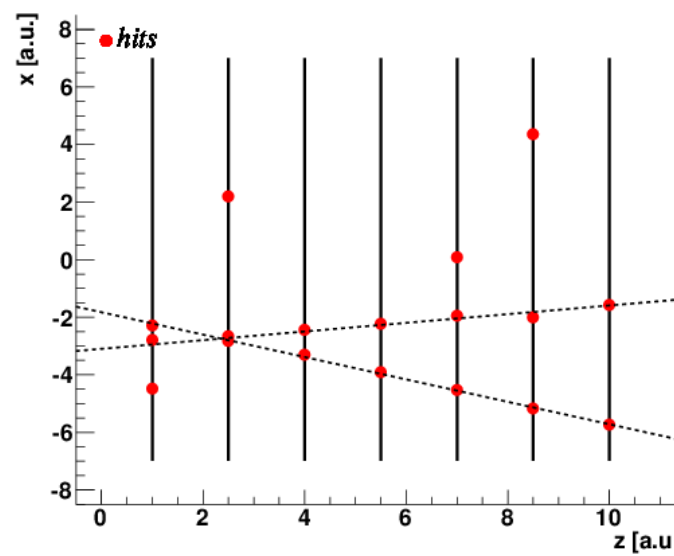
# A simple case

Reconstruction of tracks in absence of magnetic field (straight lines) using single-coordinate parallel detector layers.

$$x = mz + q$$ → Two dimensional space parameter **(m,q)**.

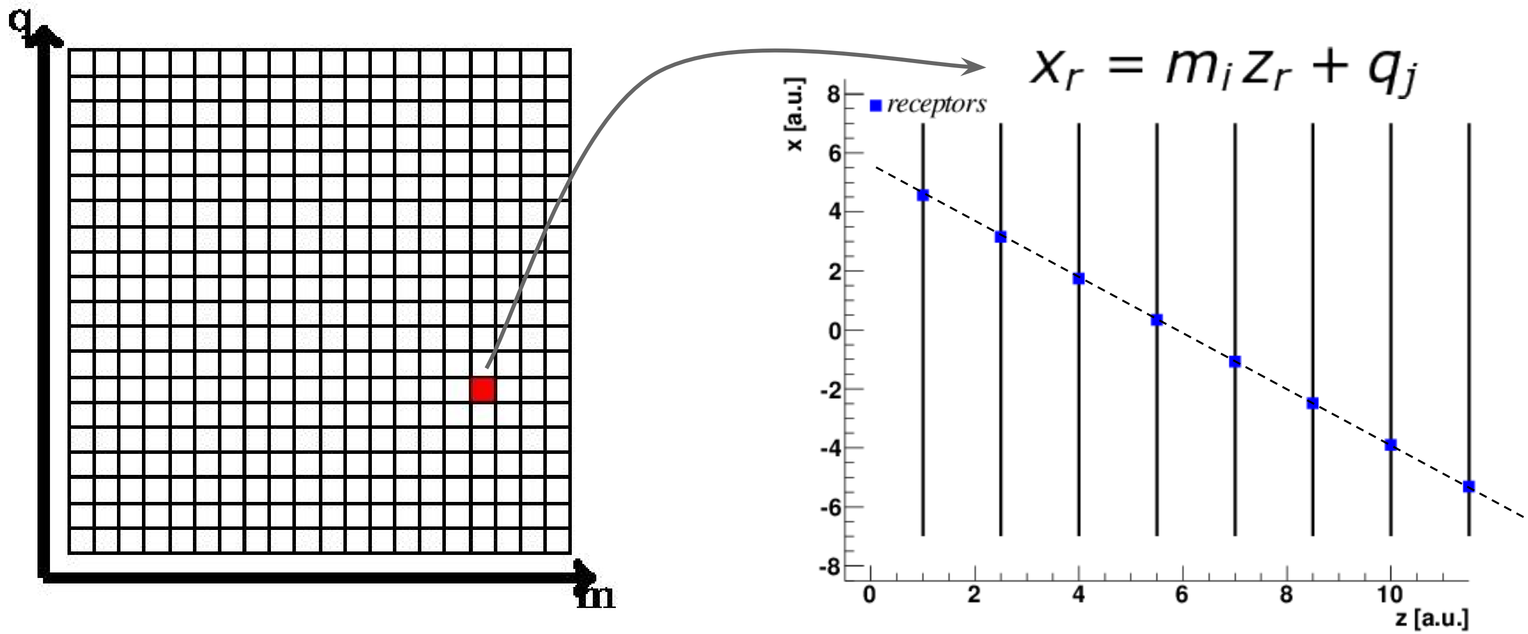
Tuning the “receptive fields” to cover all possible values of (m,q).



# Detector mapping

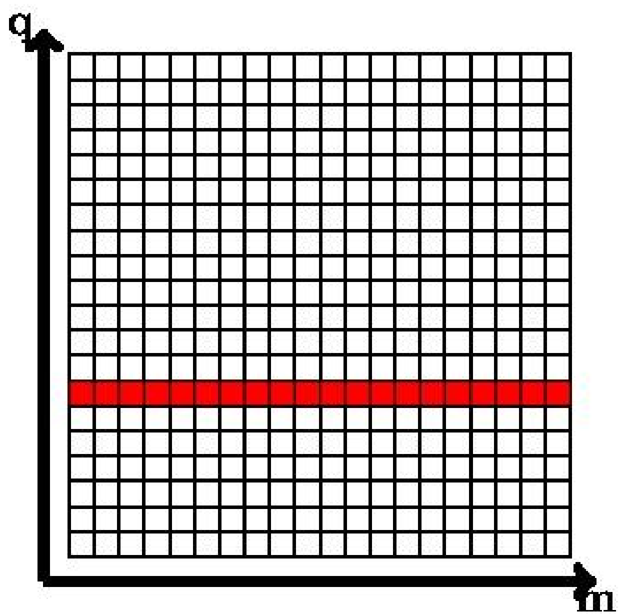
Discretize track parameter space in "cells". The center of each cell (or *d*-dim hyper-cube, in the case of a track with *d* parameters) identifies a track in the real space that intersects detector layers in "receptors".

