ACAT 2019

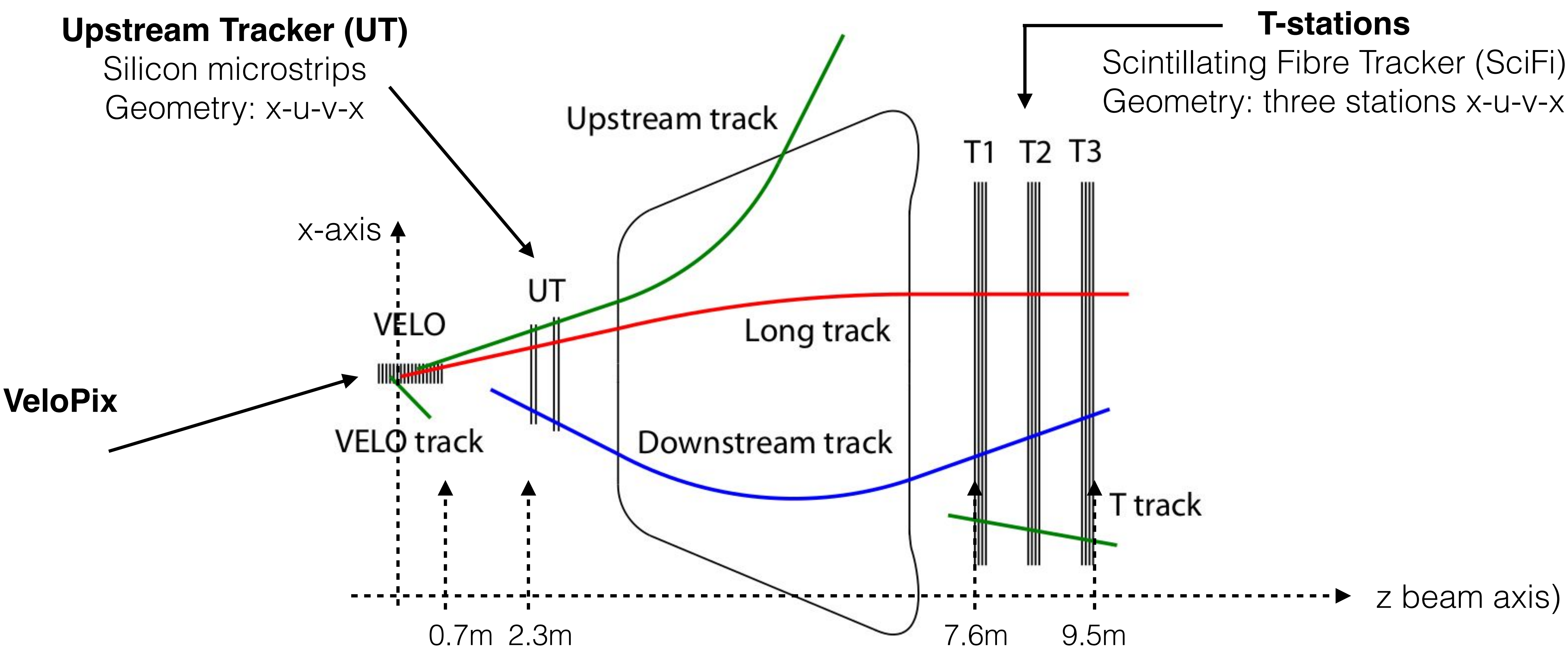
19th International Workshop on Advanced Computing and Analysis Techniques in Physics Research - March 11-15, 2019 Saas Fee (Switzerland)

Real-time reconstruction of long-lived particles at LHCb using FPGAs.

Michael J. Morello
michael.morello@sns.it
Scuola Normale Superiore and INFN, Pisa (Italy)

on behalf of the LHCb Collaboration

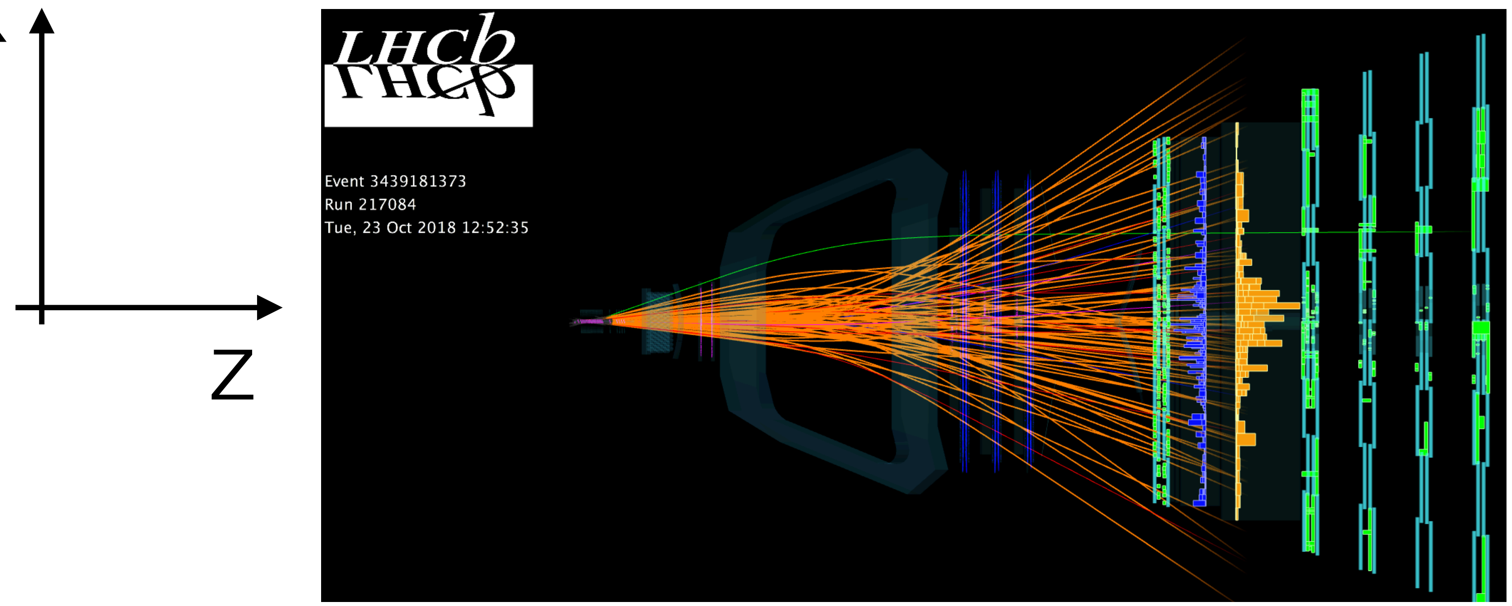
Tracking at LHCb-Upgrade



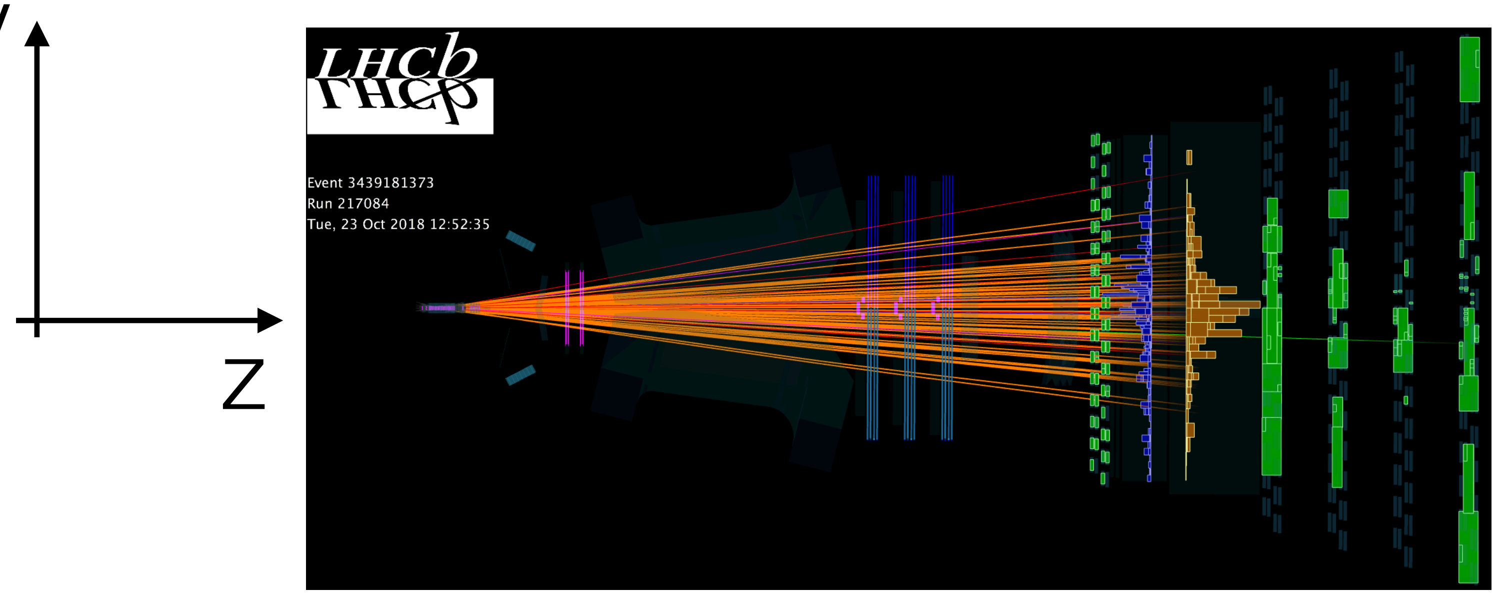
- **Long track**: reconstructible using VELO + UT + T stations. → beauty and charm core physics
- **Downstream track**: reconstructible using UT + T stations. → crucially to reconstruct for LLPs.
- **T-track**: reconstructible using T stations. → first stage of downstream tracks reconstruction.

A crowded event in 2018 ($L=4x10^{32}$)

Axial view



Stereo view



An average of about 150-200 tracks per event, up to a maximum of 400-500 tracks per event, is expected at the LHCb-Upgrade conditions $L=2x10^{33}$ $cm^{-1}s^{-1}$ (LHC Run3 and Run 4).

Downstream tracking

- Algorithm used for reconstructing standalone T-tracks in the SciFi sub-detector, called **seeding**, is crucial for reconstructing tracks generated by long-lived particles (LLPs), such as K_S,K_L,Λ.
- Seeding is a very intensive pattern recognition task and requires a significant amount of CPU-time to be executed, i.e about 500 μsec (❋) per event [*LHCb-PUB-2017-005*].
  - Adding UT hits to form a “downstream track” requires a total CPU-time even larger.
- The total budget for the LHCb-Upgrade tracking sequence, in Run 3, is expected to be 33 μsec per event, assuming 1000 Event Filter Farm nodes [*LHCB-TDR-016, LHCb-PUB-2017-005*] $\Longrightarrow$ at the moment finding tracks downstream of the magnet at the earliest trigger level is not part of the baseline trigger scheme.

See C. Fitzpatrick’s talk (Track 1) for details on the LHCb-Upgrade Tracking Sequence.

(❋) measured using a setup different from the official throughput test setup.

The Downstream Tracker

- R&D work in the context of the Future LHCb Upgrades (LHC Run 4 and beyond), to realize a downstream tracking unit that can be integrated in the DAQ architecture and act as an “embedded track-detector”.

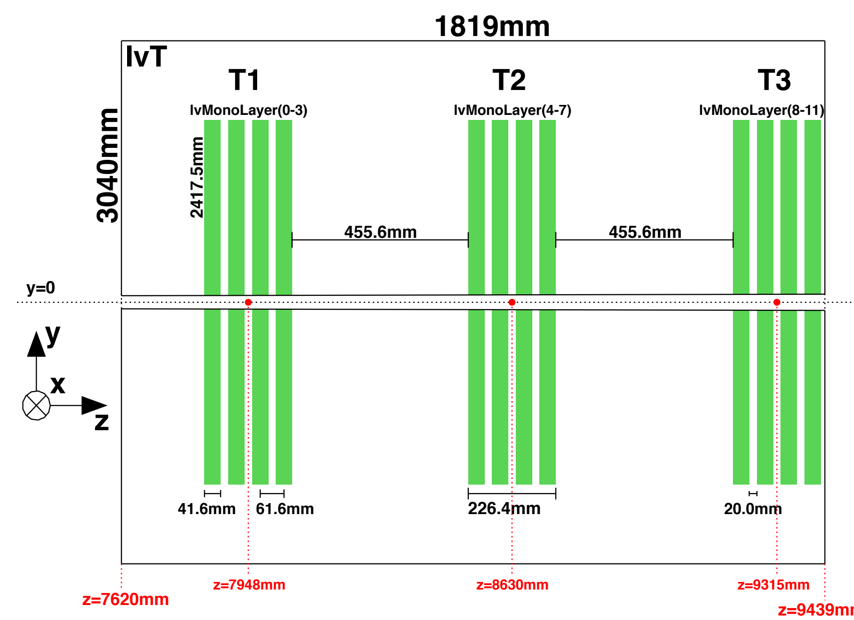
- This would make event reconstruction primitives immediately available to the event-building, and to the high-level-trigger farm.

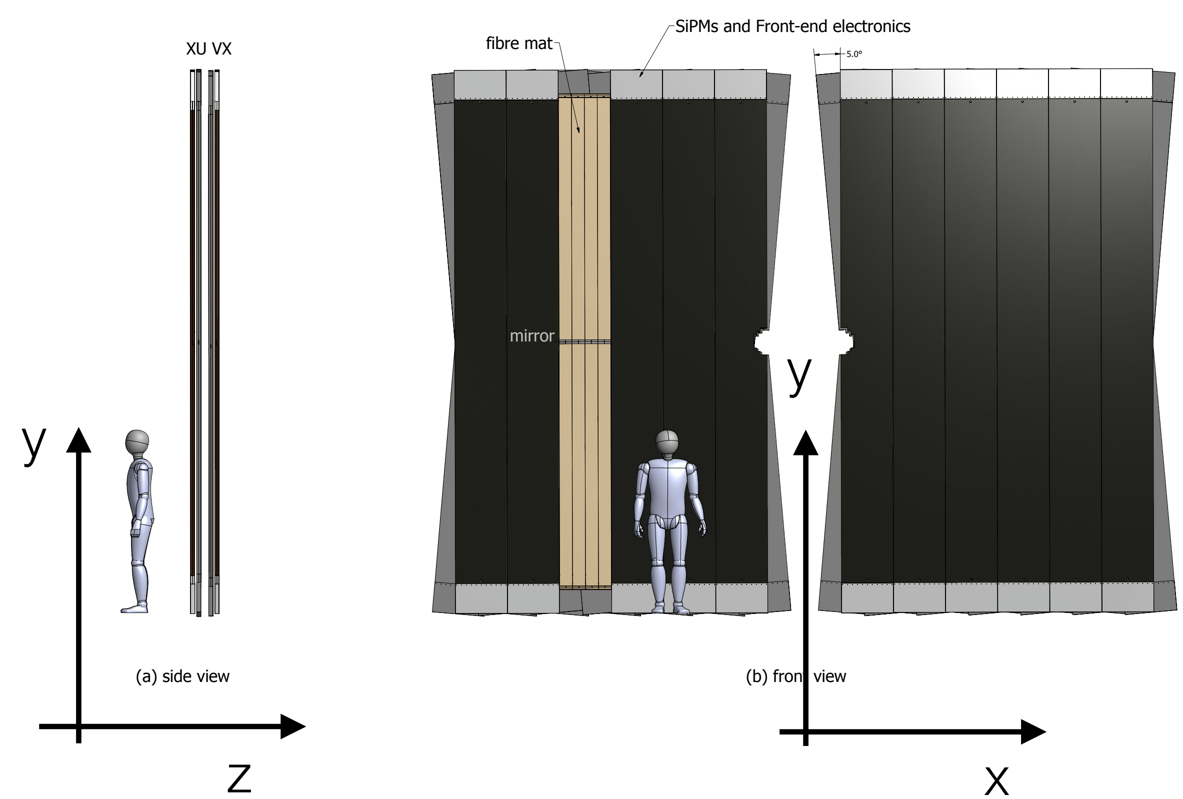
- FPGA is the appropriate technology: aim for high bandwidth and low latency, comparable with that of other elements in the detector DAQ.

