

Applications of Lipschitz neural networks to the Run 3 LHCb trigger system

Blaise Delaney^{1,2,*}, Nicole Schulte³, Gregory Ciezarek⁴, Niklas Nolte⁵, Mike Williams^{1,2}, and Johannes Albrecht³

¹Massachusetts Institute of Technology, Cambridge MA, USA

²NSF AI Institute for Artificial Intelligence and Fundamental Interactions (IAIFI)

³TU Dortmund University, Dortmund, Germany

⁴CERN, Meyrin, Switzerland

⁵Meta AI (FAIR)

Abstract. The operating conditions defining the current data taking campaign at the Large Hadron Collider, known as Run 3, present unparalleled challenges for the real-time data acquisition workflow of the LHCb experiment at CERN. To address the anticipated surge in luminosity and consequent event rate, the LHCb experiment is transitioning to a fully software-based trigger system. This evolution necessitated innovations in hardware configurations, software paradigms, and algorithmic design. A significant advancement is the integration of monotonic Lipschitz Neural Networks into the LHCb trigger system. These deep learning models offer certified robustness against detector instabilities, and the ability to encode domain-specific inductive biases. Such properties are crucial for the inclusive heavy-flavour triggers and, most notably, for the topological triggers designed to inclusively select b -hadron candidates by exploiting the unique kinematic and decay topologies of beauty decays. This paper describes the recent progress in integrating Lipschitz Neural Networks into the topological triggers, highlighting the resulting enhanced sensitivity to highly displaced multi-body candidates produced within the LHCb acceptance.

1 Introduction

The LHCb detector [1] located at the Large Hadron Collider at CERN is a single-arm forward spectrometer instrumented to achieve acceptance in the pseudorapidity range, $2 < \eta < 5$. The primary goal of the LHCb experiment is the discovery of Beyond the Standard Model (BSM) physics through the analysis of heavy-flavour processes, with particular focus placed on investigating b -hadron decays. Since its inception, however, the LHCb physics programme has grown substantially to include, among other endeavours, the search for feebly interacting dark-portal candidates produced in the LHCb geometrical acceptance [2–6].

Throughout its Run 3 data taking campaign, the LHCb experiment is tasked with operating under unprecedented conditions, namely a nominal instantaneous luminosity, $\mathcal{L} = 2 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$, amounting to a five-fold increase on the data taking conditions present in Runs 1 and 2. In order to cope with the challenging detector occupancy of Run 3, the

*e-mail: blaise.delaney@cern.ch

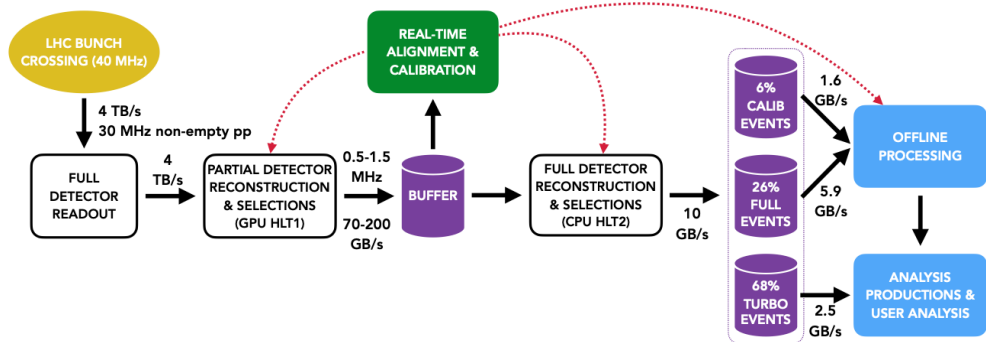


Figure 1: Schematic representation of the LHCb data flow during the Run 3 data taking period. The incoming rate of 30 MHz of non-empty bunch crossings is processed by the full detector readout and thereafter passed through the two-tiered trigger system. The selected data is stored and further processed offline for end-user analysis. Taken from [9].