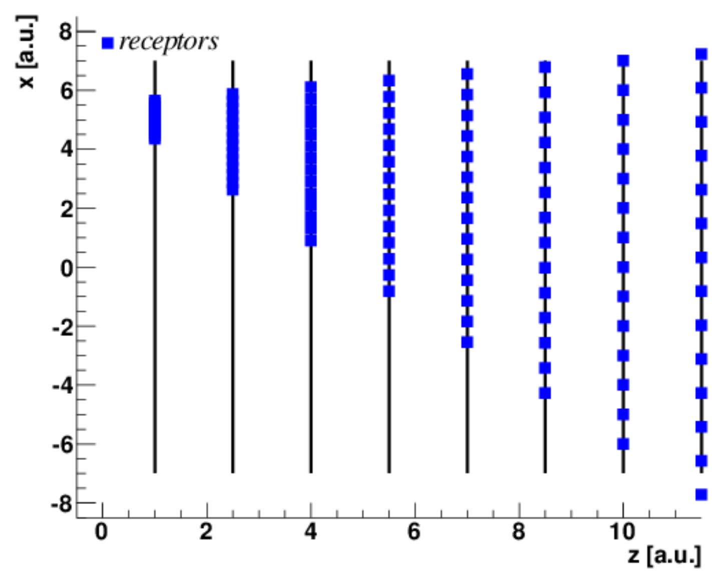



Each cellular unit corresponds to ***n*** (=number of layers) cellular receptors $(z_r, x_r)$ (*r* runs over the layers)

# Detector mapping

Discretize track parameter space in "cells". The center of each cell (or *d*-dim hyper-cube, in the case of a track with *d* parameters) identifies a track in the real space that intersects detector layers in "receptors".

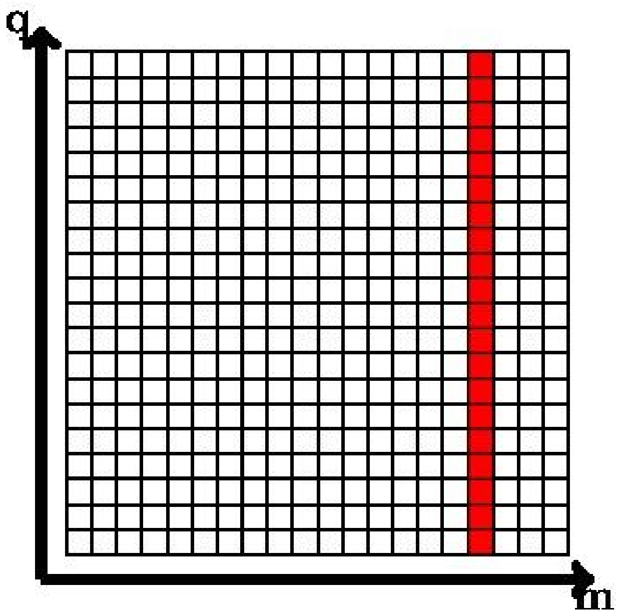


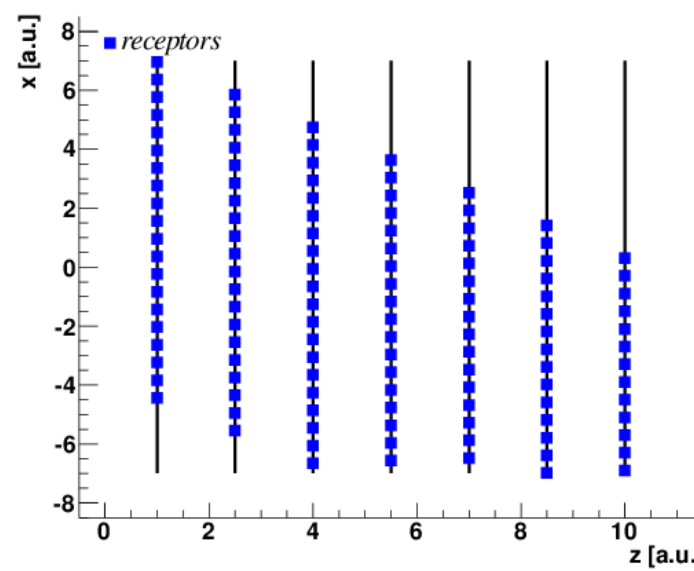




Each cellular unit corresponds to ***n*** (=number of layers) cellular receptors $(z_r, x_r)$ (*r* runs over the layers)

# Detector mapping

Discretize track parameter space in "cells". The center of each cell (or *d*-dim hyper-cube, in the case of a track with *d* parameters) identifies a track in the real space that intersects detector layers in "receptors".

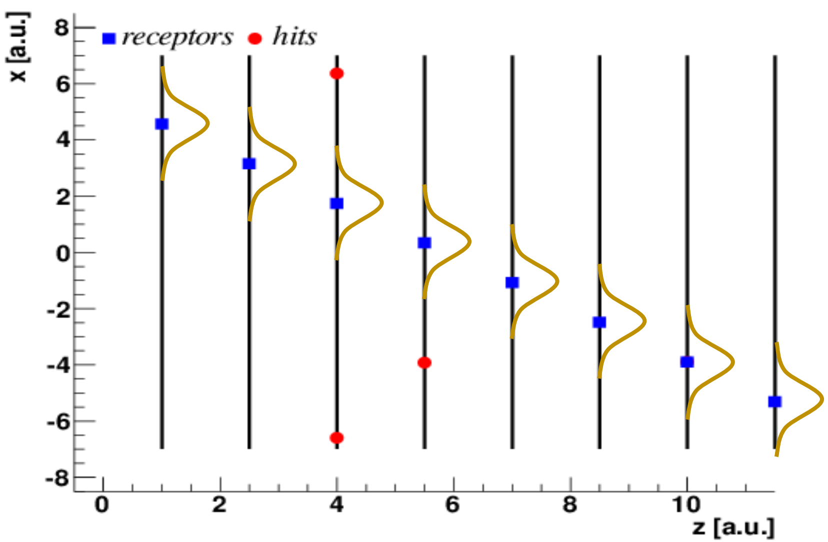



Each cellular unit corresponds to ***n*** (=number of layers) cellular receptors $(z_r, x_r)$ (*r* runs over the layers)

# Basic principle

For all the hits in the detector layers $(z_n, x_n)_k$ (due to real particles going through the detector or noise), the response $R_{ij}$ of the $(m_i, q_j)$ cellular unit is calculated summing over all hits and layers



