Spring 5-15-2022

Probing Lepton Flavour Universality With $\bz \to \kstarz \taup \taum$

Aravindhan Venkateswaran
Syracuse University

Abstract

Neutral current $b \to sll$ processes have long been known to be precise probes of physics beyond the Standard Model. A pattern of anomalies seen in $b \to s\mu^+\mu^-$ and in $b \to c\tau\nu$ data has hinted at the violation of lepton flavour universality. Effective field theories which attempt to find an explanation for these anomalies predict sizeable enhancements to the rate of $b \to s\tau^+\tau^-$ currents. We use pp collision data collected by the LHCb detector to estimate a sensitivity of $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) < 1.5 \times 10^{-3}$ at 95% C.L.

Probing lepton flavour universality with $B^0 \to K^{*0}\tau^+\tau^-$

by

Aravindhan Venkateswaran

B.Tech. and M.Tech., Indian Institute of Technology Madras, 2015

Dissertation

Submitted in partial fufillment of the requirements for a degree of

Doctor of Philosophy in Physics

Syracuse University

May 2022

Acknowledgements

When I started my undergraduate education in engineering in 2010, I did not imagine for even a second that I would one day be pursuing a PhD in physics. My journey in the world of particle physics and my growth as a scientist would not have been possible without the support of the multitude of people I met along the way.

I am very grateful to my late advisor Sheldon Stone, who guided me through my first six years of research in graduate school, and taught me a lot of what I know. I learnt the importance of developing a gut instinct for physics from him. His hands on support and willingness to discuss problems at length were invaluable to my growth.

I also thank my advisor, Matthew Rudolph, for being a constant source of support throughout the analysis which forms my thesis. I am grateful for his help in concluding my graduate studies, despite his having a full plate and other students to advise.

My gratitude goes out to Marina Artuso, Steve Blusk, Ray Mountain, Tomasz Skwarnicki and Ivan Pechenezhskiy for their generosity with time, advice and support for furthering my research career.

A vital part of my graduate school ecosystem were the fellow PhD students in my research group. Michael Wilkinson and Andy Beiter were the best officemates I could have asked for. I thank Michael Wilkinson and Matt Kelsey for putting up with my endless badgering and for never being too busy to share their hard earned knowledge. I am indebted to Swetha Bhagwat for helping my survive my first year in Syracuse and for making me appreciate the winter beauty of Upstate New York.

My time at CERN since March 2020 has been made much more pleasant thanks to Scott Ely, who never thought twice about going out of his way to help me.

I thank Harsh Dash, Hemanth Loku, Anand Parikh, Rajat Tiwari, Avinay Bhat and Eric King for their steadfast friendship.

Finally, I thank my parents Shanthi and Venkat Venkateswaran, for their love and support.

Contents

List of Tables

List of Figures

1 Introduction

The overarching objective of the field of physics is to explain the world around us. Sub-fields of physics use different frameworks in order to do this, and attempt to provide insight into diverse aspects of the natural world. Particle physics takes a largely reductionist viewpoint, and attempts to explain the universe by asking: “What are its fundamental components, and how do they interact with each other?”. This is by no means a new question. Philosophers going as far back as ancient Greece have been pondering the same. Democritus was believed to have been the first to hypothesize that matter (substance with mass) had to have a fundamental, indivisible component to it, and he dubbed this component “atom”, deriving from the Greek *atomos*, meaning uncuttable.

This elegant idea kept its grip on mankinds thinking, and it took until the 19th century for us to obtain empirical evidence for the particulate nature of matter. This came about with advances in our ability to measure and examine matter with ever increasing precision. The invention of the microscope, for example, allowed botanist Robert Brown to observe dust grains being jostled about while suspended in water. This is explained by them undergoing constant collisions with the molecules of water. Albert Einstein, and later, Francis Perrin, used this observation of Brownian motion to provide evidence of the particulate nature of matter.

Our knowledge of the forces of nature is not a new one either. Anyone who has flung a stone and witnessed it in free fall, held two magnets close to each other, or heard sound, has felt them. Our curiosity to look behind the curtain, and ask how these forces arise, has led to our current understanding of four fundamental forces (though we might have reasons to look for a fifth), that explain all the phenomena that we observe.

As our ability to look at matter on smaller and smaller length scales grew, so did our understanding of its fundamental structure. Since the 19th century, we know that atoms (despite their name) are not fundamental, but are composed of protons and electrons. Not only this, but we also now know that protons are not fundamental, but are composed of further sub-units, which we call quarks. Electrons, to the best of our current knowledge, are fundamental. Today, we know of 24 fundamental matter particles, and 5 force carrying particles which are responsible for the interactions between these matter particles. The theory that describes all of this, known as the Standard Model (SM) of particle physics, is our current best description of the natural world. But we also know that this theory does not describe everything that we see in the universe. It does not explain dark matter or dark energy. Gravity, arguably the most felt force in our day to day life, is left out of it completely. The Standard Model also stands at odds with the existence of any matter at all. All of this leads us inexorably towards the conclusion that there is physics *beyond* the Standard Model, the so called New Physics.

The job of experimental particle physicists is to take measurements of fundemantal processes and compare them to the predictions made by the Standard Model, in the tradition of empirical science. The modern day "microscopes" used to study the properties of fundamental particles are particle colliders and detectors. The Large Hadron Collider (LHC) at CERN is the worlds largest and highest energy particle collider. The LHC collides bunches of protons flying near the speed of light, and the explosion of particles coming out of this collision are studied by detectors built around the collision point. A highly oversimplified, and somewhat comical, way to explain this is to compare it to smashing clocks together to be able to understand what comprises their innards.

This thesis is organized as follows: we start by delving into the theoretical background (including a summary of the Standard Model) pertaining to the experimental measurements that follow. A description of the LHCb detector, used to record the data analyzed in this thesis, follows. The main focus of this thesis, a probe of lepton flavour universality through the $B^0 \to K^{*0}\tau^+\tau^-$ decay mode, is then presented. In the Appendix, after supplementary material pertaining to the $B^0 \to K^{*0}\tau^+\tau^-$ work, a chapter is presented on earlier work done by the author in graduate school, investigating isospin amplitudes in b baryon decays.

2 Theoretical Background

In this section, we start with a summary of the Standard Model, before presenting the aspects of theory pertinent to $B^0 \to K^{*0}\tau^+\tau^-$.

2.1 Standard Model

The Standard Model is the current best description of the fundamental constituents of the universe and their interactions, sans gravity. A theoretical framework that has evolved over the 20th century, it is a remarkably successful theory that has been validated to great precision time and time over by experimental measurements of its parameters, and confirmations of its predictions.

The Standard Model is built in the framework of a quantum field theory (QFT), in which both elementary particles and their interactions are described in terms of fields. The matter content of the SM is described in terms of spin 1/2 fields, known as fermions, while their interactions are described in terms of spin 1 fields, known as bosons.

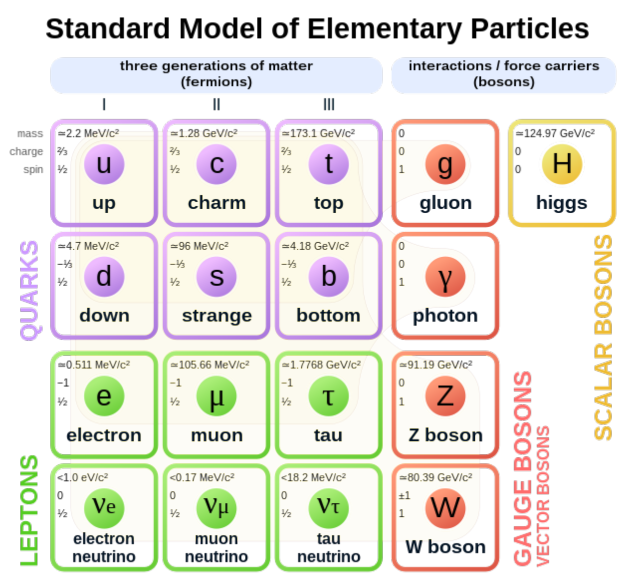


Figure 1: The particle content of the Standard Model is shown, broken into the force carrying bosons and three generations of fermionic matter. [1]

The fermionic matter can be separated into two classes, quarks and leptons, based on their ability to feel the strong force. Quarks come in six varieties (known as flavours in the jargon of the SM) (along with their anti-matter counterparts), arranged in three generations of increasing mass. Nucleons (protons and neutrons), once thought to be elementary particles, are now known to be composed of three quarks each (as depicted in Fig.2). Quarks carry fractional charges of 2/3 and 1/3, and are never seen to exist independently outside of hadrons (the collective term for composite particles made from quarks), a property known as confinement.

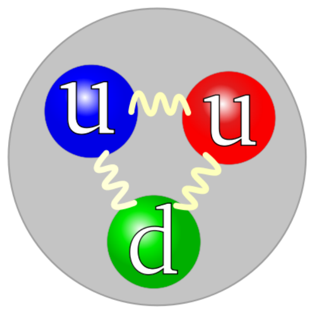


Figure 2: Quark model of a proton. It is composed of two up (u) quarks and one down (d) quark. The wavy lines joining the quarks are gluons. [2]

Leptons on the other hand, are a kind of particle that physicists have been more readily familiar with, being the family to which the electron belongs. There are 6 leptons, three of them charged: the electron, muon and tau, and three of them neutral: the electron neutrino, muon neutrino and tau neutrino (again with their antimatter counterparts). All leptons, to the best of our current knowledge, are elementary particles, and come in three generations of increasing mass just like the quarks. Neutrinos are remarkably light particles, and up until very recently were thought to be massless.

In the paradigm of QFT, the fundamental interactions between the matter particles are mediated by bosonic fields. Each interaction has a type of charge associated with it (where charge refers to a discretely valued quantity that is conserved by that interaction).

The *electromagnetic interaction* is mediated by the massless photon. It is felt by all particles which carry an electric charge, which happens to also be the conserved charge associated with this interaction. It is a long range interaction that is responsible for most of the phenomena felt in day to day life, along with gravity.

The *weak interaction* is mediated by three massive vector fields, the $W^{\pm}$ and the Z^0, and it is felt by all fermions, which carry a weak charge associated with the weak interaction. The weak interaction is a very short range force, operating only over sub-nuclear distance scales. It is named so because it is several orders of magnitude weaker than the electromagnetic force over comparable distance scales. The weak interaction is responsible for nuclear phenomena like the beta decay of the neutron, nuclear fusion and fission. Weak interactions mediated by the $W^{\pm}$ are called charged current interactions, and change the flavour of the quarks that participate; while those mediated by the Z^0 are called neutral current interactions, and do not change the flavour of quarks. Weak interactions become especially significant for neutrinos, since they do not interact via the strong and electromagnetic interactions.

One of the important features of the weak interaction is the relative rates with which the $W^{\pm}$ bosons couple to the different quarks and leptons. It is known in the SM that the gauge bosons couple differently to different pairs of quarks, a phenomenon encoded in the Cabibbo-Kobayashi-Maskawa matrix. But when it comes to leptons, a concept known as **lepton flavour universality** is deeply ingrained in the Standard Model, according to which the weak gauge bosons interact equally with the three generations of leptons, up to differences due to the mass of the lepton. At high energies, the electromagnetic force and the weak force couple together into a single electroweak force.

The *strong interaction* is mediated by massless spin 1 gluons, and it is felt by both the quarks and gluons. The strong force stands apart from the other interactions in that its mediating particle is able to participate in the interaction. The strong nuclear force is responsible for binding the quarks inside hadrons. It is a very short range force, operating only over length scales comparable to the size of the nucleon. It is so named because it is approximately 137 times stronger than the electromagnetic force, and $\sim 10^6$ times stronger than the weak force. A residual force from the interactions between the quarks also binds the protons and neutrons in a nucleus together, overcoming the electromagnetic repulsion between the protons. The charge associated with the strong interaction, carried by all quarks and gluons, is whimsically known as colour. The colour charge comes in six varieties: red, blue and green (along with their “anti”-colours), which conveniently add up to white (or colourless), since only colourless objects are allowed to exist in nature. Besides the common three quark combinations of protons and neutrons, QCD allows for meson (a combination of a quark and an antiquark), baryons (combinations of three quarks), as well as tetraquarks (4 quarks combinations) and pentaquarks (5 quark combinations).

Besides electrons and protons, all other particles in the Standard Model are unstable, meaning that they decay to lighter, more stable particles. The ways in which a particle is allowed to decay is governed by a number of conservation rules and symmetries, as well as which of the fundamental interactions the particle feels. A particular decaying particle (usually referred to as the “mother”) and a specific final state (composed of “daughter particles”) together make up a decay mode, and the probability that a mother decays in a specific mode of all those available to the mother is known as the branching fraction of that decay mode. The Standard Model often makes precise predictions for the branching fractions of specific decay modes, as well as for the ratios of branching fractions of similar modes, and measurements of the branching fractions and ratios constitute important tests of the Standard Model, since these can be modified by the presence of new particles.

2.1.1 Problems with the Standard Model

While the Standard Model is an incredibly successful theory, it is not complete. Cosmological observations of gravitational lensing [40] and galactic rotation curves [41] tell us that the SM does not explain 95% of the energy density in the universe, 68% of which is known as dark matter and 27% as dark energy. The SM does not provide candidate particles for either of these.

Further, the Standard Model is unable to provide a satisfactory explanation for the observed matter-antimatter asymmetry in the universe. To the best of our current knowledge, the universe contains orders of magnitude more matter than antimatter, and the SM sources of CP violation are not enough to explain this asymmetry, suggesting the need for sources of CP violation from beyond the Standard Model.

Neutrinos in the Standard Model were initially supposed to be massless, and fixed in flavour. But since the 1960s, we have learned through experiments that neutrinos not only must have a non zero (albeit extremely small) mass, they also oscillate from the flavour in which they were produced into other flavours. For example, an electron neutrino can oscillate into a muon neutrino. Allowing neutrinos to have mass also changes some of their other fundamental properties. It means that either right handed neutrinos must also exist, or that the neutrino is its own antiparticle. These different scenarios would have very different implications for the way in which the neutrino obtains its mass, which is an open question at the moment.

For these reasons, and more, it is known without question that the SM is incomplete, and that there must be physics beyond it at higher energies, commonly referred to as New Physics (NP). In this perspective, the SM is viewed simply as the lower energy manifestation of some more complete theory.

2.2 Effective Field Theory of flavour changing neutral currents

Quark transitions which change the flavour of the quark without changing its electric charge are known as flavour changing neutral currents (somewhat misleadingly, since they are generally mediated by the $W^{\pm}$ bosons). In the Standard Model, FCNCs are heavily suppressed by the GIM mechanism [42], and only occur at the loop level with extremely small rates. The smallness of their Standard Model rate and the loop mechanism of these transitions make them extremely sensitive to the presence of new particles at energy scales much higher than the scale of the interaction. Since direct searches for new particles have not yielded positive results after the discovery of the Higgs Boson, FCNCs have come to the fore as tools for indirect searches of NP. FCNCs that are particularly interesting to LHCb are decays of the form $b \to sll$, which can be studied through various decay channels.

From the theoretical point, it is very interesting to be able to study these decays in a model independent formalism through Effective Field Theories (EFTs). One of the defining features of EFT is the separation of physics at low energy scales from that at high energy scales. This is intuitively done by writing down an effective Hamiltonian, where the heavy degrees of freedom (top quark, W and Z bosons, Higgs, and *potential heavy new particles*) have been integrated out into high energy Wilson Coefficients C_i, leaving behind a set of operators O_i describing the physics at lower energies [35] [43].

$$H_{\text{eff}} = -4\frac{G_F}{\sqrt{2}} V_{tb} V_{ts}^* \sum_i C_i O_i \quad \text{(up to corrections prop to } V_{ub} V_{us}^*) \tag{1}$$

Each operator describes a different sort of interaction contributing to the decay, and the coefficients can be thought of as desribing the strength of that interaction. In the SM, H_{eff} contains 10 operators. In the context of $b \to sll$ decays, the operators of importance are

$$O_7 = \frac{e}{16\pi^2} m_b (\bar{s}\sigma_{\mu\nu} P_R b) F^{\mu\nu} \qquad O_{7'} = \frac{e}{16\pi^2} m_b (\bar{s}\sigma_{\mu\nu} P_L b) F^{\mu\nu} \tag{2}$$

$$O_9 = \frac{e}{16\pi^2} (\bar{s}\gamma_\mu P_L b)(\bar{l}\gamma_\mu l) \qquad O_{9'} = \frac{e}{16\pi^2} (\bar{s}\gamma_\mu P_R b)(\bar{l}\gamma_\mu l) \tag{3}$$

$$O_{10} = \frac{e}{16\pi^2} (\bar{s}\gamma_\mu P_L b)(\bar{l}\gamma_\mu \gamma_5 l) \qquad O_{10'} = \frac{e}{16\pi^2} (\bar{s}\gamma_\mu P_R b)(\bar{l}\gamma_\mu \gamma_5 l) \tag{4}$$

where $P_{L,R} = (1 \mp \gamma_5)/2$ are the projection operators for left and right handed chiralities and $m_b \equiv m_b(\mu_b)$ denotes the running b mass in the $\overline{\text{MS}}$ renormalization scheme. O_7 is the electromagnetic operator describing the interaction of the b and s with the photon. O_9 and O_{10} are semileptonic operators which describe the interaction of the b and s with the charged leptons, and the primed versions of the operators are the chirally flipped versions of the unprimed versions. The chirality flipped operators have negligible coefficients in the SM. Contributions from physics beyond the SM could manifest as modifications in the value of

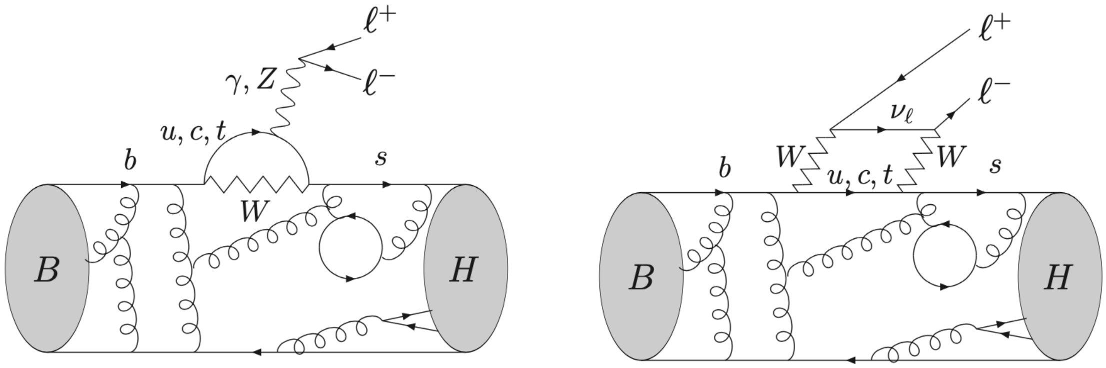


Figure 3: Illustration of two diagrams for a $b \to sl^+l^-$ transition in the Standard Model, as seen in the hadronic level, in the case of a B meson decaying into an unspecified H meson. [3]

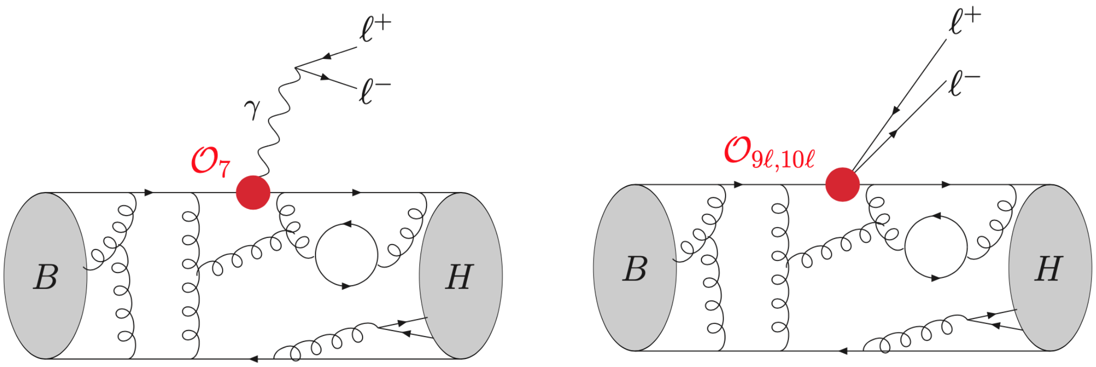


Figure 4: Illustration of two diagrams for a $b \to sl^+l^-$ transition in the effective field theory approach, in the case of a B meson decaying into an unspecified H meson. The red dots indicate the local operators. [3]

the Wilson coefficients $C_{7,9,10}$ or make other operators contribute in a significant manner (such as $O_{7',9',10'}$). Figs.3 and 4 illustrate the $b \to sl^+l^-$ transition in the SM at the hadronic level and in the effective field theory approach respectively.

