



Impact of 4D Tracking in Flavour Tagging Applications

Summer student report

Andraž Tomšič

Advisor: Valentina Cairo

Co-advisors: Lorenzo Santi, Nicole Hartmann, Pierfrancesco Butti

August 9, 2024

Abstract

The High-Luminosity Large Hadron Collider (HL-LHC) presents significant challenges for flavour tagging due to increased pile-up events. This study explores the potential integration of precise timing information into the ATLAS detector, specifically targeting future upgrades beyond the confirmed Inner Tracker (ITk), to enhance flavour tagging performance. By improving the separation between hard scatter and pile-up interactions, timing information significantly enhances the discriminating power of machine learning models. Additionally, the study investigates similar applications for the Open Data Detector (ODD) and outlines the necessary steps to address data quality and training sample size issues. The results highlight the potential benefits of timing detectors for future colliders, such as the Future Circular Collider (FCC), and other high-energy physics projects.

Contents

1	Introduction	2
2	Challenge of Flavour Tagging in HL-LHC	2
3	Importance of Timing Information	3
3.1	Improvement in Hard Scatter Time Reconstruction	3
3.2	Removal of Pile-Up Tracks Associated with Jets	4
4	Analysis Pipeline	5
4.1	DIPS Model	5
4.2	Inputs to the DIPS Model	5
5	Flavour Tagging Performance in ATLAS Detector	7
6	Flavour Tagging Performance in Open Data Detector	8
7	Conclusion	9
8	Acknowledgments	9

1 Introduction

The ATLAS Experiment at CERN [1] is continuously evolving to meet the demands of the High-Luminosity Large Hadron Collider (HL-LHC). The Inner Tracker (ITk), a full-silicon Inner Tracker, is being built for the HL-LHC and is expected to be installed and begin collecting data towards the end of this decade. After collecting about 2000 fb^{-1} of data, it is anticipated that the ITk will need to be replaced due to radiation damage. For this future replacement, one of the significant upgrades being explored is the integration of 4-dimensional (4D) trackers [2], which are designed to measure both spatial and temporal coordinates with high precision. This upgrade, if approved, is expected to enhance the tracking capabilities, especially in the challenging environment of HL-LHC where pile-up scenarios involve approximately 200 interactions per bunch crossing [3].

In the summer student program, I focused on testing the performance of trackers for different detectors, including the Inner Tracker (ITk) within the ATLAS framework [4] and the Open Data Detector (ODD) [5]. For my project, I concentrated on the application of flavour tagging, a technique crucial for identifying the origin of particle jets. Machine learning (ML) methods are employed for flavour tagging, with current efforts using open-source tools like DIPS [6]. However, more advanced graph neural networks (GNNs) such as GN1 and its variations [7] are now being explored for even better performance.

We anticipate that the inclusion of precise track timing will significantly improve the accuracy of flavour tagging. My study aimed to quantify this improvement by evaluating the performance of the trackers in various scenarios using the A Common Tracking Software (ACTS) framework [8], which is detector-agnostic and can be extended to future colliders. Different time resolutions are explored in the simulations to assess their potential impact on improving flavour tagging accuracy.

2 Challenge of Flavour Tagging in HL-LHC

The HL-LHC presents significant challenges due to the increased number of proton-proton (pp) interactions per bunch crossing, resulting from denser beams, leading to increased pile-up (PU). PU refers to the multiple pp interactions occurring within the same or nearby bunch crossings, which creates a denser environment. This elevated PU causes some PU tracks to be falsely associated with

jets, generating noise. Flavour tagging, which assigns a flavour to a quark that has hadronized into a jet measured by the detector, relies on accurately associating tracks to the correct vertices. The additional noise from PU interactions complicates this process, decreasing the performance of flavour tagging by increasing the likelihood of PU tracks being incorrectly associated with jets. This makes it more challenging to identify the origin of particle jets accurately.