LHCb experiment has pioneered the deployment of a redesigned, fully software-based trigger system for real-time data acquisition. In this paradigm, the full detector readout and event building is enacted at an incoming 30 MHz rate of visible proton-proton (pp) bunch crossings. A two-staged trigger system is deployed to select events of interest with a 100 kHz output rate suitable for storage.

The core objective of the LHCb trigger is therefore data-volume reduction within the real-time data storage constraints. This process translates to a reduction by a factor of approximately 400 of the input 4 TB/s bandwidth, at nominal instantaneous luminosity, to achieve an output bandwidth of 10 GB/s [7]. Figure 1 provides a schematic representation of the LHCb data flow in Run 3. Following the full detector readout, the GPU-enabled first trigger stage, the High Level Trigger 1 (HLT1), delivers partial event reconstruction from charged-track information, resulting in a data-volume reduction by a factor of approximately 20. Subsequently, the data is passed to a buffer stage to achieve real-time alignment and calibration, thus enabling the CPU-based High Level Trigger 2 (HLT2). This trigger stage operates selection algorithms exploiting offline-quality, full-event reconstruction observables.

To achieve the real-time rate reduction required by the Run 3 operating conditions, the LHCb trigger exploits a combination of expert systems and machine learning solutions. Among the latter, Lipschitz monotonic neural networks (NNs) [8] find optimal application in the inclusive algorithms deployed to select heavy-flavour particles, owing to their robustness, afforded by a minimally constraining weight normalisation scheme enforcing a bounded per-layer Lipschitz constant, and their ability to encode domain-specific inductive biases.

2 Applications of Lipschitz neural networks to the LHCb trigger

Implementing trigger selections is a critical step in High Energy Physics data acquisition, as discarded events are irreversibly lost. No margin for error is therefore afforded to the event-selection protocols implemented in the trigger system. Moreover, effective classification algorithms must be able to process event-level information within the memory and compute limits imposed by real-time data acquisition. The application of monotonic Lipschitz NNs to the LHCb trigger meets such mission-critical requirements whilst delivering additional desirable certified benefits: namely, robustness and interpretability.

In this work, robustness signifies mitigated sensitivity to both simulation inaccuracies and detector instabilities during data taking. Such conditions may be achieved by constraining the gradient of the function approximated by a deep learning classifier with respect to the input features. The models deployed in the LHCb experiment realise this result by bounding the Lipschitz constant of the learnt decision-boundary function. This approach effectively sets a strict upper limit on the classifier-response variation resulting from experimental instabilities of limited magnitude. Robustness is thus essential for the estimators deployed in the trigger system, serving to reduce sensitivity to resolution and calibration effects conditioning the input features in real time. Additionally, robustness simplifies the evaluation of trigger-related systematic uncertainties in end-user physics measurements.

A complementary advantage is offered by enforcing a monotonic estimator response with respect to a set of input features. In essence, this procedure expresses domain-specific inductive bias. Deploying provably monotonic networks in the trigger system enhances the selective retention of interesting outlier candidates absent from training samples and, therefore, not learnt by the trigger models. Specifically, in the context of the LHCb heavy-flavour triggers, such architectures enable enhanced sensitivity to potential yet-undiscovered feebly interacting BSM states produced within the LHCb acceptance.

By adopting the deep learning models introduced in [8], robustness and certified monotonic response with respect to a set of input features are realised through a minimal set of architecture-level constraints. Such conditions enable highly expressive classifiers capable of inference within the bandwidth-rate constraints set by the LHCb Run 3 trigger operations. Consequently, Lipschitz monotonic NNs supersede the decision-tree implementations of the inclusive heavy-flavour triggers deployed in previous data taking campaigns [10, 11]. The topological triggers, presented in the following section, exemplify the advantages obtained through this methodological advancement.

3 The LHCb topological triggers

The topological triggers are designed to inclusively select b -hadron decay processes based on the offline-quality information available at the HLT2 trigger stage. Decays of beauty particles display a distinct topology: owing to their forward boost in the detector, they exhibit a long lifetime [12], causing them to traverse distances of $O(1\text{ cm})$ before decaying within the detector acceptance. Additionally, due to their relatively high mass, beauty hadrons typically exhibit sizeable transverse momentum, p_T . Combined, both conditions make for a distinct experimental signature. The topological triggers thus aim to select displaced secondary decay-vertex candidates reconstructed in the pairwise combination of final-state charged tracks.

