Dissertation zur Erlangung des akademischen Grades
Dr. rer. nat.

Development of Flavour Tagging Algorithms for Run 3 at LHCb

From High-Energy Physics to Financial Transactions: Automation as a Unifying Principle

Micol Olocco
geboren in Carmagnola, Italien

Fakultät Physik
Technische Universität Dortmund

Dortmund, 2025

This work is part of the **SMARTHEP** project, which has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 956086.

A Nonna Iucci,
Che mi ha spiegato che fare la Ricerca è come cercare le caramelle

A Pl,
Che ha creduto in me in ogni secondo

Der Fakultät Physik der Technischen Universität Dortmund zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften vorgelegte Dissertation.

Gutachter:
Prof. Dr. Johannes Albrecht
Prof. Dr. Chris Malena Delitzsch

Vorsitzender der Prüfungskommission:
Prof. Dr. Heinz Hovel

Vertreter der wissenschaftlichen Mitarbeiter:
Dr. Jörg Debus

Datum der Einreichung der Dissertation:
19/12/2025

Datum der mündlichen Prüfung:
13/03/2026

Abstract

This thesis investigates the role of automation and machine-learning techniques as a methodological bridge between high-energy physics and industry, within the framework of the SMARTHEP European Training Network. The core scientific contribution of this thesis is the development of flavour-tagging algorithms for the LHCb Upgrade I. The performance of the developed algorithms is evaluated using data collected by the LHCb experiment in 2024 and demonstrates stable behaviour in the upgraded detector environment. These algorithms constitute the first flavour-tagging implementation specifically developed for Run 3, enabling time-dependent measurements of neutral B-meson mixing and CP violation. Automation principles are further applied to the development of an infrastructure for the deployment and validation of Trigger Configuration Keys at LHCb, designed to support Run 3 data-taking and operations. Beyond the core LHCb physics programme, this thesis presents the design of an open-source framework for the automated generation of synthetic financial transaction data using large language models. Such synthetic datasets provide a practical alternative for fraud-detection research, where access to real financial data is constrained by privacy and regulatory requirements.

Kurzfassung

Diese Dissertation untersucht die Rolle von Automatisierung und Machine-Learning-Methoden als Brücke zwischen der Hochenergiephysik und der Industrie im Rahmen des SMARTHEP European Training Network. Der zentrale wissenschaftliche Beitrag dieser Arbeit ist die Entwicklung von Flavour-Tagging-Algorithmen im Zuge des Upgrade I des LHCb-Experiments. Die Performanz der entwickelten Algorithmen wird anhand von im Jahr 2024 vom LHCb-Experiment aufgezeichneten Daten evaluiert und zeigt ein stabiles Verhalten für den aufgerüsteten Detektor. Diese Algorithmen stellen die ersten speziell für Run 3 entwickelten Flavour-Tagger dar und ermöglichen zeitabhängige Messungen der Mischung neutraler B-Mesonen sowie der CP-Verletzung. Automatisierungsprinzipien werden zudem bei der Entwicklung der Infrastruktur zur Bereitstellung und Validierung von Trigger Configuration Keys am LHCb-Experiment angewendet. Die Automatisierung ist zur Unterstützung der Datennahme und des Betriebs des Detektors während Run 3 konzipiert. Über das zentrale Physikprogramm des LHCb-Experiments hinaus präsentiert diese Dissertation die Konzeption eines Open-Source-Frameworks zur automatisierten Erzeugung synthetischer Finanztransaktionsdaten mithilfe von Large Language Models. Solche synthetischen Datensätze stellen eine praktikable Alternative für die Forschung zur Betrugserkennung dar, da der Zugang zu echten Finanzdaten durch Datenschutz- und regulatorische Anforderungen stark eingeschränkt ist.

Contents

1 Introduction

The PhD work presented in this thesis has been carried out within the **SMARTHEP Network**, a European Training Network established to bridge research in real-time data analysis between academia and industry [1, 2]. The core motivation of SMARTHEP lies in the strong parallels between the challenges faced by High-Energy Physics (HEP) experiments and those encountered in various industrial sectors. From automotive to finance and manufacturing, many domains must process massive data streams under stringent latency constraints while enabling automated, reliable decision-making. A distinctive feature of this PhD programme is the opportunity to undertake an internship with one of the industrial partners [3], the results of which will also be presented in this thesis. SMARTHEP aims to train researchers capable of operating at the interface between HEP and industry and of transferring methods and technologies across domains. A detailed description of SMARTHEP and its objectives is provided in Ref. [1, 3].

1.1 Data, Artificial Intelligence, Real-Time Analysis

Data has become the new gold of the 21st century. It fuels global economies, accelerates scientific discovery, and powers artificial intelligence (AI). The recent breakthroughs in AI—particularly deep learning and large language models—would likely not have been possible without the unprecedented availability of massive datasets. In this sense, AI is a triumph of data and computational infrastructure as much as it is of mathematical innovation. The scientific importance of AI was formally recognised in 2024, when the Nobel Prize in Physics was awarded to John J. Hopfield and Geoffrey Hinton for their pioneering work on artificial neural networks [4], underlining the transformative impact of data-driven methods across scientific domains. Data can come in different forms: structured tables, images, videos, and text. Each has brought to the development of specialised AI methods capable of handling its complexity. Convolutional neural networks [5] revolutionised image recognition, recurrent neural networks [6] and transformer models reshaped natural language processing [7], while graph neural networks have emerged as a powerful tool for relational and topological data, such as social networks, molecular structures, or particle-interaction graphs [8]. The central challenge, however, remains universal: how to efficiently transform raw data into knowledge.

One organisation that has long stood at the forefront of data challenges is CERN, the European Organisation for Nuclear Research. The World Wide Web (WWW) was invented at CERN in 1989 as a way to share experimental data among physicists around the globe, at a time when collaboration often relied on physically shipping magnetic tapes and hard disks across continents [9]. Today, the Large Hadron Collider (LHC), the world's most powerful particle accelerator located at CERN, produces data volumes that rival, and sometimes exceed, those generated by the largest digital platforms such as YouTube or Google [10]. As illustrated in Figure 1.1, the raw detector readout produced by the LHC in 2018 would correspond to hypothetical data rates of ≈ 40 ZettaBytes (ZB) per year—far beyond what any existing storage infrastructure could handle. Through multi-stage hardware and software selection systems, only a fraction of this data is retained: in 2018, for instance, the LHC experiments stored roughly 160 PB of data, comparable to the annual volumes produced in 2021 by platforms such as YouTube or Facebook. Looking ahead to the High-Luminosity LHC, the total retained data volume is expected to increase by at least a factor of five, reaching approximately 800 PB of stored data per year by 2026 [10]. This dramatic growth promises unprecedented statistical power for precision measurements, while at the same time pushing the limits of what is technically feasible in data acquisition, storage, and large-scale

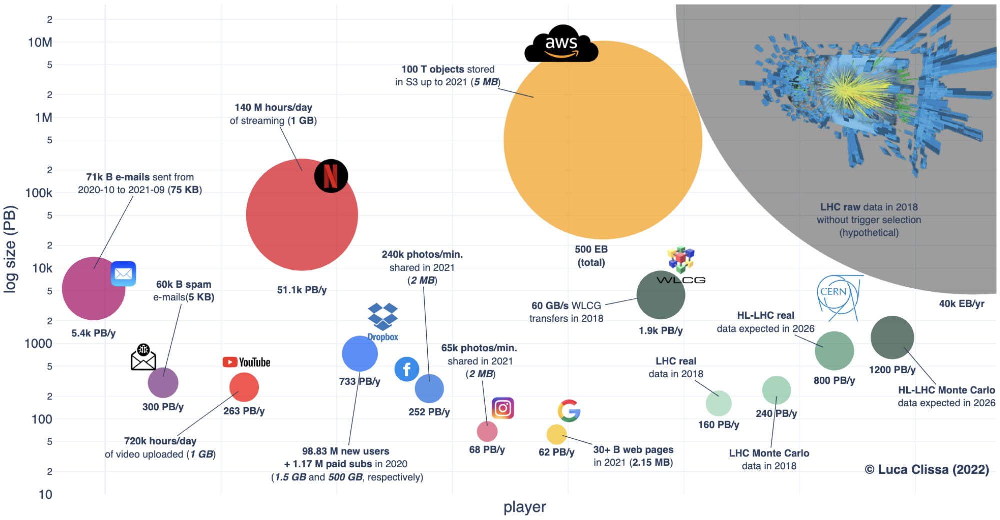


Figure 1.1: Comparison of annual data volumes generated in different domains during 2021 with data collected by the LHC experiments in 2018 (Ref. [10]).

analysis [11].

At the LHC, the challenge of data volume is inseparable from the challenge of *real-time processing.* Deciding, in real-time, which data to keep is a central component of the physics program of the LHC experiments. This task is fulfilled by the trigger and data-acquisition systems: automated processing pipelines that read out all detector components under strict timing constraints. Within only a few microseconds, these systems must convert the raw detector signals, extract physically meaningful quantities, and determine whether a collision—among the roughly 40 million proton–proton bunch crossings occurring each second—produced data that could be scientifically valuable and should be retained for further analysis. Recent advances in computing and machine learning have expanded what is possible in these resource-constrained environments [12]. In many areas of industry—ranging from high-frequency trading and fraud detection to large-scale traffic analysis—massive data streams must also be processed and acted upon under strict latency constraints [13]. The parallels between these domains open the door to cross-disciplinary exchange, where concepts developed in one field can inform and accelerate progress in another.

1.2 Automation as a Unifying Principle

As data volumes grow and analyses become more complex, manual procedures quickly become irreproducible and impossible to scale. In high-energy physics, automation is already essential to detector operations and data processing, yet the HL-LHC environment will further amplify the need for *smart* solutions [14]: increasing autonomy in areas such as early data-quality assessment, real-time monitoring, and fault diagnosis or recovery could further support the already remarkable physics output of the LHC. In industry, automated pipelines form the backbone of large-scale data processing, in part due to the substantial gains in efficiency and cost-effectiveness they provide. These parallels reinforce the broader relevance of automation as a unifying principle across scientific and industrial data ecosystems. This thesis offered the opportunity to apply automation principles in three projects:

- **Development of flavour tagging algorithms for the Run 3 at the LHCb experiment** (core thesis work, Chapters 5, 6). Automation plays a central role in the construction of reproducible and scalable workflows for feature production, particle selections, model training and calibration. This work was carried out under the supervision of Dr. Quentin Führing (TU Dortmund, Lamarr Institute for Machine Learning and Artificial Intelligence) and Dr. Sara Celani (University of Heidelberg).

- **Infrastructure to validate and develop the Trigger Configuration Keys (TCKs) at LHCb** (CERN secondment, Section 3.5.1). Development of a Continuous-Integration test that automatically validates and deploys the TCKs, ensuring smooth trigger operation. This work was conducted in collaboration with Luke Grazette (University of Warwick) and Rosen Matev (CERN).
- **Development of an open-source framework to generate synthetic financial transactions via Large-Language Models** (IBM secondment, Chapter 7). In collaboration with Pierre Feillet (IBM), automation is explored in the context of agent-based simulation by employing large language models (LLM). The produced framework aims to generate realistic financial transaction streams, including fraud scenarios, with strong requirements on temporal consistency. Beyond autonomous data generation, automated fallback mechanisms were developed to repair or recycle imperfect LLM outputs.

Each project addresses a distinct domain and employs different technologies while sharing a common philosophy: automation as a mechanism for ensuring reproducibility, scalability, and reliability in complex data-analysis pipelines.

1.2.1 Thesis Outline

The thesis is structured to provide the theoretical and methodological foundations of the work and to describe the projects undertaken during the PhD.

Chapter 2 introduces the theoretical background relevant to this thesis, including an overview of the Standard Model, *CP* violation, and the machine-learning concepts employed in the subsequent chapters.

Chapter 3 describes the LHCb experiment and its Upgrade I, with emphasis on the detector components most relevant for flavour tagging.

Chapter 4 presents the LHCb flavour tagging principles and the related performance metrics.

Chapter 5 details the development of the Run 3 flavour tagging algorithms, including tagging particle selection, model training, and calibration.

Chapter 6 evaluates the performance of the developed algorithms on a subset of Run 3 data collected by the LHCb experiment in 2024.

Chapter 7 details the framework developed at IBM France Lab for simulating legitimate and fraudulent banking transactions.

Chapter 8 provides an overall conclusion, bringing together the three projects discussed in this thesis and highlighting common methodological themes related to automation and machine-learning techniques.

Part I

Context And Foundations

---

2 Theoretical Framework

The Standard Model (SM) of particle physics provides a remarkably successful description of the known elementary particles and their interactions. At the same time, experimental observations and theoretical considerations point to limitations of the Standard Model description, motivating the exploration of physics beyond its current framework. High experimental sensitivity is achieved through the collection of large data samples produced by particle collisions at high energies. At CERN, the Large Hadron Collider (LHC) provides such collisions at unprecedented centre-of-mass energies and luminosities. The LHCb experiment at CERN is specifically optimised for the study of heavy-flavour hadrons and *CP* violation, where indirect effects of physics beyond the Standard Model (BSM) may manifest as small deviations from SM predictions.

This chapter summarises the theoretical and experimental concepts relevant to the work presented in this thesis. Section 2.1 introduces the Standard Model of particle physics and the Cabibbo–Kobayashi–Maskawa (CKM) matrix, and Section 2.2 outlines the main limitations of the Standard Model. Section 2.3 discusses the phenomenon of neutral B-meson mixing, providing the motivation for flavour tagging. Finally, Section 2.4 presents the theoretical background of the machine learning methods applied in the development of the Run 3 flavour tagging algorithms.

2.1 The Standard Model

The Standard Model [15–17] is a relativistic quantum field theory describing matter and the strong, weak, and electromagnetic interactions. It is based on the local gauge symmetry

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y,$$

where $SU(3)_C$ corresponds to the strong interaction, while $SU(2)_L \otimes U(1)_Y$ governs the electroweak sector.

Matter fields are spin-$\frac{1}{2}$ fermions arranged into three generations of quarks and leptons. Each generation contains an up-type quark (u, c, t), a down-type quark (d, s, b), a charged lepton (e, μ, τ), and a corresponding neutrino $(\nu_e, \nu_\mu, \nu_\tau)$. A defining feature of the Standard Model is the chiral structure of matter fields: left-handed fermions transform as doublets under $SU(2)_L$, while right-handed fermions are singlets, leading to maximal parity violation in charged-current weak processes.

The strong interaction is described by Quantum Chromodynamics (QCD), a non-Abelian $SU(3)_C$ Yang–Mills gauge theory coupled to quark matter fields. One of its emergent properties is colour confinement, whereby quarks are not observed as free particles but instead form colour-neutral bound states known as hadrons, including mesons (quark–antiquark states) and baryons (three-quark states).

In addition to fermions and gauge fields, the Standard Model contains a scalar field, the Higgs field. The electroweak symmetry is preserved at the level of the Lagrangian but is spontaneously broken by the non-zero vacuum expectation value of the Higgs field. Through the Higgs mechanism [18–20], this symmetry breaking generates masses for the weak gauge bosons and fermions. In the broken phase, the electroweak gauge fields recombine to form the photon γ, the massive $W^{\pm}$, Z boson. The full particle content of the Standard Model is summarised in Figure 2.1.

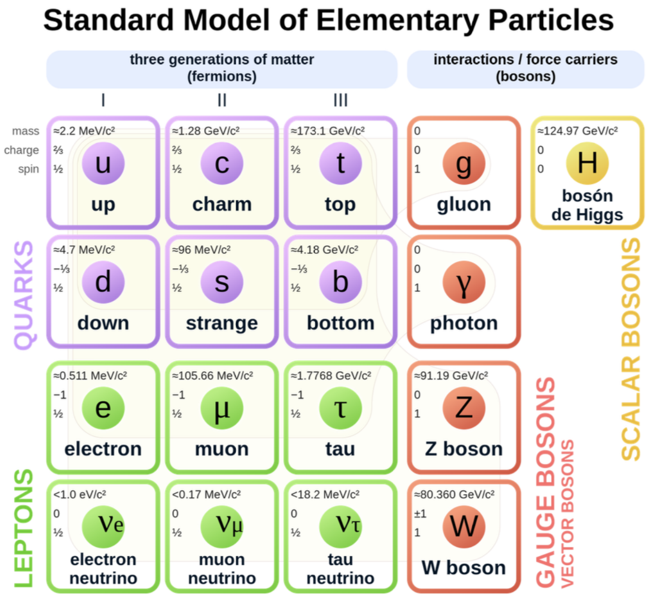


Figure 2.1: The elementary particles of the Standard Model, grouped by family, generation, and interaction type. Figure source [21].

In the Standard Model, the charged-current weak interaction couples up- and down-type quarks, and the misalignment between the mass and weak interaction eigenstates of the quarks gives rise to mixing between generations. By convention, the weak interaction eigenstates of the up-type quarks (u, c, t) are chosen to coincide with their mass eigenstates, so that the mixing is absorbed entirely into the down-type sector. The weak interaction eigenstates of the down-type quarks (d', s', b') differ from their mass eigenstates (d, s, b) and are related by the Cabibbo–Kobayashi–Maskawa (CKM) matrix [22, 23] V_{CKM}:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{\mathrm{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix} . \tag{2.1}$$

The elements V_{ij}, where $i = (u, c, t)$ and $j = (d, s, b)$ label the up- and down-type quarks respectively, give the amplitude for a transition between quark flavours i and j in a charged-current weak interaction. V_{CKM} is a unitary 3×3 matrix, ensuring conservation of probability in weak processes.

The CKM matrix can be parametrised in terms of three real mixing angles θ_{ij} and a single irreducible complex phase δ [24],

$$V_{\mathrm{CKM}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} , \tag{2.2}$$

with $s_{ij} \equiv \sin\theta_{ij}$ and $c_{ij} \equiv \cos\theta_{ij}$.

The Standard Model is invariant under the combined CPT transformation, a fundamental symmetry of local Lorentz-invariant quantum field theories. As a consequence, any observed violation of *CP* symmetry must be accompanied by a corresponding violation of time-reversal symmetry, ensuring overall CPT invariance.

A *CP* transformation is defined as the combined action of charge conjugation (C), which exchanges particles with their antiparticles, and parity (P), which corresponds to a spatial inversion $\vec{x} \to -\vec{x}$. In the Standard Model, the charged-current interaction between quarks and the charged $SU(2)_L$ gauge fields is not invariant under *CP* transformations. This non-invariance arises from the presence of the irreducible complex phase δ in the CKM matrix, such that $V_{\mathrm{CKM}} \neq V^*_{\mathrm{CKM}}$. This phase cannot be removed by field redefinitions and constitutes the unique source of *CP* violation in the quark sector of the Standard Model.

The unitarity of the CKM matrix leads to six independent relations among its elements, defining closed triangles in the complex plane. The angles of this so-called unitarity triangle are defined as

$$\alpha \equiv \arg\left(-\frac{V_{td}V^*_{tb}}{V_{ud}V^*_{ub}}\right) , \qquad \beta \equiv \arg\left(-\frac{V_{cd}V^*_{cb}}{V_{td}V^*_{tb}}\right) , \qquad \gamma \equiv \arg\left(-\frac{V_{ud}V^*_{ub}}{V_{cd}V^*_{cb}}\right) . \tag{2.3}$$

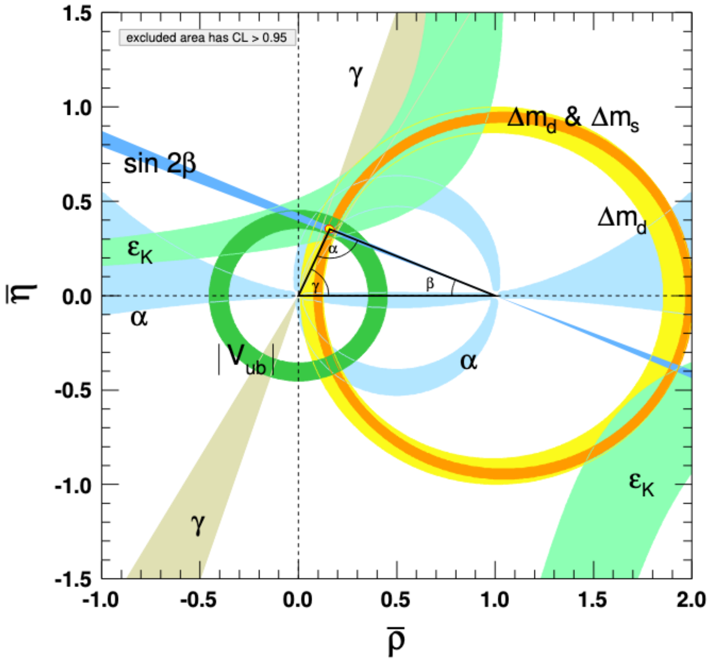


Figure 2.2: Experimental constraints on the CKM unitarity triangle. The coloured regions indicate the 95% confidence level contours resulting from global averages of flavour observables. Each band corresponds to a different process sensitive to CKM parameters, collectively providing a global test of the Standard Model flavour sector. Figure from [25].

Of particular relevance for B-meson physics is

$$V_{ud}V_{ub}^{*} + V_{cd}V_{cb}^{*} + V_{td}V_{tb}^{*} = 0, \tag{2.4}$$

that originates the triangle shown in Figure 2.2, scaled so that its base length is unity. The sides and angles of the resulting unitarity triangle can be accessed through measurements of B-meson decays, mixing, and *CP* violation, and are therefore central to experimental tests of the Standard Model in the heavy-flavour sector.

2.2 Limitations of the Standard Model

Despite its extraordinary success in describing the known elementary particles and the three fundamental forces (strong, weak, and electromagnetic), the Standard Model (SM) is recognised as limited. Several experimental observations cannot be accommodated within its framework, and several theoretical considerations strongly suggest the existence of physics beyond the SM (BSM). Firstly, the SM includes three out of the four fundamental forces in nature as no quantum-field theoretic formulation of gravity compatible with the SM can be trusted at all energy scales.

Astrophysical and cosmological measurements further point to the presence of new phenomena not captured by the SM. Observations such as galaxy rotation curves [26], gravitational lensing [27], and the cosmic microwave background indicate that only about 5% of the energy density of the Universe is composed of SM matter, while approximately 27% corresponds to dark matter and 68% to dark energy. No SM particle possesses the required properties to account for dark matter, and dark energy does not arise from any known SM interaction.

Another clear indication of physics beyond the SM comes from the observation of neutrino oscillations [28], which demonstrate that neutrinos are massive particles. The mixing among neutrino flavours can be described by introducing the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix (analogous to the CKM matrix in the quark sector), but the origin of the neutrino masses remains unclear [29, 30].

The observed dominance of matter over antimatter in the Universe provides another compelling motivation to search for new physics. If the early Universe began in a symmetric state, the emergence of a matter-dominated cosmos requires baryon-number violation, departures from thermal equilibrium, and the presence of *CP* and charge-conjugation violation. Although the SM contains a source of *CP* violation through the CKM phase, its magnitude is insufficient by many orders of magnitude to explain the observed baryon asymmetry. This discrepancy motivates precise studies of *CP* violation in heavy-flavour systems, where contributions from new physics may manifest as small deviations from SM expectations [31–34].

Further theoretical tension arises from the instability of the Higgs boson mass. Quantum corrections to the Higgs mass are sensitive to energy scales up to the Planck scale (10^{19}GeV), and reproducing the measured value of 125 GeV requires finely tuned cancellations between large contributions. This so-called hierarchy problem suggests the existence of new symmetries or mechanisms, such as supersymmetry [35], composite Higgs models [36], or theories with extra spatial dimensions, which stabilise the electroweak scale [37, 38]. Other proposed BSM scenarios introduce entirely new particle species, such as leptoquarks, which could mediate transitions between quarks and leptons and help unify the structure of fermion interactions [39, 40].

Taken together, these challenges indicate that the Standard Model cannot be the final theory of fundamental interactions. They motivate a broad experimental programme to search for new physics, combining direct exploration at the highest available energies with precision measurements sensitive to virtual effects.

The experimental search for new physics at the Large Hadron Collider (LHC) (described in Section 3.2) proceeds through two complementary strategies. Direct searches, primarily carried out by the ATLAS and CMS experiments at the LHC, probe the production of new heavy particles in proton–proton collisions, though their sensitivity is ultimately limited by the collider's centre-of-mass energy and luminosity. Indirect searches exploit precision measurements to look for virtual effects of new heavy particles, which can alter observables even when the particles themselves are too massive to be produced directly. Flavour physics provides a particularly sensitive arena for indirect searches: new physics may alter rare decay rates, neutral-meson mixing amplitudes, or *CP*-violating phases at a level accessible to modern experiments. The design of the LHCb detector is particularly well suited to flavour physics, as it combines excellent vertexing, good momentum resolution, and dedicated particle identification in the kinematic region relevant for heavy-flavour decays.

2.3 *CP* Violation and Neutral Meson Mixing

One of the objectives of the LHCb experiment is to investigate Charge-Parity (CP) violation in the heavy-flavour sector. All experimentally established sources of *CP* violation within the Standard Model originate from the non-zero complex phase δ of the CKM matrix introduced in Section 2.1. In a physical system, *CP*-violating effects can manifest through different mechanisms depending on the interplay between mixing and decay amplitudes. *Direct CP* violation arises from differences in the decay amplitudes of a particle and its antiparticle; *indirect CP* violation is closely related to *neutral meson mixing* or *oscillation*. Direct *CP* violation does not require mixing and can therefore also occur in charged meson decays, where it is the only possible *CP*-violating mechanism.

Mixing refers to the transition of neutral B_q^0 mesons ($q = d, s$) to their antiparticles via weak interaction before decaying. This transition is described at lowest order by the two Feynman box diagrams shown in Figure 2.3, involving two W-boson exchanges and virtual up-type quarks. Owing to the large top-quark mass, the dominant contribution arises from loops containing the top quark [25]. Mixing causes the flavour eigenstates ($B_q^0, \bar{B}_q^0$) to evolve into a linear combination of two mass eigenstates, $B_{q,\mathrm{H}}$ and $B_{q,\mathrm{L}}$, which differ in both mass and decay width [41, 42]. The mass difference $\mathit{\Delta} m_q = m_\mathrm{H} - m_\mathrm{L}$ governs the oscillation frequency, while the decay-width difference $\mathit{\Delta\Gamma}_q = \mathit{\Gamma}_\mathrm{L} - \mathit{\Gamma}_\mathrm{H}$ controls the relative decay rates of the two components in time. The measurements of Δm_q and $\Delta\Gamma_q$ constitute sensitive tests of the flavour sector of the Standard Model. For instance, the measurement of Δm_d and Δm_s allow to constrain the ratio of the CKM elements $|V_{ts}/V_{td}|$ that determines the apex of the unitarity triangle described in Section 2.1 [43]. For flavour-specific decays such as $B_s^0 \to D_s^- \pi^+$, the decay-time distribution in the absence of detector effects can be written as [42]

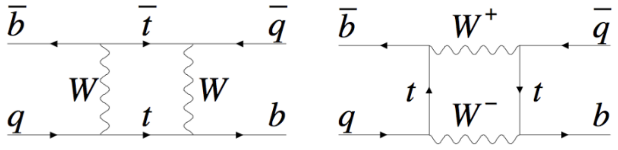


Figure 2.3: Box diagrams contributing to B_q^0–$\bar{B}_q^0$ mixing ($q = d, s$). In the SM, the dominant contribution arises from internal top-quark exchange [25].

$$P(t) \approx e^{-\Gamma_q t}\left[\cosh\left(\frac{\Delta\Gamma_q\, t}{2}\right) + C \cdot \cos(\Delta m_q\, t)\right], \tag{2.5}$$

where $\Gamma_q = (\Gamma_{\mathrm{H}} + \Gamma_{\mathrm{L}})/2$. The parameter C is 1 if the decay is unmixed (*i.e.* $B_s^0 \to D_s^- \pi^+$), -1 if the initially produced meson mixed into its antiparticle before decaying (*i.e.* $B_s^0 \to \overline{B}_s^0 \to D_s^- \pi^+$), depending on whether the meson decays with the same flavour it was produced with or after oscillating into its antiparticle.

Neglecting $\Delta\Gamma_q$ for illustration, Eq. (2.5) gives

$$A_{\mathrm{mix}}(t) = \frac{N_{\mathrm{unmixed}}(t) - N_{\mathrm{mixed}}(t)}{N_{\mathrm{unmixed}}(t) + N_{\mathrm{mixed}}(t)} \simeq \cos(\Delta m_q t),$$

The asymmetry is observable in data if the *initial flavour* of the meson—whether it was produced as a B_q^0 or a $\bar{B}_q^0$—is known: this is the purpose of flavour tagging at LHCb. Without this information, mixed and unmixed decays cannot be separated and the asymmetry would vanish, rendering the extraction of Δm_q impossible.

LHCb currently provides the most precise determination of the B_s^0–$\overline{B}_s^0$ mixing frequency Δm_s [25], based on the analysis of Run 2 data collected between 2015 and 2018. Figure 2.4 illustrates the decay-time distributions of mixed and unmixed $B_s^0 \to D_s^- \pi^+$ decays (left), together with the corresponding time-dependent mixing asymmetry (right). The oscillatory behaviour reflects the $\cos(\Delta m_s t)$ term in Eq. (2.5) and is observable thanks to the use of flavour tagging algorithms.

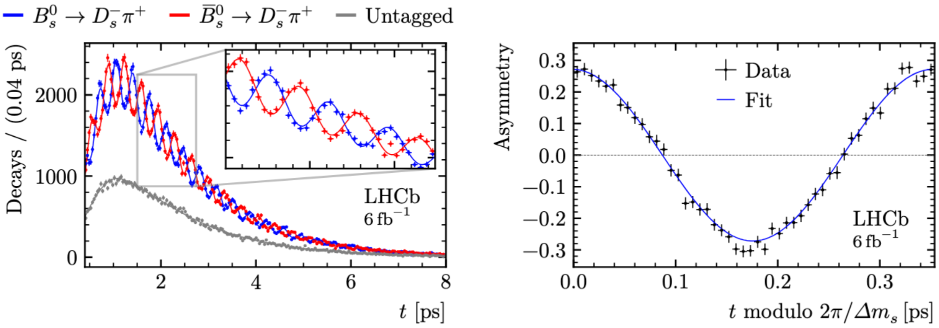


Figure 2.4: Left: decay-time distributions for unmixed (blue) and mixed (red) $B_s^0 \to D_s^- \pi^+$ decays, together with the untagged component (grey). Right: time-dependent mixing asymmetry between unmixed and mixed decays. The oscillation amplitude and phase encode the value of Δm_s. Figure adapted from [42].

The ability to resolve the oscillations experimentally demonstrates both the excellent decay-time resolution of the LHCb detector and the crucial role of flavour tagging in such measurements. The flavour tagging mechanism is discussed in detail in Chapter 4, while the concrete implementations and performance of the flavour tagging algorithms developed for Run 3 are presented in Chapter 5 and Chapter 6.

2.4 Machine Learning Methods

This section introduces the supervised machine learning methods employed to develop the flavour tagging algorithms for Run 3. In particular, decision trees and artificial neural networks are used to identify particles whose properties are correlated with the production flavour of the signal B meson, and to estimate the probability of an incorrect flavour assignment, as described in detail in Chapter 5.

2.4.1 Decision Trees

Decision trees are non–parametric models widely used for classification and regression tasks [44]. They operate by recursively partitioning the input feature space into regions associated with different target classes. Each internal node represents a binary decision based on a threshold applied to one input variable, while each terminal leaf node is associated with a class label or class probability.

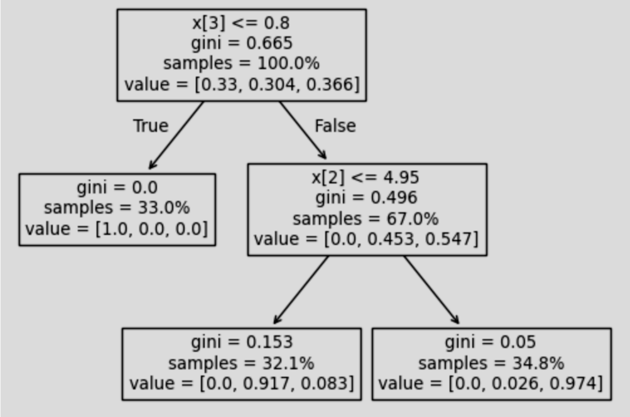


Figure 2.5: Example of a decision tree. Each node displays the Gini impurity, class distribution, and relative sample fractions. Reproduced from the `scikit-learn` documentation [45].

At each node, the splitting variable and threshold are chosen to optimise the Gini impurity criterion [44]. The Gini impurity at node t is defined as

$$G(t) = 1 - \sum_{i=1}^{C} p_i^2,$$

where p_i is the fraction of samples belonging to class i among the C possible classes at that node. Minimising the Gini impurity favours splits that produce child nodes dominated by a single class. The simple example in Figure 2.5 shows how class distributions become progressively purer as the tree branches. An advantage of decision trees is their interpretability: the classification logic can be expressed as a sequence of simple, human–readable selection cuts.

2.4.2 Artificial Neural Networks

Artificial neural networks (ANNs) are flexible models capable of learning complex, non–linear relationships from data [46, 47]. They consist of multiple layers of interconnected neurons, organised into an input layer, one or more hidden layers, and an output layer (Figure 2.6). Each neuron computes a weighted sum of its inputs, adds a bias term, and applies a non–linear activation function to the result. Neurons in one layer are connected to those in the subsequent layer through adjustable weights, which define the strength of the connections.

Neural networks are particularly well suited for flavour tagging applications, where the correlation between the available observables and the production flavour of

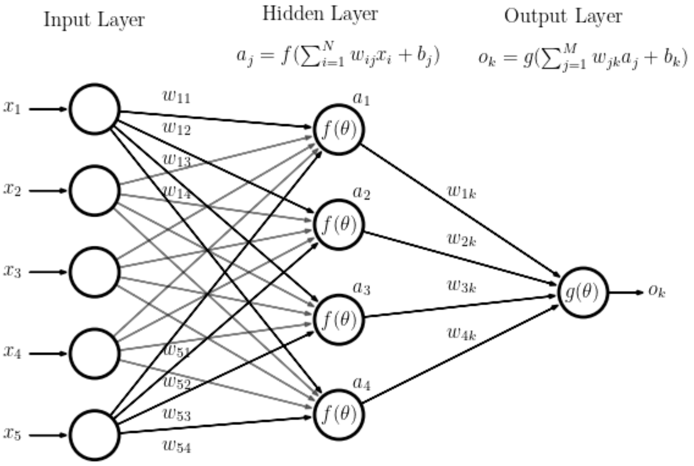


Figure 2.6: Structure of a fully connected neural network. The network maps an input vector $\mathbf{x} = (x_1, \dots, x_N)$ to an output o_k through successive affine transformations and non-linear activation functions. Hidden-layer activations are computed as $a_j = f(\sum_i w_{ij}x_i + b_j)$, while the output is given by $o_k = g(\sum_j w_{jk}a_j + b_k)$. The non-linearities enable the network to model complex, non-linear relationships between input features and target labels. Figure source [48]

the signal B meson is typically weak and distributed across many variables. By combining multiple inputs through successive non–linear transformations, neural networks can model such correlations without imposing a predefined functional form.

The training of a neural network is formulated as an optimisation problem. A loss function is defined to quantify the discrepancy between the network output and the true target labels. During training, the values of the network parameters—namely the weights and biases associated with the connections between neurons—are adjusted to minimise the loss function on a training sample. This optimisation is performed using the backpropagation algorithm [49], which computes the gradients of the loss function with respect to the network parameters and propagates them backwards from the output layer to the input layer. The parameters are then updated iteratively using a gradient-based optimisation procedure until the loss converges. For binary classification, neural networks are trained using the binary cross–entropy loss,

$$\mathcal{L}_{\mathrm{BCE}} = -\left[y \log \hat{y} + (1-y)\log(1-\hat{y})\right],$$

where $y \in \{0, 1\}$ is the true label and $\hat{y} \in (0, 1)$ is the predicted probability. This loss function naturally encourages probabilistic outputs, which are essential for mistag–probability estimation in flavour tagging applications. To prevent overfitting, early stopping is employed as a regularisation technique [47]. Training is monitored on an independent validation sample, and the model corresponding to the minimum validation loss is retained. This procedure ensures robust generalisation while avoiding overfitting and unnecessary computation. The performance of the classifiers is evaluated using the area under the receiver operating characteristic curve (ROC AUC) [50]. The ROC curve represents the trade–off between true–positive and false–positive rates as the classification threshold is varied, while the AUC provides a threshold–independent measure of separation power.

3 The Upgraded LHCb Detector

3.1 Introduction

This chapter provides an overview of the LHCb detector and its Upgrade I, placing particular emphasis on aspects that are most relevant for flavour tagging. The chapter is structured as follows. Section 3.2 introduces the LHC, Section 3.3 describes the LHCb experiment and its unique forward geometry optimised for heavy-flavour physics, while Section 3.4 discusses the motivations and design of the LHCb Upgrade I, together with its subdetectors. The trigger system is illustrated in Section 3.5, and Section 3.6 introduces the Monte Carlo simulation framework employed in LHCb analyses.

Two specific contributions of this PhD work are documented in the chapter. The first is the development of an infrastructure for the automation of the validation and deployment of the Trigger Configuration Keys (TCKs), which became a key component of LHCb's Run 3 operations (Section 3.5.1). The second investigates how varying the amount of information retained by trigger lines relevant for flavour tagging affects the event size (Section 3.5.2).

3.2 The Large Hadron Collider

The Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN) was designed to explore the high-energy frontier of particle physics, enabling stringent tests of the SM and direct searches for new phenomena [51]. It is a circular proton–proton collider with a circumference of 27 km, located about 100 m underground at the Franco–Swiss border near Geneva.

Protons (and lead ions) are accelerated in several stages of the CERN accelerator complex before being injected into the LHC ring, where two counter-rotating beams circulate in opposite directions. The beams are guided and focused by superconducting dipole and quadrupole magnets cooled to 1.9 K with superfluid helium. At designated interaction points, the beams are brought into collision at a frequency of 40 MHz, corresponding to the 25 ns bunch spacing of the LHC.

The performance of the collider is quantified by its luminosity. For colliding proton bunches, the instantaneous luminosity $L(t)$ depends on the beam parameters according to [52]:

$$L = \frac{N_b^2\, n_b\, f_{\mathrm{rev}}\, F}{4\pi\, \sigma_x \sigma_y}, \tag{3.1}$$

with N_b the number of protons per bunch, n_b the number of bunches per beam, f_{rev} the revolution frequency, $\sigma_{x,y}$ the transverse beam sizes at the interaction point, and F a geometric luminosity reduction factor accounting for the crossing angle at the interaction point. The integrated luminosity is then obtained as the time integral of $L(t)$:

$$\mathcal{L}_{\mathrm{int}} = \int L(t)\, dt \tag{3.2}$$

Expressed in units of inverse femtobarns (fb^{-1}), it is related to the number of observed events N of a process with cross-section σ:

$$N = \sigma \times \mathcal{L}_{\mathrm{int}} \tag{3.3}$$

Since its first operation in 2009, the LHC has collided protons at centre-of-mass energies up to 13.6 TeV (Run 3), producing instantaneous luminosities of the order of $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. Over its programme, the machine is expected to deliver hundreds of fb^{-1} of integrated luminosity to each of the main experiments [51, 53].

There are four primary collision points at the LHC, each hosting a major detector. ATLAS and CMS are general-purpose experiments optimised for high-p_T physics and played a central role in the discovery of the Higgs boson [54, 55]. ALICE is dedicated to the study of heavy-ion collisions and the properties of the quark–gluon plasma [56], while LHCb specialises in heavy-flavour physics, as described in Section 3.3. In addition to these flagship experiments, CERN hosts a wide range of other facilities and experiments across the accelerator complex (Figure 3.1), covering fixed-target, antimatter, and accelerator-physics programmes [57].

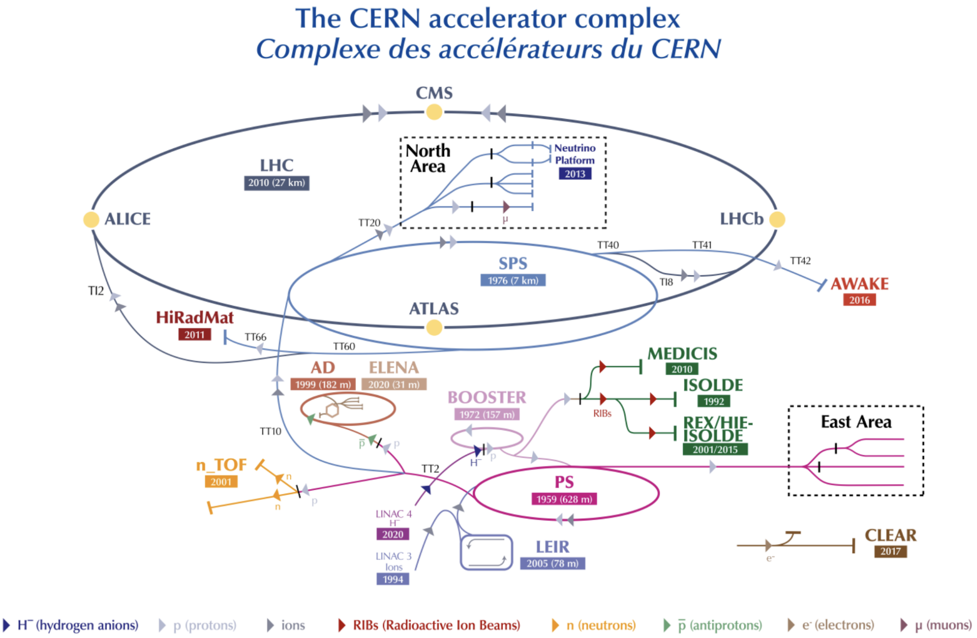


Figure 3.1: The CERN accelerator complex. Protons are pre-accelerated through LINAC, the Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS) and then are injected into the Large Hadron Collider (LHC). The four main experiments –ATLAS, CMS, ALICE, and LHCb– are located at the LHC interaction points, while many other facilities operate throughout the complex. Figure adapted from CERN [57].

3.3 The LHCb Experiment

The Large Hadron Collider beauty (LHCb) experiment at CERN is primarily dedicated to the study of heavy-flavour quark physics, in particular the bottom (b) and charm (c) quarks. The experiment is carried out by an international collaboration of about 1,700 physicists, engineers, and technicians from nearly 105 institutions in 25 countries [58]. Beyond its primary mission, the LHCb experiment has developed a broad and versatile physics programme, including the study of very rare decays [59], electroweak processes [60], heavy-ion and fixed-target physics [61–63][1].

LHCb is designed as a single-arm forward spectrometer, covering the pseudorapidity range $2 < \eta < 5$ [65]. This forward acceptance exploits the kinematic characteristics of pp collisions at the LHC, where $b\bar{b}$ and $c\bar{c}$ pairs are predominantly produced at small polar angles relative to the beamline. The asymmetry in the incoming parton momenta in the laboratory frame, together with gluon–gluon fusion and quark–antiquark annihilation as the dominant production mechanisms (Figure 3.2), leads to a forward–backward production of heavy-flavour hadrons [21].

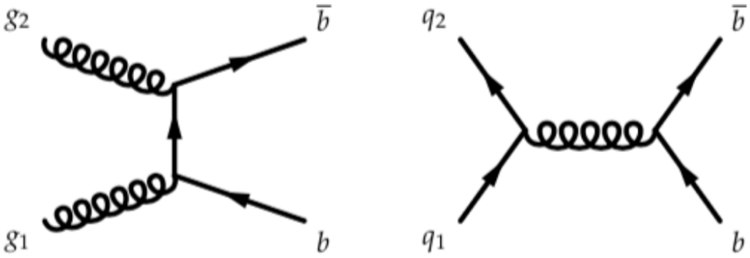


Figure 3.2: Feynman diagrams for the production of a pair of $b\bar{b}$ quarks in a pp collision at the LHC. Left: gluon–gluon fusion. Right: quark–antiquark annihilation. Figure from [21].