This talk: specifically aimed at the reconstruction of standalone T-tracks, the first and most expensive part of the reconstruction of downstream tracks.

Scintillating Fibre Tracker (SciFi)

- 3 tracking stations (T1,T2,T3) of scintillating fibre.
- 4 layers per station (x-u-v-x)
  - u/v layers tilted by a stereo angle of +5°/-5°.
  - Electronic readout at 40MHz.
  - Hit spatial resolution: ~100 μm.
  - High occupancy: an average of about 300 hits per layer, up to a maximum of 800 hits per layer.
- A small component of magnetic field (fringe field) is present in the SciFi region. Tracks are well approximated as parabola in x-z view, and as straight lines in y-z view.
- For this study SciFi divided in 4 independent quadrants. Negligible loss of efficiency because of tracks crossing different quadrants.



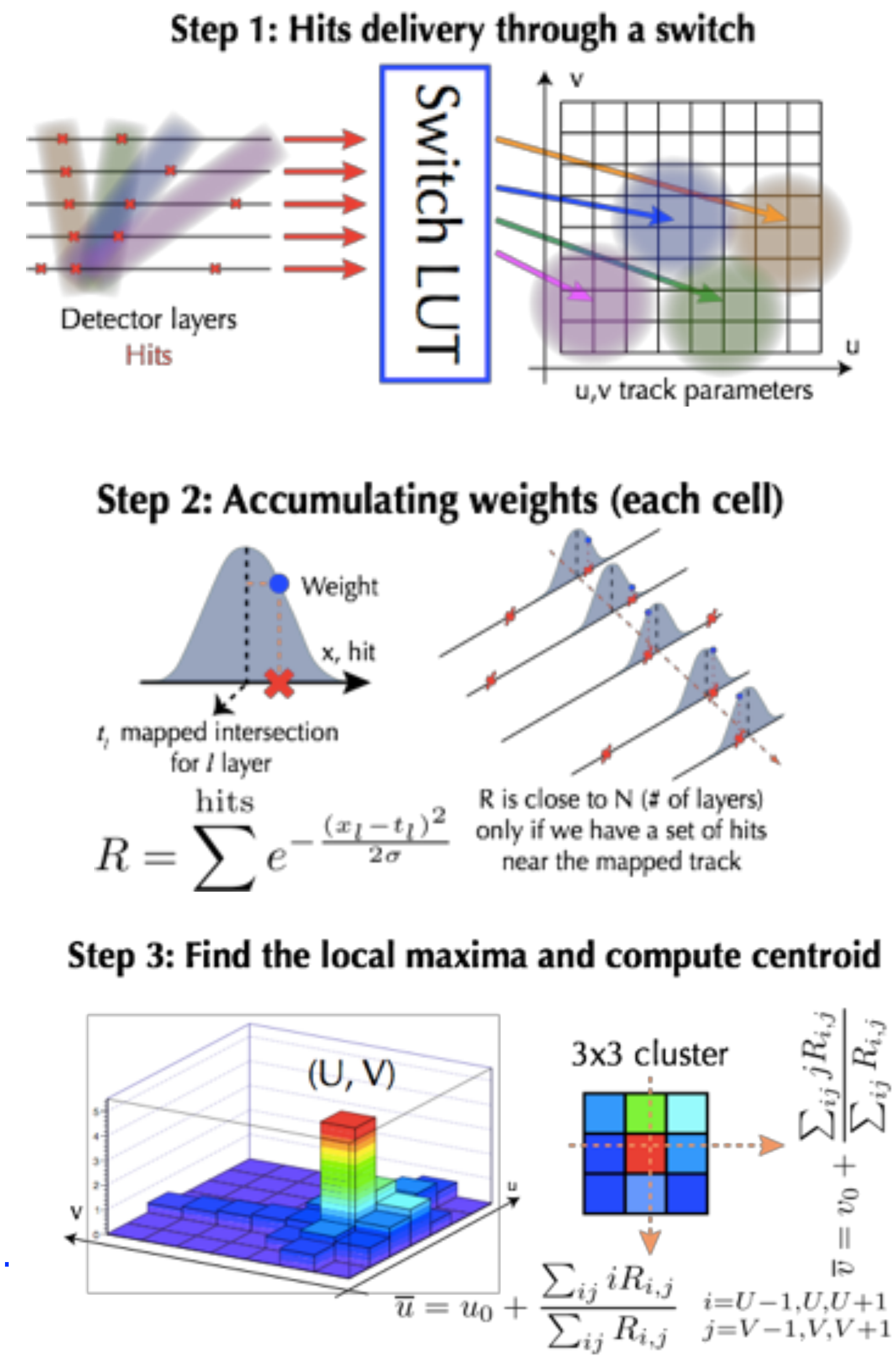


A biologically inspired architecture: the Retina Algorithm

Massively parallel architecture. Similarities with Hough transform and associative memories for pattern matching.

- **Step 0 -** Discretize space of track parameters (pattern cells) and generate track intersections with detector planes (receptors) and connect them to cells (mapping).
- **Step 1 -** Detector hits are distributed (Switching Network) only to a reduced number of cells according the mapping of Step 0 (LUT).
- **Step 2 -** A logic unit (engine) for each cell accumulates a Gaussian weight proportional to the distance with the receptors.
- **Step 3 -** Tracks are identified as local maxima of accumulated weights, above a certain threshold, over the cells grid.

Conceived for parallelism: processing before the event building.

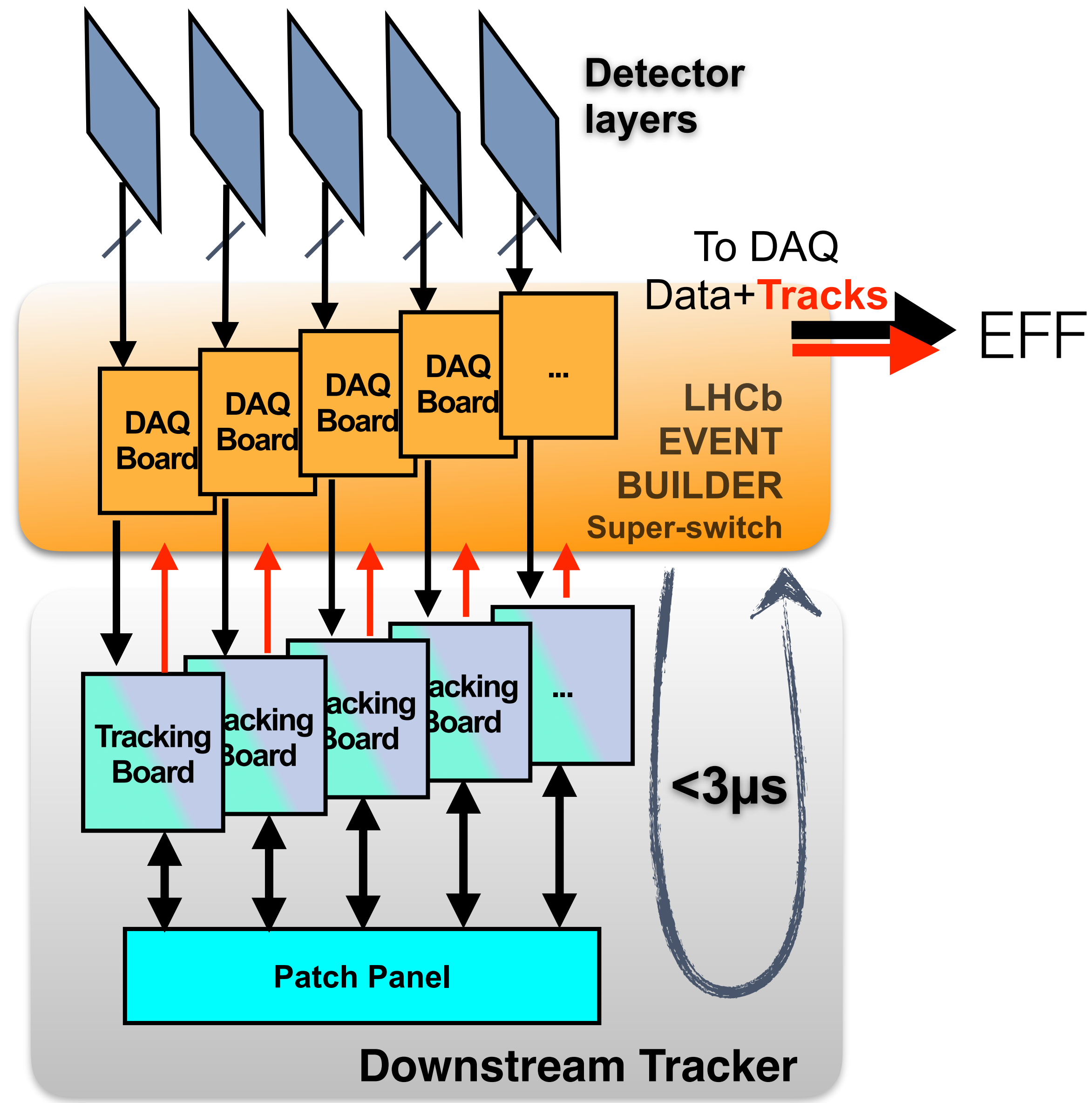


L. Ristori, NIM A453 (2000) 425-429

DAQ integration

- The LHCb-Upgrade Event Builder is made by about 500 interconnected PCs nodes, each of one equipped with a DAQ-board. Events are built and sent to the Event Filter-Farm server nodes, where the HLT will be running [*LHCB-TDR-016*].

- Event Builded PC nodes (of interest) have to be instrumented with a “tracking board” equipped with a modern FPGA connected to the node through a PCIe card (100 Gbit/s).
- A single tracking board performs both hits distribution and template matching. A Patch Panel is needed to distribute hits inter boards.
- Reads small detector portion, outputs small parameter space.
- Easier to implement large global bandwidths.
- May use standard commercial PCIexpress FPGA boards.



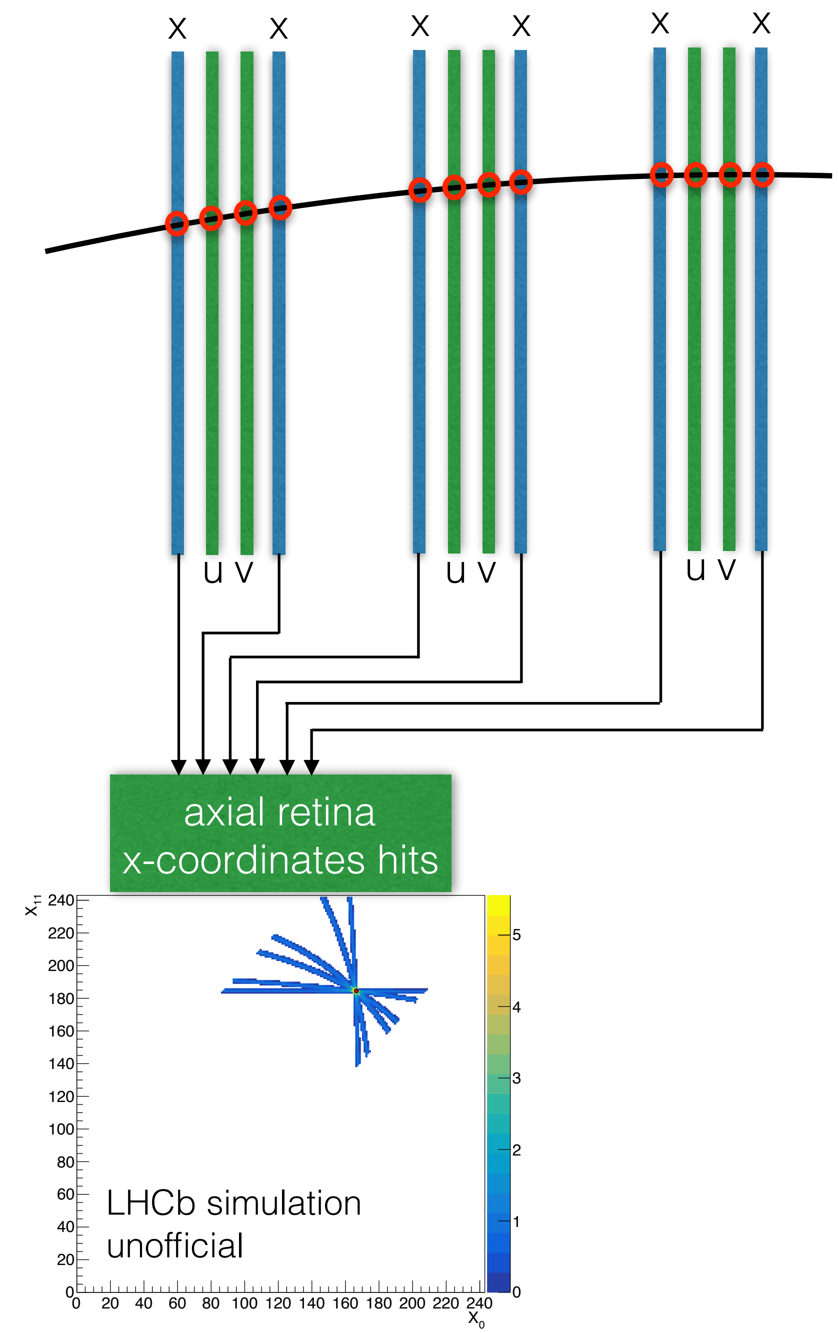
156 DAQ boards will receive raw data from SciFi sub-detector, half from the axial layers (6 layers providing x-coordinates) and half from the stereo layers (6 layers providing u- and v-coordinates).