Two variations of the Run 3 topological architectures have been incorporated in the LHCb online software stack, henceforth referred to as the two- and three-body topological triggers. These are designed to identify beauty secondary-vertex candidates reconstructed in the combination of two or three charged particles, respectively. Such a suite of topological triggers thus facilitates the identification of multi-particle beauty decays. The two-body composites are reconstructed from final-state tracks compatible with originating from a well-resolved secondary-vertex candidate appreciably displaced from the pp collision point. The three-body candidates, in turn, are reconstructed by adding a companion track to the two-body object and imposing similar vertex-fit and kinematic criteria as for the two-body counterparts. Notably, the topological triggers record the full-event level information, writing to the so-called *Full Stream*, depicted in Figure 1. As a result, n -body b -hadron decays, with $n > 3$, may be successfully reconstructed by combining additional tracks persisted in the event with the trigger-selected two- and three-body candidates. Crucially, this design maximises the selection efficiency on signal whilst meeting the HLT2 output-rate requirements.

Table 1: Features used to train the two-body classifier, along with the monotonicity requirements and the operations enacted to rescale the feature distributions to a range of order unity. The operators $\cdot\{\}$ and \sum run over the final-state tracks, and are evaluated on a per-candidate basis. The shorthand notations GeV and log signify rescaling from units of $\text{MeV}/c^{(2)}$ to $\text{GeV}/c^{(2)}$ and evaluating the natural logarithm of the observable, respectively.

Feature	Monotonicity	Rescaling to $O(1)$ procedure
$\min\{p_T\}$	Increasing	GeV
$\sum p_T$	Increasing	GeV
$\min\{\chi_{\text{IP}}^2\}$	Increasing	log
$\max\{\chi_{\text{IP}}^2\}$	—	log
Composite p_T	Increasing	GeV
Composite $\max(\text{DOCA})$	—	—
Composite m_{corr}	—	GeV
Composite flight distance χ^2	—	log
Composite χ_{vtx}^2	—	—

The HLT2 topological triggers are tasked with rejecting several sources of background: combinatorial and soft-QCD processes; interactions of particles with the LHCb detector material, exhibiting softer p_T than the beauty signal, and a comparatively larger flight distance from the pp interaction point; fake particles, typically referred to as *ghosts*, erroneously inferred from tracking-level information and exhibiting high p_T , as a linear trajectory within the detector translates to the maximal possible momentum reconstructed by the tracking system; and charm decays, rendered challenging by the high charm production cross section, approximately $O(10)$ times higher than the beauty counterpart, and topologies compatible with the signal, albeit with comparatively reduced lifetimes [12].

Each topological trigger exploits a two-staged selection to achieve effective signal isolation. Firstly, a cut-based prefilter is devised to discard prompt and soft background candidates. Subsequently, a deep learning classifier is implemented as a monotonic Lipschitz NN. The two- and three-body models share the same architecture complexity, amounting to four hidden layers comprising 16, 32, 64, and 32 internal nodes, respectively. This design choice mitigates inference-time consumption during real-time trigger operations. Model training is enacted through an offline Python-based pipeline exploiting the PyTorch [13] software package. Thereafter, the network weights are ported to the LHCb trigger software stack for event-based inference in real time.

Both topological triggers are optimised by considering a suite of exclusive simulations representative of the LHCb beauty physics programme. In this way, the respective selection criteria and architecture complexity are optimised to attain sensitivity to the characteristic decay topologies and kinematics of b -hadron decays. The exclusive signal Monte Carlo (MC) simulations are combined to contribute the same number of decay-vertex candidates to the inclusive signal sample. This, in turn, is used both to optimise the cut-based prefilter and train the NNs. Crucially, this procedure prevents potential biases in the form of heightened sensitivity to a subset of signal channels, by construction. The background is modelled by an inclusive minimum-bias MC sample, generated to represent the average content of a pp collision. Notably, events containing beauty candidates are filtered from the minimum-bias sample, making it a suitable inclusive proxy of the aforementioned background processes.