The spectrometer acceptance corresponds to angles of 10–300 mrad in the horizontal plane and 10–250 mrad in the vertical plane [21]. Figure 3.3 (left) shows the azimuthal distribution of $b\bar{b}$ pairs, with the LHCb acceptance highlighted, while the right panel compares the coverage with the central acceptance of ATLAS and CMS. Within this geometry, LHCb is exposed to approximately 20–25% of all $b\bar{b}$ pairs produced [64].

---

[1] A broad overview can be found in [64], which reviews the history and physics scope of the LHCb experiment.

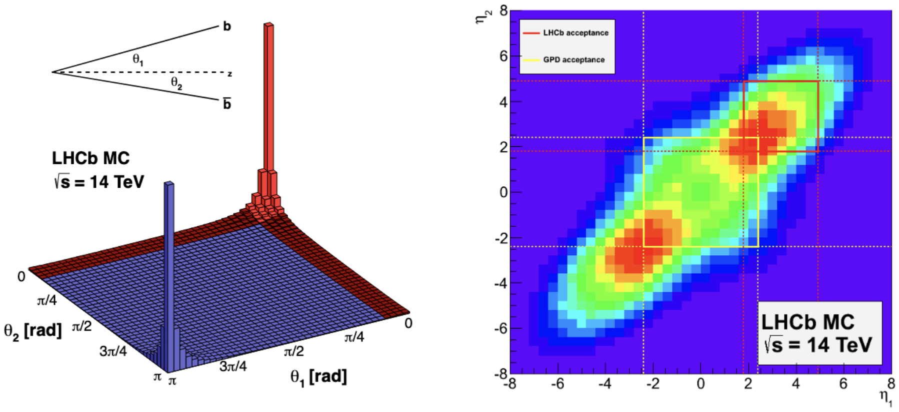


Figure 3.3: Left: azimuthal angle distribution of $b\bar{b}$ quark pairs. The red band corresponds to the geometrical acceptance of the LHCb detector [66]. Right: comparison of the geometrical acceptance of produced $b\bar{b}$ pairs between LHCb and the general-purpose detectors (ATLAS and CMS) [66].

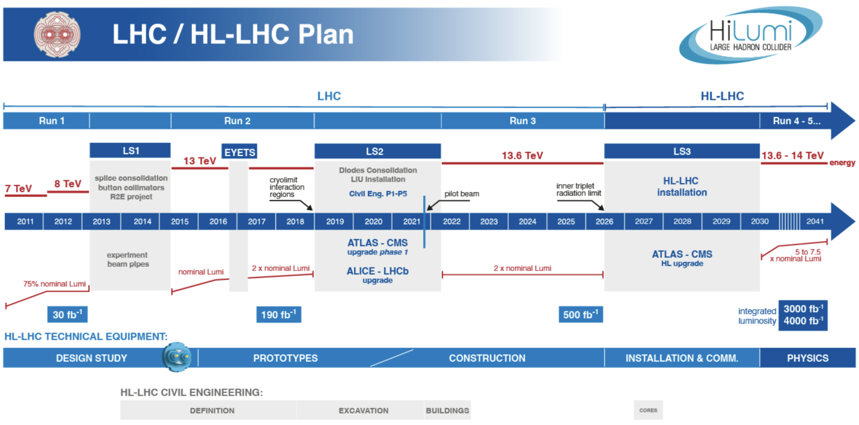
The production cross-section for $pp \to b\bar{b}X$ within the LHCb acceptance at $\sqrt{s} = 13$ TeV has been measured to be (154 ± 14) μb [67], providing exceptionally large event samples for precision flavour physics. The expected number of events containing a $b\bar{b}$ pair in the LHCb pseudorapidity range is given by

$$N_{b\bar{b}} = \sigma(pp \to b\bar{b}X) \times \mathcal{L}_{\text{int}}, \tag{3.4}$$

where $\sigma(pp \to b\bar{b}X)$ is the production cross-section and $\mathcal{L}_{\text{int}}$ the integrated luminosity.

3.4 The LHCb Upgrade I

The LHCb detector was originally designed to operate at an instantaneous luminosity of $2 \times 10^{32}\,\text{cm}^{-2}\text{s}^{-1}$, corresponding to an average of approximately one visible proton–proton interaction per bunch crossing ($\mu \simeq 1$) [65]. During Run 1 and Run 2 (2010–2018), the experiment operated successfully at luminosities significantly exceeding this design value, reaching up to $4 \times 10^{32}\,\text{cm}^{-2}\text{s}^{-1}$. Further gains in experimental sensitivity are expected from the High-Luminosity LHC (HL-LHC) programme [11], which aims to increase the collision rate by a factor of approximately 5–7.5 with respect to the original LHC design [68]. To fully exploit these conditions, all four major LHC experiments have planned extensive detector upgrades, as illustrated in Figure 3.4.



LHC/ HL-LHC Plan (last update January 2025)

Figure 3.4: Timeline of the LHC and HL-LHC programme. LHCb Upgrade I is designed to operate during Run 3 and Run 4, targeting an integrated dataset of about 50 fb^{-1}. A further Upgrade II is foreseen for Run 5 and beyond, to sustain performance in the much higher luminosity environment of the HL-LHC [69]. Figure source [53].

At LHCb, the detector upgrade strategy is implemented in two stages. The Upgrade I, installed during Long Shutdown 2, is designed to operate throughout Run 3 and Run 4, with a target integrated luminosity of approximately $50\,\text{fb}^{-1}$. A further Upgrade II is foreseen for Run 5 and beyond and aims to enable operation at the higher luminosities expected from the High Luminosity LHC (HL-LHC) [69]. In the following, an overview of the LHCb Upgrade I is presented.

The primary limitation of the Run 1 and Run 2 LHCb detector design was the hardware-based Level-0 (L0) trigger. While this system was well suited to the initial operating conditions, it would have saturated at higher luminosities, particularly for hadronic decay channels. As illustrated in Figure 3.5, trigger yields for muonic channels would scale approximately linearly with luminosity, whereas those for hadronic channels would saturate in the Run 1–2 configuration. This limitation would have prevented full exploitation of the increased luminosity delivered by the LHC and constituted a major motivation for a revised trigger strategy. Consequently, a transition to a fully software-based trigger was adopted for Run 3, as formalised in the Upgrade I Letter of Intent [70] and detailed in the Framework Technical Design Report [65].

The upgraded detector is designed to operate at a nominal instantaneous luminosity of $2 \times 10^{33}\,\text{cm}^{-2}\,\text{s}^{-1}$ and to record events at the LHC bunch-crossing rate of 40 MHz.

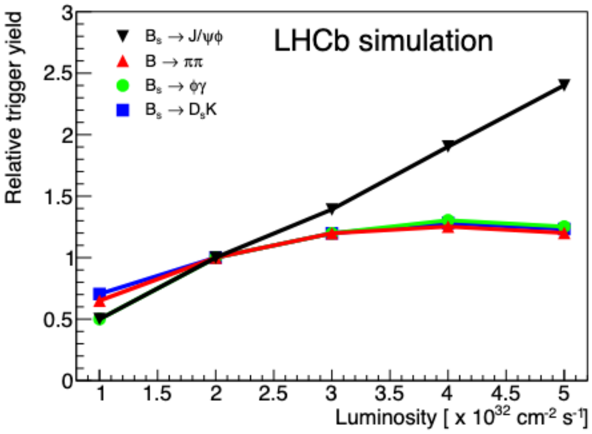


Figure 3.5: Relative trigger yields from simulation as a function of instantaneous luminosity for several benchmark B-meson decay channels. In the Run 1 and Run 2 design, the hardware-based L0 trigger would have led to a saturation of hadronic yields at high luminosities, motivating the transition to the fully software-based trigger of Upgrade I. Figure source [65].

Achieving this required a substantial renewal of the detector subsystems, enabling a 40 MHz readout, increased granularity, and improved radiation tolerance [65]. In parallel, the trigger strategy was redefined to support a *real-time analysis* paradigm, in which detector calibration and alignment are performed online, allowing event selection to be based on reconstructed physics objects within a software trigger.

The year 2024 marked the first full year of operation with the upgraded detector and proved remarkably successful. As illustrated in Figure 3.6, LHCb recorded nearly $9.6\,\mathrm{fb}^{-1}$ of data at a centre-of-mass energy of $\sqrt{s} = 13.6$ TeV, already comparable to the total luminosity accumulated during Run 1 and Run 2 over nearly a decade. Data taking continued at a similar pace in 2025, with more than $11.8\,\mathrm{fb}^{-1}$ recorded by early October. This achievement clearly highlights the increase in physics potential unlocked by the Upgrade I.

The layout of the LHCb Upgrade I detector is shown in Figure 3.7, with the subdetectors arranged along the beamline. Closest to the interaction point, the silicon-pixel Vertex Locator (VELO) provides precise measurements of track positions near the pp collisions, enabling accurate reconstruction of primary and secondary vertices. Upstream of the dipole magnet with a bending power of $4\,\mathrm{T{\cdot}m}$, the Upstream Tracker (UT) provides information to improve the momentum resolution and to reconstruct charged particles before they are deflected by the magnetic field. Downstream of the

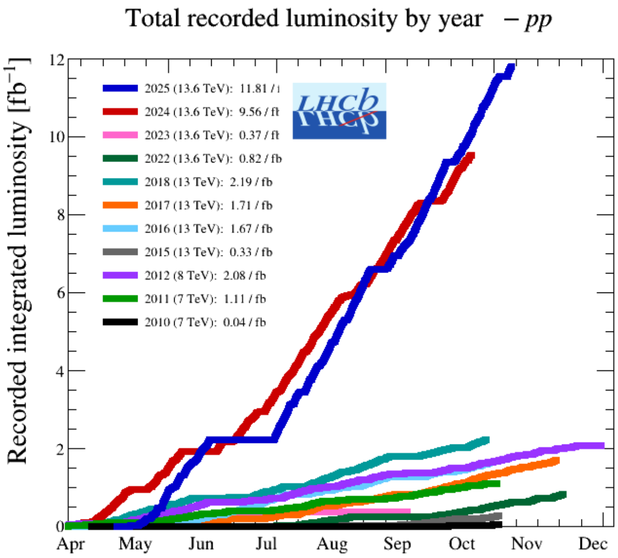


Figure 3.6: Integrated recorded luminosity by LHCb in 2024 and 2025 compared to previous years. Total recorded luminosity since 2011: 31.7 fb^{-1} [71].

magnet, three stations of scintillating fiber detectors, collectively referred to as the Scintillating Fibre (SciFi) tracker, deliver high-granularity tracking information over a large acceptance. Charged hadron identification over a wide momentum range is achieved using two Ring-Imaging Cherenkov (RICH) detectors. The electromagnetic (ECAL) and hadronic (HCAL) calorimeters measure the energies of photons, electrons, and hadrons, while the muon system, composed of alternating layers of iron and multiwire proportional chambers, provides muon identification and completes the detector setup.

In the following sections, the individual subsystems of the LHCb Upgrade I detector are described, including the upgraded trigger system discussed in Section 3.5.

3.4.1 Tracking Detectors

The Vertex Locator

The primary role of the Vertex Locator (VELO) is to reconstruct primary vertices (PVs) from *pp* collisions and to identify secondary vertices (SVs) from the decays of long-lived beauty and charm hadrons. A key observable is the impact parameter (IP), defined as the distance of closest approach between a reconstructed track

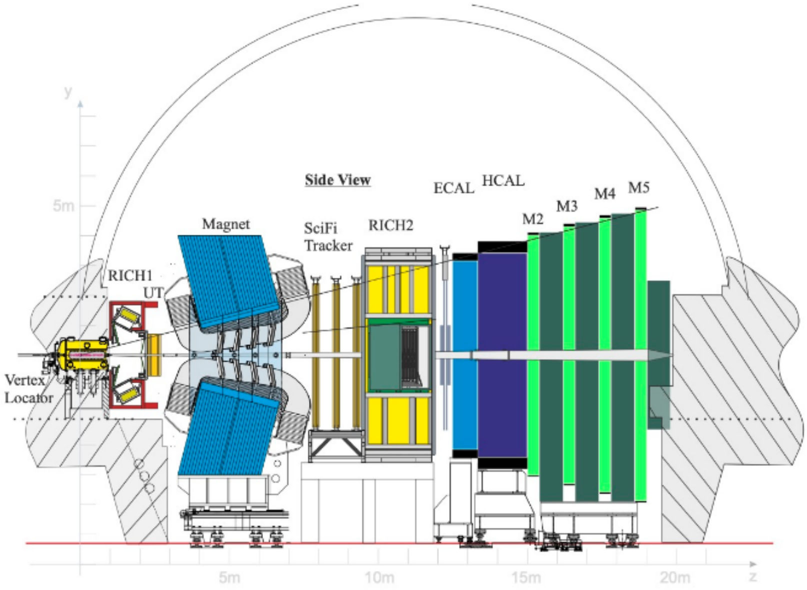


Figure 3.7: Layout of the LHCb Upgrade I detector, showing the main subdetectors arranged along the beamline: the Vertex Locator (VELO), the first Ring-Imaging Cherenkov (RICH1), detector, the Upstream Tracker (UT), the dipole magnet, the Scintillating Fibre (SciFi) tracker, the second Ring-Imaging Cherenkov (RICH2) detector, the calorimeter system (ECAL, HCAL), and the muon stations.

and the nearest PV. Tracks originating from SVs typically have large IP values, making them a primary signature of b- and c-hadron decays. The VELO is a hybrid pixel detector consisting of 52 L-shaped modules, each segmented into $55 \times 55\ \mu\text{m}^2$ pixels [72]. The detector is split into two retractable halves: these are opened during beam injection and closed to within 5.1 mm of the beam line during stable running, 3.1 mm closer than in Run 2. Together with the finer pixel granularity, this leads to a significant improvement in both vertexing and IP resolution. Two standard performance benchmarks are the PV resolution as a function of the number of tracks, N_{tracks}, and the IP resolution as a function of the inverse transverse momentum p_T. Figure 3.8 shows the PV resolution (left) and IP resolution (right) measured during Run 1 and Run 2. The IP resolution scales approximately linearly with $1/p_T$, as expected from multiple scattering. At high p_T, multiple scattering becomes negligible, and the resolution is limited by the PV reconstruction, which itself improves with the number of tracks forming the vertex.

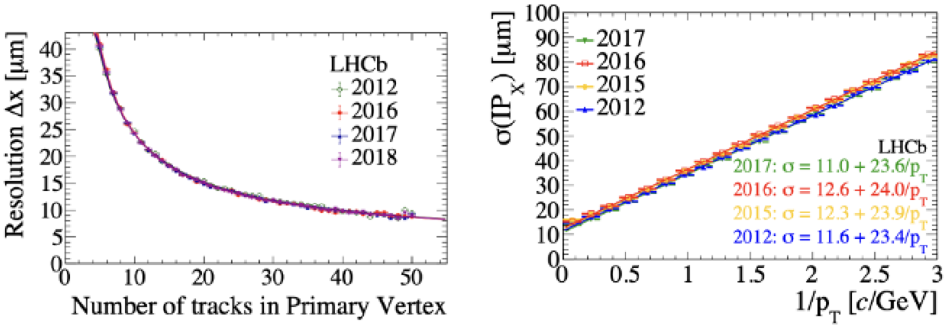


Figure 3.8: Left: primary vertex resolution along x in Run 1 and Run 2 data taking period as a function of the number of tracks composing the vertex. **Right**: impact parameter resolution along the x axis as a function of inverse transverse momentum $1/p_T$ measured on Run 1 and Run 2 data [73].

The IP resolution can be parametrised as [74]:

$$\sigma_{\text{IP}}^2 = \frac{r_1^2}{p_T^2}\left(0.0136\ \text{GeV}/c \cdot \sqrt{\frac{x}{X_0}}\left(1 + 0.038 \ln \frac{x}{X_0}\right)\right)^2 + \frac{\varDelta_{02}^2 \sigma_1^2 + \varDelta_{01}^2 \sigma_2^2}{\varDelta_{12}^2} + \sigma_{\text{extrap}}^2, \tag{3.5}$$

where r_1 is the distance form the beamline of the first hit, p_T is the track transverse momentum, x/X_0 the fractional radiation length before the second hit, $\sigma_{1,2}$ the uncertainties of the first and second hits, and $\varDelta_{ij}$ the distances between vertex and hits. The final term σ_{extrap}^2 accounts for further extrapolation uncertainties. This expression highlights the main drivers of the VELO design: minimising the distance of the first hit from the beam line (r_1), reducing material to limit multiple scattering (x/X_0), and providing multiple closely spaced measurements to improve extrapolation (Δ_{ij}).

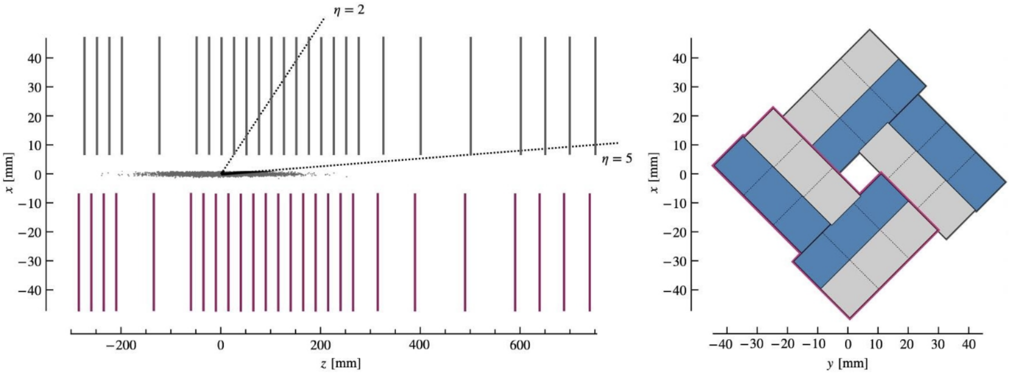


Figure 3.9: Left: schematic top view of the z–x plane at $y = 0$, showing the luminous region and the LHCb pseudorapidity acceptance, $2 < \eta < 5$. Right: schematic layout of the pixel modules in the closed VELO configuration [75].

The vertex resolution σ_{PV} directly impacts the decay time resolution, which is crucial for resolving the fast oscillations of the B_s^0 meson that mixes with a frequency $\Delta m_s \simeq 17.7$ ps^{-1}, a measurement where flavour tagging is crucial. For a Gaussian decay time resolution σ_t, the dilution of the oscillation amplitude D is given by [74]

$$D = exp(-\frac{\sigma_t^2 \Delta m^2}{2}) \tag{3.6}$$

With the upgraded VELO, the PV resolution reaches $\sigma_{\mathrm{PV}} = (11.0 + 13.1/p_T)\,\mu$m, enabling decay time resolutions of $\sigma_t = (43.4 \pm 1.6)$ fs for $B_s^0 \to \phi\phi$ and $\sigma_t = (35.3 \pm 0.3)$ fs for $B^0 \to K^{*0}\mu^+\mu^-$ [74].

The Upstream Tracker and SciFi Tracker

Beyond the VELO, the tracking system is completed by the Upstream Tracker (UT), located before the dipole magnet, and the Scintillating Fibre (SciFi) tracker downstream of the magnet.

The Upstream Tracker (UT) consists of four silicon-strip layers located upstream of the dipole magnet. It provides precise position measurements prior to the bending in the magnetic field, thereby improving the momentum resolution and facilitating the association of tracks reconstructed in the Vertex Locator (VELO) with those measured in the downstream Scintillating Fibre (SciFi) tracker. In addition, the

UT contributes to the discrimination of tracks originating from different primary vertices in the high-multiplicity environment of Run 3. The four UT layers are arranged in the characteristic x–u–v–x geometry, illustrated in Figure 3.10. The first and last layers are oriented vertically (x layers), providing precise measurements in the horizontal direction. The two intermediate layers (u and v) are tilted by $\pm 5^{\circ}$ with respect to the vertical. While a single strip layer measures only the coordinate perpendicular to its strip orientation, the combination of vertical and tilted layers enables the reconstruction of both the horizontal and vertical positions of charged particles.

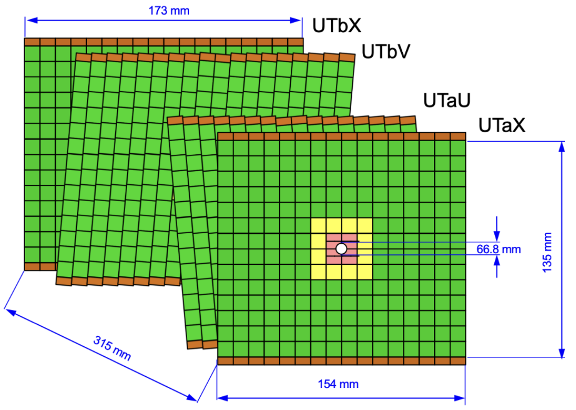


Figure 3.10: Layout of the four layers of the Upstream Tracker (UT), arranged in the x–u–v–x geometry. The x layers (vertical strips) provide horizontal measurements, while the u and v layers are tilted by $\pm 5^{\circ}$. This arrangement provides complementary information that improves the determination of both the horizontal and vertical positions of a particle.

The SciFi tracker covers about 340 m^2 and consists of scintillating fibres with a hit resolution of about 70 μm and per-layer efficiencies above 97% [76]. Together with the VELO and UT, it provides precise momentum measurements, obtained from the curvature induced by the dipole magnet between upstream and downstream tracking stations[2].

The combined tracking system reconstructs several categories of tracks, defined according to which subdetectors are traversed, as illustrated in Figure 3.11. Tracks that include hits in the VELO, UT, and SciFi detectors are classified as long tracks. These provide the most precise momentum determination and are therefore used in

---

[2] For a charged particle moving in a uniform magnetic field, the momentum is related to the radius of curvature R by: $p = qBR$ with q = electric charge, B = magnetic field strength, p = particle momentum.

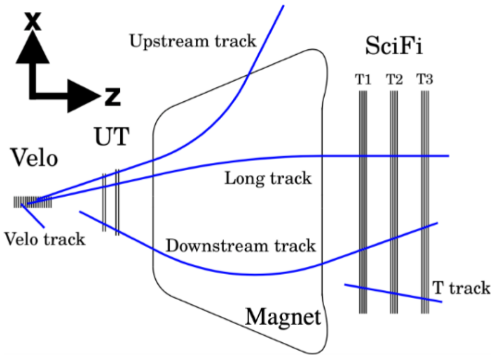


Figure 3.11: Track classification in LHCb according to the subdetectors in which they leave hits. VELO tracks: VELO only; Upstream tracks: VELO+UT; Long tracks: VELO+UT+SciFi; Downstream tracks: UT+SciFi; T tracks: SciFi only.

the majority of physics analyses. Upstream tracks are reconstructed from VELO and UT hits only and typically correspond to low-momentum charged hadrons that do not reach the SciFi tracker. Downstream tracks are formed from UT and SciFi hits and originate from particles produced outside the VELO acceptance, such as decay products of long-lived strange hadrons. In addition, T tracks are standalone track segments reconstructed exclusively in the SciFi tracker, while VELO tracks consist of hits in the VELO only.

The upgraded tracking system achieves a relative momentum resolution of approximately 0.5% at low momentum, increasing to about 1% at high momentum, as shown in Figure 3.12. The momentum resolution for long tracks in data is determined using $J/\psi \to \mu^+\mu^-$ decays. Under the assumption that the two muons have similar momenta and that the muon masses can be neglected, the relative momentum resolution σ_p/p can be approximated as

$$\frac{\sigma_p^2}{p^2} = 2\left(\frac{\sigma_m}{m}\right)^2 - 2\left(\frac{\sin\theta\,\sigma_\theta}{2 - 2\cos\theta}\right)^2, \tag{3.7}$$

where m is the reconstructed invariant mass of the J/ψ candidate and σ_m is the corresponding signal width obtained from a fit to the mass distribution. The second term accounts for the effect of the opening angle θ between the two muons, with σ_θ denoting the per-event uncertainty on θ as obtained from the individual track fits [77].

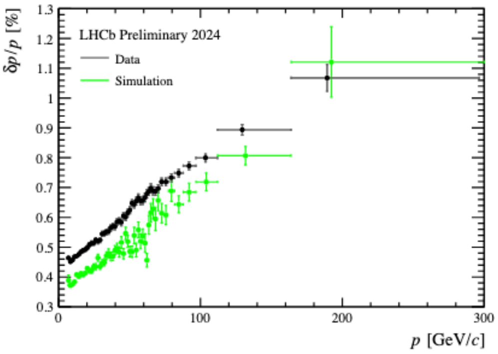


Figure 3.12: Relative momentum resolution of long tracks as a function of momentum. Black: measurement from $J/\psi \to \mu^+\mu^-$ 2024 data; Green: simulation sample. Figure from [77].

3.4.2 Particle Identification Detectors

Particle identification (PID) at LHCb is provided by three dedicated systems: the Ring Imaging Cherenkov (RICH) detectors, the calorimeter (ECAL and HACL) system, and the muon stations.

Ring Imaging Cherenkov Detectors

The RICH detectors, composed of multi-anode photomultiplier tubes (MaPMTs), play a central role in the separation of charged hadrons (π, K, p). They exploit the Cherenkov effect: a charged particle traversing a medium with refractive index n at a velocity $v > c/n$ emits photons on the surface of a cone with opening angle θ, given by

$$\cos\theta = \frac{1}{\beta n}, \qquad \beta = v/c. \tag{3.8}$$

Since the particle momentum p is provided by the tracking system, the combination of p and β, inferred from the measurement of the Cherenkov angle, allows the particle mass m to be determined and hence the particle species to be identified. The dependence of the Cherenkov angle on particle momentum for different species is shown in Figure 3.13, illustrating the separation power of the RICH system.

LHCb employs two RICH detectors using different radiator materials, providing particle-identification coverage across a broad momentum spectrum:

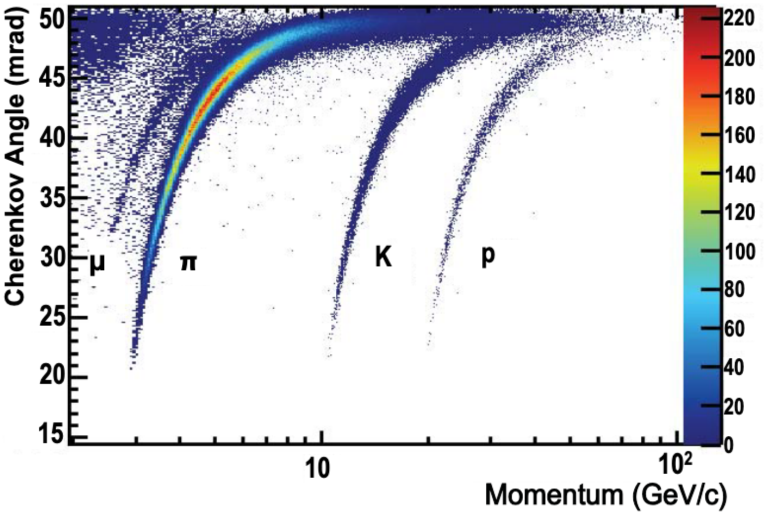


Figure 3.13: Distribution of the Cherenkov angle as a function of momentum for different particle species (muons, pions, kaons, and protons). The separation of the bands illustrates the particle identification capability of the RICH detectors across the momentum range.

- **RICH1**, located upstream of the magnet, is filled with C_4F_{10} ($n \simeq 1.0014$) and provides hadron separation in the range 2–60 GeV/c.
- **RICH2**, installed downstream of the tracking stations, is filled with CF_4 ($n \simeq 1.0005$) and extends PID coverage between 15 and 100 GeV/c.

Cherenkov photons are reflected and focused by mirrors onto arrays of photodetectors, where they are detected and used to reconstruct the Cherenkov angle. The combination of two radiators ensures effective $\pi/K/p$ separation over the full momentum range relevant for LHCb physics, including the overlap region between the two detectors [65].

Calorimeter System

The calorimeter system provides measurements of the energies of electrons, photons, and hadrons, and contributes significantly to particle identification at LHCb. It is composed of two main subsystems: the **Electromagnetic Calorimeter (ECAL)**, optimised for electrons and photons, and the **Hadronic Calorimeter (HCAL)**, designed to measure hadronic energy deposits.

When a high-energy particle interacts with the calorimeter absorber material, it initiates a cascade of secondary particles, referred to as a *shower*. The calorimeters are sampling devices consisting of alternating layers of absorber and scintillator: the absorber material induces the shower, while the scintillator layers sample it, emitting light proportional to the deposited energy. This light is collected by photomultiplier tubes, enabling the reconstruction of the particle energy. Different particle species exhibit distinct shower characteristics [78, 79]:

- **Electrons and photons** produce compact electromagnetic showers dominated by bremsstrahlung radiation and pair production, which are largely contained within the ECAL.
- **Hadrons** undergo nuclear interactions in addition to ionisation losses, resulting in broader and deeper hadronic showers that extend into the HCAL.
- **Muons** lose only small amounts of energy through ionisation and traverse the calorimeter system as minimum-ionising particles (MIPs), leaving characteristic straight-line signatures.

By exploiting these distinct shower profiles, the calorimeter system provides particle-identification information that complements the tracking and RICH detectors. The characteristic detector responses of different particle species are illustrated in Figure 3.14.

The ECAL itself is a lead/scintillator sampling calorimeter located downstream of RICH2. Its performance, measured with electron test beams, is parametrised by:

$$\frac{\sigma_E}{E} = \frac{(9.0 \pm 0.5)\%}{\sqrt{E}} \oplus (0.8 \pm 0.2)\% \oplus \frac{0.003}{E \sin\theta}, \tag{3.9}$$

The constant term accounts for effects such as calibration uncertainties, non-linearities, and energy leakage, while the final term reflects the contribution from electronic noise [65].

The HCAL consists of alternating layers of scintillating tiles and iron absorbers and is used to measure hadronic energy deposits. Its energy resolution, determined from pion test-beam measurements, is parameterised as

$$\frac{\sigma_E}{E} = \frac{(67 \pm 5)\%}{\sqrt{E}} \oplus (9 \pm 2)\%, \tag{3.10}$$

where E is the deposited energy in GeV [65].

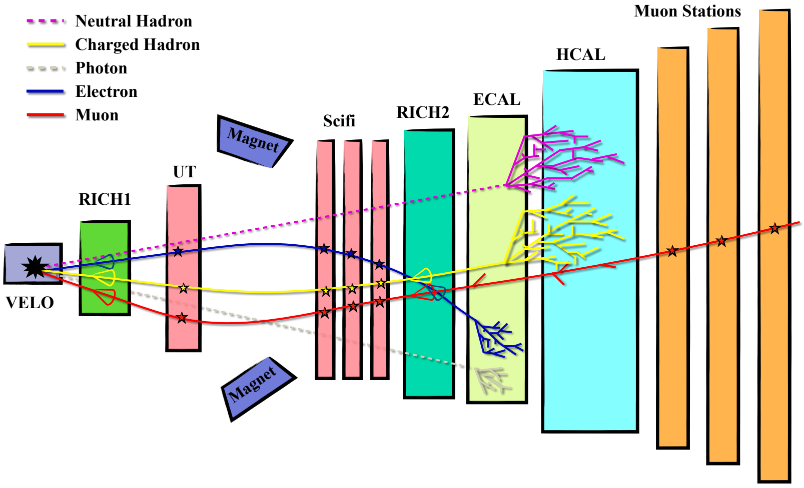


Figure 3.14: Schematic illustration of the LHCb detector response to different particle species. Charged particles are tracked in the VELO, UT, and SciFi, with the dipole magnet providing momentum measurement. The RICH detectors enable hadron separation via Cherenkov radiation. Photons and electrons produce showers in the ECAL, while HCAL maximise the energy loss for charged and neutral hadrons. Muons traverse the detector with minimal energy deposit and are identified in the muon stations. Figure adapted from [78].

Muon System

Muons are capable of traversing the entire LHCb detector with minimal energy loss and without being absorbed. To exploit this property, four muon stations (M2–M5) are installed downstream of the hadronic calorimeter. Each station consists of multi-wire proportional chambers and is separated by approximately 80 cm-thick iron absorbers, which serve to range out hadrons and other particles [65]. Only muons with momenta above 6 GeV/c are able to penetrate all absorber layers, ensuring a high-purity muon sample.

Each muon station is subdivided into four regions (R1–R4) with varying granularity. The segmentation is finest near the beam pipe, where particle occupancies are highest, and becomes progressively coarser at larger radii. This layout provides an effective compromise between spatial resolution and the overall number of readout channels [72]. Owing to their minimal interactions with detector material and their distinctive penetration signature, muons can be identified with high efficiency and purity. As a result, muonic final states play a central role in the trigger strategy and provide high-quality control samples for calibration and flavour-tagging studies.

Particle Identification Variables

Information from the tracking system, the RICH detectors, the calorimeters, and the muon stations is combined to assign each reconstructed track a set of particle-identification (PID) likelihoods under different mass hypotheses.

For each particle type h, a likelihood $L(h)$ is computed by combining the subdetector responses. As an example, the total likelihood that a track corresponds to a pion is defined as

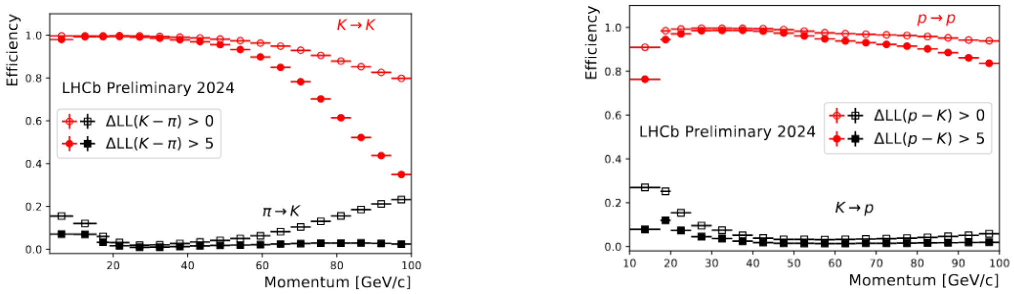
$$L(\pi) = L_{\text{RICH}}(\pi) \times L_{\text{CALO}}(!e) \times L_{\text{MUON}}(!\mu), \tag{3.11}$$

where $L_{\text{RICH}}(\pi)$ is the likelihood from the RICH system, and $L_{\text{CALO}}(!e)$ and $L_{\text{MUON}}(!\mu)$ represent the probabilities that the same track is not identified as an electron or a muon, respectively.

Since pions are the most abundant charged particles at LHCb, particle identification is usually expressed as a log-likelihood difference ($\mathit{\Delta}$LL) with respect to the pion hypothesis [80]:

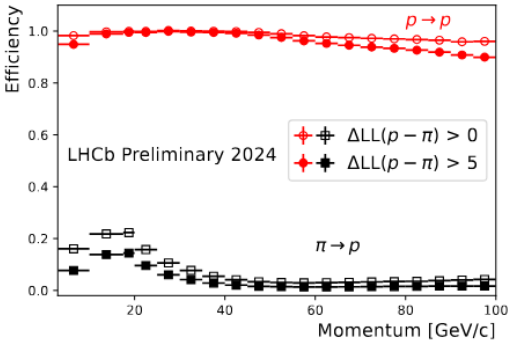
$$\text{PID}(h) = \mathit{\Delta}\text{LL} = \ln L(h) - \ln L(\pi) = \ln \frac{L(h)}{L(\pi)}, \tag{3.12}$$

where $h = K, \mu, e, p$. These $\mathit{\Delta}$LL variables provide powerful separation between particle species, but treat each subdetector independently. Figure 3.15 shows the



(a) Kaon PID performance during 2024 data taking. Kaon efficiency ($K \to K$) and misidentification ($\pi \to K$) are shown as a function of momentum, for two PID cuts $\Delta\text{LL}(K-\pi) > 0$ and > 5.

(b) Proton PID performance during 2024 data taking. Proton efficiency ($p \to p$) and misidentification ($K \to p$) as a function of momentum, for two PID cuts $\Delta\text{LL}(p-K) > 0$ and > 5.



(c) Proton PID performance during 2024 data taking. Proton efficiency ($p \to p$) and misidentification ($\pi \to p$) as a function of momentum, for two PID cuts $\Delta\text{LL}(p-\pi) > 0$ and > 5.

Figure 3.15: PID performances in terms of efficiency and misidentification for kaons, protons, and pions in 2024 data taking. The curves illustrate the trade-off between correct identification and background suppression for different ΔLL thresholds. Figures from [81].

kaon, proton, pion PID performance in terms of identification and misidentification using early Run 3 data recorded in 2024. To exploit correlations between detector responses, LHCb also employs a machine-learning-based PID variable, known as ProbNN [82]. This variable is derived from a neural network trained on simulated samples and uses as input the information from the tracking system, RICH detectors, calorimeters, and muon chambers. The neural network is used to identify a given track as a π, K, p, e, μ particle, providing improved discrimination compared to the traditional ΔLL variables as illustrated in Figure 3.16. These PROBNN variables play an important role in flavour tagging, as discussed in Chapter 5.

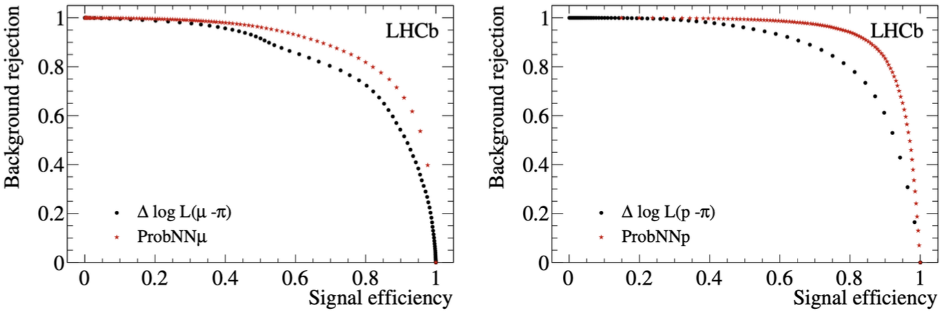


Figure 3.16: Background rejection rates as a function of muon (left) and proton (right) identification efficiency. The variables ΔLL and ProbNN are plotted respectively in black and red for comparison, using 2012 data-taking samples [82].

3.5 Trigger and Online Processing

At the LHC, proton-proton bunches cross at a rate of 40 MHz. For comparison, this rate is about 40 million times faster than a human heartbeat at rest ($\sim$1 Hz). Each pp collision produces signals in thousands of detector channels, resulting in a raw data rate that far exceeds what can be permanently stored. To illustrate the scale of this challenge, in 2018 the LHC experiments collectively generated of the order of 4×10^4 ExaBytes (EB) of raw data prior to online selection and reduction [10]. At the current level of experimental sensitivity and theoretical understanding, however, only a small fraction of these events are expected to contain signatures of physics beyond the Standard Model (BSM), such as the processes discussed in Section 2.2. As a consequence, a fundamental requirement of any high-energy physics experiment is the ability to identify and retain only those events that are potentially relevant for physics analyses, where “interesting” is defined by the observables and processes targeted by the experimental programme.

The system responsible for this real-time event selection is referred to as the *trigger*. Trigger selections can be classified as *exclusive*, when they target specific final-state topologies, or *inclusive*, when they require only the presence of one or more generic final-state objects. During Run 1, the LHCb trigger strategy was predominantly inclusive, similar to that of the general-purpose experiments ATLAS and CMS. In Run 2, the trigger remained largely inclusive, with the notable exception of charm-hadron triggers, for which the event rates became prohibitively high [83].

Studies performed in preparation for Upgrade I [83] estimated that, at Run 3 luminosities, approximately 2% of collisions would contain a reconstructible b hadron, while about 24% would contain a reconstructible c hadron. By contrast, the sustainable output rate of the trigger system was estimated to be of the order of 0.05% of the total collision rate. It was therefore not feasible to record all events containing heavy-flavour hadrons. These considerations motivated a significant evolution of the LHCb trigger strategy towards a primarily *exclusive* approach [83]. In this paradigm, online event reconstruction plays a central role in identifying heavy-flavour decays and selecting the subset of events most relevant to the experiment's physics objectives.

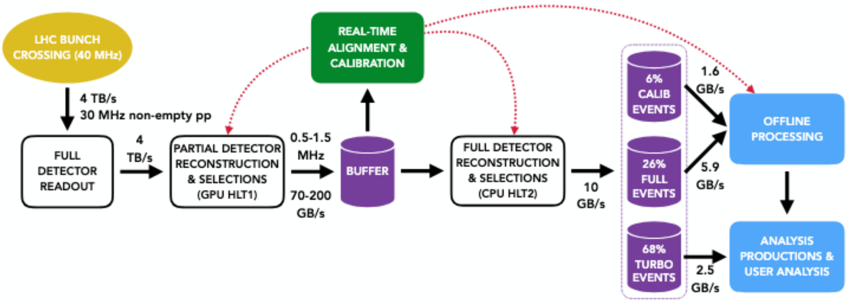


Figure 3.17: Schematic overview of the LHCb Upgrade I online processing chain [84]. All detectors are read out at 40 MHz. HLT1 (GPU-based fast reconstruction) reduces the visible interaction rate to ∼1 MHz. Events are then buffered on disk, where real-time alignment and calibration constants are computed and fed back into the online reconstruction. HLT2 (CPU-based full reconstruction) applies the complete selection menu and writes ∼10 kHz to permanent storage, divided into the FULL, Turbo, and TurCal streams.

The overall Run 3 data flow is shown schematically in Figure 3.17 [84]. At each LHC bunch crossing, all subdetectors are read out and the events are passed to the High-

Level Trigger (HLT), which operates in two software stages with an intermediate buffering step [21, 65]:

- **HLT1**: a GPU-accelerated partial reconstruction stage that performs fast tracking and vertexing. It reduces the input rate from ≈ 30 MHz of visible collisions to 0.5–1.5 MHz, applying $\mathcal{O}(100)$ selection lines while retaining high efficiency for heavy-flavour signatures.
- **Buffer and real-time alignment/calibration**: surviving events are written to a ∼40 PB disk buffer. This buffer provides sufficient latency to compute detector alignment and calibration constants. A sampled subset of events is continuously analysed, producing updated constants within minutes; these are then fed back to the trigger. HLT1 parameters are updated as needed, and HLT2 consistently uses the latest values to ensure stable offline-quality reconstruction.
- **HLT2**: an offline CPU-based full event reconstruction, applying $\mathcal{O}(1000)$ selection algorithms tuned to specific topologies. HLT2 reduces the output bandwidth to 10 GB/s. The amount of information stored per event is determined by the **Turbo paradigm**, described in Section 3.5.2.