Number of Event Builder PC nodes

Sub-detectors	VELO	UT	SciFi	RICH	Calorimeter	Muons
EB node	56	72	156	114	55	30

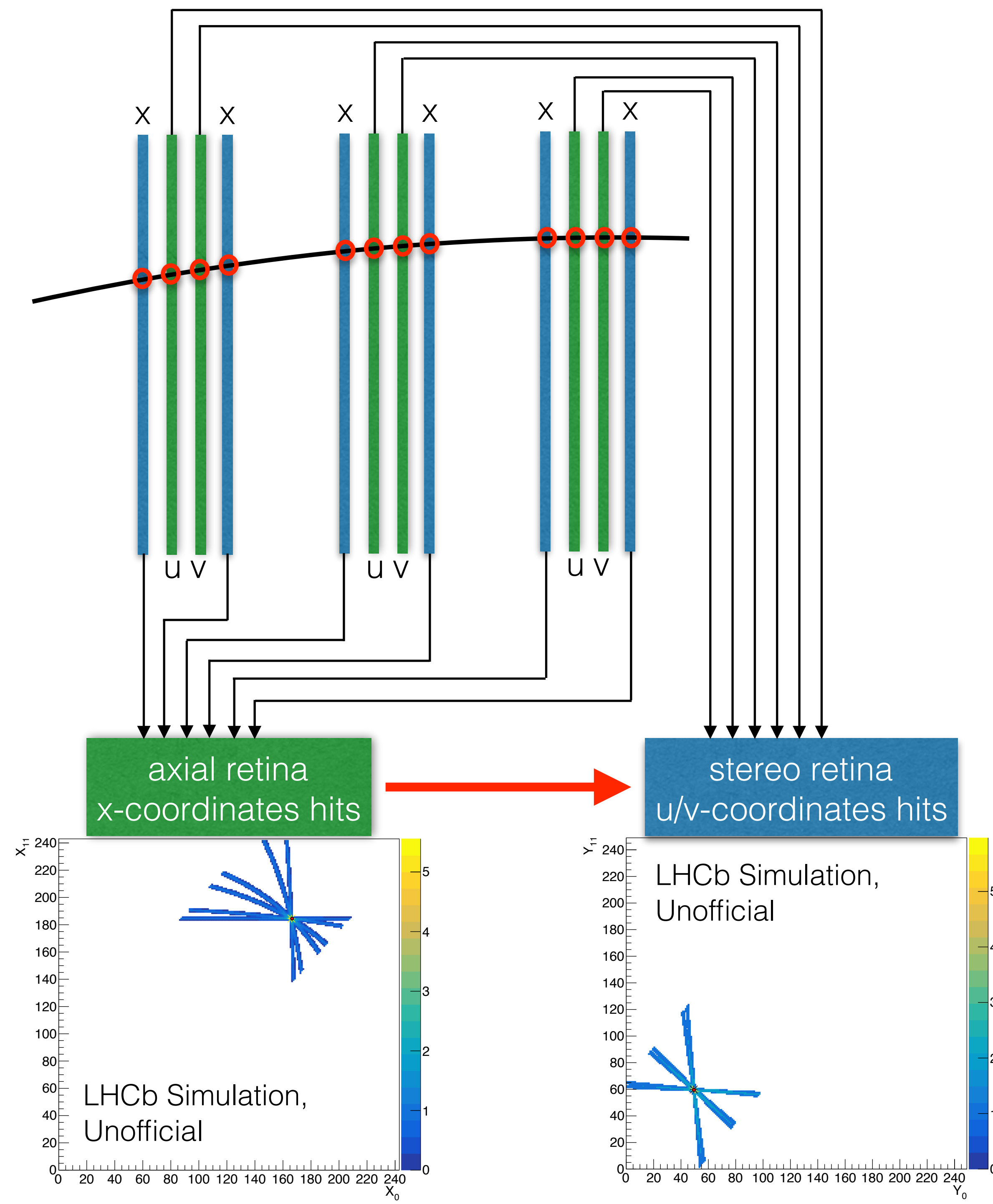
Reconstruction of 3D tracks

- Reconstruction of T-tracks is factorized in two stages.
- **Pattern recognition:** find the x-z track projection using only axial layers.
    - tracks approximated as straight lines (2-dim Retina).
    - for each local maximum found, a linear χ^2 fit to a parabola is executed (on DSP blocks of FPGAs) over a limited number of combinations of the two closest hits to the receptors, in order to kill ghost tracks and evaluate parabola parameters.



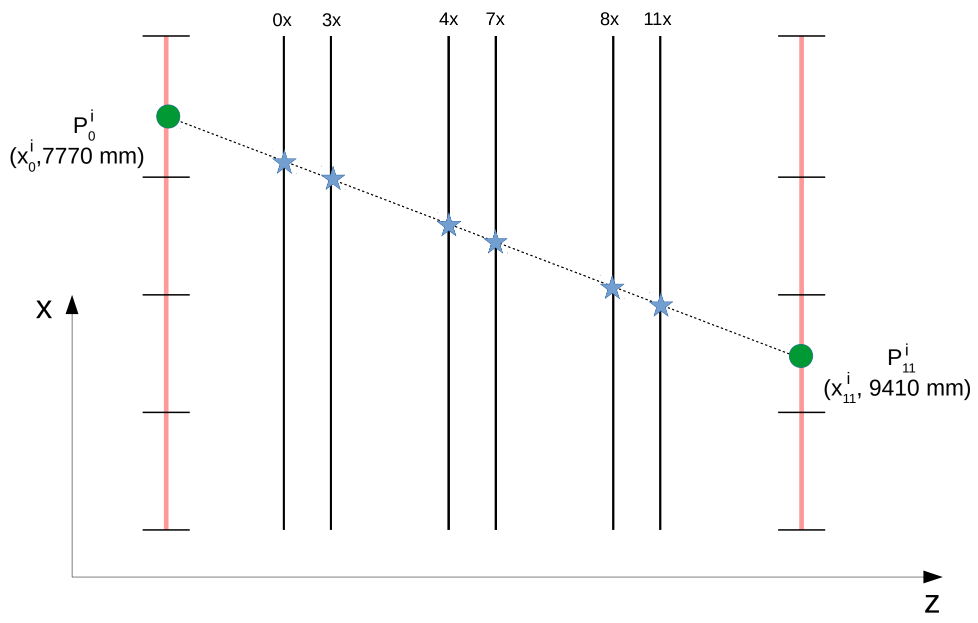
Reconstruction of 3D tracks

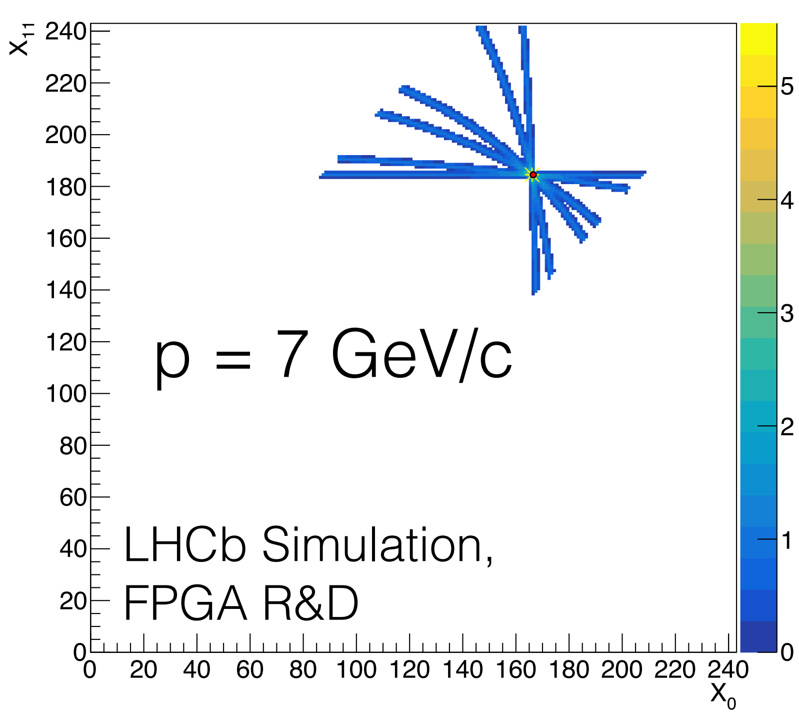
- Reconstruction of T-tracks is factorized in two stages.
- **Pattern recognition:** find the x-z track projection using only axial layers.
    - tracks approximated as straight lines (2-dim Retina).
    - for each local maximum found, a linear χ^2 fit to a parabola is executed (on DSP blocks of FPGAs) over a limited number of combinations of the two closest hits to the receptors, in order to kill ghost tracks and evaluate parabola parameters.
- **Stereo association**: x-z projection of axial track candidate is used as “seed” to extract y-coordinates from u/v layers and associate y-z track projection. The same firmware of axial Pattern Reco stage is used, but with much lower granularity.



Tracking 6-layers of SciFi

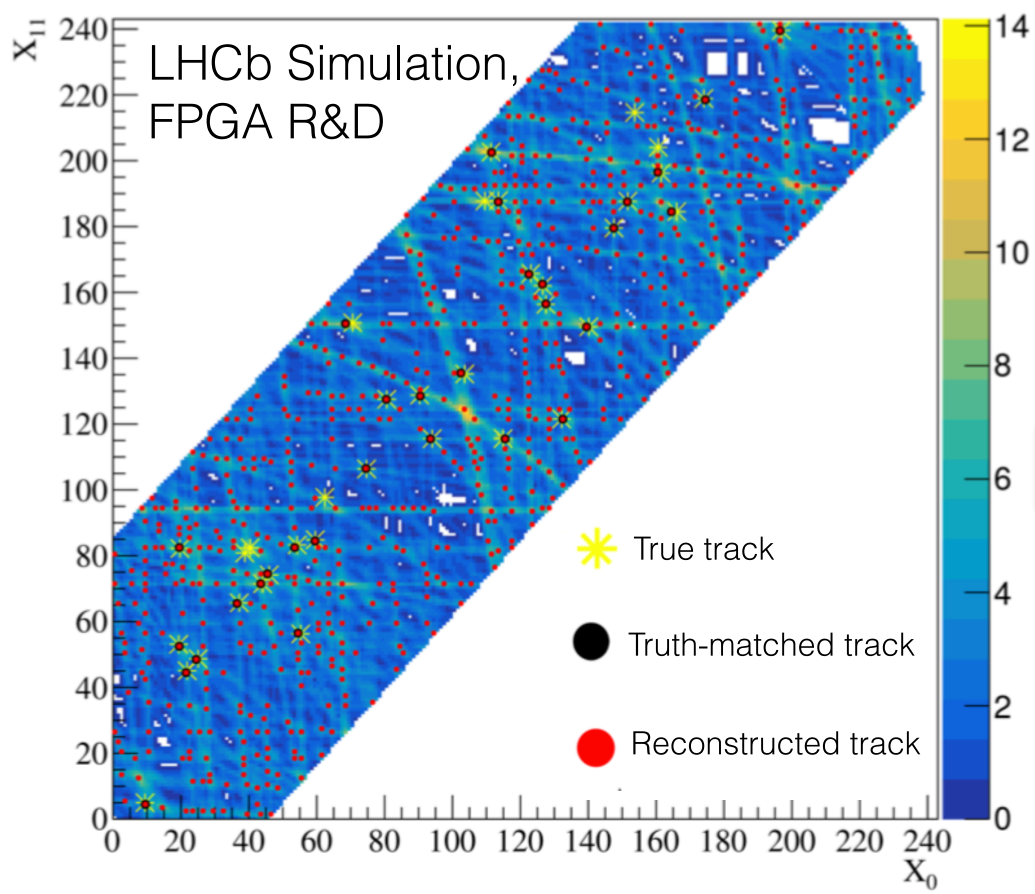
- Axial retina track parameterization:
  - x_0 (x_{11}) x-coordinates of the intersection between the first (last) axial layer;
  - similar for the stereo association y_0 (y_{11}).
- Interesting physics tracks (Long and Downstream) distributed over the diagonal region, being $x_0 \approx x_{11}$.
- About 25k cells per quadrant, for a total of about 100k cells for the whole SciFi sub-detector.
- Cells granularity of axial retina is the minimal to ensure the presence of the true track hits within the two closest hits of each receptor, once a local maximum is found.



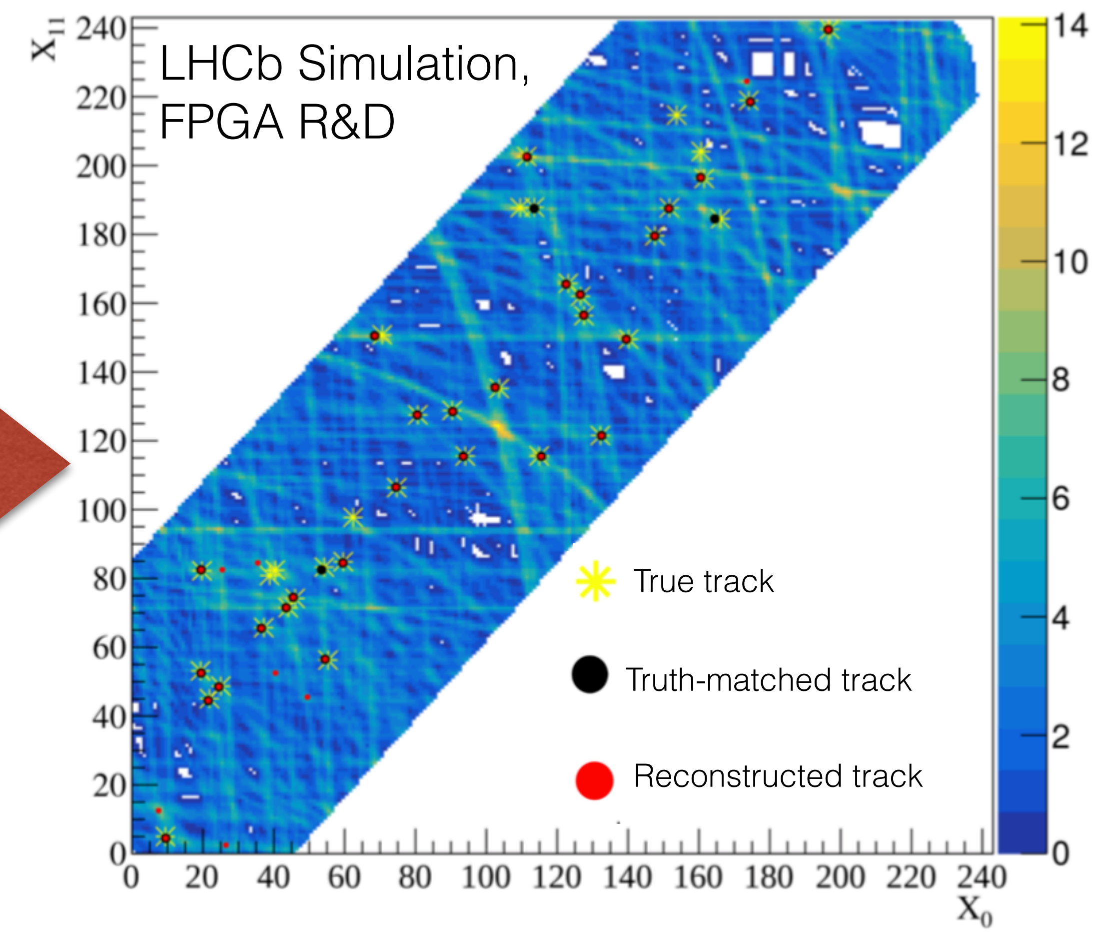


Axial retina at work

Excitation level $W_A > 4$



Excitation level $W_A > 4$
$\chi^2_A < 20$

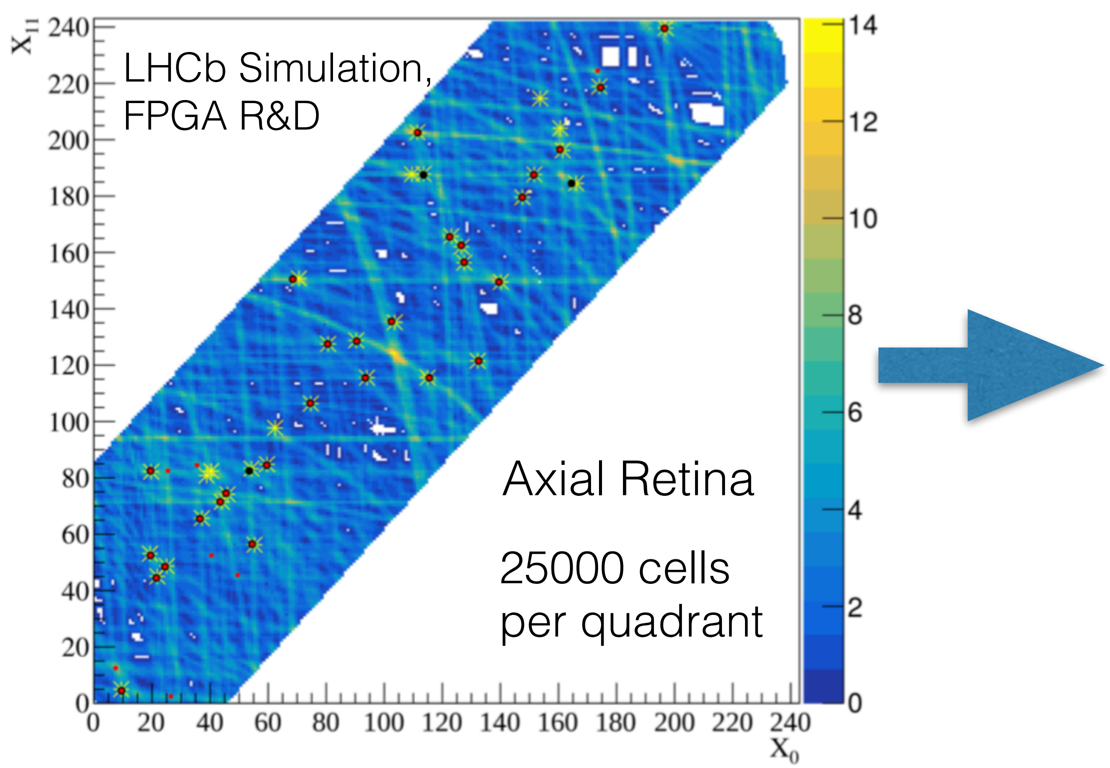


- Efficiency above 95%.
- Ghost rate ~ 90% (for instance 150 true-matched reconstructed tracks correspond to 1350 “fake” reconstructed tracks).

