Machine learning at LHCb (Run3):
from event reconstruction to physical analysis

Zan Ren (任赞)

University of Chinese Academy of Sciences

Dec 16th, 2024

The 29th LHC Mini-workshop

Fuzhou, China

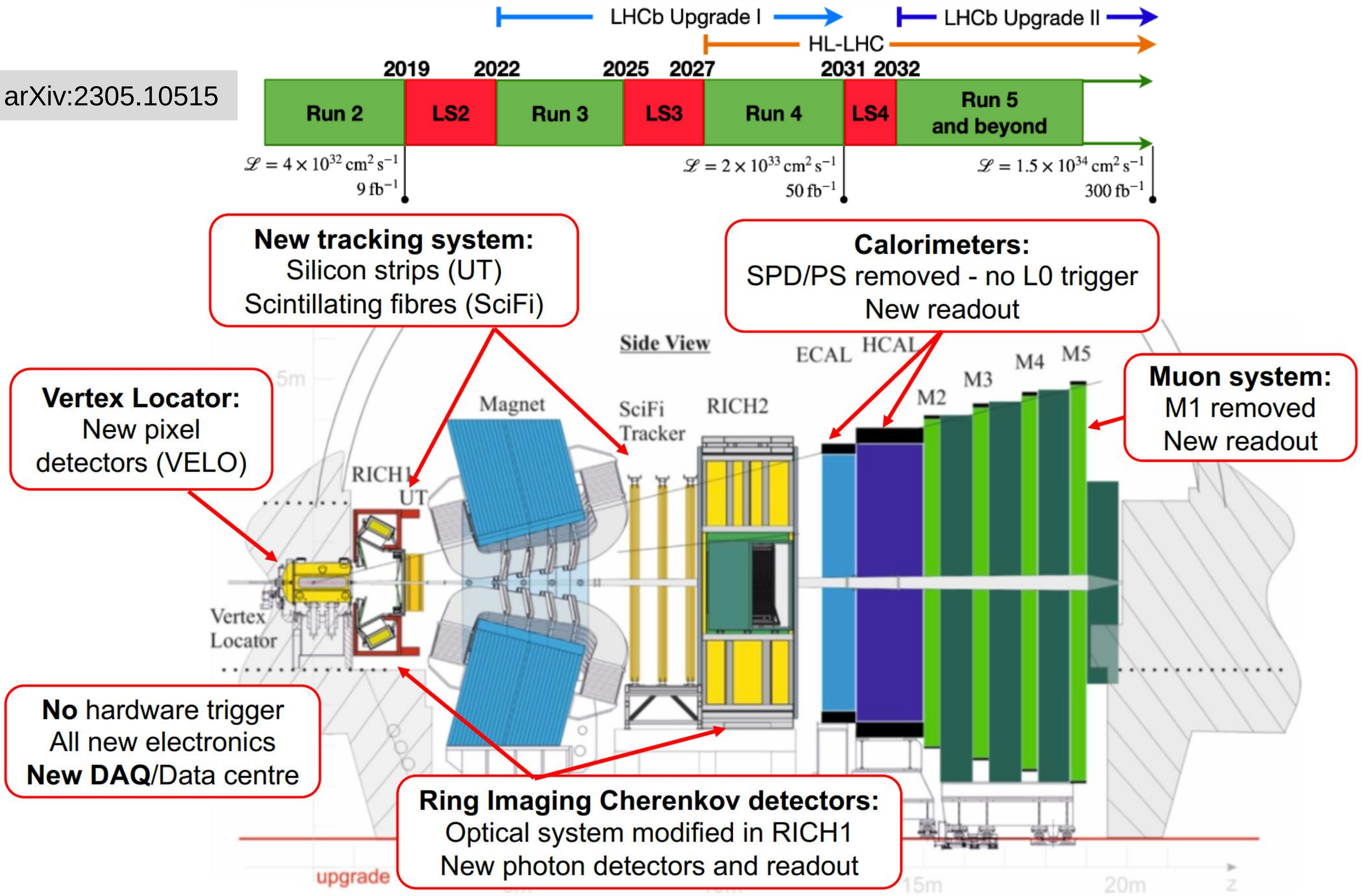
Outline

- Introduction to LHCb
  - Overview of the detector
  - Challenges in Run3 era and beyond
- ML developments at LHCb
  - ML in event reconstruction
  - ML in simulation
  - ML in physical analysis (too trivial, not included in this presentation)
  - ML related organizations at LHCb
- Summary

LHCb detector (Run3)

- Single-arm, forward. Specifically designed for heavy-flavour physics.
- The ML implementations of Run1&2 will be briefly reviewed in this talk, with a focus on Run3.

arXiv:2305.10515



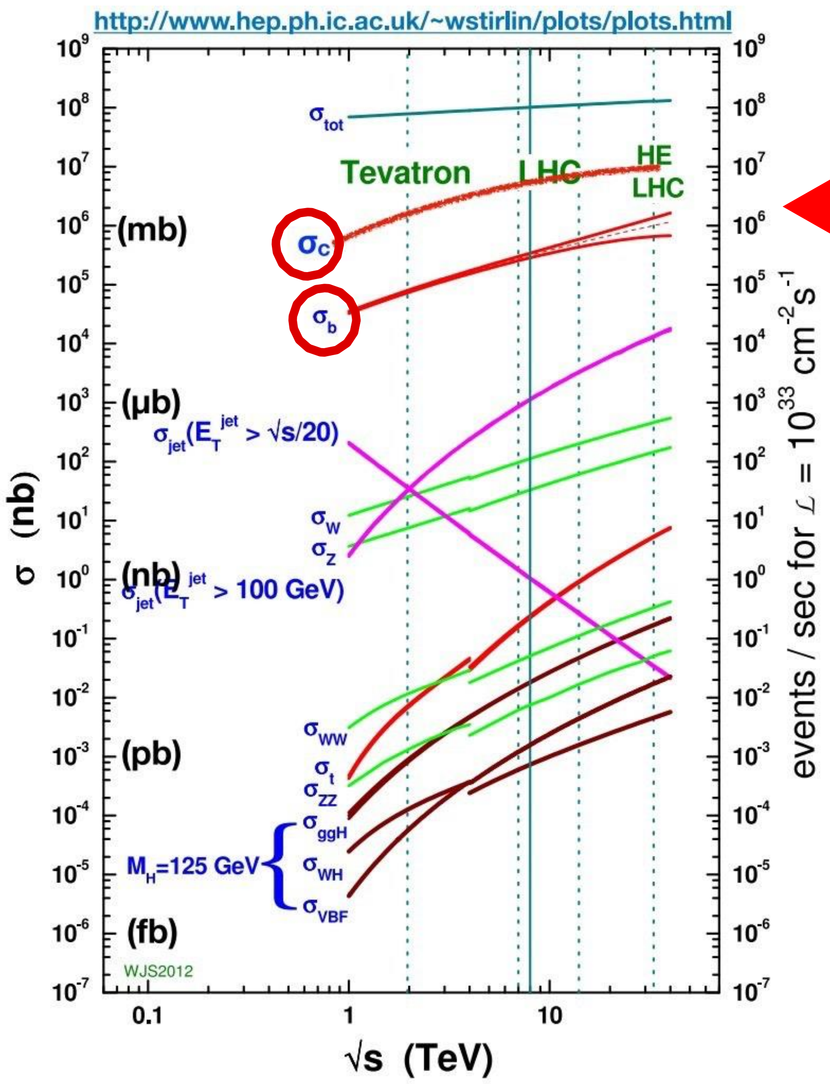
Almost a new detector!

- A factor of 5 luminosity increase.
- $L = 2 \times 10^{33}$ cm^{2}s^{-1}
- Expect 23 fb^{-1} by 2025 (Run 3)
- Expect 50 fb^{-1} by 2031 (Run 4)
- Pile-up ~6 interactions.

More data, more challenges!