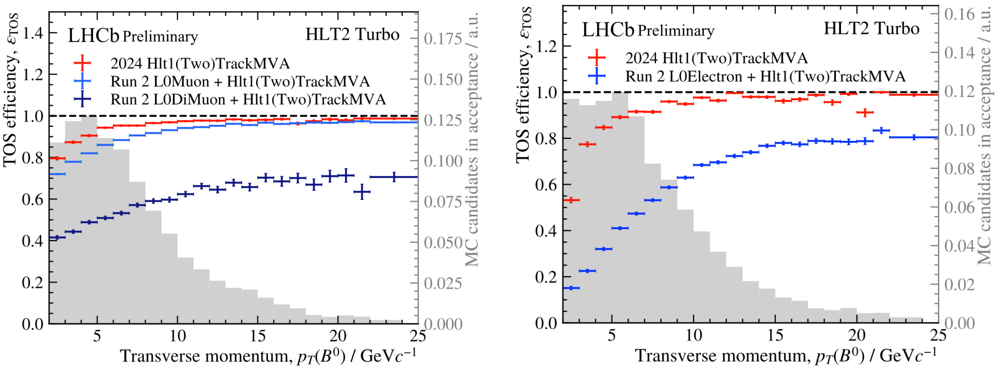
The Trigger-On-Signal (TOS) efficiency quantifies the probability that a given trigger condition selects events in which the targeted signal is genuinely present. It is commonly used as a metric to evaluate the performance of the trigger. Figure 3.18 reports the TOS efficiencies for the channels $B^+ \to J/\psi K^+$, $B^0 \to J/\psi K^{*0}$, and $B^0 \to D^-\pi^+$, which are of particular importance for flavour tagging as they provide the samples used for training and calibrating the taggers. The plots compare the efficiencies of the LHCb HLT1 trigger (specifically of the `HLT1(Two)TrackMVA` trigger lines[3]) in pp collision data taken by the LHCb detector in 2024 to the trigger efficiencies of equivalent Run 2 lines (combined efficiencies of L0 and HLT1) [85]. The removal of the L0 stage in Upgrade I results in substantially improved efficiency at low transverse momentum.

3.5.1 Trigger Configuration Keys (TCKs)

Each configuration of the LHCb trigger is uniquely identified by a Trigger Configuration Key (TCK). The TCK encodes the sequence of algorithms, their parameters, and the control flow defining the trigger at a given point in time. This mechanism is essential for reproducibility: every event stored by LHCb can be associated

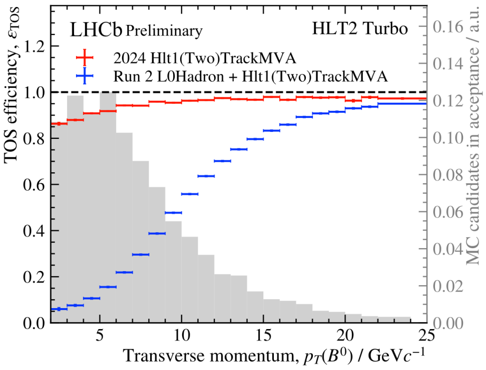
---

[3] The `Hlt1(Two)TrackMVA` lines use kinematic and geometrical selections to identify events containing a single (or a pair of) displaced track(s) consistent with originating from a heavy-flavour decay.



(a) $B^+ \to J/\psi K^+$

(b) $B^0 \to J/\psi K^{*0}$

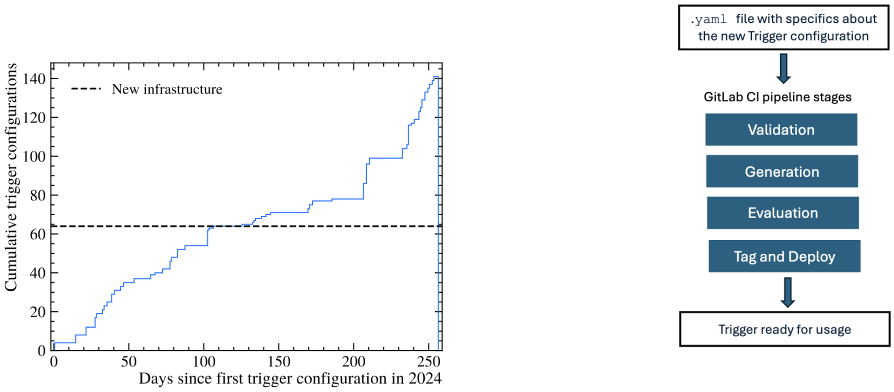


(c) $B^0 \to D^- \pi^+$

Figure 3.18: Efficiencies of the LHCb HLT1 trigger (namely of the `HLT1(Two)TrackMVA` trigger lines) in pp collision data taken by the LHCb detector in 2024 (red). Direct comparisons to the equivalent trigger efficiencies in Run 2 (combined efficiencies of L0 and HLT1) are established (blue). Generator-level MC distributions within the LHCb acceptance are shown in grey. The Upgrade I trigger achieves nearly flat efficiency across the full p_T range, with notable improvements at low p_T where Run 2 suffered losses. This gain is especially important for flavour tagging, since many tagging particles are produced at low transverse momentum.

with the TCK active during its selection, ensuring that the corresponding trigger configuration can be retrieved for later offline analysis.

During Run 1 and Run 2, generating new trigger configurations was a manual process that required following detailed instructions. With the transition to a fully software-based trigger in Upgrade I, and the increasing frequency with which new configurations must be prepared, a more scalable solution was required. Part of this PhD work has been the development of a new infrastructure to automate the validation and deployment of TCKs.



(a) Cumulative TCKs deployed in 2024. The dashed line marks the transition to the new automated infrastructure, which deployed over 70 additional configurations thereafter.

(b) Schematic of the TCK production pipeline. User specifications (YAML) are processed through GitLab CI stages (validation, generation, evaluation) before being tagged and deployed.

Figure 3.19: Overview of the automated validation of the Trigger Configuration Keys (TCKs) TCKs [86].

The process, described in [86], is implemented as a GitLab continuous integration (CI) pipeline. User specifications are provided in YAML format, which is processed through multiple CI pipeline stages (validation, generation, evaluation, tagging, and deployment). If successful, the new TCK is tagged and deployed into production, becoming available for use in data taking (Figure 3.19b). Figure 3.19a shows the cumulative number of trigger configurations deployed in 2024. The new infrastructure became the default in mid-2024, after which dozens of new configurations were automatically validated and deployed, ensuring smooth trigger operation throughout the year.

3.5.2 Turbo Model

The *Turbo* model [87–89], first introduced during Run 2, revolutionised the LHCb data processing strategy by enabling *selective persistence* of reconstructed objects. This approach allows the experiment to store varying levels of information for each event, ranging from the minimal subset required for a specific physics analysis to the complete reconstructed event. Figure 3.20 illustrates the different persistence levels available within the LHCb trigger model.

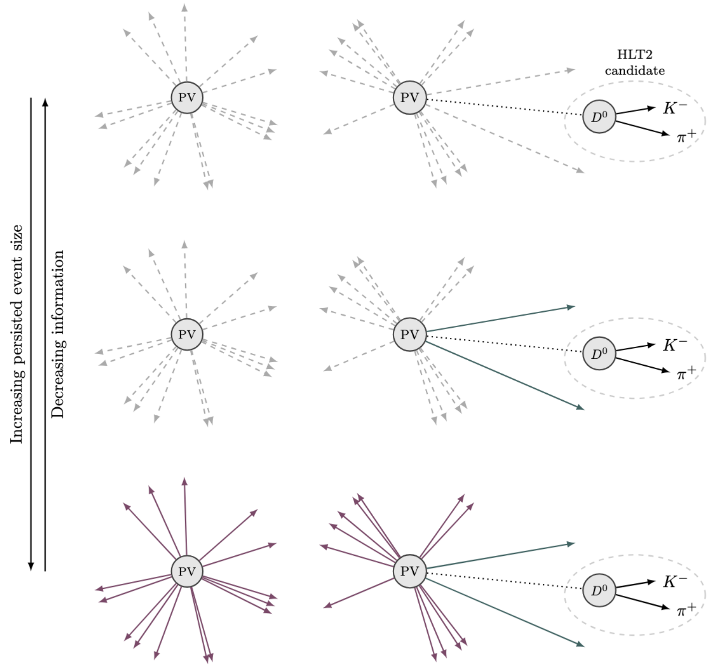


Figure 3.20: Illustration of the LHCb trigger persistence model. The same reconstructed event can be saved with varying levels of persistence: **Top:** *Turbo persistence*, where only the decay candidate selected by the HLT2 trigger (e.g. $D^0 \to K^-\pi^+$) and the corresponding Primary Vertex (PV) are persisted. **Middle:** *Selective persistence*, where additional reconstructed objects (e.g. the soft pion from $D^{*+} \to D^0\pi^+$) are stored. **Bottom:** *Complete reconstruction persistence*, where the full event is saved, including raw subdetector data banks. Solid lines and objects indicate persisted information in each case. Figure adapted from [90].

In Run 3, the vast majority of events are stored using the Turbo format, while a smaller fraction is retained in FULL format to allow for exploratory or future analyses. A dedicated calibration stream, known as *TurCal*, is also maintained to enable precise detector and physics calibration studies. Figure 3.21 illustrates the relative contributions of the Turbo, FULL, and TurCal streams to the Run 3 output. While the majority of events are recorded in the Turbo format, FULL events dominate the overall data volume written to tape and disk due to the larger size.

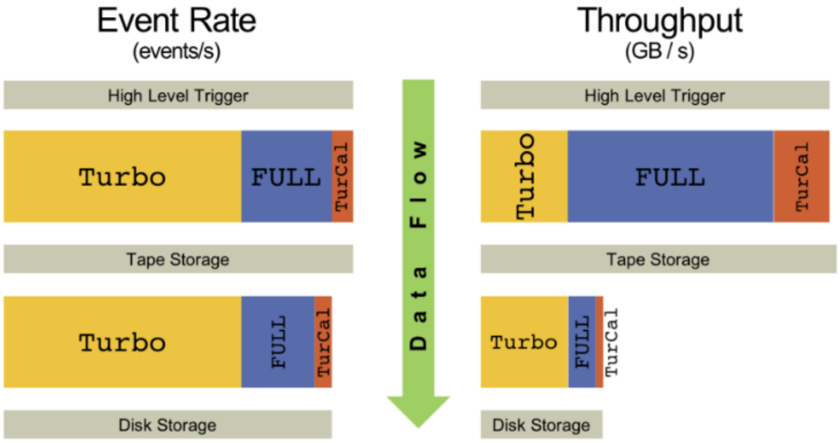


Figure 3.21: Relative contributions of the different output streams (Turbo, FULL, TurCal) to the event rate (left) and data throughput (right) in Run 3. Most events are stored in Turbo format, while the larger size of FULL events means they dominate storage throughput. TurCal provides calibration samples at low rate and size. Figure from [72].

The impact of different persistence levels can be quantified in terms of the bandwidth consumption:

$$BW = \frac{\text{event size} \times \text{event rate}}{10^6} \quad [\text{MB/s}],$$

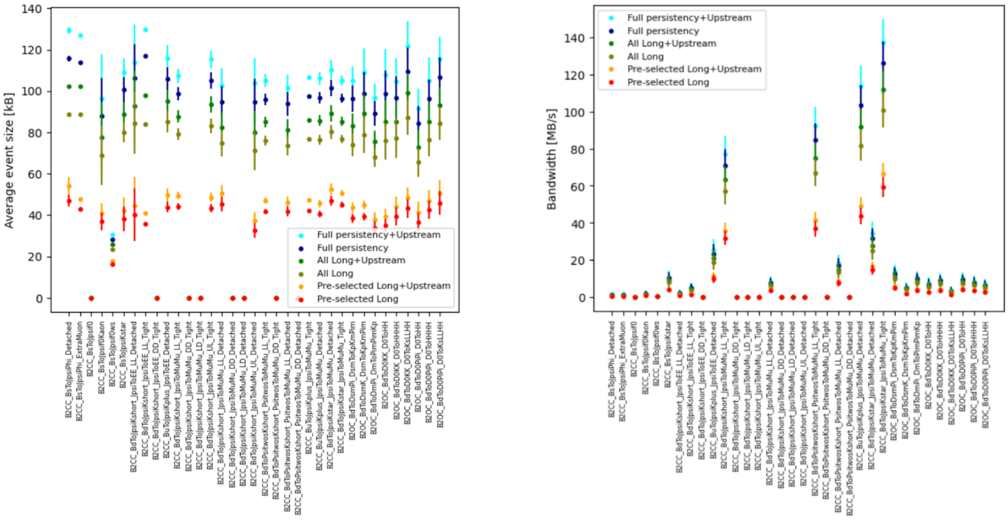
where the event size is expressed in bytes and the event rate in Hz. Given the finite bandwidth available for data storage and transfer, optimising the event size is essential to balance physics performance and computing constraints.

Within the scope of this PhD thesis, a dedicated study was performed to evaluate the event size and bandwidth associated with different persistence configurations for trigger lines relevant to flavour tagging. This study used Run 3 simulated samples of B decays into open-charm or charmonium final states, corresponding to a total of 10^5 events. Three main persistence configurations were explored:

- **Preselected long tracks:** only long tracks satisfying specific selections are saved. These include requiring that the track originates from the same primary vertex (PV) as the signal B candidate, or that the minimum χ^2_{IP} with respect to all PVs exceeds six. These empirical criteria correspond to loose selections designed to retain all potential tracks used by the flavour tagging algorithms.

- **All long tracks:** all long tracks in the event are persisted, without any additional selection.
- **Full persistence:** the complete reconstructed event is retained. This configuration is particularly important for inclusive flavour tagging algorithms, which use all reconstructed tracks in the event.

For each configuration, the effect of including upstream tracks was also quantified. Upstream tracks are especially relevant for same-side taggers, as low-momentum hadrons associated with the signal B meson are often reconstructed only in the tracking stations located upstream of the magnet. The results of these studies, summarised in Figure 3.22, demonstrate that selective persistence can significantly reduce bandwidth consumption while retaining the information required for flavour tagging. In particular, even when the event size is comparable across different trigger lines, higher trigger rates can lead to a substantial increase in bandwidth usage, as observed for channels such as $B^+ \to J/\psi K^+$ and $B^0 \to J/\psi K^{*0}$.



(a) Event size vs persistency.

(b) Bandwidth vs persistency.

Figure 3.22: Impact of different persistence levels and of including upstream tracks on event size (left) and bandwidth (right) for trigger lines relevant to flavour tagging. Higher trigger rates can lead to a substantial increase in bandwidth usage, as observed for channels such as $B^+ \to J/\psi K^+$ and $B^0 \to J/\psi K^{*0}$. An event size equal to zero indicates that the corresponding trigger line did not fire in the simulation.

3.6 Monte Carlo Simulation at LHCb

Monte Carlo (MC) simulation plays a central role in the LHCb physics programme, providing the essential link between theoretical predictions, detector response, and the reconstruction algorithms used in data analysis. In particular, the development of the Run 3 flavour-tagging algorithms presented in this thesis relies critically on simulated events, as illustrated in Chapter 5.

The simulation chain begins with the description of the proton–proton collisions provided by the LHC. The hard scattering processes are generated with [91, 92], tuned to LHCb kinematics. Heavy-flavour hadron decays are simulated with EvtGen [93], which models decay amplitudes and angular distributions with inputs from both theoretical calculations and experimental measurements. Final-state QED corrections, such as photon emission, are simulated with Photos [94]. The passage of particles through the LHCb detector is modelled using the Geant4-based LHCb simulation framework [95–97], which describes their interactions with matter and the resulting detector signals. This modelling includes effects such as multiple scattering, energy loss, shower development, and detector inefficiencies. The full Upgrade I detector geometry is implemented using the `DD4hep` framework [98], which provides a unified description of the detector for both simulation and reconstruction. The simulated detector signals are subsequently digitised to produce raw data objects equivalent to those recorded in real collisions. These are processed using the same reconstruction algorithms as data, ensuring consistent treatment of simulated and real events.

4 Flavour Tagging at LHCb

One of the primary objectives of the LHCb experiment at CERN is the study of *CP* violation in the heavy-flavour sector. As discussed in Section 2.3, measurements of mixing and time-dependent *CP*-violating asymmetries require the knowledge of the flavour of neutral B meson at the time of its production. The set of techniques developed for this purpose is collectively referred to as *flavour tagging*. It is important to note that the term *flavour tagging* has a different meaning in general-purpose experiments such as ATLAS and CMS, where it usually denotes the identification of the flavour of reconstructed jets [99] (for example, distinguishing b-jets from light-flavour jets). In the context of LHCb, flavour tagging refers exclusively to the determination of the production flavour of a neutral B meson, namely, whether the signal meson contained a b or a $\bar{b}$.
Two main categories of flavour tagging algorithms (hereafter also indicated as *taggers* for brevity) are employed at LHCb. The **classical taggers** exploit the correlation between the electric charge of a specific particle, referred to as the *tagging particle*, and the flavour of the signal B meson. The inclusive tagger, instead, uses information from all particles in the event (excluding those associated with the signal decay) to infer the production flavour of the signal B meson [100]. In both approaches, the goal is to provide, on an event-by-event basis, a tagging decision d indicating the inferred production flavour of the signal meson, together with a predicted mistag probability η, which estimates the probability that the tagging decision is incorrect.

This thesis focuses on the development of the classical taggers for the LHCb Upgrade I. Their implementation and optimisation are described in Chapter 5. The purpose of this chapter is to provide the foundation of how the classical taggers work (Section 4.1) and introduce the metrics used to evaluate their performance (Section 4.4).

4.1 Classical Taggers

Classical taggers use specific particles to infer the B flavour at production time, named as tagging particles, that carry an electric charge that is correlated with the flavour of the signal B meson at production. Depending on the process that produced this tagging particle, the algorithm is classified as an **opposite-side (OS)** or a **same-side (SS)** tagger.

Figure 4.1 provides a schematic illustration of the OS and SS tagging mechanisms and their connection to the quark content of the signal B meson.

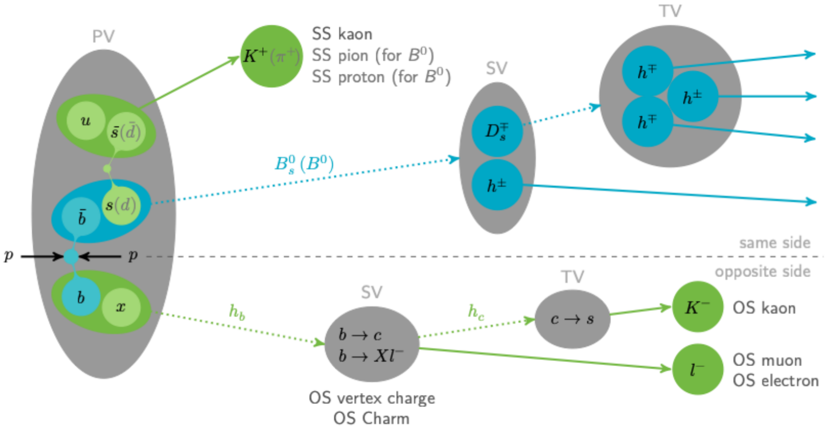


Figure 4.1: Schematic illustration of opposite-side (OS) and same-side (SS) flavour tagging algorithms at LHCb. OS taggers rely on the decay products of the B hadron produced in association with the signal, while SS taggers exploit charged hadrons produced during the fragmentation of the signal B meson. Image source [101]

.

OS taggers exploit the $b\bar{b}$ pair production. While one quark hadronises into the signal B meson, the other forms an OS B hadron whose decay product charges can be used to infer its production flavour and, as a consequence, the production flavour of the signal candidate. The most commonly used particles for OS tagging are kaons, muons, and electrons (**OSKaon tagger, OSMuon tagger, OSElectron tagger**). In Run 1-2, additional algorithms such as the **OSVertexCharge** and **OSCharm** taggers were also employed. These rely on the reconstruction of secondary vertices or charm hadrons from the opposite-side B decay.

SS taggers exploit particles produced in the hadronisation of the signal B meson. During hadronisation, the signal B meson is formed together with charged particles that carry correlated flavour information. The tagging particle typically originates from the fragmentation of the light quark (kaon, proton, or pion) that combines with the $\bar{b}$ (or b) quark to form the signal B meson. The charge of this particle provides information on the flavour of the B meson at production. Different SS taggers are used depending on the signal meson species: the **SSKaon tagger** for B_s^0 mesons, the **SSPion tagger** and **SSProton tagger** for B^0 mesons. SS taggers are particularly sensitive to the modelling of the hadronisation environment.

From the considerations above, the inferred flavour of the signal B at production, denoted as *tagging decision* d, can be derived from the electric charge of the tagging particle Q_{track} as:

- for charged decays (e.g. $B^+ \to J/\psi K^+$) and SS/OS taggers: $d = -1 \cdot Q_{\text{track}}$
- for neutral decays (e.g. $B^0 \to J/\psi K^{*0}$, $B_s^0 \to D_s^- \pi^+$):
  - SSPion/SSKaon taggers: $d = +1 \cdot Q_{\text{track}}$
  - SSProton/OS taggers: $d = -1 \cdot Q_{\text{track}}$

The tagging decision can take the values $d \in \{+1, -1, 0\}$, where the convention adopted in this work is

$$d = +1 \quad \longrightarrow \quad \text{signal meson contains a } \bar{b} \quad \longrightarrow \quad B, \tag{4.1}$$

$$d = -1 \quad \longrightarrow \quad \text{signal meson contains a } b \quad \longrightarrow \quad \overline{B}, \tag{4.2}$$

while $d = 0$ is assigned when the flavour can not be determined.

The probability that the assigned tagging decision is incorrect is denoted by the *predicted mistag probability* (or *mistag estimate*) η, with $\eta \in [0, 0.5]$. A value of $\eta = 0$ corresponds to a perfectly reliable tagging decision, while $\eta = 0.5$ corresponds to a random assignment with no discriminating power. The mistag probability is estimated using a multivariate classifier, which provides an event-by-event prediction by exploiting kinematic and topological properties of the selected tagging particle and the signal decay.

4.2 Mistag Calibration

The predicted mistag probability η does not coincide exactly with the true mistag probability ω. Several effects can lead to systematic deviations between these two quantities. Since flavour tagging algorithms are typically trained using simulated events, residual mismodelling of detector effects or of the underlying physics processes can bias the mistag prediction when the algorithm is applied to data. In addition, the performance of a tagger may depend on the kinematic properties or selection criteria of the decay channel used for training; if these differ from those of the target analysis channel, the predicted mistag probability can become systematically biased. Finally, multivariate classifiers may partially overfit statistical fluctuations in the training sample, leading to a mistag estimate that does not generalise to independent data.

To account for these effects, a dedicated calibration procedure is performed to relate the predicted mistag probability η to the true mistag probability ω. In practice, the calibration is performed by grouping events into bins of the predicted mistag probability η and measuring, in each bin, the fraction ω of incorrectly tagged events N_{wrong},

$$\omega = \frac{N_{\text{wrong}}}{N_{\text{tagged}}}, \tag{4.3}$$

where $N_{\text{tagged}} = N_{\text{right}} + N_{\text{wrong}}$ and N_{right} is the number of correctly tagged events. This procedure requires knowledge of the true production flavour of the signal B meson. In simulated samples, the true production flavour is available for all decays. In data, however, it can only be determined for specific calibration channels. For self-tagging charged B decays, such as $B^+ \to J/\psi K^+$, the charge of the final-state kaon uniquely identifies the flavour of the B meson at production. Since charged B mesons do not undergo flavour oscillations, their production flavour is known unambiguously. For flavour-specific neutral B decays, such as $B_s^0 \to D_s^- \pi^+$ and $B^0 \to J/\psi K^{*0}$, the charge of the decay products identifies the flavour at decay time. In these cases, statistical corrections must be applied to account for neutral-meson oscillations as a function of the decay time.

The calibration of flavour tagging algorithms is performed using the dedicated `lhcb_ftcalib` software package [102]. This framework implements a Generalised Linear Model (GLM) to map the predicted mistag probability η to the calibrated mistag probability ω. The calibration function is expressed as a polynomial in a transformed variable, defined through a link function $g(\eta)$, allowing for a non-linear mapping of η. The general calibration model can be written as:

$$\omega(\eta) = g\left(g^{-1}(\eta) + \sum_{i=0}^{m} p_i P_i(\eta) \right), \tag{4.4}$$

where p_i are calibration parameters, $P_i(\eta)$ are orthogonal polynomial basis functions, and g is the link function. Two link functions are commonly used in practice: the *identity (mistag)* link, $g(x) = x$, and the *logit* link, defined as $g(x) = \ln\left(\frac{x}{0.5-x}\right)$. The latter guarantees that the calibrated ω always lies within the interval $[0, 1]$.

`lhcb_ftcalib` employs a convention that parameterizes the deviation from the identity $\omega(\eta) = \eta$ and define the polynomial basis so that the parameters p_0 and p_1 are uncorrelated [102]. This choice simplifies the interpretation of the calibration parameters in terms of bias and slope corrections.

For the simplest case (linear calibration with the identity link $g(x) = x$), expression 4.2 simplifies to:

$$\omega(\eta) = \eta + p_0 + p_1 (\eta - \langle\eta\rangle) = p_0 - p_1 \langle\eta\rangle + \eta(p_1 + 1), \tag{4.5}$$

where $\langle\eta\rangle$ denotes the average predicted mistag probability in the calibration sample. In this formulation, p_0 quantifies the offset of the mistag fraction, while p_1 measures the deviation of the calibration slope from unity.

Although the first-order polynomial form is commonly used due to its simplicity and interpretability, higher-order polynomial terms can also be included in the GLM framework when the data exhibits more complex dependencies between the predicted and true mistag probabilities. This allows the calibration model to capture subtle non-linearities, particularly at the extremes of the η distribution.

To account for possible differences in tagging performance between B and $\bar{B}$ mesons, the calibration model can be extended to include flavour-dependent parameters. Such differences may arise from production asymmetries in proton–proton collisions or from charge-dependent detection effects, and can lead to small but non-negligible biases if not properly accounted for.

In this approach, separate calibration functions are defined for candidates tagged as B and $\bar{B}$, while preserving a common functional form. For a first-order calibration, this can be achieved by introducing asymmetry parameters Δp_0 and Δp_1, defined such that

$$p_0^B = p_0 + \tfrac{1}{2}\,\Delta p_0, \qquad p_0^{\bar{B}} = p_0 - \tfrac{1}{2}\,\Delta p_0, \tag{4.6}$$

$$p_1^B = p_1 + \tfrac{1}{2}\,\Delta p_1, \qquad p_1^{\bar{B}} = p_1 - \tfrac{1}{2}\,\Delta p_1. \tag{4.7}$$

Using the identity link function for illustration, the calibrated mistag probability for B and $\bar{B}$ mesons can then be written as

$$\omega^B(\eta) = \eta + p_0^B + p_1^B\,(\eta - \langle\eta\rangle), \tag{4.8}$$

$$\omega^{\bar{B}}(\eta) = \eta + p_0^{\bar{B}} + p_1^{\bar{B}}\,(\eta - \langle\eta\rangle). \tag{4.9}$$

In this parameterisation, p_0 and p_1 describe the average calibration behaviour, while Δp_0 and Δp_1 quantify the asymmetry between B and $\bar{B}$ mesons in the offset and slope of the mistag calibration, respectively. A perfectly symmetric tagging performance corresponds to $\Delta p_0 = \Delta p_1 = 0$, whereas non-zero values indicate flavour-dependent effects that must be taken into account to avoid biases in time-dependent *CP*-violation and mixing measurements.

4.3 Tagger Combination

Multiple taggers can be combined to obtain a single prediction of the initial B flavour d_{comb} with an associated mistag estimate η_{comb}.

The combination procedure is based on the calculation of the relative probabilities that the candidate originated from a b or $\bar{b}$ quark, following a likelihood product over all individual taggers. The combined probability P_b ($P_{\bar{b}}$) for a b ($\bar{b}$) quark is defined as:

$$P_b = \frac{p_b}{p_b + p_{\bar{b}}}, \qquad P_{\bar{b}} = 1 - P_b, \tag{4.10}$$

where p_b and $p_{\bar{b}}$ represent the unnormalised probabilities that the tagging decisions correspond to a b or $\bar{b}$ quark, respectively.

Each individual tagger contributes to these probabilities through:

$$p_b = \prod_i \left(\frac{1 + d_i}{2} - d_i \left[1 - \omega_i(\eta_i) \right] \right), \tag{4.11}$$

$$p_{\bar{b}} = \prod_i \left(\frac{1 - d_i}{2} + d_i \left[1 - \omega_i(\eta_i) \right] \right), \tag{4.12}$$

where $\omega_i(\eta_i)$ denotes the calibrated mistag probability of the i-th tagger, evaluated at its predicted η_i.

From the relative probabilities P_b and $P_{\bar{b}}$, the combined tagging decision and mistag estimate are obtained as:

$$d_{\text{comb}} = \text{sign}(P_b - P_{\bar{b}}) \qquad \eta_{\text{comb}} = 1 - \max(P_b, P_{\bar{b}}) \tag{4.13}$$

The combined tagger output is then calibrated following the same procedure described in Section 4.2, ensuring that the final combined mistag probability $\omega_{\text{comb}}(\eta_{\text{comb}})$ correctly reflects the true mistag fraction observed in data. In this thesis work, the combination is performed via the `lhcb-ftcalib` [102] toolset that handles the error propagation of the uncertainties originating from the calibration of the individual taggers.

4.4 Performance Metrics

The performance of a flavour tagging algorithm is characterised by three key quantities that allow the determination of the sensitivity to *CP* asymmetry [103]: the *tagging efficiency*, the *mistag probability*, and the derived *tagging power*.

The **tagging efficiency** ϵ_{tag} represents the fraction of signal events for which the tagger is able to provide a flavour decision, regardless of whether the decision is correct or not. It is defined as

$$\varepsilon_{\text{tag}} = \frac{N_{\text{wrong}} + N_{\text{right}}}{N_{\text{wrong}} + N_{\text{right}} + N_{\text{untagged}}} \tag{4.14}$$

where N_{wrong} is the number of incorrectly tagged events, N_{right} is the number of correctly tagged events and N_{untagged} is the number of signal events for which the tagger do not provide a tagging decision.

The **average mistag probability** ω (already informally introduced in the previous section) is quantified by the fraction of tagged events for which the tagging decision is incorrect. It is expressed as

$$\omega = \frac{N_{\text{wrong}}}{N_{\text{wrong}} + N_{\text{right}}} \tag{4.15}$$

where N_{wrong} is the number of incorrectly tagged events.

The impact of mis-tagging on the overall tagging performance is described by the **dilution factor** D, defined as

$$D = 1 - 2\omega \tag{4.16}$$

A perfect tagger exhibits $D = 1$, while a completely random tagger yields $D = 0$. The dilution factor encapsulates how much the tagging decision reduces the observed oscillation or asymmetry amplitude.

The tagging efficiency and the mistag are combined in a single figure of merit the **tagging power** ϵ_{eff}, or **effective tagging efficiency**, defined as

$$\varepsilon_{\text{eff}} = \varepsilon_{\text{tag}}\, D^2 = \varepsilon_{\text{tag}}(1 - 2\omega)^2 \tag{4.17}$$

The tagging power allows to compare the performance of different taggers and to define the *effective* statistical power $N_{\text{eff}} = \varepsilon_{\text{eff}} \cdot N$ of a flavour-tagged sample with N events.

4.4.1 Impact on *CP* Asymmetry Measurements

For a sample of size N, the measured time-dependent *CP* asymmetry A_{CP}^{meas} is reduced with respect to the true asymmetry A_{CP}^{true} by a dilution factor depending on the mistag [103]:

$$A_{CP}^{\text{meas}}(t) = \frac{N(\bar{B}_q^0 \to f)(t) - N(B_q^0 \to f)(t)}{N(\bar{B}_q^0 \to f)(t) + N(B_q^0 \to f)(t)} = (1-2\omega) \cdot A_{CP}(t) = D \cdot A_{CP}^{\text{true}}(t) \quad (4.18)$$

Therefore, the A_{CP}^{true} and its statistical error can be obtained as:

$$A_{CP}^{\text{true}} = \frac{A_{CP}^{\text{meas}}}{1-2\omega} \qquad \sigma_{A_{CP}^{\text{true}}} \propto \frac{1}{\sqrt{\epsilon_{\text{tag}} N}(1-2\omega)} \equiv \frac{1}{\sqrt{\epsilon_{\text{eff}} N}} \quad (4.19)$$

This expression shows explicitly that the statistical sensitivity can be improved both by higher statistics and higher tagging power. While the High-Luminosity LHC will provide substantially larger datasets, maximising the tagging power remains a central objective in the development of flavour tagging algorithms.

Part II

Main Results

---

5 Development of the Run 3 Classical Taggers

5.1 Introduction

The start of Run 3 in 2022 marked a major shift in the operating conditions of the LHCb experiment [104, 105]. The instantaneous luminosity nearly doubled compared to Run 2, increasing the number of pp interactions per bunch crossing and raising the average event occupancy. While the higher luminosity is beneficial for precision measurements by providing larger signal yields, it also represents a significant challenge for flavour tagging. The probability that an unrelated track satisfies the kinematic or topological requirements of a tagging candidate increases, degrading the tagging power ϵ_{eff} (Figure 5.1).

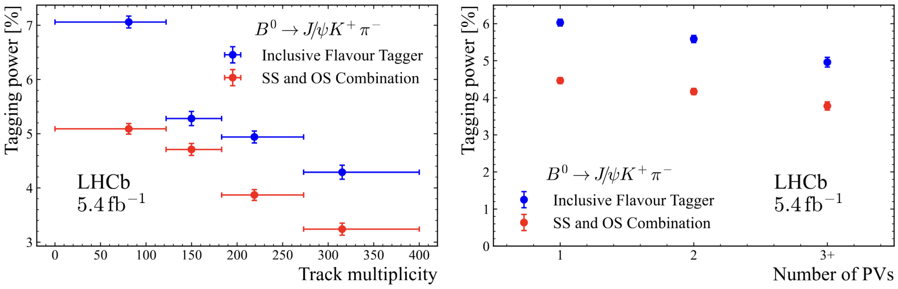


Figure 5.1: Tagging power of the Run 2 classical tagger combination (OS and SS) and Run 2 inclusive tagger as a function of track multiplicity (left) and number of primary vertices (right) using 2018-2017-2016 LHCb data ($B^0 \to J/\psi K^{*0}$) [100].

To address the Run 3 operating conditions, the LHCb detector underwent a full upgrade (Section 3.4), including new tracking systems, improved particle identification, and a re-designed trigger system (as described in Section 3.4). These modifications motivate the development of new flavour-tagging algorithms (*taggers*) specifically designed for Run 3. In particular, this chapter focuses on the development of the single-track taggers (OSKaon, OSMuon, OSElectron, SSKaon, SSPion, and SSProton). The algorithms developed combine physics-motivated principles with machine-learning techniques. Each tagger consists of a decision-tree-based selection to identify the relevant tagging particles (described in Section 5.3), followed by a neural-network-based classifier used to estimate the mistag probability (described in Section 5.4).

5.2 Samples and Event Selections

The development of the Run 3 flavour-tagging algorithms relies primarily on simulated samples produced with the framework described in Section 3.6, while the evaluation of their performance is done on a subset of data collected in 2024 (25 June to 6 August). The detector configuration and data-taking conditions are the same for simulation and real-data: pp collisions at a centre-of-mass energy of $\sqrt{s} = 13.6$ TeV, an average $\mu \approx 4.4$ of visible interactions per bunch crossing and magnet polarity up.

The decay channels used for the tagger development are listed in Table 5.1, together with the respective final state particles. For notational convenience, references to a meson such as B^0, B^0_s, or B^+ should be understood to include the respective anti-particle ($\overline{B}^0$, $\overline{B}^0_s$, and B^-). Any discussion requiring a distinction between the particle and anti-particle will use explicit notation.

Table 5.1: Decay channels used for developing the Run 3 flavour-tagging algorithms.

Taggers	**Training channel**
OSMuon, OSKaon, OSElectron	$B^+ \to J/\psi K^+$ $(J/\psi \to \mu^+\mu^-)$
SSKaon	$B^0_s \to D^-_s \pi^+$ $(D^-_s \to K^+K^-\pi^-)$
SSPion, SSProton	$B^0 \to J/\psi K^{*0}$ $(J/\psi \to \mu^+\mu^-$, $K^{*0} \to K^+\pi^-)$

The decay $B^+ \to J/\psi K^+$ is a high-yield and experimentally clean channel at LHCb and is therefore used for the development of the opposite-side (OS) flavour taggers. Since OS taggers exploit information from the hadron containing the opposite b produced in the event, their performance is largely independent of the kinematic properties of the signal decay. This allows the use of charged $B^\pm$ mesons, which do not undergo flavour oscillations. The flavour at decay time is fixed by the charge of the K^+.

In contrast, same-side (SS) taggers exploit correlations between the signal B meson and particles produced in its hadronisation. These correlations are more sensitive to the kinematic properties of the signal species and therefore differ between B^+, B^0, and B^0_s mesons. For this reason, SS taggers are trained on neutral modes representative of their target applications. The SSKaon tagger, used for B^0_s tagging, is trained on $B_s^0 \to D_s^- \pi^+$, where the D_s^- ($D_s^- \to K^+ K^- \pi^-$) charge determines the flavour of the B_s at decay. SSPion/SSProton taggers, used for B^0 tagging, are trained on $B^0 \to J/\psi K^{*0}$ with $J/\psi \to \mu^+\mu^-$ and the K^{*0} decay ($K^{*0} \to K^+\pi^-$) fixes the B^0 flavour at decay time. The presence of the $J/\psi \to \mu^+\mu^-$ provides a clean and efficiently triggered final state, in close analogy to the $B^+ \to J/\psi K^+$ mode.

The HLT2 trigger lines used to select the $B^+ \to J/\psi K^+$, $B^0 \to J/\psi K^{*0}$, and $B_s^0 \to D_s^- \pi^+$ decays are summarised below and are the same for simulation and data. These lines apply kinematic, vertex-quality, mass-window, and particle-identification requirements to the intermediate and final-state particles to identify the signal candidates.

For the $B^+ \to J/\psi K^+$ mode, a $J/\psi \to \mu^+\mu^-$ candidate is required within a tight mass window around the nominal J/ψ mass (typically $3046 < m_{\mu\mu} < 3146$ MeV at the vertex level), with good vertex quality ($\chi^2/\text{ndf} < 8$) and the muons are required to have an associated $PID_\mu > 0$. The accompanying kaon must satisfy minimal kinematic thresholds (e.g. $p_T > 0.8$ GeV, $p > 3$ GeV) and loose PID cuts ($PID_K > -1$ for the kaon). The resulting B^+ candidate must have a reconstructed mass compatible with the B mass ($5080 < m_B < 5620$ MeV) and a decay time exceeding 0.2 ps.

The $B^0 \to J/\psi K^{*0}$ line follows an analogous strategy, replacing the bachelor kaon with a $K^{*0} \to K^+\pi^-$ candidate. The K^{*0} is required to lie within its nominal mass window ($826 < m_{K\pi} < 966$ MeV), with modest kinematic requirements on its daughters ($p_T > 500$ MeV, $p > 1$ GeV) and appropriate kaon and pion PID selections (for the pion $PID_K < 3$, for the kaon $PID_K > -2$). The B^0 candidate is then formed from the J/ψ and K^{*0} and required to have a decay time exceeding 0.2 ps.

For the $B_s^0 \to D_s^- \pi^+$ mode, the $D_s^- \to K^+K^-\pi^-$ candidate is selected within a narrow mass window around the nominal D_s mass ($1930 < m_{KK\pi} < 2025$ MeV), with loose PID requirements (e.g. $PID_K > -5$ for kaons, $PID_K < 5$ for pions) and kinematic thresholds on its decay products (e.g. $p > 2$ GeV, $p_T > 250$ MeV). The bachelor pion is required to be harder ($p > 5$ GeV, $p_T > 500$ MeV). The B_s^0 candidate is constructed in a wide mass window ($5000 < m_B < 6000$ MeV), after which a dedicated multivariate classifier is applied, to select displaced B decays and suppress combinatorial background.

5.3 Tagging Particle Selection via a Decision Tree Classifier

To identify the tagging particles associated to each tagger in data, a decision tree classifier is trained on a simulated dataset comprising the decays $B^+ \to J/\psi K^+$, $B^0 \to J/\psi K^{*0}$, and $B_s^0 \to D_s^+ \pi^-$, thereby guaranteeing that all tagging particles are represented. The selection problem is framed as a multilabel classification task with a total of seven classes: one for each tagging particle category (OSElectron, OSMuon, OSKaon, SSKaon, SSPion, SSProton) and an extra class, notSamePV, introduced to exclude tracks not originating from the same primary vertex as the signal B meson. Tracks are assigned to the classes by combining two Monte Carlo variables: the particle's true ID and its origin flag. The true ID corresponds to the particle type identifier defined by the Particle Data Group (PDG) numbering scheme [25]. The origin flag encodes the process from which the particle originated in the simulated event allowing to distinguish SS and OS particles (detailed explanation of the origin flag in Section A.1.1 of the Appendix). Table 5.2 shows the class composition of the training dataset ($\approx 3.8 \cdot 10^7$ tracks). To prevent the classifier from being dominated by tracks coming from other PVs (93%), each class is assigned a weight w_c

Table 5.2: Decision tree class composition for the training dataset ($\approx 3.8 \cdot 10^7$ tracks), containing simulated samples of $B^0 \to J/\psi K^{*0}$, $B^+ \to J/\psi K^+$, $B_s^0 \to D_s^- \pi^+$).

Class	**Composition (%)**
NotSamePV	93.36
SSPion	4.36
SSKaon	0.71
OSKaon	0.70
SSProton	0.49
OSMuon	0.22
OSElectron	0.17

$$w_c = \frac{N_{\text{tot}}}{n_{\text{classes}} \times N_c}\,, \tag{5.1}$$

with N_{tot} the total number of tracks, n_{classes} the number of classes, and N_c the number of tracks in class c. This reweighting ensures that the algorithm gives equal importance to all classes during the training.

In terms of input variables for the training, the decision tree is provided with a broad feature set (65 features), including geometrical, kinematic and PID variables.

This allows the algorithm to identify the most relevant variables for maximizing the class purity in the splitting. The full list of input features is reported in Table A.1 in the Appendix, together with a brief description of each and their distributions (Figures A.2, A.3). To control the model complexity, the tree depth is limited to six levels and a minimum impurity decrease of 0.009 is required for a split to be created, ensuring that additional splits are introduced only when they yield a significant improvement in discrimination. Once the decision tree is trained, the tagging particle selections are derived by following the successive branch splits leading to the terminal leaves. Each terminal leaf is associated with a distinct tagging-particle class, determined by the majority of samples populating that leaf. An additional requirement is imposed to exclude particles originating from the signal decay ($OriginFlag \neq 0$). The complete structure of the trained tree is provided in Figure A.4 of Appendix A, while the resulting cuts are presented below in readable format as a set of logical conditions.