- Efficiency ~ 80-90%. (approaching 90% with some minimal momentum request).
- Ghost rate ~ 15% (comparable with standard offline seeding algorithm).

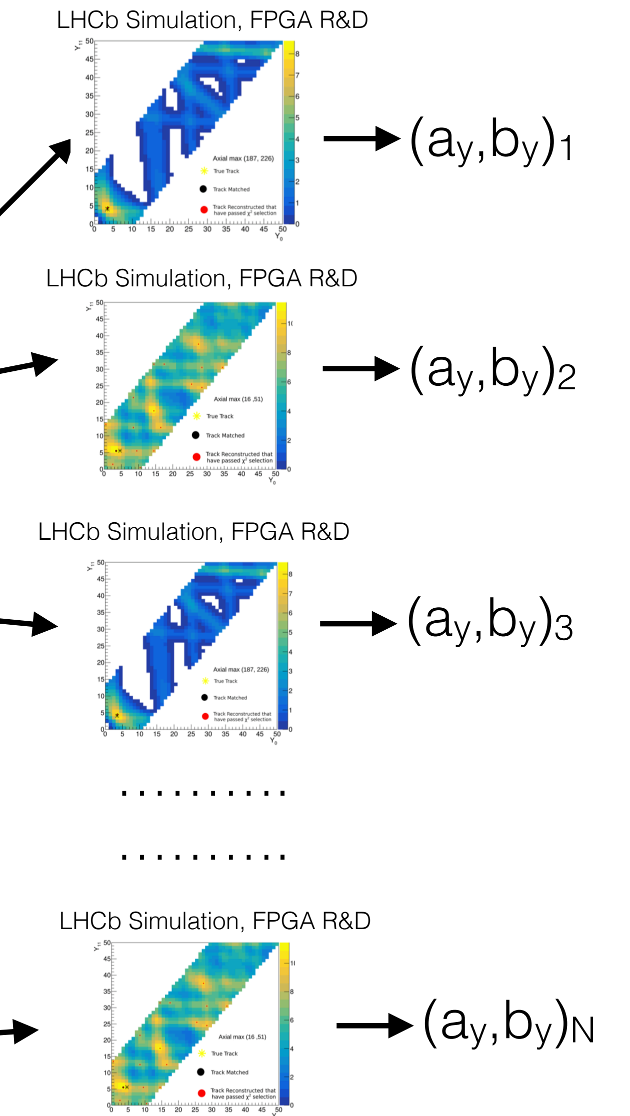
Stereo Association

- For each reconstructed axial tracks candidate, the u/v-hits from tilted layers are transformed in y-coordinates and accumulated into the corresponding stereo retina (y_0, y_{11}).
- Same processing of the x-z view is performed (hits distribution, accumulation, search of local maxima and linear fit of the combinations of hits accumulated in the local maxima).
- The y-hits combination with the best χ^2 is promoted as the y-z track projection.



$(a_x,b_x,c_x)_1$
$(a_x,b_x,c_x)_2$
$(a_x,b_x,c_x)_3$
$(a_x,b_x,c_x)_4$
…………………
…………………
$(a_x,b_x,c_x)_i$
…………………
………………
………………
$(a_x,b_x,c_x)_N$

Stereo Retinas
(355 cells each)



Each stereo tracking board hosts at least 3 stereo retinas (355 cells each), for a total of 355 x 3 ≈ 1k cells per FPGA.

Hardware prototypes

- First prototypes assembled and tested within the INFN Retina CSN5 project.

- Used an ASIC prototyping board by DiniGroup equipped with two large FPGA's:

  - 2 Stratix V (28 nm), 1M of LEs, with (0.6+0.6)Tb/s I/O bandwidth, and maximum clock of 700 MHz;

  - on-board CPU, DDR memory;

  - 96 inter-FPGA LVDS connections;

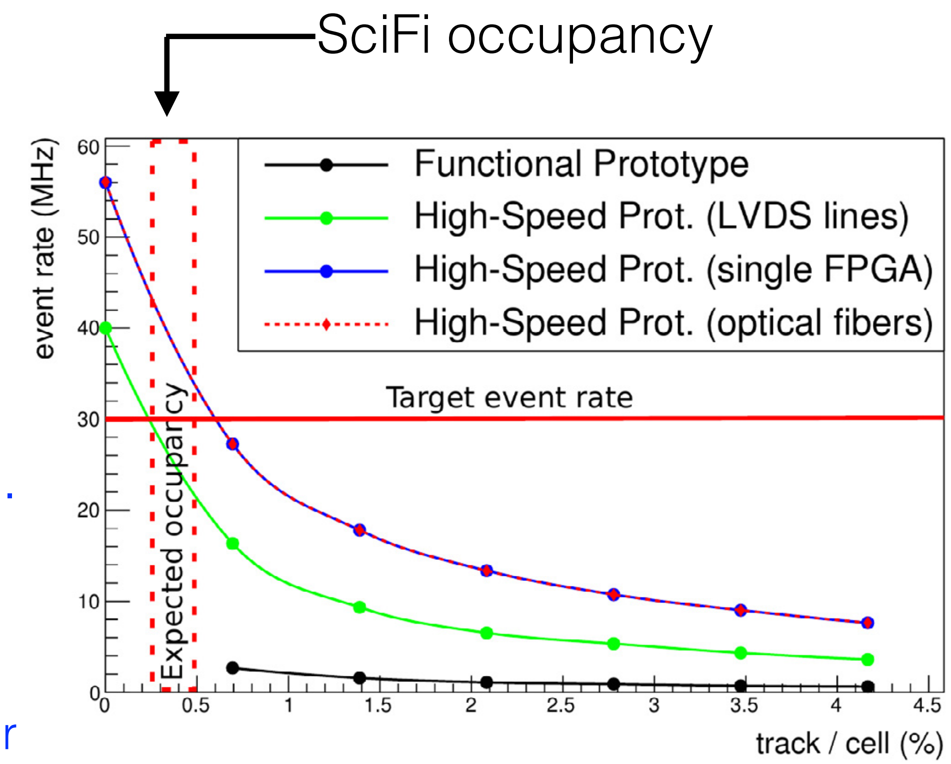
  - 96 high-speed SerDes I/O (12 Gb/s).

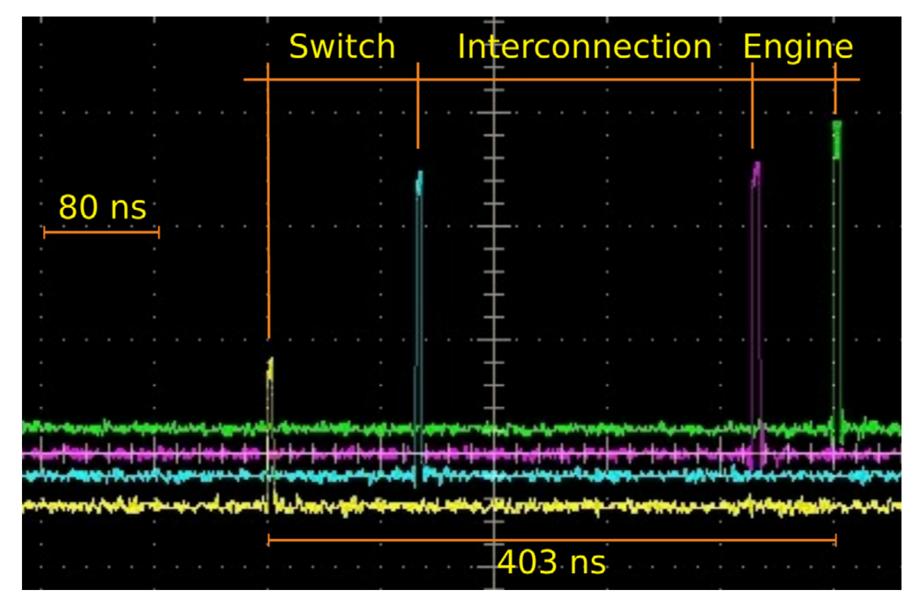
- Switch and engine can be tested in basic standalone configurations, but more are possible connecting other boards through optical fibers.

- Can be used as building block for an entire tracker. Results on this prototype therefore readily extrapolate to real systems size.

Measured performance

- Various configurations tested. Here, results on 6 axial layers of Scintillating Fibre Tracker as a function of retina occupancy (#track/#cell).
- Achieved a throughput above 30MHz and latencies <1μs.
- Determined engine size: about 1k LEs.
  - Pattern Reco. (axial): needed about 100k engines distributed over 78 FPGAs, corresponding to ~1.2M LEs per FPGAs.
  - Stereo Association: needed about 78k engines distributed over second half of EB PC nodes, corresponding to ~1M LEs per FPGAs.
- Such a specialized processor is already cost effective and can be built today using FPGAs off the shelf.
  - Hardware cost < 0.1 €/KHz of tracks.
  - Power cost: 0.2 mW/kHz of tracks.





R.Cenci et al., PoS(TWEPP-17)136
F. Lazzari et al., NIMA.2018.08.025

Physics performance

- Simulated realistic pp collisions at the LHCb-Upgrade conditions:
  - sqrt(s)=14 TeV;
  - LHC bunch spacing = 25 ns;
  - L=2 x 10^{33}cm^{-2}s^{-1};
  - ν = 7.6
- Events processed using the official LHCb-Upgrade simulation of the Scintillating Fibre sub-detector (energy deposits, spillover, noise, inefficiencies, light attenuation in the fibre, clustering, read-out electronic chain, etc.)
- Performances comparable with those obtained with the offline seeding algorithm. Efficiency within the 80-90% range, approaching the 90% level with a minimal momentum threshold.

$D^{*+} \rightarrow D^0\pi^+ \rightarrow [K_S\pi^+\pi^-]\pi^+$

MinBias

$B_s \rightarrow \phi\phi$

	Track type	*Sample 1* ε_A	*Sample 1* ε_{AS}	*Sample 2* ε_A	*Sample 2* ε_{AS}	*Sample 3* ε_A	*Sample 3* ε_{AS}
T-track	hasT	75.0	71.4	74.4	70.0	73.9	67.4
	hasT, $p > 1\mathrm{GeV/c}$	77.6	73.9	77.7	73.1	76.4	69.7
	hasT, $p > 3\mathrm{GeV/c}$	87.0	83.0	85.9	80.8	85.1	77.2
	hasT, $p > 5\mathrm{GeV/c}$	90.3	85.7	88.2	82.7	86.6	77.4
Long	Long	81.7	78.8	84.1	79.5	84.2	77.2
	Long, $p > 1\mathrm{GeV/c}$	81.7	78.8	84.1	79.5	84.2	77.2
	Long, $p > 3\mathrm{GeV/c}$	87.3	84.2	87.1	82.3	87.3	79.8
	Long, $p > 5\mathrm{GeV/c}$	90.6	86.9	88.1	83.1	88.1	79.9
Downstream	Down	80.1	77.7	83.0	78.6	82.6	76.2
	Down, $p > 1\mathrm{GeV/c}$	80.1	77.7	83.0	78.6	82.6	76.2
	Down, $p > 3\mathrm{GeV/c}$	87.0	84.4	87.1	82.5	86.5	79.3
	Down, $p > 5\mathrm{GeV/c}$	90.5	87.5	88.8	83.6	87.9	80.2
Downstream from strange	Down strange	-	-	84.7	82.8	-	-
	Down strange, $p > 1\mathrm{GeV/c}$	-	-	84.7	82.8	-	-
	Down strange, $p > 3\mathrm{GeV/c}$	-	-	89.4	86.7	-	-
	Down strange, $p > 5\mathrm{GeV/c}$	-	-	93.0	87.2	-	-
	ghost rate	12.1	15.7	16.3	20.2	18.4	24.7

LHCb Simulation, FPGA R&D

ε_A = efficiency of finding axial projection of the track
ε_{AS} = efficiency of finding 3D track

Conclusions

- Reconstruction of standalone tracks in the forward LHCb SciFi sub-detector in realtime at 30MHz, with latencies of the order of 1 μs, is doable with a system based on modern commercial PCIe FPGA cards.

- Such a specialized processor is cost effective and might be built already today using FPGAs off the shelf.