- Storage (space)
- Bandwidth (speed)
- Algorithm (data quality, UE, …)

Challenges from the MHz era



Run 3: Luminosity of 2×10^{33} $cm^{-2}s^{-1}$, $\sqrt{s}$ = 14 TeV

LHCb Run3 is here! (@MHz level).

Bandwidth [MB/s] ~ Trigger output rate [kHz] x average event size [kB]

- Read out the full detector (**hardware trigger removed**)
  - No “simple” local selection criteria
- Selective persistency events as output to storage
  - Up to 100 kB event size, can only transfer 10 GB/s to long-term storage

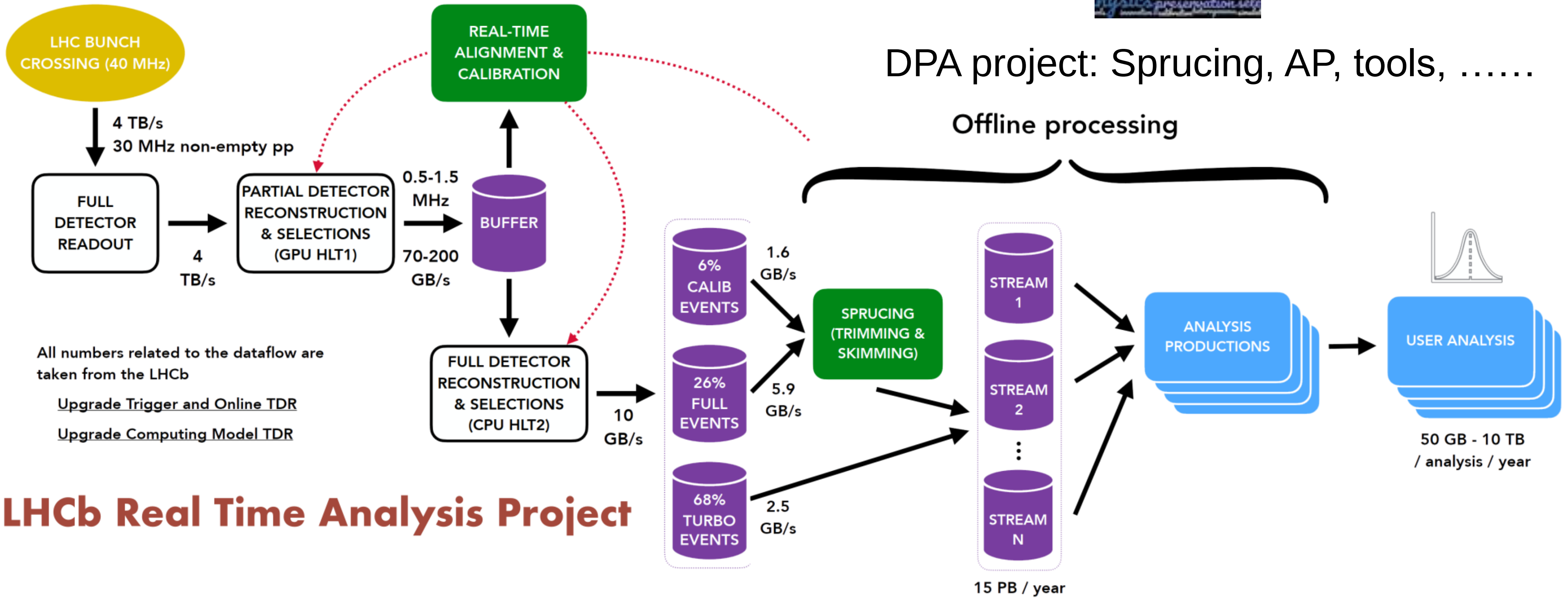
Novel opportunities to improve and expand the physics program.

Excellent setup for ML solutions!

LHCb data flow in Run3

- LHCb Run 3 data flow

DPA project: Sprucing, AP, tools, ……



LHCb Real Time Analysis Project

ML can implement almost everywhere!

ML developments at LHCb

ML in event reconstruction

Topological triggers based on Lipschitz network

- **Topological triggers** in HLT2, aimed at identifying *b*-hadron vertices. arXiv:1510.00572
- Run1&2: bonsai boosted decision tree (BBDT), converted from MatrixNet(MN)
- Run3: Improved by introducing Monotonic **Lipschitz network** arXiv:2112.00038
- **Impose desired constraints** in the behavior of the network **by construction**

- **Robustness** against detector instabilities and simulation inaccuracies.
  - Technically done via weight-normalisation scheme during training.

Suppose W^m is the weight matrix of m-th layer with activation σ

$$g(\boldsymbol{x}) = W^m \sigma(W^{m-1}\sigma(\ldots\sigma(W^1\boldsymbol{x} + b^1)\ldots) + b^{m-1}) + b^m,$$

$g(x)$ satisfies Lipschitz condition if $\prod_{i=0}^{m} \|W^i\|_1 \leq \lambda$

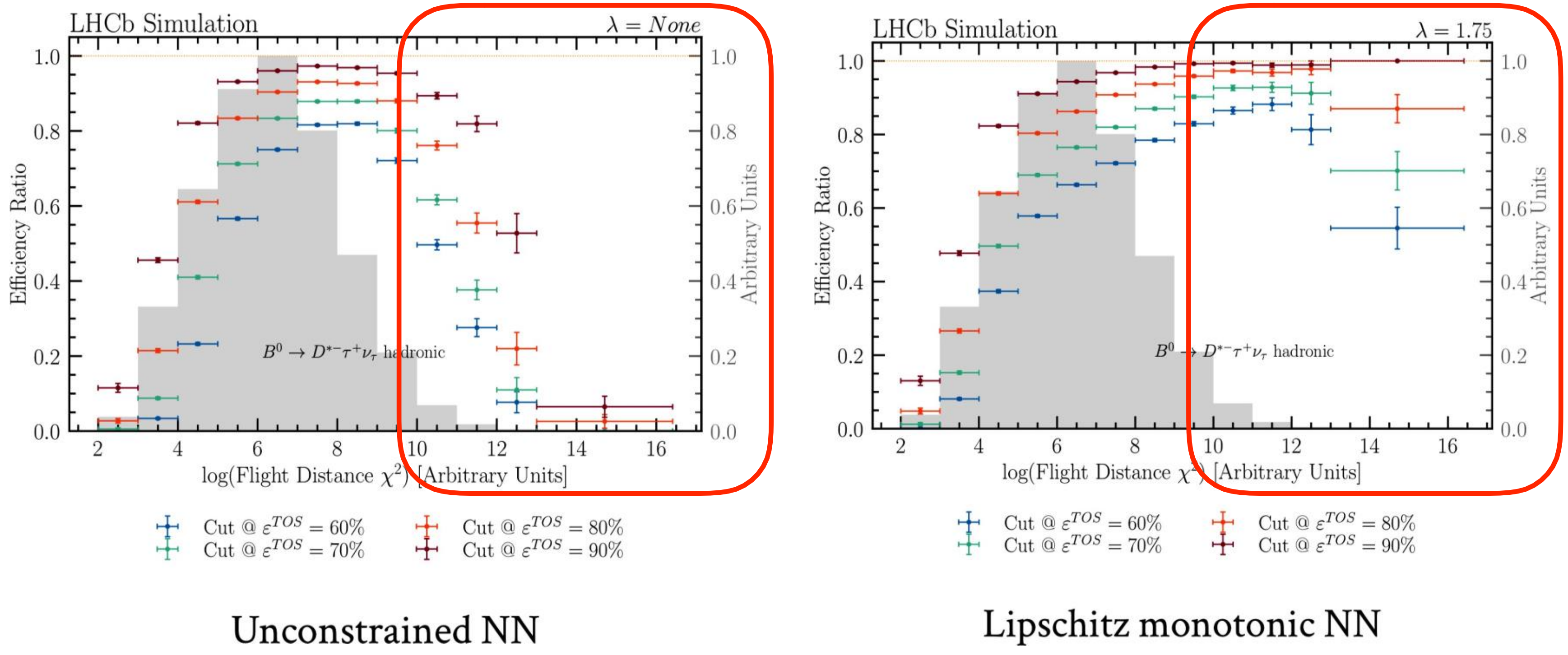
- **Monotonicity** in certain features for out-of-distribution guarantees.
  - Technically done by adding a residual connection to the network.
  - Monotonicity imposed in the IP χ^2 and the p_T

$$f(\boldsymbol{x}) = g(\boldsymbol{x}) + \lambda \sum_{i \in I} x_i, \qquad \frac{\partial f}{\partial x_i} = \frac{\partial g}{\partial x_i} + \lambda \geq 0 \quad \forall i \in I\,.$$