The NNs are trained on feature sets maximising the post-prefilter signal-to-background divergence whilst maintaining low pairwise feature correlation. The two- and three-body

Table 2: Features used to train the three-body classifier, along with the monotonicity requirements and the operations enacted to rescale the feature distributions to a range of order unity. The operators $\cdot\{\}$ and \sum run over all final-state tracks, and are evaluated on a per-candidate basis. Conversely, the $\cdot_{2\text{body}}\{\}$ operations run over the two-body children only. The shorthand notations GeV and log signify rescaling from units of MeV/c⁽²⁾ to GeV/c⁽²⁾ and evaluating the natural logarithm of the observable, respectively.

Feature	Monotonicity	Rescaling to $O(1)$ procedure
$\min\{p_T\}$	Increasing	GeV
$\sum p_T$	Increasing	GeV
$\min_{2\text{body}}\{p_T\}$	Increasing	GeV
$\sum_{2\text{body}} p_T$	Increasing	GeV
$\min\{\chi_{\text{IP}}^2\}$	Increasing	log
$\max\{\chi_{\text{IP}}^2\}$	—	log
Three-body p_T	Increasing	GeV
Two-body child p_T	Increasing	GeV
Three-body max(DOCA)	—	—
Two-body child max(DOCA)	—	—
Two-body child χ_{IP}^2	—	log
Two-body child m_{corr}	—	GeV
Three-body m_{corr}	—	GeV
Three-body flight distance χ^2	—	log
Two-body child flight distance χ^2	—	log
Three-body χ_{vtx}^2	—	log
Two-body child χ_{vtx}^2	—	—

feature sets are presented in Tables 1 and 2. Crucially, the adopted inputs must capture the decay topology and kinematics of multi-body beauty composite candidates. To this end, the NNs are trained on a suite of kinematic and geometric observables comprising the candidate transverse momentum, p_T ; the decay-vertex fit quality, χ_{vtx}^2 ; flight distance χ^2 ; the charged-tracks distance of closest approach (DOCA); the multi-body corrected mass value, m_{corr} [14]; and the IP χ^2 , the impact parameter χ^2 with respect to the primary vertex of the multi-body decay-vertex objects. Similar kinematic and geometric features are extracted from the charged final-state tracks.

The features are subjected to a preprocessing stage constraining each input observable distribution to a range of $O(1)$. Additionally, a 5σ window is imposed about the mean of each feature distribution. Input values exceeding the retention boundaries are mapped onto the last value accepted in the per-feature selection interval. Combined, these preprocessing steps stabilise the performance of the classifier without depleting the statistical population of the training samples. Moreover, comparing similarly ranged distributions facilitates the choice of the Lipschitz-constant bound, λ , of the learnt decision functions.

Separate upper bounds are imposed on the two- and three-body Lipschitz constants, respectively. These are individually selected to maximise reconstruction efficiency on the multi-body signal candidates whilst maintaining compatibility of the resolution expected of the LHCb detector. In this way, the Lipschitz-bound assignments prevent significant variations in the classification score produced by the estimators, in the limit of the inputs varying within their respective resolution scale.