In the Standard Model, $C_9 \approx 4.1$ and $C_{10} \approx -4.3$ are equal for the electron, muon and tau leptons. This reflects the idea of *lepton flavour universality* built into the SM theory. But it is conceivable to have physics beyond the Standard Model coupling preferentially to some generation, which would break the equality of the Wilson coefficients for the three lepton generations. NP contributions to the Wilson coefficients pertaining to lepton l can simply be parametrized as

$$C_{il} = C_{il}^{\mathrm{SM}} + C_{il}^{\mathrm{NP}}. \tag{5}$$

The C_i^{NP} contributions to the Wilson coefficients are often taken as real [3], stemming from the non-observation of CP violation in these processes. While NP sources of CP violation are broadly required in physics, the CP asymmetries in $b \to s l^+ l^-$ such as $B^+ \to K^+ \mu^+ \mu^-$ and $B^0 \to K^{*0} \mu^+ \mu^-$ have been measured to be compatible with zero.

3 LHCb Experiment

3.1 Large Hadron Collider

The Large Hadron Collider (LHC) is a high energy particle collider, situated astride the border between France and Switzerland, housed in CERN, the European Organization for Nuclear Research. The LHC is a circular proton-proton collider [1], measuring 27 km in circumference, buried 175m below ground level. The machine accelerates two proton beams, one moving clockwise and the other counter-clockwise in separate beam pipes, each to an energy of 6.5 TeV currently, allowing for a 13 TeV center-of-mass collision energy. The energy available at the point of collision converts into a plethora of different particles which fly outwards and are captured by detectors surrounding the collision point (Einstein's mass energy equivalence teaches us that energy can be converted into mass, which is what happens here). There are 4 such collision points around the ring, corresponding to the ALICE, ATLAS, CMS and LHCb experiments.

The acceleration of the protons does not start with the LHC. A number of acceleration stages precede it. Hydrogen anions (hydrogen atoms with an extra electron) are accelerated to 160 MeV by the Linear accelerator, Linac4 (Linac2 up to 2020), following which the two electrons in the ions are stripped, leaving only protons to be injected into the Proton Synchrotron Booster (PSB), where they are accelerated to 2 GeV. Subsequently, the Proton Synchrotron (PS) and Super Proton Synchrotron (SPS) take over, accelerating the beams to 26 GeV and 450 GeV respectively. It is at this point that the beams are ready to be injected into the LHC. A schematic depicting the accelerator complex is shown in Fig. 5.

The acceleration of the proton beams inside the LHC is achieved through oscillating electric fields in radio frequency (RF) cavities. It is important to note that the beam is not a continuous set of protons, but is rather split up into bunches. The oscillation of the electric

[1] for most of the time, a part of the time is spent colliding heavy ions

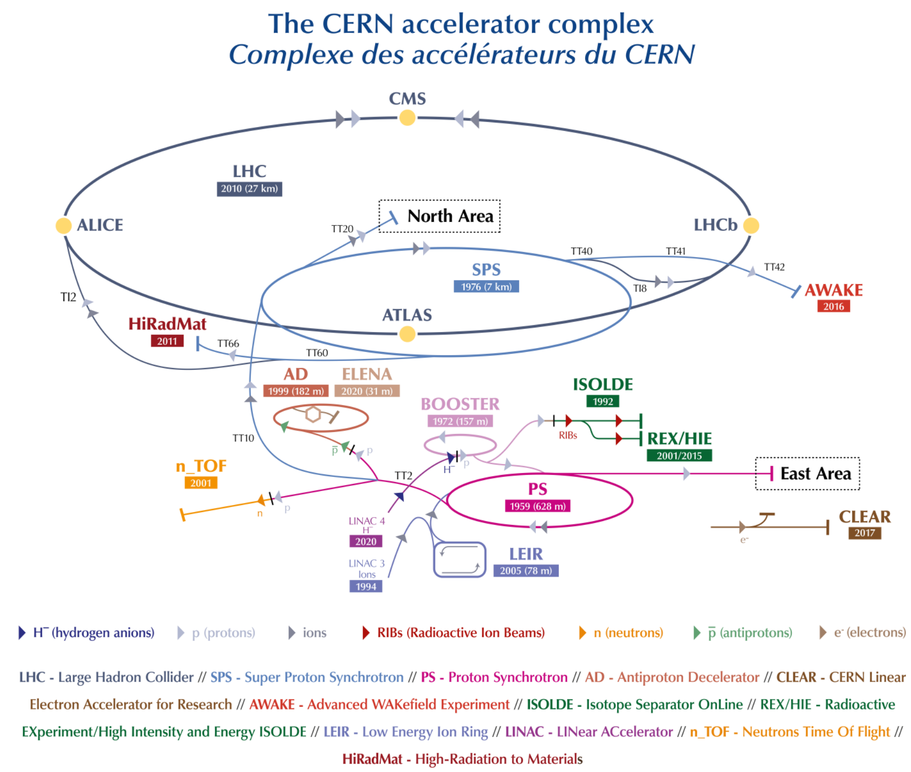


Figure 5: A schematic of the LHC accelarator complex as of 2019, detailing that various stages of proton acceleration. [4]

field is tuned to be in phase with the arrival of the bunches. The beams are guided around the ring by use of a magnetic field maintained by thousands of superconducting electromagnets of different varieties. These include dipole magnets which bend the beams and quadrupole magnets which focus them. The electromagnets are cooled to an extremely low temperature of -271.3 Celsius using liquid helium, in order to enable superconductivity.

During Run 1 of the LHC, the beams were collided at a center-of-mass energy of 7 TeV in 2011 and 8 TeV in 2012. After two years of maintainence and upgrades, between 2015-2018, in Run 2 of the LHC, the center-mass-energy was increased to 13 TeV.

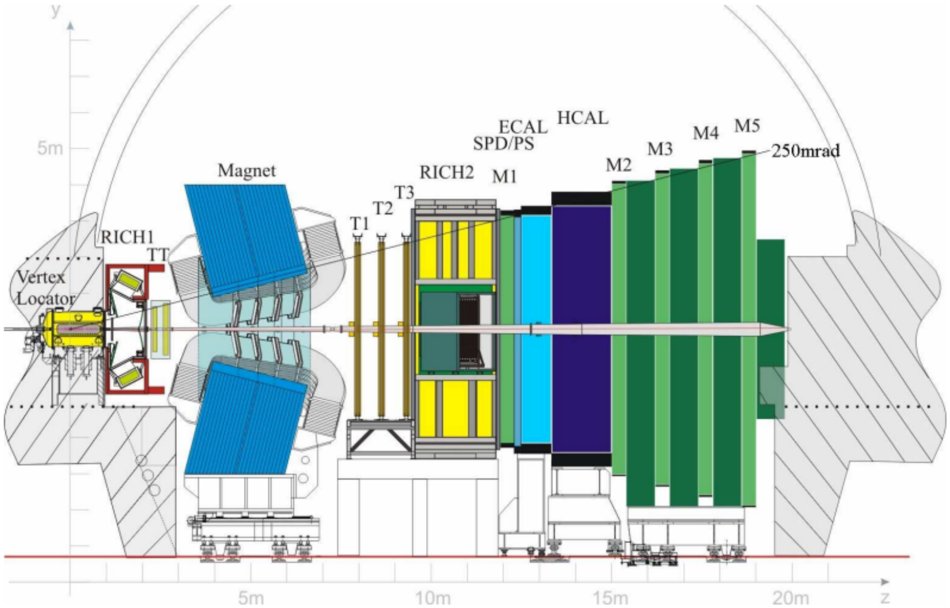


Figure 6: A schematic of the LHCb detector, shown in side view [5]

3.2 LHCb Detector

As mentioned in the previous section, the LHCb detector is one of the four detectors on the LHC ring. It is constructed as a forward arm spectrometer to make precision measurements of b and c decays and CP violation phenomena. An introduction to the primary features of the detector are given in this section, and its subdetectors are further expanded on in the following sections.

Fig. 6 shows the schematic of the LHCb detector. It is constructed to have an angular acceptance of 10-300 mrad in the magnet bending plane, and up to 250 mrad in the vertical plane, corresponding to a pseudorapidity range of $2-5$, where pseudorapidity (depicted by η) is defined as

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right], \tag{6}$$

where θ is the angle w.r.t. the beam axis. Compared to the CMS and ATLAS detectors, which are hermetic, the LHCb detector is constructed to operate in only the “forward” region. This exploits the production dynamics of b quarks at LHC collision energies, where they are predominantly produced along the beam direction, as demonstrated by PYTHIA simulations, illustrated in 7.

The LHCb detector receives pp collisions at an instantaneous luminosity ranging from $(2-4)\times10^{32}\text{cm}^{-2}\text{s}^{-1}$. During the Run 1 data collection at 7 and 8 TeV, a dataset corresponding to an integrated luminosity of around $3\,\text{fb}^{-1}$ was collected, and during the Run 2 data collection at 13 TeV (which is the data used in this thesis), a dataset corresponding to an integrated luminosity of around $5.5\,\text{fb}^{-1}$ was collected.

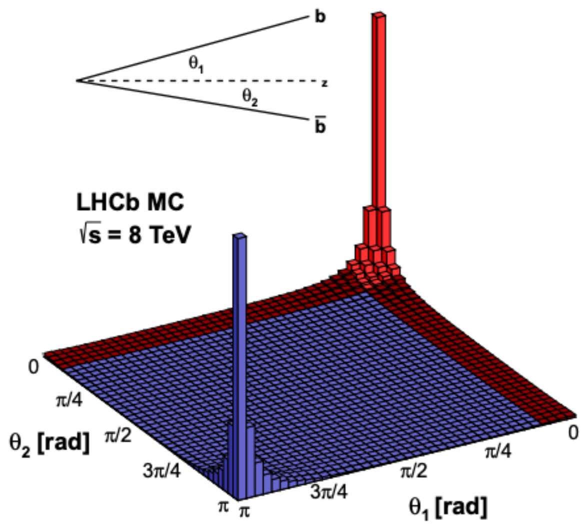


Figure 7: The production angles for $b\bar{b}$ quark pairs with respect to the beam line in pp collisions, as simulated with Pythia. The LHCb acceptance is highlighted in red.

3.2.1 Magnet

One of the key measurements made by the LHCb detector is the momenta of tracks. In order to do this, the detector employs a magnet to bend these charged tracks. By measuring the curvature of the tracks, it is possible to measure their momenta. LHCb uses a warm dipole magnet design with saddle-shaped coils in a window-frame yoke with sloping poles to match the design acceptance described in the earlier section. The magnetic field is vertically aligned (along the y axis in the LHCb coorinate system). Charged tracks of 10 m length in

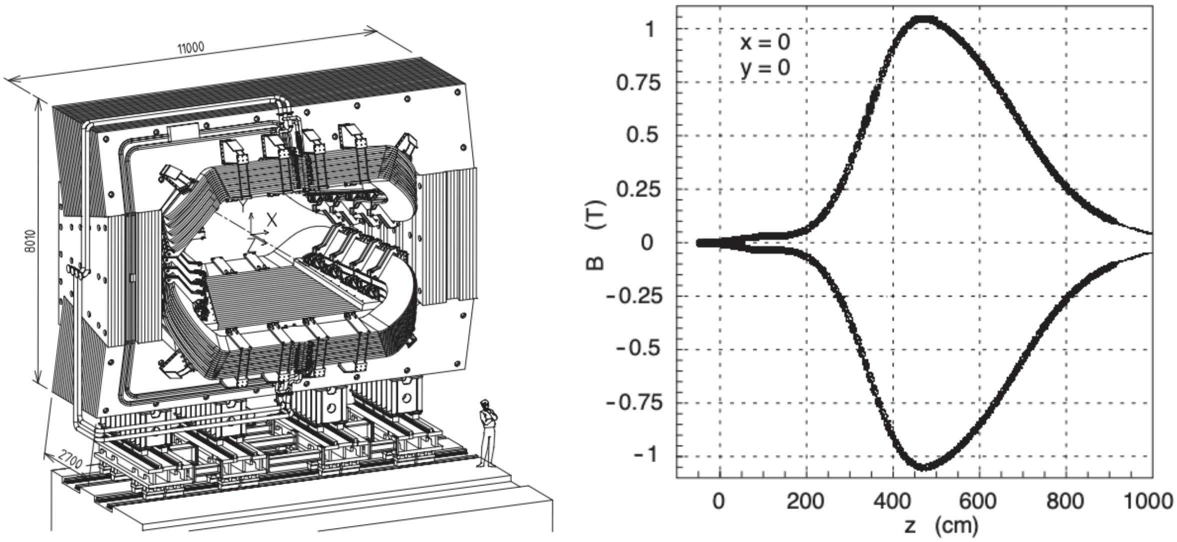


Figure 8: (Left)A schematic of the LHCb magnet, in perspective view. The interaction point is situated behind the magnet in this view. (Right) The vertical (y) component of the magnetic field is shown as a function of z (along the beam) [5]

the field see an integrated magnetic field of 4 Tm. The variation of the vertical component of the magnetic field along the z axis of the detector (parallel to the beam line) is shown in Fig 8. During data taking, the magnet spends equal amounts of time in the "up" and "down" polarities to avoid systematic biases, especially important in studies of CP violation observables, which can be affected by detector asymmetries.

3.2.2 Tracking System

The LHCb tracking system is composed of the VELO detector around the *pp* collision point (also known as the primary vertex or PV) and four planar detectors: the Tracker Turicensis (TT) before the magnet and the T1, T2 and T3 stations after it. Silicon based sensors are employed in the VELO, TT, and the regions of T1, T2 and T3 close to the beam pipe. The TT and silicon based regions of the three downstream trackers were developed collectively under the umbrella of the Silicon Tracker (ST). The outer regions of the downstream tracking stations, known as the Outer Tracker (OT) use straw tubes. The job of the tracking system is to capture the hits of the charged tracks passing through them. The precise positional measurements of the tracking system, in conjunction with the magnet allow the measurement of the momenta of the charged tracks. The system is capable of providing a momentum resolution of 0.4% on 5 GeV momenta tracks and 0.6% on 100 GeV momenta tracks. The descriptions of the different sub-components of the tracking system are provided in the following sections.

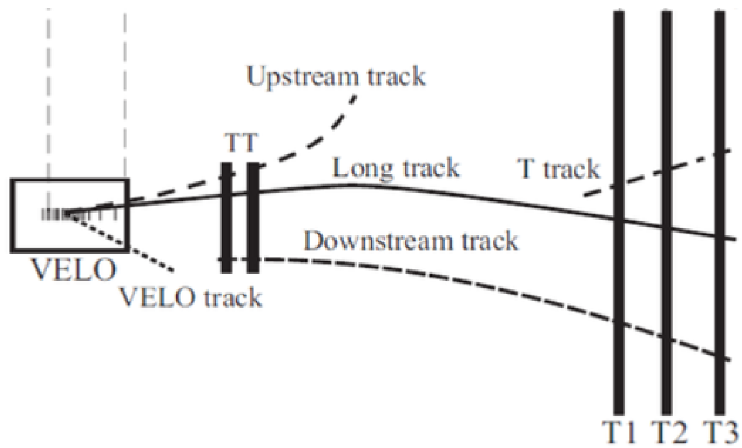


Figure 9: The LHCb tracking system is shown schematically, along with the terminology for different types of LHCb tracks. [6]

3.2.3 Vertex Locator

The Vertex Locator (VELO) detector sits around the pp interaction point and provides measurements of the charged tracks originating from this interaction region. These coordinates are key to measuring the production and decay vertices of b and c hadrons produced in the primary collision. b-hadrons especially are relatively long lived, leading to decay vertices displaced from the interaction region. The precise measurements of these displaced vertices play an important role in the High Level Trigger (HLT) where they are used to selectively retain data that contains b-hadron decays.