OSElectron

Tracks are selected as OSElectron if they satisfy:

$$\text{ProbNN}_e > 0.6606 \ \wedge \ \text{OriginFlag} \neq 0\,.$$

OSKaon

Tracks are selected as OSKaon if they satisfy (A) or (B):

$$\begin{aligned}
\text{(I)} \quad & \text{ProbNN}_e \leq 0.6606 \ \wedge \ \text{ProbNN}_\mu \leq 0.6472 \ \wedge \ \Delta z \leq 6.7063 \\
& \wedge \ \text{ProbNN}_\pi \leq 0.9496 \ \wedge \ \chi^2_{\text{IP}}(\text{PV}_{\text{best}}(B)) > 9.2305 \\
& \wedge \ \text{ProbNN}_K > 0.4192 \wedge \ \text{OriginFlag} \neq 0, \\[1ex]
\text{(II)} \quad & \text{ProbNN}_e \leq 0.6606 \ \wedge \ \text{ProbNN}_\mu \leq 0.6472 \ \wedge \ \Delta z > 6.7063 \\
& \wedge \ \chi^2_{\text{IP}}(\text{PV}_{\text{own}}) > 16.7833 \ \wedge \ \text{OriginFlag} \neq 0\,.
\end{aligned}$$

OSMuon

Tracks are selected as OSMuon if they satisfy (A) or (B):

$$\begin{aligned}
\text{(I)} \quad & \text{ProbNN}_e \leq 0.6606 \ \wedge \ \text{ProbNN}_\mu \leq 0.6472 \ \wedge \ \Delta z \leq 6.7063 \\
& \wedge \ \text{ProbNN}_\pi \leq 0.9496 \ \wedge \ \chi^2_{\text{IP}}(\text{PV}_{\text{best}}(B)) > 9.2305 \\
& \wedge \ \text{ProbNN}_K \leq 0.4192 \wedge \ \text{OriginFlag} \neq 0, \\[1ex]
\text{(II)} \quad & \text{ProbNN}_e \leq 0.6606 \ \wedge \ \text{ProbNN}_\mu > 0.6472 \ \wedge \ \text{OriginFlag} \neq 0\,.
\end{aligned}$$

SSPion

Tracks are selected as SSPion if they satisfy (A) or (B):

$$\begin{aligned}&\mathrm{ProbNN}_e \leq 0.6606 \ \wedge\ \mathrm{ProbNN}_\mu \leq 0.6472 \ \wedge\ \mathit{\Delta} z \leq 6.7063\\ &\wedge\ \mathrm{ProbNN}_\pi > 0.9496\\ &\wedge\ \mathrm{OriginFlag} \neq 0,\end{aligned}$$

SSProton

Tracks are selected as SSProton if they satisfy:

$$\begin{aligned}&\mathrm{ProbNN}_e \leq 0.6606 \ \wedge\ \mathrm{ProbNN}_\mu \leq 0.6472 \ \wedge\ \mathit{\Delta} z \leq 6.7063\\ &\wedge\ \mathrm{ProbNN}_\pi \leq 0.9496 \ \wedge\ \chi^2_{\mathrm{IP}}(\mathrm{PV}_{\mathrm{best}}(B)) \leq 9.2305\\ &\wedge\ \mathrm{ProbNN}_p > 0.3049 \ \wedge\ \mathrm{OriginFlag} \neq 0\,.\end{aligned}$$

SSKaon

Tracks are selected as SSKaon if they satisfy:

$$\begin{aligned}&\mathrm{ProbNN}_e \leq 0.6606 \ \wedge\ \mathrm{ProbNN}_\mu \leq 0.6472 \ \wedge\ \mathit{\Delta} z \leq 6.7063\\ &\wedge\ \mathrm{ProbNN}_\pi \leq 0.9496 \ \wedge\ \chi^2_{\mathrm{IP}}(\mathrm{PV}_{\mathrm{best}}(B)) \leq 9.2305\\ &\wedge\ \mathrm{ProbNN}_p \leq 0.3049 \ \wedge\ \mathrm{OriginFlag} \neq 0\,.\end{aligned}$$

notSamePV

Tracks belong to another PV if:

$$\begin{aligned}&\mathrm{ProbNN}_e \leq 0.6606 \ \wedge\ \mathrm{ProbNN}_\mu \leq 0.6472 \ \wedge\ \mathit{\Delta} z > 6.7063\\ &\wedge\ \chi^2_{\mathrm{IP}}(\mathrm{PV}_{\mathrm{own}}) \leq 16.7833 \ \wedge\ \mathrm{OriginFlag} \neq 0\,.\end{aligned}$$

Overall, the PID-related variables (ProbNN, PID_K) contribute to particle identification, while geometric variables are crucial for distinguishing their origin (OS, SS, or other PVs). As can be observed in Figure 5.2, the IP χ^2 w.r.t the best PV associated to the B signal is typically large for OS candidates, which are displaced relative to the B production PV, and smaller for SS candidates, which tend to originate closer to the PV. Since both OS and SS tracks are produced in the same pp interaction as the signal B meson, they populate the region of small $\mathit{\Delta} z(\mathrm{PV}_{\mathrm{tag}}, \mathrm{PV}_B)$, defined as absolute value of the difference in $z-$position between the tagging track best PV and the B best PV. Ideally, this quantity would peak at zero, but detector effects and reconstruction uncertainties contribute to broadening the distribution. In contrast, tracks classified as notSamePV exhibit a broader $\mathit{\Delta} z$ distribution and

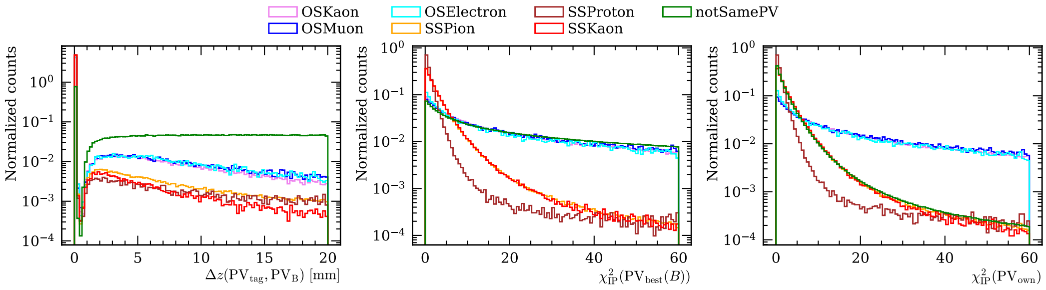


Figure 5.2: Distributions by tagging particles (OSKaon, OSMuon, OSElectron, SSPion, SSProton, SSKaon) and particles associated to other PVs than the B signal. (Left) Absolute value of the difference in $z-$position between the tagging track best PV and the B signal best PV. (Middle) IP χ^2 w.r.t. the B signal best PV. (Right) IP χ^2 w.r.t. own best PV. The sample is made of $\approx 3.8 \cdot 10^7$ tracks (MC simulation of $B^0 \to J/\psi K^{*0}$, $B^+ \to J/\psi K^+$, $B^0_s \to D^-_s \pi^+$).

dominate at large separations ($\mathit{\Delta} z > 6.7$ mm), reflecting their origin from different PVs. It is interesting to note that the OSKaon selection is made of two cuts (denoted as branches I and II). Branch I is used to discriminate OSKaon from SSKaon and SSProton, while branch II drives the separation OSKaon-notSamePV. The cut $\mathit{\Delta} z > 6.7$ mm in branch II retains OSKaons originating from the decay chain $b \to c \to s$, which can be significantly displaced from their PV. The additional requirement on the $\chi^2_{\mathrm{IP}}(\mathrm{PV}_{\mathrm{own}})$ (defined as the IP χ^2 w.r.t. the own best PV) ensures that these displaced OSKaons are distinguished from particles associated with other PVs, which have large $\mathit{\Delta} z$ but exhibit small $\chi^2_{\mathrm{IP}}(\mathrm{PV}_{\mathrm{own}})$ values.

The confusion matrix (Table 5.3) quantifies the selection efficiency (contamination) for each tagging particle class by computing:

$$\epsilon(A \to B) = \frac{N(\mathrm{true} = A \,\&\, \mathrm{predicted} = B)}{N(\mathrm{true} = A)_{\mathrm{total}}}$$

where $N(\mathrm{true} = A \,\&\, \mathrm{predicted} = B)$ is the number of particle of type A predicted as of type B and $N(\mathrm{true} = A)_{\mathrm{total}}$ is total number of true A particles. The diagonal elements (in bold) therefore represent the selection efficiency for each tagging particle class after applying the cuts derived from the decision tree.

OSElectron and OSMuon have relatively high selection efficiency, 87.4% and 87.6% respectively. For hadronic tagging particles, the efficiencies are lower: OSKaon class has selection 66.8% efficiency, with a notable fraction of true OSKaons being misclassified as SSKaons (15.9%). The lowest efficiency is observed for SSProton at 52.8%, with significant confusion with SSKaon (25.6%) and SSPion (7.5%). This pattern reflects a well-known challenge in tagging, where separating charged

Table 5.3: Confusion matrix (in %) for the decision tree particle classification. The efficiency is computed as $\frac{N(\text{true}=A\,\&\,\text{predicted}=B)}{N(\text{true}=A)_{\text{total}}}$. The diagonal values (in bold) indicate the efficiency for each particle type after applying the selection cuts.

True \Predicted	OSElectron	OSKaon	OSMuon	SSKaon	SSPion	SSProton	notSamePV
OSElectron	**87.4**	2.9	5.4	2.4	0.9	0.1	1.0
OSKaon	0.3	**66.8**	7.0	15.9	5.4	1.9	2.6
OSMuon	0.9	3.9	**87.6**	1.9	4.7	0.1	0.9
SSKaon	0.4	8.8	1.9	**70.7**	8.9	8.3	1.0
SSPion	0.8	5.0	2.8	13.3	**75.5**	1.0	1.6
SSProton	0.4	9.6	3.8	25.6	7.5	**52.8**	0.4
notSamePV	3.5	8.8	2.1	0.9	6.2	0.2	**78.2**

hadrons (kaons, pions, and protons) with similar kinematic properties and partially overlapping PID responses is intrinsically difficult.

Alternative classification strategies were explored in this thesis, including the removal of the notSamePV class during training or the introduction of a dedicated "other" particle class to collect candidates that do not satisfy any tagging definition such as prompt particles from the pp collision, particles originated from interaction with the detector material or coming from the opposite-b-fragmentation (schema A.1 in Appendix). The configuration exposed in this section is the one ultimately adopted because it provided the best compromise between selection efficiency and purity across all taggers. Further optimisation of misidentification patterns remains an interesting avenue for future improvements, with potential benefits for the mistag estimation. In particular isolation techniques could improve the OSKaon contamination from tracks associated with other primary vertices. These techniques, aiming at isolating SS and OS tagging tracks, are currently under development within the Flavour Tagging Working Group and represent a promising direction.

5.4 Mistag Estimation via Neural Network

As described in Section 4.1, the goal of flavour tagging is to provide, for each event, a tagging decision d indicating the inferred production flavour of the B signal, together with a mistag probability η, which estimates the probability that d is incorrect. The mistag estimation is achieved by training a neural network (NN) to predict whether the tagging decision derived from tagging particle charge is correct ($label = 1$) or not ($label = 0$). The mistag estimate is thus $\eta = 1 -$ NN output, where the NN output is the probability of getting $label = 1$. Given that the mistag is defined in $[0, 0.5]$, if $\eta > 0.5$, d is inverted ($d_{\text{new}} = -d$) and η is recomputed as $\eta_{\text{new}} = 1 - \eta$. This manipulation ensures that η always represents the probability that the tagging decision is wrong, avoiding information loss when the NN predicts the opposite decision with high confidence. When more than a tagging particle for the same tagger species is present in an event, only the tagging decision with the lowest predicted mistag is retained.

5.4.1 Neural Network Setup, Training, Grid Search

Table 5.4 collects the variable used as input features for the NN training. Two sets are defined, based on the variables used for the Run 2 taggers: one for the OS taggers and one for the SS taggers, and kept identical within the same tagger family. This choice simplifies the initial development stage by ensuring that performance differences among taggers within the same family are not attributable to differences in NN input variables. An explanation of each variable can be found in Table A.3 in Appendix.

Table 5.4: OS and SS input feature sets. Unless specified, the variables refer to the tagging particle. The feature ΔQ_X is computed under the mass hypothesis of the specific tagging particle associated with the tagger being trained: e.g. ΔQ_π for the SSPion.

Tagger Type	**Features**
OSKaon, OSElectron, OSMuon	nTracks, nPVs, p, p_T, $p_\mathrm{T}(B)$, $\sqrt{\chi^2_\mathrm{IP}(\mathrm{PV_{own}})}$, χ^2/ndf, ProbNN$_K$, ProbNN$_p$, ProbNN$_e$, ProbNN$_\mu$, ProbNN$_\pi$, GhostProb, $\lvert\mathrm{IP}(\mathrm{PV_{own}})\rvert$, ΔR, ΔQ_X
SSPion, SSProton, SSKaon	nTracks, nPVs, p, p_T, $p_\mathrm{T}(B)$, $\sqrt{\chi^2_\mathrm{IP}(\mathrm{PV_{own}})}$, χ^2/ndf, ProbNN$_K$, ProbNN$_p$, ProbNN$_e$, ProbNN$_\mu$, ProbNN$_\pi$, GhostProb, $\lvert\mathrm{IP}(\mathrm{PV_{own}})\rvert$, ΔR, ΔQ_X, $\Delta\phi$, $\Delta\eta$, $p_\mathrm{T}(B+tag)$

For the OS taggers, the feature set includes event-level observables (number of reconstructed primary vertices, number of tracks, kinematics of the B signal), and tagging particle variables (momentum p, transverse momentum p_T, track-fit quality χ^2/ndf, and the ProbNN variables). Geometric information is also included: the absolute value of the IP w.r.t. the best PV, its significance, and the quadratic angular separation between the tagging track and the B signal (ΔR). Additional inputs are the ghost (fake-track) probability and the mass–hypothesis variable ΔQ, defined as the difference between the invariant mass of the combined system (tagging particle and B) and the sum of the individual rest masses (more in Section A.1.2 of the Appendix). In addition to the OS feature set, the SS taggers exploit variables describing the topological correlation between the tagging track and the B signal track. These include the differences in azimuthal angle $\Delta\phi$ and pseudorapidity

$\Delta\eta$, and the transverse momentum of the combined system B-tagging particle. Before the NN training, all input variables are preprocessed. The features are first standardised to have zero mean and unit variance, and are subsequently transformed using a power transformation to reduce skewness and the impact of long tails in their distributions. This preprocessing improves the numerical stability of the training and ensures that all variables contribute on comparable scales to the optimisation procedure [46].

For each tagger, the dataset is split into a 60% training set and a 40% test set. The training subset is further divided into 80% for actual training and 20% for validation. The validation set is used exclusively for monitoring model performance and guiding early stopping, while the test set is reserved for calibrating the mistag estimate and evaluating the calibrated tagging power. Table 5.5 reports the number of events and tracks used for training each tagger after applying the tagger-specific selection (as described in Section 5.3). The statistics for the validation and calibration sets can be checked in Table A.4 of the Appendix.

Table 5.5: Number of events and associated tracks used for the training of each tagger after applying the tagger-specific selection obtained from the decision tree.

Tagger	**Events**	**Tracks**
OSElectron	$1.5 \cdot 10^6$	$7.2 \cdot 10^6$
OSKaon	$1.6 \cdot 10^6$	$1.1 \cdot 10^7$
OSMuon	$1.3 \cdot 10^6$	$3.6 \cdot 10^6$
SSKaon	$2.2 \cdot 10^5$	$1.5 \cdot 10^6$
SSPion	$2.0 \cdot 10^5$	$3.7 \cdot 10^6$
SSProton	$1.4 \cdot 10^5$	$3.2 \cdot 10^5$

Different NN architectures and hyperparameter combinations are explored by varying:

- the number of hidden layers (2, 3, 8);
- the number of neurons per layer (3, 32, 256);
- the learning rate ($1 \cdot 10^{-4}$, $1 \cdot 10^{-3}$, $1 \cdot 10^{-2}$);
- the batch size (128, 4096, 8192).

Each NN is implemented using the `PyTorch` deep-learning library [106]. All hidden layers use the Exponential Linear Unit (ELU) activation fucntion, and a dropout rate of 0.1 is applied after every hidden layer to reduce overfitting. The loss function is the binary cross-entropy, reflecting the binary nature of the mistag classification task, and the Adam algorithm [107] is used a optimizer. The training process is limited to a maximum of 500 epochs, and an early-stopping patience of 50 epochs is applied to terminate the process once the validation loss no longer improves, avoiding unnecessary computation and reducing the risk of overfitting.

After training the NN, the mistag calibration is performed as described in Section 4.2, using the dedicated calibration sample and the `lhcb-ftcalib` toolkit [102]. This procedure provides the calibrated mistag ω, tagging efficiency, and tagging power, while correctly accounting for events with no valid tagging candidate, assigned $d = 0$ and $\eta = 0.5$. For each model, the mistag is calibrated across multiple configurations, varying both the calibration function (two- or three-parameters) and the link function (identity or logit). Because the quantity of ultimate interest is the tagging power, from the 81 models trained in the grid search, the final model for each tagger is selected as the one with the highest calibrated tagging power. Performing calibration across multiple settings ensures a fair comparison among models and prevents discarding a potentially superior classifier solely because of a suboptimal calibration setup.

5.4.2 Evaluation of the Neural Network Training

After selecting the model with the highest calibrated tagging power for each tagger, the quality of the NN training is verified using a common set of validation plots: the evolution of the training and validation losses, the ROC curves, and the NN output distributions. As an illustration, the SSKaon tagger is discussed in detail below; the corresponding results for the other taggers are provided in Appendix A.2.1.

The left panel of Figure 5.3 shows the training and validation losses as a function of the number of epochs. Both curves decrease during the initial training phase and the validation loss reaches a minimum at epoch 33, indicated by the vertical dashed line. Over the subsequent 50 epochs, corresponding to the chosen patience, no further improvement is observed; instead, the validation loss exhibits an increasing trend, suggesting a tendency towards overfitting. As a result, the early-stopping criterion is triggered, and the model parameters corresponding to the lowest validation loss are retained. The training and validation ROC curves in the right panel of Figure 5.3 provide a complementary view of the classifier performance. The AUC scores are both approximately 52%. Although from machine–learning perspective these values may seem modest, they reflect the intrinsic difficulty of the classification task: as illustrated in Figure 5.4, the input-feature distributions exhibit significant overlap between the two classes, except for $\Delta\phi$, which provides discriminating power.

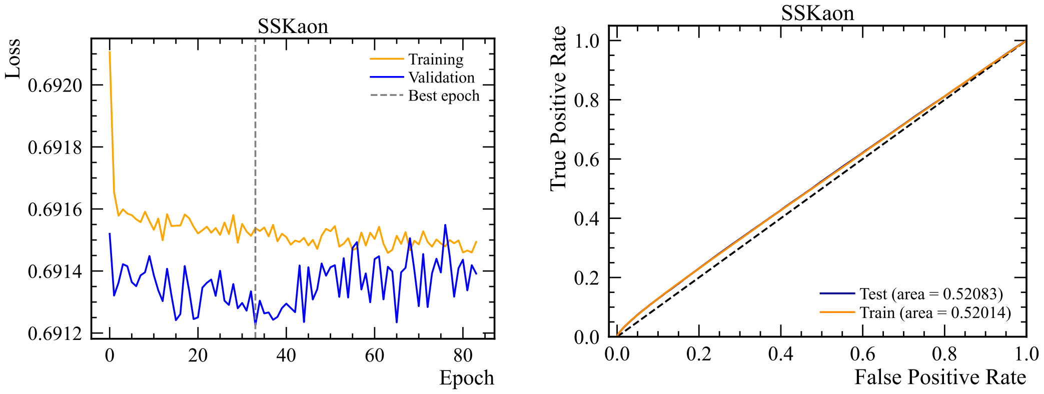


Figure 5.3: Training evaluation for the SSKaon. **Left:** Evolution of the training and validation losses (binary cross-entropy), illustrating the epoch at which early stopping is triggered. **Right:** Receiver Operating Characteristic (ROC) curve evaluated on the training and validation samples (“Train, Test” in the plot), together with the corresponding Area Under the Curve (AUC) score, which quantifies the raw separation achievable before calibration.

As with the input feature distributions, the (1 − NN output) distribution shows substantial overlap between the two classes (Figure 5.5). The pronounced peak near 0.5 is expected, as it corresponds to a random-guess regime where the classifier has limited discriminating power. Nevertheless, a long tail extending towards lower values (down to about 0.1) is clearly visible. In this region, the separation between correct and wrong tagging decisions becomes more apparent, illustrating the residual correlations that the NN learns to exploit between the input variables and the tagging outcome.

The three-parameter calibration curve is shown in the right panel of Figure 5.5, and separately for B_s^0 and $\overline{B}_s^0$ candidates in Figure 5.6. As indicated by the low Δp parameters, the SSKaon tagger provides a symmetric mistag estimate for the flavour states. The obtained calibrated tagging power is $(2.97 \pm 0.07)\%$, demonstrating that, even when the raw classifier separation is intrinsically limited, the combination of the NN output and the calibration procedure still yields a measurable tagging power ϵ_{eff}.

While the SSKaon tagger exhibits no significant flavour–dependent behaviour, both the OSKaon and SSProton taggers show a visible difference between the calibration curves for B^+/B^- (OSKaon) and $B^0/\overline{B}^0$ (SSProton) (Figure 5.7). Since the training sample is constructed to contain equal proportions of B and $\overline{B}$ with equal fractions of correctly and incorrectly tagged tracks, these effects cannot originate from an imbalance in the training labels. Their origin is therefore more

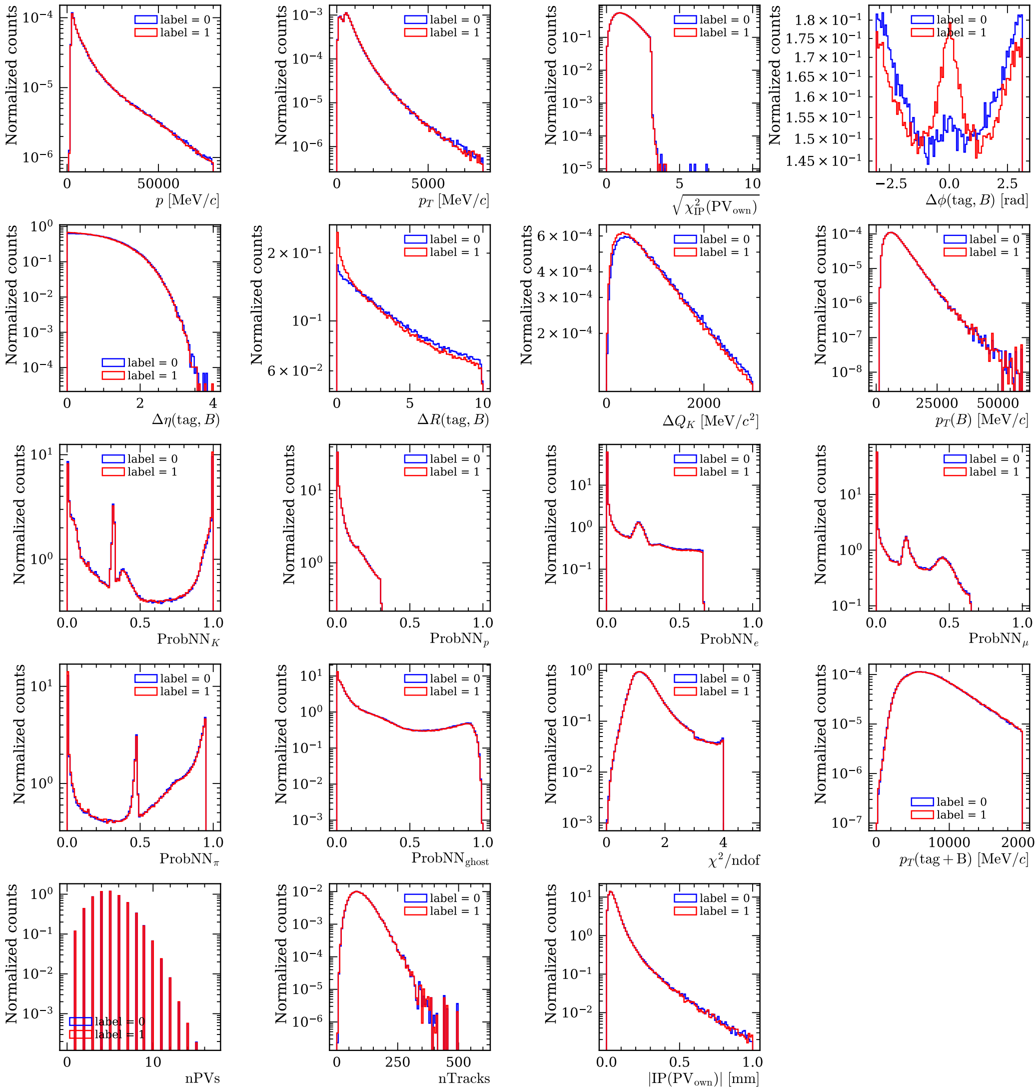


Figure 5.4: Distributions of the input features used for the SSKaon training.

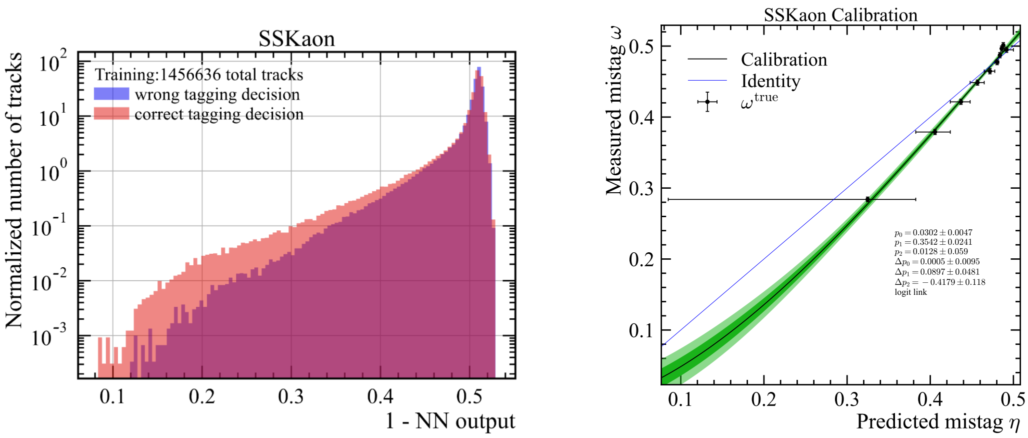


Figure 5.5: NN output distribution and global calibration for the SSKaon tagger. **Left:** Normalised distribution of (1 − NN output) for the SSKaon tagger, shown separately for correctly tagged tracks (blue) and incorrectly tagged tracks (red). A small but systematic separation appears in the high–output region. **Right:** Global calibration curve converting the predicted mistag probability η into the measured mistag ω, evaluated on the independent calibration sample. The green band represents the uncertainty in the calibration fit.

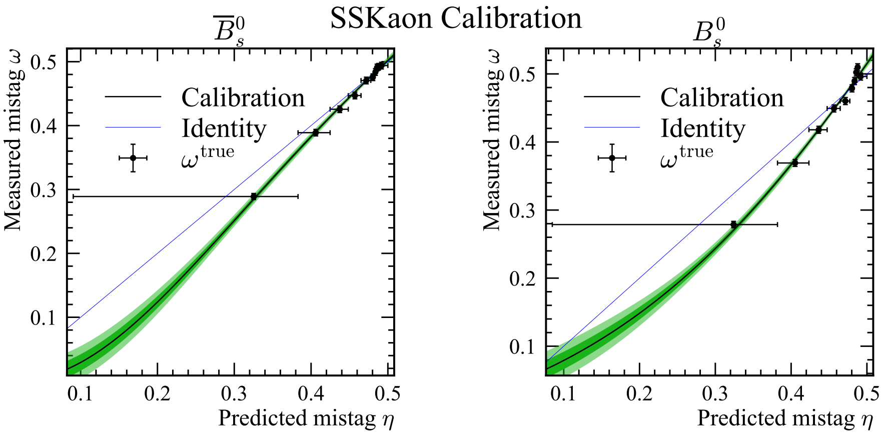


Figure 5.6: Split calibration curves for the SSKaon tagger, shown separately for B_s^0 and $\overline{B}_s^0$ candidates. These plots allow the assessment of possible species–dependent effects in the mistag calibration.

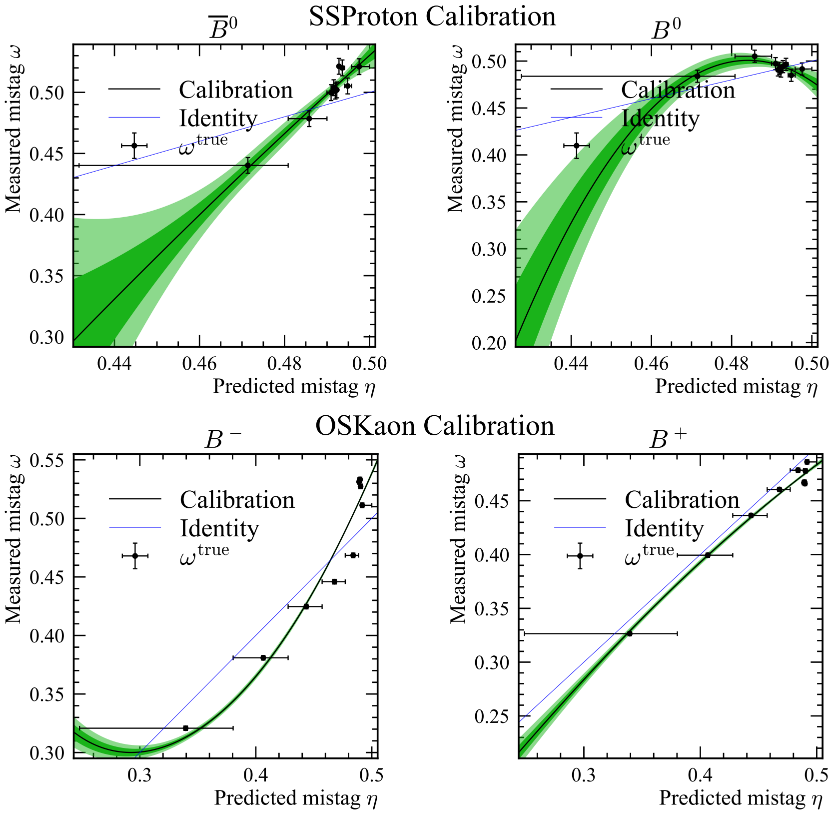


Figure 5.7: Split calibration curves for the SSProton (top) and OSKaon (bottom) taggers. For each tagger, the calibration is shown separately for B and $\overline{B}$ candidates to assess possible flavour-dependent effects in the mistag response.

subtle and likely cannot be attributed to a single cause. A plausible explanation involves a combination of misidentified tagging particles with different underlying charge–flavour correlations and what the NN learns from these mixed samples. The OSKaons majority originates from the decay chain of the *other* B hadron and OS decay chains of B and $\overline{B}$ are not perfectly flavour–symmetric. Differences in the production asymmetry, detector acceptance, and reconstruction efficiencies can lead to small differences in the effective mixture of OS decay topologies entering the OSKaon category for reconstructed B and $\overline{B}$ candidates. This asymmetric mixture could induce species–dependent differences in the mistag estimation and calibration. Furthermore, both the OSKaons and SSProtons are subjected to substantial misidentification from SSKaons and SSPions at tagging-particle selection level, as shown by the confusion matrix in Table 5.3. These misidentified tracks carry opposite charge-flavour correlation (see Section 4.1), and their relative fractions may vary between B and $\overline{B}$ events. In this scenario, the NN could learn different effective mappings between the input features and the mistag probability for the two flavours, leading to the observed differences in the calibration curves. Consistently with this interpretation, no asymmetry is observed for SSKaon and SSPion taggers. These taggers benefit from higher selection efficiencies and, although misidentification occurs between these two categories, their charge correlation with the B flavour follows the *same* assignment rule: a misidentified SSPion can still provide correct flavour information for the SSKaon tagger, and vice-versa. The arguments presented above are intended as a qualitative interpretation rather than a definitive explanation, and the origin of the observed effect is currently under investigation.

Tables 5.6 and 5.7 summarise the final hyperparameter configurations and calibration settings for the six Run 3 taggers, together with their calibrated tagging powers. The values for tagging efficiency, mistag probability, and tagging power before and after calibration are provided in Table A.5 of Appendix A.

Overall, the OSKaon and SSKaon taggers emerge as the most powerful individual algorithms, while the OSElectron and SSProton provide the lowest standalone performance. Nevertheless, all taggers contribute non-trivially to the global flavour tagging information: even relatively weak taggers can add complementary information in events where the dominant taggers fail or are unavailable. The full potential is therefore realised only through their combination. The evaluation of their combination performance is the core subject of the next chapter and provides the foundation for the subsequent discussion of potential avenues for further optimisation.

Table 5.6: Final NN hyperparameter and calibration configurations for the Run 3 OS taggers (OSKaon, OSMuon, OSElectron). The calibration block reports the link function, the number of parameters of the fit function, and the calibrated tagging power.

	OSKaon	OSMuon	OSElectron
NN Hyperparameters			
Learning rate	0.0001	0.001	0.001
Batch size	8192	4096	4096
N. layers	3	3	3
N. neurons per layer	32	256	32
Calibration			
N. parameters	3	3	3
Link function	Logit	Identity	Identity
Calibrated ϵ_{eff} (%)	1.99 ± 0.02	1.047 ± 0.016	0.477 ± 0.011

Table 5.7: Final NN hyperparameter and calibration configurations for the Run 3 SS taggers (SSPion, SSProton, SSKaon). The calibration block reports the link function, the number of parameters of the fit function, and the calibrated tagging power.

	SSPion	SSProton	SSKaon
NN Hyperparameters			
Learning rate	0.001	0.001	0.01
Batch size	4096	128	4096
N. layers	3	3	3
N. neurons per layer	256	3	32
Calibration			
N. parameters	2	3	3
Link function	Identity	Identity	Logit
Calibrated ϵ_{eff} (%)	1.19 ± 0.05	0.070 ± 0.013	2.97 ± 0.07

5.5 Automated Workflow for Flavour Tagging

As discussed in Section 1.2, automation is an essential aspect of this thesis. For the development of the Run 3 taggers, this principle is realised through an automated workflow that orchestrates the full chain, from processing the samples for training each tagger to the inclusion of the tagging information in the samples. The workflow, illustrated in Figure 5.8, is implemented using `Snakemake` [108], with each processing step defined as a separate rule. Each rule specifies well-defined inputs and outputs, and `Snakemake` automatically resolves the task dependencies and determines which rules must be executed, and in which order, using the specified computational resources.

The workflow begins with the processing of the raw samples, which are enriched with the additional features required for the NN training. In a second step, the tagger-specific selections are applied, producing dedicated training datasets for the six taggers: OSKaon, OSMuon, OSElectron, SSPion, SSProton, and SSKaon. In the subsequent stage, the NNs, defined through dedicated configuration files, are trained independently for each tagger and for each hyperparameter configuration. For the relatively lightweight architectures employed in this work, the training time typically ranges from approximately two to nine hours, depending on the tagger. The mistag calibration is performed asynchronously within a separate `Snakemake` workflow, allowing the calibration to start as soon as the trained models become available. Once the hyperparameter scan is completed, the models with the highest calibrated tagging power are extracted (one for each tagger). Finally, the selected taggers are applied to hold-out samples, and the resulting tagging information—namely the tagging decision and the calibrated mistag probability—is attached to the samples for use in the subsequent tagger-combination stage.

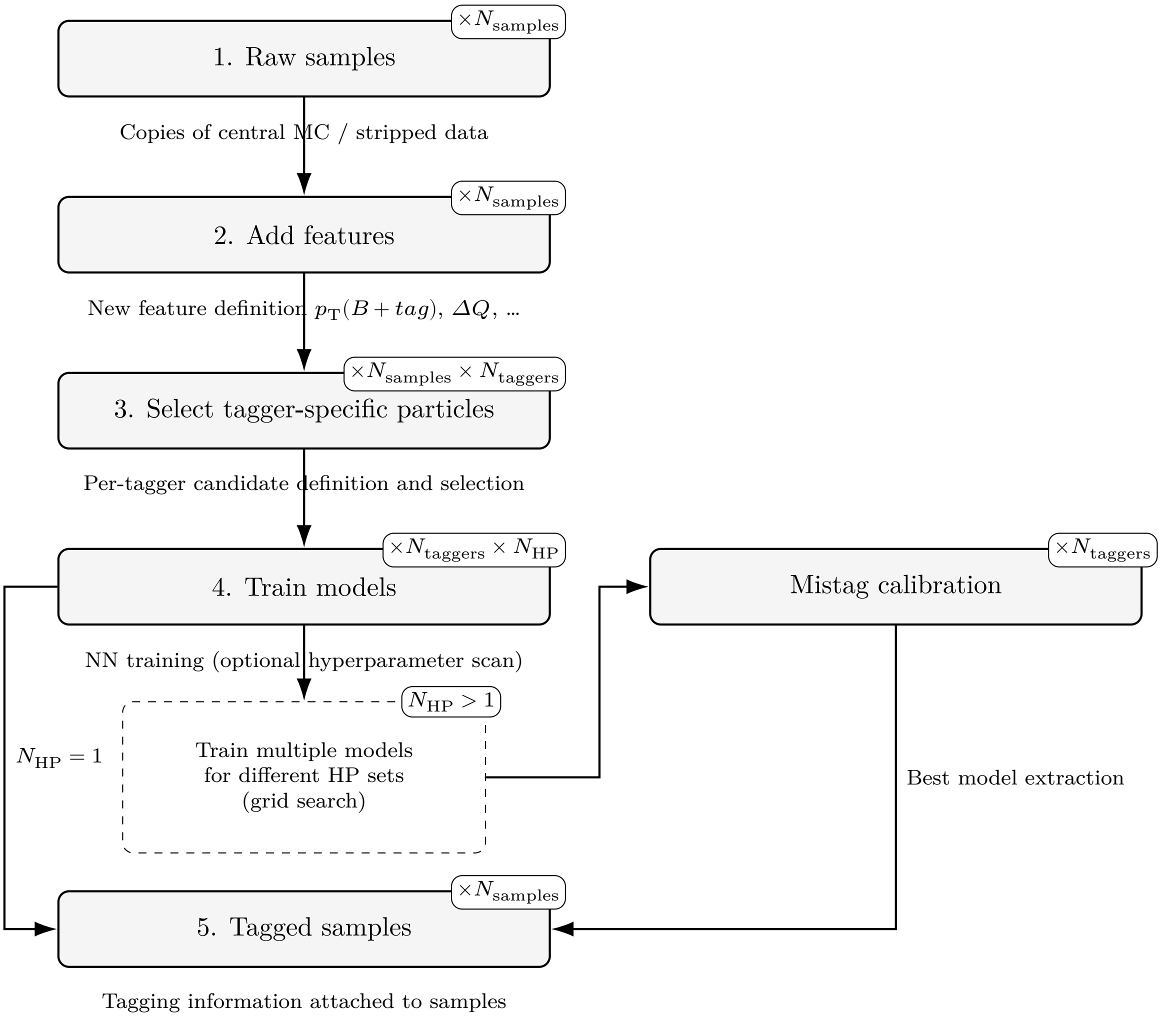


Figure 5.8: Schematic overview of the `Snakemake` workflow used for the development of the Run 3 taggers. Feature construction and tagger-specific particle selections are applied on the input samples. Each tagger is trained separately. In case of a grid search, multiple hyperparameter configurations are given as input (quantified by N_{HP}). N_{HP} models are trained for each tagger and the models achieving the best calibrated tagging power are retained. When a single configuration is used ($N_{\text{HP}} = 1$), the trained model is applied directly to holdout samples and the mistag is calibrated. In both scenarios, the tagging decision and the associated mistag probability are attached to the output samples. The multiplicity labels shown on each block (N_{samples}, N_{taggers}, N_{HP}) indicate how often the corresponding rule is executed.

6 Performance of the Run 3 Taggers

6.1 Introduction

Chapter 5 described the strategy and implementation of the Run 3 flavour-tagging algorithms, resulting in a suite of six taggers: OSKaon, OSMuon, OSElectron, SSKaon, SSPion, and SSProton. These constitute the first flavour-tagging algorithms specifically designed for the upgraded LHCb detector. As discussed in Section 4.3, the outputs of the individual taggers (tagging decision and mistag) can be combined to obtain a single optimised tagging decision. The tagging power of this "combined tagger" represents the performance of an idealised tagger that exploits the full OS and SS information simultaneously.

In Sections 6.2 and 6.3 of this chapter, the performance of the combination of the newly developed taggers (referred to as *Run 3 taggers*) is evaluated on both simulated samples and data collected by the LHCb experiment in 2024, with the simulation corresponding to the same data-taking conditions. The performance of the Run 2 taggers ported to the Run 3 software (referred to as *Benchmark taggers*) is used as a reference. Indeed, despite being optimised for the Run 2 data-taking conditions, these algorithms retain competitive tagging performance when applied to Run 3 data.

Section 6.5 discusses possible avenues for further optimisation of the tagging performance, while Section 6.6 concludes the chapter with a comparison to the performance of the Run 2 taggers on 2018 data.

For notational convenience, references to a meson such as B^0, B_s^0, or B^+ should be understood to include their respective anti-particles ($\overline{B}^0$, $\overline{B}_s^0$, and B^-) as well. Any discussion requiring a distinction between the particle and anti-particle will use explicit notation.

6.2 Combined Tagger Performance on Monte Carlo samples

In this section, the tagging power for the combination of the Run 3 taggers is evaluated calibrating[1] on simulated samples of $B^+ \to J/\psi K^+$ ($N_{events} \approx 2.4 \cdot 10^5$), $B^0 \to J/\psi K^{*0}$ ($N_{events} \approx 6.6 \cdot 10^4$) and $B_s^0 \to D_s^- \pi^+$ decays ($N_{events} \approx 9.4 \cdot 10^4$), produced as described in Section 5.2 and not used for the tagger training. To preserve readability, the calibration plots for the combinations are collected in Figure B.1 of Section B.2.

For the $B^0 \to J/\psi K^{*0}$ channel, the combination includes the OSKaon, OSMuon, OSElectron, SSPion, and SSProton taggers, while for the $B_s^0 \to D_s^- \pi^+$ channel, it includes the OSMuon, OSElectron, OSKaon, and SSKaon taggers. These choices reflect the tagging particles expected to have a strong charge correlation with the signal B flavour. In addition, the $B^+ \to J/\psi K^+$ mode provides a clean environment to evaluate the performance of the OS taggers alone, and to verify that the calibration procedure effectively removes any decay-dependent bias.

The calibrated tagging power values for the single taggers and their combination are reported in Table B.1 of the Appendix. Figure 6.1 summarises the obtained values. Each plot corresponds to a decay mode and shows the calibrated tagging power, ϵ_{eff}, for the individual taggers that contribute to the final combination (Benchmark taggers in blue, Run 3 taggers in red). The horizontal lines with shaded bands represent the calibrated tagging power of the combined tagger and its associated $\pm 1\sigma$ uncertainty, values which are reported in the plot as well. Dashed lines denote the simple sum of the individual tagging powers: if this sum is noticeably larger than the calibrated combination, it indicates that the taggers are overlapping.

It is possible to observe that the Run 3 taggers exhibit larger mutual overlap than the Benchmark taggers in the $B_s^0 \to D_s^- \pi^+$ and $B^+ \to J/\psi K^+$ channels, while showing a smaller overlap in the $B^0 \to J/\psi K^{*0}$ mode. The SS Run 3 taggers tend to underperform relative to their benchmark counterparts. Despite this, the Run 3 tagger combination achieves performance that is overall higher or comparable to that of the Benchmark combination across all decay modes.

Unlike the OSMuon and OSElectron taggers, the OSKaon tagging power shows only modest improvement in $B^+ \to J/\psi K^+$ and remains comparable to the benchmark performance in both $B^0 \to J/\psi K^{*0}$ and $B_s^0 \to D_s^- \pi^+$. Inspection of the calibration parameters reveals that the Δp_3 coefficient is notably large in the $B^+ \to J/\psi K^+$ mode (Figure 6.5a), pointing to a calibration asymmetry between B and $\overline{B}$ mesons.