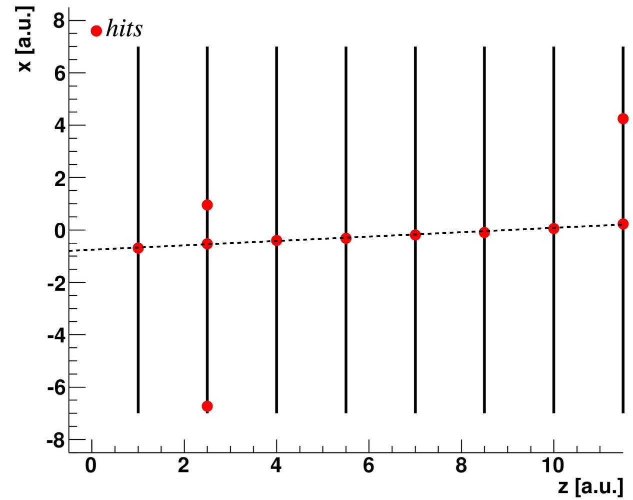

$$s^2_{ijkr} = (x^k_n - x^{ij}_r)^2$$

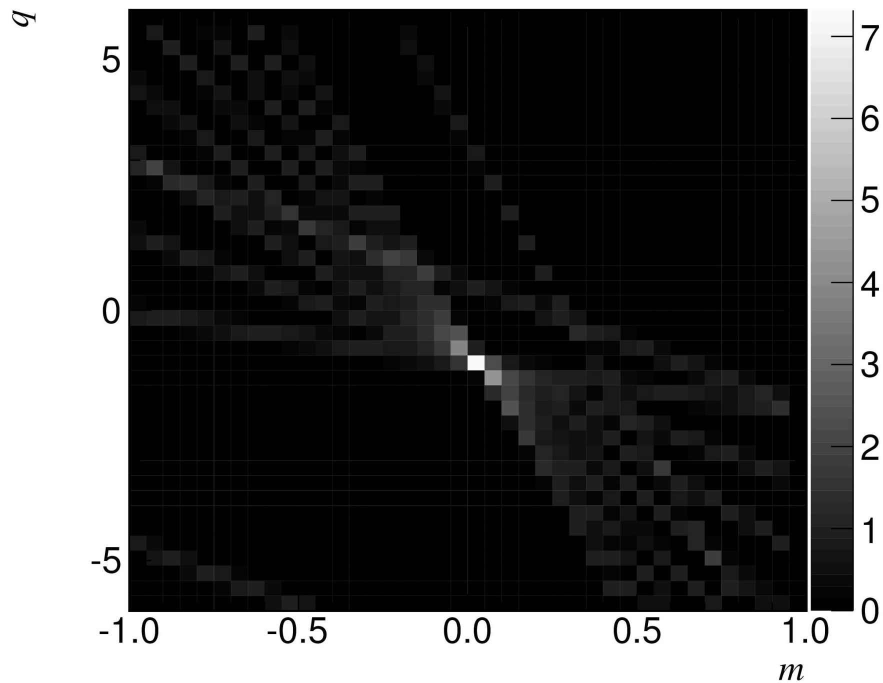
$$\boxed{R_{ij} = \sum_{kr} \exp\left(-\frac{s^2_{ijkr}}{2\sigma^2}\right)}$$

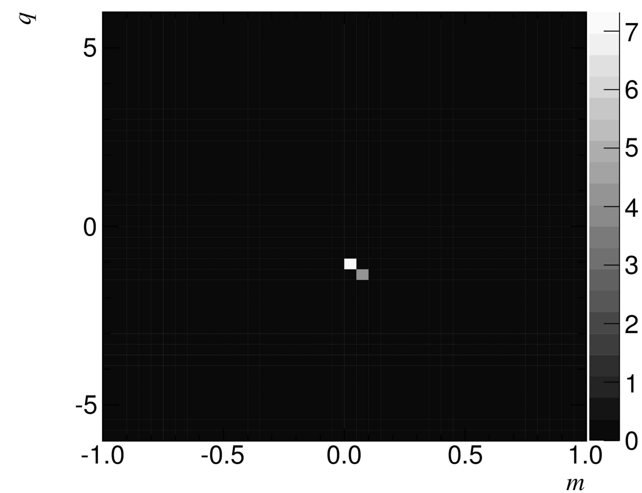
$R_{ij}$ represents the “excitation” of the the receptive field.

# The retina response

Once all cells are excited with the $R_{ij}$ , a track is identified by a local maximum (in the parameter space).



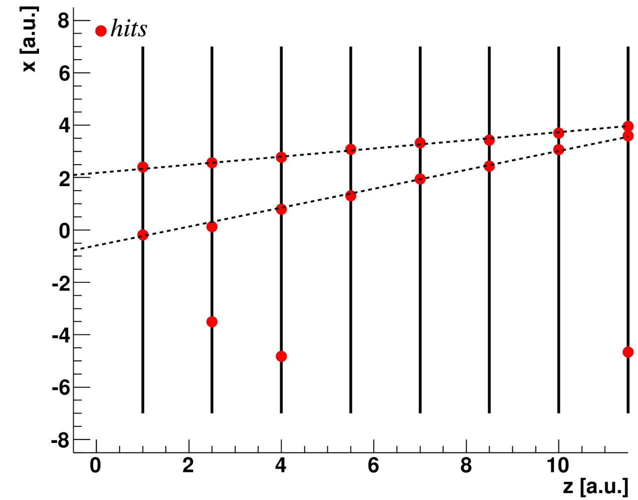



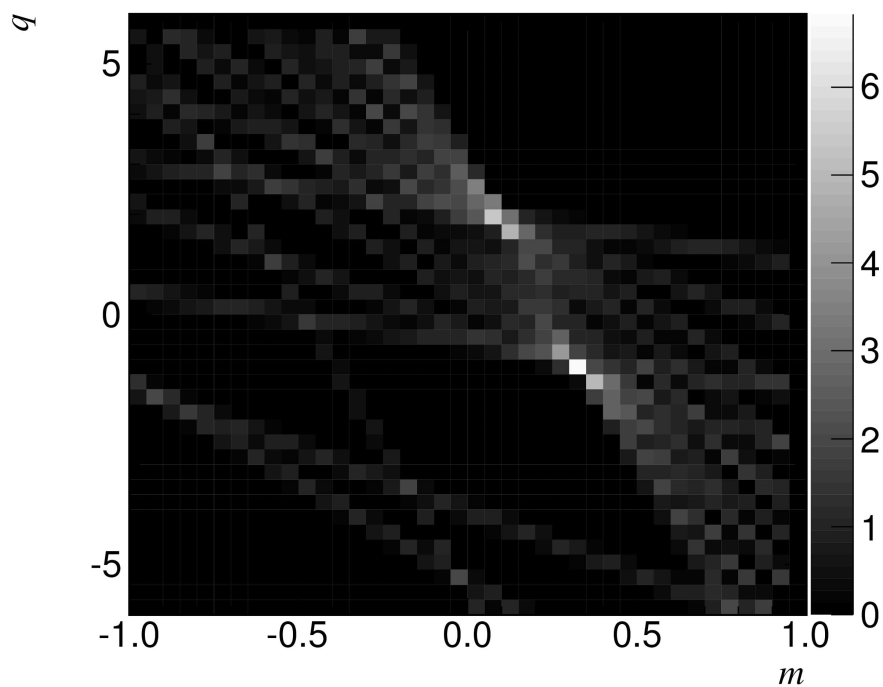



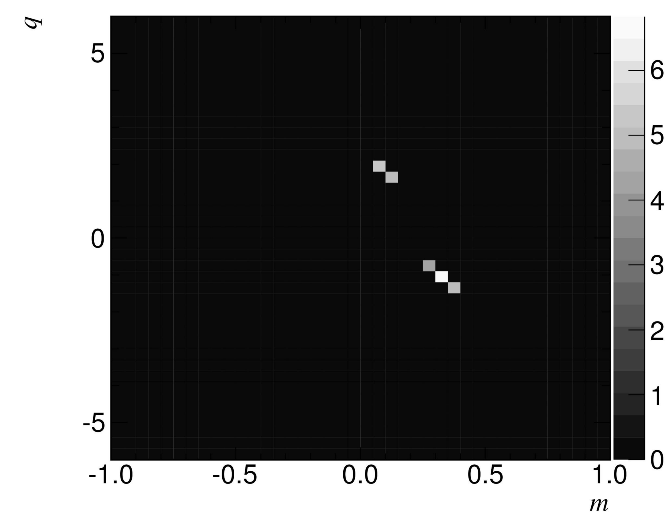

over threshold

# The retina response

Once all cells are excited with the $R_{ij}$ , a track is identified by a local maximum (in the parameter space).