Figure 1 shows the longitudinal view of a simulated $t\bar{t}$ event with $\langle\mu\rangle = 200$ in the ATLAS Inner Tracker. Figure 1(a) shows the tracks associated with vertices when no track-timing information is available. When track-timing information is available, as depicted in Figure 1(b), the association of tracks to vertices becomes more unambiguous.

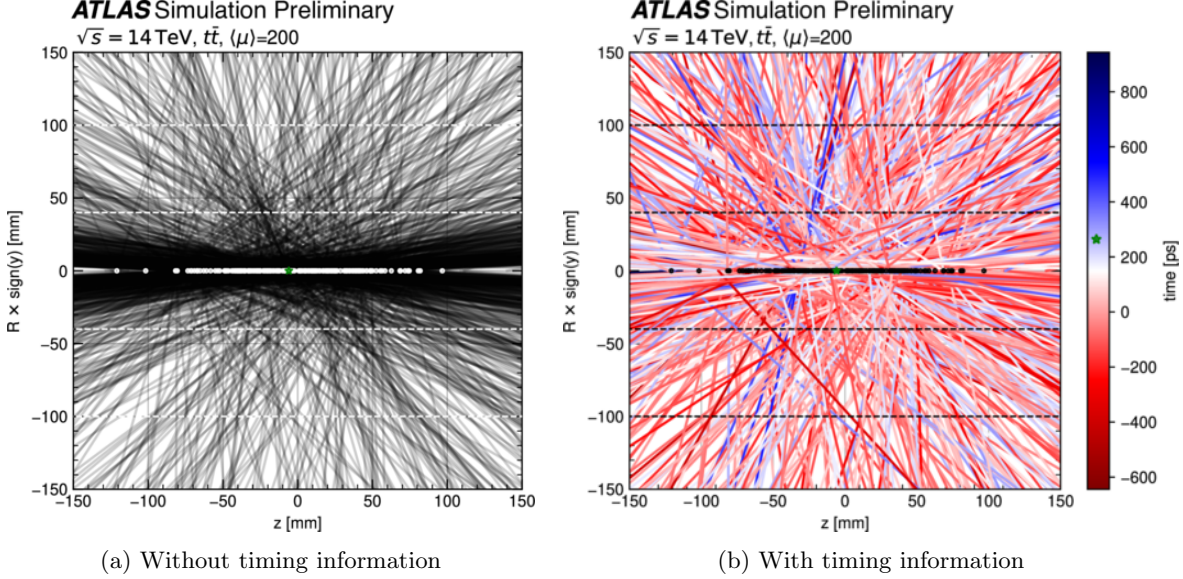


Figure 1: Longitudinal view of a simulated $t\bar{t}$ event with $\langle\mu\rangle = 200$. The truth hard-scatter is indicated with a green star, while the truth pile-up vertices are displayed as solid circles. The reconstructed tracks associated with the truth vertices are displayed. Taken from [9].

3 Importance of Timing Information

The integration of precise timing information in the ATLAS detector can help with distinguishing between tracks originating from HS and PU interactions thereby improving the accuracy of event reconstruction.

3.1 Improvement in Hard Scatter Time Reconstruction

Timing information enables better clustering of tracks and more precise reconstruction of the HS time [9]. By incorporating timing data, the reconstruction algorithms can more effectively identify and separate tracks belonging to different interactions, leading to a clearer distinction between HS and PU events. This improvement is illustrated in Figure 2, which shows the distribution of the difference between the reconstructed and true vertex times. The use of timing information allows the reconstruction of the HS time to approach the performance where the truth is known.