---

[1] Calibration done using `lhcb_ftcalib` toolset [102]

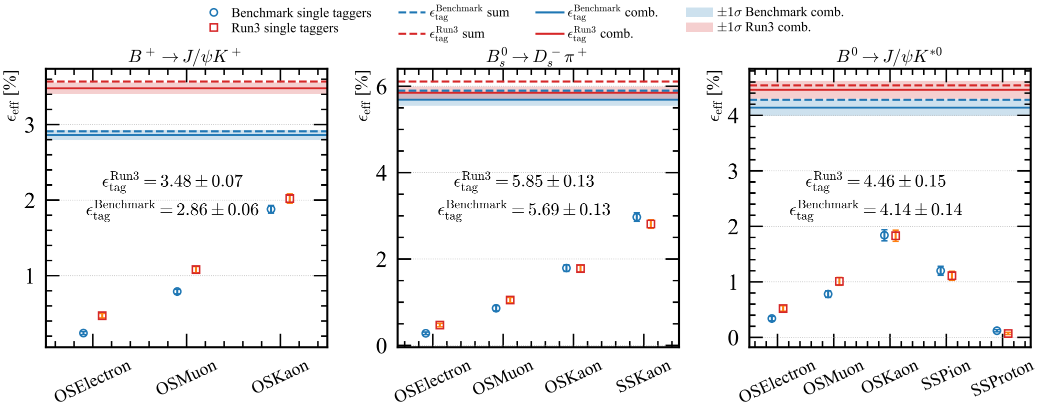
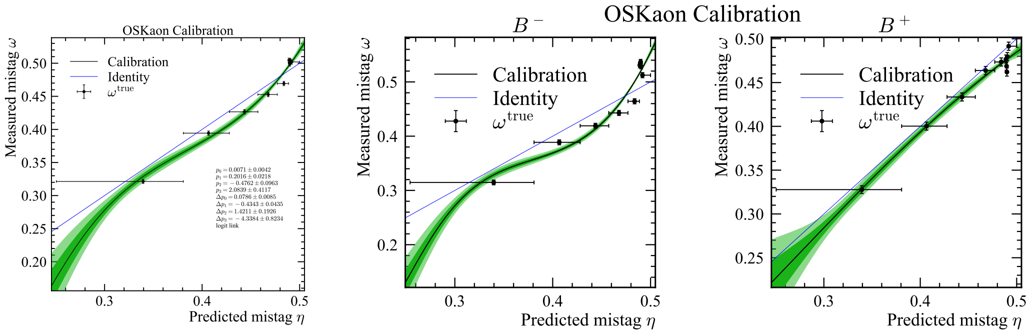


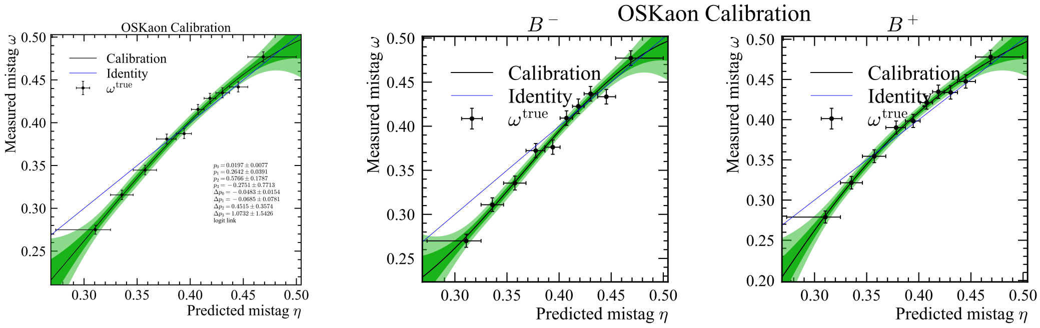
Figure 6.1: Calibrated tagging power ϵ_{eff} for individual and combined taggers, comparing Benchmark (blue) taggers and Run 3 (red) taggers evaluated on the simulated samples for the channels $B^+ \to J/\psi K^+$, $B^0 \to J/\psi K^{*0}$, and $B_s^0 \to D_s^- \pi^+$. The horizontal lines and shaded bands represent the calibrated tagging power of the combined tagger and its $\pm 1\sigma$ uncertainty. Dashed lines indicate the sum of individual tagging powers, showing that the statistical overlap among taggers is minimal.

This behaviour is clearly illustrated in Figure 6.5b, where the B^+ and B^- calibration curves show visibly different responses. This effect is not observed for the Benchmark OSKaon (Figures 6.5c and 6.5d). This asymmetry appears to be specific to the OSKaon tagger and was already evident during the development phase. Possible origins of this behaviour are discussed in Section 5.4.2, and a more detailed investigation is currently ongoing within the LHCb Flavour Tagging Working Group.



(a) Run 3 OSKaon

(b) Run 3 OSKaon: split by B^+/ B^-



(c) Benchmark OSKaon

(d) Benchmark OSKaon: split by B^+/ B^-

Figure 6.2: Calibration of the Run 3 OSKaon (top row) and Benchmark OSKaon (bottom row) in $B^+ \to J/\psi K^+$ simulated samples. **Left**: global calibration curves. **Right**: corresponding flavour-split calibrations, separating the two flavour samples. While Benchmark OSKaon exhibits consistent behaviour between the flavour samples, the Run 3 OSKaon split reveals a clear asymmetry, as indicated by the large Δp_3 fit parameter in the Run 3 calibration.

6.3 Combined Tagger Performance on 2024 data

After evaluating the expected calibrated tagging power on simulated samples, the same combinations of taggers are applied to the corresponding control channels in data collected by the LHCb experiment in 2024 (06–25 August) (details about the samples used in Section 5.2).

The calibration of the taggers on $B_s^0 \to D_s^- \pi^+$, and consequently the validation of the SSKaon tagger on data, is not included in this work. In the case of B_s^0 decays, due to the very fast oscillation frequency, the per-event decay-time uncertainty require a precise determination prior to its use in the flavour-tagging calibration. A dedicated decay-time resolution model for Run 3 data is currently under development within the LHCb collaboration, and, once this model is finalised, a consistent calibration of the SSKaon tagger using B_s^0 decays will be performed.

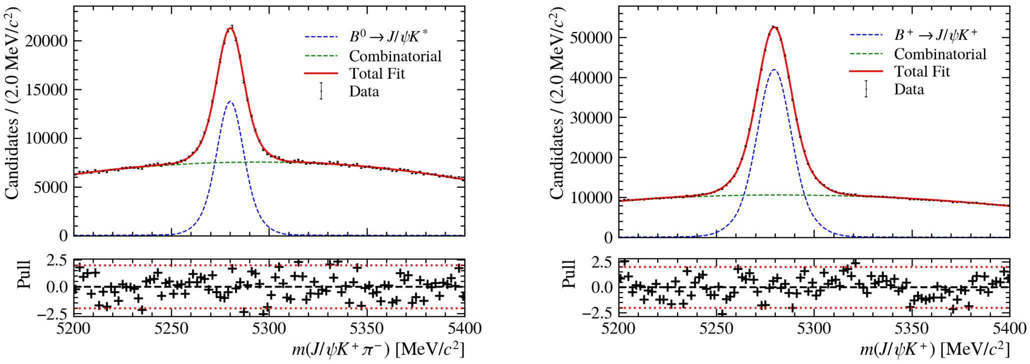


Figure 6.3: Invariant mass distributions for the control channels used in the analysis. **Left:** $B^0 \to J/\psi K^{*0}$ candidates. **Right:** $B^+ \to J/\psi K^+$ candidates. Data are shown as black points, the total fit as a red line, and the signal and combinatorial background components as dashed blue and dashed green lines, respectively.

Before calibrating, background-subtraction weights (*sWeights*) are applied to isolate the signal component [109]. This technique statistically subtracts the combinatorial background and ensures that the measured tagging efficiency and mistag rate correspond to the signal only. The signal yields extracted from the mass fits are $\approx 5.1 \cdot 10^5$ candidates for $B^+ \to J/\psi K^+$ and $\approx 1.3 \cdot 10^5$ for $B^0 \to J/\psi K^{*0}$ (Figure 6.3).

Table 6.1 reports the calibrated tagging power obtained for the Benchmark and Run 3 tagger combinations, as well as for the individual taggers. Last column shows the relative improvement $\Delta_{\%}$ (Run 3 taggers with respect to Benchmark taggers) in terms of calibrated tagging power ε_{eff}

$$\Delta_{\%} = \frac{\varepsilon_{\text{eff}}^{\text{Run 3}} - \varepsilon_{\text{eff}}^{\text{Benchmark}}}{\varepsilon_{\text{eff}}^{\text{Benchmark}}} \times 100 \tag{6.1}$$

For the OS combination, the nominal relative improvement amounts to +5.7% in $B^+ \to J/\psi K^+$ and +12.5% in $B^0 \to J/\psi K^{*0}$. The Run 3 OS calibration in $B^+ \to J/\psi K^+$ yields results consistent with the $B^0 \to J/\psi K^{*0}$ calibration, indicating that the OS calibration exhibits no strong dependence on the decay channel (although a residual dependence due to the specific decay selections cannot be completely excluded).

Table 6.1: Calibrated tagging power ϵ_{eff} (%) for individual taggers (Run 3 and Benchmark) and their combinations on 2024 data (25 June to 6 August). The rows OS, OS+SS, SS report the relative change $\Delta_{\%}$ in calibrated tagging power (Run 3 taggers with respect to Benchmark taggers), together with the uncertainty evaluated in the limiting cases of no correlation ($\rho = 0$, upper value) and full correlation ($\rho = 1$, lower value). These two limits are quoted using a compact asymmetric notation, with the upper and lower uncertainties indicated by the superscripts + and −, respectively.

Decay / Tagger	**Benchmark ϵ_{eff} [%]**	**Run 3 ϵ_{eff} [%]**	$\Delta_{\%}$
$B^+ \to J/\psi K^+$			
OSElectron	0.203 ± 0.015	0.34 ± 0.02	
OSMuon	0.72 ± 0.03	0.82 ± 0.03	
OSKaon	1.01 ± 0.03	0.88 ± 0.03	
OS	$\mathbf{1.92 \pm 0.05}$	$\mathbf{2.03 \pm 0.05}$	$\mathbf{+5.7}^{+3.8}_{-0.15}$
$B^0 \to J/\psi K^{*0}$			
OSElectron	0.24 ± 0.05	0.46 ± 0.08	
OSMuon	0.52 ± 0.07	0.62 ± 0.09	
OSKaon	1.03 ± 0.11	0.89 ± 0.11	
SSPion	1.08 ± 0.08	0.95 ± 0.08	
SSProton	0.33 ± 0.06	0.26 ± 0.04	
OS	$\mathbf{1.76 \pm 0.14}$	$\mathbf{1.98 \pm 0.17}$	$\mathbf{+12.5}^{+13.1}_{-0.64}$
SS	$\mathbf{1.38 \pm 0.10}$	$\mathbf{1.20 \pm 0.09}$	$\mathbf{-13.0}^{+9.8}_{-0.5}$
OS+SS	$\mathbf{3.06 \pm 0.17}$	$\mathbf{3.16 \pm 0.19}$	$\mathbf{+3.3}^{+8.5}_{-0.44}$

A reduced tagging power is observed for the Run 3 SS combination (−13%), but this is not a major concern, since the metric that ultimately matters for physics analyses is the performance of the full OS+SS combination. Indeed, despite the lower SS-only performance, the Run 3 OS+SS still achieve an improvement of approximately +3%.

This behaviour suggests that the Benchmark SS taggers had a non-negligible overlap with the Benchmark OS taggers, whereas the Run 3 design, benefiting from exclusive preselection criteria, suppresses this overlap. Such an effect is clearly visible in Figure 6.4 with basically no separation between red solid and dashed lines in the right-hand plot.

While this reduction in OS–SS overlap is a desirable feature of the Run 3 strategy, the modest net gain in the $B^0 \to J/\psi K^{*0}$ channel indicates that further optimisation of the SS taggers would be beneficial.

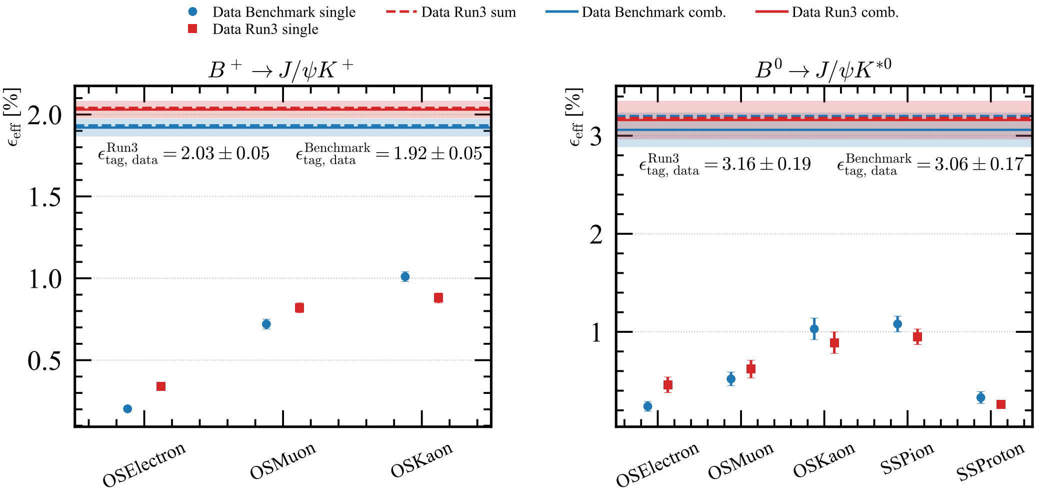
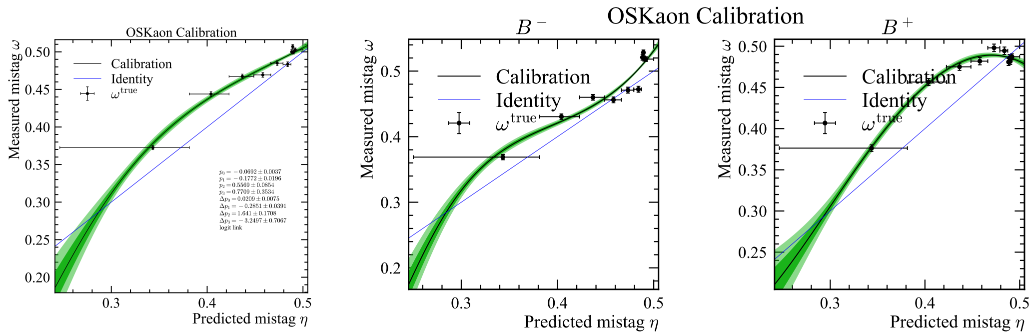


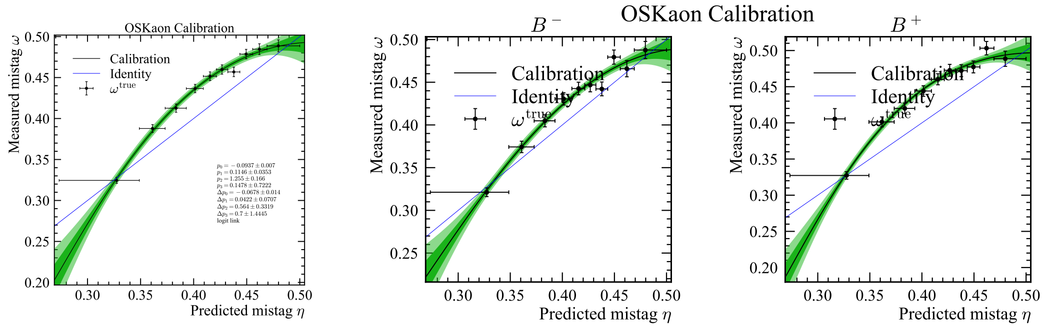
Figure 6.4: Calibrated tagging power ϵ_{eff} for individual and combined taggers, comparing Benchmark (blue) taggers and Run 3 (red) taggers evaluated on the 2024 data, for the channels $B^+ \to J/\psi K^+$ and $B^0 \to J/\psi K^{*0}$. The horizontal lines and shaded bands represent the calibrated tagging power of the combined tagger and its $\pm 1\sigma$ uncertainty. Dashed lines indicate the sum of individual tagging powers.

In addition, Figure 6.4 shows a decrease in the performance of the Run 3 OSKaon tagger. As already note in simulation, the Run 3 OSKaon calibration on data exhibits an asymmetry between B and $\overline{B}$ mesons as well. The calibration curve shown in Figure 6.5a features a large Δp_3 coefficient (-3.25 ± 0.71), indicating a flavour-dependent effect in the mistag response. The flavour-split calibrations in Figure 6.5b further confirm this, revealing different trend for the B^+ and B^- calibration curves. Again, this effect is not observed for the Benchmark OSKaon.



(a) Run 3 OSKaon

(b) Run 3 OSKaon: split by B^+/ B^-



(c) Benchmark OSKaon

(d) Benchmark OSKaon: split by B^+/ B^-

Figure 6.5: Calibration of the Run 3 OSKaon (**top** row) and Benchmark OSKaon (**bottom row**) in $B^+ \to J/\psi K^+$ data. **Left**: global calibration curves. **Right**: corresponding flavour-split calibrations, separating the two flavour samples B^-/B^+. While Benchmark OSKaon exhibits consistent behaviour between the flavour samples, the Run 3 OSKaon split reveals a clear asymmetry, as indicated by the large Δp_3 fit parameter in Figure (a).

6.4 Simulation and Data comparison

An overview of the results presented in Section 6.2 and Section 6.3 is provided in Figure 6.6, showing the calibrated tagging power achieved by the newly developed Run 3 taggers (OSElectron, OSMuon, OSKaon, SSPion, and SSProton), evaluated both on 2024 data and on Monte Carlo samples produced under Run 3 conditions, and compared to their respective Benchmark tagger counterparts. This comparison sets the stage for the discussion that follows, highlighting the gains achieved through the Run 3 retraining as well as the limitations that remain.

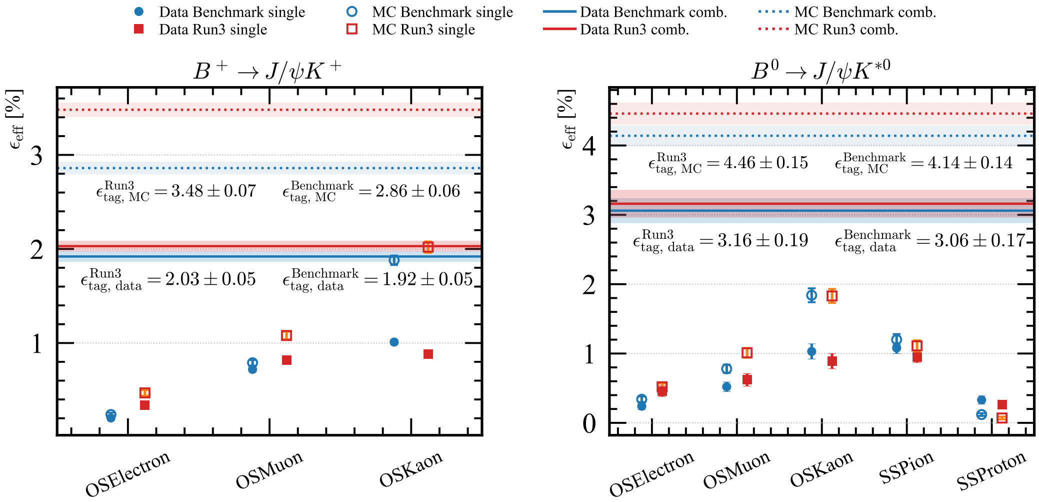


Figure 6.6: Calibrated tagging power $\epsilon_{\rm eff}$ for the combinations of Benchmark and Run 3 taggers, evaluated on Run 3 data (filled markers) and corresponding Monte Carlo (MC) simulated samples (open markers) for the channels $B^+ \to J/\psi K^+$ and $B^0 \to J/\psi K^{*0}$. Solid/dashed horizontal lines and shaded bands represent the calibrated tagging power of the tagger combinations with their associated $\pm 1\sigma$ uncertainties.

A general trend visible in both channels is that the tagging performance observed in simulation is higher than in data. This behaviour can be explained considering that the simulation exhibits lower pile-up and reduced track multiplicities (Figure 6.7), providing a cleaner environment in which flavour tagging performs more effectively, regardless of Benchmark or Run 3 taggers. Moreover, the Run 3 taggers are trained exclusively on simulated samples, which further contributes to their better performance in that domain. At single tagger level, the Run 3 OSKaon is outperformed by its Benchmark counterpart on data, in contrast to the OSMuon and OSElectron taggers, which benefit from the retraining performed under Run 3

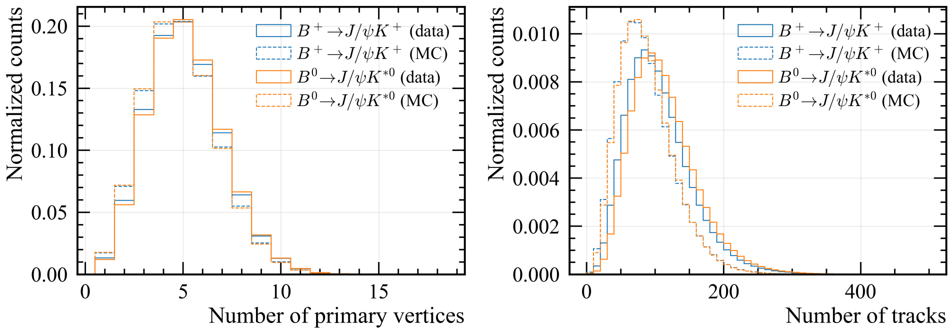


Figure 6.7: Comparison of the number of PVs and the number of tracks distributions for simulated samples (dashed line) and real data (solid line) in the channels $B^0 \to J/\psi K^{*0}$ (orange) and $B^+ \to J/\psi K^+$ (blue).

detector conditions. The Run 3 OSKaon calibration exhibits a pronounced flavour asymmetry (see Figure 6.5b), affecting its performance and indicating that further scrutiny is required. A likely contributing factor is that the Run 3 taggers are trained exclusively on magnet–up samples, as no magnet–down simulated samples were available at the time. Repeating the training with both polarities, as was done for Run 2 tagger development, may help mitigate or eliminate this asymmetry.

A noteworthy observation is that both the Benchmark and Run 3 SSProton taggers achieve higher tagging power on data than on simulation. For the Run 3 SSProton, which is trained exclusively on simulation, this behaviour might look counter-intuitive. A plausible explanation lies in the nature of the tagger: the SSProton primarily acts as a “recovery” tagger, capturing events where the true same-side particle has been misidentified as a proton. Since particle-ID misclassification is generally more frequent in data than in simulation, these recovery opportunities occur more often in real data, which may lead to its enhanced performance.

6.5 Future Prospects

Beyond the recommendations discussed in Section 6.4, the future development of the LHCb flavour-tagging framework should proceed along two complementary directions: the systematic optimisation of the current algorithms, and the exploration of more advanced machine-learning approaches capable of exploiting the full event information.

6.5.1 Optimisation of Current Algorithms

A first line of improvement concerns the refinement of the existing taggers and their associated selections.

Selection Optimisation. Throughout this thesis, it has become clear that the impact of the tagging particle selection is crucial. Particular attention should be given to increasing the OSKaon selection efficiency and purity. Although selection efficiency and purity are naturally in tension, a systematic optimisation including tagger-specific or channel-specific selections could yield measurable gains.

Global Same-Side Tagger. The persistent cross-contamination between pion, kaon, and proton categories in the same-side regime suggests that replacing the separate SSPion, SSKaon, and SSProton taggers with a unified "global" same-side model may be advantageous. Such a tagger could explicitly resolve the pion–kaon–proton ambiguity using a single classifier, thereby reducing the misidentification observed in the current approach.

Feature Set Expansion. Beyond the Run 2-inspired baseline, additional sets of input features should be investigated. Feature-selection techniques such as Principal Component Analysis (PCA) or SHAP-based variable importance could help identify the most discriminating inputs. PCA identifies the dominant directions of variance in the feature space by constructing orthogonal linear combinations of the inputs [110]. Shapley values provide a measure of feature importance by quantifying how much each feature contributes to a model's prediction on average across all possible feature combinations [111].

Hyperparameter Tuning. While the current taggers underwent targeted hyperparameter scans (Section 5.4.1), a wider grid search may identify better-performing configurations and yield further improvements in ϵ_{eff}.

Data-driven retraining. Training the neural networks for the mistag estimation has already proven beneficial for the opposite-side taggers, resulting in a measurable performance gain [112]. This could be extended to SSPion and SSProton.

6.5.2 Advanced Machine Learning Techniques

In parallel to the optimisation of the current algorithms, more advanced machine-learning architectures may be explored, drawing inspiration from recent developments in the HEP-ML communities. These approaches have the potential to capture complex correlations that are difficult to model with simpler neural networks.

Transformer Architectures. Originally developed for natural-language processing, Transformer models [7] have recently gained considerable traction in high-energy physics due to their ability to handle variable-length inputs and to learn long-range correlations through attention mechanisms. Recent studies [113, 114] have demonstrated that Transformers can dynamically focus on the most informative particles in an event, a property highly relevant for flavour tagging where both the identity and quality of the tagging tracks vary significantly from event to event. A notable advantage of attention-based models in the context of flavour tagging is their ability to disentangle the competing information provided by OS and SS tagging candidates. Classical taggers treat OS and SS categories independently and rely on selections to minimise cross-contamination. Transformers, on the other hand, can learn to assign different attention weights to OS-like and SS-like tracks based on their kinematic and topological signatures. This allows the model to suppress misleading candidates automatically while enhancing the contribution of the genuinely informative ones. Moreover, this ability to assign weights is also highly relevant for Primary Vertex association, as the model can learn to suppress tracks based on topological signatures that indicate an origin inconsistent with the B-meson decay vertex.

Inclusive Deep-Set Taggers. A particularly promising avenue is the inclusive tagger based on the DeepSet architecture [115], that can take as input a list of variable length (no fixed number of particles between events) and is invariant under permutation of its inputs (particle order does not matter to infer the B flavour at production time). The inclusive Run 2 tagger [100] DeepSet-based demonstrated substantial performance gains over the classical taggers, improving the tagging power by approximately 35% for B^0 mesons and 20% for B_s^0 mesons when evaluated on Run 2 data. Moreover, as shown in [116], the DeepSet architecture is computationally efficient: it can be trained on a GPU in roughly one hour on statistically representative samples, and achieves an evaluation time of $\mathcal{O}(7\,\mu\text{s})$ per event. Inclusive algorithms naturally leverage the full event record, enabling them to discover correlations that classical taggers cannot exploit. However, this strength also makes them sensitive to bandwidth constraints: using the complete set of event-level information increases the amount of data that must be persisted. For future operation at higher luminosities, it may therefore be advantageous to design hybrid inclusive taggers that learn to focus on a restricted, physics-motivated subset of features, preserving performance while reducing the data footprint.

6.6 Conclusions

Overall, the evaluation of the calibrated tagging power of the Run 3 taggers on both 2024 data and simulation provides an assessment of the current flavour-tagging performance. The results highlight areas of progress, while also identifying aspects where further optimisation is both necessary and promising. In particular, addressing calibration asymmetries, refining pre-selections, improving PV–track association, and pursuing data-driven retraining are expected to yield significant gains. The results also validate the application of the Benchmark taggers to Run 3 data, as their performance remains competitive when compared to the independently developed Run 3 algorithms.

In Figure 6.8, the combination of OSKaon, OSMuon, OSElectron, SSProton and SSPion taggers in the $B^0 \to J/\psi K^{*0}$ channel is compared with the corresponding Run 2 tagger combination evaluated on 2018 data. The Run 2 calibrated tagging power on 2018 data is $(3.95 \pm 0.08)\%$, approximately 25% higher than the value obtained for the Run 3 taggers on Run 3 data $(3.16 \pm 0.19)\%$. While a reduction in Run 3 is expected due to the higher pile-up and multiplicity of the Run 3 data-taking environment [69], a dedicated optimisation of the Run 3 taggers could help reduce the gap. Using the Run 3 OSElectron and OSMuon taggers together with the higher-performing Benchmark OSKaon, SSPion and SSProton taggers already increases the tagging power to $(3.43 \pm 0.19)\%$. This intermediate result suggests that there is space for improvement.

The work presented in this thesis details the development of flavour-tagging algorithms for Run 3 and provides a first evaluation of their performance using benchmark calibration modes. Following the studies described in this chapter, the taggers have been implemented in the LHCb software framework, enabling the tagging decision and mistag probability to be stored directly in the samples for use in physics analyses. A complete validation will be achieved through their deployment in physics analyses and the assessment of their impact on the sensitivity of key measurements, such as the determination of the neutral B-meson oscillation frequency Δm_s. These studies will play a central role in fully exploiting the physics potential of Run 3 data and in guiding further developments towards the Upgrade II era.

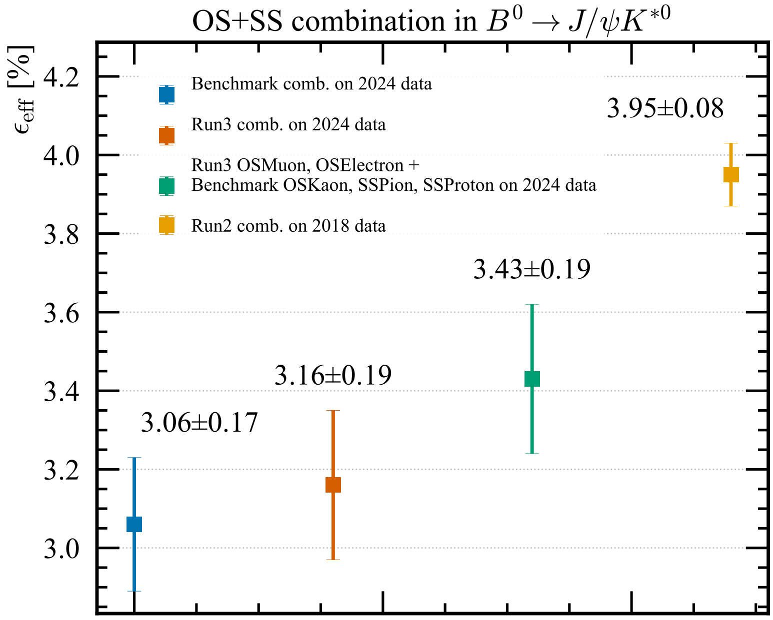


Figure 6.8: Tagging power $\epsilon_{\rm eff}$ for OS+SS tagger combination in $B^0 \to J/\psi K^{*0}$ decay. Results are shown for the combination of the taggers developed in Run 2 and evaluated on 2018 data (orange), and for three combinations evaluated on 2024 Run 3 data, specifically of the: Benchmark taggers (blue), Run 3 taggers (red), and Run 3 OSMuon/OSElectron combined with the Benchmark OSKaon/SSPion/SSProton (green). The Benchmark taggers are the Run 2 taggers ported to the Run 3 software.

7 Financial Transaction Simulation with LLMs at IBM

This chapter describes the work carried out during a three-month secondment at the IBM France Lab in Saclay, conducted within the framework of the SMARTHEP European Training Network. The objective of this project was to develop an open-source framework [117] for the generation of synthetic financial transaction logs, representing both legitimate and fraudulent user behaviour. Such datasets are of interest in fraud detection research, where access to real transaction data is restricted by confidentiality requirements, privacy regulations, and operational constraints.

The framework incorporates two modelling approaches. A baseline simulation is implemented using classical Markov chains, in which user behaviour is modelled by probabilistic state transitions. In parallel, a simulation based on Large Language Models (LLMs) is explored as an alternative approach for generating sequences of financial activities, with the aim of representing behavioural patterns that are not straightforward to encode within a purely Markovian framework. Together, these components provide a means of producing synthetic transaction datasets with configurable parameters, including fraud prevalence, behavioural profiles, and temporal structure, which can be used for development and testing in fraud-detection studies.

7.1 Introduction and Motivations

Financial fraud represents a significant challenge for banks, payment providers, and regulatory organisations. Over recent decades, increasingly complex fraud schemes—such as account takeover, money laundering, and coordinated multi-agent activity—have led to the development of detection systems based on machine learning and artificial intelligence [118–120]. Progress in this area, however, is constrained by the limited availability of realistic transaction datasets. Real-world financial logs are highly sensitive and subject to strict confidentiality and privacy requirements, which restrict their use outside of the operational environments.

Synthetic transaction data is therefore commonly employed in fraud-detection research as an alternative to real data. For such datasets to be useful, they must capture both the diversity of legitimate user behaviour and the evolving nature of fraudulent activity. Many existing data-generation approaches rely on hand-crafted rules or static heuristics, which may not reflect the complex and context-dependent patterns observed in practice. In particular, they often fail to capture adaptive decision-making and long-range temporal dependencies that arise from interactions between multiple agents. This motivates the exploration of simulation frameworks that combine established probabilistic models with more flexible generative approaches, capable of producing coherent sequences of events at the level of individual agents.

Large Language Models are a class of machine-learning models designed to process and generate natural language by learning statistical patterns from large-scale text corpora. Modern LLMs are predominantly based on the transformer architecture [7], which relies on self-attention mechanisms to capture long-range dependencies in sequential data. Through large-scale pre-training on heterogeneous data, these models acquire general representations that can be adapted to a wide range of downstream tasks, including text generation, classification, and structured data extraction [121].

In recent years, LLMs have gained significant attention in both academia and industry due to their strong generalisation capabilities and reduced reliance on task-specific supervision. In industrial settings, they provide a flexible framework for automating complex workflows and supporting scenario-based simulations when access to real data is limited by availability or confidentiality constraints. Within the context of the IBM secondment, LLMs are investigated as a generative component embedded within an agent-based simulation framework. This approach is described in detail in Section 7.3.

7.2 Baseline Simulation with Markov Chains

Markov chains provide a natural tool for modelling user activities as stochastic processes with state transitions governed by a probability matrix [122, 123]. In the developed framework, each user is assigned a behaviour profile characterised by a transition matrix encoding the likelihood of performing one activity after another. These profiles represent broad categories such as *legitimate*, *traveller*, *business user*, or common fraud scenarios like *account takeover*, *money laundering*, *card skimming*. The simulator also incorporates a temporal model in which each behavioural state is associated with a characteristic time scale. For example, legitimate users may exhibit hour- or day-scale separations between transactions, whereas fraudulent actions—such as repeated attempts to exploit compromised credentials—tend to occur at much shorter intervals (minutes).

Transaction amounts are sampled from log-normal distributions, reflecting empirical observations that financial amounts are typically skewed towards small values with occasional large outliers [124]. Legitimate and fraudulent behaviours use different parameterisations to reflect their distinct statistical properties. For example, legitimate transactions are generated with mean $\mu = 5$ and standard deviation $\sigma = 1$, yielding amounts centered around 150 USD, while fraudulent transactions use $\mu = 8$ and $\sigma = 1.2$, producing amounts averaging around 3000 USD. Figure 7.1 illustrates these differences.

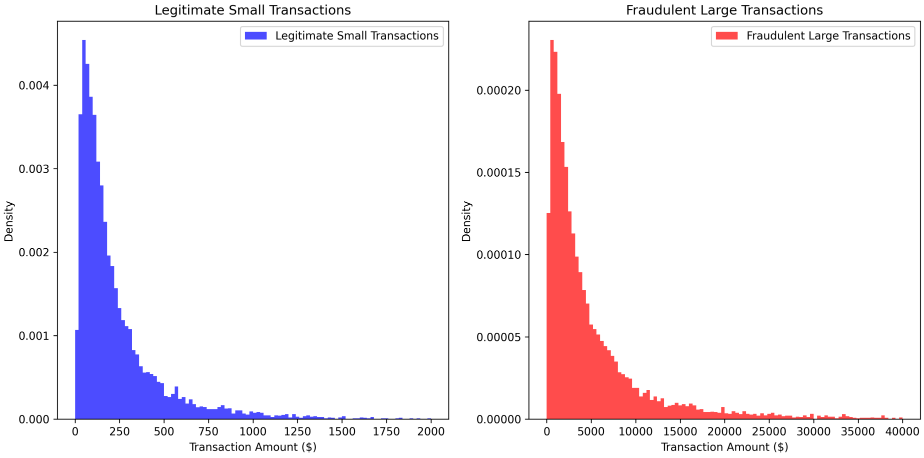


Figure 7.1: Comparison of transaction amount distributions for legitimate (left) and fraudulent (right) activities, sampled from log-normal distributions. Legitimate transactions predominantly involve smaller amounts, while fraudulent transactions exhibit significantly larger transaction values.

Markov-chain-based simulation offers several advantages:

- **Transparency**: Transition matrices are interpretable and easy to validate.
- **Reproducibility**: Simulations behave deterministically for fixed random seeds.
- **Controlled variation**: Researchers can manually adjust transitions or time scales to explore "what-if" scenarios.

Despite their usefulness, Markov chains are fundamentally limited in expressivity. They depend on the Markov assumption—future states depend only on a finite history—making it difficult to model long-range dependencies or adaptive behaviours. Higher-order Markov models partially mitigate this, but the state space grows exponentially. Moreover, domain expertise is required to construct realistic transition matrices, reducing scalability and introducing bias.

These limitations motivate the transition to the LLM-based simulation described in the following section.

7.3 Two-Step LLM-Based Simulation Framework

While Markov-chain-based simulations provide a well-defined baseline for modelling user behaviour, their formulation requires explicit specification of state transitions. As a complementary approach, Large Language Models (LLMs) can be used to generate sequences of actions conditioned on broader contextual information and longer histories. In this work, an LLM-based simulation is explored using a two-step workflow (Figure 7.2) that separates the generation of high-level behavioural descriptions from the generation of individual transaction events.

The workflow consists of the following stages:

1. **Strategy generation**: an LLM produces a high-level behavioural profile, encoded as a structured JSON object, describing typical transaction types, activity frequency, amount ranges, devices, geographical patterns, and additional contextual information.
2. **Activity generation**: a second LLM generates individual transactions sequentially, conditioned on the assigned strategy and on the previously generated simulation state.

The simulation allows the specification of different fraud levels, corresponding to distinct risk environments (e.g. high-, medium-, and low-risk scenarios), as well as the total number of transaction records to be generated. The simulation iterates over a population of agents and assigns a behavioural profile to each user (*strategy*), selected from the collection of strategies produced by the *Strategist LLM* (as described in Section 7.3.1). The simulation assigns fraudulent behaviours until the target fraud level is reached, then it proceeds with legitimate behaviour only. The assigned behaviour is provided as input to the *Agent LLM*, which generates a sequence of activities as described in Section 7.3.2.

The output produced by the *Agent LLM* is subsequently parsed and subjected to a series of validation and correction steps, as described in Section 7.3.3. Once validated, each generated activity is stored as a row in a dataframe. The simulation proceeds iteratively until the desired total number of activities has been reached.

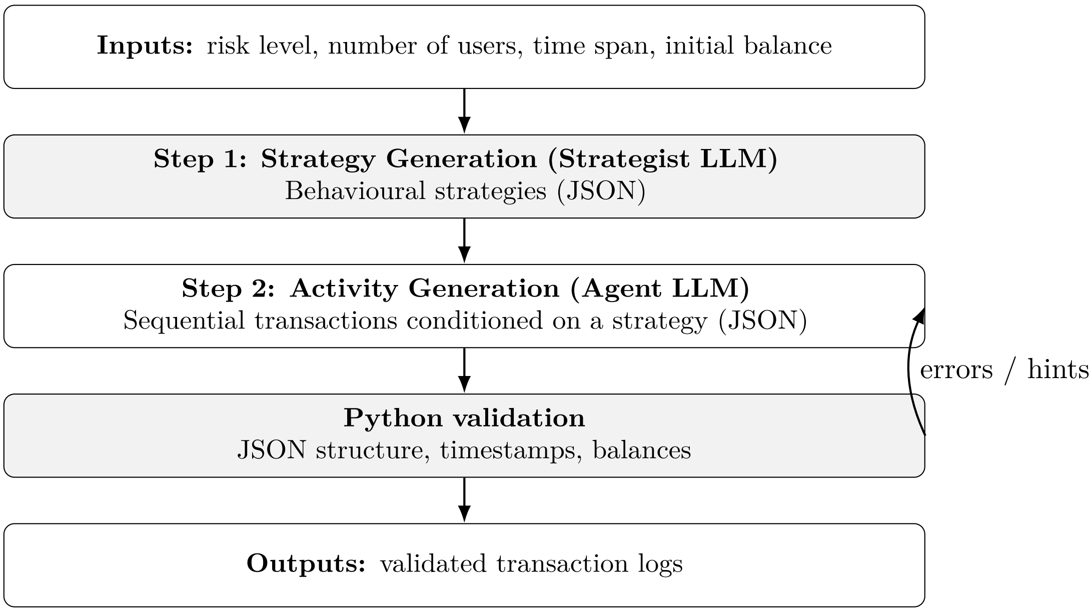


Figure 7.2: Two-step LLM-based workflow. Behavioural strategies are used to condition the generation of transaction sequences, while a Python-based validation step enforces structural and temporal consistency.

A central aspect of the simulation concerns the treatment of time. A global simulation clock is defined, and for each agent, the timestamp of the first activity is randomly initialised relative to this reference. Subsequent activities generated by the *Agent LLM* are produced with timestamps defined relative to the initial event, ensuring temporal consistency within each agent's activity sequence. After the simulation is complete, all activities are merged and ordered chronologically, yielding a combined transaction log rather than user-ordered sequences.

The workflow can be executed either locally, using the `Ollama` framework for running LLMs on local hardware [125], or in a cloud-based setup via IBM Watsonx [126], which supports the use of models with a larger number of parameters and facilitates larger-scale data generation.

7.3.1 Strategy Generation

In the first stage, an LLM—executed either locally (`deepseek-r1:7b` [127]) or in a cloud-based configuration (`llama-3-405b-instruct` [128])—is used to generate behavioural strategies encoded as structured JSON objects. The LLM output is required to be enclosed strictly within the delimiters ```` ```json …```end_json ```` in order to ensure deterministic parsing.

Each generated object defines a dictionary that characterises a user's behaviour through a set of high-level attributes, including:

- the number of accounts associated with a user,
- the types of transactions typically performed,
- expected transaction-amount ranges,
- geographical regions of activity,
- network and device types,
- typical recipients or merchants,
- characteristic time gaps between consecutive transactions.

An example of a strategy corresponding to a cryptocurrency-related fraud profile is shown in Listing 7.1.

```
{
  "profile_or_fraud_type": "Cryptocurrency Fraud",
  "n_accounts": 5,
  "involves_hijacking": true,
  "transaction_types_involved": ["Purchase", "Transfer Out",
    ↪ "Withdrawal"],
  "typical_amount_range": "$100 - $5000",
  "geographic_focus": ["Domestic US", "East Asia", "Southeast Asia"],
  "velocity": "3-5 transactions per day",
  "expected_time_gap": "5-30 minutes between transactions",
  "network_types": ["Wi-Fi", "Cellular", "VPN Connection"],
  "common_devices": ["iPhone 13", "MacBook Pro", "Samsung Galaxy S22"],
  "ip_ranges": ["73.x.x.x", "203.x.x.x", "45.x.x.x"],
  "common_merchant_names": ["Binance", "Coinbase", "Kraken"],
  "common_recipient_ids": ["ACC-774683nf", "ACC-836gfu98",
    ↪ "ACC-467382hf"],
  "common_recipient_banks": ["Bank of America", "Wells Fargo", "HSBC"],
  "context": "Fraudulent use of compromised accounts for cryptocurrency
    ↪ transactions across exchanges and banks."
}
```

Listing 7.1: Example of a generated fraudulent-strategy JSON.

Strategies are generated once prior to the simulation and are subsequently assigned to users during the activity-generation stage.

7.3.2 Activity Generation

A second LLM (*Agent LLM*), an instruction-following model such as `Mistral:7b` ([129]) or `mistral-large`[130], is used to generate sequences of transaction events. For each user, the Agent LLM is prompted with the JSON strategy corresponding to the behavioural profile assigned during the simulation.