A set of 42 silicon modules form the VELO, half measuring the r coordinates of the tracks and the other half measuring the ϕ coordinates. The modules are arranged in stations, each consisting of an r module and a ϕ module mounted back to back. The r modules have sensors arranged at constant radius, while the ϕ modules have their sensors arranged radially. The position of the station which records the hit provides the z coordinate of the hit, allowing for the complete three dimensional measurement of the hit in space. The stations are arranged on both sides of the beam pipe in a staggered manner along the beam direction as shown in Fig. 10. A requirement that tracks within the acceptance of the detector cross atleast 3 stations dictates the design of the detector. A thin walled corrugated aluminium sheet separates the vacuum in which the VELO sensors are placed from the LHC machine vacuum, in order to minimize the multiple scattering that charged particles undergo when they pass through matter.

In order to maximize the acceptance of the detector, the VELO was designed to sit very close to the beam pipe. It sits so close radially that it must be retracted (as shown in Fig 10) during the injection of the beam, as the LHC focuses the beam over time. Since it sits so close to the beam pipe, the VELO is designed to operate in a harsh radiation environment with the radiation in its most susceptible region being equal to that of 1 MeV neutrons with a flux of $1.3 \times 10^{14} \mathrm{n}_{\mathrm{eq}}/\mathrm{cm}^2$. In order to dissipate the heat generated by the sensors electronics and to ameliorate the effects of radiation induced damage, the VELO cooling system is designed to keep the sensors at a temperature between -10C and 0C. The performance of the VELO can be summarized by the $20\mu m$ precision with which it measures the impact parameter of high p_{T} tracks.

3.2.4 Silicon Tracker

As mentioned earlier, the Tracker Turicensis (TT) and the Inner Tracker (IT) collectively comprise the Silicon Tracker, so named because it employs silicon micro-strip detectors. The TT is $1.5m \times 1.3m$ and sits upstream of the magnet, while the IT forms a cross shaped region 120 cm wide and 40 cm high in the inner regions of the three downstream tracking stations. Each station of the Silicon Tracker consists of four planes, with the outer two planes having their strips oriented vertically, and the inner two planes having their strips rotated by a

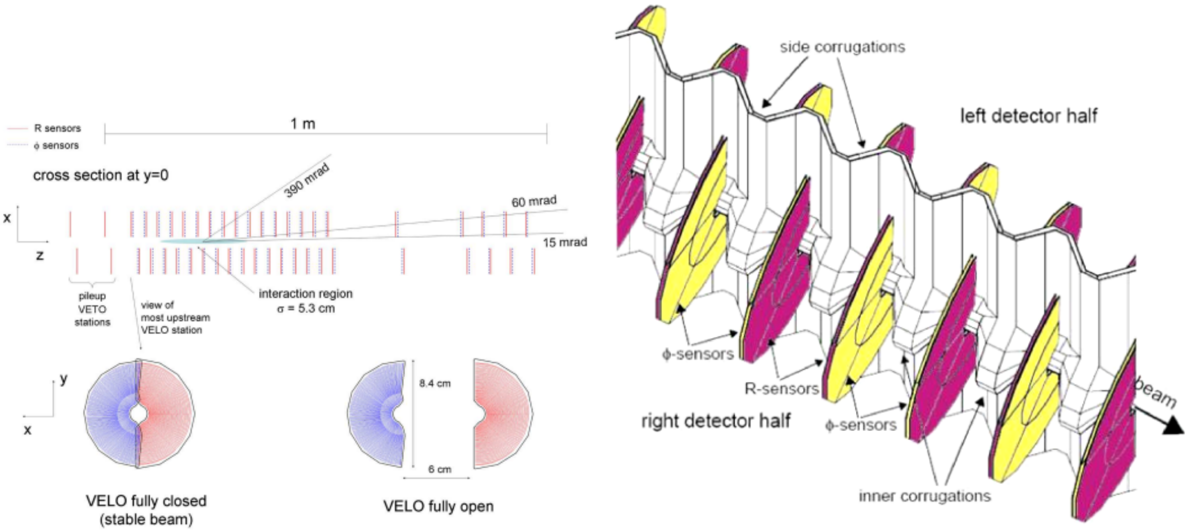


Figure 10: (Left) The top image shows the view of the VELO detector in the x−z plane, and the bottom image shows the front view of its modules in the x−y plane, in both open and closed positions. (Right) A perspective view of the VELO sensors, around the beam pipe.

stereo angle of 5°. The vertically oriented strips provide x information on the hits of charged tracks, while the stereo strips combine with the vertical strips to provide the y coordinates of the hits. As before with the VELO, the position of the plane provides the z coordinates of the hits. Fig. 11 shows on the left a stereo plane of the TT and on the right a vertical plane of the IT.

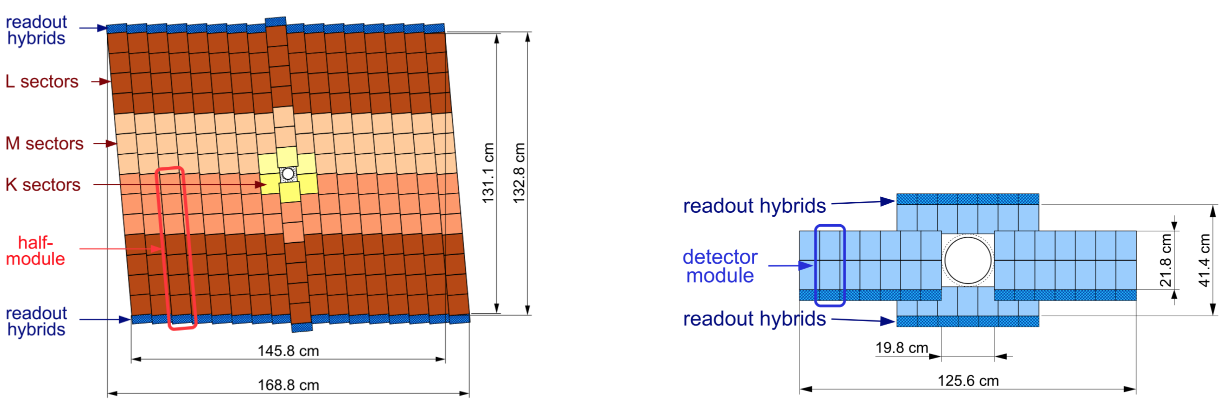


Figure 11: (Left) A schematic of the third detection layer of the TT [5]. (Right) A schematic of an x detection layer in the IT [5].

Based on simulation studies, a $200\mu m$ pitch was chosen for the sensors in the ST, to satisfy the requirement of a $50\mu m$ single hit resolution in the sensors. The strip geometries are designed to satisfy requirements of maximum particle occupancies while minimizing the number of readout channels. The sensors were designed to measure signals with a signal to noise ratio (SNR) of 12:1, even after the radiation damage accrued from ten years of nominal running. The front end electronics of the detector were chosen to be speedy enough that the 25 ns bunch crossing interval of the LHC would not result in pileups. As with the VELO, the sensors of the ST are designed to survive in a harsh radiation environment, with the inner regions of the TT seeing $5 \times 10^{14}\mathrm{n_{eq}/cm^2}$ 1 MeV neutron equivalent flux over 10 years and the IT seeing $9 \times 10^{12}\mathrm{n_{eq}/cm^2}$ 1 MeV neutron equivalent flux over the same period.

The four detection planes of the TT are surrounded by a thermally and electrically insulated volume, which also shields the detector from light. In order to siphon away the heat due to electronics, the detector is maintained at a temperature below 5C. To prevent condensation occurring on the sensitive electronics at these low temperatures, the detector volume is continuously flushed with dry nitrogen. The basic building block of the TT is a half module, half the height of the LHCb acceptance, composed of 7 sensors. Neighbouring modules are staggered in z and overlap in x to avoid gaps in the acceptance of the detector and to aid in the alignment of the modules. In order to minimize multiple scattering of the charged tracks (which is the dominant effect on the momentum resolution of the detector), the material in the active area of the detector is reduced by keeping the front end electronics and all the cooling infrastructure and module supports outside the active area of the detector.

The IT is formed by four detector boxes arranged around the beam pipe, downstream of the magnet. Each box contains four detection layers in a dry, light tight and electrically and thermally insulated environment. Each detection layer consists of seven modules, with adjacent modules being staggered in z and overlapping in x for the same reasons as in the TT. The modules in the boxes above and below the beam pipe are formed by a single sensor and its readout hybrid, whereas the modules in the boxes on the sides of the beam pipe consist of two sensors each along a readout hybrid. Unlike in the TT, it was not possible in the IT to keep readout electronics and mechanical supports outside the detector acceptance, so a significant effort was invested to ensure that the material budget here was as small as possible. The layout and dimensions of a single IT detection layer is shown in Fig. 11.

The boundary between the IT and OT has been chosen based on considerations of track reconstruction efficiency and detector occupancy.

3.2.5 Outer Tracker

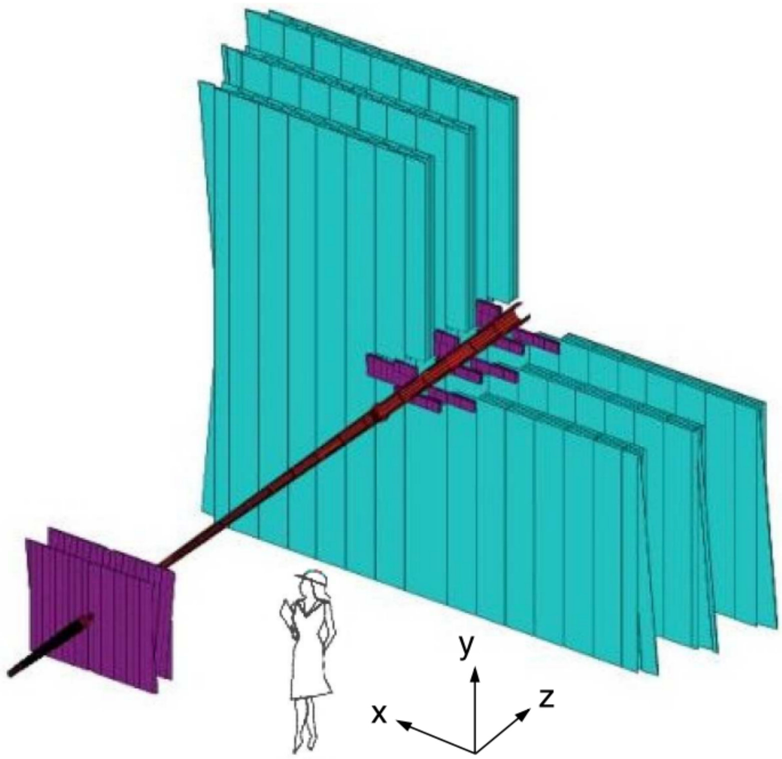


Figure 12: A perspective view of the LHCb tracking system is presented, with a cross sectional cut out in the $-x$,$+y$ quadrant of the downstream tracking stations. The stations of the ST are shown in purple, while the OT is shown in blue [5].

The outer tracker is a drift time detector built as an array of gas tight straw tube modules, constructed around the IT. The straw tubes are filled with a counting gas mixture of Argon (70%) and CO_2 (30%). Each tube has a diameter of 4.9 mm, and contains an anode wire in its center and an outer cathode layer between which an electric field acts. Charged particles which pass through the gas ionize it and the resulting electrons and ions are collected (along with the resulting avalanche) with a maximum drift time of 50 ns (equivalent to two bunch crossings) provide a drift coordinate resolution of 200 μm.

The three OT stations each consist of 4 layers arranged in the familiar (x-u-v-x) theme, with the outer layers being vertical and the inner layers being stereo rotated by $5°$ with respect to the vertical. The total active area of a station is $5.97 \times 4.85 m^2$. Each station is split into two halves, one on each side of the beam pipe, supported by aluminium structures in a manner that allows them to be retracted away from the beam pipe. Like other parts of the detector, the OT too is built to be radiation hard, and to deliver quality data over ten years of nominal running. The layout of the OT can be seen in Fig. 12, in perspective with the TT, IT and beam pipe.

3.2.6 Ring Imaging Cherenkov Detectors

Particle identification at LHCb is done via the Ring Imaging Cherenkov Detectors (RICH). There are two such detectors, one located upstream (called RICH1) of the magnet capable of handling lower momentum tracks within the range $1 - 60$ GeV, and the second located downstream of the magnet (called RICH 2) covering the higher momentum range from 15 GeV to 100 GeV. Both detectors rely on the Cherenkov radiation emitted by charged particles travelling through a dielectric medium at speeds greater than the phase velocity of light in that medium. The radiation is emitted in the form of a cone of light (usually from the blue end of the visible spectrum), whose opening angle θ depends on the velocity of the particle $\theta = \frac{c}{nv}$, where n is the refractive index of the medium. The larger the velocity of the particle, the smaller the cone. Hybrid Photon Detectors (HPDs) are used to detect the Cherenkov light and measure the velocity of the particle. In combination with the momentum of the particle measured by the tracking system, the mass of the particle can be estimated, allowing physicists to be able to separate protons, kaons, pions and muons from one another with some success. The level of separation attainable depends on the momentum of the tracks.

The schematics of the RICH1 and RICH2 detectors are shown in Fig.13. RICH1 uses aerogel and C_4F_{10} as its media while RICH2 uses CF_4. The hadron separation produced by RICH1 as a function of the track momentum is shown in Fig.14. This particle identification is a cornerstone of LHCb's ability to make precision measurements of b and c hadron decays.

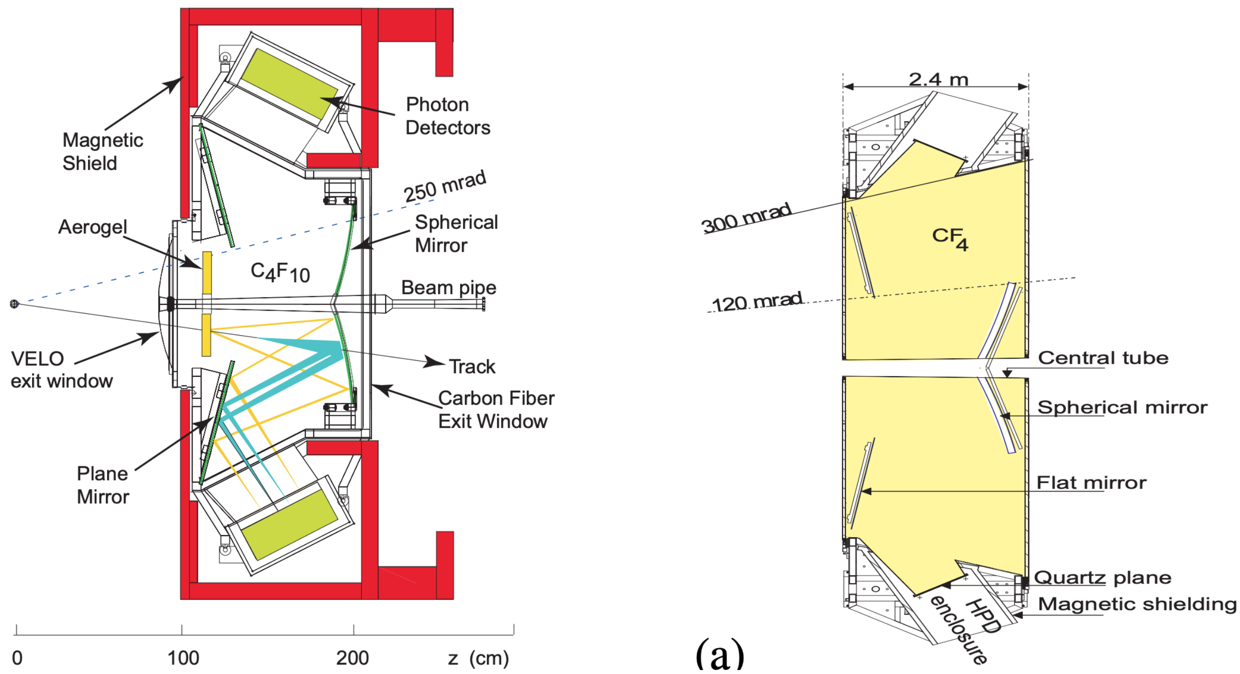


Figure 13: (Left) A side-view schematic of the RICH1 detector. (Right) A top-view schematic of the RICH2 detector [5].

3.2.7 Calorimeters

The calorimeter system is responsible for the identification and measurement of the energies and position of hadron, photon and electron candidates. The principle of operation of the calorimeter is based on stopping the particles as they pass through the material of the calorimeter, and measuring the energy they release. This information plays a vital role in the first stage of the LHCb trigger, which has run in real time during Run 1 and Run 2, in helping it decide which events are worth keeping. The calorimeter is located downstream of the magnet, between the first and second muon stations. It consists of four sub-detectors: a

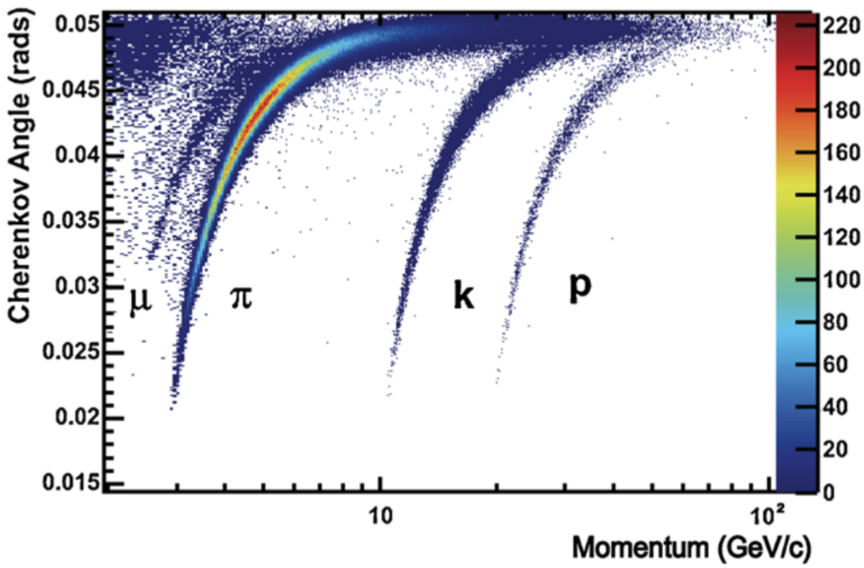


Figure 14: The RICH1 Cherenkov angle as a function of track momentum for different species of tracks is shown, demonstrating the particle identification capabilities of the detector. [7].

scintillating pad detector (SPD), a pre-shower detector (PS), an electromagnetic calorimeter (ECAL) and a hadron calorimeter (HCAL), in order from upstream to downstream.