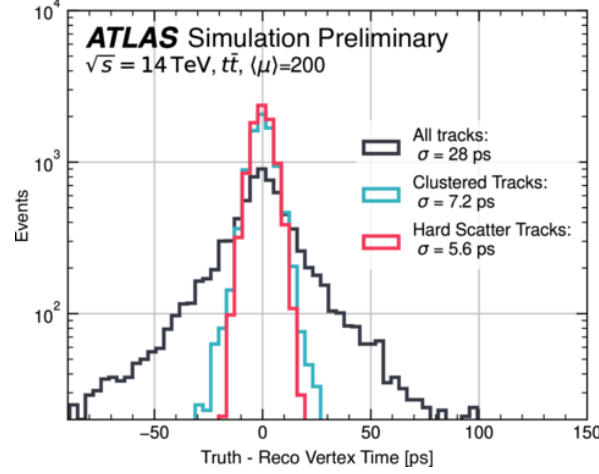
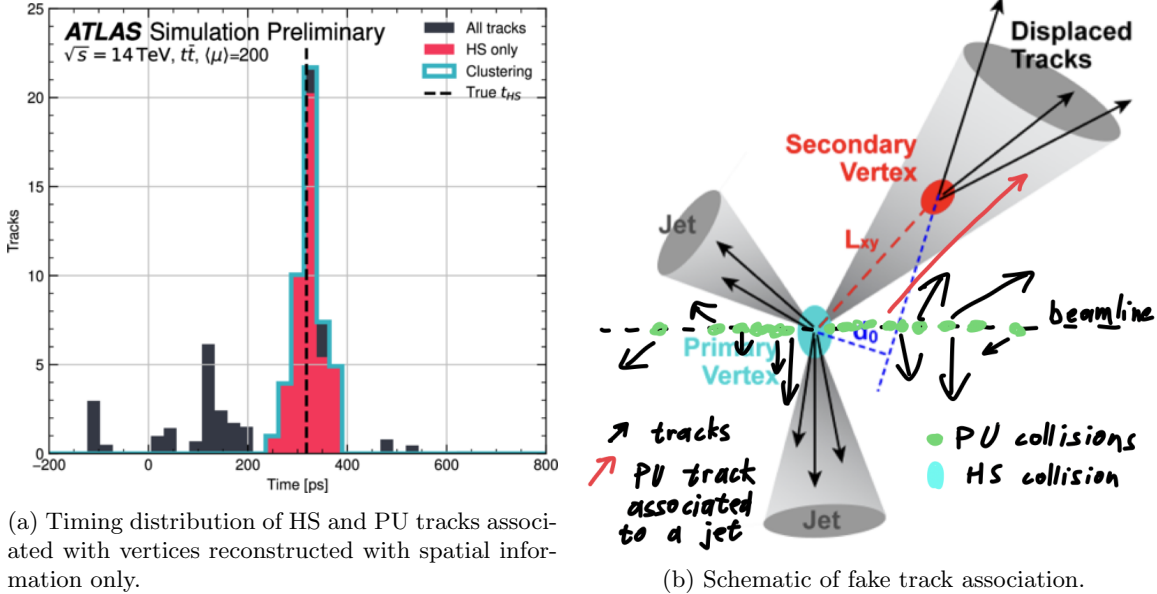


Figure 2: Distribution of the difference between reconstructed and true vertex times for simulated $t\bar{t}$ events with $\langle\mu\rangle = 200$. Timing information enables better clustering and more precise HS time reconstruction of each event. Taken from [9].

3.2 Removal of Pile-Up Tracks Associated with Jets

Another significant benefit of timing information is the removal of PU tracks that are falsely associated with jets [9]. In the high-density environment of the HL-LHC, the spatial separation alone is less successful in correctly associating tracks with their originating vertices. Timing information adds an additional layer of discrimination, allowing for a more accurate identification of tracks from PU interactions.



(a) Timing distribution of HS and PU tracks associated with vertices reconstructed with spatial information only.

(b) Schematic of fake track association.

Figure 3: Removal of PU tracks associated with jets using timing information. Subfigure 3a shows the timing distribution of HS and PU tracks associated with vertices and Subfigure 3b provides a schematic of how timing information helps in the precise identification and removal of PU tracks. Adapted from [9].