The strategy provides the contextual information required for activity generation, including transaction types, amount ranges, temporal characteristics, and environmental attributes. The LLM is instructed to generate transaction records conforming strictly to a predefined schema and to populate all required fields without producing additional free-text content. This constraint is necessary to ensure compatibility with downstream parsing and validation steps.

Each generated activity is described by a set of financial and contextual fields, which together define the structure of the simulated transaction logs. These fields are listed in Table 7.1.

Field	**Explanation**
transaction_id	Unique identifier (UUID).
bank_timestamp	UTC timestamp of the transaction (ISO 8601).
local_timestamp	Local timestamp including timezone offset.
user_id	Unique user identifier.
account_id	Identifier of the involved bank account.
type	Transaction type (e.g. Purchase, Withdrawal, Transfer).
amount	Monetary amount of the transaction.
currency	Currency code (e.g. USD).
balance_before / after	Account balance before and after the transaction.
location	City and country where the transaction occurred.
ip_address	IP address associated with the transaction.
device_id	Device identifier.
network_type	Type of network connection.
merchant_name	Merchant name, or `null` if not applicable.
recipient_id	Recipient identifier for transfer transactions.
recipient_bank	Recipient bank name for transfers.
granted	Boolean flag indicating transaction approval.
fraud_label	Binary indicator of fraudulent activity.
behavior_type	Assigned behavioural profile.

Table 7.1: Fields generated during activity simulation with relative explanation.

7.3.3 LLM Output Validation and Error Handling

As LLM-generated outputs may contain structural inconsistencies or invalid JSON, a dedicated validation layer is implemented to perform a series of systematic checks. These include:

- verification of JSON formatting integrity (e.g. balanced brackets, consistent quoting, valid separators),
- presence and type checking of all required fields,
- consistency of timestamps and account balances,
- rejection of invalid or implausible values (e.g. non-monotonic timestamps, out-of-range amounts, incompatible device, network, or merchant attributes, and inconsistent timezones).

Several correction approaches were explored, including post-processing, sequence regeneration, adaptive prompt reinforcement, and the use of lightweight validator or corrector LLMs. This last approach, tested on local hardware (MacBook Pro, Apple M1 Pro, 16 GB RAM), proved to be impractical due to increased computational overhead, added prompt complexity, and repeated validation–correction cycles that occasionally introduced new structural errors.

For these reasons, a simpler approach based on hard-coded validation and sequence repair was adopted, triggering a sequence regeneration if necessary. The final workflow consists of the following steps:

1. **Generation**: the Agent LLM produces a JSON sequence conditioned on the assigned strategy and simulation parameters.
2. **Parsing**: extraction of the JSON sequence.
3. **Validation**: enforcement of hard constraints on schema, timestamps, balances, and consistency with the assigned strategy.
4. **Deterministic repair**: correction of minor violations where safe.
5. **Re-validation and eventual regeneration**: acceptance if all checks pass, or regeneration.

7.4 Outlook and Future Directions

This chapter presented an open-source framework for the generation of synthetic financial transaction datasets. The framework provides two alternatives: a transparent, Markov-chain-based baseline simulation or a more expressive, LLM-driven approach.

From a practical perspective, the Markov-chain simulation proved to be more stable and computationally efficient. Its deterministic behaviour, low computational cost, and fully controlled output make it a reliable baseline for large-scale data generation. As such, it provides a solid foundation upon which more sophisticated behavioural models can be incrementally introduced. Future improvements could enrich this baseline by incorporating more realistic geographical modelling, for example by assigning transaction locations according to probability distributions centred around a user-specific barycentre, with decreasing likelihood at increasing distances. Such extensions would preserve interpretability while increasing realism and interoperability with downstream analysis tools.

The LLM-based simulation, while more constrained in terms of output reliability, demonstrated a strong potential for generating realistic and diverse behavioural sequences. However, this approach comes with practical limitations. Local execution is constrained by available computational resources and is not well suited for large-scale generation. Cloud-based deployment using IBM Watsonx enabled testing at a larger scale (exceeding 10^3 generated activities), but introduces operational costs associated with token-based usage and external dependencies.

Beyond computational considerations, a key challenge of the LLM-based approach lies in the variability of model outputs, which can occasionally violate structural or semantic constraints and disrupt the simulation pipeline. As observed in this work, the development of robust fallback mechanisms—combining strict validation, deterministic correction, and controlled regeneration—is therefore essential for maintaining the workflow stability.

An open question concerns the evaluation of the realism of the generated datasets. While qualitative inspection suggests that LLM-generated sequences capture plausible behavioural dynamics, defining quantitative realism metrics remains an open research problem. Possible directions include the construction of composite realism scores based on statistical distances, rule-based consistency checks, or LLM-as-a-judge approaches (another LLM is used to evaluate the realism). Realism metrics could be optimised using reinforcement learning techniques, or employed in stress-testing scenarios aimed at probing the limits of fraud-detection algorithms.

Despite these challenges, the use of LLMs for financial transaction simulation appears promising, particularly in the context of developing fraud-detection systems that are robust to emerging and previously unseen fraud schemes. The ability of LLMs to generate adaptive, context-dependent sequences offers a valuable complement to traditional simulation techniques, and motivates further exploration, including domain-specific fine-tuning and extensions to coordinated multi-agent fraud scenarios.

8 Summary and Perspectives

The different application domains addressed in this thesis provided a valuable opportunity for scientific and technical development. Although the individual projects were motivated by distinct objectives, they are connected by a common theme: the exploration of automation as a means of improving scalability and reproducibility in complex data-processing environments.

In the context of the LHCb trigger operations, automation was driven by the need to reduce manual intervention during data taking, a phase that is inherently fast-paced and operationally demanding. The development of an automated workflow for the generation, deployment, and validation of Trigger Configuration Keys (TCKs) contributed to standardising trigger configurations. This standardisation is particularly relevant for offline analyses, as it facilitates the retrieval and interpretation of the trigger conditions under which a given dataset was recorded.

The automation of flavour-tagger training represented a second, complementary application of these principles. By formalising the training procedure within a reproducible workflow, it became possible to scale the development process and to train multiple taggers in parallel. The resulting infrastructure enabled the simultaneous development of six taggers and provided a consistent framework for their optimisation, calibration, and evaluation.

The industrial secondment at IBM extended these ideas beyond the domain of high-energy physics and reinforced the transferability of methodological concepts. In particular, the simulation of financial transaction data shares conceptual similarities with particle reconstruction, in which a sequence of observations (e.g. detector hits or transactions) is interpreted as the manifestation of an underlying process. This analogy motivated the exploration of probabilistic and generative approaches for modelling sequential behaviour in a financial context.

The introduction of Large Language Models (LLMs) within this framework opened additional perspectives. While their use in this thesis was exploratory, it highlighted the potential of such models. This is especially relevant in light of the growing interest in LLMs across multiple scientific disciplines, including physics. Possible applications range from operational support and data quality monitoring to more speculative questions concerning the representation of physical knowledge within data-driven models. At the same time, the results emphasise the need for careful validation and supervision when deploying LLM-based methods. Assessing the reliability, stability, and limitations of such models remain a prerequisite for their adoption in scientific workflows, and automated systems should not replace rigorous downstream testing and expert oversight.

Part III

Supplements

A Supplementary Material for Chapter 5

A.1 Decision Tree for Tagging Particle Selection

A.1.1 Particle Origin Classification

When working with simulated samples, the identification of the physical origin of each reconstructed particle is provided by the `OriginFlag` Monte Carlo variable, which assigns an integer label to every track according to its production process. This classification enables a clear distinction between signal particles, SS and OS, and other sources such as prompt tracks. The `OriginFlag` scheme is illustrated in Figure A.1 and summarised below:

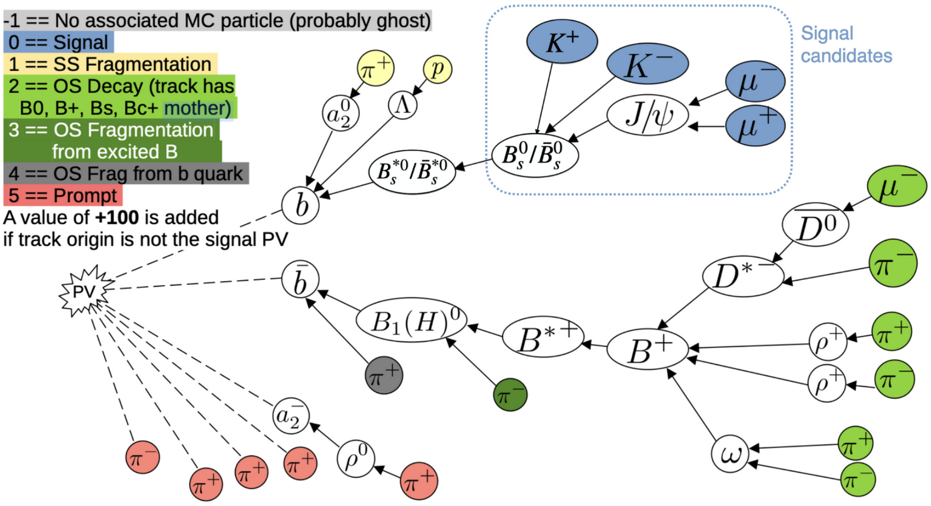


Figure A.1: Schematic representation of the `OriginFlag` classification in a simulated B_s^0 event. Each reconstructed track is assigned an integer value corresponding to its physical origin within the event. The diagram highlights signal particles, same-side (SS) and opposite-side (OS) components, and prompt tracks from the primary vertex. Courtesy of *Vukan Jevtić* (LHCb)

0: **Signal particles.** Particles originating from the decay chain of the signal meson, e.g. the muons and kaons from $B_s^0 \to J/\psi\,\phi$.

1: **SS Fragmentation.** Particles produced from the hadronisation of the same quark jet that contains the signal meson. These are not part of the signal decay itself but are correlated with its flavour.

2: **OS Decay.** Particles produced in the decay chain of the B hadron on the opposite side of the event (such as kaons or leptons from semileptonic decays).

3: **OS Fragmentation from an excited B meson.** Particles originating from the decay of an excited B meson on the opposite side.

4: **OS Fragmentation from a b.** Particles directly produced in the fragmentation process of the opposite-side b.

5: **Prompt particles.** Tracks originating directly from the primary interaction vertex.

A constant value of +100 is assigned to the tracks that do not originate from the same primary vertex as the signal candidate. This classification provides the foundation for identifying the tagging particles.

A.1.2 Input Features

To follow, all variables used for decision tree training are listed, along with their naming conventions and explanations.

Table A.1: Input features used for training the decision tree that identifies the tagging particle for the Run 3 classical taggers (OSKaon, OSMuon, OSElectron, SSPion, SSProton). If a variable refers to the B signal, it is indicated with B. Tagging track refers to the tagging particle track (or a potential tagging particle if before applying selections).

Variable	**Description**
	Particle Identification (PID) Features
ProbNN$_{K,p,e,\mu,\pi}$	A family of particle PID probabilities. Each ProbNN is built using multivariate techniques by combining tracking and PID information for a specific particle hypothesis (e.g., Kaon, Proton, Electron, etc.).
PID$_K$	Kaon PID response.
PID$_e$	Electron PID response.
PID$_\mu$	Muon PID response.

Variable	**Description**
PID_p	Proton PID response.
isMuon	Muon identification flag.
P_{ghost}	Low-level ghost probability.
Kinematic Features (Tagging Candidate)	
p [MeV/c]	Total momentum.
p_T [MeV/c]	Transverse momentum.
E [MeV]	Energy.
η	Pseudorapidity.
p_x, p_y, p_z [MeV/c]	Momentum components.
$\arctan(p_T/p_z)$	Polar-angle estimator.
$p_T(\mathrm{tag} + B)$ [MeV/c]	Transverse momentum of the system (tagging particle, B).
Vertexing and Spatial Association	
$\mathrm{IP}(\mathrm{PV}_{\mathrm{own}})$ [mm]	IP w.r.t. best own PV.
$\chi^2_{\mathrm{IP}}(\mathrm{PV}_{\mathrm{own}})$	IP χ^2 w.r.t. best own PV.
$\lvert\mathrm{IP}(\mathrm{PV}_{\mathrm{own}})\rvert$ [mm]	Absolute IP w.r.t. best own PV.
$\sqrt{\chi^2_{\mathrm{IP}}(\mathrm{PV}_{\mathrm{own}})}$	IP significance w.r.t. best own PV.
$\mathrm{IP}(\mathrm{PV}_{\mathrm{best}}(B))$ [mm]	IP w.r.t. best PV of the B signal.
$\chi^2_{\mathrm{IP}}(\mathrm{PV}_{\mathrm{best}}(B))$	IP χ^2 w.r.t best PV of the B signal.
MINIP [mm]	Minimum IP w.r.t. any PV.
χ^2_{MINIP}	χ^2 associated with MINIP.
EVIP [mm]	Distance from the PV to the point of the track that is the closest to the B decay vertex.
log(EVIP)	Logarithm of EVIP.
$\Delta z(\mathrm{PV}_{\mathrm{own}}, PV_B)$ [mm]	Difference in z position between own best PV and best PV of the B signal (absolute value).
$\Delta\eta(\mathrm{tag}, B)$	Difference in pseudorapidity between tagging track and B signal (absolute value).

Variable	**Description**
$\Delta R(\text{tag}, B)$	Separation in the η–ϕ plane between tagging track and B signal.
$\Delta\phi(\text{tag}, B)$ [rad]	Azimuthal angle difference between tagging track and B signal computed as described in ††.
$\cos\Delta\phi(\text{tag}, B)$	Cosine of the azimuthal angle difference between tagging track and B signal.
$\min\Delta\phi(\text{tag}, B)$	Minimum azimuthal separation between tagging track and B signal.
Kinematic Differences	
Δp [MeV/c]	Momentum difference between B signal and tagging track.
ΔQ_π [MeV/c^2]	ΔQ under pion mass hypothesis †.
ΔQ_μ [MeV/c^2]	ΔQ under muon mass hypothesis †.
ΔQ_e [MeV/c^2]	ΔQ under electron mass hypothesis †.
ΔQ_p [MeV/c^2]	ΔQ under proton mass hypothesis †.
ΔQ_K [MeV/c^2]	ΔQ under kaon mass hypothesis †.
B Signal Properties	
$E(B)$ [MeV]	Energy
$\eta(B)$	Pseudorapidity.
$m(B)$ [MeV/c^2]	Invariant mass.
$p(B)$, $p_T(B)$ [MeV/c]	Momentum and transverse momentum.
$p_x(B), p_y(B), p_z(B)$ [MeV/c]	Momentum components.
$\phi(B)$ [rad]	Azimuthal angle.
$x_{\text{PV}}(B), y_{\text{PV}}(B), z_{\text{PV}}(B)$ [mm]	Coordinates of B PV.
$x_{\text{end}}(B), y_{\text{end}}(B), z_{\text{end}}(B)$ [mm]	B Decay–vertex coordinates.
nPVs	Number of primary vertices.
nTracks	Number of reconstructed tracks in the event.
Tagging Track Properties	
isLong	Track type flag (Long track or not).
x, y, z [mm]	Spatial coordinates.
χ^2/ndof	Track–fit χ^2 per degree of freedom.

†The variable ΔQ represents a mass difference, computed under the hypothesis that the tagging track is a specific particle species $X \in \{\pi, K, p, e, \mu\}$. It quantifies how much the invariant mass of the B candidate would change if it were combined with the tagging track, assumed to have mass M_X. The ΔQ_X is defined as:

$$\Delta Q_X = \sqrt{(E_X + E_B)^2 - |\vec{p}_X + \vec{p}_B|^2} - M_B - M_X,$$

where:

- M_X is the assumed mass of the tagging particle,
- $\vec{p}_X = (p_x^X, p_y^X, p_z^X)$ is the 3-momentum of the tagging track,
- $E_X = \sqrt{M_X^2 + |\vec{p}_X|^2}$ is the energy under that mass hypothesis,
- $\vec{p}_B$ and E_B are the momentum and energy of the B candidate,
- M_B is the invariant mass of the B candidate.

††The azimuthal separation between the tagging particle and the B candidate is quantified through the variable $\Delta\phi$. Since the azimuthal angle is a periodic coordinate, the separation is computed as the *minimum* angular distance on the circle. This is obtained by evaluating the signed difference using trigonometric identities rather than a simple subtraction, ensuring a well-defined value in the interval $[-\pi, \pi]$ even when the angles lie on opposite sides of the 2π boundary. Explicitly, if ϕ_{tag} and ϕ_B denote the azimuthal angles of the tagging track and the B candidate, respectively, the minimal angular difference is defined as

$$\Delta\phi = \text{atan2}\left(\sin\bigl(\phi_{\text{tag}} - \phi_B\bigr), \cos\bigl(\phi_{\text{tag}} - \phi_B\bigr)\right).$$

This convention ensures a continuous and physically meaningful measure of azimuthal separation, independent of the branch cut of the raw angles.

A.1.3 Output Schema

Figure A.4 shows the schema of the decision tree trained to obtain the Run 3 cuts used to select the different tagging particles.

A.1.4 Final cuts

The decision tree cuts — corresponding to the terminal leaves of the decision tree — are summarised in Table A.2, while the full structure of the trained model is provided in Figure A.4. A further condition is applied to exclude particles originating from the signal decay chain (OriginFlag $\neq 0$) [1].

---

[1] In data, the equivalent is IsInTree $\neq 1$

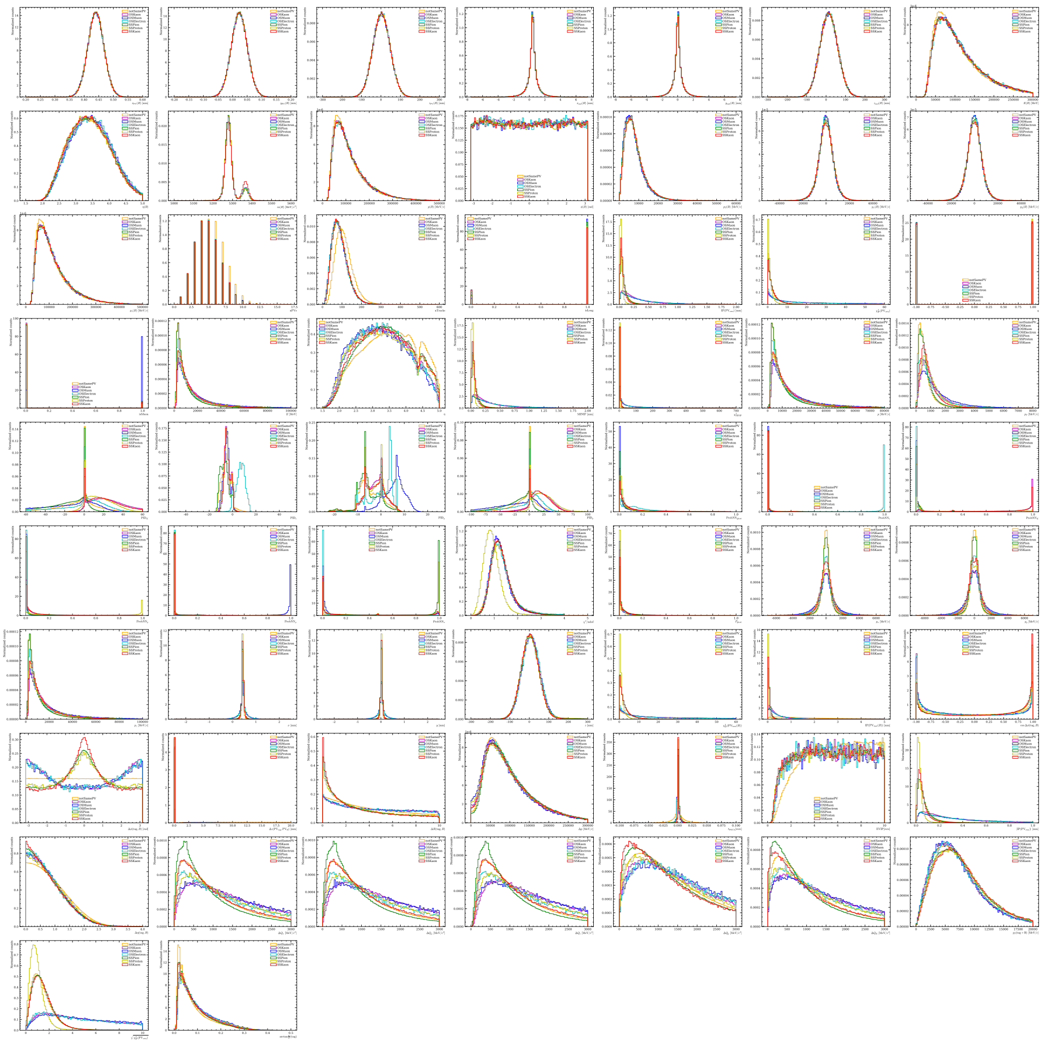


Figure A.2: Decision tree (DT) input feature distributions. The DT is trained to identify the tagging particle (OSKaon, OSMuon, OSElectron, SSPion, SSProton, SSKaon). The label ”tag” refers to the tagging track. The label B refers to the B signal (signal). The class notSamePV identifies particles belonging to a different PV than the B signal.

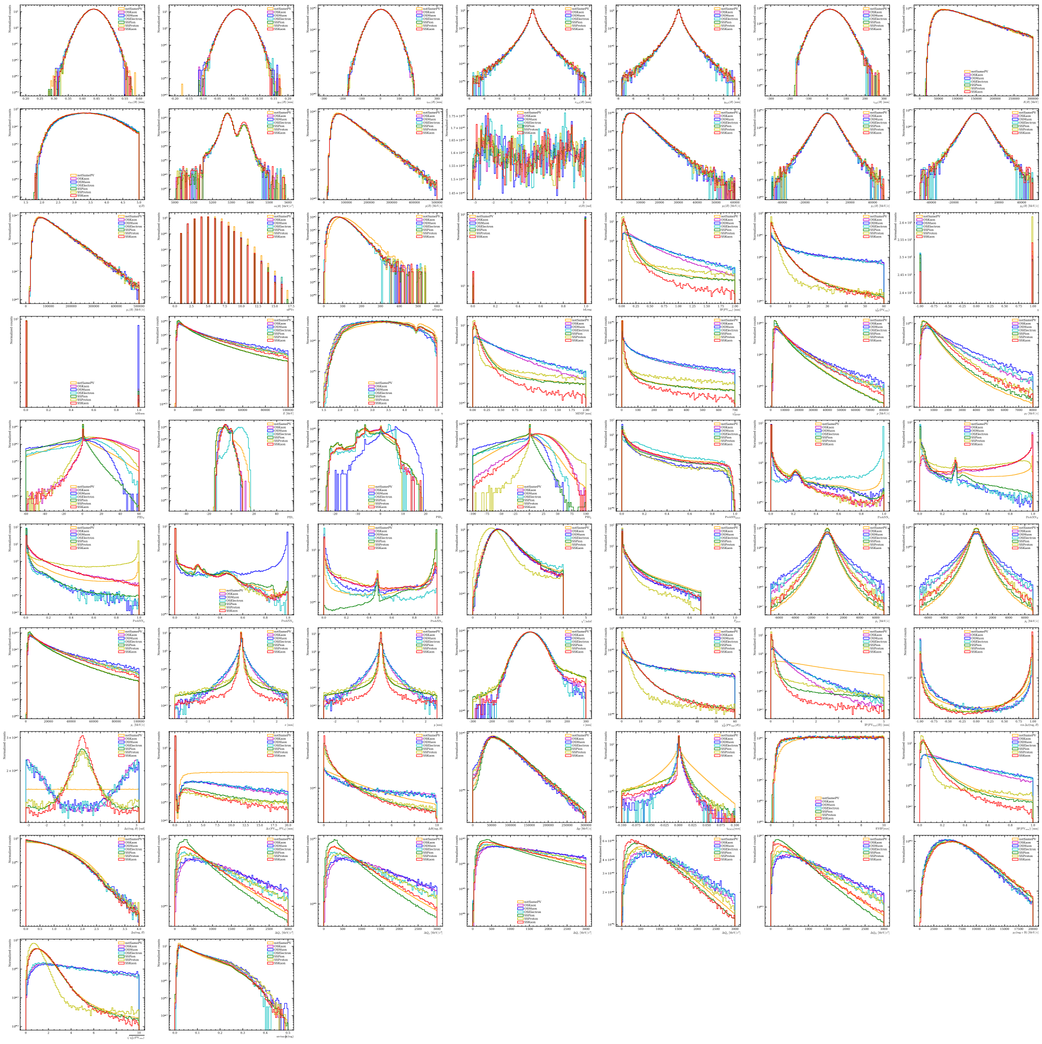


Figure A.3: Decision tree (DT) input feature distributions in logarithmic scale. The DT is trained to identify the tagging particle (OSKaon, OSMuon, OSElectron, SSPion, SSProton, SSKaon). The label ”tag” refers to the tagging track. The label B refers to the B signal (signal). The class notSamePV identifies particles belonging to a different PV than the B signal.

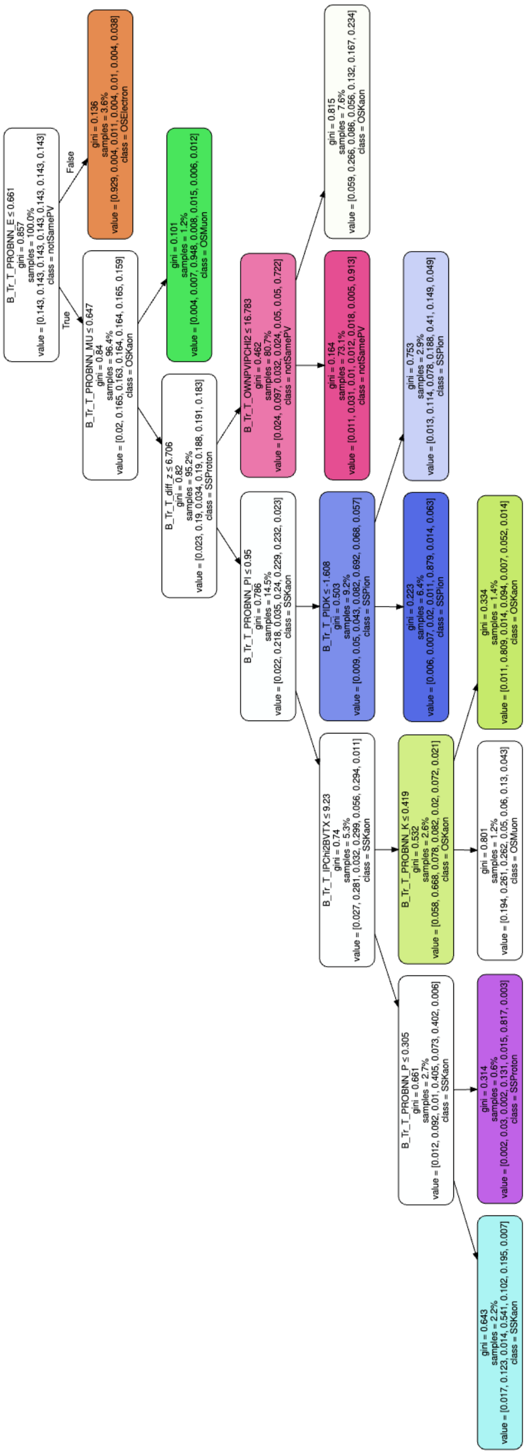


Figure A.4: Full structure of the decision tree trained for tagging particle selection in Run 3. Decays used: $B^+ \to J/\psi K^+$, $B^0 \to J/\psi K^{*0}$, $B_s^0 \to D_s^- \pi^+$. The obtained selections are used to select the identify the tracks associated to each tagger.

Table A.2: Selection cuts for tagging particles obtained by training a decision tree with balanced classes, depth set to 6, minimum impurity decrease set to 0.009.

Tagging Particle	**Relative Cut**
OSElectron	(ProbNN$_e$ $>$ 0.6606) & (OriginFlag !=0)
OSKaon	((ProbNN$_e$ <= 0.6606) & (ProbNN$_\mu$ <= 0.6472) & ($\mathit{\Delta}z$ <= 6.7063) & (ProbNN$_\pi$ <= 0.9496) & ($\chi^2_{\mathrm{IP}}(tag, \mathrm{PV}_{\mathrm{best}}(B))$ $>$ 9.2305) & (ProbNN$_K$ $>$ 0.4192) & (OriginFlag !=0)) | ((ProbNN$_e$ <= 0.6606) & (ProbNN$_\mu$ <= 0.6472) & ($\mathit{\Delta}z$ $>$ 6.7063) & ($\chi^2_{\mathrm{IP}}(\mathrm{PV}_{\mathrm{own}})$ $>$ 16.7833) & (OriginFlag !=0))
OSMuon	((ProbNN$_e$ <= 0.6606) & (ProbNN$_\mu$ <= 0.6472) & ($\mathit{\Delta}z$ <= 6.7063) & (ProbNN$_\pi$ <= 0.9496) & ($\chi^2_{\mathrm{IP}}(tag, \mathrm{PV}_{\mathrm{best}}(B))$ $>$ 9.2305) & (ProbNN$_K$ <= 0.4192) & (OriginFlag !=0)) | ((ProbNN$_e$ <= 0.6606) & (ProbNN$_\mu$ $>$ 0.6472) & (OriginFlag !=0))
SSPion	((ProbNN$_e$ <= 0.6606) & (ProbNN$_\mu$ <= 0.6472) & ($\mathit{\Delta}z$ <= 6.7063) & (ProbNN$_\pi$ $>$ 0.9496) & (PID$_{\mathrm{K}}$ <= -1.6081) & (OriginFlag !=0)) | ((ProbNN$_e$ <= 0.6606) & (ProbNN$_\mu$ <= 0.6472) & ($\mathit{\Delta}z$ <= 6.7063) & (ProbNN$_\pi$ $>$ 0.9496) & (PID$_{\mathrm{K}}$ $>$ -1.6081) & (OriginFlag !=0))
SSProton	(ProbNN$_e$ <= 0.6606) & (ProbNN$_\mu$ <= 0.6472) & ($\mathit{\Delta}z$ <= 6.7063) & (ProbNN$_\pi$ <= 0.9496) & ($\chi^2_{\mathrm{IP}}(tag, \mathrm{PV}_{\mathrm{best}}(B))$ <= 9.2305) & (ProbNN$_p$ $>$ 0.3049) & (OriginFlag !=0)
SSKaon	(ProbNN$_e$ <= 0.6606) & (ProbNN$_\mu$ <= 0.6472) & ($\mathit{\Delta}z$ <= 6.7063) & (ProbNN$_\pi$ <= 0.9496) & ($\chi^2_{\mathrm{IP}}(tag, \mathrm{PV}_{\mathrm{best}}(B))$ <= 9.2305) & (ProbNN$_p$ <= 0.3049) & (OriginFlag !=0)
notSamePV	(ProbNN$_e$ $>$ 0.6606) & ($\chi^2_{\mathrm{IP}}(\mathrm{PV}_{\mathrm{own}})$ $>$ 16.7833)& (ProbNN$_K$ $>$ 0.4192) & (OriginFlag 0)

A.2 Neural Network for Mistag Estimation

A.2.1 Tagger Training and Validation Plots

This section contains the full set of training and validation plots for each of the flavour taggers developed in this work. For each tagger, the plots show the training loss, the ROC curve, the neural network output distributions, and the calibration on a hold-out test sample (for both: distinguishing by B species and merged calibration).

Table A.3: Input features used for OS and SS taggers, with brief explanations. Unless specified, all the variables refer to the tagging particle.

Feature	**Explanation**
nTracks	Number of reconstructed tracks in the event.
nPVs	Number of reconstructed primary vertices.
p	Momentum magnitude.
p_{T}	Transverse momentum.
$p_{\mathrm{T}}(B)$	Transverse momentum of the B signal.
$p_{\mathrm{T}}(B + tag)$	p_T of the B–tagging particle system.
$\sqrt{\chi^2_{\mathrm{IP}}(\mathrm{PV_{own}})}$	Impact-parameter significance w.r.t. the own best PV.
χ^2/ndf	Track-fit χ^2 per degree of freedom.
$\mathrm{ProbNN}_{K,p,e,\mu,\pi}$	A family of particle PID probabilities. Each ProbNN is built using multivariate techniques by combining tracking and PID information for a specific particle hypothesis (e.g., Kaon, Proton, Electron, etc.).
GhostProb	Probability that the track is a ghost (fake).
$\lvert\mathrm{IP}(\mathrm{PV_{own}})\rvert$	Absolute impact parameter w.r.t. the own best PV.
$\Delta\phi$	Azimuthal angle difference between tagging track and B.
$\Delta\eta$	Difference in pseudorapidity between tagging track and B.
ΔR	Distance in η–ϕ space between tagging track and B.
$\Delta Q_{K,p,e,\mu,\pi}$	ΔQ computed assuming a specific particle mass hypothesis.

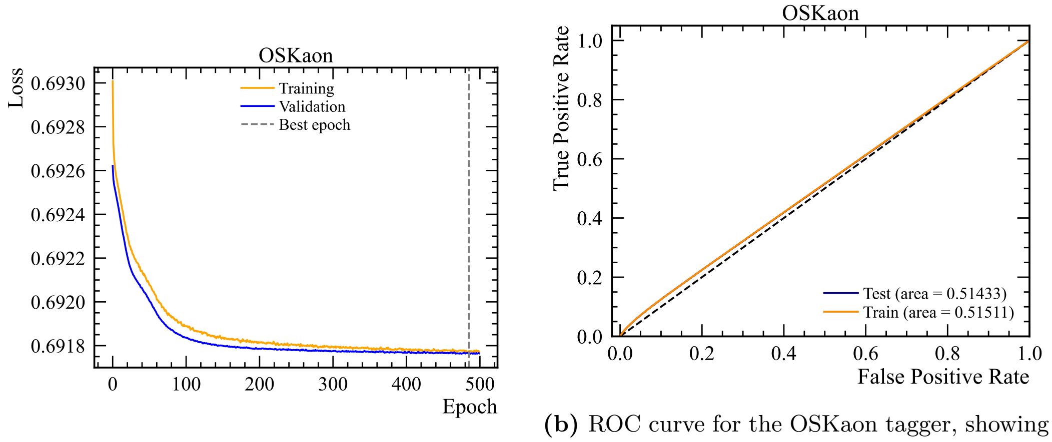
Table A.4: Breakdown of the number of events and associated tracks for each Run 3 tagger at various stages of the development pipeline. The "Before selection" and "After selection" columns reflect the counts before and after the application of the DT-based exclusive cuts designed for that specific tagger. The "After selection" sample is then split: 60% is used for the combined Training (80%) and Validation (20%) sets, and 40% is reserved for the calibration sample. The "Calibration (only selected)" column refers to the 40% sample with the DT-cut applied, while "Calibration (total)" represents the entire initial 40% sample used for calibration, regardless of whether a tagging particle was successfully selected. OS taggers and SS taggers (pion, proton) share the same initial event statistics as they originate from common merged simulation samples.

	Before selection		**After selection**		**Training**		**Validation**		**Calibration (sel.)**		**Calibration (total)**	
Tagger	Events	Tracks	Events	Tracks	Events	Tracks	Events	Tracks	Events	Tracks	Events	Tracks
OSElectron	$3.4 \cdot 10^6$	$3.0 \cdot 10^8$	$3.2 \cdot 10^6$	$1.5 \cdot 10^7$	$1.5 \cdot 10^6$	$7.2 \cdot 10^6$	$3.8 \cdot 10^5$	$1.8 \cdot 10^6$	$1.3 \cdot 10^6$	$6.1 \cdot 10^6$	$1.4 \cdot 10^6$	$1.2 \cdot 10^8$
OSKaon	$3.4 \cdot 10^6$	$3.0 \cdot 10^8$	$3.3 \cdot 10^6$	$2.4 \cdot 10^7$	$1.6 \cdot 10^6$	$1.1 \cdot 10^7$	$3.9 \cdot 10^5$	$2.8 \cdot 10^6$	$1.3 \cdot 10^6$	$9.7 \cdot 10^6$	$1.4 \cdot 10^6$	$1.2 \cdot 10^8$
OSMuon	$3.4 \cdot 10^6$	$3.0 \cdot 10^8$	$2.8 \cdot 10^6$	$7.9 \cdot 10^6$	$1.3 \cdot 10^6$	$3.6 \cdot 10^6$	$3.3 \cdot 10^5$	$9.1 \cdot 10^5$	$1.1 \cdot 10^6$	$3.2 \cdot 10^6$	$1.4 \cdot 10^6$	$1.2 \cdot 10^8$
SSKaon	$4.7 \cdot 10^5$	$4.1 \cdot 10^7$	$4.6 \cdot 10^5$	$3.1 \cdot 10^6$	$2.2 \cdot 10^5$	$1.5 \cdot 10^6$	$5.5 \cdot 10^4$	$3.7 \cdot 10^5$	$1.8 \cdot 10^5$	$1.2 \cdot 10^6$	$1.9 \cdot 10^5$	$1.6 \cdot 10^7$
SSPion	$4.1 \cdot 10^5$	$3.6 \cdot 10^7$	$4.1 \cdot 10^5$	$7.7 \cdot 10^6$	$2.0 \cdot 10^5$	$3.7 \cdot 10^6$	$4.9 \cdot 10^4$	$9.1 \cdot 10^5$	$1.6 \cdot 10^5$	$3.1 \cdot 10^6$	$1.6 \cdot 10^5$	$1.4 \cdot 10^7$
SSProton	$4.1 \cdot 10^5$	$3.6 \cdot 10^7$	$3.0 \cdot 10^5$	$6.9 \cdot 10^5$	$1.4 \cdot 10^5$	$3.2 \cdot 10^5$	$3.6 \cdot 10^4$	$8.1 \cdot 10^4$	$1.2 \cdot 10^5$	$2.8 \cdot 10^5$	$1.6 \cdot 10^5$	$1.4 \cdot 10^7$

These plots provide a detailed visual overview of the taggers' performance and training behavior.

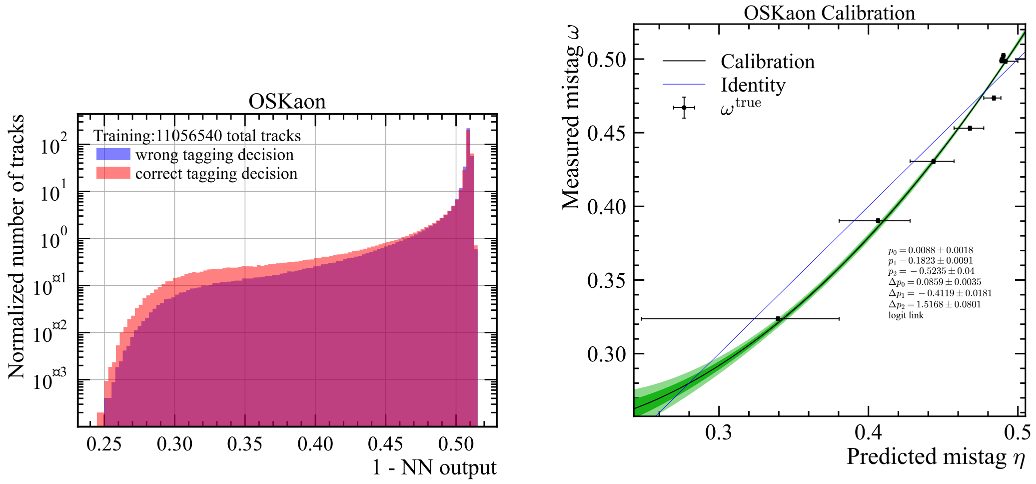
Table A.5: Pre- and post-calibration tagging performance for all Run 3 taggers. Tagging efficiency ϵ_{tag}, mistag ω_{eff} and tagging power ϵ_{eff} are given in percent. Where two uncertainties are quoted, the first uncertainty is statistical and the second one is the propagated uncertainty from the calibration parameters.

Tagger	ϵ_{tag} (%)	$\epsilon_{\text{tag}}^{\text{cal}}$ (%)	ω_{eff} (%)	$\omega_{\text{eff}}^{\text{cal}}$ (%)	ϵ_{eff} (%)	$\epsilon_{\text{eff}}^{\text{cal}}$ (%)
Opposite-side taggers						
OSKaon	96.35 ± 0.02	95.19 ± 0.02	43.63 ± 0.01	42.77 ± 0.01 (stat) ± 0.04 (cal)	1.563 ± 0.003	1.990 ± 0.003 (stat) ± 0.023 (cal)
OSMuon	81.76 ± 0.03	71.23 ± 0.04	44.65 ± 0.01	43.94 ± 0.01 (stat) ± 0.05 (cal)	0.935 ± 0.003	1.047 ± 0.003 (stat) ± 0.016 (cal)
OSElectron	92.71 ± 0.02	79.83 ± 0.03	46.70 ± 0.01	46.14 ± 0.01 (stat) ± 0.04 (cal)	0.404 ± 0.002	0.477 ± 0.002 (stat) ± 0.011 (cal)
Same-side taggers						
SSPion	99.90 ± 0.01	99.90 ± 0.01	45.43 ± 0.01	44.53 ± 0.02 (stat) ± 0.12 (cal)	0.836 ± 0.005	1.194 ± 0.007 (stat) ± 0.052 (cal)
SSProton	72.43 ± 0.11	31.93 ± 0.11	48.78 ± 0.00	47.66 ± 0.01 (stat) ± 0.002 (cal)	0.043 ± 0.000	0.070 ± 0.001 (stat) ± 0.013 (cal)
SSKaon	97.13 ± 0.04	95.27 ± 0.05	42.90 ± 0.02	41.17 ± 0.02 (stat) ± 0.10 (cal)	1.957 ± 0.011	2.973 ± 0.017 (stat) ± 0.070 (cal)



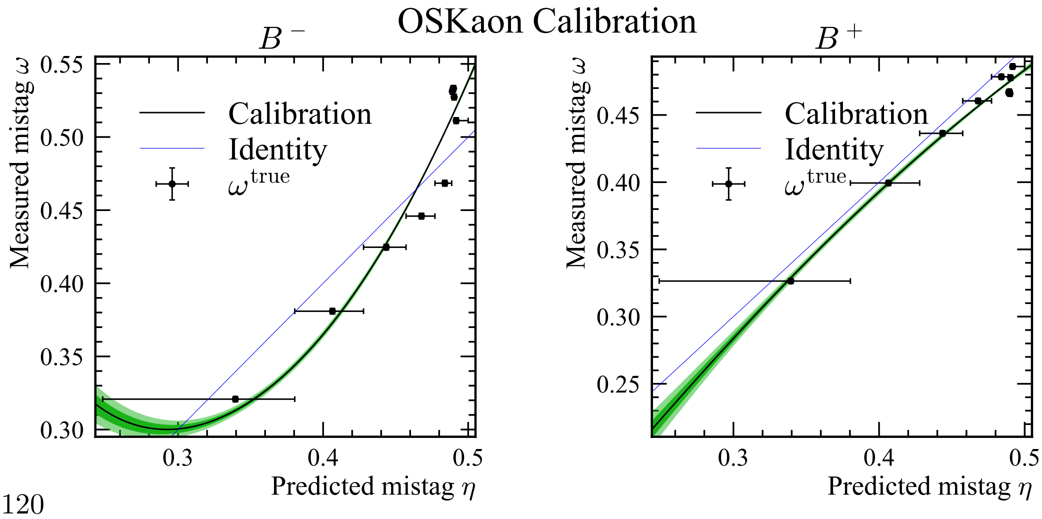
(a) Loss versus epoch for the OSKaon tagger.

(b) ROC curve for the OSKaon tagger, showing the Area Under the Curve (AUC).



(c) Normalized distribution of the tagger output, separated by tagging decision.