The pre-shower detector provides information on the electromagnetic character of the particles (i.e. whether it is a photon ot an electron), while the SPD indicates whether the particles are charged or neutral. As their names suggest, the electromagnetic calorimeter is responsible for measuring the energy of photon and electron candidates, and the hadron calorimeter measures the energy of hadronic particles. The particles slowing down in the different layers interact with the detector material (which is made of a scintillating material), releasing light, which is picked up by wavelength shifting fibers and transmitted to photomultiplier sensors. All the layers of the calorimeter are segmented into cells along the x and y directions, with finer segmentation in the regions closer to the beam pipe to handle the higher particle occupancies there. This is shown in Fig. 15.

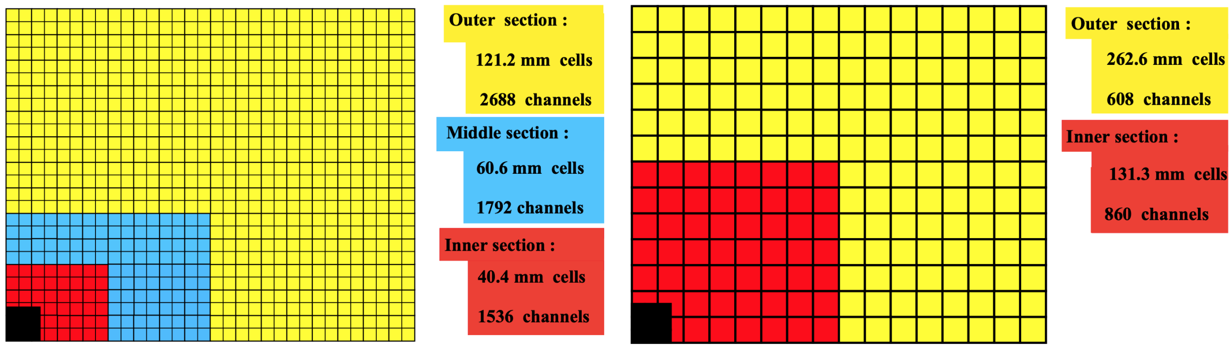


Figure 15: (Left) The segmentation of a quarter plane of SPD/PS and ECAL is shown, with the numbers providing the cell dimensions of the ECAL. (Right)The segmentation of a quarter plane of HCAL is shown, with the numbers for its cell sizes.

3.2.8 Muon System

The efficient detection and reconstruction of muons is key to the LHCb experiment, given the number of important final states in which they appear. LHCb's muon detectors appear in the form of five stations (M1 - M5) at the most downstream end of the detector, so positioned because of the penetrating power of muons. The five stations are rectangular and cover a combined area of 453m^2, and are composed of 1380 chambers in total, with the chambers being filled with a cocktail of three gases: carbon dioxide, argon and tetrafluoromethane. Muons passing through the gas ionize it, and the resulting eletrons are picked up by wire electrodes. The first muon station is placed upstream of the calorimeters to provide a more accurate p_{T} measurement of the tracks to the trigger.The four remaining stations are downstream of the calorimeter, and have 80 cm thick iron absorbers placed between them, to absorb all particles that are not muons.

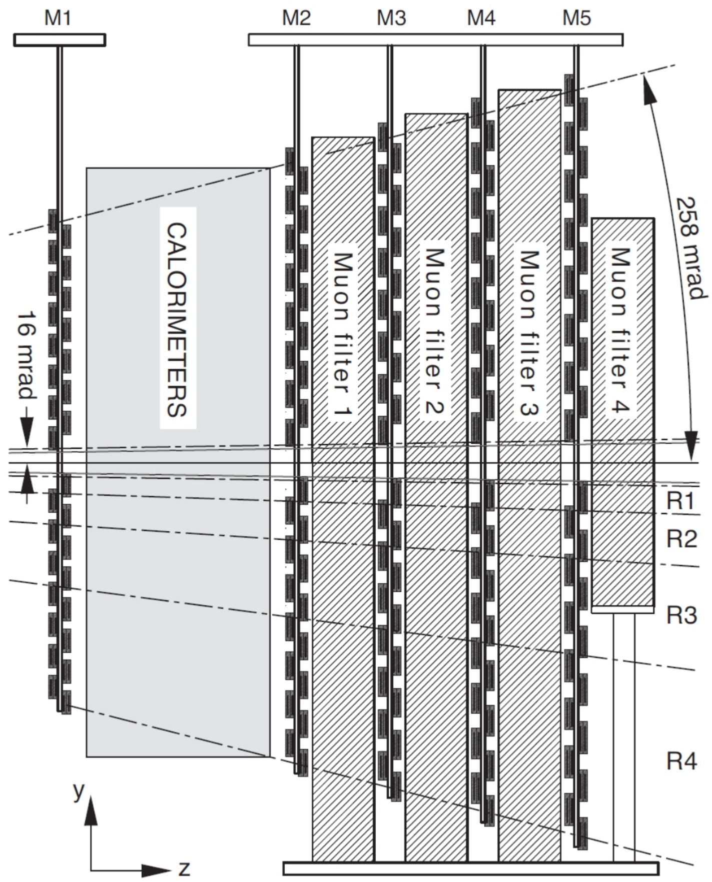


Figure 16: A side view schematic of the muon detector.

3.2.9 Trigger

The LHC provides *pp* collisions to the LHCb detector at a rate of 40 MHz. Of these collisions, the ones that produce at least two charged tracks that are visible to the tracking system (called visible interactions) occur at a rate of 10 MHz. Due to disk storage constraints, it is not possible to write out the information resulting from all of these collisions; nor is it desirable, since many of them will be uninteresting for physics analyses. The trigger system is responsible for reducing the rate from 10 MHz to $2-5$ kHz, by attempting to select events with interesting signatures to be written to disk. This reduction of rate occurs in stages. The first stage is the L0 trigger, which runs synchronously with the 40 MHz collisions, and reduces the data rate to 1 MHz. This is followed by the High Level Trigger (HLT), which runs asynchronously on a processor farm, and brings the data rate down to the requisite $2-5$ kHz.

3.2.9.1 L0 Trigger

The L0 trigger utilizes information from the calorimeter and muon subsystems of the LHCb detector. The energies deposited in the SPD, PS, ECAL and HCAL detectors are used by the L0 calorimeter system to trigger the selection of events. These detectors are segmented into cells transverse to the beam axis. The trigger decision is based on the transverse energy deposited in clusters of 2×2 cells in the ECAL and HCAL detectors, with the PS and SPD detectors providing information about the electromagnetic character of the candidate. The transverse energy deposited in a cluster is defined as

$$E_T = \sum_{i=1}^{4} E_i \sin\theta_i \tag{7}$$

where E_i is the energy deposited in cell i and θ_i is the angle between the z-axis and a neutral particle assumed to be coming from the mean position of the interaction envelope hitting the center of the cell. From these calorimeter clusters, three types of candidates are built:

- Hadron Candidate (`L0Hadron`): this is chosen as the HCAL cluster with the highest E_T. In the event of there being a "highest-E_T" ECAL cluster located in front of the aforementioned HCAL cluster, that E_T of the hadron candidate is taken as the sum of the E_T from both ECAL and HCAL.

- Photon Candidate (`L0Photon`): this is chosen as the ECAL cluster with the highest E_T such that 1 or 2 PS cells have a hit in front of this ECAL cluster, and that there are no SPD hits in the cells corresponding to the PS cells. In the area of the ECAL closer to the beam pipe, a highest E_T ECAL cluster with 3-4 PS cells hit in front of it is also accepted as the photon candidate. The E_T of the photon candidate is taken as the E_T

deposited in the ECAL alone.

- Electron Candidate (`L0Electron`): this is chosen based on the same requirements as the photon candidate, with the additional requirement that at least one SPD cell is hit in front of the PS cells.

The E_T of the candidates is compared to a threshold (see 1), and events containing at least one candidate above threshold are retained by the L0 trigger. The L0-Muon trigger utilizes information from the 5 muons stations (M1-M5), each of which are divided into four quadrants, with each quadrant having its own L0 processor. Each processor identifies the two muon tracks with the highest transverse momentum in the quadrant. The `L0Muon` trigger retains candidates based on comparing the largest p_T of the eight candidates with a threshold (see 1). A requirement is also placed on the maximal number of SPD hits in order to reduce the complexity of events and hence to enable a faster reconstruction in the subsequent software trigger (HLT).

Table 1: L0 trigger thresholds. E_T thresholds apply to Hadron, Photon and Electron decisions. The p_{T} threshold applies for the Muon decision.

L0 trigger	E_T/p_T threshold			SPD threshold
	2015	2016	2017	
Hadron	$> 3.6\,\mathrm{GeV}$	$> 3.7\,\mathrm{GeV}$	$> 3.46\,\mathrm{GeV}$	< 450
Photon	$> 2.7\,\mathrm{GeV}$	$> 2.78\,\mathrm{GeV}$	$> 2.47\,\mathrm{GeV}$	< 450
Electron	$> 2.7\,\mathrm{GeV}$	$> 2.4\,\mathrm{GeV}$	$> 2.11\,\mathrm{GeV}$	< 450
Muon	$> 2.8\,\mathrm{GeV}$	$> 1.8\,\mathrm{GeV}$	$> 1.35\,\mathrm{GeV}$	< 450

3.2.9.2 HLT trigger

The High Level Trigger is software based and runs on CPUs on the Event Filter Farm (EFF). It is split into two stages: HLT1 and HLT2, for reasons of timing. In the first stage, a partial reconstruction of the event is done (using tracks with $p_{\mathrm{T}} > 1\,\mathrm{GeV}$), allowing for inclusive event selections based on one or two track signatures. The concept of a "trigger line" comes into play in the HLT. A trigger line is composed of a sequence of reconstruction algorithms and selections. The HLT1 lines are general purpose triggers, designed to select tracks which are displaced from the PV, a signature of b decays. The HLT1 reduces the data rate to 30 kHz.

At this rate, the HLT2 is able to perform a more thorough reconstruction (using all tracks with momentum $p_{\mathrm{T}} > 500\,\mathrm{MeV}$). A large part of the HLT2's output comes from inclusive topological trigger lines, which are constructed to select b-hadron decays based on the presence of a displaced vertex associated with at least two charged tracks [44]. These lines are inclusive in the sense that they only require a part of the B decay to be reconstructed, allowing them to have a high efficiency across a range of b decay types. While in Run 1 of the LHC, a simple reconstruction was performed in the HLT2 stage, followed by a complete event reconstruction offline, in Run 2, the reconstruction performed in the HLT2 was vastly improved to make it identical to that which is performed offline.

3.3 Data Flow

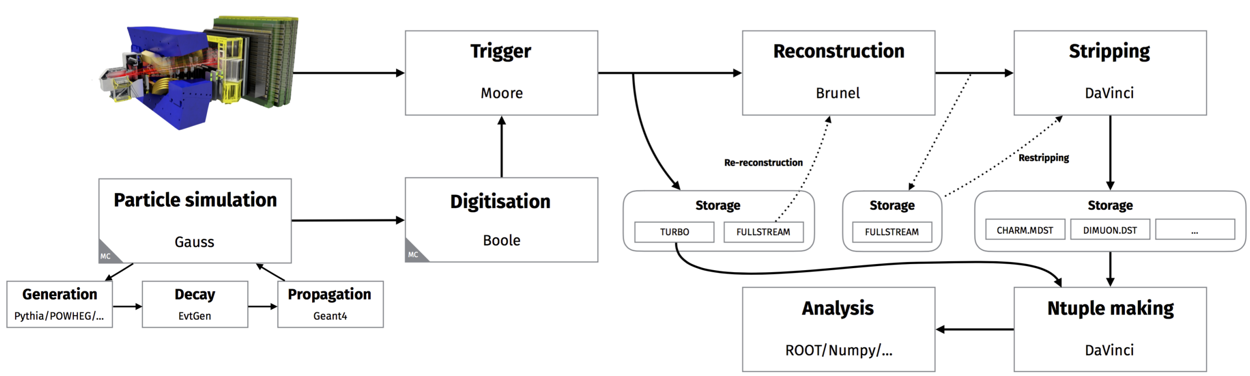


Figure 17: The LHCb data flow in Run 2.

Before the data recorded by the detector can be analyzed by physicists, it must undergo a number of processing stages to form the physics objects of interest, as well as to make the size of the data manageable. From a computational point of view, it would be extremely wasteful to have every analysis process all the data written by the HLT trigger to disk. The HLT is run by the `MOORE` software package [45], which is built around Python packages that configure algorithms, tools, data flow, and control flow in order to run a `Gaudi`-based application. `Gaudi` [46] is a software framework used for building High Energy Physics data processing applications.

The data at this stage (in Run 1) exists in the form of the responses of the various LHCb sub-detectors. The offline reconstruction, used to convert the detector responses (for example, track hits) into physics objects (such as tracks, vertices and calorimeter clusters), is performed by the `BRUNEL` [47] software package. The data after reconstruction is still too voluminous to analyze practically, so a Stripping stage is conducted, where selections are applied, and the data is separated into different physics streams (such as charm events, events with $J/\psi \to \mu\mu$, semileptonic events). The `DaVinci` [47] software application is used to handle the Stripping campaigns, which are conducted centrally by the collaboration.

It is after the Stripping stage that the data becomes available to the users of the collaboration, who process it with `DaVinci` to reconstruct the decays of their interest and make n-tuples.

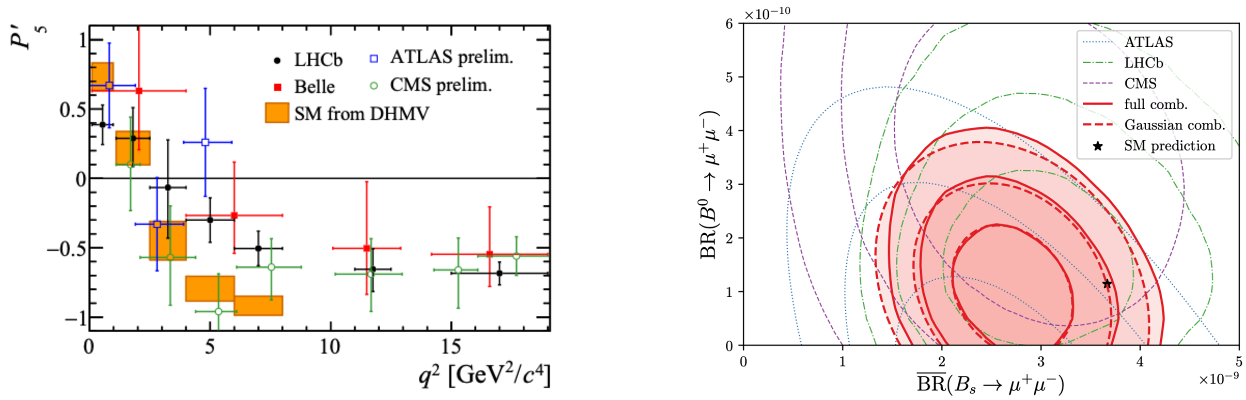
4 Measurement of $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-)$

4.1 Introduction

In recent years a pattern of deviations from the Standard Model have been observed in experimental measurements concerning $b \to sll$ processes and lepton flavour universality ratios. As it stands today, none of these deviations have crossed the traditional "5 σ" threshold in particle physics. But the fact that these deviations are seen in the same direction in similar but independent processes, and by more than one experiment in some cases, stands as one of the biggest signposts to the direction of New Physics. In this subsection, we first detail the experimentally observed deviations that we are interested in. Next, we discuss the possible explanations provided by theorists for these deviations. Finally, we provide the connection between these deviations and the process of interest: $B^0 \to K^{*0}\tau^+\tau^-$.

First, the $b \to s\mu^+\mu^-$ current has shown deviations, measured by multiple experiments, in branching fractions and angular observables. It can be seen in Fig.18a that LHCb [8], BELLE [9] and ATLAS [10] see deviations from the SM in the CP averaged angular observable P_5' in the decay $B^0 \to K^{*0}\mu^+\mu^-$. Further, it can be seen from Fig.18b that the SM prediction for the branching fraction of $B_s^0 \to \mu^+\mu^-$ stands at roughly 2σ tension with respect to the combination of measurements from ATLAS [14], LHCb [13] and CMS [14] (Note that both of these decays are mediated by $b \to s\mu^+\mu^-$).

Secondly, the LHCb collaboration sees deviations from lepton flavour universality in the measurements of the parameters R_K [16] and R_{K^*} [48] (defined below), which compare $b \to se^+e^-$ with $b \to s\mu^+\mu^-$. The measurement of R_K is at a 3.1σ tension with the SM expectation of $R_K = 1.00 \pm 0.01$ [43], while the measurement of R_{K^*} is at $2.1 - 2.3\sigma$ and $2.4 - 2.5\sigma$ tension with the SM expectation in the two q^2 bins defined below. It is also interesting to note that the measurements of R_K and R_{K^*} both deviate from the SM expectation in the downward direction, indicating that muons are being produced less often



(a) (b)

Figure 18: (Left) A combination of the LHCb [8], BELLE [9], ATLAS [10] and CMS [11] measurements of the CP averaged angular variable P_5' in $B^0 \to K^{*0}\mu^+\mu^-$, shown in bins of the dilepton invariant mass squared q^2, overlaid with the theoretical prediction from the SM [12]. (Right)A combination of the LHCb [13] , ATLAS [14] and CMS [14] simultaneous measurements of $\mathcal{B}(B^0 \to \mu^+\mu^-)$ and $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ shown as two dimensional likelihood contours, compared to the SM prediction [15]

relative to electrons in $B^0 \to K^{(*)}ll$ decays.

$$R_{K^{(*)}} = \frac{\mathcal{B}(B^0 \to K^{(*)}\mu^+\mu^-)}{\mathcal{B}(B^0 \to K^{(*)}e^+e^-)} \tag{8}$$

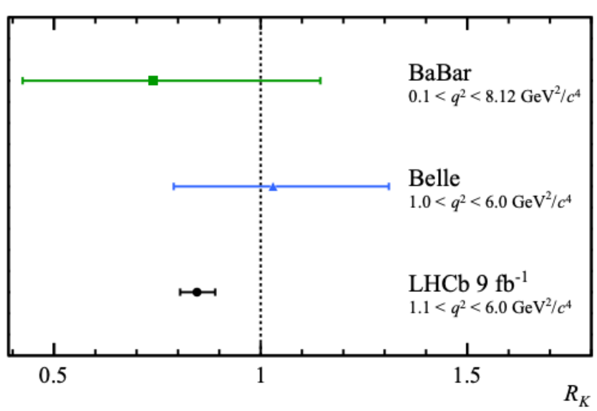


Figure 19: A comparison between R_K measurements by LHCb [16], BaBar [17] and Belle [18] is shown. The SM expectation of $R_K = 1.00 \pm 0.01$ is shown by the dotted line.