Figure 3 demonstrates this benefit. Subfigure 3a shows the distribution of tracks in time, highlighting the distinction between HS and PU tracks associated with vertices reconstructed with spatial information only. Subfigure 3b, which includes a detailed schematic of the vertex and track associations, shows how timing information helps in the precise identification and removal of PU tracks from jets. Specifically, the red arrow in Subfigure 3b indicates a PU track associated with a jet, which can be removed if timing information is available, as the track time of PU tracks is likely different from that of HS tracks associated with a jet. This discrimination is learned by neural networks (NN), which can use the timing information to identify tracks with times significantly different from HS tracks, improving the performance of flavour tagging by removing PU tracks falsely associated with jets.

4 Analysis Pipeline

The analysis workflow used to evaluate flavour tagging performance consists of several key steps. Initially, A Common Tracking Software (ACTS) is employed for Monte Carlo simulations of $t\bar{t}$ events for the Open Data Detector (ODD). ACTS handles the simulation, tracking, vertexing, and track-to-jet association, although these aspects are not the primary focus of our study. For the ATLAS ITk data, the samples were generated through the ATLAS Athena infrastructure.

The primary objective is to evaluate the performance of a neural network for flavour tagging, both with and without timing information. The NNs are trained to classify jets into different categories, such as light-jets, charm-jets, and bottom-jets. By incorporating timing information, the goal is to improve the accuracy and efficiency of b -jet reconstruction.

A simplified Graph Neural Network (GNN), specifically the DIPS model, is utilized in this analysis. The DIPS model uses observables of tracks associated with a jet to make its predictions. The core idea of the GNN is to use a small neural network to extract information from each track and then combine this information into a single vector, either by summing or averaging, to classify the jets.

4.1 DIPS Model

The DIPS model is based on the Deep Sets architecture [6], utilizing the observables of tracks associated with a jet. The primary concept is to use a small neural network to extract information from each track and then aggregate this information into a single vector, either by summing or averaging, to classify the jets as illustrated in Figure 4.

DIPS can be thought of as a simplified GNN where the input is represented as a graph. Each jet has multiple tracks, and one graph represents one sample (one jet to be classified). Each track associated with a jet is a node in this graph, with up to 40 nodes per graph. If there are fewer than 40 tracks in a jet, a mask is used for those inputs so they do not contribute to the result. This forms a fully connected graph without edge features, where after computing the embeddings, the messages are aggregated via a mean operation.

The DIPS model retains a similar concept to the Message Passing Neural Network from [10], but it simplifies the structure by having only one full message passing layer and no edge features. This simplicity allows it to effectively utilize the track-level information for flavour tagging tasks.

4.2 Inputs to the DIPS Model

The inputs to the DIPS model are derived from various observables associated with the tracks of jets. Key observables include impact parameters such as d_0 and z_0 , which are crucial for identifying b -jets due to their longer decay time, resulting in larger impact parameters on average as depicted in Figure 5. Additionally, track time relative to the hard scatter is used, where tracks with larger times are likely to originate from PU events. The timing distribution shows two peaks: the narrow peak represents tracks from the HS, while the broader peak indicates tracks likely originating from PU. These inputs are derived from ATLAS simulations.