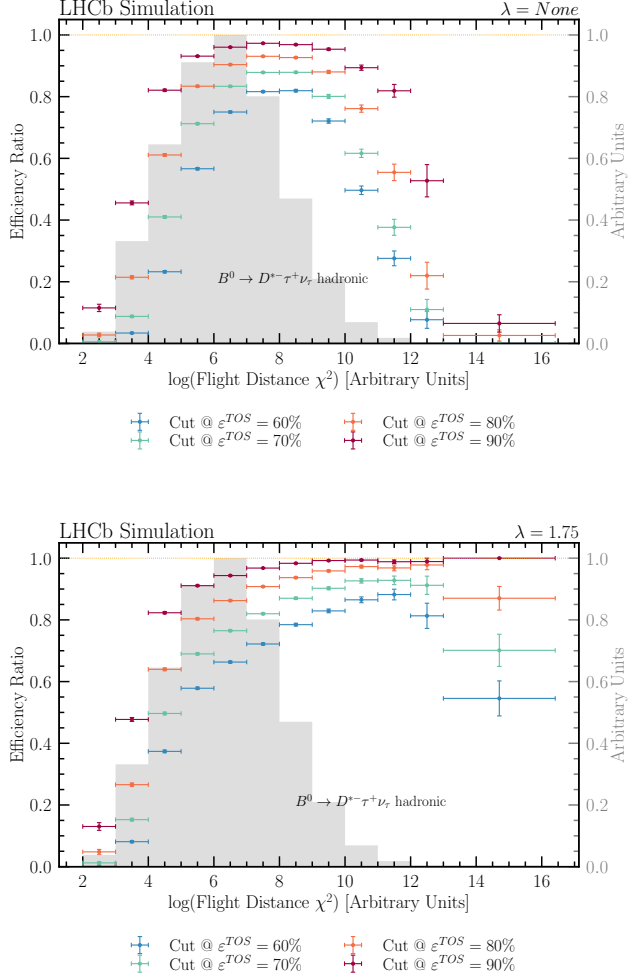


Figure 2: Selection Efficiency on simulated $B^0 \rightarrow D^{*-}\tau^+\nu_\tau$ candidates reconstructed from two final-state tracks. Top: performance delivered by a feed-forward, unconstrained network. Bottom: efficiency values obtained by the two-body monotonic Lipschitz network. The normalised B^0 logarithmic flight distance χ^2 distribution is shown in the shaded grey histogram. Binwise efficiencies for response cuts yielding global trigger-on-signal efficiency values, ε^{TOS} , at the 60%-90% level, are marked. The Lipschitz bound, λ , is shown where relevant.

Finally, a monotonically increasing response is enforced with respect to a subset of features, identified independently for the two- and three-body classifiers. Monotonicity is imposed in the IP χ^2 and the p_T budget of the composite candidates and their constituents. This conservative choice prevents the introduction of unwanted biases in selection efficiency due to monotonicity. Furthermore, this approach delivers enhanced sensitivity to highly displaced outliers compared to a baseline unconstrained model with identical network complex-

ity, as demonstrated by Figure 2. The per-candidate selection efficiency delivered by the two-body topological trigger is evaluated for simulated $B^0 \rightarrow D^{*-}\tau^+\nu_\tau$ decays, where the τ lepton decays hadronically, excluded by the training MC samples. The efficiency extraction is performed in bins of the signal logarithmic χ^2 fit-quality of the partially reconstructed B^0 candidate flight distance with respect to the pp collision point, taken to be a proxy for the decay-vertex candidate lifetime. Compared to the performance attained by the baseline unconstrained NN, the topological trigger delivers comparatively higher selection efficiencies for highly displaced outliers. Concurrently, it delivers compatible performance in the remainder of the observed range. Such a trend is evident when requiring thresholds on the respective network responses delivering overall efficiency values at the 60%-, 70%-, 80%-, and 90%-level on signal. The efficiency drop evident at very high displacement is due to the fact that monotonicity is enforced in a subset of the input features. Consequently, all else being equal, the distributions of other features can be much more background-like at large flight distance, leading to a marginal efficiency drop in the high tail of the flight distance χ^2 projection. Nevertheless, the results presented in Figure 2 demonstrate the capacity of the two-body topological trigger to efficiently select signal b -hadron decays and enhance sensitivity to highly displaced candidates. This property bolsters the capacity to retain outlier multi-body objects, and thus feebly interacting BSM candidates produced in the LHCb acceptance.

4 Summary and Discussion

LHCb is pioneering the deployment of a fully software-based trigger in the LHC Run 3 data taking campaign. Such a paradigm shift facilitates the adoption of Lipschitz monotonic NNs in the LHCb inclusive heavy-flavour triggers. These models are particularly well suited to the real-time selection of events containing decay-vertex candidates exhibiting topologies and kinematics compatible with heavy-flavour decays.