Topological triggers based on Lipschitz network

- Performance:

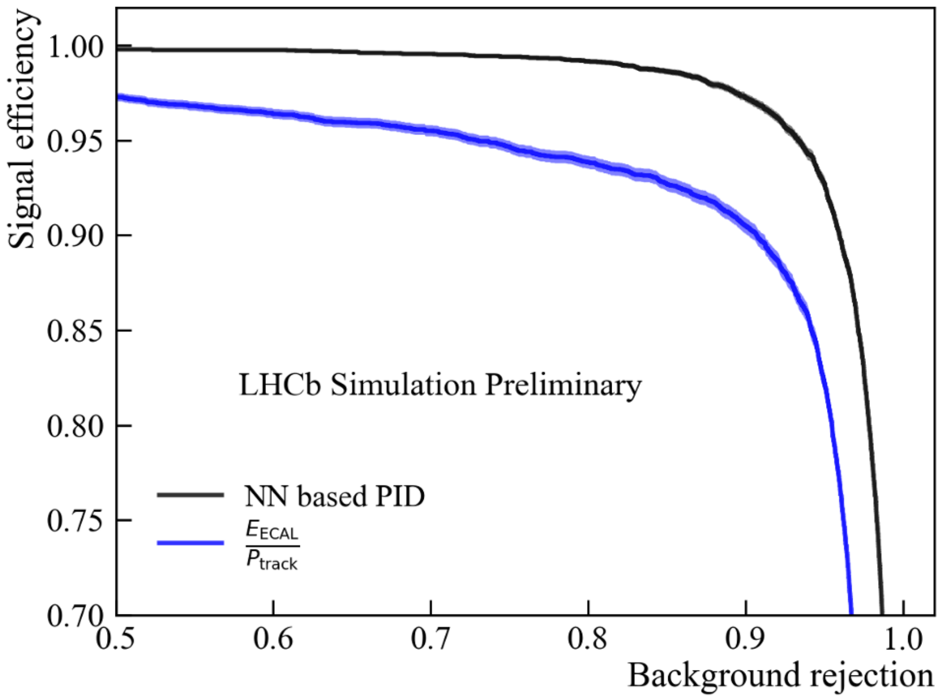
Enhanced sensitivity to long-lived candidates, particularly useful for searches of SL decays.



PID algorithm based on Lipschitz network

- The **ANN-based PID** has been used in LHCb since Run1 (not go in to detail here).
- This Lipshitz network is now also used for **electron ID** at the HLT1 level, implemented in Allen (a GPU-based HLT1 project).
  - Be able to remove 50% of the background without affecting the electron efficiency

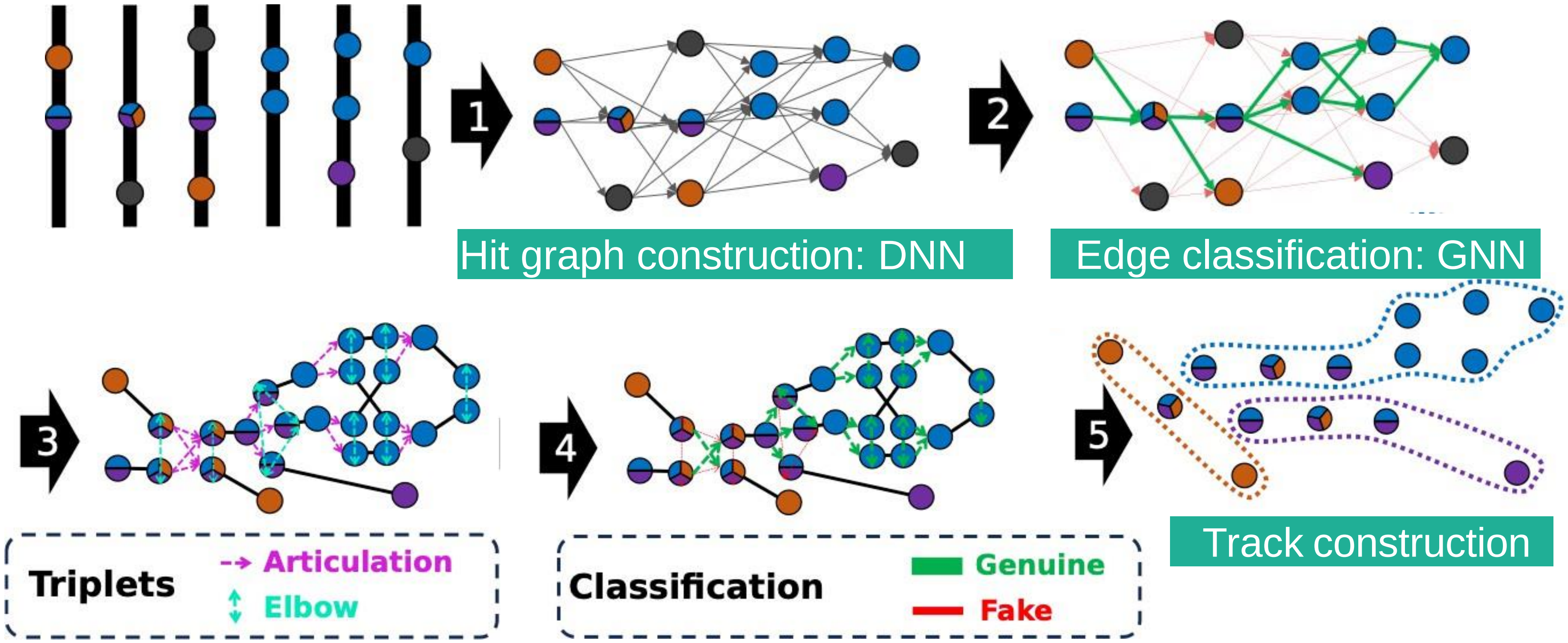
Large improvement with respect to the conventional (not ML based) algorithm.



LHCB-FIGURE-2024-003

ML-based tracking finding: The ETX4VELO project

- Based on the Exa.TrkX approach, to reconstruct forward tracks without a magnetic field, accounting for hit overlaps and inefficiencies. Eur. Phys. J. C 81, 876 (2021)
  - ETX4VELO introduces new triplet-related stages compared to the original one, to handle tracks with shared hits. (Combined usage of GNN and DNN)



Generate triplets (edge-edge connections)

Triplet classification: extension to the GNN (in step 2) with a DNN per triplet

arXiv:2406.12869

ML-based tracking finding: The ETX4VELO project

- Performance, compared to the default algorithm in LHCb:
  - Similar efficiency.
  - Improved reconstruction for electrons.
  - **Lower ghost (fake-track) rate.**

Long category	Efficiency	
	Allen	ETX4VELO
No electrons	99.26	99.28 (99.51)
Electrons	97.11	98.80 (99.22)
From strange	97.69	97.50 (98.06)

Velo-only category	Efficiency	
	Allen	ETX4VELO
No electrons	96.84	97.03 (97.86)
Electrons	67.81	85.10 (86.69)
From strange	93.53	93.07 (96.05)

	Allen	ETX4VELO	
		$d^2_{\max} = 0.010$	$d^2_{\max} = 0.020$
Ghost rate	2.18%	0.76%	0.81%