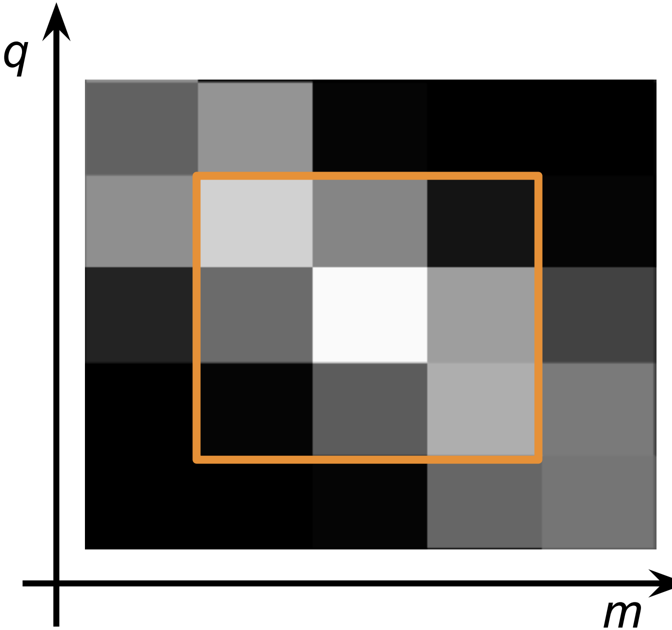

over threshold

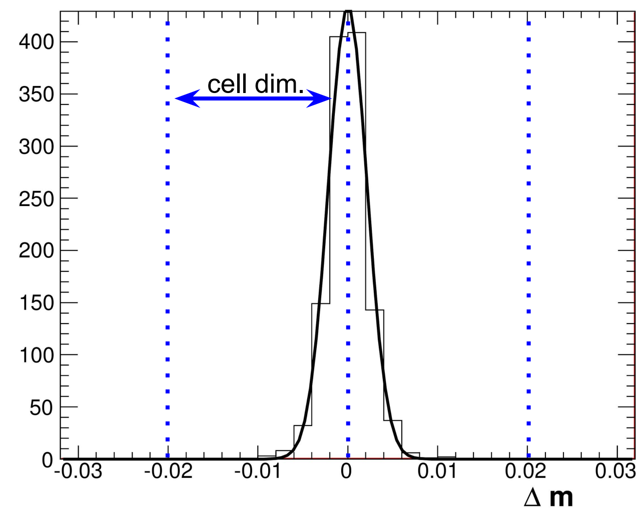
# Parameter extraction

Once local maxima (=tracks) are found parameter values are extracted by performing the centroid of the nearest cells.

$$m = \frac{\sum_{ij} m_i w_{ij}}{\sum_{ij} w_{ij}} \qquad q = \frac{\sum_{ij} q_j w_{ij}}{\sum_{ij} w_{ij}}$$

A subcell resolution is achieved by interpolation. Particularly important since it allows a coarse space granularity → limited number of cells.



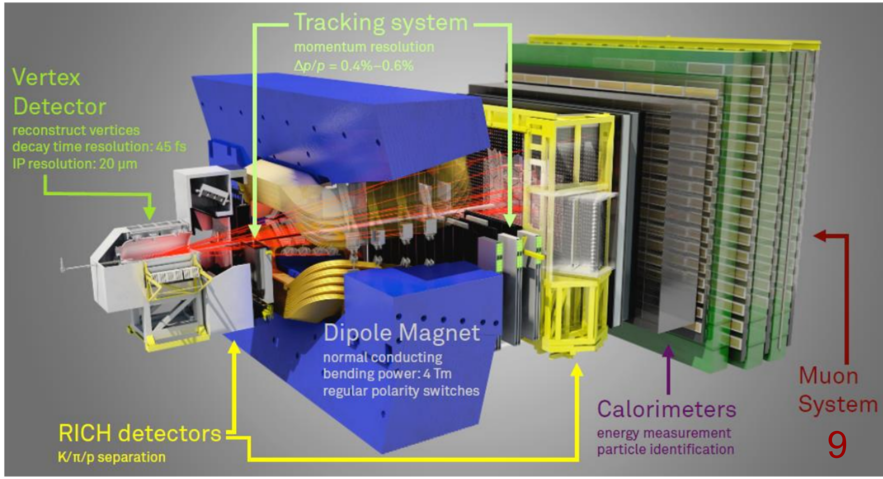
# Moving forward: a realistic case

The artificial retina algorithm seems **very promising** having interesting features for a very fast and high-quality tracking. However it has not been ever implemented in a real HEP detector (particles in magnetic field, high track multiplicity, multiple scattering, 3D geometry, **noise** and so on).

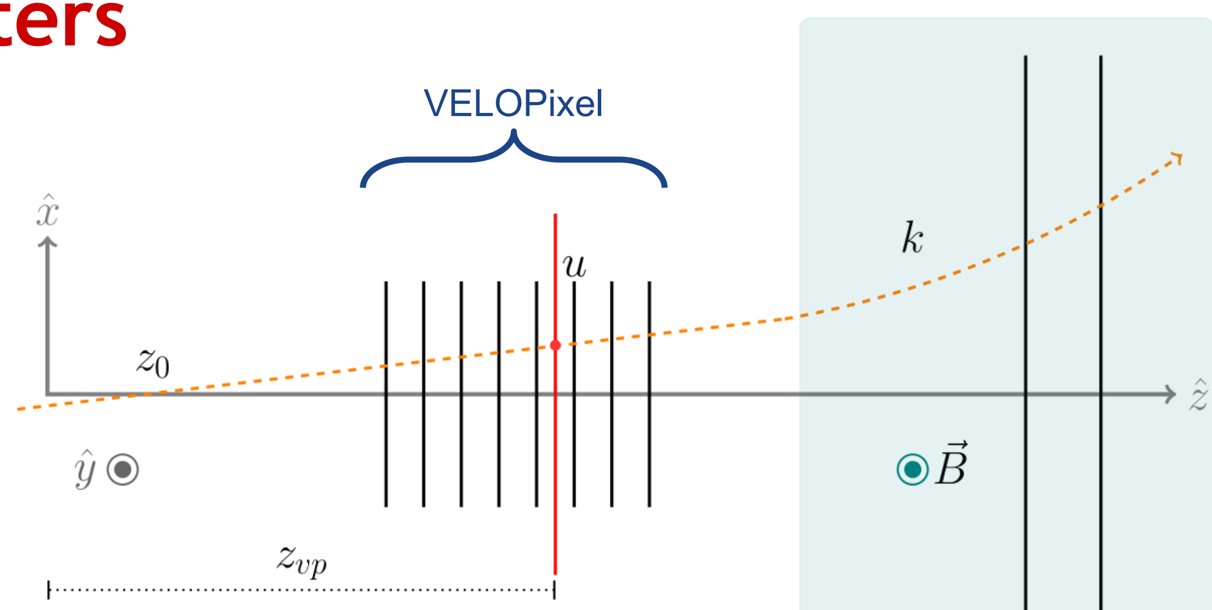
We studied features and performances of the retina algorithm in reconstructing real tracks passing through 8 realistic parallel pixel layers (without any magnetic field) and 2 strip layers (sink into the fringe field of magnet).