Finally, signs of lepton flavour universality violation have also been observed in charged current $b \to cl\nu_l$ transitions, through the variables R_D and R_{D^*} (defined below), which

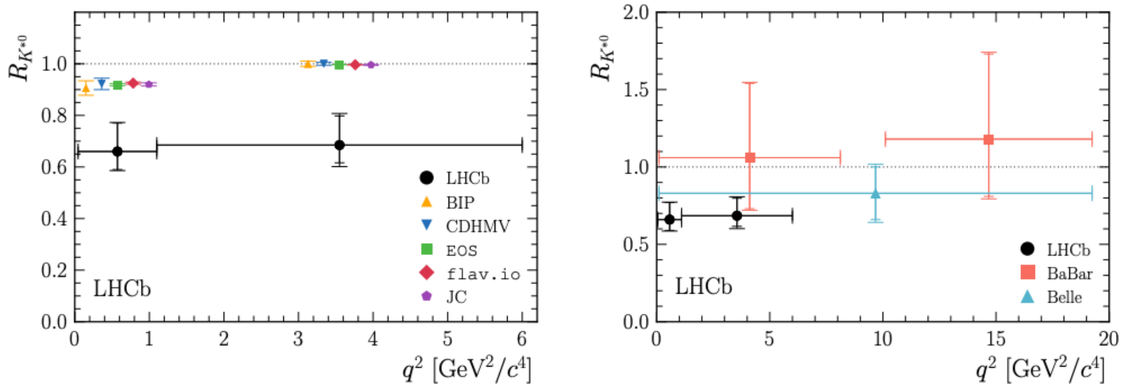


Figure 20: (Left) A comparison of the LHCb R_{K^*} measurement [19] with the SM theoretical predictions: BIP [20], CDHMV [21], EOS [22], flav.io [23] and JC [24], which have been displaced horizontally for presentation. (Right) Comparison of the LHCb R_{K^*} measurements with previous experimental results from BaBar [17] and Belle [25]. In the case of the B factories the specific vetoes for charmonium resonances are not represented.

compare $B \to D^{(*)}\tau\nu_\tau$ with $B \to D^{(*)}l\nu$ (where l is either e or μ). Unlike the FCNC $b \to sll$ transitions which are suppressed in the SM (occurring at rates O(10^{-7})), the charged current $b \to cl\nu_l$ are mediated at the tree level in the SM with branching fractions of $O(1\%)$. As a result, a rather large NP contribution would be required in order to compete with the SM processes. The combination of the experimental measurements, shown in Fig. 21, are in $\sim 3\sigma$ tension with the SM predictions.

$$R_{D^{(*)}} = \frac{\mathcal{B}(B \to D^{(*)}\tau\nu)}{\mathcal{B}(B \to D^{(*)}l\nu)}, l = e, \mu \tag{9}$$

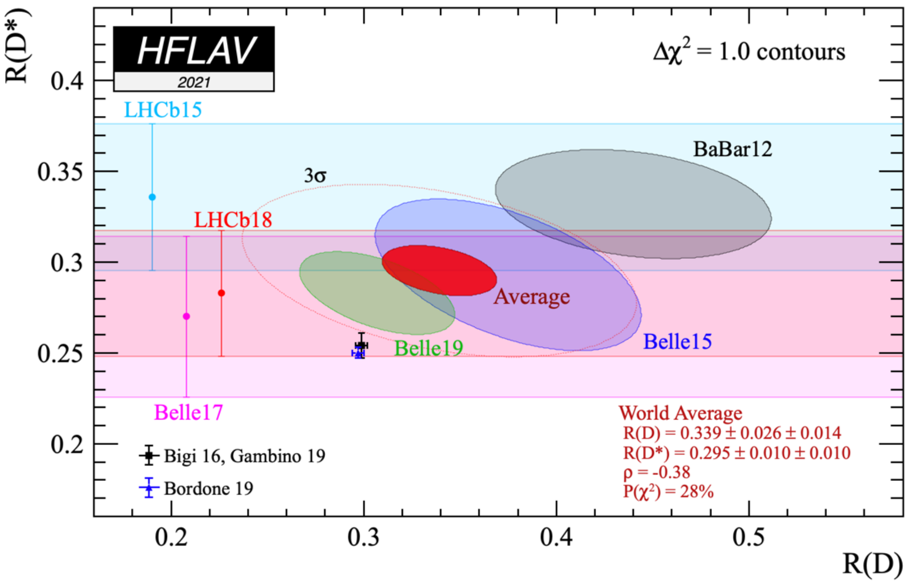


Figure 21: The experimental measurements of R_D and R_{D^*} by BaBar [26], Belle [27] [28] [29], LHCb [30] [31] and their two dimensional average (denoted by the dark red ellipse) are shown (with contours corresponding to 68% confidence level for bands and 39% confidence level for ellipses), compared to the theoretical SM model predictions [32] [33] [34] shown as the black and blue points with error bars.

Table 2: LHCb measurement of the LFU variables, shown along with their SM prediction and the tension between measurement and expectation.

Observable	SM prediction	LHCb measurement	Tension
R_K	1 ± 0.01	$0.846^{+0.042+0.013}_{-0.039-0.012}$ [16]	3.1σ
$R_{K^*}(0.045 < q^2 < 1.1\,\mathrm{GeV}^2/c^4)$	$(0.906 - 0.925) \pm 0.028$	$0.66^{+0.11}_{-0.07} \pm 0.03$ [19]	$2.1 - 2.3\sigma$
$R_{K^*}(1.1 < q^2 < 6.0\,\mathrm{GeV}^2/c^4)$	$(0.996 - 1.000) \pm 0.01$	$0.69^{+0.11}_{-0.07} \pm 0.05$ [19]	$2.4 - 2.5\sigma$
R_{D^*}(muonic tau)	0.258 ± 0.005	$0.336 \pm 0.027 \pm 0.030$ [30]	1.9σ
R_{D^*}(hadronic tau)	0.258 ± 0.005	$0.291 \pm 0.019 \pm 0.029$ [49]	1.0σ
R_{D^*}(combined)	0.258 ± 0.005	$0.310 \pm 0.016 \pm 0.022$	2.2σ

Since LFU violation occurs naturally in many extensions to the SM, this pattern of anomalies has prompted theorists to search for explanations in a model independent manner

([35], [36], [50]), using an effective field theory (EFT) formalism. The neutral current anomalies can be explained by NP contributions to the Wilson coefficients associated with operators describing $b \to s\mu^+\mu^-$ transitions. A global analysis [43] of all the neutral current variables (pertaining either to $b \to s\mu^+\mu^-$ alone or to $R_{K^{(*)}}$) points to NP structures which describe the observed data significantly better than the Standard Model. Importantly, a strong consistency is seen in the pattern of deviations observed in $b \to s\mu^+\mu^-$ and those seen in the LFUV variables. It is understood that the semileptonic Wilson coefficient C_9 plays a central role in all these NP scenarios, requiring a negative contribution of $\sim 25\%$ with respect to the SM value. The analysis found several scenarios with one or two free parameters exhibiting a pull of more than 4σ with respect to the SM. When *allowing the NP to break LFU* by having contributions of different sizes in the muon and electron sectors, the analysis concludes that the data requires NP contributions in the muon sector much more strongly than in the electron sector, and thus generally disfavours lepton flavour universal NP couplings.

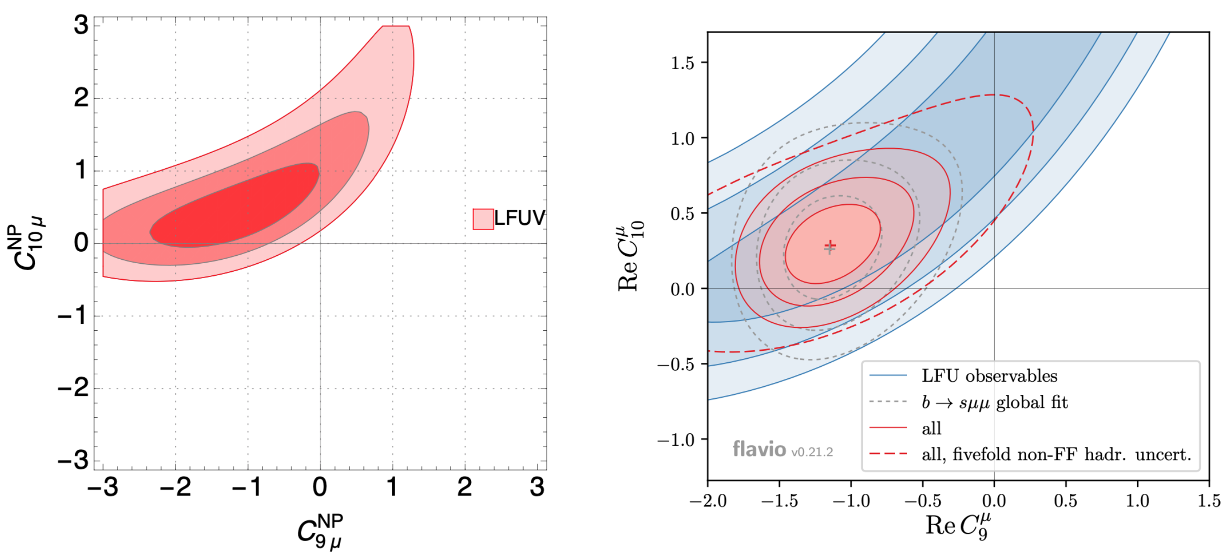


Figure 22: Constraints on the NP contributions to the Wilson coefficients $C_{9\mu}$ and $C_{10\mu}$ for $b \to sl^+l^-$ using only LFU variables (left) and using all $b \to sl^+l^-$ data (right).

Capdevila et.al. [35] indicate that a solution of the $R_{D^{(*)}}$ anomaly requires a sizeable NP contribution of O(20%) to the branching ratio of $B \to D^{(*)}\tau\overline{\nu}_\tau$. External constraints from measurements of the B_c lifetime and the q^2 distribution of $R_{D^{(*)}}$ point the NP contribution to the SM operator $[\overline{c}\gamma^\mu P_L b][\overline{\tau}\gamma_\mu P_L \nu_\tau]$ in such a manner that there is interference with the SM. If the scale of NP contributing to $R_{D^{(*)}}$ is much higher than the electroweak symmetry breaking scale, the semileptonic decays involving only left handed quarks and leptons are described by the two operators:

$$\begin{aligned} O^{(1)}_{ijkl} &= [\overline{Q}_i\gamma_\mu Q_j][\overline{L}_k\gamma^\mu L_l] \\ O^{(3)}_{ijkl} &= [\overline{Q}_i\gamma_\mu \sigma^I Q_j][\overline{L}_k\gamma^\mu \sigma^I L_l] \end{aligned}$$

with the Pauli matrices σ^I acting on the weak isospin components of the quark (Q) and lepton (L) doublets. These operators affect semileptonic $b \to c(s)$ decays involving charged tau leptons and tau neutrinos after electroweak symmetry breaking. The SU(2) components of the above operators pertaining to the third generation leptons can be written as

$$\begin{aligned} C^{(1)}O^{(1)} &\to C^{(1)}_{23}([\overline{s}_L\gamma_\mu b_L][\overline{\tau}_L\gamma_\mu \tau_L] + [\overline{s}_L\gamma_\mu b_L][\overline{\nu}_\tau\gamma_\mu \nu_\tau]), \\ C^{(3)}O^{(3)} &\to C^{(3)}_{23}(2V_{cs}[\overline{c}_L\gamma_\mu b_L][\overline{\tau}_L\gamma_\mu \nu_\tau] + [\overline{s}_L\gamma_\mu b_L][\overline{\tau}_L\gamma_\mu \tau_L] \\ &\qquad - [\overline{s}_L\gamma_\mu b_L][\overline{\nu}_\tau\gamma_\mu \nu_\tau]) + C^{(3)}_{33}(2V_{cb}[\overline{c}_L\gamma_\mu b_L][\overline{\tau}_L\gamma_\mu \nu_\tau]). \end{aligned}$$

with $C^{(n)}_{ij}$ denoting the Wilson coefficient for the operator O^n_{ij33}. We see from the above the there are two ways of enhancing $b \to c\tau^-\overline{\nu}_\tau$. One way involves an NP contribution to $C^{(3)}_{33}$, in scenarios with NP being aligned to the third generation, which would avoid affecting the down quark FCNCs. But the smallness of the V_{cb} CKM element forces the contribution to $C^{(3)}_{33}$ to be rather large, conflicting with constraints from direct LHC searches for $\tau^+\tau^-$ final states [51] and electroweak precision data [52].

Therefore the solution to the $R_{D^{(*)}}$ has to be through $C^{(1,3)}_{23}$, which generates large contributions to $b \to s\tau^+\tau^-$ and/or $b \to s\nu_\tau\overline{\nu}_\tau$. In order to account for the severe constraints on NP in $B \to K^{(*)}\nu\overline{\nu}$ [53], the contribution from $C^{(3)}_{23}$ needs to be approximately cancelled by that from $C_{23}^{(1)}$, implying $C_{23}^{(1)} \approx C_{23}^{(3)}$. Capdevila et.al. [35] use this assumption to correlate $b \to c\tau^-\overline{\nu}_\tau$ and $b \to s\tau^+\tau^-$, and find the following formulation of NP contributions to the Wilson coefficients $C_{9(10)}^{\tau\tau}$

$$C_9^{\tau\tau} = C_9^{\mathrm{SM}} - \Delta \tag{10}$$

$$C_{10}^{\tau\tau} = C_{10}^{\mathrm{SM}} + \Delta \tag{11}$$

$$\Delta = \frac{2\pi}{\alpha}\frac{V_{cb}}{V_{tb}V_{ts}^*}\left(\sqrt{\frac{R_{D^{(*)}}}{R_{D^{(*)}}^{\mathrm{SM}}}} - 1\right) \tag{12}$$

It is worth noting that the factor multiplying the brackets in Eq.12 is very large (around 860). The authors go on to provide predictions for the branching fraction of $B^0 \to K^{*0}\tau^+\tau^-$ (neglecting the SM contribution since it is completely overwhelmed by the NP effects), taking these enhancements into account:

$$\begin{aligned} \mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-)^{\mathrm{NP}} &= (10.1 \pm 0.8) \times 10^{-9} \times \Delta^2 \\ &\approx 0.008\left(\sqrt{\frac{R_{D^{(*)}}}{R_{D^{(*)}}^{\mathrm{SM}}}} - 1\right)^2 \end{aligned} \tag{13}$$

The predictions for various $b \to s\tau^+\tau^-$ processes are visualised in Fig.23 as a function of the enhancement in the charged current processes.

The EFT analysis provided above is model independent, i.e. it does not depend on a specific type of particle beyond the Standard Model causing these enhancements. But generally, models for such particles fall into one of two camps: colour singlets such as charged Higgses or W' bosons, and leptoquarks. The constraints of colour singlet models from measurement of the B_c lifetime and direct searches are quite severe. In this regard, leptoquark models are less constrained, making them the favoured explanation for such anomalies. Regardless of the exact NP model, all attempts to find a minimalistic explanation for the LFU anomalies agree on one thing: sizeable enhancements to the rate of $b \to s\tau\tau$ processes, up to a factor of $O(10^3)$ over the SM. Alonso et. al. [36] correlate the $R_{K^{(*)}}$ anomaly with the enhancement in $b \to s\tau^+\tau^-$, as shown in Fig.24.

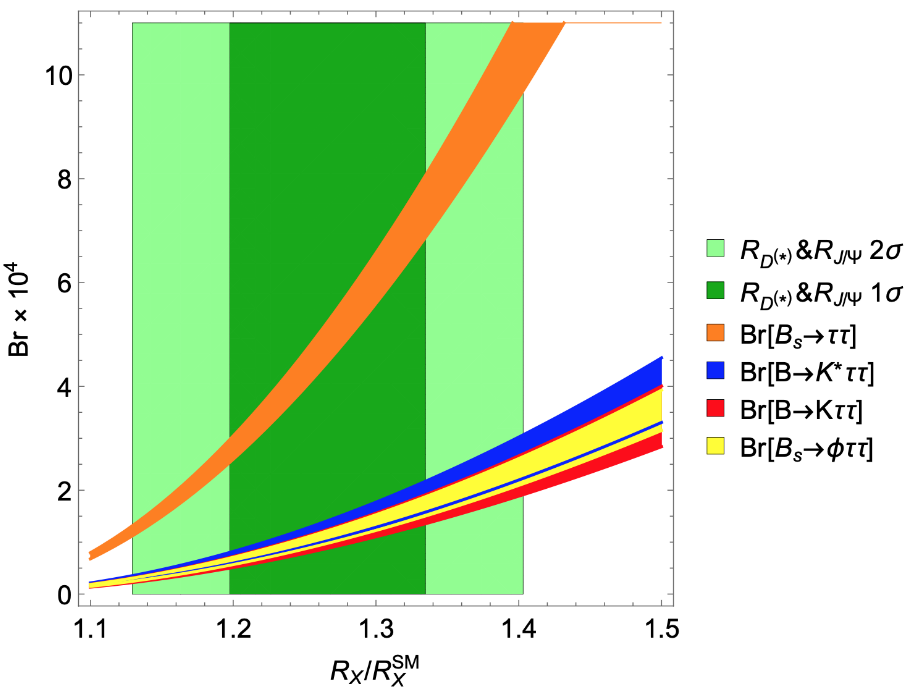


Figure 23: Correlation between the predicted branching fractions for various $b \to s\tau^+\tau^-$ processes as a function of the enhancements in the charged current anomalies [35]. X is used as a placeholder for D or D^*. The dark green and light green bands depict the measured 1σ and 2σ intervals for R_X/R_X^{SM}.