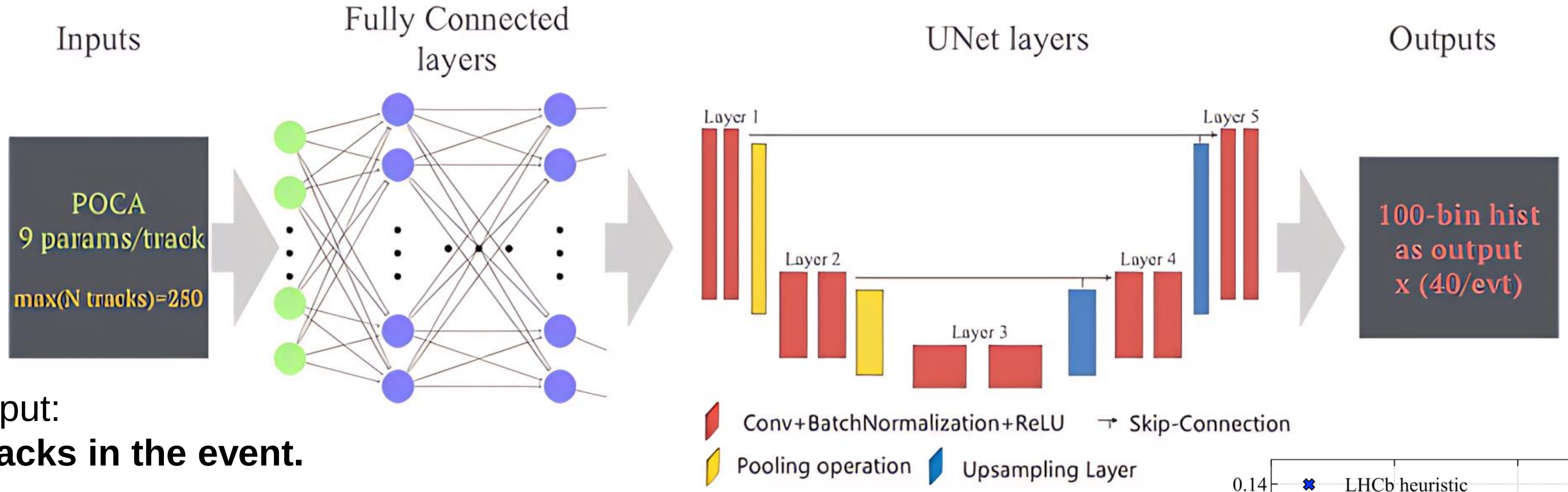
arXiv:2406.12869

Next: optimise the throughput for usage in HLT1.
Batching over events in the GPU recently achieved.

ML-based vertexing: candidate algorithms

- The hybrid approach:
  - Cooperate with ATLAS collaboration.
  - LHCb uses a hybrid model, composed of DNN and CNN.

Not yet implemented in the HLT framework.

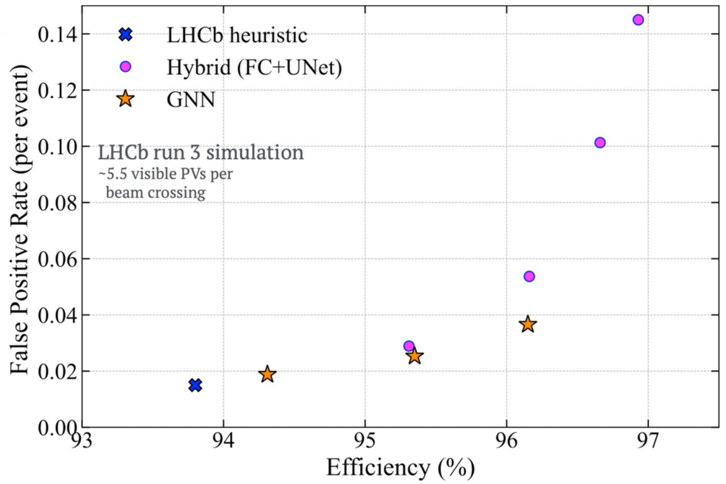


Target:
Gaussian distribution

heights and widths reflect the expected PV resolutions

Input:
tracks in the event.

- The GNN approach:
  - GNN model based on the ETX4VELO one
- GNN achieves **slightly better physics performance**
- can also provide **track-to-PV association**.



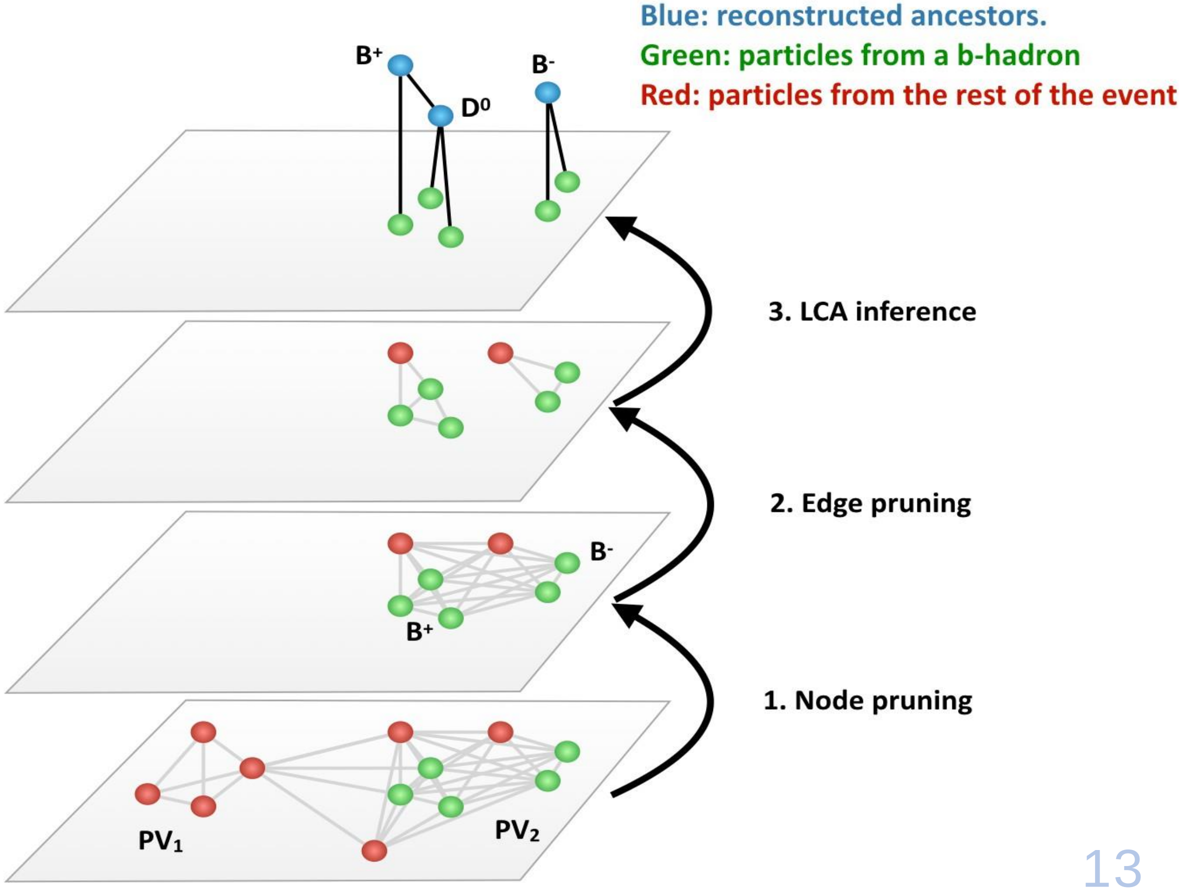
ML-based full event reconstruction

Comput Softw Big Sci 7, 12 (2023)]

- Deep-learning based Full Event Interpretation (**DFEI**):
  - One-go inclusive event filtering

  → Alternative to current approach: OR between HLT2/Sprucing lines + selective persistency of other associated objects in the event.

- First prototype:
  - Based on three sequential **GNN** modules.
  - Restricted to b-hadron decays and charged stable particles.
  - Only considers target ancestors which are "topologically" reconstructible.
  - Trained on custom simplified simulation in Run3-like conditions.