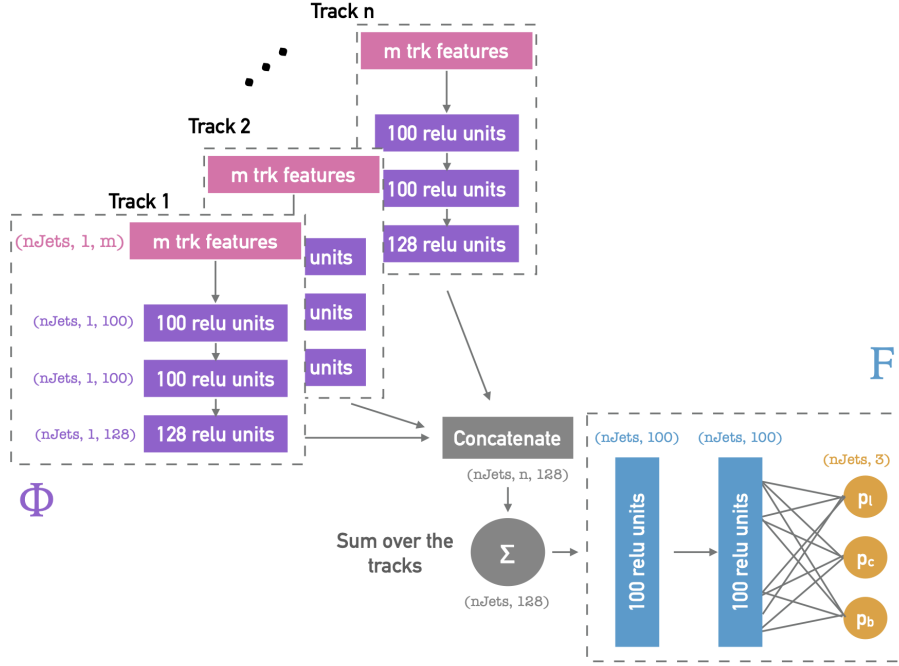


Figure 4: Schematic diagram of the Deep Sets model which consists of one embedding layer and one output layer. The DIPS model follows this structure with an aggregation step for up to 40 tracks per jet. Taken from [6].

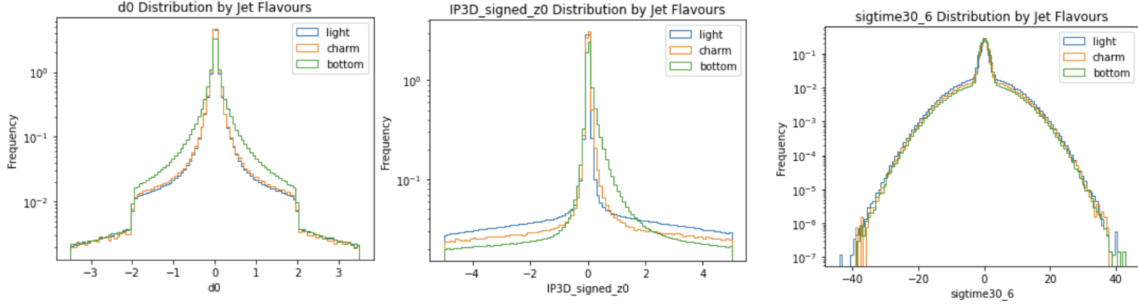


Figure 5: Distributions of impact parameters d_0 and signed z_0 by jet flavours (left) and distribution of signed track time relative to the hard scatter by jet flavours (right) from ATLAS simulations. These are some of the most important variables used in flavour tagging tasks.

Signed variables such as d_0 and z_0 are used. The signed d_0 is defined as (source [11]):

$$\text{sign}_{d_0} = \text{sign}(\sin(\phi_{\text{jet}} - \phi_{\text{trk}}) \cdot d_{0,\text{trk}})$$

and the signed z_0 is defined as (source [11]):

$$\text{sign}_z = \text{sign}((\eta_{\text{jet}} - \eta_{\text{trk}}) \cdot z_{0,\text{trk}})$$

These signed variables help in distinguishing the direction of the displacement, providing additional information that aids in the identification of b -jets.

Other important features include the number of pixel hits (`numPix`), the number of SCT hits (`numSCT`), the transverse momentum fraction (`ptfrac`), and the distance in the η - ϕ plane (`deltaR`).

These features are normalized in different ways to enhance their utility in the model. It is important to note that jet features are not directly used in the DIPS model.

5 Flavour Tagging Performance in ATLAS Detector

In this section, we evaluate how the inclusion of time information can enhance the discriminating power of flavour tagging models and improve background rejection for charm and light-jets. The analysis begins with the evaluation of the discriminant D , which is defined as follows:

$$D = \ln \left(\frac{f_c p_c + (1 - f_c) p_l}{p_b} \right).$$

Here, p_f represents the probability that the NN thinks a jet belongs to flavour f , which is equivalently the value of the output neuron corresponding to a certain flavour. The parameter f_c is a tunable parameter that assigns weight to charm flavour vs light flavour. We used a value of 0.1 because we focused on light-jet rejection mostly. The discriminant combines the outputs of the neural network models, with higher values of D indicating that the model predicts the jet to be a b -jet, while lower values suggest the jet is a c -jet or a light-jet. A successful model will exhibit strong discrimination power, indicated by minimal overlap between the distributions of D for different jet flavours. Figure 6 shows that the outputs for b -jets are generally higher than those for c -jets and light-jets, indicating effective separation by the model.

