Real-time track reconstruction with FPGAs in LHCb Scintillating Fibre Tracker beyond Run 3

Ao Xu et al.
on behalf of the LHCb Collaboration

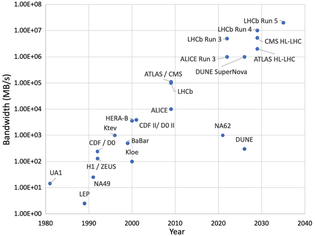
22nd International Workshop on
Advanced Computing and Analysis Techniques in Physics Research

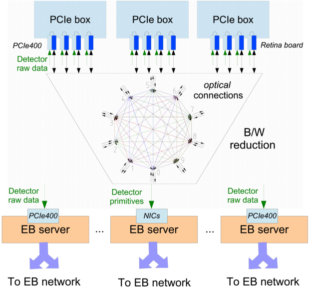
March 14, 2024

Motivation

- LHCb: largest throughput in HEP
- High memory-access tasks more efficiently performed in dedicated devices
  - Envision a system where pattern recognition is performed within readout
  - Kick off High Level Trigger with **pre-reconstructed primitives** (array of aligned hits)
  - Free HLT for higher-level tasks
- **FPGA** most suitable technology
  - Sufficient programmable logic
  - High bandwidth
  - Can be more energy-efficient

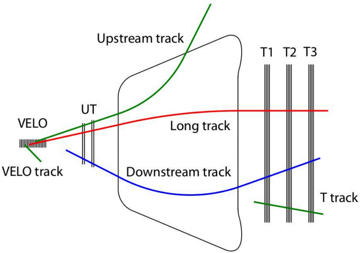
[EPJ Plus 138 1005 (2023)]

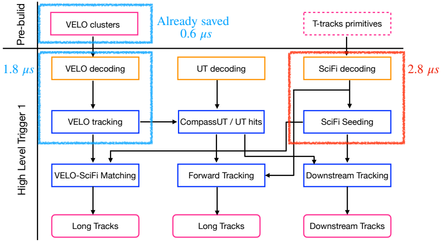




LHCb Upgrade tracking

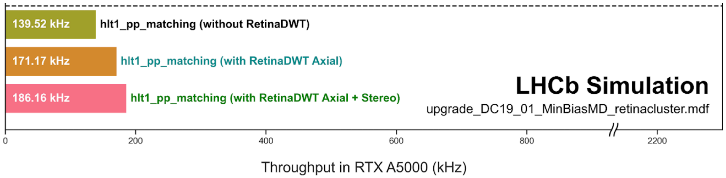
- Current HLT1 reconstructions focus on Long tracks
  - Run 2 based on Forward tracking
  - Run 3 benefits also from Matching
- Add Downstream tracks to HLT1
  - Expand the LHCb physics program
  - See previous talk [V. Svintozelskyi]
- **DoWnstream Tracker** will provide HLT1 with pre-formed T-track primitives in FPGA
  - Make room for Downstream tracking and other desirable enhancements





Throughput gain with DoWnstream Tracker

- HLT1 sequence `hlt1_pp_matching`

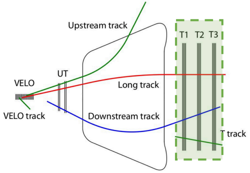


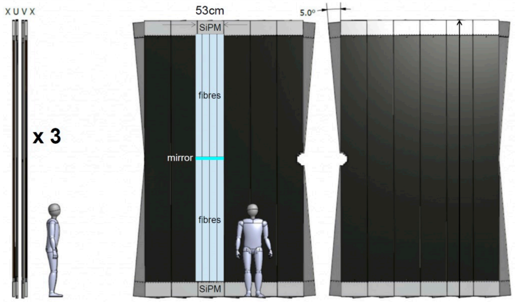
- Default sequence
  - Total (T-track reconstruction): $7.2\,\mu s$ ($1.5\,\mu s$)
- With T-track primitives from DWT
  - Total (Primitives decoding and refitting): $5.4\,\mu s$ ($0.06\,\mu s$)
- Throughput increased by a factor of **1.33**

LHCb Scintillating Fibre Tracker

[CERN-LHCC-2014-001]

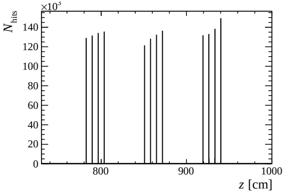
- **Three** tracking stations: T1, T2, T3
- Each consists of **four** detection planes: oriented $(0°, +5°, -5°, 0°)$
  - Modules have 2.5 m long scintillating fibres with a diameter of $250\,\mu\text{m}$ read out by SiPMs
  - Measurements of the co-ordinates (x, u, v, x)