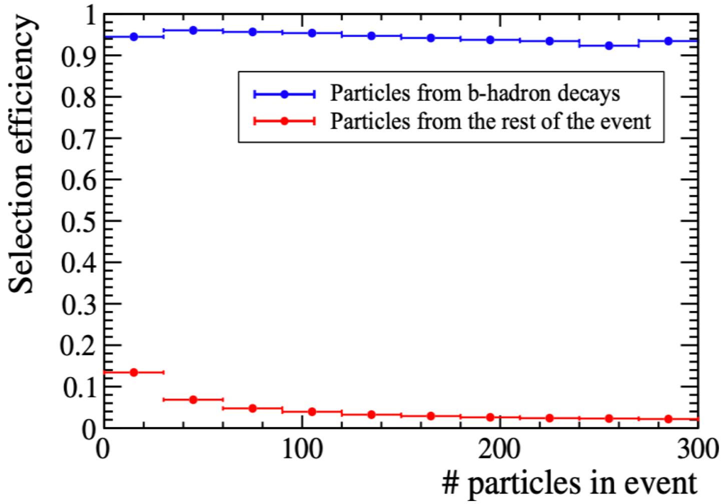


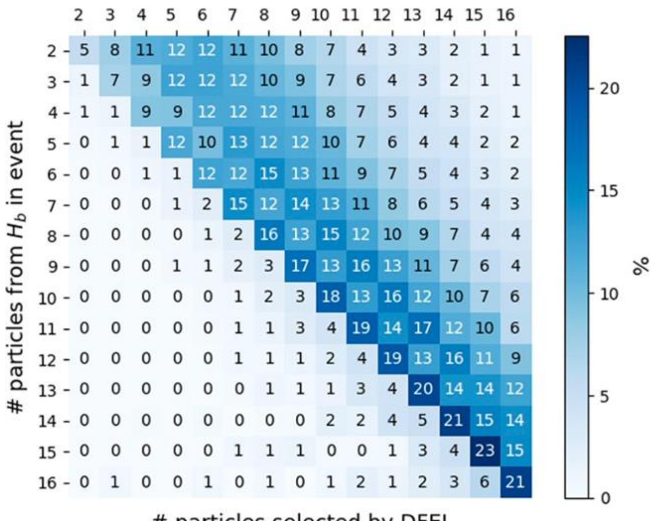
DFEI performance

Comput Softw Big Sci 7, 12 (2023)]

- In inclusive b-hadron simulation:
  - Acceptable perfect reconstruction rate
  - **Selection efficiency independent of the event multiplicity**

Decay mode	Perfect (%)	Wrong hierarchy (%)	Not iso. (%)	Part. reco. (%)
Inclusive H_b decay	4.6 ± 0.1	5.9 ± 0.1	76.0 ± 0.2	13.4 ± 0.1
$B^0 \to K_0^*[K\pi]\mu^+\mu^-$	35.8 ± 0.7	19.2 ± 0.6	44.9 ± 0.7	<0.02
$B^0 \to K^+\pi^-$	38.0 ± 0.7	–	54.7 ± 0.7	7.2 ± 0.4
$B_s^0 \to D_s^-[K^-K^+\pi^-]\,\pi^+$	32.8 ± 0.7	7.1 ± 0.4	53.7 ± 0.8	6.4 ± 0.4
$B^0 \to D^-[K^+\pi^-\pi^-]D^+[K^-\pi^+\pi^+]$	22.7 ± 0.6	22.4 ± 0.6	54.9 ± 0.8	<0.02
$B^+ \to K^+K^-\pi^+$	35.7 ± 0.7	10.2 ± 0.4	46.4 ± 0.7	7.7 ± 0.4
$\Lambda_b^0 \to \Lambda_c^+[pK^-\pi^+]\,\pi^-$	21.7 ± 1.0	8.9 ± 0.7	36.8 ± 1.2	32.6 ± 1.1
$B_s^0 \to J/\psi[\mu^+\mu^-]\,\phi[K^+K^-]$	26.9 ± 0.6	20.5 ± 0.5	52.5 ± 0.6	<0.02





Future:

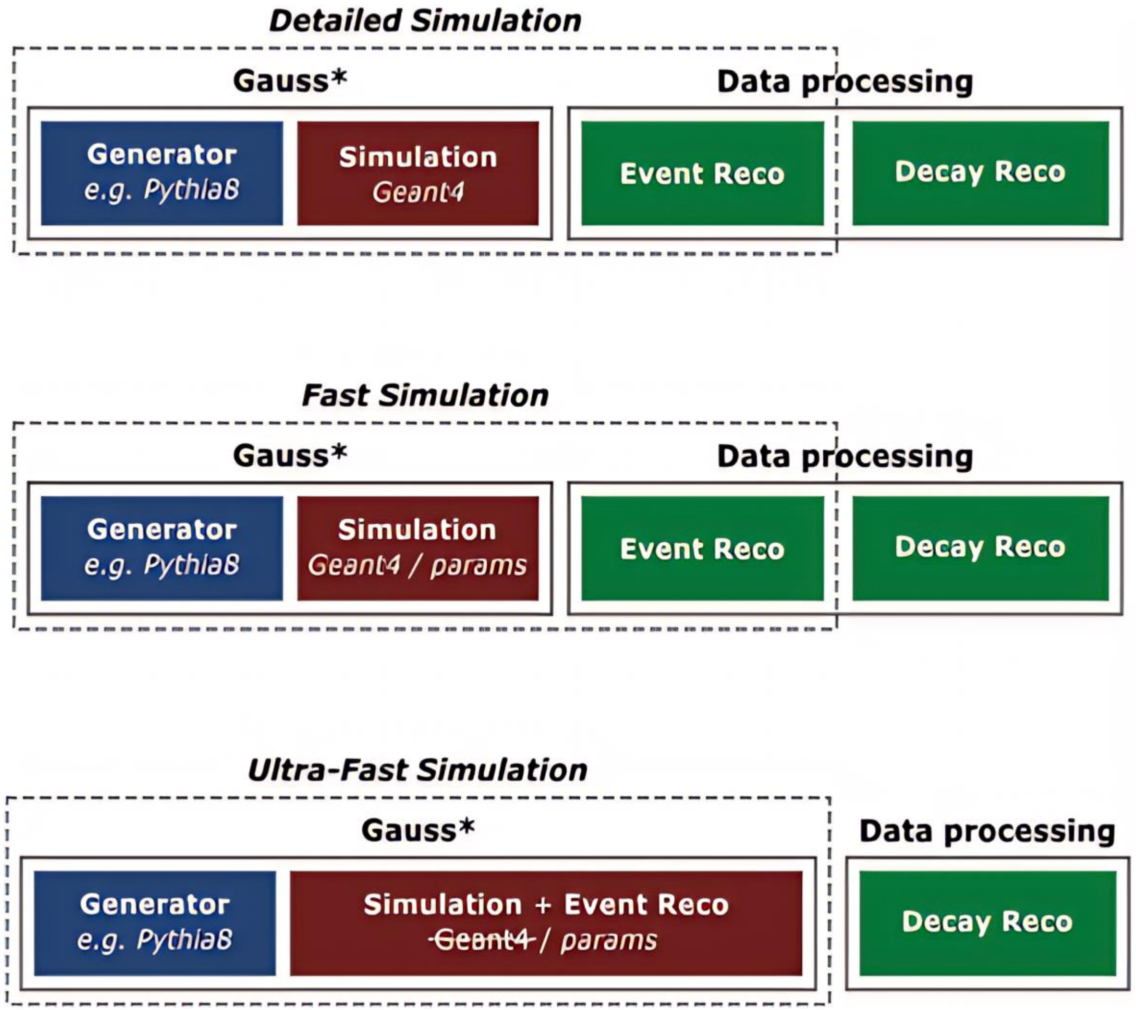
- **Implementation in the LHCb software stack**: targeting Sprucing for the near future and HLT2 in the long term.
- Detailed performance studies with simulation and with data.