In the SM, $\mathcal{B}(B^0 \to K^*\tau^+\tau^-) = (0.98 \pm 0.10) \times 10^{-7}(15 < q^2 < 19\,\mathrm{GeV}^2/c^4)$ [35]. Past experimental constraints [54] [55] on $b \to s\tau\tau$ are

$$\mathcal{B}(B_s^0 \to \tau^+\tau^-) < 6.8 \times 10^{-3} \text{ at } 95\% \text{ C.L. (LHCb)}$$

$$\mathcal{B}(B^+ \to K^+\tau^+\tau^-) < 2.25 \times 10^{-3} \text{ at } 95\% \text{ C.L. (BaBar)}$$

The Belle collaboration conducted a recent search for $B^0 \to K^{*0}\tau^+\tau^-$ [56]. They did not find evidence for signal, resulting in the constraint $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) < 2 \times 10^{-3}$ at 90% C.L.

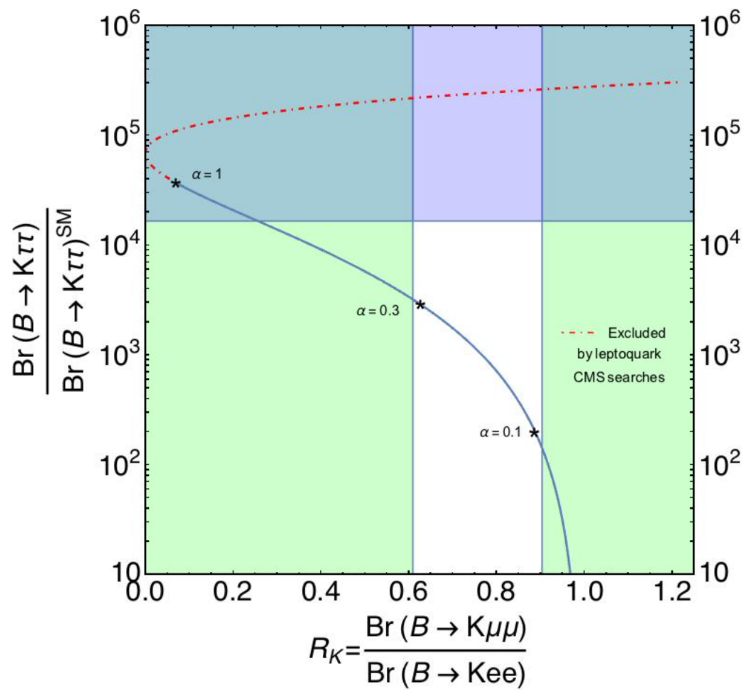


Figure 24: Prediction for the correlation between the enhancement in $B^+ \to K^+\tau^+\tau^-$ and the R_K anomaly, shown as the blue lineshape. The vertical white band is the experimentally allowed region in R_K [36].

The decay we consider in this analysis is $B^0 \to K^* \tau^+ \tau^-, K^{*0} \to K^+ \pi^-, \tau^+ \to \pi^+ \pi^- \pi^+ (\pi^0) \nu_\tau$. Some pertinent branching fractions of the intermediate decays are:

$$\mathcal{B}(K^{*0} \to K^+ \pi^-) = 66\%$$
$$\mathcal{B}(\tau^+ \to \pi^+ \pi^- \pi^+ \nu_\tau) = (9.02 \pm 0.05)\%$$
$$\mathcal{B}(\tau^+ \to \pi^+ \pi^- \pi^+ \pi^0 \nu_\tau) = (4.49 \pm 0.05)\%$$

Finally, it is worth noting that $B^0 \to (\psi(2S) \to \tau^+ \tau^-)(K^{*0} \to K^+ \pi^-)$ is not quite capable of faking the signal we are searching for. The branching fractions of the relevant decays have been measured [57]:

$$\mathcal{B}(B^0 \to \psi(2S) K^{*0}) = (5.9 \pm 0.4) \times 10^{-4}$$
$$\mathcal{B}(\psi(2S) \to \tau^+ \tau^-) = (3.1 \pm 0.4) \times 10^{-3}$$
$$\mathcal{B}(B^0 \to (\psi(2S) \to \tau^+ \tau^-)(K^{*0} \to K^+ \pi^-)) = (1.2 \pm 0.2) \times 10^{-6}$$

Since this branching fraction is so small, if the efficiency of reconstructing and selecting these decays is anything like the efficiency for our signal mode (since it is so similar to our signal, the efficiencies cannot be drastically different), the contribution of this mode to our data sample will be essentially zero. Therefore we will not consider it any further.

4.2 Analysis Strategy

The parameter of interest in this analysis the branching fraction of the decay $B^0 \to K^{*0}\tau^+\tau^-$. In order to avoid bias, the most signal rich region of the data is blinded in the mass distribution. Before we are allowed to unblind, we must fix all our selections, document the analysis thus far and the procedure post unblinding, and go through an internal review and approval process within the LHCb collaboration. For the purpose of this thesis, we will not unblind. Instead, we will quote a sensitivity to $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-)$ based on toy studies.

We use the $B^0 \to K^{*0}\tau^+\tau^-$ decay, mediated by the $b \to s\tau^+\tau^-$ current, instead of $B^+ \to K^+\tau^+\tau^-$ because the position of the $K^{*0} \to K^+\pi^-$ decay vertex also gives us the B^0 decay vertex. Since the τ's are not fully reconstructed, this would not be possible with $B^+ \to K^+\tau^+\tau^-$. Following a similar thought process, we reconstruct the τ in its hadronic decay to $\pi^+\pi^-\pi^+(\pi^0)\nu_\tau$ because the three pion charged tracks allow us to reconstruct the τ decay vertex. This would not be possible with the purely leptonic decay $\tau^+ \to \mu^+\nu_\tau\nu_\mu$. The topology of the signal decay is shown in Fig.25.

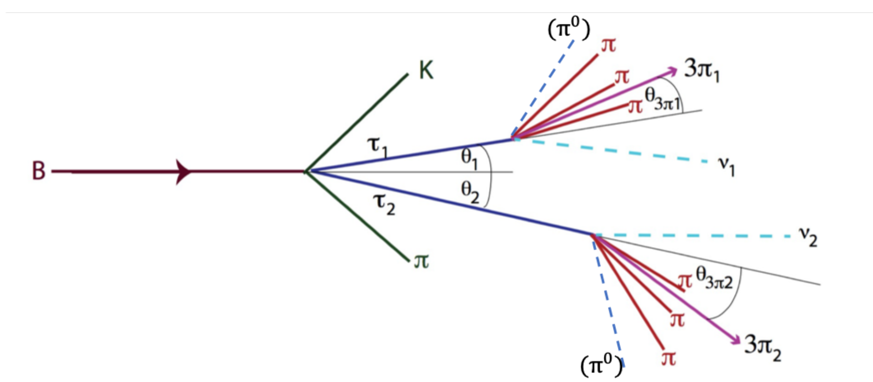


Figure 25: An illustration of the topology of the signal decay. The tracks shown by dashed lines are not reconstructed. The flight distances of the intermediate τ's have been exaggerated for illustration.

The branching fraction of $B^0 \to K^{*0}\tau^+\tau^-$ can be expressed by definition as

$$\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) = \frac{N_{\text{corr}}(B^0 \to K^{*0}\tau^+\tau^-)}{N(B^0)}. \tag{14}$$

The numerator of the above fraction will be the efficiency corrected yield of the signal that we measure in data. The denominator is the initial number of B^0's that were present in the data sample. While this is calculable in principle from measured b-quark production cross section ($\sigma_{b\bar{b}}$), luminosity ($\mathcal{L}$), and the B^0 fragmentation fraction (f_d), using a formula such as

$$N_{B^0} = 2 \times \mathcal{L} \times \sigma_{b\bar{b}} \times f_d \tag{15}$$

but the numbers used in the right hand side have substantial uncertainties that will feed into the final result if we use this method of calculation. In order to get around this, we use the $B^0 \to (D^- \to K^+\pi^-\pi^-)(D^0 \to K^-\pi^+\pi^+\pi^-)K^{*0}$ decay for normalization purposes. Since both decays share the same parent hadron, their corresponding denominators from 14 will cancel out.

$$\frac{N_{\text{corr}}(B^0 \to K^{*0}\tau^+\tau^-)}{N_{\text{corr}}(B^0 \to D^-D^0K^+)} = \frac{\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-)}{\mathcal{B}(B^0 \to D^-D^0K^+)} \tag{16}$$

This equation can be expanded, including the decays of the intermediate resonances, and replacing N_{corr} by $N \times \epsilon$:

$$\frac{N(B^0 \to K^{*0}\tau^+\tau^-, K^{*0} \to K^+\pi^-, \tau^+ \to \pi^+\pi^-\pi^+(\pi^0)\nu_\tau)}{N(B^0 \to D^-D^0K^+, D^- \to K^+\pi^-\pi^-, D^0 \to K^-\pi^+\pi^+\pi^-)} =$$
$$\frac{\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) \cdot \mathcal{B}(K^{*0} \to K^+\pi^-) \cdot [\mathcal{B}(\tau^+ \to \pi^+\pi^-\pi^+\nu_\tau) + \mathcal{B}(\tau^+ \to \pi^+\pi^-\pi^+\pi^0\nu_\tau)]^2 \cdot \epsilon_{\text{sig}}}{\mathcal{B}(B^0 \to D^-D^0K^+) \cdot \mathcal{B}(D^- \to K^+\pi^-\pi^-) \cdot \mathcal{B}(D^0 \to K^-\pi^+\pi^+\pi^-) \cdot \epsilon_{\text{norm}}} \tag{17}$$

ϵ_{sig} and ϵ_{norm} refer to the total efficiencies for the signal mode and normalization mode respectively. They are determined from simulation samples of these decays which are described below. One of the benefits of using a normalization mode that has the same number of charged hadron tracks as the signal mode is that significant portion of the systematic uncertainties pertaining to the efficiencies cancel out in the ratio.

We can rearrange Equation 17 to have our parameter of interest (POI), $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-)$,

Parameter	Value
$\mathcal{B}(K^{*0} \to K^+\pi^-)$	66%
$\mathcal{B}(\tau^+ \to \pi^+\pi^-\pi^+\nu_\tau)$	$(9.02 \pm 0.05)\%$
$\mathcal{B}(\tau^+ \to \pi^+\pi^-\pi^+\pi^0\nu_\tau)$	$(4.49 \pm 0.05)\%$
$\mathcal{B}(B^0 \to D^-D^0K^+)$	$(0.107 \pm 0.011)\%$
$\mathcal{B}(D^- \to K^+\pi^-\pi^-)$	$(9.38 \pm 0.16)\%$
$\mathcal{B}(D^0 \to K^-\pi^+\pi^+\pi^-)$	$(8.22 \pm 0.14)\%$

Table 3: External branching fraction inputs to the measurement of $\mathcal{B}(B^0 \to K^*\tau^+\tau^-)$.

on the left hand side

$$\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) =$$
$$\frac{N(B^0 \to K^{*0}\tau^+\tau^-)}{N(B^0 \to D^-D^0K^+)} \cdot \left[\frac{\mathcal{B}(B^0 \to D^-D^0K^+) \cdot \mathcal{B}(D^- \to K^+\pi^-\pi^-) \cdot \mathcal{B}(D^0 \to K^-\pi^+\pi^+\pi^-)}{\mathcal{B}(K^{*0} \to K^+\pi^-) \cdot [\mathcal{B}(\tau^+ \to \pi^+\pi^-\pi^+\nu_\tau) + \mathcal{B}(\tau^+ \to \pi^+\pi^-\pi^+\pi^0\nu_\tau)]^2}\right] \cdot \left[\frac{\epsilon_{\text{norm}}}{\epsilon_{\text{sig}}}\right]$$

The branching fractions which go into the right hand side of the above equation are obtained from the PDG [58], and summarized in Table 3. This analysis also relies on inputs from a parallel analysis which measures the branching fractions of modes of the form $B \to DDK^{*0}$, which form dangerous physical backgrounds in the $K^*\tau^+\tau^-$ data sample, as well as the normalization mode $B^0 \to D^-D^0K^+$. The measurement of the $B \to DDK^{*0}$ is summarised in Sec.4.8, and that of the normalization mode in Sec. 4.7.

4.2.1 Modified `DecayTreeFitter` fit

Since the two neutrinos (and sometimes also neutral pions) in the final state of our signal decay are not reconstructed, and the τ lepton decays quickly (with a mean lifetime of 0.29 ps), the invariant mass distribution of the visible part of the decay attains a wide shape and loses discriminating power. In order to recover some of this lost discrimination, we exploit the kinematic information at hand to over-constrain the decay and create a new mass variable, using a modified version of the `DecayTreeFitter` (DTF) [59] tool that is commonly used in LHCb.

DTF performs a global least squares fit of a decay chain involving multiple decay vertices, extracting the decay time, position and momentum parameters of all the particles in the chain, along with their uncertainties and correlations. This global fit takes into account internal kinematic constraints such as the conservation of momentum at each vertex. Additionally, users are able to impose external constraints, such as a mass constraint on an intermediate particle (e.g. constraining the fitted invariant mass of a J/ψ reconstructed from $\mu^+\mu^-$ to the known J/ψ mass), or a pointing constraint on the head of a decay chain (e.g. requiring that the B^+ at the head of a $B^+ \to J/\psi K^+$ decay originate from the primary vertex).

At the end of the day, what we want is the invariant mass of the $K^{*0}\tau^+\tau^-$ combination

$$\begin{aligned} m_{K^{*0}\tau^+\tau^-} &= \sqrt{(E_{K^{*0}} + E_{\tau_1} + E_{\tau_2})^2 - (\vec{P}_{K^{*0}} + \vec{P}_{\tau_1} + \vec{P}_{\tau_2})^2} \\ &= \sqrt{\left(\sqrt{P_K^2 + m_K^2} + \sqrt{P_\pi^2 + m_\pi^2} + \sqrt{P_{\tau_1}^2 + m_\tau^2} + \sqrt{P_{\tau_2}^2 + m_\tau^2}\right)^2 - (\vec{P}_{K^{*0}} + \vec{P}_{\tau_1} + \vec{P}_{\tau_2})^2} \end{aligned} \tag{18}$$

The unknowns in the above equation are the 3-momenta of the τ's. If we consider the $B^0 \to K^{*0}(\tau^+ \to \pi^+\pi^-\pi^+\nu_\tau)(\tau^- \to \pi^-\pi^+\pi^-\nu_\tau)$ decay, we can demonstrate that the decay is kinematically overconstrained. Let us temporarily align the Z axis of the coordinate system with the B flight direction. We can write one energy constraint and three momentum constraints:

$$\begin{aligned} E_B &= E_K + E_\pi + E_{\tau_1} + E_{\tau_2} \\ P_B = P_{Bz} &= P_{Kz} + P_{\pi z} + P_{\tau_1 z} + P_{\tau_2 z} \\ 0 &= P_{Kx} + P_{\pi x} + P_{\tau_1 x} + P_{\tau_2 x} \\ 0 &= P_{Ky} + P_{\pi y} + P_{\tau_1 y} + P_{\tau_2 y} \end{aligned}$$

We have 6 unknowns which are the three momentum components of each of the τ's, and 4 equations. Further, with the knowledge of the τ production and decay vertices, we have two additional constraints, in the form of the directions of the τ momenta. Only the magnitudes of the τ momenta are unknown. This already allows for a constrained fit. Further, a mass constraint is imposed on each τ. This supplies two more constraints. This gives us an overconstrained fit problem.

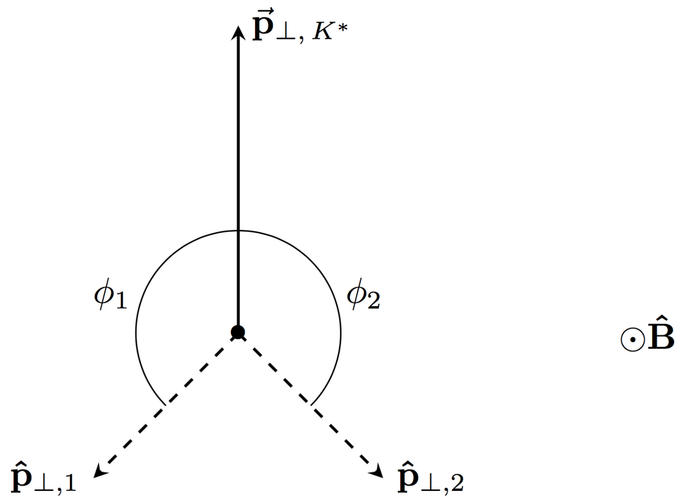


Figure 26: A visualization of the components of the K^{*0}, τ^+, τ^- momenta perpendicular to the B^0 momentum, denoted as $p_{\perp,K^*}$, $p_{\perp,1}$ and $p_{\perp,2}$ respectively.

We set up the kinematic constraints assuming only missing neutrinos without the missing π^0. The components of the K^*, τ^+ and τ^- momenta in the plane perpendicular to the B momentum are shown in Fig. 26, denoted as $p_{\perp,K^*}$, $p_{\perp,1}$ and $p_{\perp,2}$ respectively. We can balance these components along the X and Y directions

$$|\vec{p}_{\perp,1}| \times \sin\phi_1 = |\vec{p}_{\perp,2}| \times \sin\phi_2$$

$$|\vec{p}_{\perp,1}| \times \cos\phi_1 + |\vec{p}_{\perp,2}| \times \cos\phi_2 = -|\vec{p}_{\perp,K^*}|$$

These allow us to solve for $p_{\perp,1}$ and $p_{\perp,2}$

$$|\vec{p}_{\perp,1}| = \frac{-|\vec{p}_{\perp,K^*}|}{\cos\phi_1 + \cos\phi_2 \times \sin\phi_1 / \sin\phi_2}$$
$$|\vec{p}_{\perp,2}| = |\vec{p}_{\perp,1}| \times \frac{\sin\phi_1}{\sin\phi_2}$$

We are able to make these constraints with the knowledge of the B flight direction (which comes from the knowledge of the PV and the K^* vertex), the K^* momentum and the τ flight directions (which comes from the knowledge of the K^* vertex and the 3π vertex).