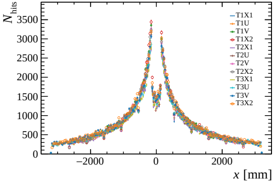


Simulation infrastructure for physics performance

- LHCb Upgrade simulation
  - Default Run 3 and 4 condition: $E = 7\,\mathrm{TeV}$, bunch $25\,\mathrm{ns}$, $\nu = 7.6$
  - Samples: Minimum Bias, $D^0 \to K^0_\mathrm{S}\pi^+\pi^-$, $B^0_s \to \phi\phi$



Hit distribution z



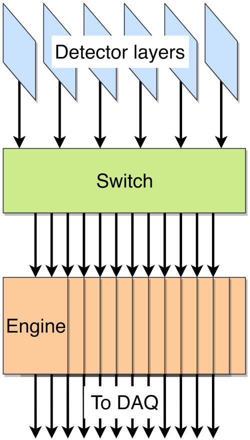
Hit distribution x

- DoWnstream Tracker emulator
  - C++ software emulator of an FPGA-based system for reconstruction of T-track primitives
  - Use integers to emulate the firmware implementation at **bit-level**

Artificial retina architecture

[NIM A453 (2000) 425-429]

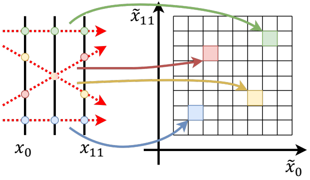
- Architecture for real-time reconstruction by **extreme parallelism** and **high connectivity**
  - Computation similar to Hough transform
- Data flow
  - Input from detector and data preparation
  - Distribution network
    - Switch: routes hits only to appropriate cells using lookup tables
    - Optical communication: exchanges hits between boards
  - Cell engine and max-finder
  - Primitive tracks are forwarded to the Event Builder

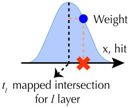


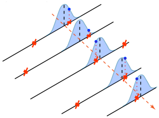
Reconstruction of axial T-track primitives

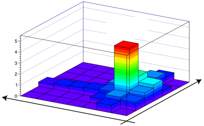
[NIM A453 (2000) 425-429]

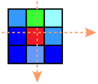
- Emulate in detail the same steps of the hardware system
  1. Axial (x-z plane) track parametrisation
     - $(\tilde{x}_0, \tilde{x}_{11})$: x-coordinates at the **first** and **last** SciFi layer
     - $\#$ of pattern cells for SciFi: **2×73k**
  2. Weight accumulation
     - $w = \sum_{hits} \exp\left(-\frac{(x_l - t_l)^2}{2\sigma}\right)$ for $|x_l - t_l| < d_s$
  3. Identification of local maxima (axial track primitives)
     - **Maximum** above **threshold** in the centered 3×3 cluster









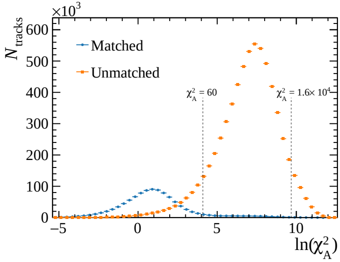


Ghost removal with axial track fit

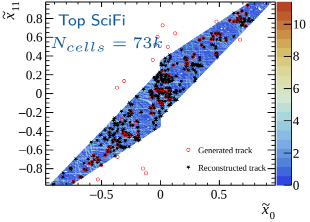
- Linearised χ^2 fit for false maxima removal
- Parabolic model with cubic correction [1, 2]

$$x(z) = a_x + b_x \times z + c_x \times z^2 \times (1 + \mathrm{dRatio} \times z)$$

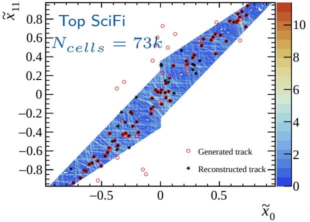
- For each local maximum determine the best fit over combinations of
  - 5 different axial layers out of 6
  - 1 out of 2 candidate hits on each layer



Before χ^2_A requirement



After $\chi^2_A < 60$



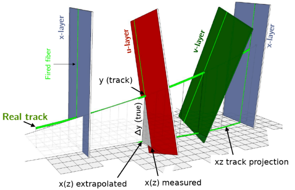
Reconstruction of 3D T-track primitives

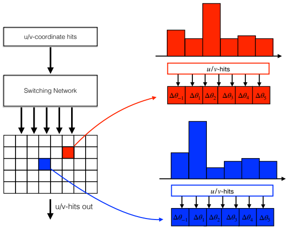
[CERN-LHCC-2024-001, pending]