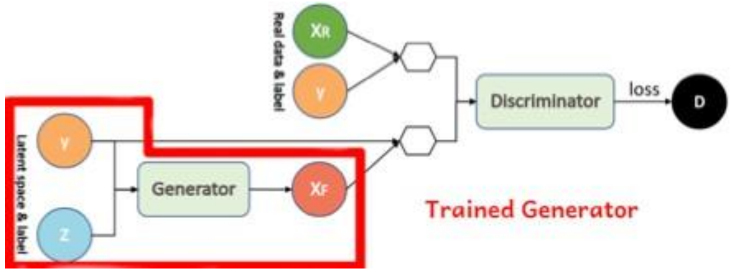
ML developments at LHCb

ML in simulation

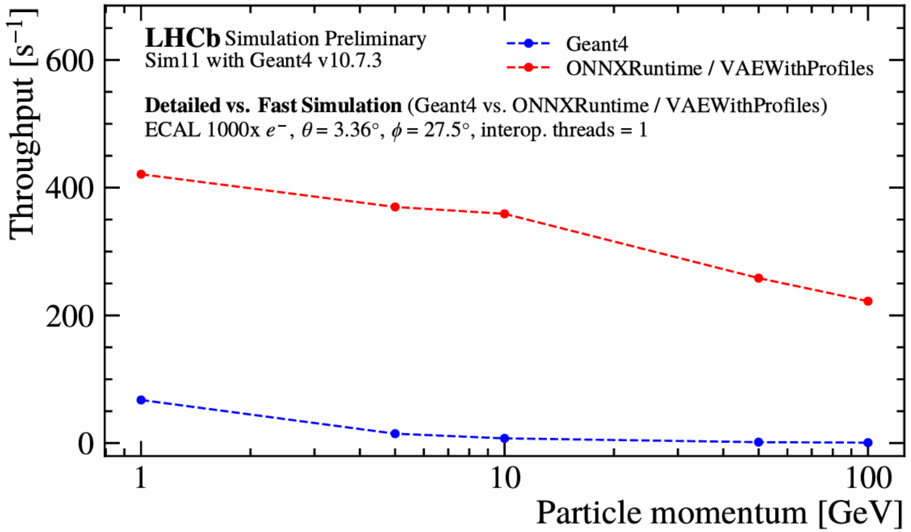
ML-based fast simulation at LHCb

- Multiple complementary techniques to speed up the simulation process.