Detector geometry, event topology, sub-detectors occupancies, etc. are taken from the LHCb-Upgrade experiment [LHCb-TDR-12]:

- ➜ VELO pixel: 55x55μm, ~15μm hit resolution
- ➜ 2 UT mini-strip axial layers, ~50μm hit resolution



# Track parameters




Tracks can be described with 5 parameters, we chose:

- **(*u*,*v*)**: spatial coordinates of the intersection of a track on a "virtual plane" placed at $z_{vp}$ .
- $\boldsymbol{z_0}$: z coordinate of the point of closest approach to the z-axis
- ***d***: transverse impact parameter
- ***k***: signed track curvature, defined as $q/\sqrt{p_x^2 + p_z^2}$

# Retina in 5 dimensions

The general approach for the retina algorithm requires to discretize the 5-dimensional parameter space. Large number of cells $\rightarrow$ large size hardware device.

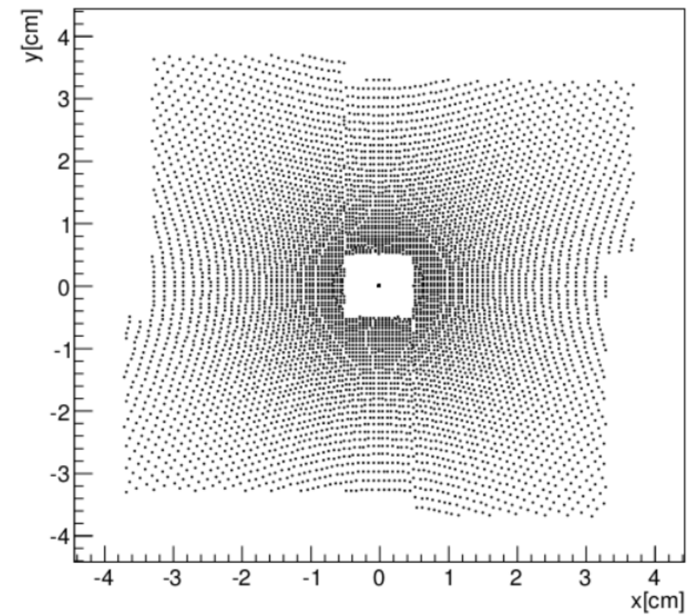
Track parameters do not have in general the same "relative weight". This allows a collapse of the retina dimensionality in performing the most relevant and time consuming task: the track finding.

For instance in LHCb the 5-dim space can be factorized into two subspaces:

$$\underbrace{(u,v)}_{\text{Main parameters}} \otimes \underbrace{(d,z_0,k)}_{\text{small perturbations}}$$

This allows performing the pattern recognition using only a 2D retina in the $(u,v)$ (with $d = z_0 = k = 0$). The $(d,z_0,k)$ parameters are corrections to the main $(u,v)$ parameters.

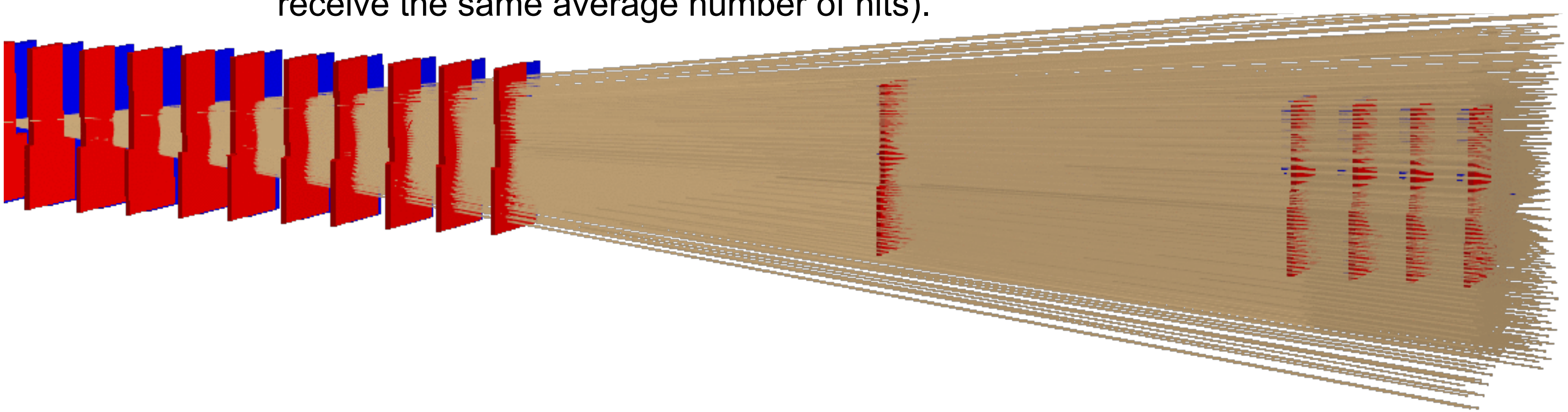
# Detector mapping




➜ Map all the 25,000 (*u,v*) cells units in the detector.

➜ Granularity of (*u,v*) space cells has chosen accordingly hits density on detector layers to optimize the computing power usage.

- In such a way all engines (cells) will have in average the same activity (or better, they will receive the same average number of hits).