The utility of monotonic Lipschitz NNs is exemplified by the topological triggers deployed in the highest level of the LHCb Run 3 trigger system, HLT2. The preliminary results presented in this contribution demonstrate the capacity to efficiently select two- and three-body signal candidates. Enforcing a monotonically increasing response with respect to a subset of input features yields enhanced sensitivity to highly displaced decay-vertex candidates, as evaluated on a probe decay channel. From a physics standpoint, this performance strengthens sensitivity to beauty candidates at high lifetime and, notably, to potential BSM states produced within the LHCb detector. Furthermore, the Lipschitz bound applied to the learnt decision function provides certified protection against instabilities during data taking, thereby easing the evaluation of the relevant systematic uncertainties in end-user measurements.

The combined output of the topological triggers dominates the HLT2 bandwidth allocation. The threshold imposed on the classifier response for real-time inference on data, in turn, is fixed by the maximum bandwidth allocated to the HLT2 trigger stage when writing data to storage. Optimisation studies are currently underway to improve the classification power of these triggers. The aim is to maximise selection efficiency for standard-candle beauty decay modes whilst satisfying the HLT2 output-rate constraints.

Finally, as robustness and monotonicity are ideal inductive biases in experimental particle physics, investigations are progressing to deploy Lipschitz NNs in the Run 3 LHCb tracking, lepton identification and ghost-rejection algorithms.

5 Acknowledgements

The authors would like to thank the LHCb computing and simulation teams for their support and for producing the simulated LHCb samples used in the paper. The authors would also like to thank the LHCb online, DPA and RTA team for providing most of the software that this project was built upon and for all the support and guidance that was given. BD and MW were supported by NSF grant PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, <http://iaifi.org/>). BD and MW were also supported by U.S. NSF grant PHY-2209181. NS and JA acknowledge funding from the German Science Foundation DFG, within the Collaborative Research Center SFB1491 “Cosmic Interacting Matters - From Source to Signal.” JA acknowledges funding from the German Federal Ministry of Education and Research (BMBF, grant no. 05H21PECL1) within ErUM-FSP T04.

References

- [1] A.A. Alves, Jr. et al. (LHCb), *JINST* **3**, S08005 (2008)
- [2] R. Aaij et al. (LHCb), *Phys. Rev. Lett.* **120**, 061801 (2018), 1710.02867
- [3] R. Aaij et al. (LHCb), *Phys. Rev. Lett.* **124**, 041801 (2020), 1910.06926
- [4] R. Aaij et al. (LHCb), *JHEP* **10**, 156 (2020), 2007.03923
- [5] R. Aaij et al. (LHCb), *Phys. Rev. Lett.* **115**, 161802 (2015), 1508.04094
- [6] R. Aaij et al. (LHCb), *Phys. Rev. D* **95**, 071101 (2017), 1612.07818
- [7] R. Aaij et al. (LHCb), *The LHCb upgrade I* (2023), 2305.10515
- [8] O. Kitouni, N. Nolte, M. Williams, *Mach. Learn. Sci. Tech.* **4**, 035020 (2023), 2112.00038
- [9] *RTA and DPA dataflow diagrams for Run 1, Run 2, and the upgraded LHCb detector* (2020), <https://cds.cern.ch/record/2730181>
- [10] V.V. Gligorov, M. Williams, *JINST* **8**, P02013 (2013), 1210.6861
- [11] T. Likhomanenko, P. Ilten, E. Khairullin, A. Rogozhnikov, A. Ustyuzhanin, M. Williams, *J. Phys. Conf. Ser.* **664**, 082025 (2015), 1510.00572
- [12] R.L. Workman, Others (Particle Data Group), *PTEP* **2022**, 083C01 (2022)
- [13] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimeshein, L. Antiga et al., *Advances in neural information processing systems* **32** (2019)
- [14] K. Abe et al. (SLD), *Phys. Rev. Lett.* **80**, 660 (1998), hep-ex/9708015