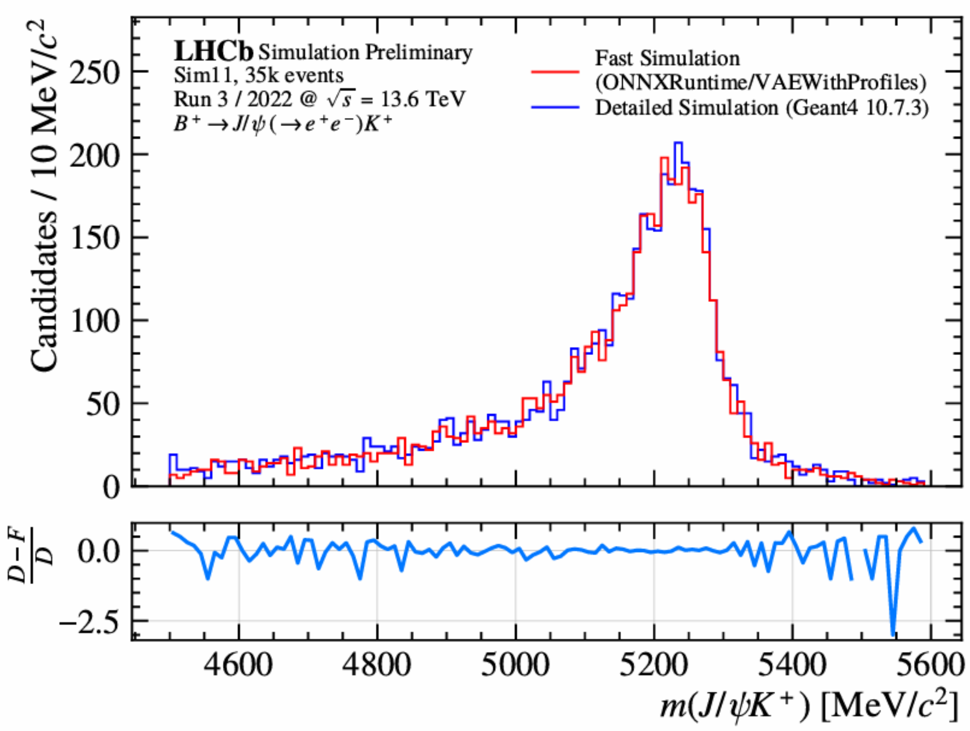


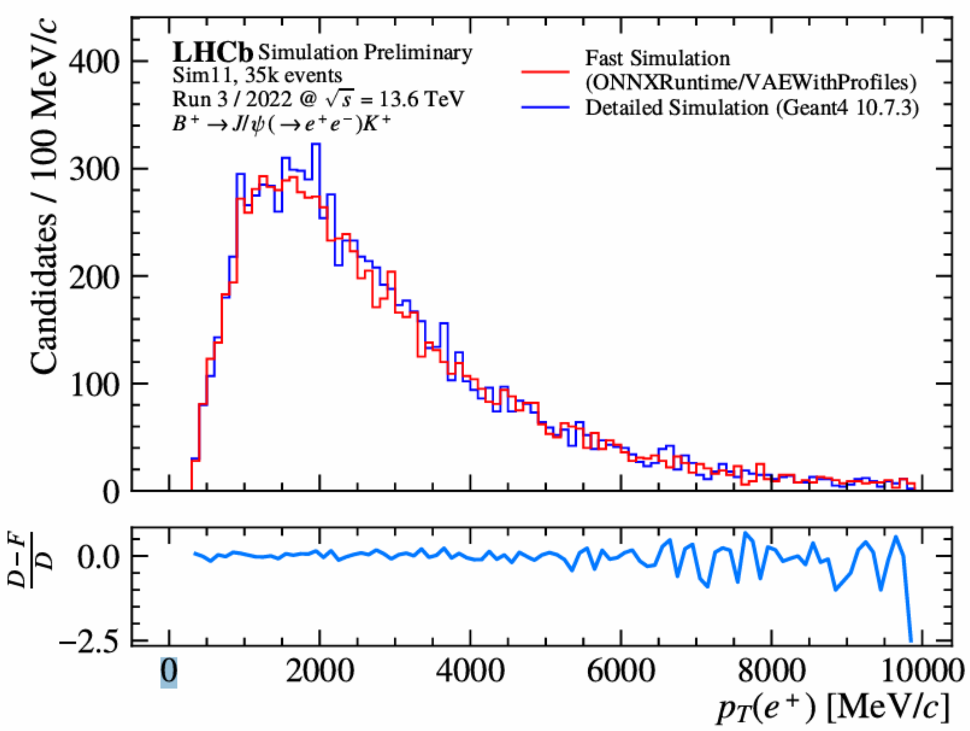
- CALO shower simulation with GANs



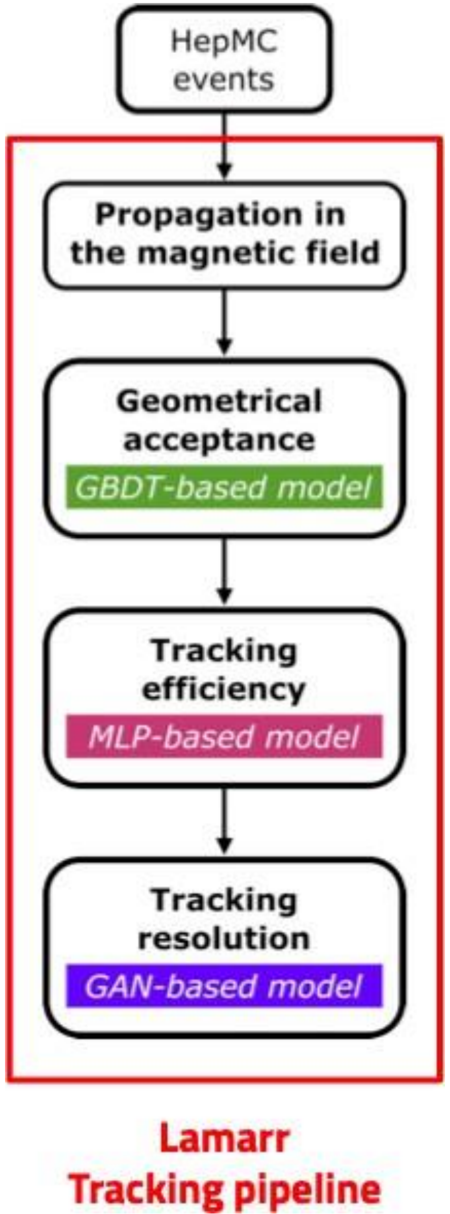
ML-based fast simulation at LHCb

- One model only for e^+, e^- and γ in the electromagnetic calorimeter
  - Detector simulation **speed up by 20x**.
  - **Around 1-4% energy difference** v.s. Geant4-based simulation
- Not official integrated yet. Still WIP.

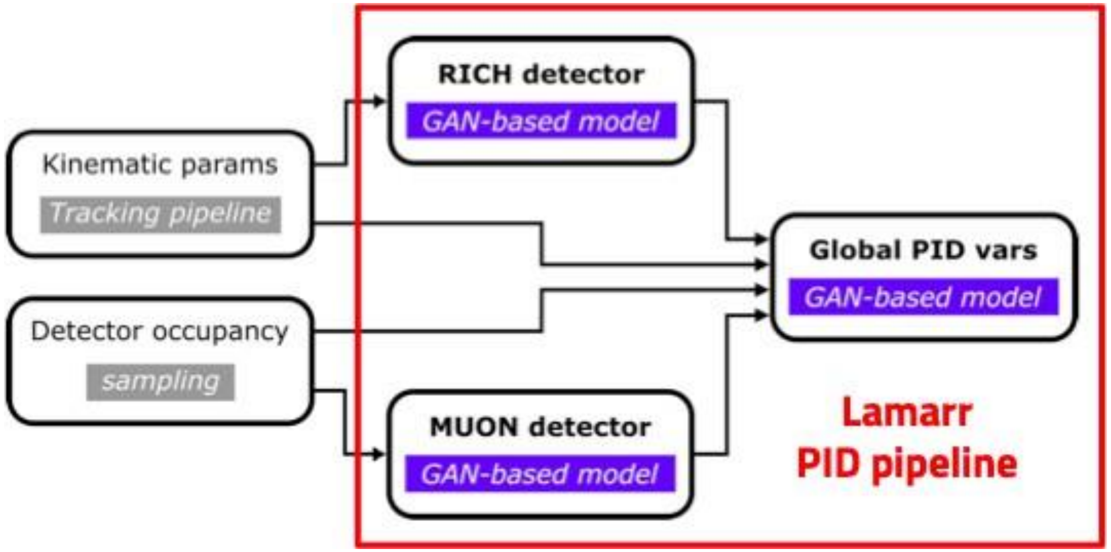




LAMARR: ML-based ultra-fast simulation at LHCb



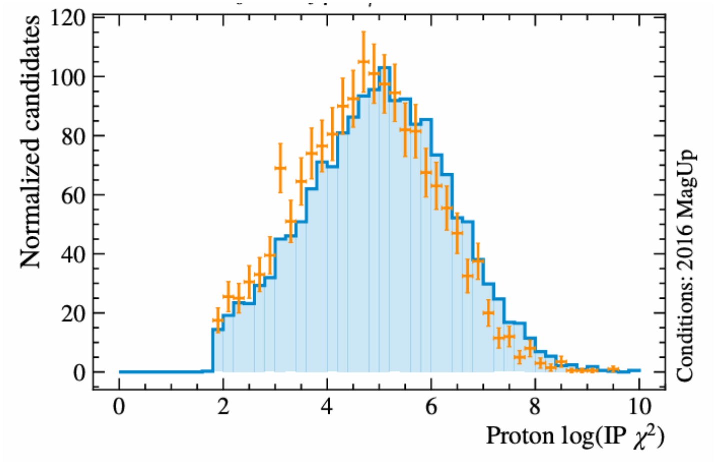
- **LAMARR**, a ultra-fast simulation using ML-based parametrizations deployed within Gauss.
  - Pipeline of modules parameterizing **both the detector response and the reconstruction algorithms** of the LHCb experiment.
- **Output high-level quantities directly**, including uncertainties on reconstructed quantities.
  - Detector simulation speed up by ~100x

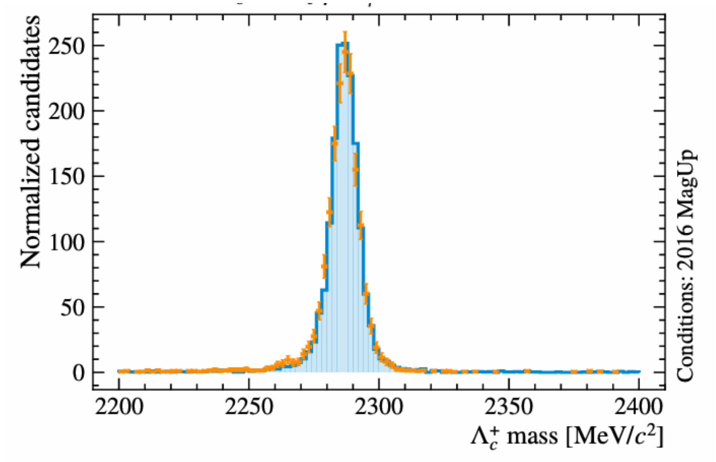


arXiv:2309.13213

LAMARR: ML-based ultra-fast simulation at LHCb

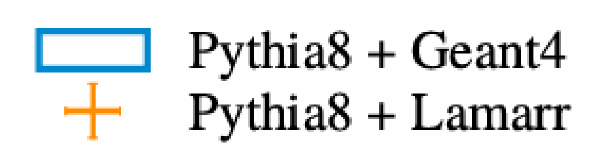
- Preliminary validation studies show good performance.
- **LAMARR is built within the LHCb simulation framework.**
  - Next: integration in the MC production system.

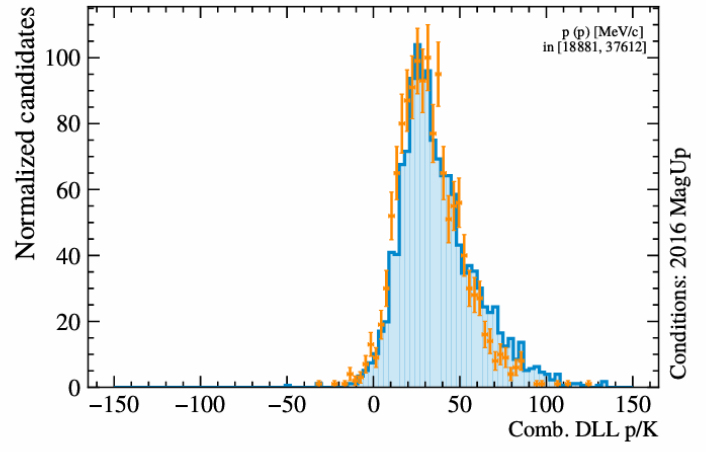


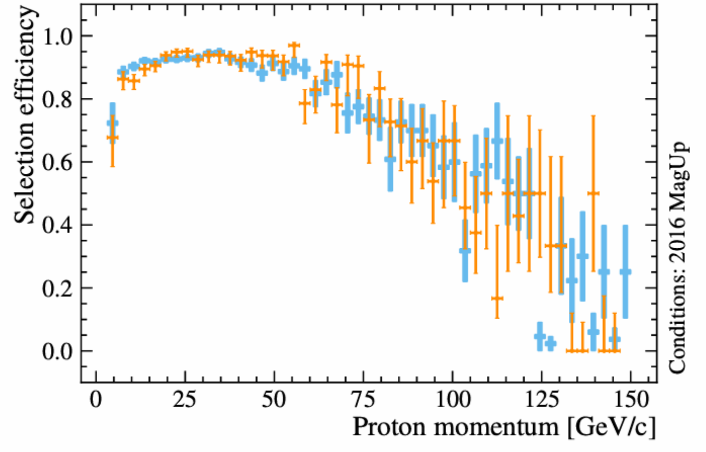


LHCb Simulation Preliminary
Protons from $\Lambda_b^0 \to \Lambda_c^+ \mu^- \bar{\nu}_\mu$