(d) Calibration plot for the OSKaon tagger on a hold-out test sample.



(e) Calibration plot for the OSKaon tagger on a hold-out test sample distinguishing by B species.

Figure A.5: Training and validation plots for the OSKaon tagger.

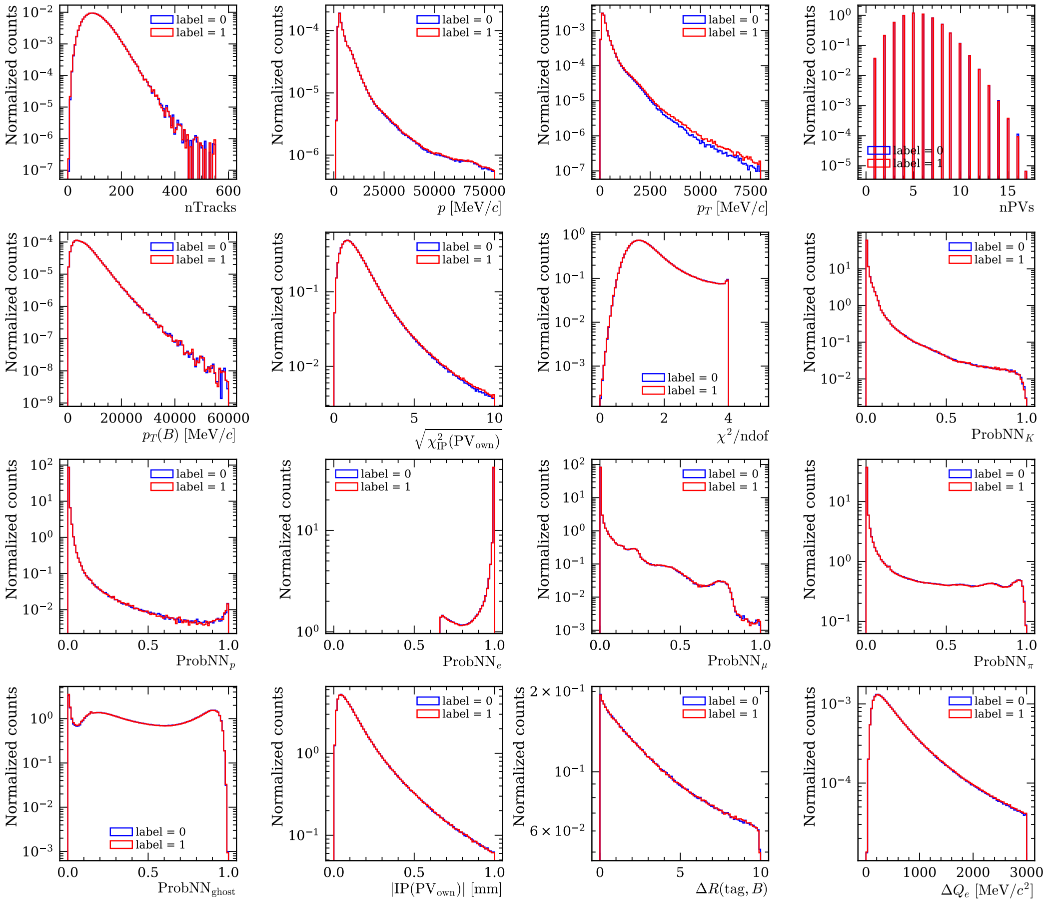
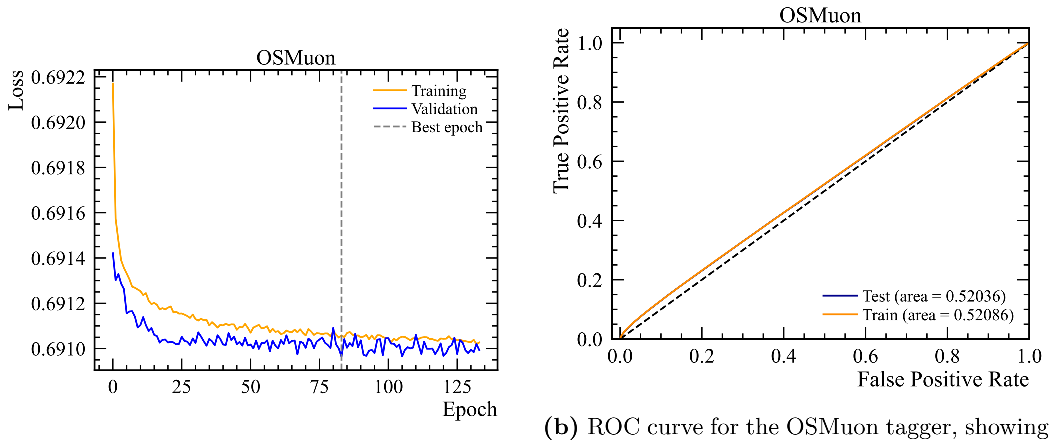
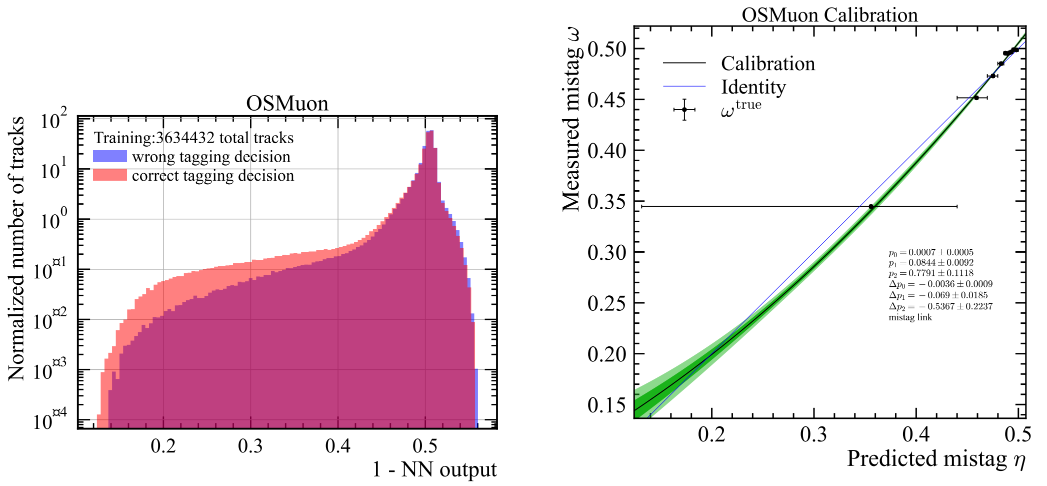


Figure A.6: Distributions of the input features used for the OSKaon training.



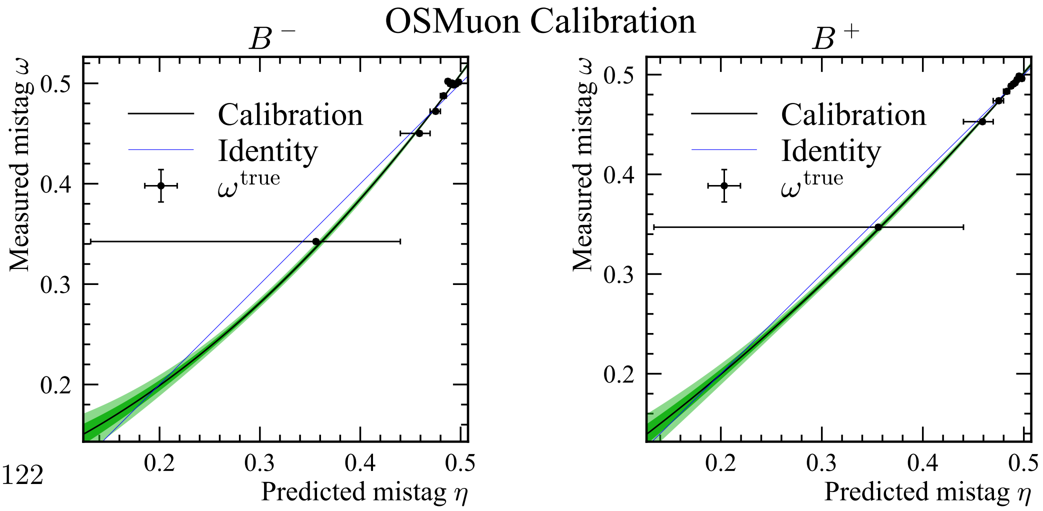
(a) Loss versus epoch for the OSMuon tagger.

(b) ROC curve for the OSMuon tagger, showing the Area Under the Curve (AUC).



(c) Normalized distribution of the tagger output, separated by tagging decision.

(d) Calibration plot for the OSMuon tagger on a hold-out test sample.



(e) Calibration plot for the OSMuon tagger on a hold-out test sample distinguishing by B species.

Figure A.7: Training and validation plots for the OSMuon tagger.

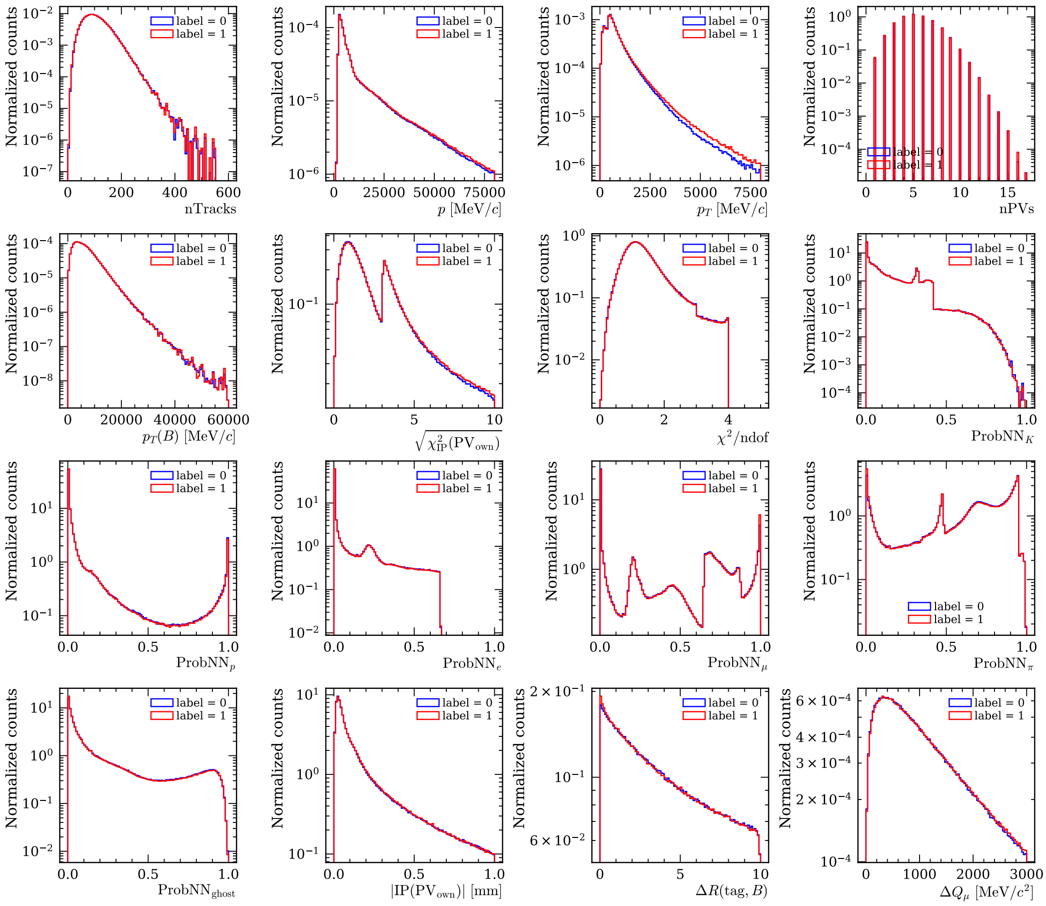
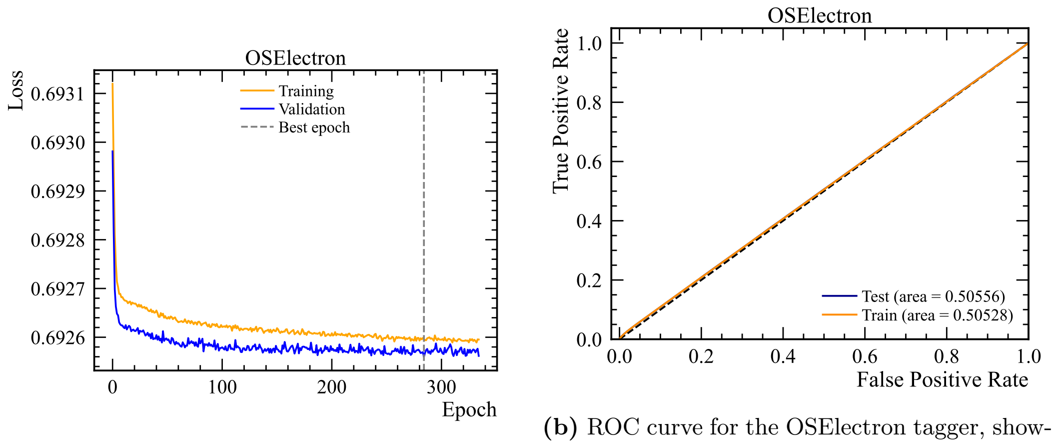
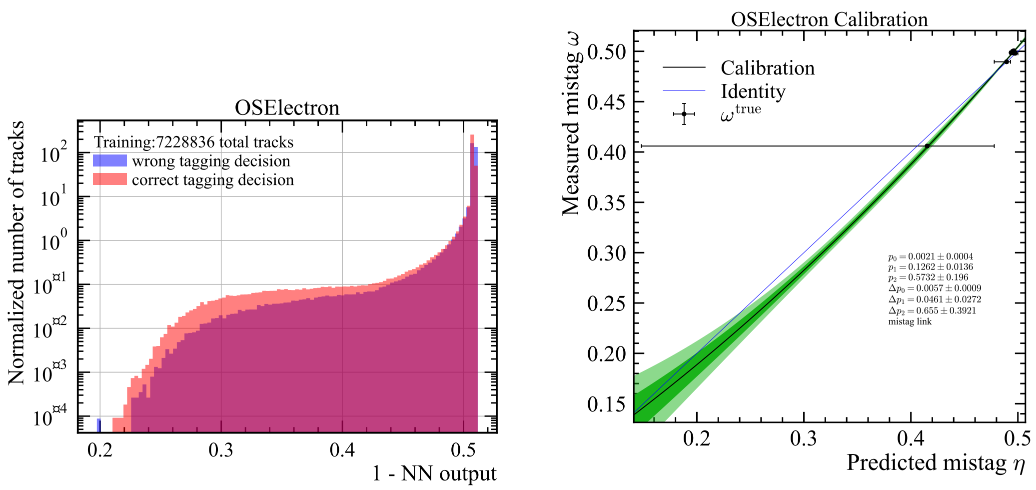


Figure A.8: Distributions of the input features used for the OSMuon training.



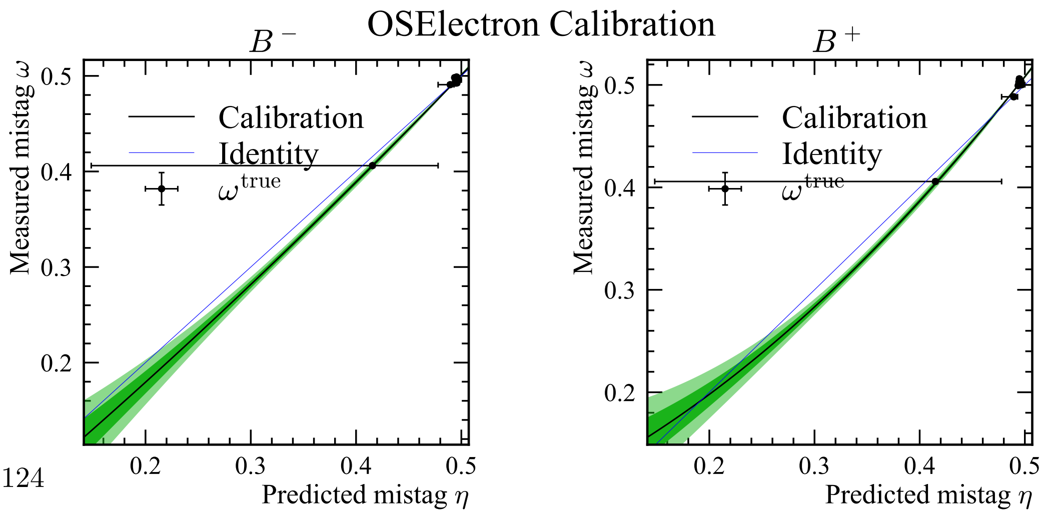
(a) Loss versus epoch for the OSElectron tagger.

(b) ROC curve for the OSElectron tagger, showing the Area Under the Curve (AUC).



(c) Normalized distribution of the tagger output, separated by tagging decision.

(d) Calibration plot for the OSElectron tagger on a hold-out test sample.



(e) Calibration plot for the OSElectron tagger on a hold-out test sample distinguishing by B species.

Figure A.9: Training and validation plots for the OSElectron tagger.

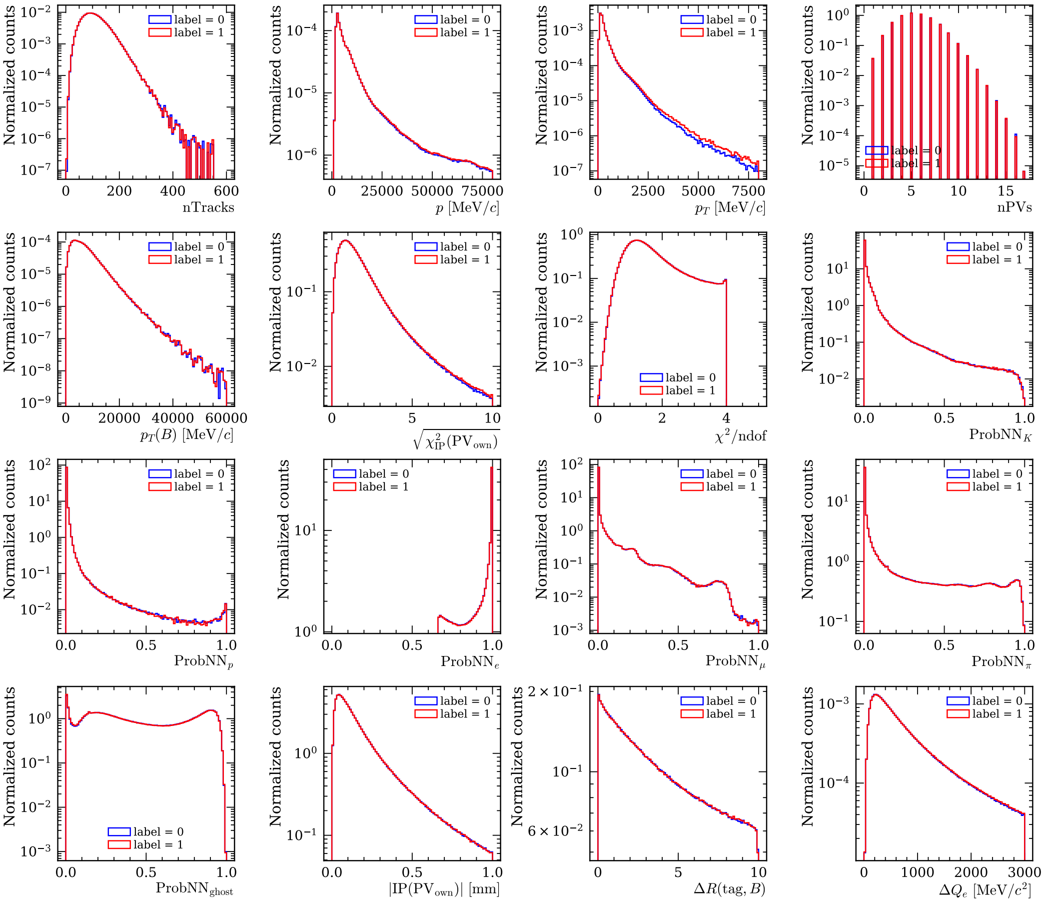
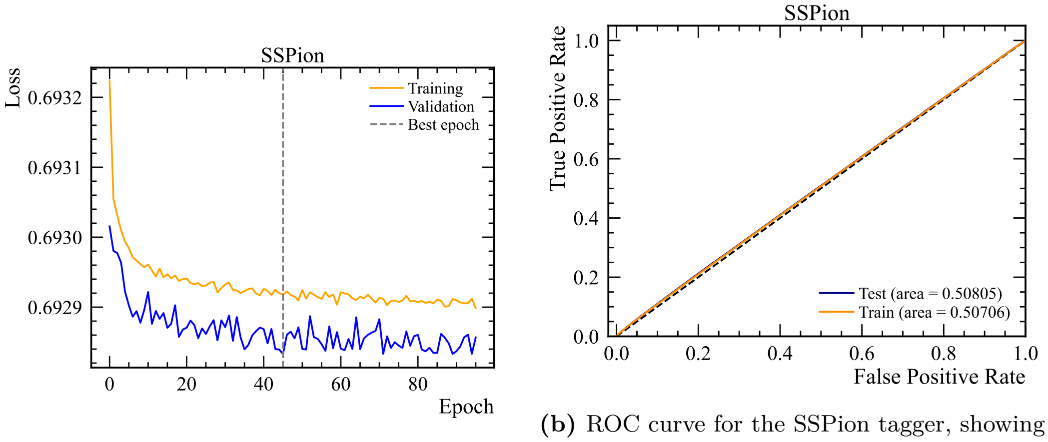
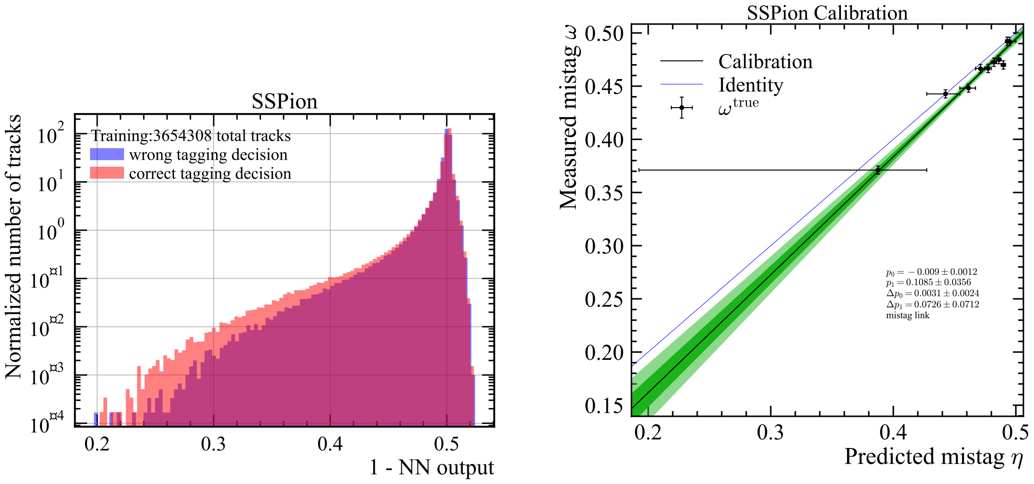


Figure A.10: Distributions of the input features used for the OSElectron training.



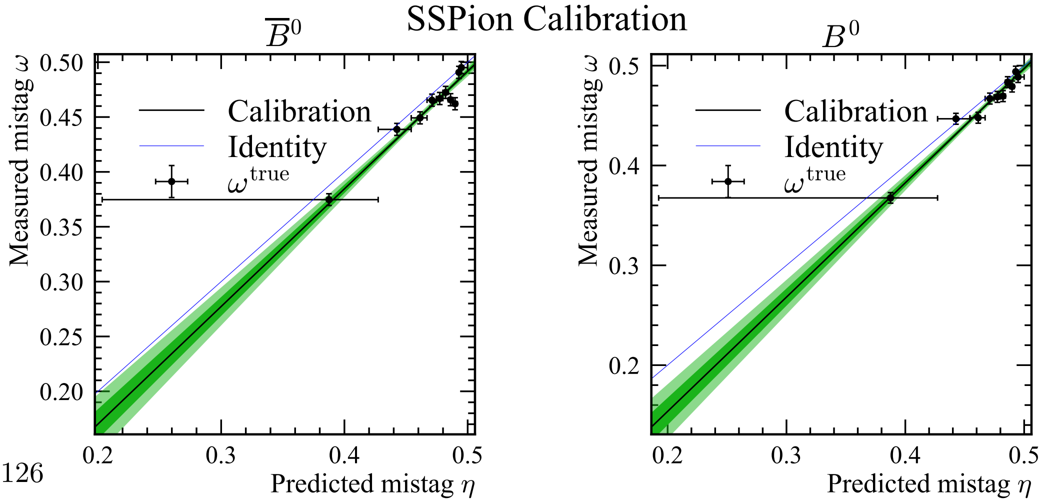
(a) Loss versus epoch for the SSPion tagger.

(b) ROC curve for the SSPion tagger, showing the Area Under the Curve (AUC).



(c) Normalized distribution of the tagger output, separated by tagging decision.

(d) Calibration plot for the SSPion tagger on a hold-out test sample.



(e) Calibration plot for the SSPion tagger on a hold-out test sample distinguishing by B species.

Figure A.11: Training and validation plots for the SSPion tagger.

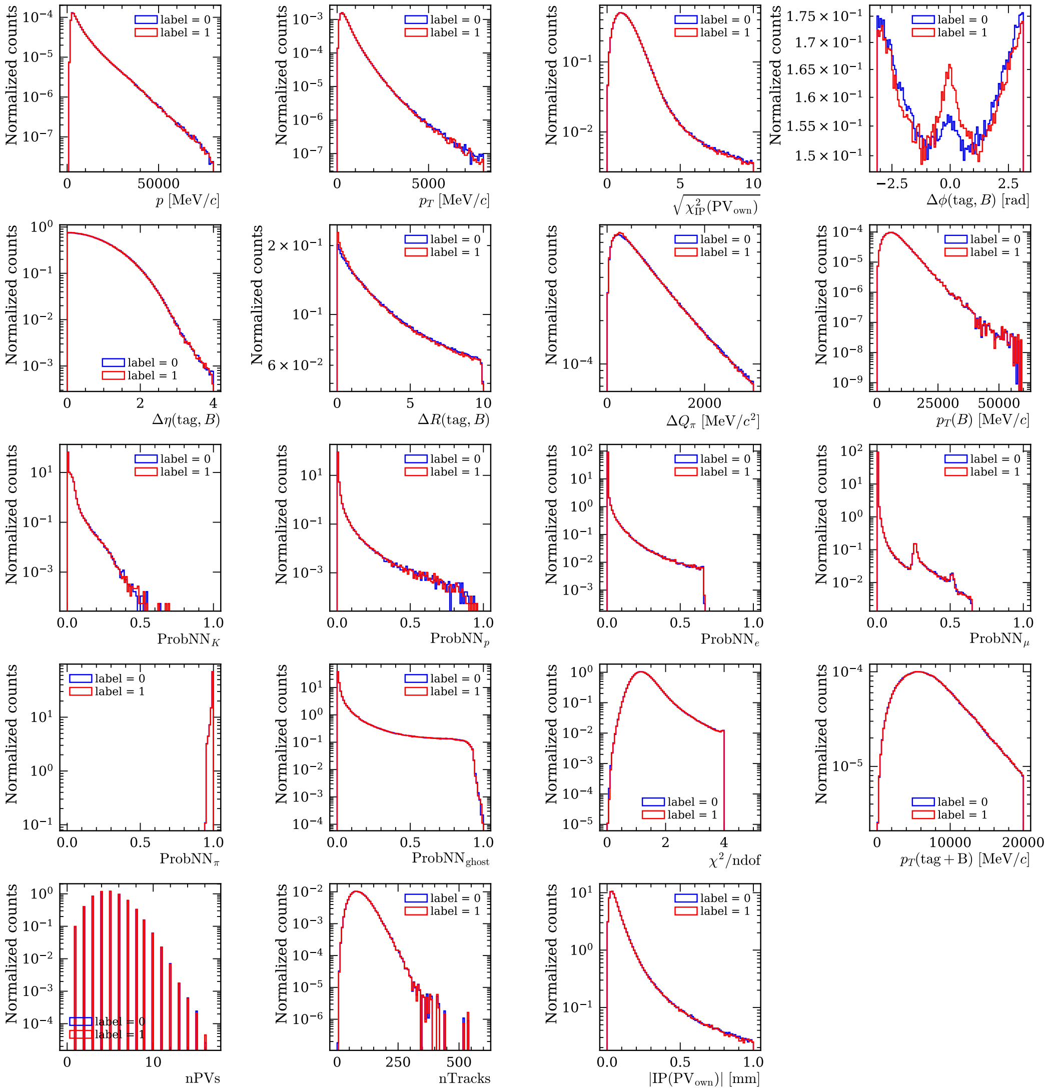
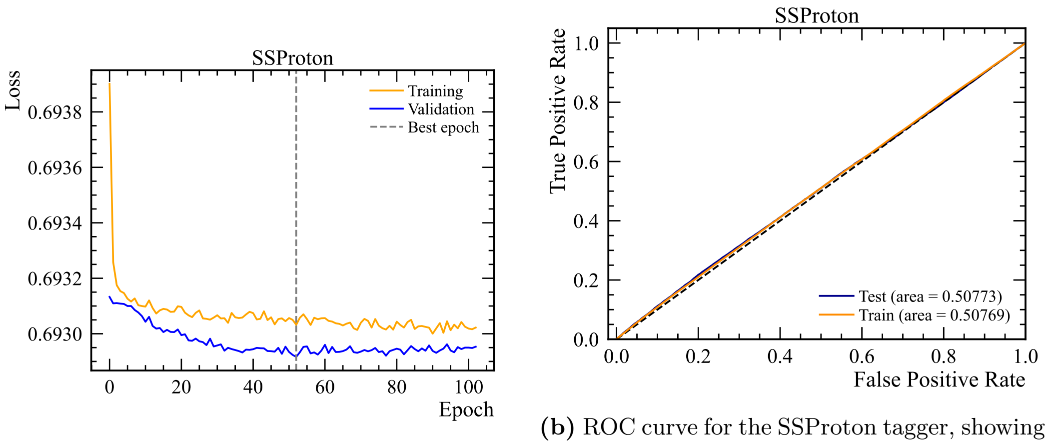
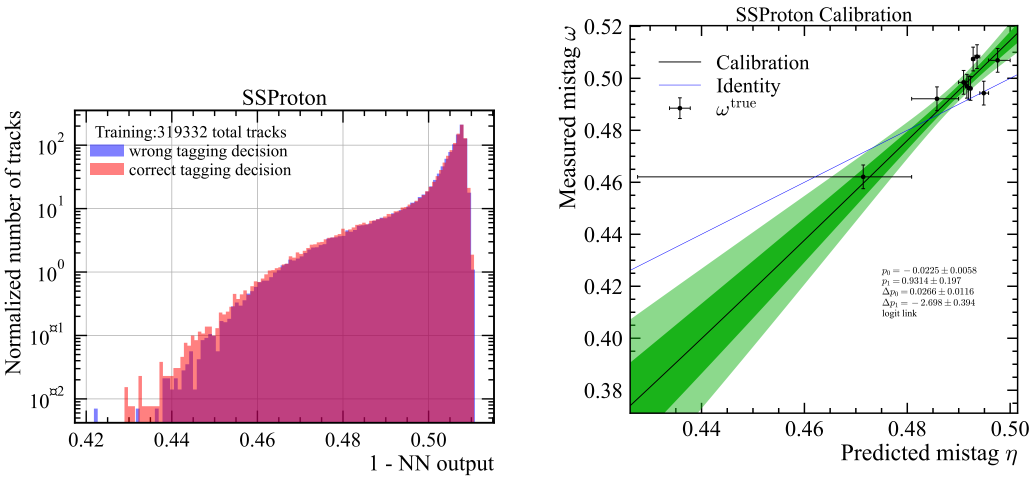


Figure A.12: Distributions of the input features used for the SSPion training.



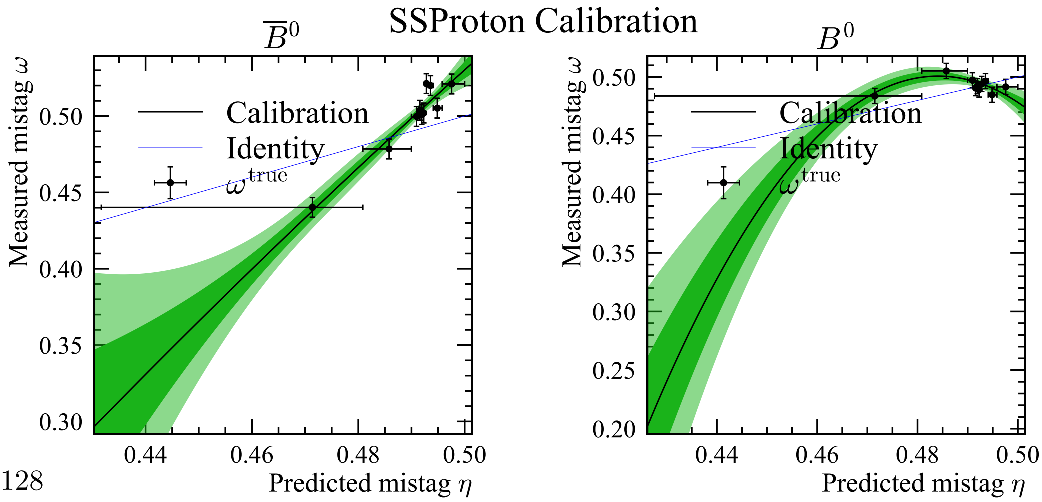
(a) Loss versus epoch for the SSProton tagger.

(b) ROC curve for the SSProton tagger, showing the Area Under the Curve (AUC).



(c) Normalized distribution of the tagger output, separated by tagging decision.

(d) Calibration plot for the SSProton tagger on a hold-out test sample.



(e) Calibration plot for the SSProton tagger on a hold-out test sample distinguishing by B species.

Figure A.13: Training and validation plots for the SSProton tagger.

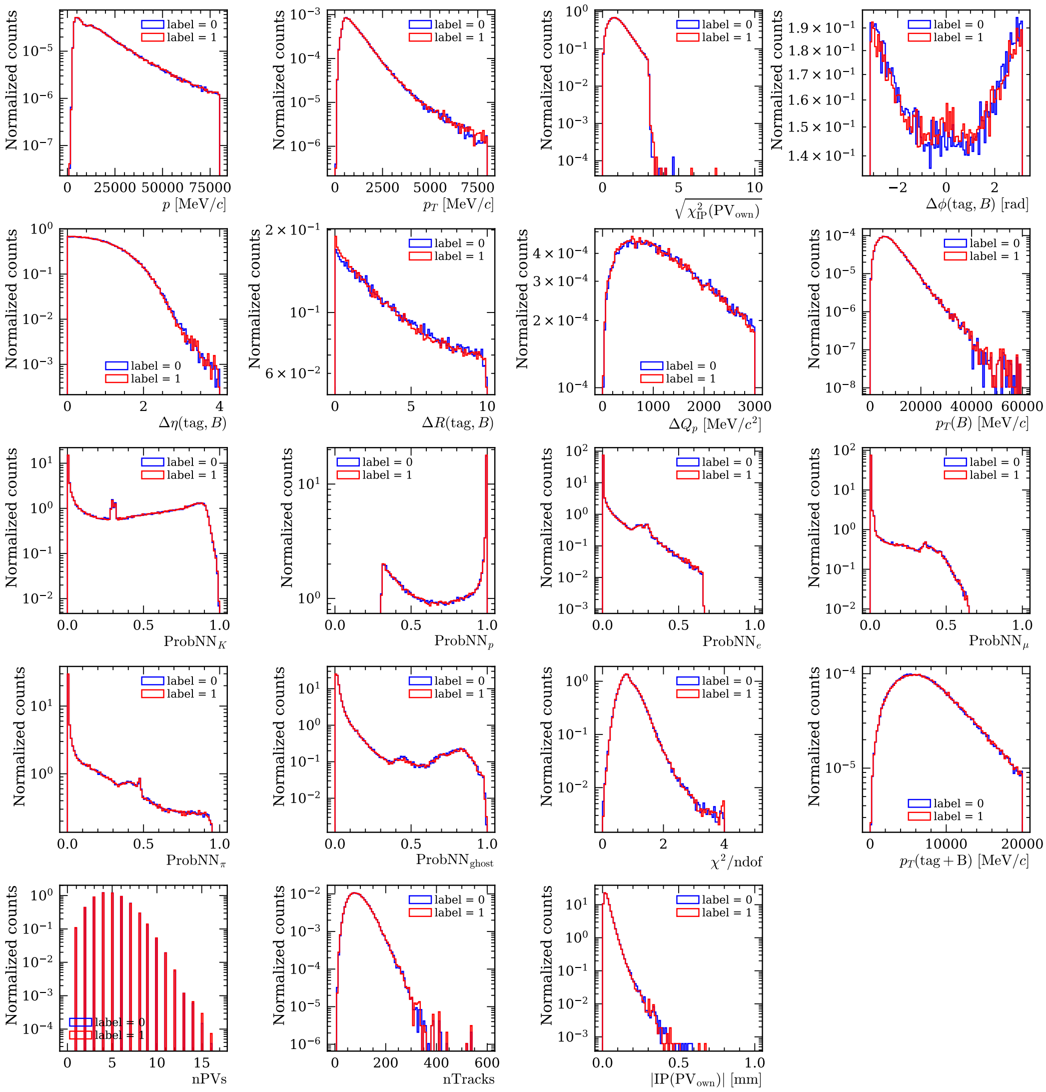


Figure A.14: Distributions of the input features used for the SSProton training.

B Supplementary Material for Chapter 6

B.1 Combined taggers on 2024 MC

Table B.1: Calibrated tagging power ϵ_{eff} (%) for individual Run 3/Benchmark taggers and their combinations. Calibration on Run 3 Monte Carlo samples of $B_s^0 \to D_s^- \pi^+$, $B^+ \to J/\psi K^+$, and $B^0 \to J/\psi K^{*0}$. The combined rows (OS, OS+SS) report the relative change $\Delta_{\%}$ between the Run 3 and Benchmark, together with its uncertainty evaluated in the limiting cases of no correlation ($\rho = 0$, upper value) and full correlation ($\rho = 1$, lower value). These two limits are quoted using a compact asymmetric notation, with the upper and lower uncertainties indicated by the superscripts + and −, respectively.

Decay / Tagger	**Benchmark ϵ_{eff} [%]**	**Run 3 ϵ_{eff} [%]**	$\Delta_{\%}$
$B_s^0 \to D_s^- \pi^+$			
OSElectron	0.28 ± 0.03	0.47 ± 0.04	
OSMuon	0.86 ± 0.06	1.05 ± 0.06	
OSKaon	1.79 ± 0.08	1.78 ± 0.08	
SSKaon	2.97 ± 0.10	2.81 ± 0.10	
OS	$\mathbf{2.89 \pm 0.10}$	$\mathbf{3.22 \pm 0.10}$	$\mathbf{+11.4^{+5.2}_{-0.4}}$
OS+SS	$\mathbf{5.69 \pm 0.13}$	$\mathbf{5.85 \pm 0.13}$	$\mathbf{+2.8^{+3.28}_{-0.06}}$
$B^+ \to J/\psi K^+$			
OSElectron	0.24 ± 0.02	0.47 ± 0.03	
OSMuon	0.79 ± 0.03	1.08 ± 0.04	
OSKaon	1.88 ± 0.05	2.02 ± 0.06	
OS	$\mathbf{2.86 \pm 0.06}$	$\mathbf{3.48 \pm 0.07}$	$\mathbf{+21.7^{+3.54}_{-0.11}}$
$B^0 \to J/\psi K^{*0}$			
OSElectron	0.34 ± 0.04	0.52 ± 0.05	
OSMuon	0.78 ± 0.06	1.01 ± 0.07	
OSKaon	1.84 ± 0.10	1.83 ± 0.10	
SSPion	1.20 ± 0.08	1.11 ± 0.08	
SSProton	0.12 ± 0.02	0.07 ± 0.02	
OS	$\mathbf{2.90 \pm 0.12}$	$\mathbf{3.32 \pm 0.12}$	$\mathbf{+14.5^{+5.8}_{-0.6}}$
OS+SS	$\mathbf{4.14 \pm 0.14}$	$\mathbf{4.46 \pm 0.15}$	$\mathbf{+7.7^{+5.14}_{-0.02}}$

Table B.1 reports the calibrated tagging power for the single taggers and their combination. The last column quantifies the performance gain $\mathit{\Delta}_{\%}$ of the Run 3 combination with respect to their Benchmark counterpart as:

$$\mathit{\Delta}_{\%} = \frac{\varepsilon_{\text{eff}}^{\text{Run 3}} - \varepsilon_{\text{eff}}^{\text{Benchmark}}}{\varepsilon_{\text{eff}}^{\text{Benchmark}}} \times 100 \tag{B.1}$$

where $\varepsilon_{\text{eff}}^{\text{Benchmark}}$ and $\varepsilon_{\text{eff}}^{\text{Run 3}}$ are the calibrated tagging power of the Benchmark and Run 3 tagger combination. Introducing ρ as the correlation coefficient, the propagated uncertainty $\sigma_{\mathit{\Delta}_{\%}}$ is:

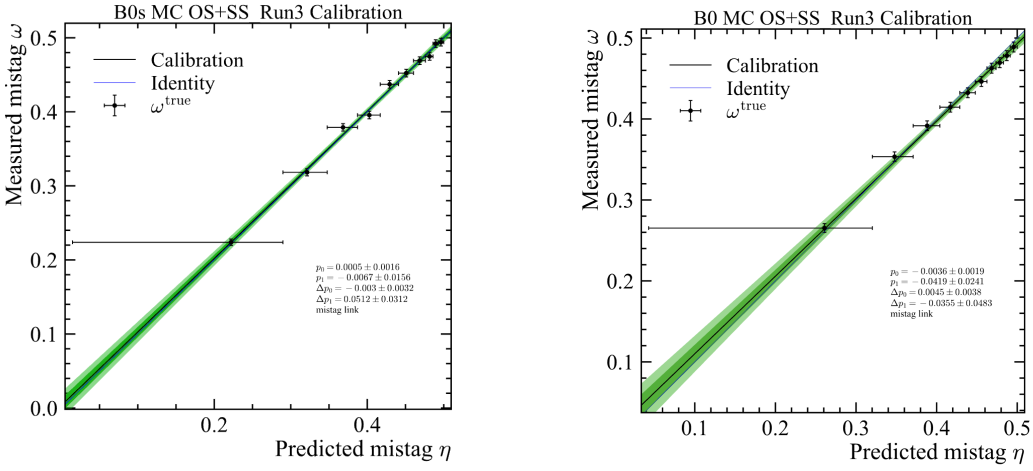
$$\sigma_{\mathit{\Delta}_{\%}} = 100 \cdot \frac{\varepsilon_{\text{eff}}^{\text{Run 3}}}{\varepsilon_{\text{eff}}^{\text{Bench.}}} \sqrt{\left[\frac{\sigma\left(\varepsilon_{\text{eff}}^{\text{Run 3}}\right)}{\varepsilon_{\text{eff}}^{\text{Run 3}}}\right]^2 + \left[\frac{\sigma\left(\varepsilon_{\text{eff}}^{\text{Bench.}}\right)}{\varepsilon_{\text{eff}}^{\text{Bench.}}}\right]^2 - 2\rho \left[\frac{\sigma\left(\varepsilon_{\text{eff}}^{\text{Run 3}}\right)}{\varepsilon_{\text{eff}}^{\text{Run 3}}}\right]\left[\frac{\sigma\left(\varepsilon_{\text{eff}}^{\text{Bench.}}\right)}{\varepsilon_{\text{eff}}^{\text{Bench.}}}\right]} \tag{B.2}$$

Given that the taggers are calibrated on the same samples, the tagging powers are expected to be correlated. The uncertainty on $\mathit{\Delta}_{\%}$ is therefore reported as a range: no correlation $\rho = 0$, full correlation $\rho = 1$. The $\sigma_{\mathit{\Delta}_{\%}}$ is expected to lie between the two limiting scenarios.