- Emulate in detail the same steps of the hardware system
  1. Stereo (y-z plane) track parametrisation
     - $\tilde{y}$: y-coordinate at the middle of SciFi
     - # of bins per axial track: 45
  2. u/v hits distribution
     - Good axial track candidate $\longleftrightarrow$ Binned parametric space

$$x_{\text{pred},u/v} \xrightarrow{x_{\text{pred},u/v} - y \times \tan\alpha} x_{\text{meas},u/v}$$

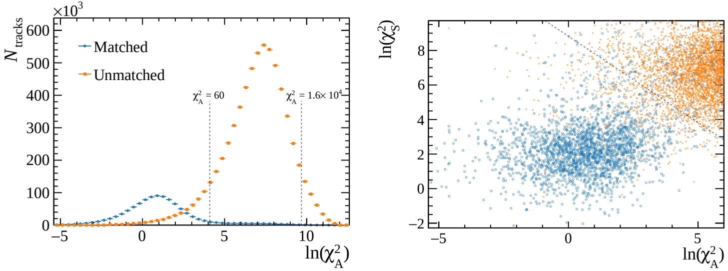
  3. Identification of local maxima (stereo track primitives)
     - **Maximum** above **threshold** in 1D histrogram





Ghost removal with stereo track fit

- Linearised χ^2 fit for false maxima removal
- Straight line: $y(z) = a_y + b_y \times z$
- For each local maximum determine best fit over combinations of
  - 5 different stereo layers out of 6
  - 1 out of all candidate hits on each layer
- 3D track primitives filtered with (χ^2_A, χ^2_S) requirement
  - Linear cut for illustration of performance



Physics tracking performance of T-track primitives

- Fiducial requirements: $p_{\mathrm{T}} > 200\,\mathrm{MeV}$ and $2 < \eta < 5$
- Efficiencies **comparable** with GPU-HLT1 and CPU-HLT2 Seeding
  - Higher efficiencies could be achieved with looser χ^2 requirements
- Ghost rate is **under control**
  - As a reference: below 15% (6%) for GPU-HLT1 tracking

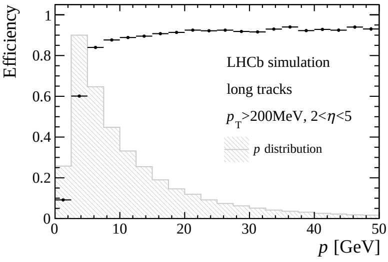
Track type	MinBias	$D^0 \to K^0_{\mathrm{S}}\pi^+\pi^-$	$B^0_s \to \phi\phi$
Long, $p > 3\,\mathrm{GeV}/c$	85 (86)	83 (84)	84 (85)
Long, $p > 5\,\mathrm{GeV}/c$	**90 (91)**	**89 (90)**	**89 (89)**
Long from B not $e^\pm$, $p > 3\,\mathrm{GeV}/c$	-	-	88 (87)
Long from B not $e^\pm$, $p > 5\,\mathrm{GeV}/c$	-	-	90 (90)
Down, $p > 3\,\mathrm{GeV}/c$	84 (85)	83 (84)	83 (84)
Down, $p > 5\,\mathrm{GeV}/c$	**89 (91)**	**88 (89)**	**88 (89)**
Down from strange not $e^\pm$, $p > 3\,\mathrm{GeV}/c$	-	83 (83)	-
Down from strange not $e^\pm$, $p > 5\,\mathrm{GeV}/c$	-	88 (88)	-
Down from strange not long not $e^\pm$, $p > 3\,\mathrm{GeV}/c$	-	83 (83)	-
Down from strange not long not $e^\pm$, $p > 5\,\mathrm{GeV}/c$	-	88 (89)	-
ghost rate	**16 (10)**	**17 (12)**	**17 (13)**
ghost per real track	0.2 (0.1)	0.2 (0.1)	0.2 (0.1)

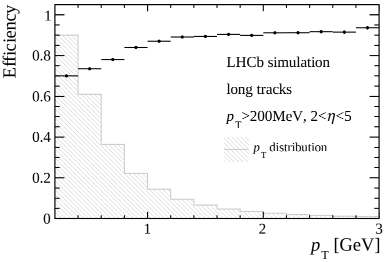
Event-averaged values shown in brakets

Performance as a function of kinematics

[CERN-LHCC-2024-001, pending]