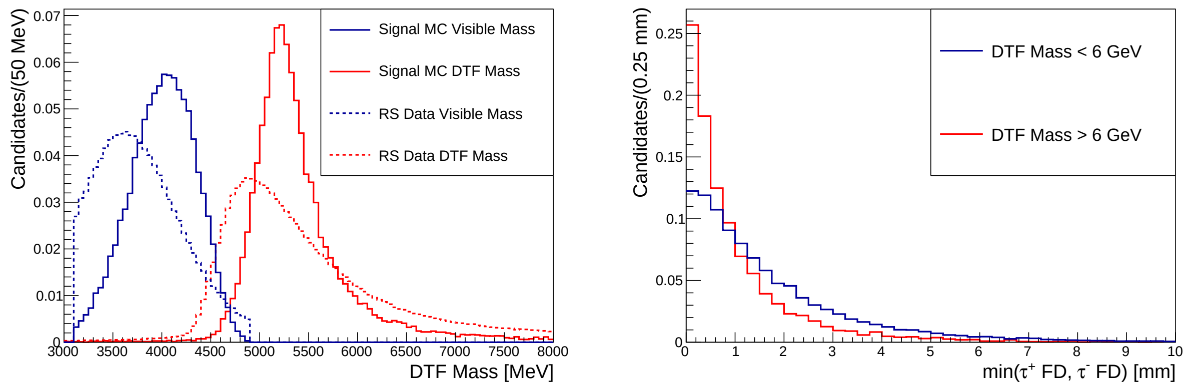
We also calculate the components of the τ momenta parallel to the B^0 flight direction. If θ_1 and θ_2 are the angles made by the τ_1 and τ_2 momenta with respect to the B^0 flight direction, then we have

$$|\vec{p}_{\parallel,1}| = |\vec{p}_{\perp,1}| \times \frac{\cos\theta_1}{\sqrt{1-\cos^2\theta_1}} \tag{19}$$

$$|\vec{p}_{\parallel,2}| = |\vec{p}_{\perp,2}| \times \frac{\cos\theta_2}{\sqrt{1-\cos^2\theta_2}} \tag{20}$$

By virtue of now knowing the components of the τ momenta parallel and perpendicular to the B^0 flight direction, we now know the 3-momentum of both τ's. Note that our re-calculation of the τ momenta is independent of whether the τ decays into $3\pi\nu_\tau$ or $3\pi\pi^0\nu_\tau$. Now, in the DTF code, a massless neutrino particle is added to the daughters of each τ. The 3-momentum of this neutrino is set to the difference between the recalculated τ momentum and the original 3π momentum. The τ momenta are set to the recalculated values. The B^0 momentum is set to the vector sum of the momenta of the K^*, τ_1 and τ_2. Once this initialization is done, the usual DTF fit is run.

The effect of recalculating the mass with this modified version of DTF is shown in Fig. 27a. The recalculated mass distribution is shifted upwards compared to the visible mass (because we have effectively added back the momentum from the missing neutrinos), and is also narrower in shape, albeit with a long tail extending to 8 GeV. It is important to note that some fraction of the time, the DTF fit fails (meaning that it fails to converge on a local minimum). This failure usually happens when the τ production and decay vertices are so close to each other that the τ flight direction cannot be reliably determined. In order to understand what kind of candidates end up in the upper tail, it is instructive to see (Fig.27b) that the minimum τ flight distance (of the two τ's in the event) is lesser on average for the signal candidates in the upper tail of the DTF mass compared to those in the core. For such



(a) (b)

Figure 27: (a) A comparison of visible $(K^{*0}(3\pi)^{+}(3\pi)^{-})$ mass with DTF recalculated mass for signal MC and data after offline pre-selections. (b) A comparison of the minimum τ flight distance between signal candidates which have a DTF mass below 6 GeV with those that have a mass above it

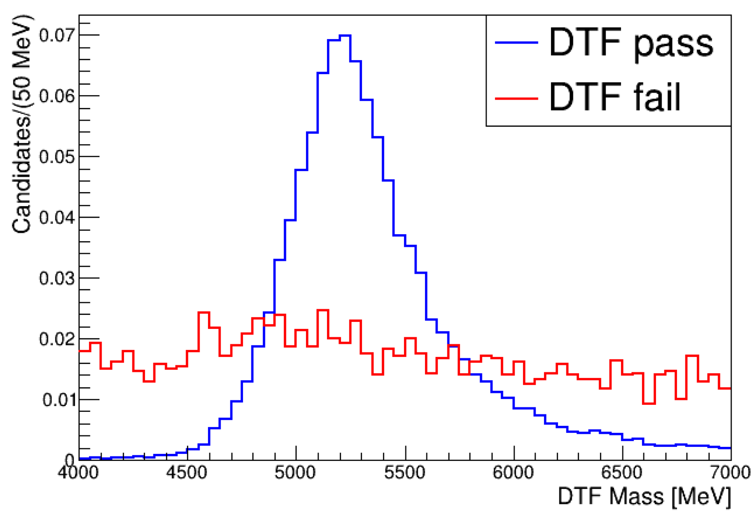


Figure 28: A comparison of the DTF mass distribution in simulation for signal candidates which pass the fit (blue lineshape) with those that fail (red lineshape).

candidates, the uncertainty on the τ momentum is larger. A comparison of the shape of signal candidates which pass the DTF fit with those that fail is shown in Fig. 28. No peaking structure is seen in the candidates which fail the fit.

All our data and MC samples are put through this modified DTF fitter. The resulting mass distribution in data is what we fit to after all selections are made.

4.3 Methods

4.3.1 Boosted Decision Trees

Boosted Decision Trees (BDTs) are a key part of this analysis, used to make multivariate selections on data to discriminate against background. In this section, we provide an overview of what BDTs are and how they work.

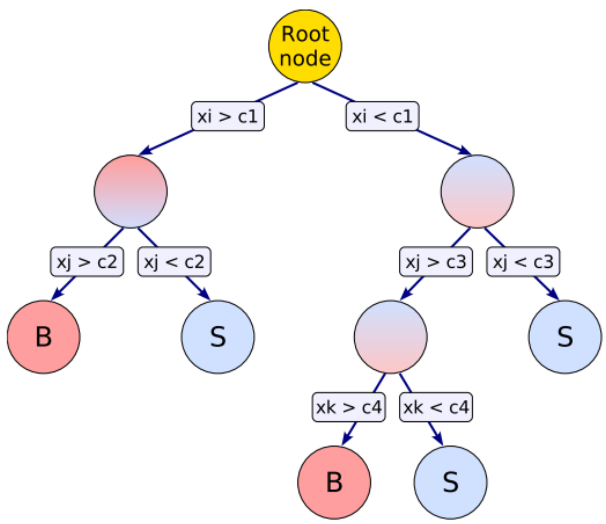


Figure 29: Schematic of a decision tree [37]

A decision tree, in the context of classification problems, is a machine learning algorithm that uses a set of features to recursively split data into different categories. In essence, a decision tree is a series of if-else conditions. A schematic of a decision tree is shown in Fig. 29. The topmost node in a tree is referred to as the root node, while the bottommost ones are the leaf nodes. At each node except for the leaf nodes, a binary split is applied to the data using the feature x_i that gives the best separation between signal and background at that point. The splitting is continued until a predefined stopping criterion (for example in terms of the maximum number of levels allowed) is fulfilled. The leaf nodes are then classified as either signal (S) or background (B) depending on the majority of events that end up in them. In this way, the n dimensional phase space (n being the number of features) is split up into signal and background regions. Compared to selections based on rectangular cuts on the features, which would normally select one hypercube of the phase space against another, decision trees are capable of splitting up the phase space into many hypercubes and designating each as signal or background.

A single decision tree by itself is a weak learner, i.e. it would not be very powerful in separating signal from background. Single trees are also unstable with respect to statistical fluctuations in the training sample from which the tree is constructed. Both of these issues are overcome by constructing a forest of decision trees using the so called *boosting* technique, and combining them into a single classifier. Through boosting, each subsequent tree is trained based on a reweighted version of the same training sample seen by the first tree (we go into specifics of boosting below).

The training of a decision tree is the process that determines for each node the most discriminating feature as well as the cut on that feature, based on a sample of training data that is provided with tagged signal and background events. The training starts with the root node, where the feature that provides the most seperation between signal and background is selected and cut on, resulting in two subsets of the original sample which go through the same procedure, until the stopping criterion is reached. The choice of feature and cut at a node is based on the Gini index, defined as

$$G = p(1 - p) \tag{21}$$

where p is the purity of a node ($\frac{S}{S+B}$). The feature and cut value which maximize the increase in the Gini index between the parent node and the sum of the two daughter nodes (weighted by their relative fraction of events) are chosen. While in principle, the splitting could be continued until every leaf node contained purely signal and background events, such a tree would be strongly overtrained (meaning that it would be heavily tuned to the statistical fluctuations present in the training sample, and would not perform well on unseen data). Therefore, individual trees are usually limited to a maximum depth of two or three, and the leaf nodes invariably contain some fraction of misclassified events. The misclassification rate, err, of a node is defined as

$$1 - \max(p, 1-p). \tag{22}$$

The gradient boosting technique is used to combine many weak decision trees to form a much more powerful Boosted Decision Tree (BDT) which is also stable against fluctuations in the training sample. Signal events are assigned a target value $y = +1$, while background events are assigned $y = -1$. The training procedure aims to build a model which can predict the target response through a function $F(\mathbf{x})$ (where $\mathbf{x}$ is a vector of the input features), which is a combination of M weak learners $f(x)$

$$F(\mathbf{x}; \beta, \alpha) = \sum_{m=1}^{M} \beta_m f(x; \alpha_m). \tag{23}$$

The boosting procedure aims to find the values of the parameters β_m such that the disagreement between the true response y and the model response $F(\mathbf{x})$ is minimized. The objective function used for this minimization is known as the loss function $L(F, y)$. The $f(x; \alpha_m)$ are the individual decision trees, and α_m are the parameters that define the decision trees, such as the features and cuts at each node. The loss function used in TMVA's implementation of gradient boosting is the binomial log likelihood

$$L(F, y) = \ln\bigl(1 + e^{-2F(\mathbf{x})y}\bigr). \tag{24}$$

The first decision tree is built in the usual fashion, and the β_m parameter multiplying the response of each subsequent tree is tuned to minimize the loss function. A gradient descent approach is numerically used in the minimization process.

Overtraining (or overfitting) is a common concern in the training of multivariate classifiers. This is when a classifier is over tuned on the training data, and ends up learning the statistical fluctuations present in this sample and interpreting them as features. When this happens, the classifier will seem to perform well on the training data, but will underperform when tested against unseen data. Therefore, we always withhold a subset of data from training, and use it to judge the performance of the BDT (for example, to obtain the efficiency of a BDT selection). Boosted decision trees are generally resistant to overtraining, but it is straightforward to check for overtraining in our classifier. We compare the response of the BDT on the training sample to its response on an unseen sample of data (referred to as the testing sample). In the presence of overtraining in the BDT, we would expect these responses to be incompatible.

4.3.2 Hypothesis Testing

Statistical techniques are necessary to interpret the results of a search, and to define a sensitivity to the parameter of interest. In the context of a search for new physics, two competing hypotheses (or models) can be defined. The analysis of the search results can then be formulated in terms of a hypothesis test. The background only hypothesis, B, describes only the known physics processes, while the signal plus background hypothesis, $S+B$, describes the sought after signal on top of the known background. The unknown signal strength (which could be parameterized as a branching fraction, or cross section for example) μ is the continuous parameter of interest which defines a family of signal plus background models like $\mu S + B$.

One must then define an observable (or set of observables) which encompass the results of the search. The observable could be a reconstructed quantity like the momentum of some particle, or an invariant mass (as in this work), or it could be a composite feature such as the output of a multivariate classifier. The next step is to define a test statistic, which is a function of the observables and the model parameters, to rank experiments from least to most signal like (or vice verse, depending on convention; the monotonicity is the important feature). Finally, one specifies rules for exclusion or discovery, by defining ranges of the test statistic in which observables will favour one hypothesis over the other. It is desirable to be able to provide a significance associated with discovery (or a confidence level for an exclusion) rather than simply choosing a hypothesis in a binary manner. In case of an exclusion, a limit is defined as the value of the parameter of interest (like the signal strength) which is excluded at the specified confidence level. The limit is an upper (lower) limit if the exclusion confidence is lesser (greater) than the specified confidence for all values of the parameter of interest above (below) the specified limit.

The agreement of a given dataset with a hypothesis H can be summarized in terms of a p-value, i.e. a probability under the assumption of hypothesis, H of obtaining data with equal or less compatibility with the predictions of H. The hypothesis can be then rejected if the p-value is less than a specified threshold. This p-value is conventionally converted into an equivalent significance Z such that a variable following a standard Gaussian distribution found Z standard deviations above its mean will have an upper tail probability of p i.e.

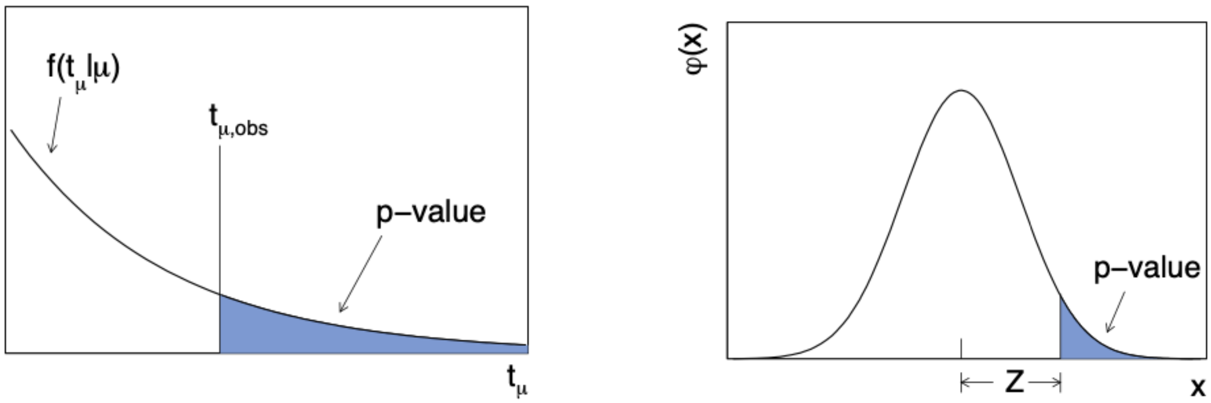
$$Z = \Phi^{-1}(1 - p) \tag{25}$$

where Φ is the cumulative distribution function of the standard Gaussian. The convention in particle physics for declaring discovery is to set $Z = 5$, corresponding to $p = 2.87 \times 10^{-7}$. This is the probability of the observed data if the background only hypothesis was true. For exclusion of a signal hypothesis, a conventional threshold p-value of 0.05 (corresponding to a 95% confidence level, and $Z = 1.64$) is used. This is shown visually in Fig.30.

Let the parameter of interest (POI) be denoted as μ, and all the other parameters in the model, known as nuisance parameters, be collectively denoted as $\boldsymbol{\theta}$. To perform hypothesis testing, we use the profile likelihood ratio

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})}, \tag{26}$$

where L denotes the likelihood. $\hat{\hat{\boldsymbol{\theta}}}$ in the numerator is the value of $\boldsymbol{\theta}$ that maximizes L for the specified μ. This is the conditional maximum likelihood (ML) estimate of $\boldsymbol{\theta}$, and is a



(a) (b)

Figure 30: (Left) Illustration of the relation between the p-value and the observed value of the test statistic. (Right) The standard normal distribution, showing the relation between the significance Z and the p-value. [38]

function of μ. The denominator is the usual (unconditional) maximized likelihood function, with $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ being the ML estimates for μ and $\boldsymbol{\theta}$ respectively.

From the above definition, we see that $0 \leq \lambda(\mu) \leq 1$ with larger values of $\lambda(\mu)$ denoting better agreement of the specified μ with the observed data. The test statistic q_μ appropriate for setting upper limits is defined as

$$q_\mu = \begin{cases} -2 \ln \lambda(\mu) & \hat{\mu} < \mu, \\ 0 & \hat{\mu} > \mu \end{cases} \tag{27}$$

with smaller values of q_μ now denoting greater agreement between data and the specified value of μ. q_μ is set to 0 for $\hat{\mu} > \mu$ because in the case of setting upper limits, data with $\hat{\mu} > \mu$ is not regarded as less compatible with μ than the data obtained. Note from the definition of the profile likelihood ratio in Eq.26 that μ in the numerator is a free parameter, not determined by data. As a result, the observed value ($q_{\mu,\text{obs}}$) as well as the sampling distributions of the test statistic depend on μ.

Now that the test statistic is decided, a p-value can be calculated to measure the compatibility of the observed data with a hypothesized value of μ

$$p_\mu = \int_{q_{\mu,\text{obs}}}^{\infty} f(q_\mu|\mu) dq_\mu, \tag{28}$$

where $f(q_\mu|\mu)$ is the pdf of q_μ under the hypothesized value of μ (also known as the sampling distribution of the test statistic). Normally, the sampling distribution is obtained through Monte Carlo calculations using the known pdf's of the signal and background components. But often, this is computationally intensive. Cowan et. al. [60] present asymptotic formulae for the distributions of various test statistics. This allows the calculation of the p-value without recourse to extensive MC calculations. The asymptotic formula is based on the approximation of the profile likelihood ratio provided by Wald [61].