The nominal relative improvement is +2.8% for the OS+SS combination in $B_s^0 \to D_s^- \pi^+$, +7.7% for the OS+SS combination in $B^0 \to J/\psi K^{*0}$, and +21.7% for the OS combination in $B^+ \to J/\psi K^+$. It is nonetheless important to note that the SS Run 3 taggers tend to underperform relative to their benchmark counterparts. This reduction affects the relative gain $\mathit{\Delta}_{\%}$ when adding SS taggers to the OS combination. For example, in $B_s^0 \to D_s^- \pi^+$, the improvement decreases from approximately +11% (OS only) to about +3% once the SSKaon tagger is included. A plausible explanation is the higher-track-multiplicity environment in Run 3, which may particularly challenge SS taggers that rely on identifying nearby fragmentation particles.

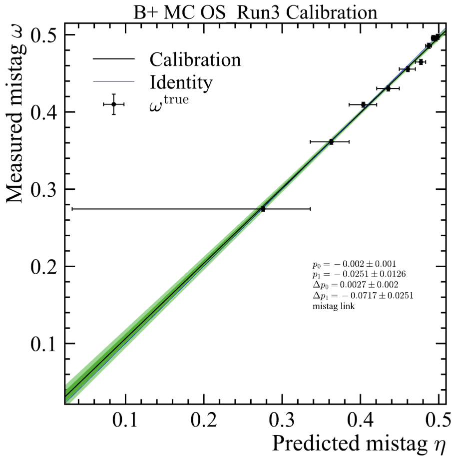
B.2 Calibration Plots: Run 3 combination

This section collects the calibration plots for the tagger combinations evaluated on simulated samples (exposed in Section 6.2) and on 2024 data (exposed in Section 6.3). The corresponding calibrations plots for the combination of the Benchmark taggers are documented in Appendix B (Figure B.3 for simulated samples, Figure B.4 for block 1 data). In all cases, a linear fit is sufficient, and the $\mathit{\Delta}p$ parameters do not report pathological values.



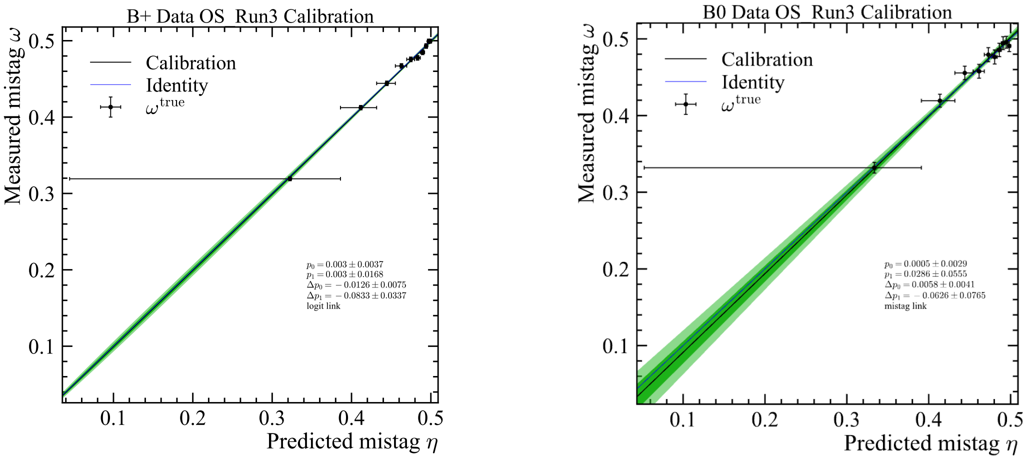
(a) $B_s^0 \to D_s^- \pi^+$ OS+SS combination

(b) $B^0 \to J/\psi K^{*0}$ OS+SS combination

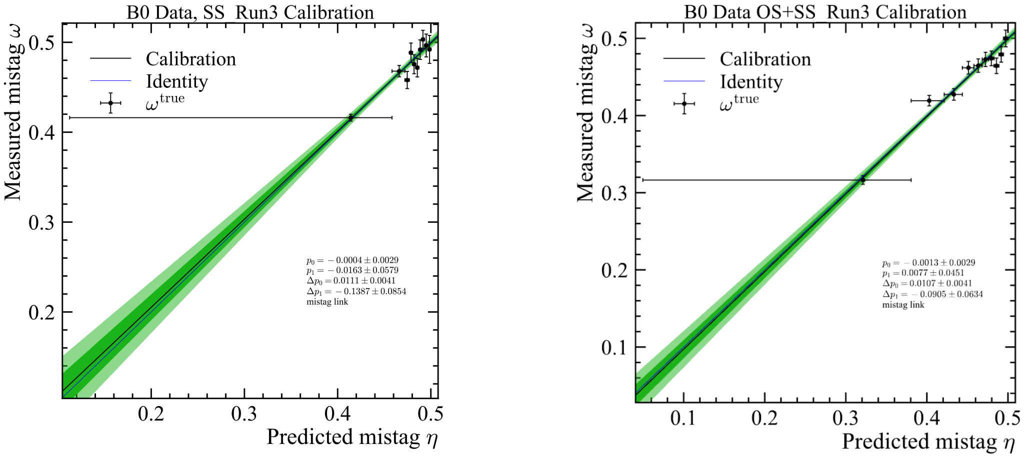


(c) $B^+ \to J/\psi K^+$ OS combination

Figure B.1: Calibration plots for the Run 3 tagger combinations on Monte Carlo simulations of $B_s^0 \to D_s^- \pi^+$, $B^0 \to J/\psi K^{*0}$, and $B^+ \to J/\psi K^+$ decays. The black line shows the first–order polynomial calibration in the link space, with parameters p_0 and p_1 controlling the average offset and slope, and Δp_0 and Δp_1 accounting for flavour asymmetries between B and $\overline{B}$ mesons. The shaded band reflects the fit uncertainty, while the blue line shows the identity $\omega(\eta) = \eta$.



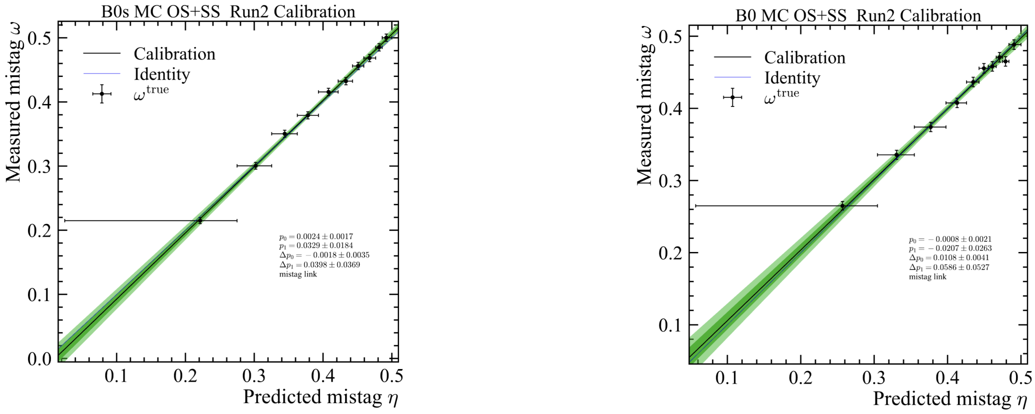
(a) OS combination in $B^+ \to J/\psi K^+$ **(b)** OS combination in $B^0 \to J/\psi K^{*0}$



(c) SS combination in $B^0 \to J/\psi K^{*0}$ **(d)** OS+SS combination in $B^0 \to J/\psi K^{*0}$

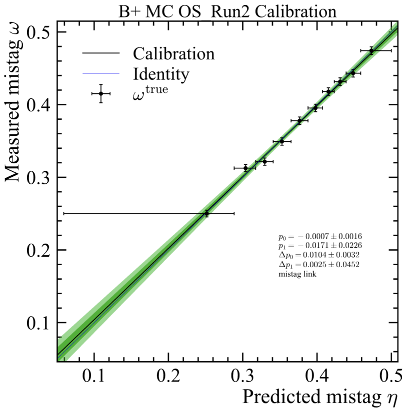
Figure B.2: Calibration plots for Run 3 tagger combinations on 2024 data. Top: OS (OSKaon, OSMuon, OSElectron) in $B^+ \to J/\psi K^+$ and $B^0 \to J/\psi K^{*0}$ decays. Bottom: SS and SS+OS (SSPion, SSProton, OSKaon, OSMuon, OSElectron) in $B^0 \to J/\psi K^{*0}$. The black line shows the first–order polynomial calibration in the link space, with parameters p_0 and p_1 controlling the average offset and slope, and Δp_0 and Δp_1 accounting for flavour asymmetries between B and $\overline{B}$ mesons. The shaded band reflects the fit uncertainty, while the blue line shows the identity $\omega(\eta) = \eta$.

B.3 Calibration Plots: Benchmark taggers



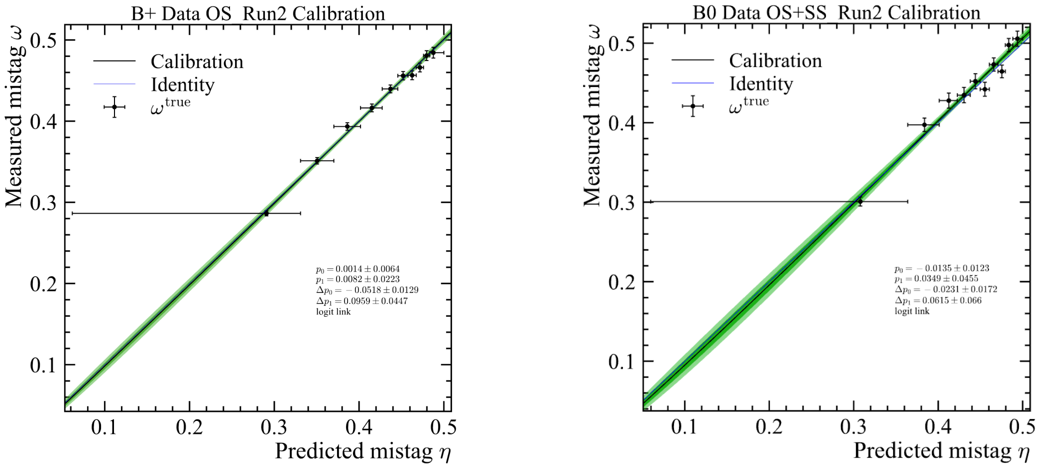
(a) $B_s^0 \to D_s^- \pi^+$, Benchmark tagger combination: OSKaon, OSElectron, OSMuon, SSKaon.

(b) $B^0 \to J/\psi K^{*0}$, Benchmark tagger combination: OSKaon, OSElectron, OSMuon, SSPion, SSProton.



(c) $B^+ \to J/\psi K^+$, Benchmark tagger combination: OSKaon, OSElectron, OSMuon.

Figure B.3: Calibration plots for the Benchmark tagger combinations on Run 3 Monte Carlo simulations of $B_s^0 \to D_s^- \pi^+$, $B^0 \to J/\psi K^{*0}$, $B^+ \to J/\psi K^+$ decays. The black line shows the calibration fit, the blue dashed line indicates perfect calibration ($\omega_{\mathrm{true}} = \eta$), and the green shaded area represents the fit uncertainty.



(a) OS combination in $B^+ \to J/\psi K^+$

(b) OS+SS combination in $B^0 \to J/\psi K^{*0}$

Figure B.4: Calibration plots for Benchmark tagger combinations on 2024 data. Left: OS (OSKaon, OSMuon, OSElectron) in $B^+ \to J/\psi K^+$. Right: SS+OS (SSPion, SSProton, OSKaon, OSMuon, OSElectron) in $B^0 \to J/\psi K^{*0}$. The black line shows the first–order polynomial calibration in the link space, with parameters p_0 and p_1 controlling the average offset and slope, and Δp_0 and Δp_1 accounting for flavour asymmetries between B and $\bar{B}$ mesons. The shaded band reflects the fit uncertainty, while the blue line shows the identity $\omega(\eta) = \eta$.

List of Figures

List of Tables

Bibliography

[1] J. A. Gooding et al. “The SMARTHEP European Training Network.” In: *EPJ Web of Conferences* 295 (2024). Ed. by R. De Vita et al., p. 08022. ISSN: 2100-014X. DOI: `10.1051/epjconf/202429508022`. URL: `http://dx.doi.org/10.1051/epjconf/202429508022`.

[2] *SMARTHEP ETN Project Website*. Accessed: 2025-08-16.

[3] J. Albrecht et al. *SMARTHEP: training PhD students in real-time analysis at the LHC and in industry*. 2026. arXiv: `2601.07089 [physics.ed-ph]`. URL: `https://arxiv.org/abs/2601.07089`.

[4] The Nobel Prize. *The Nobel Prize in Physics 2024*. `https://www.nobelprize.org/prizes/physics/2024/press-release/`. Awarded to John J. Hopfield and Geoffrey Hinton for pioneering work on artificial neural networks. 2024.

[5] Y. Lecun et al. “Gradient-based learning applied to document recognition.” In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324. DOI: `10.1109/5.726791`.

[6] S. Hochreiter and J. Schmidhuber. *Long Short-Term Memory*. 1997. DOI: `10.1162/neco.1997.9.8.1735`.

[7] A. Vaswani et al. *Attention Is All You Need*. 2023. arXiv: `1706.03762 [cs.CL]`. URL: `https://arxiv.org/abs/1706.03762`.

[8] F. Scarselli et al. “The Graph Neural Network Model.” In: *IEEE Transactions on Neural Networks* 20.1 (2009), pp. 61–80. DOI: `10.1109/TNN.2008.2005605`.

[9] T. Berners-Lee et al. *WorldWideWeb: Proposal for a HyperText Project*. CERN Internal Document. Accessed: 2025-12-04. 1990. URL: `https://www.w3.org/History/19921103-hypertext/hypertext/WWW/Proposal.html`.

[10] L. Clissa, M. Lassnig, and L. Rinaldi. “How big is Big Data? A comprehensive survey of data production, storage, and streaming in science and industry.” In: *Frontiers in Big Data* 6 (Oct. 2023). ISSN: 2624-909X. DOI: `10.3389/fdata.2023.1271639`. URL: `http://dx.doi.org/10.3389/fdata.2023.1271639`.

[11] A. G. et al. *High-Luminosity Large Hadron Collider (HL-LHC)*. CERN Yellow Reports: Monographs. Geneva: CERN, 2017. DOI: `10.23731/CYRM-2017-004`. URL: `https://cds.cern.ch/record/2284929`.

[12] S. Astrand et al. “Perspective on machine learning for real-time analysis at the Large Hadron Collider experiments ALICE, ATLAS, CMS and LHCb.” In: *Machine Learning: Science and Technology* 7.1 (2026), p. 013001. DOI: `10.1088/2632-2153/ae35cc`. URL: `https://doi.org/10.1088/2632-2153/ae35cc`.

[13] A. Bifet et al. *Machine Learning for Data Streams: with Practical Examples in MOA*. MIT Press, 2015.

[14] A. Di Girolamo et al. "Preparing Distributed Computing Operations for the HL-LHC Era With Operational Intelligence." In: *Front.Big Data* 4 (2022), p. 753409. DOI: `10.3389/fdata.2021.753409`. URL: `https://hal.science/hal-03554926`.

[15] S. L. Glashow. "Partial-symmetries of weak interactions." In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: `https://doi.org/10.1016/0029-5582(61)90469-2`. URL: `https://www.sciencedirect.com/science/article/pii/0029558261904692`.

[16] S. Weinberg. "A Model of Leptons." In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: `10.1103/PhysRevLett.19.1264`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.19.1264`.

[17] A. Salam and J. C. Ward. "Weak and electromagnetic interactions." In: *Il Nuovo Cimento* 11 (1959), pp. 568–577. DOI: `10.1007/BF02726525`.

[18] P. W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons." In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: `10.1103/PhysRevLett.13.508`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.13.508`.

[19] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons." In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: `10.1103/PhysRevLett.13.321`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.13.321`.

[20] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles." In: *Phys. Rev. Lett.* 13 (20 Nov. 1964), pp. 585–587. DOI: `10.1103/PhysRevLett.13.585`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.13.585`.

[21] A. Scarabotto. *Search for rare four-body charm decays with electrons in the final state and long track reconstruction for the LHCb trigger*. Nov. 2023.

[22] N. Cabibbo. "Unitary Symmetry and Leptonic Decays." In: *Phys. Rev. Lett.* 10 (12 June 1963), pp. 531–533. DOI: `10.1103/PhysRevLett.10.531`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.10.531`.

[23] M. Kobayashi and T. Maskawa. "CP-Violation in the Renormalizable Theory of Weak Interaction." In: *Progress of Theoretical Physics* 49.2 (Feb. 1973), pp. 652–657. ISSN: 0033-068X. DOI: `10.1143/PTP.49.652`. eprint: `https://academic.oup.com/ptp/article-pdf/49/2/652/5257692/49-2-652.pdf`. URL: `https://doi.org/10.1143/PTP.49.652`.

[24] L.-L. Chau and W.-Y. Keung. "Comments on the Parametrization of the Kobayashi-Maskawa Matrix." In: *Phys. Rev. Lett.* 53 (19 Nov. 1984), pp. 1802–1805. DOI: `10.1103/PhysRevLett.53.1802`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.53.1802`.

[25] S. Navas et al. "Review of particle physics." In: *Phys. Rev. D* 110.3 (2024), p. 030001. DOI: `10.1103/PhysRevD.110.030001`.

[26] Y. Sofue and V. Rubin. “Rotation Curves of Spiral Galaxies.” In: *Annual Review of Astronomy and Astrophysics* 39.1 (Sept. 2001), pp. 137–174. ISSN: 1545-4282. DOI: `10.1146/annurev.astro.39.1.137`. URL: `http://dx.doi.org/10.1146/annurev.astro.39.1.137`.

[27] R. Massey, T. Kitching, and J. Richard. “The dark matter of gravitational lensing.” In: *Reports on Progress in Physics* 73.8 (July 2010), p. 086901. DOI: `10.1088/0034-4885/73/8/086901`. URL: `https://doi.org/10.1088/0034-4885/73/8/086901`.

[28] Q. R. Ahmad et al. “Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory.” In: *Phys. Rev. Lett.* 89 (1 June 2002), p. 011301. DOI: `10.1103/PhysRevLett.89.011301`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.89.011301`.

[29] Z. Maki, M. Nakagawa, and S. Sakata. “Remarks on the Unified Model of Elementary Particles.” In: *Progress of Theoretical Physics* 28.5 (Nov. 1962), pp. 870–880. ISSN: 0033-068X. DOI: `10.1143/PTP.28.870`. eprint: `https://academic.oup.com/ptp/article-pdf/28/5/870/5258750/28-5-870.pdf`. URL: `https://doi.org/10.1143/PTP.28.870`.

[30] B. Pontecorvo. “Inverse Beta Processes and Nonconservation of Lepton Charge.” In: *Sov. Phys. JETP* 7 (1958), pp. 172–173.

[31] A. D. Sakharov. “Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe.” In: *Soviet Physics Uspekhi* 34.5 (May 1991), p. 392. DOI: `10.1070/PU1991v034n05ABEH002497`. URL: `https://doi.org/10.1070/PU1991v034n05ABEH002497`.

[32] G. R. Farrar and M. E. Shaposhnikov. “Baryon asymmetry of the Universe in the standard model.” In: *Physical Review D* 50.2 (July 1994), pp. 774–818. ISSN: 0556-2821. DOI: `10.1103/physrevd.50.774`. URL: `http://dx.doi.org/10.1103/PhysRevD.50.774`.

[33] P. Huet and E. Sather. “Electroweak baryogenesis and standard modelCPviolation.” In: *Physical Review D* 51.2 (Jan. 1995), pp. 379–394. ISSN: 0556-2821. DOI: `10.1103/physrevd.51.379`. URL: `http://dx.doi.org/10.1103/PhysRevD.51.379`.

[34] M. Gavela et al. “Standard model CP-violation and baryon asymmetry (I). Zero temperature.” In: *Nuclear Physics B* 430.2 (Nov. 1994), pp. 345–381. ISSN: 0550-3213. DOI: `10.1016/0550-3213(94)00409-9`. URL: `http://dx.doi.org/10.1016/0550-3213(94)00409-9`.

[35] S. P. MARTIN. “A SUPERSYMMETRY PRIMER.” In: *Perspectives on Supersymmetry*. WORLD SCIENTIFIC, July 1998, pp. 1–98. DOI: `10.1142/9789812839657_0001`. URL: `http://dx.doi.org/10.1142/9789812839657_0001`.

[36] M. Schmaltz and D. Tucker-Smith. “LITTLE HIGGS THEORIES.” In: *Annual Review of Nuclear and Particle Science* 55.1 (Dec. 2005), pp. 229–270. ISSN: 1545-4134. DOI: `10.1146/annurev.nucl.55.090704.151502`. URL: `http://dx.doi.org/10.1146/annurev.nucl.55.090704.151502`.

[37] L. Randall and R. Sundrum. “Large Mass Hierarchy from a Small Extra Dimension.” In: *Physical Review Letters* 83.17 (Oct. 1999), pp. 3370–3373. ISSN: 1079-7114. DOI: `10.1103/physrevlett.83.3370`. URL: `http://dx.doi.org/10.1103/PhysRevLett.83.3370`.

[38] L. Randall and R. Sundrum. "An Alternative to Compactification." In: *Physical Review Letters* 83.23 (Dec. 1999), pp. 4690–4693. ISSN: 1079-7114. DOI: `10.1103/physrevlett.83.4690`. URL: `http://dx.doi.org/10.1103/PhysRevLett.83.4690`.

[39] I. Doršner et al. "Physics of leptoquarks in precision experiments and at particle colliders." In: *Physics Reports* 641 (June 2016), pp. 1–68. ISSN: 0370-1573. DOI: `10.1016/j.physrep.2016.06.001`. URL: `http://dx.doi.org/10.1016/j.physrep.2016.06.001`.

[40] W. Buchmüller, R. Rückl, and D. Wyler. "Leptoquarks in lepton-quark collisions." In: *Physics Letters B* 191.4 (1987), pp. 442–448. ISSN: 0370-2693. DOI: `https://doi.org/10.1016/0370-2693(87)90637-X`. URL: `https://www.sciencedirect.com/science/article/pii/037026938790637X`.

[41] R. Aaij et al. "A precise measurement of the $B^0 - \bar{B}^0$ meson oscillation frequency." In: *The European Physical Journal C* 76.7 (July 2016). ISSN: 1434-6052. DOI: `10.1140/epjc/s10052-016-4250-2`. URL: `http://dx.doi.org/10.1140/epjc/s10052-016-4250-2`.

[42] R. Aaij et al. "Precise determination of the B^0_s–$\overline{B}^0_s$ oscillation frequency." In: *Nature Physics* 18.1 (Jan. 2022), pp. 1–5. ISSN: 1745-2481. DOI: `10.1038/s41567-021-01394-x`. URL: `http://dx.doi.org/10.1038/s41567-021-01394-x`.

[43] S. Vecchi. "Measurements of $\Delta m_{d,s}$ and $\Delta\Gamma_d$ at LHCb." In: *PoS* CKM2016 (2017), p. 087. DOI: `10.22323/1.291.0087`.

[44] L. Breiman et al. *Classification and Regression Trees.* Wadsworth International Group, 1984.

[45] S.-l. developers. *Understanding the decision tree structure.* Accessed: 2025-08-03. 2024. URL: `https://scikit-learn.org/stable/auto_examples/tree/plot_unveil_tree_structure.html`.

[46] C. M. Bishop. *Pattern Recognition and Machine Learning.* Springer, 2006.

[47] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning.* MIT Press, 2016. URL: `http://www.deeplearningbook.org/`.

[48] Ž. Ivezić et al. *Statistics, Data Mining and Machine Learning in Astronomy.* Princeton, NJ: Princeton University Press, 2014.

[49] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. "Learning representations by back-propagating errors." In: *Nature* 323.6088 (1986), pp. 533–536. DOI: `10.1038/323533a0`.

[50] T. Fawcett. "An introduction to ROC analysis." In: *Pattern Recognition Letters* 27.8 (2006). ROC Analysis in Pattern Recognition, pp. 861–874. ISSN: 0167-8655. DOI: `https://doi.org/10.1016/j.patrec.2005.10.010`. URL: `https://www.sciencedirect.com/science/article/pii/S016786550500303X`.

[51] L. Evans and P. Bryant. "LHC Machine." In: *JINST* 3 (2008), S08001. DOI: `10.1088/1748-0221/3/08/S08001`.

[52] W. Herr and B. Muratori. *Concept of luminosity.* 2006. DOI: `10.5170/CERN-2006-002.361`. URL: `https://cds.cern.ch/record/941318`.

[53] CERN. *High-Luminosity LHC (HL-LHC) Project.* 2025. URL: `https://hilumilhc.web.cern.ch/content/hl-lhc-project` (visited on 09/03/2025).

[54] A. Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC." In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: `10.1016/j.physletb.2012.08.020`.

[55] C. Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC." In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: `10.1016/j.physletb.2012.08.021`.

[56] A. Collaboration. "The ALICE experiment at the CERN LHC." In: *JINST* 3 (2008), S08002. DOI: `10.1088/1748-0221/3/08/S08002`.

[57] *The CERN accelerator complex.* Accessed: 2025-09-05. 2024. URL: `%5Curl%7Bhttps://home.cern/science/accelerators/accelerator-complex%7D`.

[58] LHCb Collaboration. *LHCb Collaboration Membership List.* 1789 members from 105 institutes in 25 countries. Reference date: 2025-09-03. 2025. URL: `https://lbfence.cern.ch/membership/report/members-list` (visited on 09/03/2025).

[59] L. CMS. "Observation of the rare Bs0 →µ+µ− decay from the combined analysis of CMS and LHCb data." In: *Nature* 522.7554 (May 2015), pp. 68–72. ISSN: 1476-4687. DOI: `10.1038/nature14474`. URL: `http://dx.doi.org/10.1038/nature14474`.

[60] R. Aaij et al. "A study of the Z production cross-section in pp collisions at $\sqrt{s} = 7\,\text{TeV}$ using tau final states." In: *Journal of High Energy Physics* 2013.1 (Jan. 2013). ISSN: 1029-8479. DOI: `10.1007/jhep01(2013)111`. URL: `http://dx.doi.org/10.1007/JHEP01(2013)111`.

[61] R. Aaij et al. *Coherent photoproduction of* ρ^0, ω *and excited vector mesons in ultraperipheral PbPb collisions.* Tech. rep. 22 pages, 7 figures. All figures and tables, along with any supplementary material and additional information, are available at https://cern.ch/lhcbproject/Publications/p/LHCb-PAPER-2024-042.html (LHCb public pages). Geneva: CERN, 2025. arXiv: `2506.06250`. URL: `https://cds.cern.ch/record/2935461`.

[62] R. Aaij et al. *Measurement of the* $\psi(2S)$ *to* J/ψ *cross-section ratio as a function of centrality in PbPb collisions at* $\sqrt{s_{NN}}$ *= 5.02 TeV.* All figures and tables, along with any supplementary material and additional information, are available at https://cern.ch/lhcbproject/Publications/p/LHCb-PAPER-2024-041.html (LHCb public pages). 2025. DOI: `10.1007/JHEP07(2025)235`. arXiv: `2411.05669`. URL: `https://cds.cern.ch/record/2916825`.

[63] T. L. collaboration. "Precision luminosity measurements at LHCb." In: *Journal of Instrumentation* 9.12 (Dec. 2014), P12005–P12005. ISSN: 1748-0221. DOI: `10.1088/1748-0221/9/12/p12005`. URL: `http://dx.doi.org/10.1088/1748-0221/9/12/P12005`.

[64] I. Belyaev et al. "The history of LHCb." In: *The European Physical Journal H* 46.1 (Mar. 2021). ISSN: 2102-6467. DOI: `10.1140/epjh/s13129-021-00002-z`. URL: `http://dx.doi.org/10.1140/epjh/s13129-021-00002-z`.

[65] R. Aaij et al. "The LHCb upgrade I." In: *JINST* 19.05 (2024). All figures and tables, along with any supplementary material and additional information, are available at http://lhcbproject.web.cern.ch/lhcbproject/Publications/LHCbProjectPublic/LHCb-DP-2022-002.html (LHCb public pages), P05065. DOI: `10.1088/1748-0221/19/05/P05065`. arXiv: `2305.10515`. URL: `https://cds.cern.ch/record/2859353`.

[66] C. Elsasser. *$\bar{b}b$ production angle plots*. Accessed:05/09/2025. URL: `%5Curl%7Bhttps://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb%5C_ProductionAngles.htm%7D`.

[67] R. Aaij et al. "Measurement of the b-Quark Production Cross Section in 7 and 13 TeV pp Collisions." In: *Physical Review Letters* 118.5 (Feb. 2017). ISSN: 1079-7114. DOI: `10.1103/physrevlett.118.052002`. URL: `http://dx.doi.org/10.1103/PhysRevLett.118.052002`.

[68] CERN. *High-Luminosity LHC (HL-LHC) Frequently Asked Questions*. Accessed: 2025-09-12. 2025. URL: `https://www.home.cern/resources/faqs/high-luminosity-lhc`.

[69] L. Collaboration. *LHCb Upgrade II: Letter of Intent*. Feb. 2017. URL: `https://cds.cern.ch/record/2250162`.

[70] L. Collabortaion. *Letter of Intent for the LHCb Upgrade*. Tech. rep. Geneva: CERN, 2011. URL: `https://cds.cern.ch/record/1333091`.

[71] B. Pietrzyk. *End of the 2025 proton-proton collision run*. LHCb Outreach, CERN. Nov. 2025. URL: `https://lhcb-outreach.web.cern.ch/2025/11/11/end-of-the-2025-proton-proton-collision-run/` (visited on 12/12/2025).

[72] I. Bediaga, J. M. De Miranda, F. Ferreira Rodrigues, et al. *Framework TDR for the LHCb Upgrade: Technical Design Report*. Tech. rep. LHCb Collaboration, 2012. URL: `https://cds.cern.ch/record/1443882`.

[73] R. Aaij et al. "Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC." In: *Journal of Instrumentation* 14.04 (Apr. 2019), P04013–P04013. ISSN: 1748-0221. DOI: `10.1088/1748-0221/14/04/p04013`. URL: `http://dx.doi.org/10.1088/1748-0221/14/04/P04013`.

[74] L. Collaboration. *LHCb VELO Upgrade Technical Design Report*. Tech. rep. LHCb Collaboration, 2013. DOI: `10.17181/CERN.4DGI.MZN4`. URL: `https://cds.cern.ch/record/1624070`.

[75] F. Ferrari. "LHCb upgrades." In: *PoS* LHCP2023 (2024), p. 224. DOI: `10.22323/1.450.0224`.

[76] S. Jakobsen. *LHCb SciFi: From performance requirements to an operational detector*. June 3, 2022. URL: `https://videos.cern.ch/record/3015689` (visited on 02/12/2025).

[77] Y. Zhao et al. *Track momentum resolution measurement for Run 3 at LHCb*. Tech. rep. Geneva: CERN, 2024. URL: `https://cds.cern.ch/record/2920030`.

[78] S. Celani. *Testing lepton flavour universality in $B^+ \to K^+\pi^+\pi^-\ell^+\ell^-$ decays with LHCb.* Aug. 2023.

[79] C. Grupen. *Physics of particle detection.* 1998. URL: `https://cds.cern.ch/record/332290`.

[80] M. Atzeni. “Performance of the particle identification system at LHCb.” In: *PoS* LHCP2024 (2024), p. 149. DOI: `10.22323/1.478.0149`.

[81] *Hadron PID Performance in 2024.* 2024. URL: `https://cds.cern.ch/record/2913359`.

[82] D. Derkach, M. Hushchyn, and N. Kazeev. “Machine Learning based Global Particle Identification Algorithms at the LHCb Experiment.” In: *EPJ Web Conf.* 214 (2019), p. 06011. DOI: `10.1051/epjconf/201921406011`. URL: `https://cds.cern.ch/record/2728397`.

[83] C. Fitzpatrick and V. V. Gligorov. *Anatomy of an upgrade event in the upgrade era, and implications for the LHCb trigger.* Tech. rep. Geneva: CERN, 2014. URL: `https://cds.cern.ch/record/1670985`.

[84] *RTA and DPA dataflow diagrams for Run 1, Run 2, and the upgraded LHCb detector.* 2020. URL: `https://cds.cern.ch/record/2730181`.

[85] *HLT1 trigger efficiencies in 2024 data.* 2024. URL: `https://cds.cern.ch/record/2912743`.

[86] Grazette, Luke, Olocco, Micol, and Matev, Rosen. “Infrastructure for deployment and evaluation of LHCb trigger configurations.” In: *EPJ Web Conf.* 337 (2025), p. 01116. DOI: `10.1051/epjconf/202533701116`. URL: `https://doi.org/10.1051/epjconf/202533701116`.

[87] R. Aaij et al. “Tesla: An application for real-time data analysis in High Energy Physics.” In: *Computer Physics Communications* 208 (Nov. 2016), pp. 35–42. ISSN: 0010-4655. DOI: `10.1016/j.cpc.2016.07.022`. URL: `http://dx.doi.org/10.1016/j.cpc.2016.07.022`.

[88] R. Aaij et al. “A comprehensive real-time analysis model at the LHCb experiment.” In: *Journal of Instrumentation* 14.04 (Apr. 2019), P04006–P04006. ISSN: 1748-0221. DOI: `10.1088/1748-0221/14/04/p04006`. URL: `http://dx.doi.org/10.1088/1748-0221/14/04/P04006`.

[89] L. collaboration. *Computing Model of the Upgrade LHCb experiment.* Tech. rep. Geneva: CERN, 2018. DOI: `10.17181/CERN.QOP4.57ON`. URL: `https://cds.cern.ch/record/2319756`.

[90] R. Aaij et al. “A comprehensive real-time analysis model at the LHCb experiment.” In: *Journal of Instrumentation* 14.04 (Apr. 2019), P04006–P04006. ISSN: 1748-0221. DOI: `10.1088/1748-0221/14/04/p04006`. URL: `http://dx.doi.org/10.1088/1748-0221/14/04/P04006`.

[91] T. Sjostrand, S. Mrenna, and P. Z. Skands. “PYTHIA 6.4 Physics and Manual.” In: *JHEP* 05 (2006), p. 026. DOI: `10.1088/1126-6708/2006/05/026`. arXiv: `hep-ph/0603175`.

[92] T. Sjostrand, S. Mrenna, and P. Z. Skands. "A Brief Introduction to PYTHIA 8.1." In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: `10.1016/j.cpc.2008.01.036`. arXiv: `0710.3820 [hep-ph]`.

[93] D. J. Lange. "The EvtGen particle decay simulation package." In: *Nucl. Instrum. Meth. A* 462 (2001), pp. 152–155. DOI: `10.1016/S0168-9002(01)00089-4`.

[94] P. Golonka and Z. Was. "PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays." In: *Eur. Phys. J. C* 45 (2006), pp. 97–107. DOI: `10.1140/epjc/s2005-02396-4`. arXiv: `hep-ph/0506026`.

[95] S. Agostinelli et al. "GEANT4: A Simulation toolkit." In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303. DOI: `10.1016/S0168-9002(03)01368-8`.

[96] J. Allison et al. "Geant4 developments and applications." In: *IEEE Trans. Nucl. Sci.* 53 (2006), pp. 270–278. DOI: `10.1109/TNS.2006.869826`.

[97] S. Miglioranzi et al. *The LHCb Simulation Application, Gauss: Design, Evolution and Experience.* Tech. rep. Geneva: CERN, 2011. URL: `https://cds.cern.ch/record/1322402`.

[98] M. Frank, F. Gaede, and P. Mato. "DD4hep: A Detector Description Toolkit for High Energy Physics Experiments." In: *J. Phys.: Conf. Ser.* 513 (2014), p. 022010. DOI: `10.1088/1742-6596/513/2/022010`. URL: `https://cds.cern.ch/record/1670270`.

[99] A. Malara. *Exploring jets: substructure and flavour tagging in CMS and ATLAS.* 2024. arXiv: `2410.14330 [hep-ex]`. URL: `https://arxiv.org/abs/2410.14330`.

[100] L. collaboration et al. *Inclusive B-meson flavour-tagging algorithm at LHCb.* 2025. arXiv: `2508.20180 [hep-ex]`. URL: `https://arxiv.org/abs/2508.20180`.

[101] J. Wishahi. *Flavour tagging conference plots: Schematic overview of the underlying principles of LHCb's flavour tagging algorithms.* Accessed via CERN Twiki. 2016. URL: `https://twiki.cern.ch/twiki/bin/view/LHCb/FlavourTaggingConferencePlots`.

[102] Q. Führing and V. Jevtić. *lhcb-ftcalib: A software package for the calibration of flavour-tagged LHCb data.* 2024. DOI: `10.5281/zenodo.12156328`. URL: `https://gitlab.cern.ch/lhcb-ft/lhcb_ftcalib`.

[103] D. Fazzini. "Flavour Tagging in the LHCb experiment." In: *PoS* LHCP2018 (2018), p. 230. DOI: `10.22323/1.321.0230`. URL: `https://cds.cern.ch/record/2668185`.

[104] CERN. *LHCb upgrades during LS2.* CERN. 2022. URL: `https://home.cern/press/2022/LHCb-upgrades-LS2` (visited on 11/18/2025).

[105] CERN. *LHC Run 3: physics at record energy starts tomorrow.* The page covers the start of Run 3, which began on 5 July 2022. CERN. July 2022. URL: `https://home.cern/press/2022/run-3` (visited on 11/18/2025).

[106] A. Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library." In: *Advances in Neural Information Processing Systems 32.* Ed. by H. Wallach et al. Vol. 32. Curran Associates, Inc., 2019, pp. 8024–8035. URL: `http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf`.

[107] D. P. Kingma and J. Ba. “Adam: A method for stochastic optimization.” In: *International Conference on Learning Representations.* 2015.

[108] F. Mölder et al. *Sustainable data analysis with Snakemake.* Oct. 2020. DOI: `10.5281/zenodo.4240139`. URL: `https://doi.org/10.5281/zenodo.4240139`.

[109] M. Pivk and F. L. Diberder. “: A statistical tool to unfold data distributions.” In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 555.1-2 (Dec. 2005), pp. 356–369. DOI: `10.1016/j.nima.2005.08.106`.

[110] M. Greenacre, P. J. F. Groenen, T. Hastie, et al. “Principal component analysis.” In: *Nature Reviews Methods Primers* 2.1 (2022), p. 100. DOI: `10.1038/s43586-022-00184-w`. URL: `https://doi.org/10.1038/s43586-022-00184-w`.

[111] S. Lundberg and S.-I. Lee. *A Unified Approach to Interpreting Model Predictions.* 2017. arXiv: `1705.07874 [cs.AI]`. URL: `https://arxiv.org/abs/1705.07874`.

[112] T.-C. Ogasa. *Optimization of Opposite-Side Flavor-Tagging Algorithms for the LHCb Upgrade.* Master Defense Presentation, Technische Universität Dortmund. Presented on 29 September 2025; Accessed: 17 November 2025. Sept. 2025. URL: `https://indico.e5.physik.tu-dortmund.de/event/4085/sessions/2659/attachments/3130/6020/Mastertalk_Thomas_Ogasa.pdf`.

[113] A. Collaboration. *Transforming jet flavour tagging at ATLAS.* 2025. arXiv: `2505.19689 [hep-ex]`. URL: `https://arxiv.org/abs/2505.19689`.

[114] H. Qu and L. Gouskos. “Jet tagging via particle clouds.” In: *Physical Review D* 101.5 (Mar. 2020). ISSN: 2470-0029. DOI: `10.1103/physrevd.101.056019`. URL: `http://dx.doi.org/10.1103/PhysRevD.101.056019`.

[115] M. Zaheer et al. *Deep Sets.* 2018. arXiv: `1703.06114 [cs.LG]`. URL: `https://arxiv.org/abs/1703.06114`.

[116] C. Prouve, N. Nolte, and C. Hasse. *Fast Inclusive Flavour Tagging at LHCb.* 2024. arXiv: `2404.14145 [hep-ex]`. URL: `https://arxiv.org/abs/2404.14145`.

[117] P. F. M. Olocco. *Fraud Detection Simulator.* `https://github.com/micolocco/fraud-detection-simulator`. Accessed: 2026-03-30. 2026.

[118] E. Ngai et al. “The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature.” In: *Decision Support Systems* 50.3 (2011). On quantitative methods for detection of financial fraud, pp. 559–569. ISSN: 0167-9236. DOI: `https://doi.org/10.1016/j.dss.2010.08.006`. URL: `https://www.sciencedirect.com/science/article/pii/S0167923610001302`.

[119] S.-H. Li et al. “Identifying the signs of fraudulent accounts using data mining techniques.” In: *Computers in Human Behavior* 28.3 (May 2012), pp. 1002–1013. ISSN: 0747-5632. DOI: `10.1016/j.chb.2012.01.002`. URL: `http://dx.doi.org/10.1016/j.chb.2012.01.002`.

[120] F. Carcillo et al. “Combining unsupervised and supervised learning in credit card fraud detection.” In: *Information Sciences* 557 (2021), pp. 317–331. ISSN: 0020-0255. DOI: `https://doi.org/10.1016/j.ins.2019.05.042`. URL: `https://www.sciencedirect.com/science/article/pii/S0020025519304451`.

[121] T. B. Brown et al. *Language Models are Few-Shot Learners*. 2020. arXiv: `2005.14165 [cs.CL]`. URL: `https://arxiv.org/abs/2005.14165`.

[122] J. Norris. *Markov Chains*. Cambridge University Press, 1997. ISBN: 9780521633963. DOI: `10.1017/CBO9780511810633`.

[123] L. R. Rabiner. "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition." In: *Proceedings of the IEEE* 77.2 (1989), pp. 257–286. DOI: `10.1109/5.18626`.

[124] A. Dal Pozzolo et al. "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy." In: *IEEE Transactions on Neural Networks and Learning Systems* 29.8 (Aug. 2018), pp. 3784–3797. ISSN: 2162-2388. DOI: `10.1109/TNNLS.2017.2736643`.

[125] J. Morgan and the Ollama Contributors. *Ollama: Get up and running with large language models locally*. Version 0.5.1. 2024. URL: `https://github.com/ollama/ollama`.

[126] IBM Corp. *IBM watsonx: The AI and Data Platform*. Accessed: 2025-12-17. 2024. URL: `https://www.ibm.com/watsonx`.

[127] DeepSeek-AI et al. *DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning*. 2025. arXiv: `2501.12948 [cs.CL]`. URL: `https://arxiv.org/abs/2501.12948`.

[128] A. Grattafiori et al. *The Llama 3 Herd of Models*. 2024. arXiv: `2407.21783 [cs.AI]`. URL: `https://arxiv.org/abs/2407.21783`.

[129] A. Q. Jiang et al. *Mistral 7B*. 2023. arXiv: `2310.06825 [cs.CL]`. URL: `https://arxiv.org/abs/2310.06825`.

[130] M. AI. *Mistral Large*. `https://mistral.ai/news/mistral-large/`. Accessed: 2025-08-21. 2024.