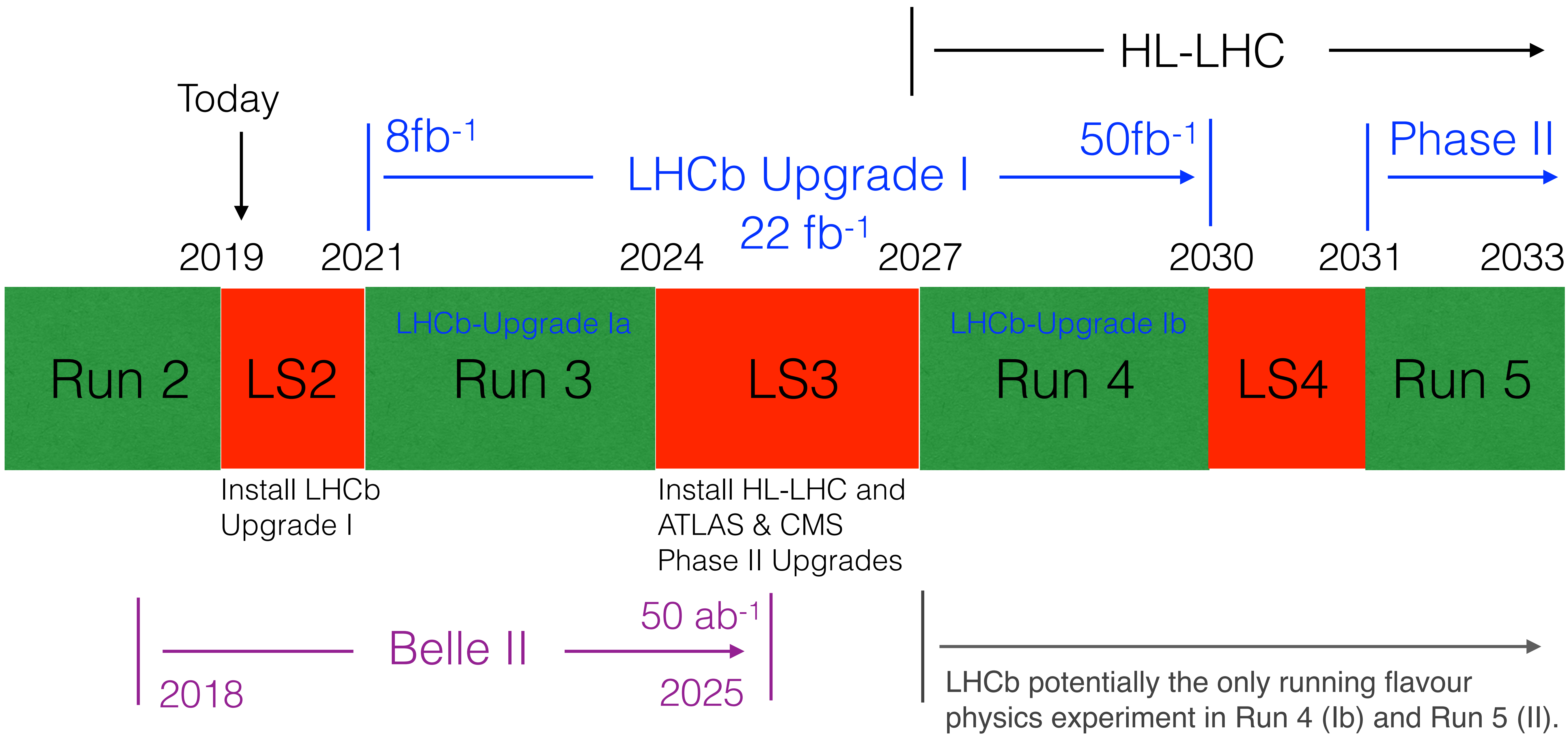
- Achieved a crucial milestone in the R&D development for a future realization of a downstream tracking unit to operate during LHC Run 4 (2026-2029) and beyond.

Backup

Selected list of references

- F. Lazzari et al, Performance of a high-throughput tracking processor implemented on Stratix-V FPGA, 2018, NIM A (2018) 08,025.
- R. Cenci et al., Development of a High-Throughput Tracking Processor on FPGA Boards, 2017, PoS(TWEPP-17)136 .
- R. Cenci et al. First results of an "artificial retina" processor prototype, 2016 https://doi.org/10.1109/MOCAST.2016.7495111.
- R, Cenci et al, First Results of an “Artificial Retina” Processor Prototype, 2016, EPJ Web Conf. 127 (2016) 00005.
- G. Punzi et al., A Specialized Processor for Track Reconstruction at the LHC Crossing Rate, 2014, JINST 9, C09001 (2014).
- A. Abba et al., A specialized track processor for the LHCb upgrade, 2014, LHCb-PUB-2014-026.
- A Di Luca, Real-time reconstruction of tracks in the Scintillating Fibre Tracker of the LHCb Upgrade, 2018, CERN-THESIS-2018-272.
- F. Lazzari, Development of a real-time tracking device for the LHCb Upgrade 1b, 2017, CERN-THESIS-2017-442 .
- A. Piucci, Reconstruction of tracks in real time at high luminosity environment at LHC, 2014, etd-06242014-055001.
- D. Ninci, Real-time track reconstruction with FPGA at LHC, 2014, etd-11302014-212637 (in italian).

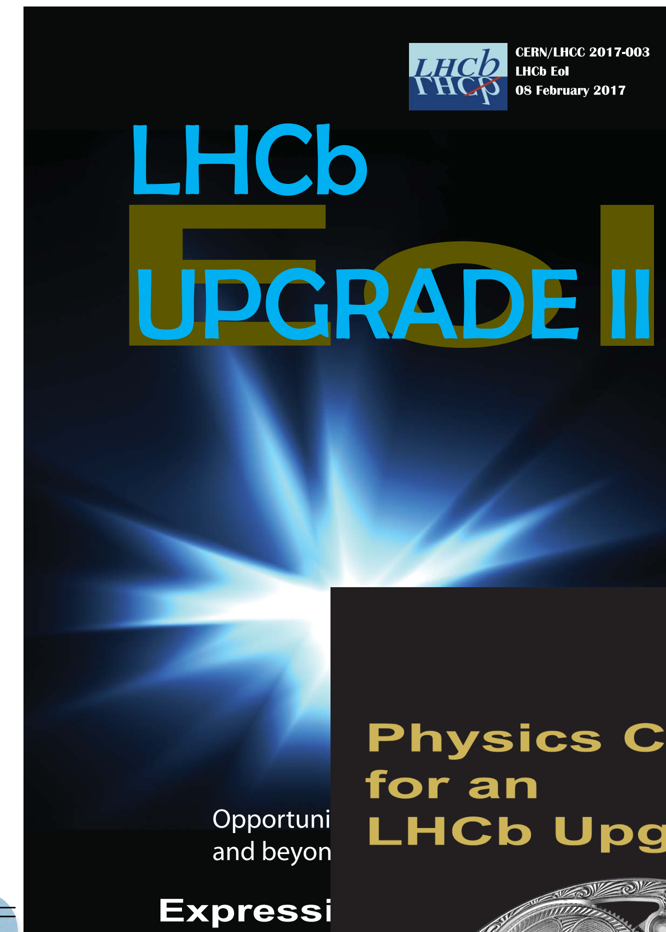
LHCb timeline in the next decades



The LHCb Upgrade I will enable to integrate about 22 fb^{-1} by end of Run 3 and 50 fb^{-1} by end of Run 4.

Intensity frontier

- LHCb Upgrade Ia in Run-3 (2021-2023)
  - L_{inst}= 2 x 10^{33} cm^{-2} s^{-1}.
- LHCb Upgrade Ib Run-4 (2026-2029)
  - Integrate 50 fb^{-1} by the end of Run 4.
  - Profit from LS3 for a “consolidation”.
- LHCb Upgrade II in Run 5 (2031-2033) and beyond.
  - New experiment to be installed in LS4 to integrate > 300 fb^{-1}.



CERN-LHCC-2017-003

CERN-LHCC-2018-027

	2010–12	2015–18	2021–23	2026–29	2031-33
LHC Run	1	2	3	4	5
$E_{\rm cm}$ (TeV)	$7-8$	13	14	14	14
LHC $L_{\rm peak}$ ($\rm cm^{-2}\,s^{-1}$)	$7.7\cdot10^{33}$	$1.7\cdot10^{34}$	$2\cdot10^{34}$	$7\cdot10^{34}$	$7\cdot10^{34}$
LHCb $L_{\rm peak}$ ($\rm cm^{-2}\,s^{-1}$)	$2-4\cdot10^{32}$	$2-4\cdot10^{32}$	$2\cdot10^{33}$	$2\cdot10^{33}$	$>10^{34}$

This talk: R&D developed in the context of the LHCb Upgrade Ib (LHC Run 4 and beyond), specifically aimed at the reconstruction of long-lived particles.

LHCb Upgrade I (Run 3)

VErtex LOcator

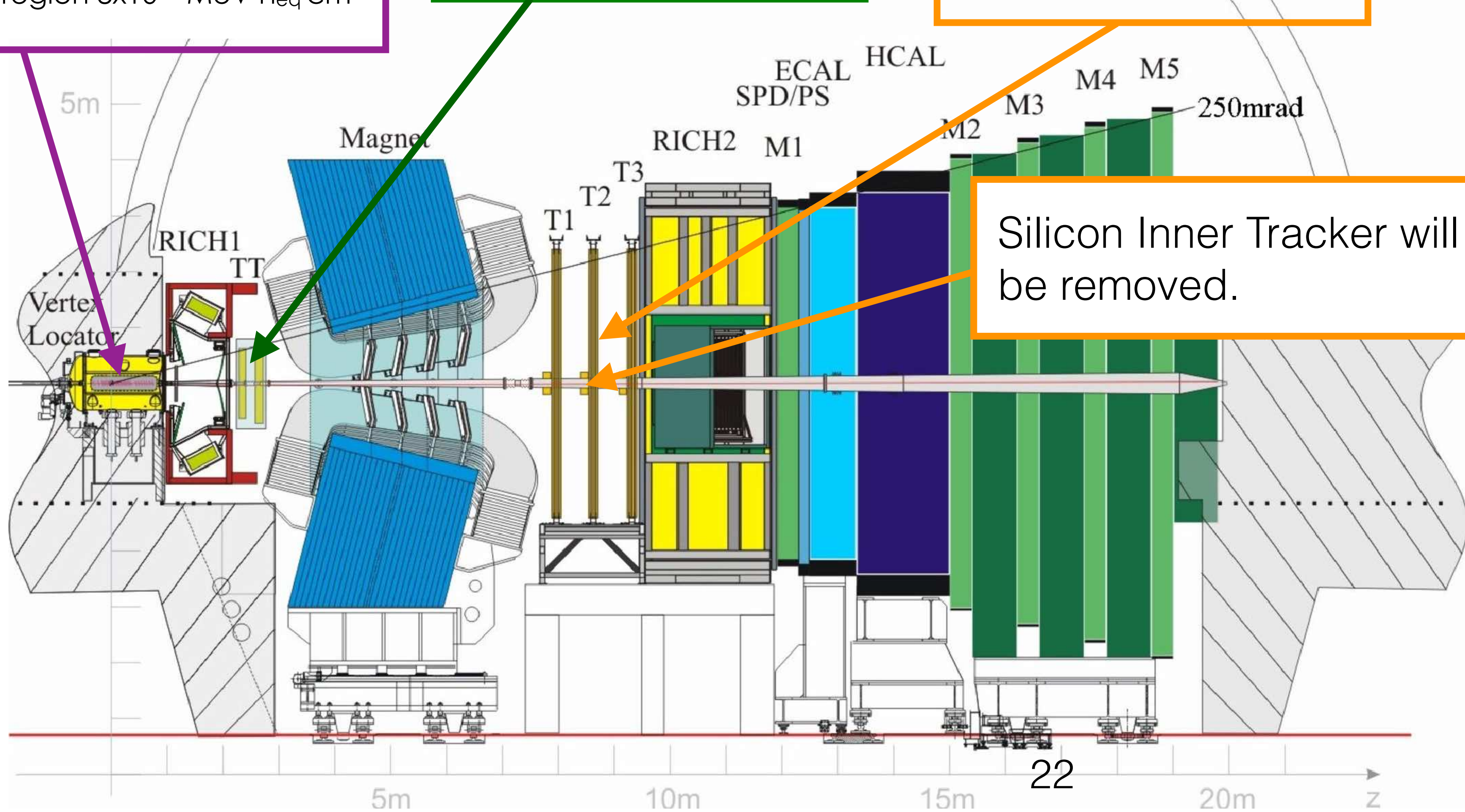
- From strip sensors to hybrid pixel detectors
- Closer to the beam (from 8.2 mm to 5.1 mm)
- CO_2 cooling in micro-channel substrate
- New RF box
- Fluence in the innermost region 8×10^{15} MeV n_{eq} cm^{-2}

Upstream Tracker

- 4 planes (x-u-v-x) of Si strips sensors
- Staves staggered: overlap in x
- Closer to the beam pipe
- Finer granularity
- Bi-phase CO2 cooling in stave support

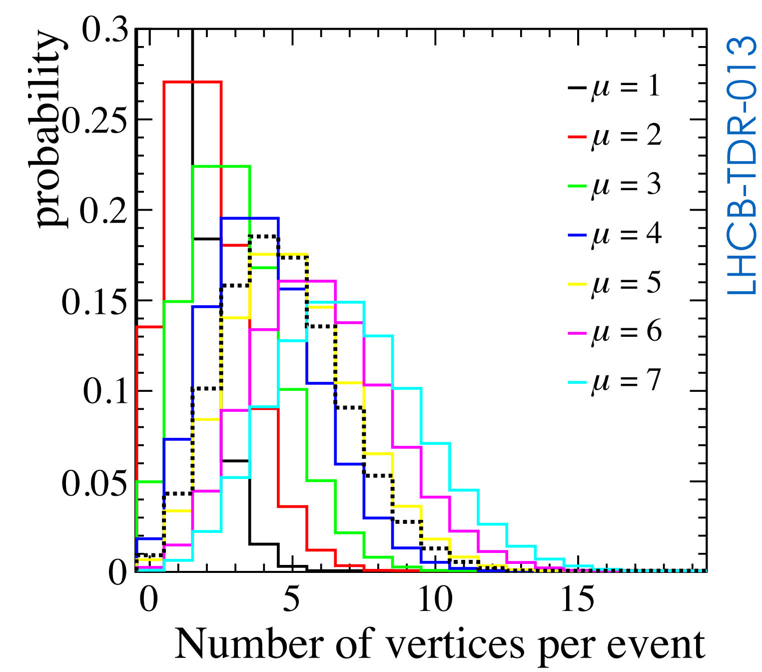
SCIntillating FIbres

- Scintillating fibres as active detector elements
- 3 stations with 4 detection layers (x-u-v-x)
- 2 x 2.5 m long modules with mirror in the middle
- Readout with SiPMs at the outer edge



Upgrade (current) Conditions

- visible interactions=5.5 (1.1)
- $\sqrt{s}$=14 TeV (13 TeV)
- lumi: 2×10^{33} cm^{-2} s^{-1} (4×10^{32} cm^{-2} s^{-1})
- expected integrated lumi: 50 fb^{-1} (8 fb^{-1})



Major upgrade of the electronics to allow the read-out of all sub-detectors at 40MHz.

LHCb acceptance of LLPs

- Considering only VELO, LHCb il long about 50 cm.
- First UT layer is 2.3 m from the interaction point.
- Assuming an average momentum of 20GeV for LLPs, LHCb acceptance increases
  - K_S: $N_{DD} \approx 1.3\ N_{LL}$ (x2.3 in stat.)
  - K_L: $N_{DD} \approx 3.4\ N_{LL}$ (x4.4 in stat.)
  - Λ: $N_{DD} \approx 1.7\ N_{LL}$ (x2.7 in stat.)