# Simulation ingredients

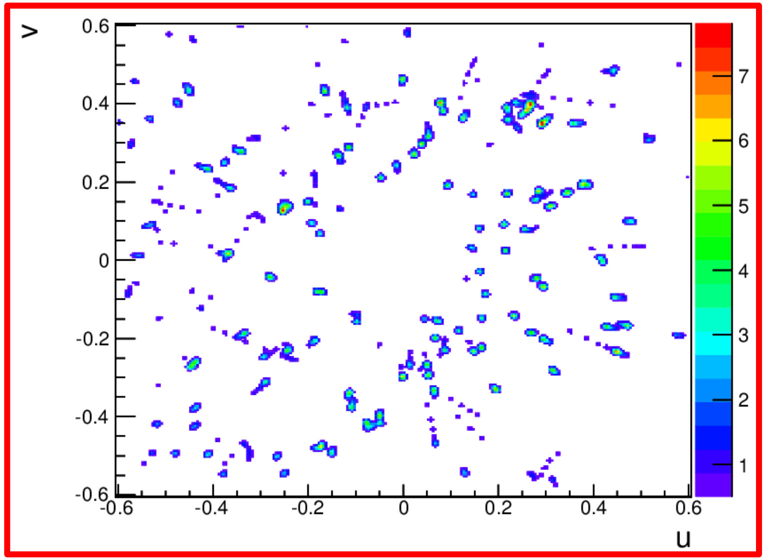
➜ We use a sample of minimum-bias events generated with PYTHIA8, with beam energy $E_{\text{beam}} = 7\,\text{TeV}/c^2$ , in two scenarios:

- $L = 2 \times 10^{33}\text{cm}^{-2}\text{s}^{-1} (\nu = 7.6)$
- $L = 3 \times 10^{33} cm^{-2} s^{-1} (\nu = 11.4)$

➜ Fiducial cuts are applied on the *reconstructable* tracks:

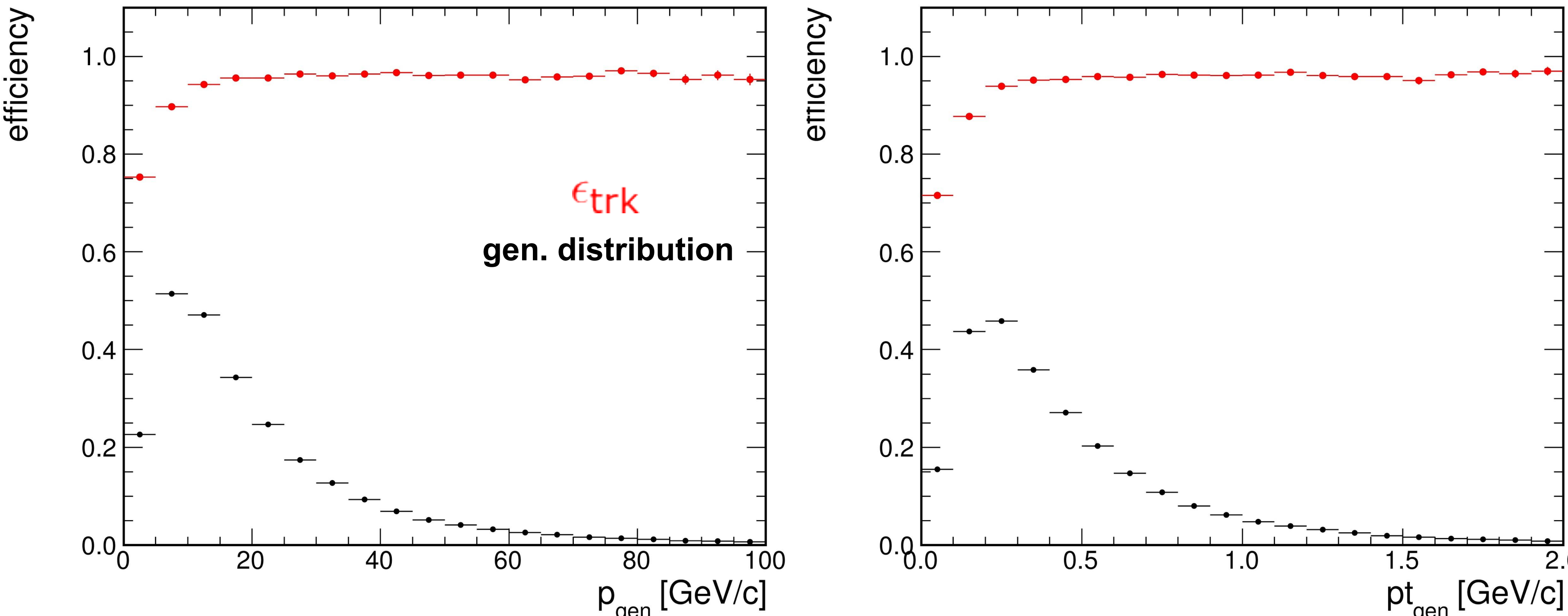
- acceptance cuts (max(|*u*|, |*v*|) < 0.35 ($\theta \approx 50\,\text{mrad}$ ), |z| < 150 mm) ,
- at least 3 hits on the VELO layers
- and 2 hits on the UT layers

A typical event has hundreds of charged particles.



# Tracking efficiency vs. p, pt

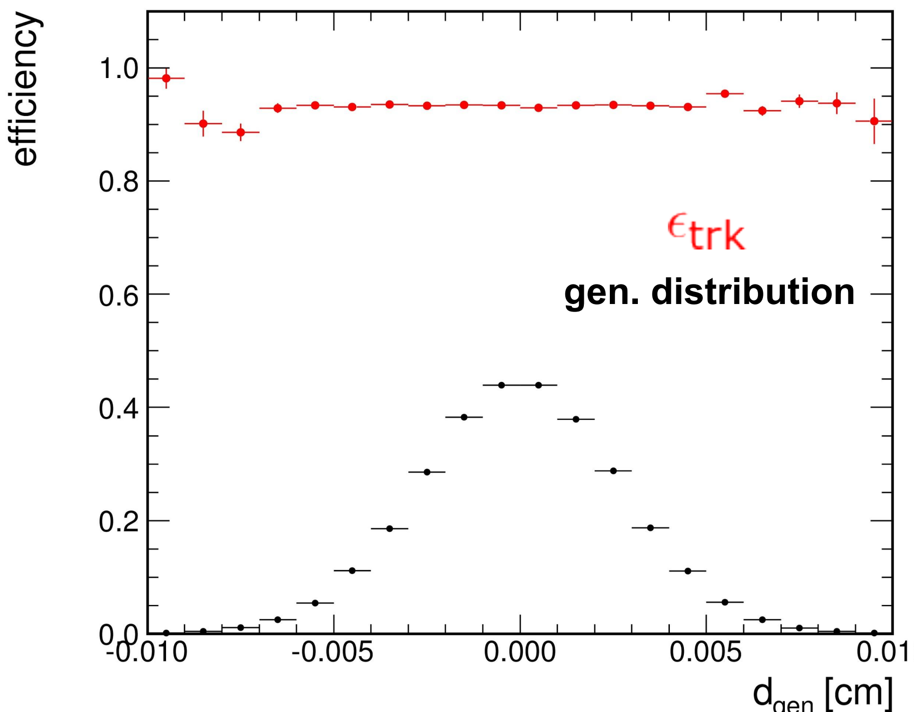
High efficiency (~95%) and uniformity in response.
The same as the full offline reconstruction algorithm.