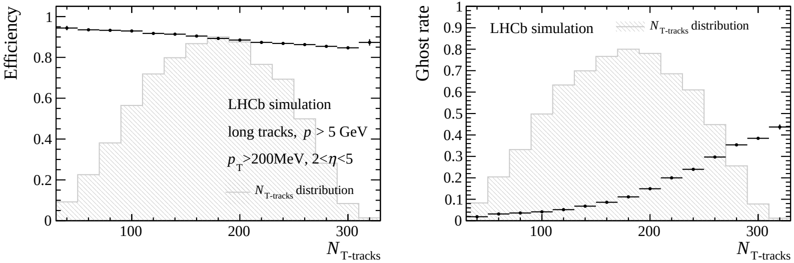
- Efficiency above 90% for high-momentum tracks
- Good efficiency for low-momentum ($p < 5\,\mathrm{GeV}$) tracks
  - Essential for downstream tracks (K^0_S and Λ)





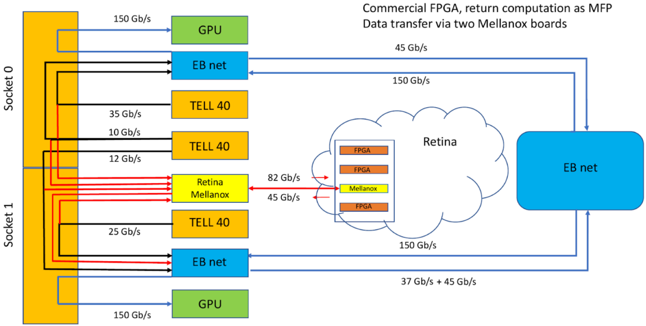
[CERN-LHCC-2024-001, pending]

- Robust scaling with occupancy



Resources and integration in LHCb Run 4 DAQ

- Number of FPGAs: 64 (axial) + 32 (stereo)
- DWT Boxes (up to 6 FPGA each) connected to SciFi EB nodes
- Modular, scalable, and minimal disturbance to current DAQ



Summary and outlook

- **DoWnstream Tracker** is an FPGA-based tracking system running at 30 MHz to reconstruct T-track primitives at pre-build stage
  - Initial study [ACAT2019] and preliminary result [CTD2023]
  - Aim to accelerate Downstream tracking by providing T-track primitives of good quality
  - Good physics performance of T-track primitives can be achieved
- Hardware **demonstrator installed and tested** with live data
  - Reconstruct a quadrant of the VELO detector in real-time
- TDR submitted to LHCC as part of LHCb DAQ Enhancement
  1. Develop the technique for Run 5 and beyond
  2. Bring physics enhancement to Run 4
- Optimisation ongoing with **integration of LHCb trigger system**
  - FPGA-based DWT + GPU-based HLT1 + CPU-based HLT2

Backup

LHCb RETINA team

Wander Baldini[1,2], Giovanni Bassi[3,4], Andrea Contu[5], Riccardo Fantechi[3], Jibo He[6,7], Brij Kishor Jashal[8], Sofia Kotriakhova[1,9], Federico Lazzari[3,10], Maurizio Martinelli[11,12], Diego Mendoza[8], Michael J. Morello[3,4], Arantza De Oyanguren Campos[8], Lorenzo Pica[3,4], Giovanni Punzi[3,10], Qi Shi[6], Francesco Terzuoli[3,13], Giulia Tuci[14], Ao Xu[3], Jiahui Zhuo[8]

[1] *INFN Sezione di Ferrara, Ferrara, Italy*
[2] *European Organization for Nuclear Research (CERN), Geneva, Switzerland*
[3] *INFN Sezione di Pisa, Pisa, Italy*
[4] *Scuola Normale Superiore, Pisa, Italy*
[5] *INFN Sezione di Cagliari, Monserrato, Italy*
[6] *University of Chinese Academy of Sciences, Beijing, China*
[7] *Hangzhou Institute for Advanced Study, UCAS, Hangzhou, China*
[8] *Instituto de Fisica Corpuscular, Centro Mixto Universidad de Valencia - CSIC, Valencia, Spain*
[9] *Università di Ferrara, Ferrara, Italy*
[10] *Università di Pisa, Pisa, Italy*
[11] *INFN Sezione di Milano-Bicocca, Milano, Italy*
[12] *Università di Milano Bicocca, Milano, Italy*
[13] *Università di Siena, Siena, Italy*

[14] *Physikalisches Institut, Ruprecht-Karls-Universitat Heidelberg, Heidelberg, Germany*