	Mass (MeV)	cτ	L=βγcτ (at p=20GeV)
D^0	1864	123μm	1.3mm
B^0	5279	456μm	1.7mm
K_S	497	2.7cm	108cm
K_L	497	15.3m	615m
Λ	1115	7.9cm	141cm

Long Lived Particles at LHCb

LHCb-Upgrade readout system

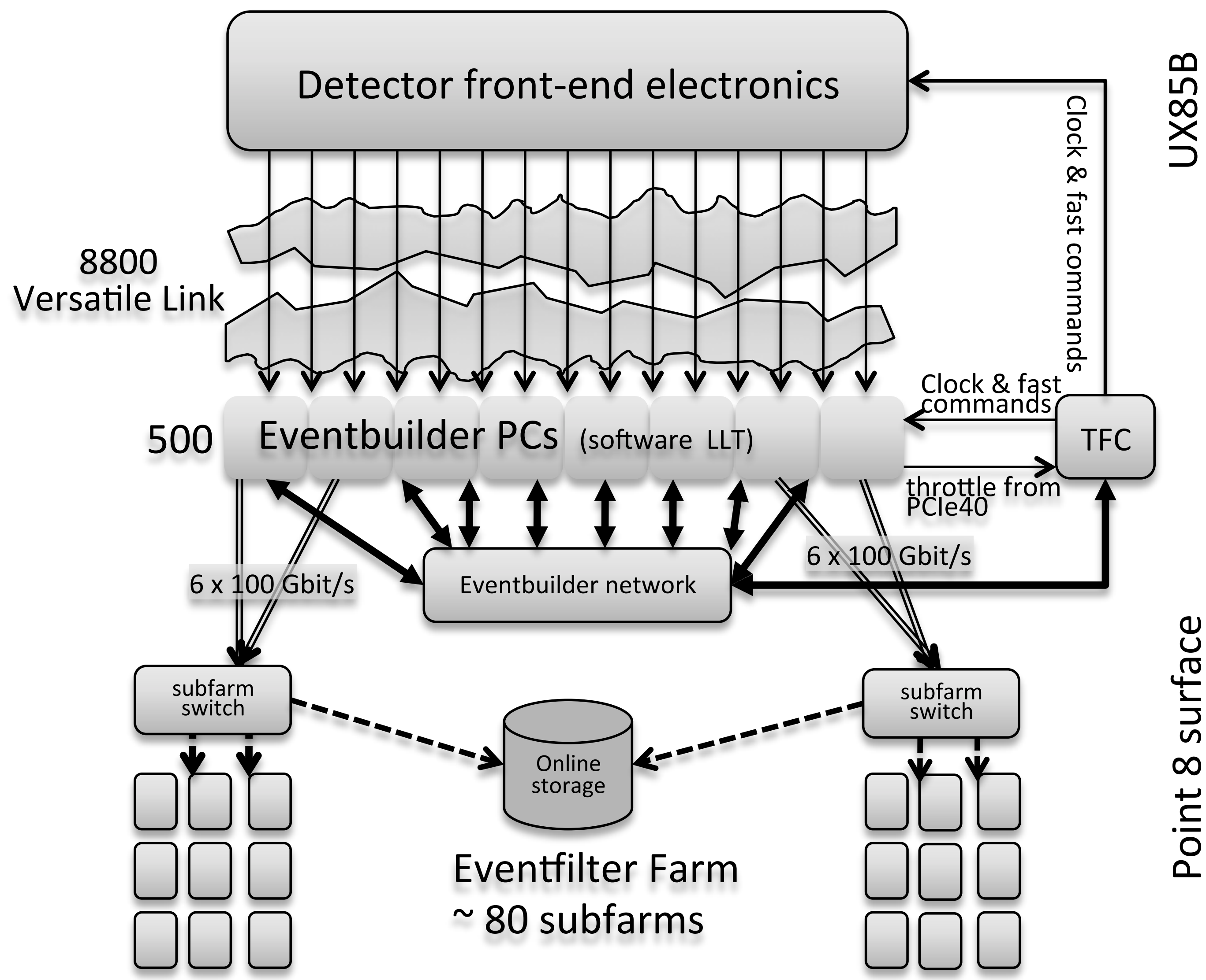


Figure 2.23: The architecture of LHCb Upgrade-I readout-system.

DAQ Integration

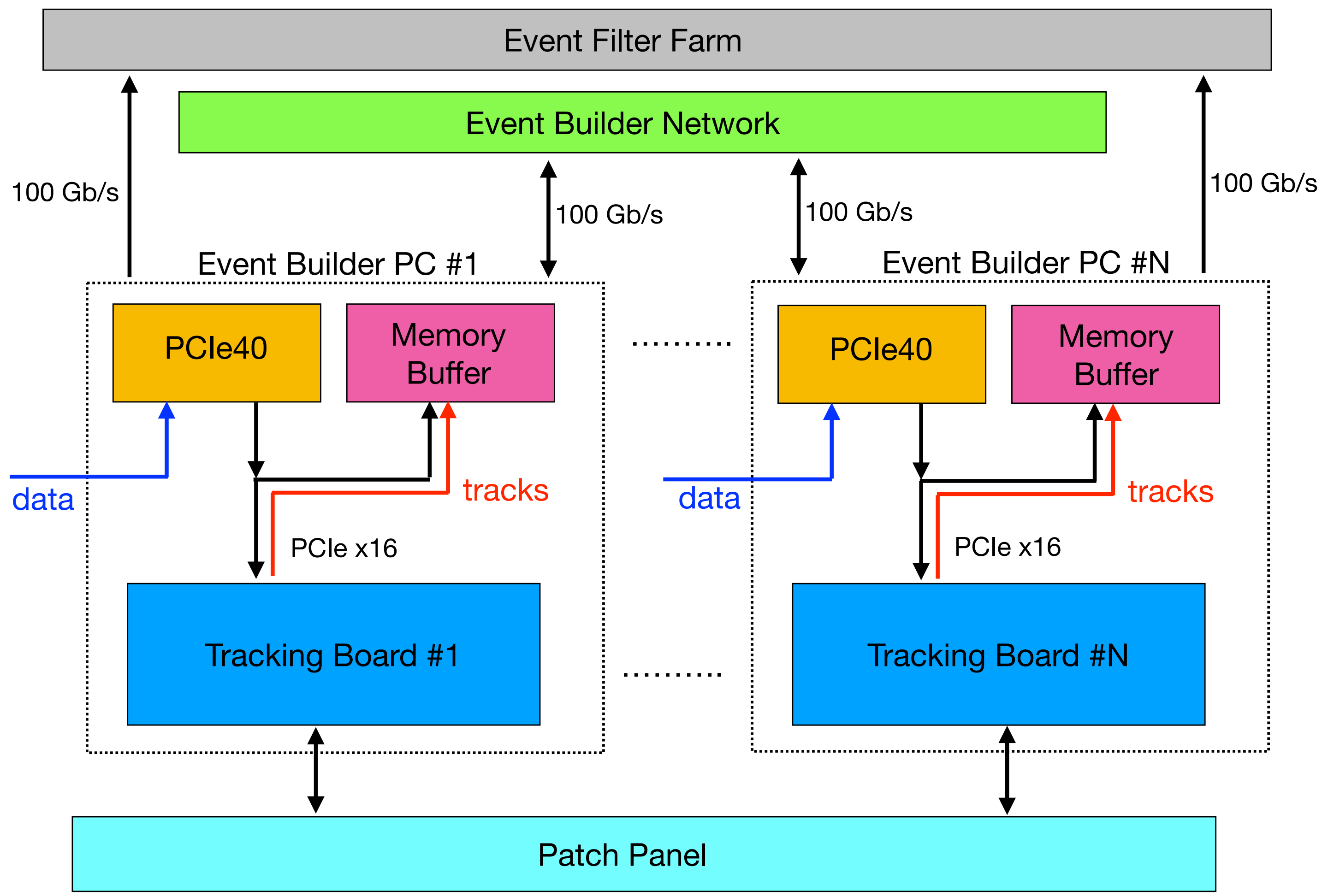
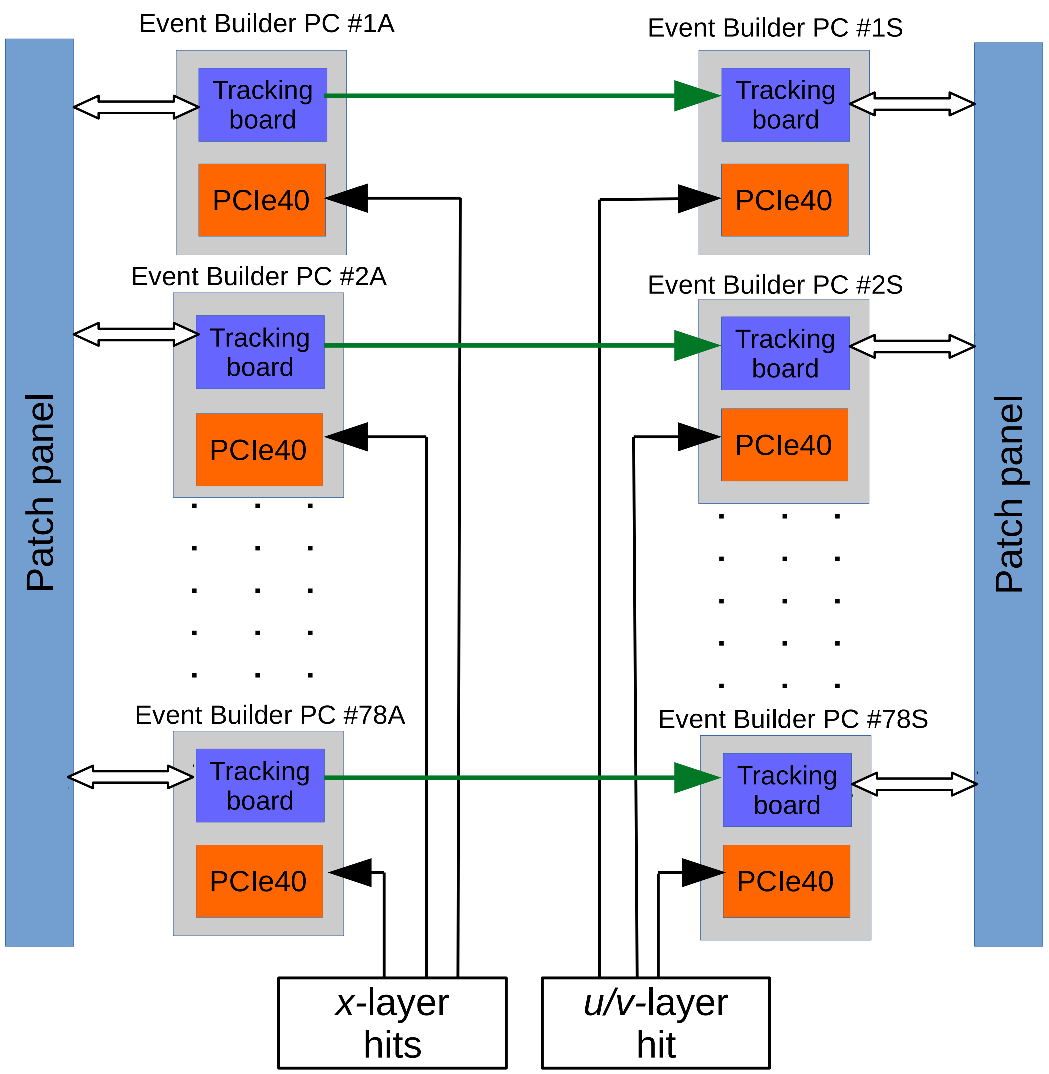


Figure 3.5: Integration scheme of the Downstream Tracker into the LHCb Upgrade Event Builder system.

Envisioned architecture

- 156 Event Builder PC nodes dedicated to the SciFi tracker.
    - half receives raw hits from the x-layers and the second half receives raw hits from the tilted u/v-layers.
- Each EB node has to be instrumented with a “tracking board” equipped with a modern FPGA (at least ~1MLEs per chip is needed) interconnected through PCIe card (100 Gbit/s).
- Two optical Patch Panels and Switching Networks needed to distribute x-hits and u/v-hits inter boards.
- The axial retina (10^5 cells) is distributed over 78 tracking boards (left side), while each stereo tracking board (right side) should host at least 3 different stereo retinas (355 cells each).
- Once axial tracks are ready, they are sent (**green arrows**) to the corresponding stereo retinas for the “stereo association”, with a total latency of the process of ~ 1 μs, well below the Event Builder timing constraints.



Track model

- A small component of magnetic field (fringe field) is present in the SciFi region.

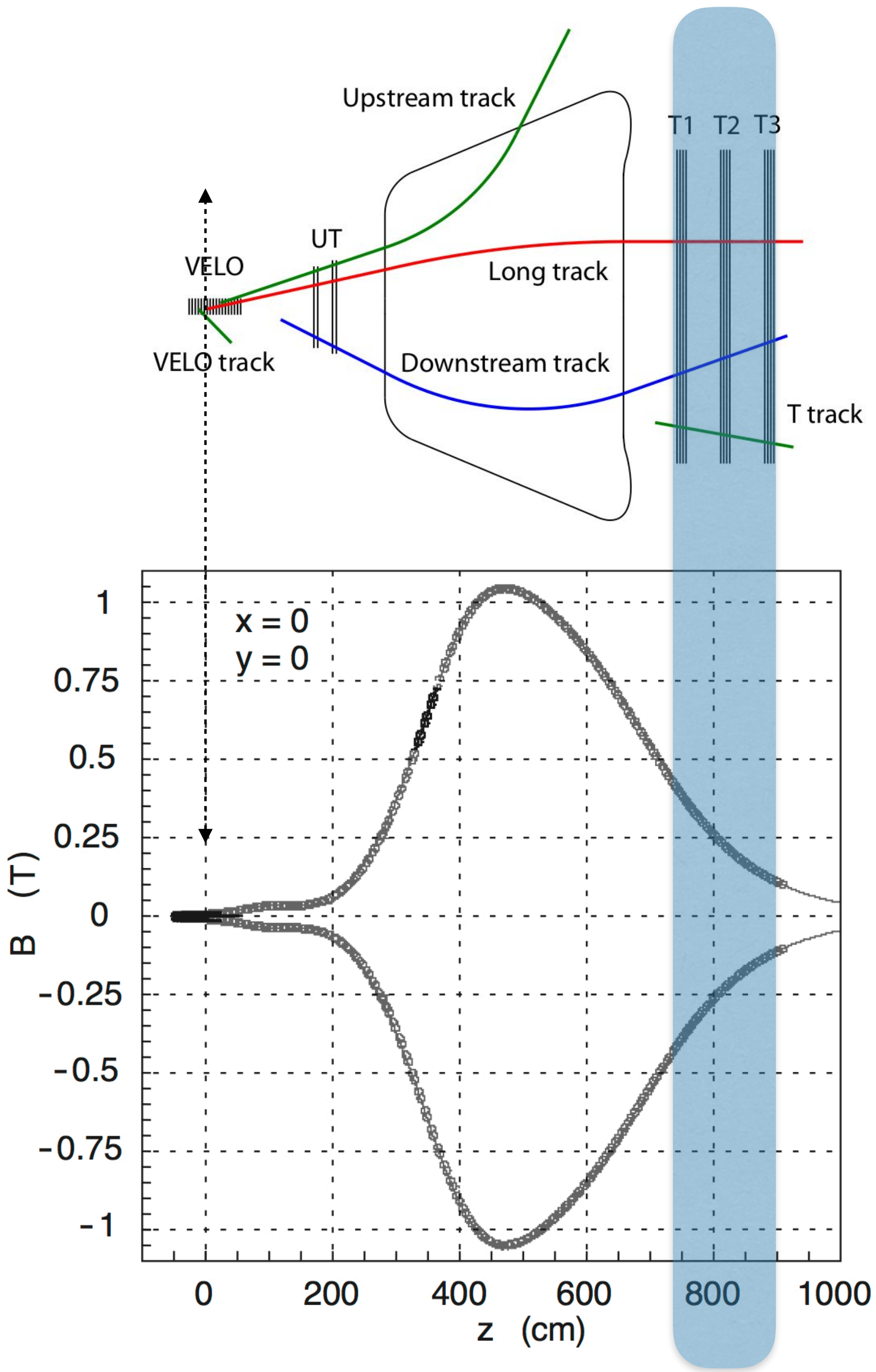
$$\frac{\mathrm{d}\vec{p}}{\mathrm{d}t} = q\vec{v} \times \vec{B}$$

- Assuming B_y small and constant over the x-z plane:

$$x(z) = x_0 + t_x(z - z_0) + \frac{q}{p}\frac{B_y}{2}(z - z_0)^2$$

- Assuming null B_x and B_z components:

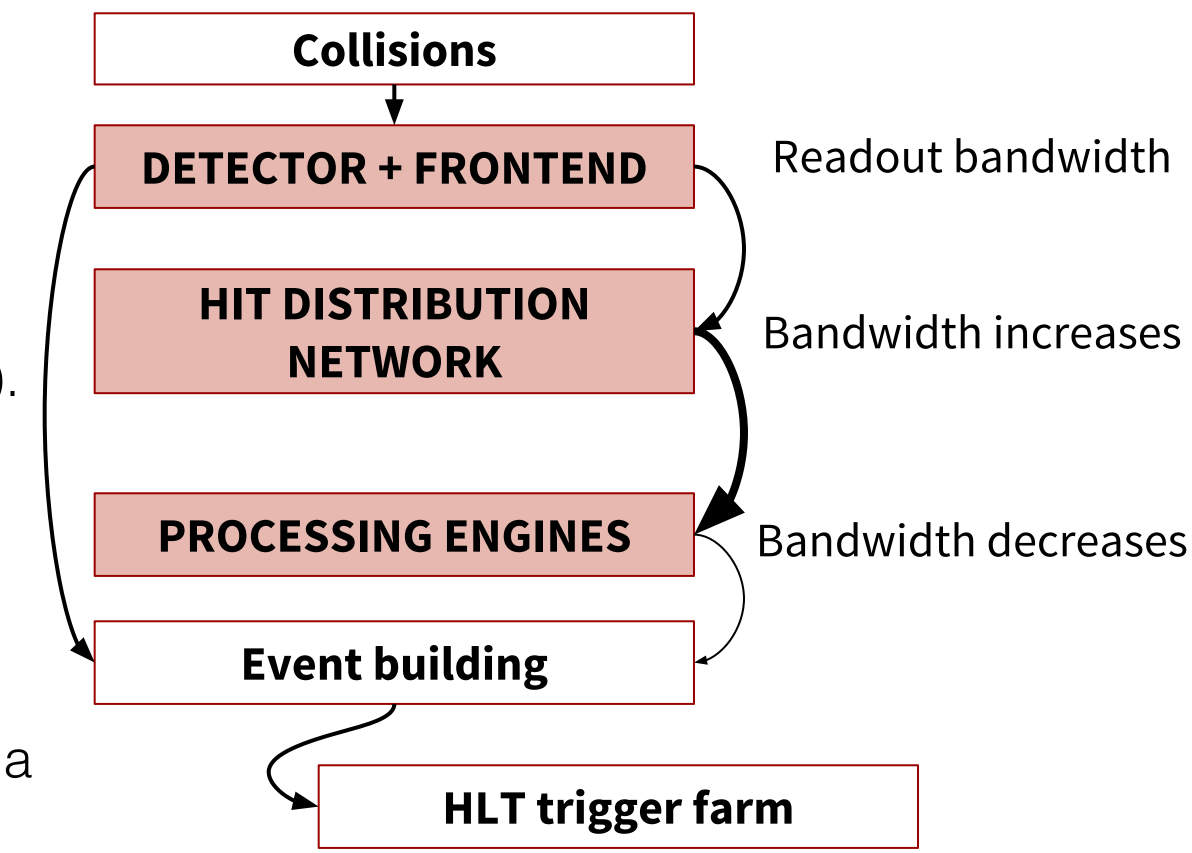
$$y(z) = y_0 + t_y(z - z_0)$$



A biologically inspired architecture: the Retina Algorithm

Massively parallel architecture. Similarities with Hough transform and associative memories for pattern matching.

- **Step 0 -** Discretize space of track parameters (pattern cells) and generate track intersections with detector planes (receptors) and connect them to cells (mapping).
- **Step 1 -** Detector hits are distributed (Switching Network) only to a reduced number of cells according the mapping of Step 0 (LUT).
- **Step 2 -** A logic unit (engine) for each cell accumulates a Gaussian weight proportional to the distance with the receptors.
- **Step 3 -** Tracks are identified as local maxima of accumulated weights, above a certain threshold, over the cells grid.



Conceived for parallelism: processing before the event building.

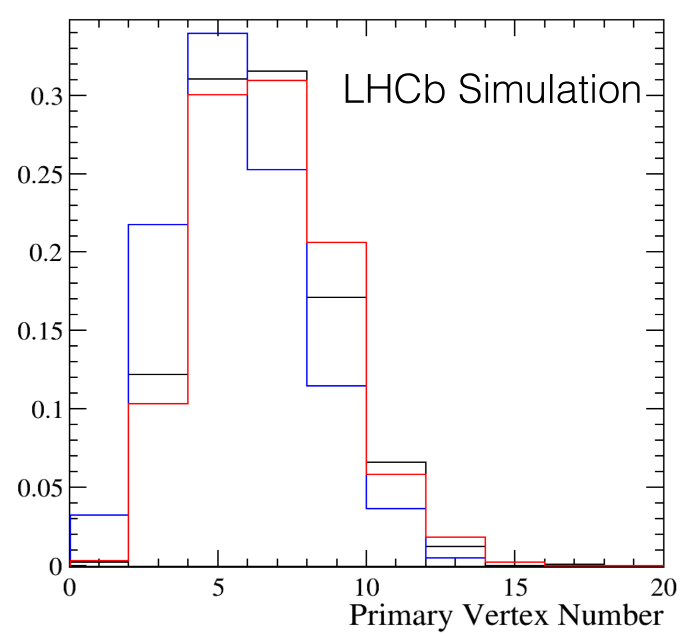
L. Ristori,NIM A453 (2000) 425-429

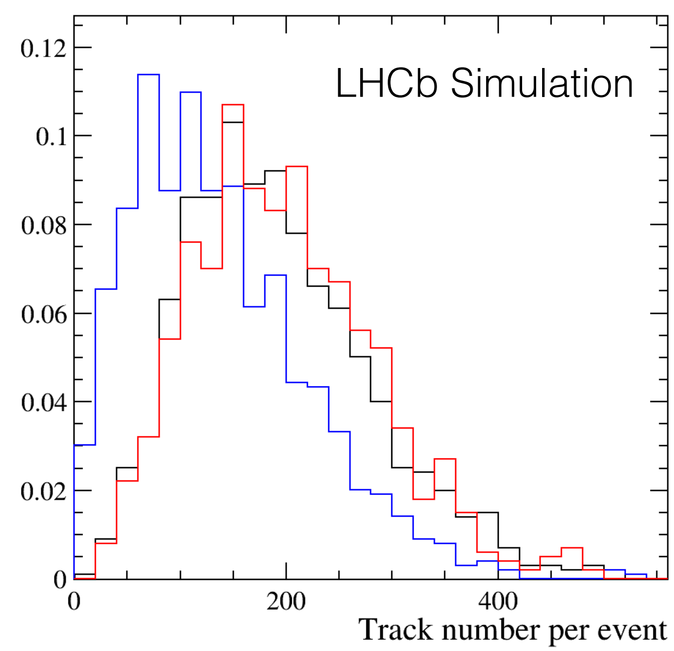
LHCb-Upgrade SciFi Simulation

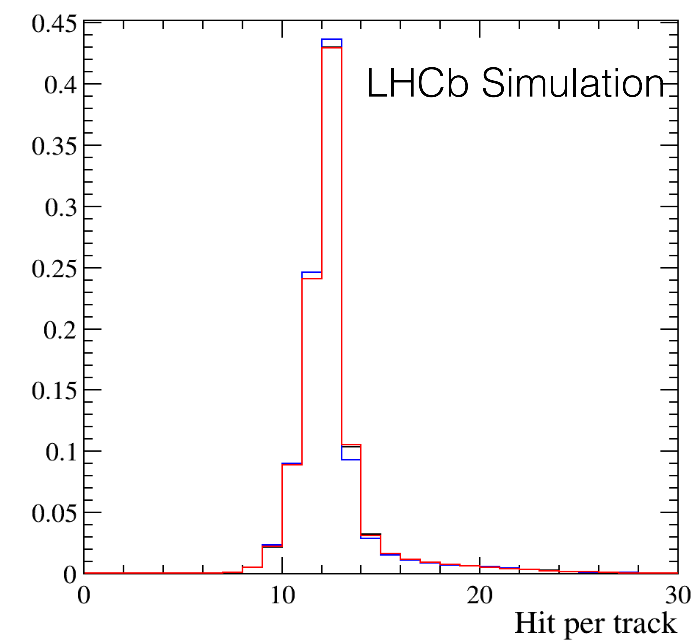
- Official LHCb-Upgrade simulation:
  - pp collisions with PYTHIA;
  - Particle interaction with detect layers usign GEANT4 toolkit;
  - Accurate simulation of hits digitization (energy deposits, spillover, noise, inefficiencies, light attenuation in the fibre, clustering, read-out electronic chain, etc)

- Simulated LHCb-Upgrade conditions:
  - sqrt(s)=14 TeV;
  - LHC bunch spacing = 25 ns;
  - L=2 x 10^{33} $cm^{-2}s^{-1}$;
  - ν = 7.6

Sample	Nb. Events	Decay Mode
Sample 1	1,000	Minimum-Bias
Sample 2	1,000	$D^{*+} \to D^{*+} \to D^0\pi^+ \to [K^0_S\pi^+\pi^-]\pi^+$
Sample 3	1,000	$B^0_s \to \phi\phi \to [K^+K^-][K^+K^-]$







Occupancy

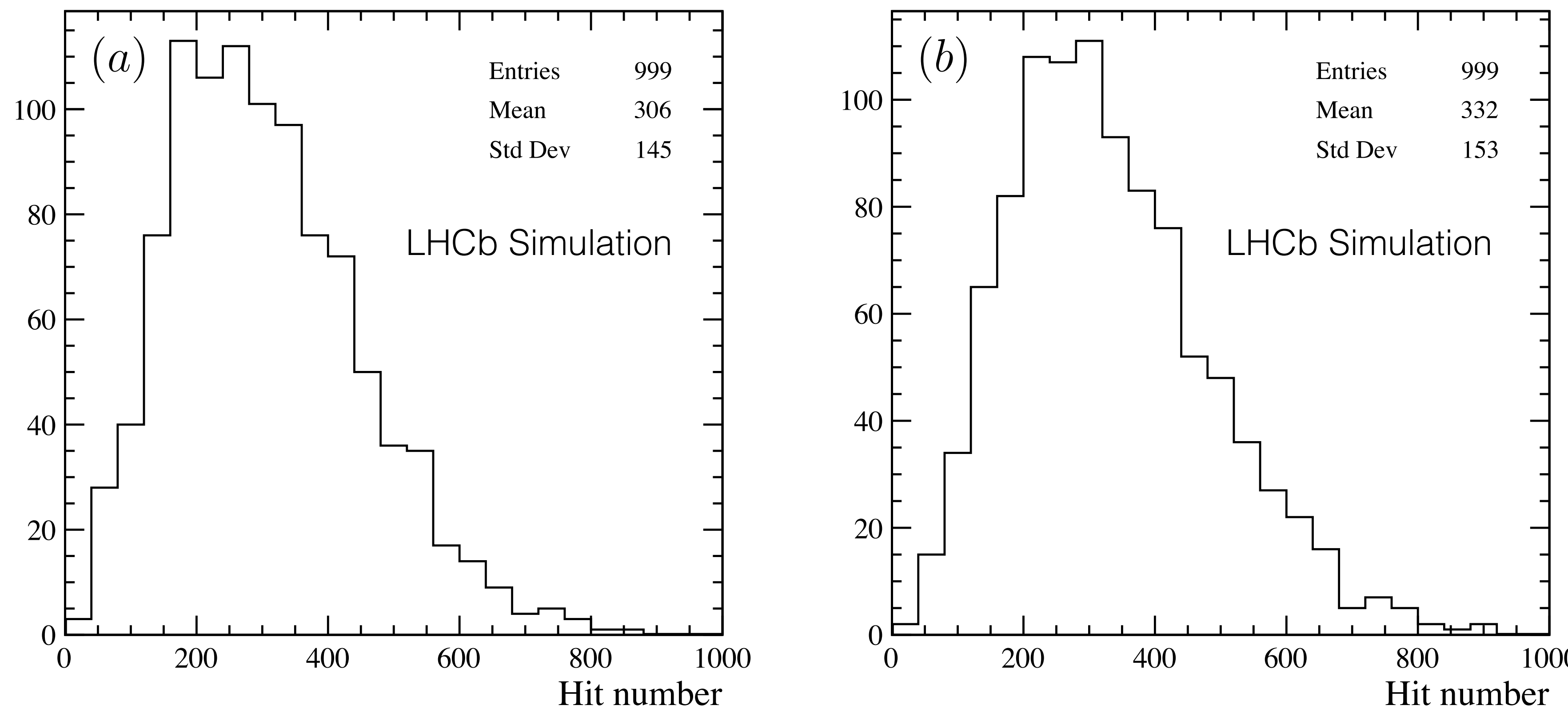


Figure 4.1: Distribution of the number of hits for the first axial layer (n. $0x$) and the last stereo layer (n. $10v$) of the SciFi subdetector.

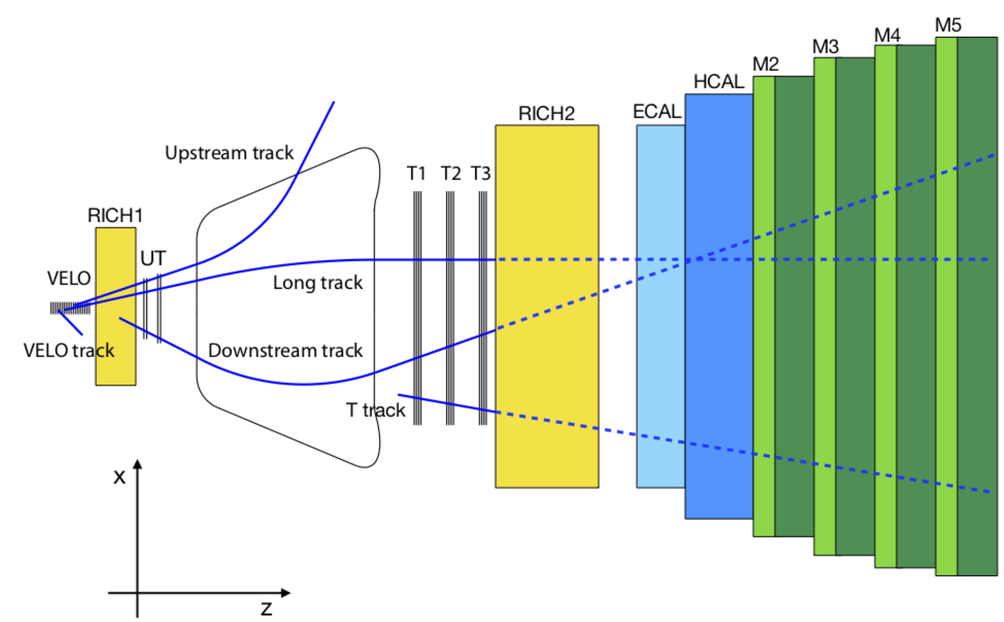
Tracking indicators

Efficiency: $$\varepsilon_{\text{Tracking}} = \frac{\text{reconstructed \& matched}}{\text{reconstructible}}$$

Ghost rate: $$ghost\ rate = \frac{\text{reconstructed not matched}}{\text{reconstructed}}$$

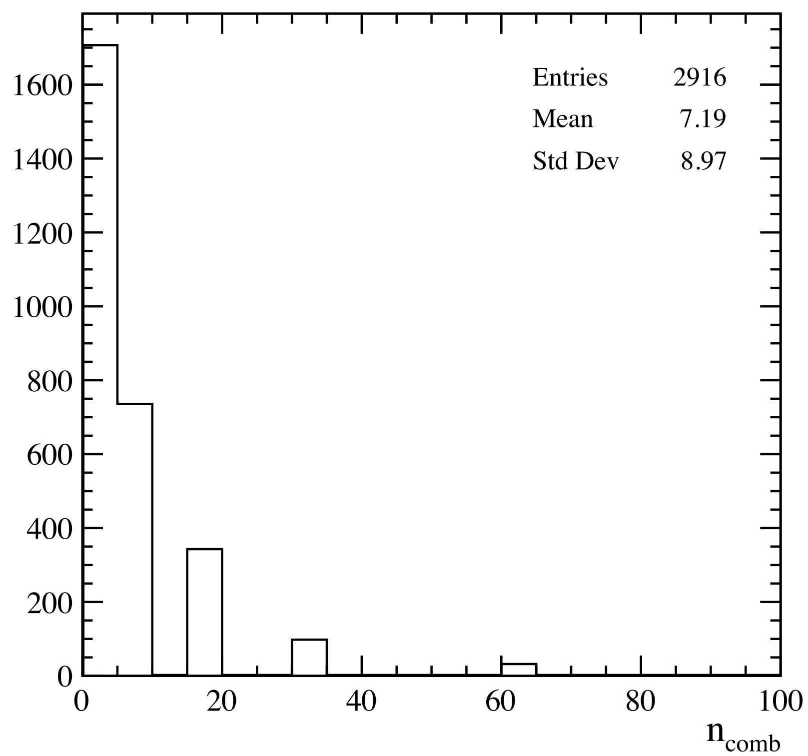
A reconstructed track is said to be "truth-matched" to a simulated track if both the axial and stereo maxima are contained in a 3x3 square centered at the true track parameters position.