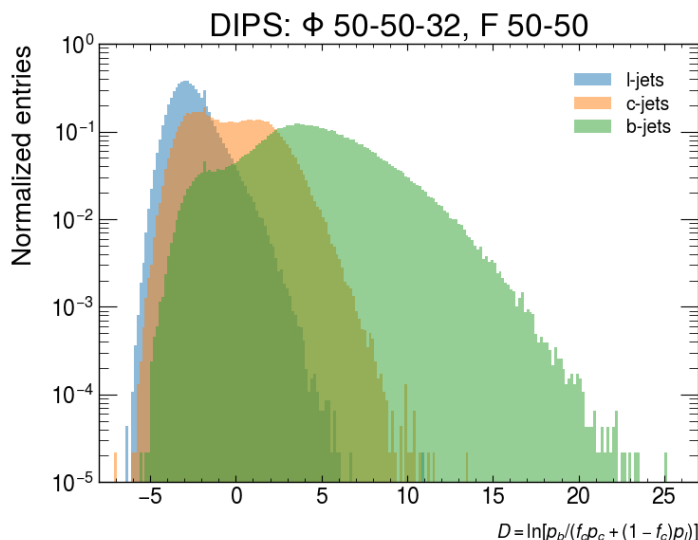


Figure 6: Typical distribution over NN score for flavour tagging. Higher D values indicate the model thinks it is a b -jet, while lower values indicate c -jets or light-jets. Successful models show minimal overlap between distributions for different flavours.

The performance of flavour tagging models is typically assessed using Receiver Operating Characteristic (ROC) curves, which plot the background rejection rate against the b -jet efficiency. Background rejection is defined as the inverse of the false positive rate, indicating how well non- b -jets are correctly identified as such. A curve that extends further to the upper right indicates better performance, with more background jets being rejected at the same b -jet efficiency. The improvement in rejection rates for the same efficiency shows the benefits of incorporating timing information.

The ROC curves in Figure 7 demonstrate that the inclusion of timing information significantly improves the background rejection rate, especially for light-jets. The performance ratio at the bottom