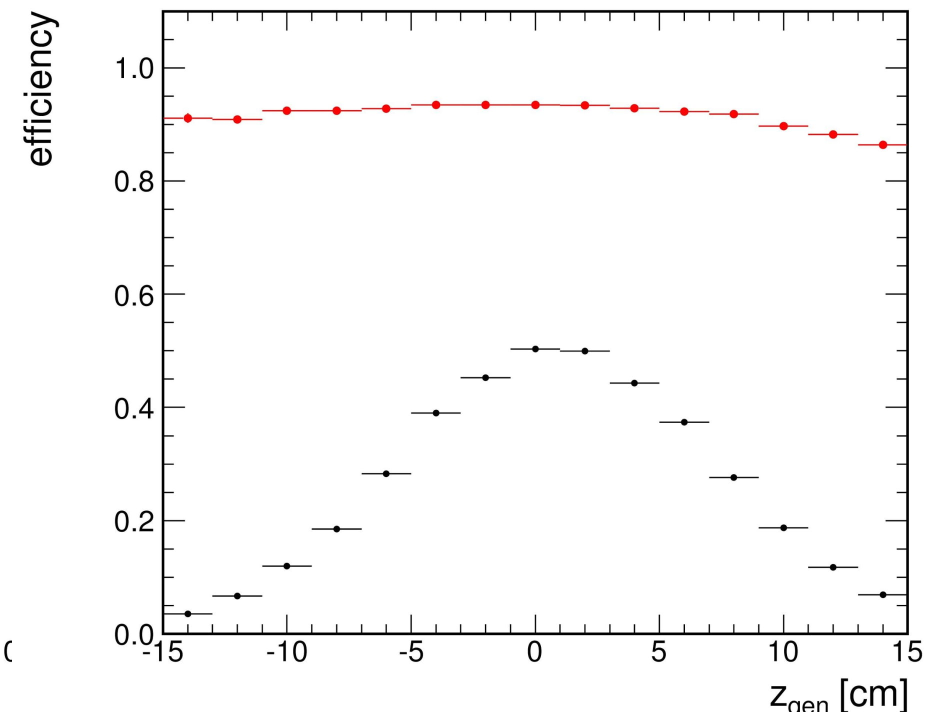



$$L = 2 \times 10^{33} cm^{-2} s^{-1} (\nu = 7.6)$$

# Tracking efficiency vs. d, z

High efficiency (~95%) and uniformity in response.
The same as the full offline reconstruction algorithm.

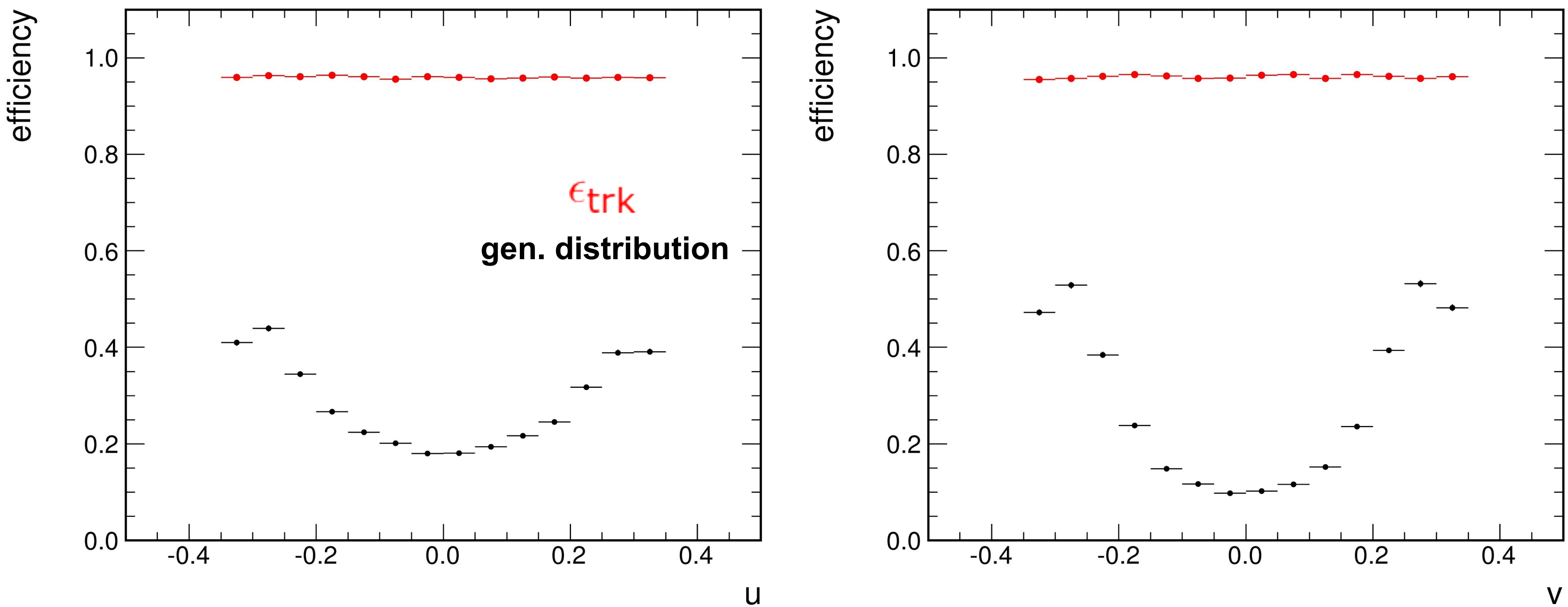



$$L = 2 \times 10^{33} \mathrm{cm}^{-2} \mathrm{s}^{-1} (\nu = 7.6)$$

# Tracking efficiency vs. u, v

High efficiency (~95%) and uniformity in response.
The same as the full offline reconstruction algorithm.




$L = 2 \times 10^{33} \mathrm{cm}^{-2} \mathrm{s}^{-1} (\nu = 7.6)$

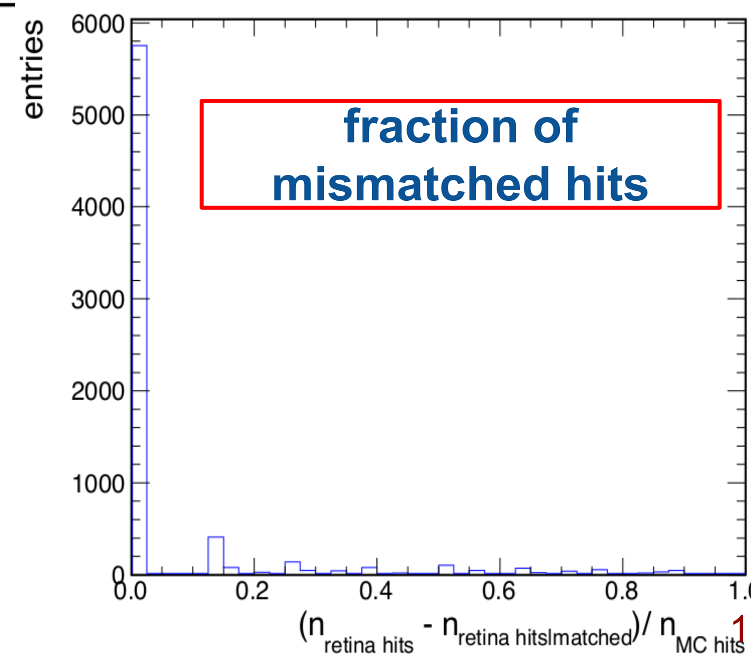
# Pattern recognition performance

| | $2 \times 10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ | $3 \times 10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ |
|---|---|---|
| Number of hits | 880 | 1220 |
| Number of clusters (over thld) | 121 | 223 |
| Number of hits per engine | 1.3 | 1.95 |
| Ghost rate | 0.08 | 0.12 |

per event

➜ Ghost rate under control, at the same level as the offline reconstruction algorithm.