Name	Property
hasT	reconstructible in SciFi
down (or downstream)	reconstructible in SciFi and UT
long	reconstructible in VELO and SciFi
noVelo	not reconstructible in the VELO
strange	daughter of a strange particle (K^0 , Λ, ..)
from B	belongs to the decay chain of a b hadron
from D	belongs to the decay chain of a c hadron

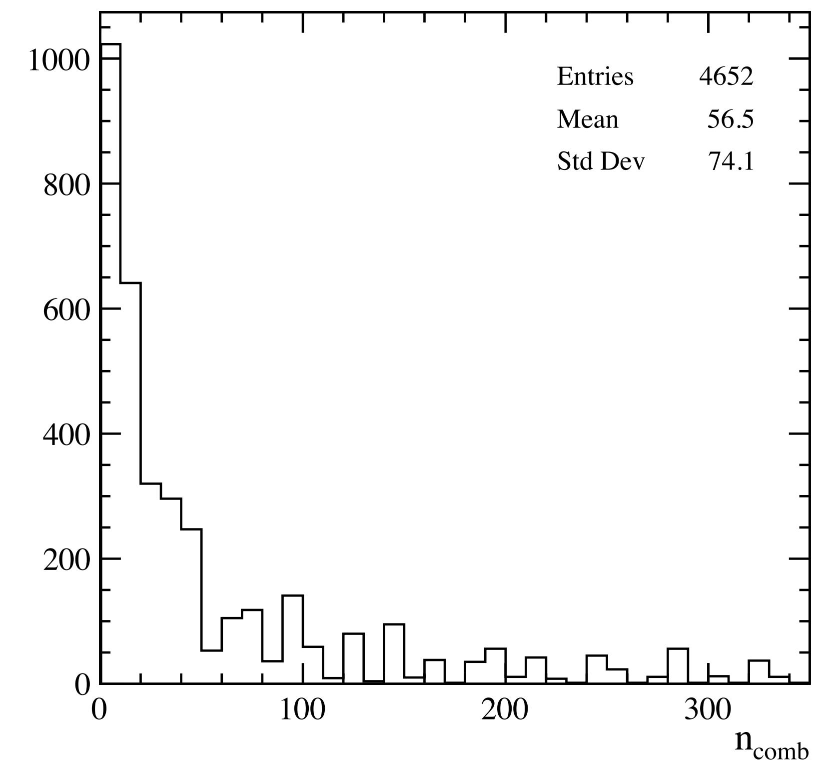


Number of combinations

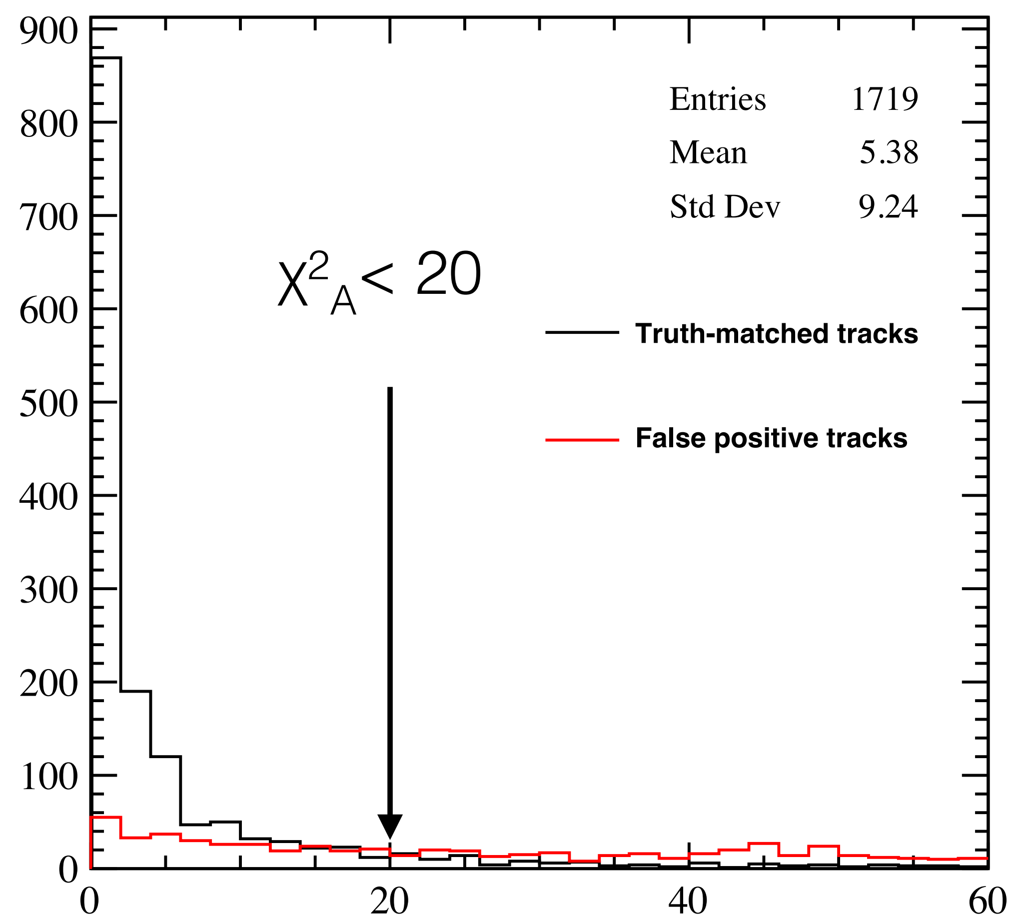
Axial



Stereo

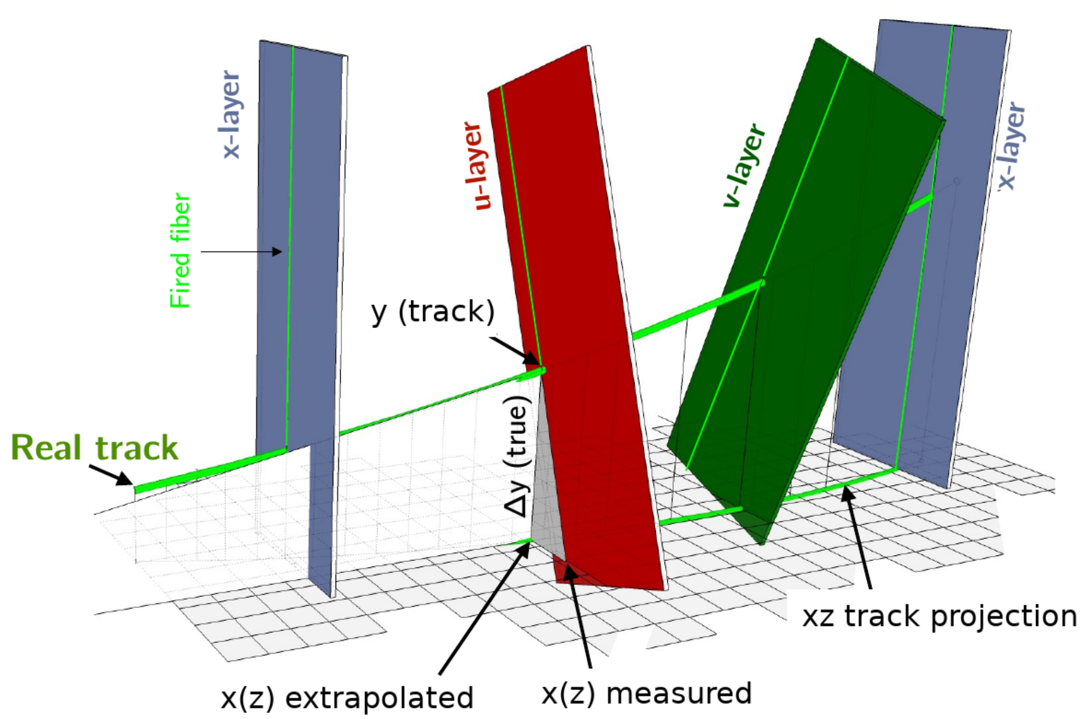


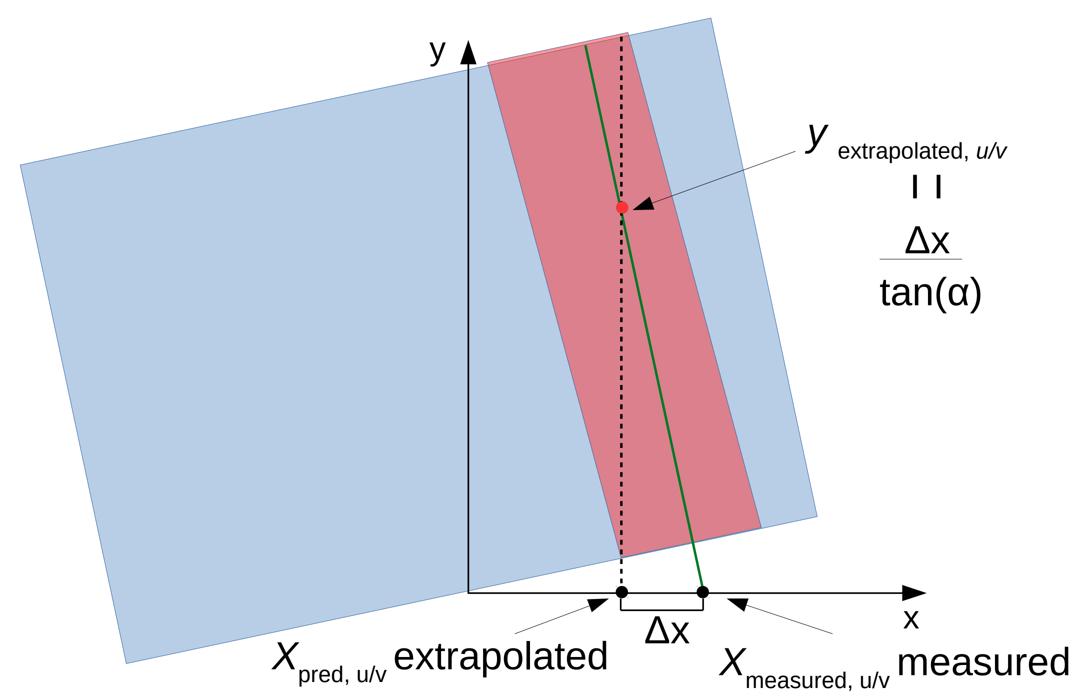
Remove axial false positive



axial track χ^2_A

Stereo Retina





Stereo retina

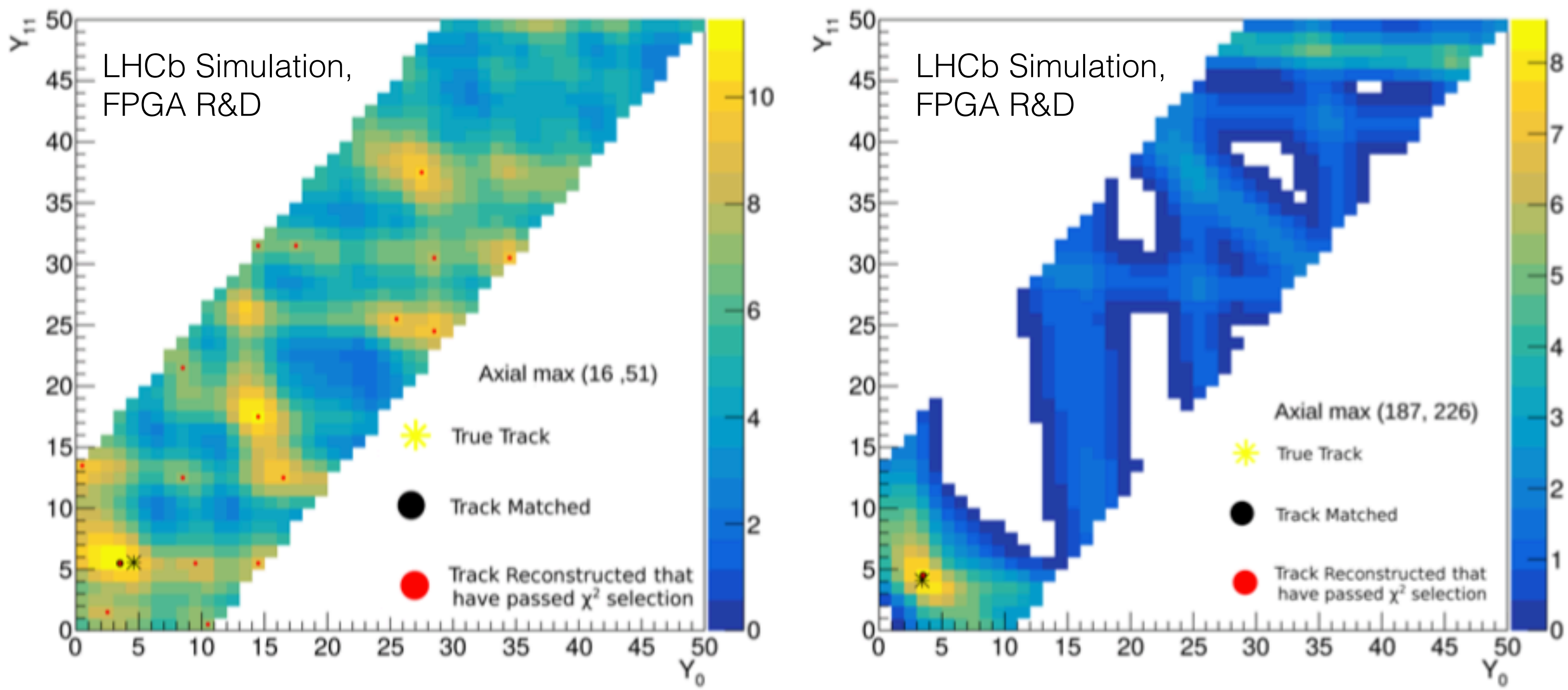


Figure 6.4: Excitation level of two stereo retinas filled with SciFi subdetector hits from fully simulated LHCb Upgrade events. True tracks (yellow stars), reconstructed track candidates (red dots), and truth-matched reconstructed track candidates (black dots) are superimposed.