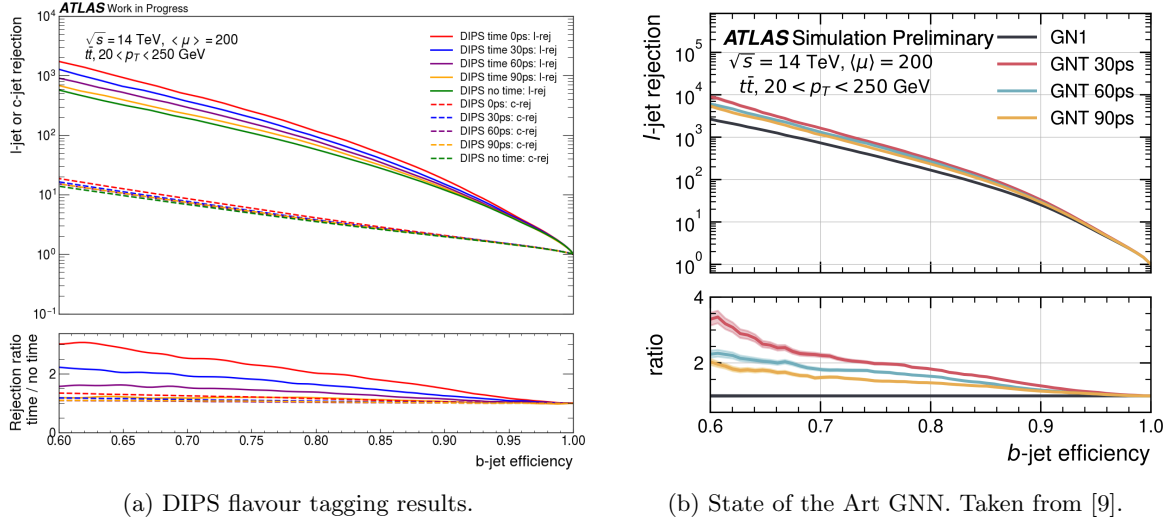


Figure 7: ROC curves comparing the performance of different models. (a) Shows the DIPS flavour tagging results with different time resolutions. (b) Shows the state-of-the-art GNN results currently used for ATLAS flavour tagging, which can be used for comparison.

of each ROC plot further quantifies this improvement. For the DIPS model, the rejection ratio plot shows that timing improves background rejection by up to 3 times for perfect resolution.

Comparing the DIPS model with the state-of-the-art GNN, we observe that the GNN also benefits significantly from the inclusion of timing information. The improvement is even more pronounced for the GNN, as seen in the ratio plot, which could be attributed to the GNN's ability to utilize more complex data and the additional track label used during training. This track label information includes identifying which flavour jet a track is associated with or if it is a pile-up (PU) track associated with a jet or another source. Although predicting track labels is not the primary goal, it helps the model better understand track associations and improves its overall performance. Additionally, the GNN was trained on approximately 50% more statistics than the DIPS model under the same experimental conditions, which could lead to slightly more improvement when including time in a GNN scenario.

Furthermore, the more sophisticated structure of the GNN allows it to exploit timing information more effectively, leading to better discrimination between b -jets and background jets. This comparison highlights the interplay between advanced machine learning techniques and innovative detector technologies such as 4D trackers in enhancing the accuracy of flavour tagging in the challenging environment of the HL-LHC.

6 Flavour Tagging Performance in Open Data Detector

In this section, we evaluate the flavour tagging performance using the Open Data Detector (ODD). Initially, we encountered issues with the data, particularly with simulated $t\bar{t}$ events that had an average pile-up (PU) of 200. The data exhibited spikes in the distributions of track observables, likely due to an indexing issue, which was not fully resolved.

To address this, we implemented a quick fix by removing events that had jets with more than 40 tracks. Some jets in the simulated data had unrealistic counts of thousands of tracks, indicating an issue with the data. After this filtering, the distributions appeared relatively fine, albeit with some outliers, especially in the impact parameter distributions of tracks. Although we could have continued the analysis with cuts on these distributions, we decided to identify and resolve the core issue and

resimulate the data. This resimulation has not yet been completed and remains a part of further work.

Our goal was to perform the same analysis for different time resolutions as conducted with the ATLAS detector data. After training DIPS on the ODD data, we did not observe any statistically significant improvement when including time information at any resolution. This lack of improvement could be attributed to several factors:

1. **Data Quality:** The unresolved indexing issue likely affected the quality of the data, thereby influencing the performance of the flavour tagging models.

2. **Insufficient Statistics:** For the ATLAS data, the DIPS model was trained on approximately 2 million jets. In contrast, the ODD training dataset consisted of only around 200,000 jets. When the ATLAS detector data was trained with a similar number of samples (200,000), only a minor improvement in flavour tagging performance was observed. However, increasing the sample size to 2 million for the ATLAS data resulted in an improvement in performance with timing information by an order of five compared to without.

Given these observations, it is probable that the lack of improvement in the ODD analysis is due to a combination of insufficient training sample statistics and unresolved data quality issues. To conclusively analyze the impact of timing information on flavour tagging performance using the ODD, the following steps are essential:

- Resolve the core issue with the data indexing to ensure high-quality, realistic simulated data.
- Resimulate the data with at least 1 million training samples to provide sufficient statistics for robust model training and evaluation, and train the DIPS model on this data.

By addressing these points, we anticipate that future analyses will yield more definitive and positive results, highlighting the potential benefits of timing detectors in improving flavour tagging performance.