➜ Fraction of mismatched hits is limited.



# Efficiency on signal benchmarks

| | $2 \times 10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ | $3 \times 10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ |
|---|---|---|
| $B_s^0 \to \phi\phi$ (signal tracks) | 0.97 | 0.97 |
| $D^{*+} \to D^0\pi^+$ (signal tracks) | 0.97 | - |
| $B^0 \to K^*\mu\mu$ (signal tracks) | 0.98 | - |

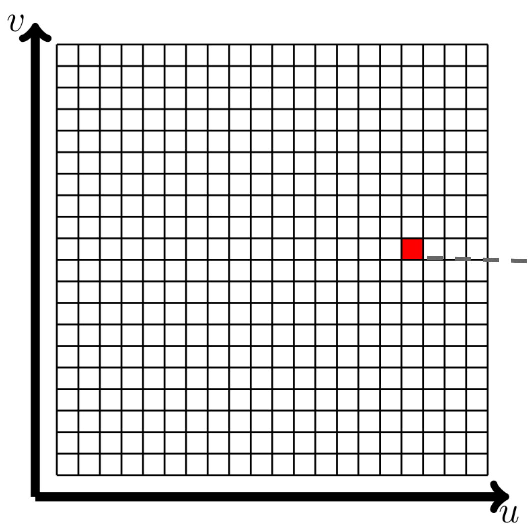
(Momentum criteria of p > 3 GeV/c and pT > 0.5 GeV/c are applied)

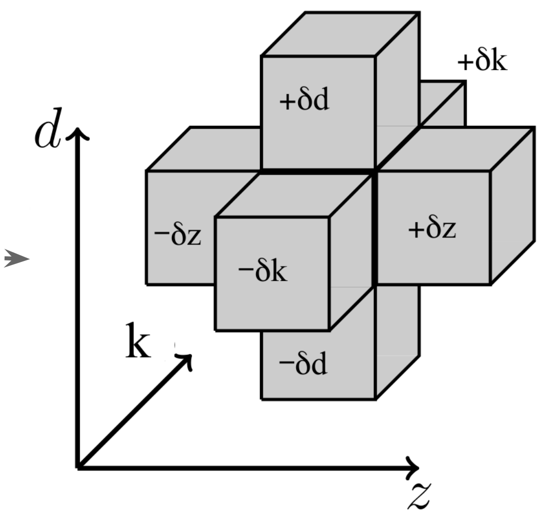
# Parameters extraction

The *u*,*v* parameters are directly extracted from clusters centroid

➜ for the other 3 parameters:

- add “lateral cells” to each cellular unit
- and interpolate their response



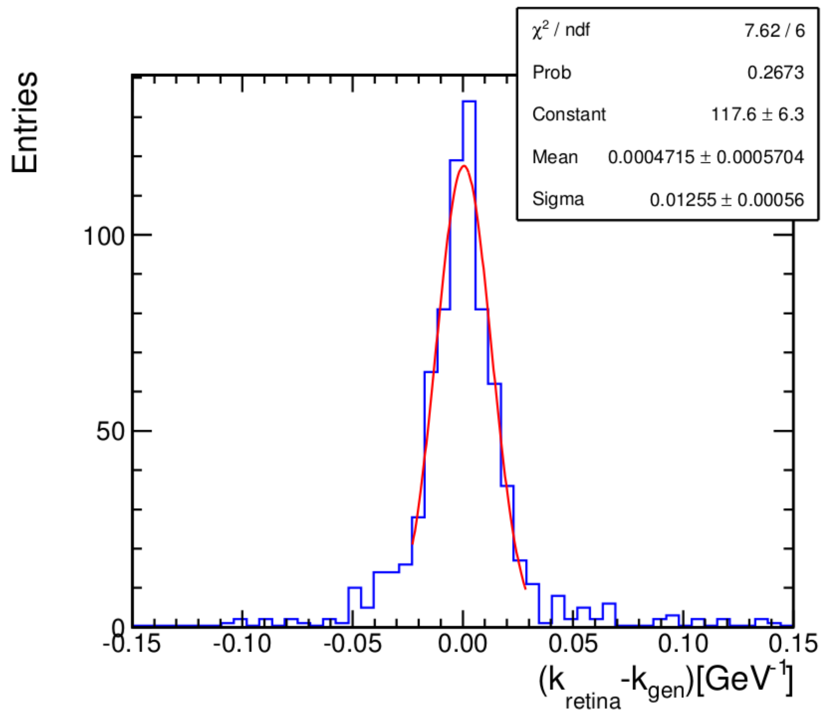

Interpolation of the lateral cells response provides an estimation of the correspondent parameter e.g.:

$$d_{rec} = \frac{w_{\delta d} - w_{-\delta d}}{w_{\delta d} + w_{d=0} + w_{-\delta d}} \cdot \delta d$$

Correlation among parameters affects the achievable resolution.

# Linearized fit

Further refinement of the track parameters estimation is achieved with a linearized track fitting algorithm [SVT TDR].




$\mathbf{x}$ = vector hits

$p_i$ = i-th parameter

$$p_i = p_i(\mathbf{x})$$

$$p_i \approx \mathbf{w}_i \cdot (\mathbf{x} - \mathbf{x}_0) + p_i(\mathbf{x}_0)$$

$$\left.\begin{matrix} \mathbf{w}_i \\ p_i(\mathbf{x}_0) \end{matrix}\right\}$$ Constants calculated from a sample of tracks with known parameters

**Comparable resolution with a full fit of the tracks.**

# Conclusions

- ➜ For the first time the artificial retina algorithm was developed in a real and complex experimental apparatus,
  - and its performances are established.
    - >~95% efficiency, uniform response.

- ➜ The retina system is technology feasible (D. Tonelli talk)
  - and able to reconstruct track at the LHC crossing rate,
  - latency < 1μs

- ➜ From a very promising idea we moved to a feasible, 40MHz, offline-quality tracker system for an intelligent and massively parallel tracking.

- ➜ Reference: retina LHCb public note [CERN-LHCb-PUB-2014-026].

… thank you!

# Backup

# LHCb Magnetic field