Physics performance of axial T-track primitives

- Working point set for $\varepsilon = 90\%$ **of long tracks with** $p > 5\,\mathrm{GeV}$
  - Number of pattern cells for SciFi: 2×73k
  - Efficiencies comparable with CPU-HLT2 Hybrid Seeding and GPU-HLT1 Seeding
  - Ghost rate about 35% (25%) $\Longrightarrow$ 0.5 (0.4) fake track for each real track
    - For reference 22% of (axial-only) GPU-HLT1

Track type	$\varepsilon(\mathrm{MinBias})$	$\varepsilon(D^0 \to K^0_S\pi^+\pi^-)$	$\varepsilon(B^0_s \to \phi\phi)$ [%]
T-track, $p > 3\,\mathrm{GeV}$	83 (85)	82 (83)	83 (84)
T-track, $p > 5\,\mathrm{GeV}$	90 (91)	89 (90)	88 (89)
Long, $p > 3\,\mathrm{GeV}$	86 (87)	84 (85)	85 (86)
Long, $p > 5\,\mathrm{GeV}$	**91 (92)**	**90 (91)**	**89 (90)**
Long from B not $e^\pm$, $p > 3\,\mathrm{GeV}$	-	-	89 (88)
Long from B not $e^\pm$, $p > 5\,\mathrm{GeV}$	-	-	92 (91)
Down, $p > 3\,\mathrm{GeV}$	85 (86)	83 (84)	84 (85)
Down, $p > 5\,\mathrm{GeV}$	90 (91)	89 (90)	89 (90)
Down from strange not $e^\pm$, $p > 3\,\mathrm{GeV}$	-	83 (83)	-
Down from strange not $e^\pm$, $p > 5\,\mathrm{GeV}$	-	89 (89)	-
ghost rate [%]	32 (22)	35 (28)	35 (27)
ghost per real track	0.5 (0.3)	0.5 (0.4)	0.5 (0.4)

Event-averaged values are shown in parenthesis

Definition of efficiency and ghost rate

- Event-integrated quantity

$$\varepsilon \equiv \frac{\sum_i n^i_{\texttt{tracks,matched}}}{\sum_i n^i_{\texttt{tracks,reconstructible}}}$$

$$\texttt{ghost rate} \equiv \frac{\sum_i n^i_{\texttt{tracks,unmatched}}}{\sum_i n^i_{\texttt{tracks,reconstructed}}}$$

$$= \sum_i \frac{n^i_{\texttt{tracks,reconstructed}}}{\sum_i n^i_{\texttt{tracks,reconstructed}}} \times \frac{n^i_{\texttt{tracks,unmatched}}}{n^i_{\texttt{tracks,reconstructed}}}$$

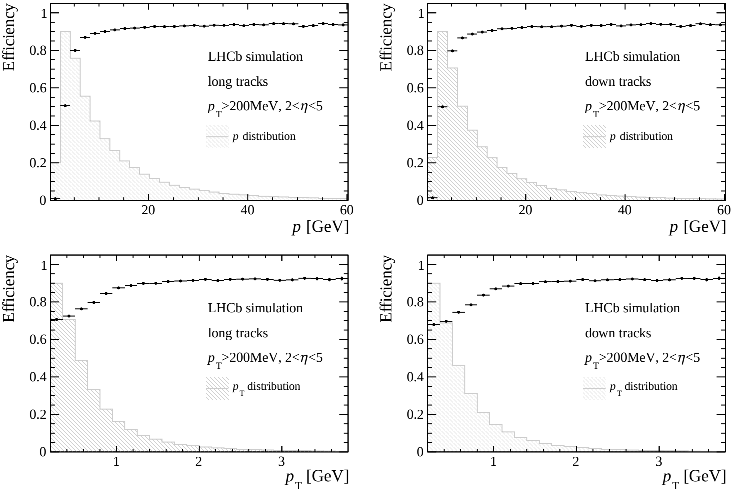
- Event-averaged quantity

$$\varepsilon \equiv \sum_i \frac{1}{N_{\texttt{evt}}} \times \frac{n^i_{\texttt{tracks,matched}}}{n^i_{\texttt{tracks,reconstructible}}}$$

$$\texttt{ghost rate} \equiv \sum_i \frac{1}{N_{\texttt{evt}}} \times \frac{n^i_{\texttt{tracks,unmatched}}}{n^i_{\texttt{tracks,reconstructed}}}$$

Physics performance (axial): effciency VS momentum

- Working point set for $\varepsilon = 90\%$ of long tracks with $p > 5\,\mathrm{GeV}$



Physics performance (axial): efficiency VS η and ϕ

- Working point set for $\varepsilon = 90\%$ of long tracks with $p > 5\,\mathrm{GeV}$