7 Conclusion

This study demonstrated the significant benefits of integrating precise timing information into the ATLAS detector for flavour tagging. The improvements observed with the ATLAS data pave the way for flavour tagging studies with the Open Data Detector (ODD) and other future detectors. Future studies in this direction are essential to extend the algorithms to generic detector layouts and transfer knowledge from the experiments at the HL-LHC to future hadronic machines. These include the hadronic Future Circular Collider (FCC-hh) and other similar collider projects that rely on effective pile-up rejection to maximize their physics reach.

8 Acknowledgments

A big thanks to the members of the DIPactS team, including Lorenzo Santi, Nicole Hartmann, Pierfrancesco Butti, and Valentina Cairo, for their countless advice and guidelines on this extremely exciting summer school project.

References

- [1] ATLAS Collaboration, G. Aad, E. Abat, *et al.*, “The ATLAS Experiment at the CERN Large Hadron Collider,” *Journal of Instrumentation*, vol. 3, no. 08, S08003, Aug. 2008. DOI: 10.1088/1748-0221/3/08/S08003. [Online]. Available: <https://dx.doi.org/10.1088/1748-0221/3/08/S08003>.

- [2] D. Berry, V. Cairo, A. Dragone, *et al.*, “4-Dimensional Trackers,” Tech. Rep., 2022. arXiv: 2203.13900. [Online]. Available: <https://cds.cern.ch/record/2807541>.
- [3] B. Petersen, C. Ohm, A. Sfyrta, *et al.*, “Expected performance of the ATLAS detector under different High-Luminosity LHC conditions,” CERN, Geneva, Tech. Rep., 2021. [Online]. Available: <https://cds.cern.ch/record/2765851>.
- [4] L. Gonella, “The ATLAS ITk detector system for the Phase-II LHC upgrade,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 1045, p. 167 597, 2023, ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2022.167597>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0168900222008890>.
- [5] P. Gessinger-Befurt, A. Salzburger, and J. Niermann, “The Open Data Detector Tracking System,” *Journal of Physics: Conference Series*, vol. 2438, no. 1, p. 012 110, Feb. 2023. DOI: 10.1088/1742-6596/2438/1/012110. [Online]. Available: <https://dx.doi.org/10.1088/1742-6596/2438/1/012110>.
- [6] ATLAS Collaboration, “Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS,” CERN, Geneva, Tech. Rep., 2020. [Online]. Available: <https://cds.cern.ch/record/2718948>.
- [7] ATLAS Collaboration, “Graph Neural Network Jet Flavour Tagging with the ATLAS Detector,” CERN, Geneva, Tech. Rep., 2022. [Online]. Available: <https://cds.cern.ch/record/2811135>.
- [8] X. Ai, C. Allaire, N. Calace, *et al.*, “A Common Tracking Software Project,” *Computing and Software for Big Science*, vol. 6, no. 1, p. 8, Apr. 2022, ISSN: 2510-2044. DOI: 10.1007/s41781-021-00078-8. [Online]. Available: <https://doi.org/10.1007/s41781-021-00078-8>.
- [9] “Investigating the impact of 4D Tracking in ATLAS Beyond Run 4,” CERN, Geneva, Tech. Rep., 2023. [Online]. Available: <https://cds.cern.ch/record/2870326>.
- [10] M. Abdughani, D. Wang, L. Wu, J. M. Yang, and J. Zhao, “Probing the triple Higgs boson coupling with machine learning at the LHC,” *Phys. Rev. D*, vol. 104, p. 056 003, 5 Sep. 2021. DOI: 10.1103/PhysRevD.104.056003. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevD.104.056003>.
- [11] G. Piacquadio, “Identification of b -jets and investigation of the discovery potential of a Higgs boson in the $WH \rightarrow \ell \nu b \bar{b}$ channel with the ATLAS experiment,” Ph.D. dissertation, Freiburg U., 2010. [Online]. Available: <https://cds.cern.ch/record/1243771>.