LHCB-FIGURE-2022-014







ML developments at LHCb

ML related organizations at LHCb

ML-related organizational structure in LHCb

- LHCb has 3 internal organizations related to machine learning.
- Stats. & ML WG：
  - ML discussion at the R&D level.
  - **ML for analysis**
- Co-coordinator of the Inter-Experimental LHC Machine Learning (IML):
  - Interface LHCb with the LHC community and helps organizing IML meetings
- LHCb ML Forum: (New!)
  - ML discussion at the production level on aspects which are either cross-project or LHCb-common (including common ML interfaces and pipelines, developments of ML for FPGA, usage of Large Language Models (LLM) for documentation, …).
  - Discussion of external ML opportunities for LHCb (requests of LHCb speakers for project-unspecific ML overview talks, new multi/interexperiment ML networks, available hardware infrastructure, ...)

Other ML developments in LHCb

- **Anomaly detection** in the muon system [arXiv:2105.05735]
- Reinforcement Learning from Human Feedback (RLHF) in **Data Quality Monitoring** [arXiv:2405.15508]
- Inclusive **Flavour Tagging** with DeepSets [arXiv:2404.14145]
- Robust Neural Networks for **particle identification** [arXiv:2212.07274]

Summary

ML is starting a new era at LHCb experiment!

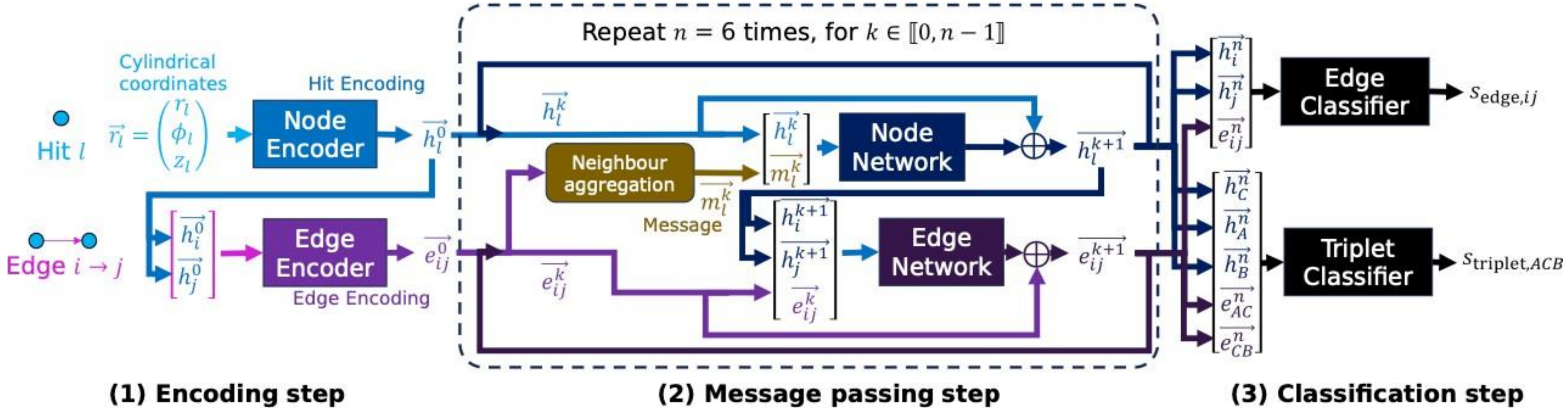
- Higher data statistics, tighter bandwidth limits, and more complex physical processes provide a wide range of applications for ML at LHCb.
- The application of ML in LHCb runs through the process of event reconstruction **from the lowest level to the highest leve**l and will be used as the main algorithm in the future upgrade process.
- A variety of neural networks, especially GNN, are widely used and most of the time perform better than traditional algorithms.

Thanks!

Backup

GNN model for ETX4VELO project

[arXiv:2406.12869]



Neutral particles in LAMARR

- To extend the LAMARR simulation to photons and electrons, an accurate simulation of the high-level ECAL response is required.
- Technical challenge:

number of generated particles ≠ number of reconstructed objects

(due to bremsstrahlung radiation, converted photons, and merged π^0)

- Two complementary approaches:
  - **Signal photons** (produced in decay modes under study): one-to-one relation possible.

    → Similar treatment as for charged particles.
  - **Secondary photons**: event-level description.

    → Two types of algorithms under study: **Transformers** and **GNNs**.

Further work needed for improving the performance.

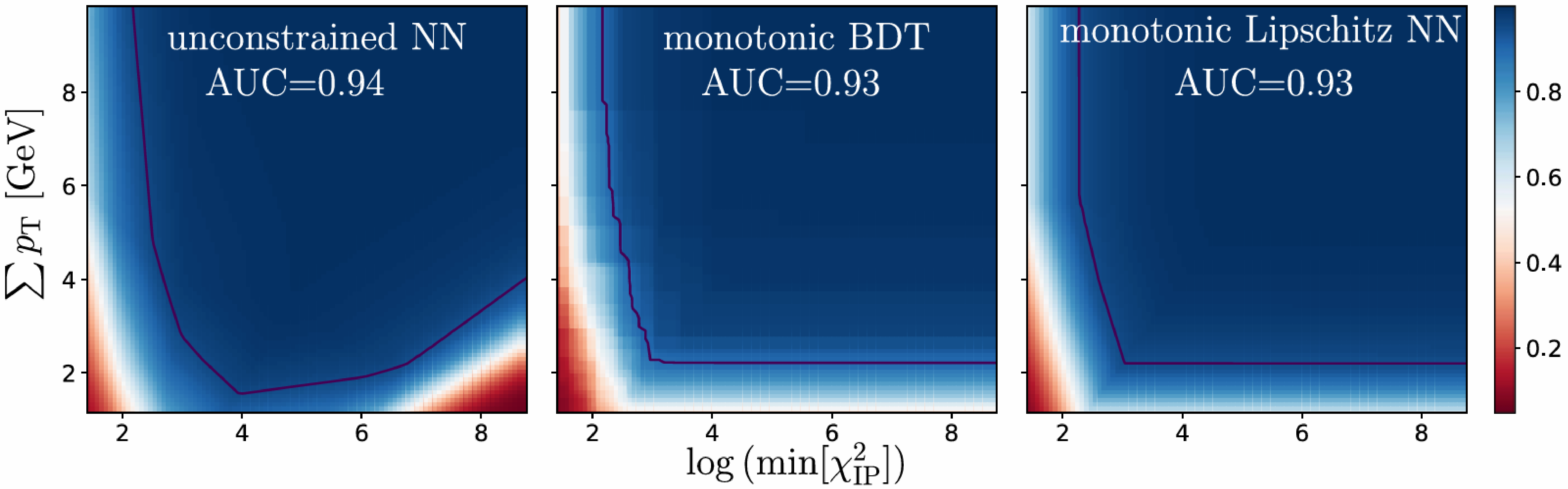


Figure 4. Simplified version of the LHCb inclusive heavy-flavor trigger problem using only 2 inputs, which permits displaying the response everywhere in the feature space; shown here as a heat map with more signal-like (background-like) regions colored blue (red). The dark solid line shows the decision boundary predicted to give the required output bandwidth in Run 3.