Wald's approximation of $\lambda(\mu)$ is as follows: Consider a test of some hypothesized value of the parameter μ. If the data is distributed according to a value μ', then

$$-2 \ln \lambda(\mu) = \frac{(\mu - \hat{\mu})^2}{\sigma^2} + O(1/\sqrt{N}), \tag{29}$$

where $\hat{\mu}$ follows a Gaussian distribution with mean μ' and standard deviation σ. N represents the size of the data sample. σ can be determined either from the second derivative of the log likelihood, or from a special data set known as the Asimov data set [38]. The Asimov data set is defined such that when the maximum likelihood estimators of all the parameters are evaluated on it, their true values are obtained. From this approximation, the "asymptotic" p value is given as

$$p_\mu = 1 - \Phi(\sqrt{q_\mu}). \tag{30}$$

Now, it is possible to determine an upper limit on the parameter of interest by finding the $S + B$ hypothesis whose p-value (which we can denote as p_{S+B}) is below a specific threshold

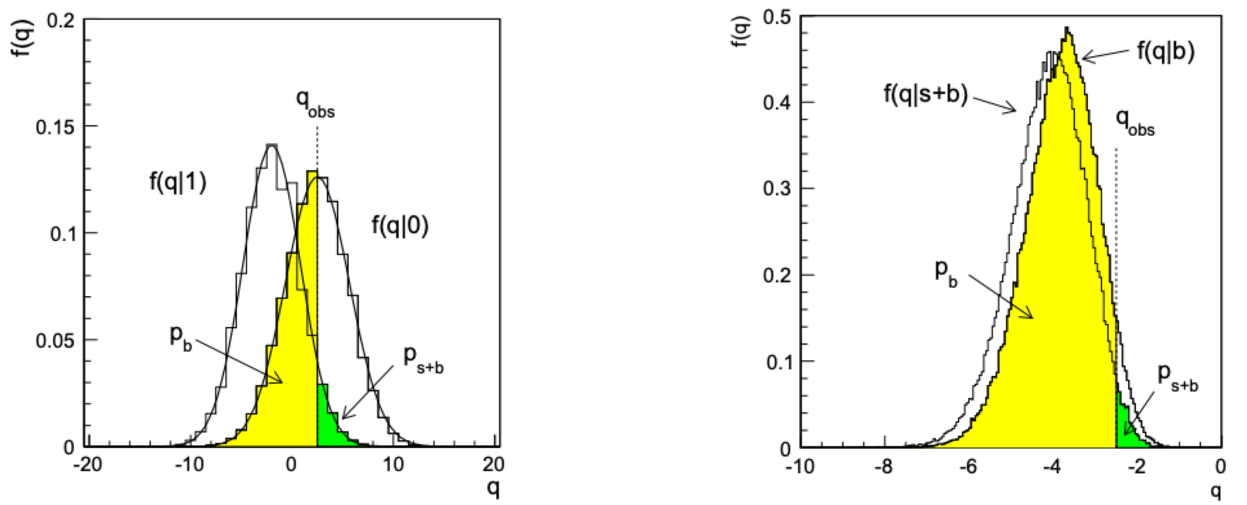
α $p_{S+B} < \alpha$. This is the "$\mathrm{CL_{S+B}}$" method.

$$p_{S+B} = P(q \leq q_{\text{obs}}|S+B) = \int_{q_{\text{obs}}}^{\infty} f(q|S+B)dq \tag{31}$$

Similarly, the p-value under the background only hypothesis, p_B, can be calculated

$$p_B = P(q \leq q_{\text{obs}}|B) = \int_{-\infty}^{q_{\text{obs}}} f(q|B)dq \tag{32}$$

These p values are visualized in Fig.31a. The problem with the "$\mathrm{CL_{s+b}}$" method is that it will exclude, with a probability close to α, hypotheses to which we have little to no sensitivity. This corresponds to the cases where the expected number of signal events are much less than background, meaning that the distributions of q_{S+B} and q_B are almost overlapping, as in Fig. 31b.



(a) (b)

Figure 31: (Left) Distributions of the test statistic q under the $s + b$ and b hypotheses, (Right) Distributions of the test variable q under the $s + b$ and b hypotheses in an example where one has very little sensitivity to the signal model [39].

For example, if the expected signal and background counts are given by s and b, and $s << b$, and the observed number of events has a significant downward fluctuation relative to $s + b(\approx b)$, then this small value of s will be excluded. In this corner case, we would desire the probability of exclusion to be zero, but with the $\mathrm{CL_{s+b}}$ method, this approaches α. To protect against this, the $\mathrm{CL_s}$ method [62] is used, where a signal model is excluded if

$$CL_s = \frac{CL_{s+b}}{CL_b} = \frac{p_{s+b}}{1 - p_b} < \alpha. \tag{33}$$

In cases where the $f(q|s + b)$ and $f(q|b)$ are widely separated, $1 - p_b \approx 1$, meaning that the models excluded by $\mathrm{CL_s}$ are almost the same as those excluded by $\mathrm{CL_{s+b}}$. However, if the distributions are almost overlapping, as in Fig. 31b, $1 - p_b$ becomes small and p_{s+b} is increased, preventing these cases from being excluded when sensitivity is low.

4.4 Data Samples

We analyze data collected from pp collisions at a 13 TeV center-of-mass energy by the LHCb detector in the years 2016, 2017 and 2018, during Run 2 of the LHC operation. This sample corresponds to an integrated luminosity of approximately $6\,\mathrm{fb}^{-1}$. An event is defined in LHCb as all the activity that occurs in 1 pp bunch crossing. We utilize events that fire specific L0, HLT1, and HLT2 trigger lines. The selections pertaining to our lines of interest are detailed in Sec. 4.5.1.

4.4.1 Simulated samples

Simulated samples (also referred to as "Monte Carlo" (MC) samples) of the signal and relevant physical background decay processes are crucial to analyses of particle physics data. In this thesis, the simulated samples are used to extract the efficiencies of detection and reconstruction of the relevant decays, as well as the efficiencies of the selections applied over the course of the analysis. The knowledge of the signal efficiency is essential for our goal for measuring or constraining the branching fraction for $B^0 \to K^{*0}\tau^+\tau^-$. The knowledge of these efficiences is also paramount in order to be able to optimize the selections to maximize signal efficiency and background rejection simultaneously. These samples are also used to extract the shapes of the mass distribution for the signal and physical background processes, for use in the final fit. Since these simulated samples play such a central role in this measurement, it is crucial that they model the processes close to reality.

In LHCb, the simulated samples are put through the same data flow as the real data, with some extra steps at first. Simulated events are generated in a three part process, using the GAUSS software package [63]. The first step involves the **generation** of proton-proton collisions via the event generator PYTHIA [64] with an LHCb specific configuration [65]. This is followed by the **decay** of the unstable particles generated by the collision, handled by the EVTGEN library, in which final state radiation is generated using PHOTOS [66]. The third and final step of the particle simulation is the **propagation** of the generated particles through a simulated version of the LHCb detector, in order to obtain the simulated detector response to the event. This is the most time consuming step since modeling the interactions of particles with detector materials is computationally very intensive. This step is implemented using the GEANT4 toolkit [67] [68]. Following this, the simulated detector response is put through the same steps as the real data. The various simulated samples used in this analysis are listed in Table 4. We generate MC samples for our signal, as well as for three sets of dangerous physical backgrounds. In order to reduce the computational time for generation, the simulated samples are generated with some cuts (referred to as generator level cuts) to model the effect of detector acceptance and to discard low momentum tracks. The generator level cuts used in this analysis are listed in Table5. Generated events which don't pass these cuts are discarded without being put through the detector interaction phase. These cuts also have an efficiency associated with them (referred to as the generator level efficiency), which is provided centrally by the collaboration.

Table 4: Details of simulated datasets.

Decay	Type	Generated Yield
$B^0 \to (K^{*0} \to K^+\pi^-)(\tau^+ \to \pi^+\pi^-\pi^+(\pi^0)\nu_\tau)(\tau^- \to \pi^-\pi^+\pi^-(\pi^0)\nu_\tau)$	Signal	$\sim$ 24M
$B^0 \to (K^{*0} \to K^+\pi^-)(D^{(*)} \to 3\pi X)(\overline{D}^{(*)} \to 3\pi X)$	Background	$\sim$ 10M
$B^+ \to (K^{*0} \to K^+\pi^-)(D^{(*)} \to 3\pi X)(\overline{D}^{(*)} \to 3\pi X)$	Background	$\sim$ 4M
$B_s^0 \to (K^{*0} \to K^+\pi^-)(D^{(*)} \to 3\pi X)(\overline{D}^{(*)} \to 3\pi X)$	Background	$\sim$ 4M

Another procedure used to speed up the simulation is called ReDecay [69]. In this method, the underlying event (all the constituents of the event except for the signal decay) is reused a set number of times (lets call it N), along with the origin vertex and kinematic of the signal particle. The decay time of the signal particle, and hence its decay vertex, and the final state particles are varied in the usual manner. In doing so, the computational cost of propagating the quasi-stable particles of the underlying event through the detector is reduced by a factor of N. This approach produces efficiencies and resolutions which are identical to that of the original simulation procedure.In this analysis, ReDecay is used with $N = 100$ only for the

Table 5: Generator level cuts. The θ cuts on the charged tracks model the geometrical acceptance of the LHCb detector. p_T denotes transverse momentum i.e. momentum of a track perpendicular to the beam direction.

Variable	Sample	Cut
p_T of charged tracks	Signal	$> 250\,\mathrm{MeV}$
p_T of charged tracks	Background	$> 250\,\mathrm{MeV}$
θ of charged tracks	Signal	$0.01 - 0.4\,\mathrm{rad}$
θ of charged tracks	Background	$0.01 - 0.4\,\mathrm{rad}$
p of charged tracks	Background	$> 2000\,\mathrm{MeV}$
Ancestor of pions from Charm	Background	not K^0

background MC samples.

The $(K^{*0} \to K^+\pi^-)$ decay in all the simulated samples is modelled by a vector-to-scalar-scalar (VSS) decay model, which generates the correct angular distributions for the daughters. The decay of $\tau^+ \to \pi^+\pi^-\pi^+(\pi^0)\nu_\tau$ is modelled using the `TAUOLA` model in EvtGen, which uses the Resonance Chiral Lagrangian model [70] with a tuning based on BaBar data for $\tau^+ \to \pi^+\pi^-\pi^+\nu_\tau$ [71]. Figure 32 shows the BaBar measurements of the $\tau^+ \to \pi^+\pi^-\pi^+\nu_\tau$ with the model overlaid for comparison.

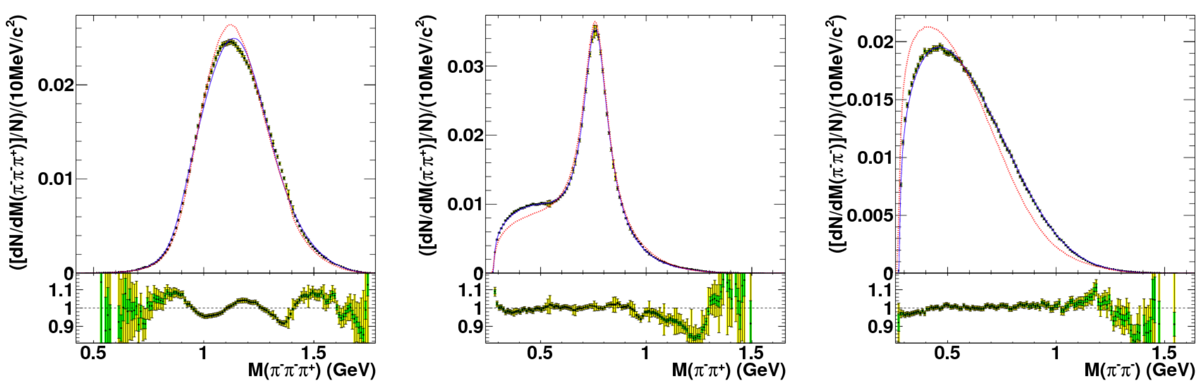


Figure 32: The BaBar measurements of the $\tau^- \to \pi^-\pi^+\pi^-\nu_\tau$ are shown (data points) overlaid with the Resonance Chiral Lagrangian model tuned to the data (blue line) and the old model from CLEO (red line) for $m(\pi^-\pi^-\pi^+)$ (left), $m(\pi^-\pi^+)$ (middle) and $m(\pi^-\pi^-)$ (right).

The three background MC samples are made up of a cocktail of the possible $Bx \to$

$K^{*0}D^{(*)}\overline{D}^{(*)}$ decays, weighted according to measurements of their branching fractions made in a parallel analysis (see Sec. 4.8), as well as the known $D^{(*)} \to 3\pi X$ decay modes, extracted from the PDG.

4.5 Selections

The signal decay of interest in this analysis is expected to be a rare one. According to the SM, approximately one in 10 million B^0 mesons will decay to $K^{*0}\tau^+\tau^-$. As a result, most of the data that we collect will be background that we must minimize in order to gain sensitivity to the signal. This task is made challenging by the large number of light hadrons (7 pions and 1 kaon) present in the final state of our signal decay. We sift through the data by putting it through different stages of selection, which are outlined in this chapter. The data must first pass through the event reconstruction and selections which happen in the trigger, reconstruction and stripping stages, as explained in Sec.3.3. The specific trigger and stripping lines that we employ are outlined in subsections below. These selections are applied to maximize signal efficiency and background rejection by exploiting differences in the signature of the signal and background candidates. The signal decays are characterized by a $K^+\pi^-$ vertex of 2 charged tracks detached from the PV and two $\pi^+\pi^-\pi^+$ vertices downstream of the $K^+\pi^-$ vertex. After applying loose selections to the data, we train and apply two Boosted Decision Trees (BDTs) to the data to intelligently discriminate signal from background. We then select a "best candidate" per event in the data based on the output of the second stage BDT. Only this best candidate is retained per event, while the rest of the candidates are rejected. The selection on the output of these two BDTs is optimized by using background-only toy studies to maximize the expected sensitivity to the signal.

4.5.1 Trigger selections

We utilize events that fire specific L0, HLT1, and HLT2 trigger lines. At the L0 level, we require that the event fire one of L0HadronDecision, L0MuonDecision, L0ElectronDecision and L0PhotonDecision. Events which pass the L0 level must pass one of the HLT1 lines HLT1TrackMVADecision and HLT1TwoTrackMVADecision. Finally, events surviving the HLT1 stage must pass one of the HLT2 lines HLT2Topo2BodyDecision, HLT2Topo3BodyDecision and HLT2Topo4BodyDecision. Each trigger line is just an algorithm formed by a set of selections designed to retained interesting events.

An event is capable of producing multiple candidates (a candidate is the jargon used for the head of a decay in data which has passed the selections at a given stage). Candidates can classified based on the role their tracks played in firing the trigger. If tracks in the event belonging to the decay chain of the signal candidate alone are responsible for firing the trigger, then the candidate is said to be TOS ("triggered on signal") for that particular line. In other words, if a trigger line which fired on an event would otherwise not have fired if we took out the tracks belonging to the signal candidate, then the candidate is TOS on that line. If the trigger line would have fired even without the tracks belonging to the signal candidate, then the candidate is said to be TIS ("triggered independent of signal") on that line. It is also possible for a candidate to be TOB ("triggered on both"), where tracks belonging to the signal candidate as well as other tracks are necessary to have fired the trigger, but this category is not relevant in our analysis.

At the L0 hardware trigger, we accept candidates which are either TOS or TIS on any of the lines: `L0Hadron`, `L0Muon`, `L0Electron` or `L0Photon`. The details of how these are constructed in the L0 trigger are outlined in Sec.3.2.9.1. At the HLT1 stage, two inclusive trigger lines are used in this analysis; `HLT1TrackMVA` and `HLT1TwoTrackMVA`, with the candidate required to be TOS on any one of these. These are general purpose triggers designed to select tracks which are displaced from the PV, a signature of b-hadron decays.

- `HLT1TrackMVA` selects tracks with a track $\chi^2/\mathrm{ndf} < 2.5$, ghost probability (the probability of a fake track) < 0.2. After these selections, tracks which fall into one of two categories are accepted. They must either satisfy $\chi^2_{\mathrm{IP}} > 7.4$ and $p_{\mathrm{T}} > 2.5\,\mathrm{GeV}$ or satisfy the 2D hyperbolic selection(with p_{T} in GeV):

$$\log \chi^2_{\mathrm{IP}} = \frac{1}{(p_{\mathrm{T}} - 1)^2} + \frac{b}{25}(25 - p_{\mathrm{T}}) + \log(7.4).$$

The parameter b is defined with different values (2.3 and 1.1), corresponding to tight and loose configurations, with the tighter configuration being used when the buffer between HLT1 and HLT2 gets too full. That majority of the data used in this analysis is collected with the loose configuration.

- `HLT1TwoTrackMVA` selects a pair of tracks consistent with originating from the same displaced vertex. It requires each track to have $p_{\mathrm{T}} > 500\,\mathrm{MeV}$, $p > 5\,\mathrm{GeV}$, track $\chi^2/\mathrm{ndf} < 2.5$, $\chi^2_{\mathrm{IP}} > 4$ w.r.t the PV. It also employs a MatrixNet classifier [72] to select candidates based on their vertex fit quality and displacement, the scalar sum of the two tracks' p_{T} and the displacement of each track w.r.t. the PV.

The HLT2 topological trigger lines used in this analysis (requiring the candidate to be TOS): `HLT2Topo2Body`, `HLT2Topo3Body` and `HLT2Topo4Body`, are based on a novel Bonsai Boosted Decision Tree (BBDT) technique [73]. This is a BDT based multivariate technique where the input variables have been discretized into predefined intervals which are designed to be larger than the detectors resolution in those variables, ensuring that the regions that the classifier decides to "keep" are not smaller than detector resolution. This allows the classifier to learn the general traits of b-hadron decays, rather than learning a large set of specific traits. The input variables used to train the BDT are :$\sum |p_{\mathrm{T}}|$, p_T^{min}, mass, $m_{\mathrm{corr}} = \sqrt{m^2 + |p_T^{\mathrm{miss}}|^2} + |p_T^{\mathrm{miss}}|$, distance of closest approach, candidate χ^2_{IP} and flight distance χ^2. The cuts applied on the BBDT outputs for the Topo2Body, Topo3Body and Topo4Body lines are 0.4, 0.4 and 0.3 respectively.

We define the following two disjoint trigger categories for later use:

`TOS` = L0_TOS & HLT1_TOS & HLT2_TOS

`TISnotTOS` = (L0_TIS & !L0_TOS) & HLT1_TOS & HLT2_TOS

where & denotes the logical AND operator, and ! denotes the logical NOT operator. L0_TOS and L0_TIS are simply defined as

L0_TOS = L0_Photon_TOS ‖ L0_Hadron_TOS ‖ L0_Photon_TOS ‖ L0_Electron_TOS, and

L0_TIS = L0_Photon_TIS ‖ L0_Hadron_TIS ‖ L0_Photon_TIS ‖ L0_Electron_TIS where $\|$ denotes the logical OR operation.

4.5.2 Stripping selections

We utilize events passing the `B2KstTauTau_TauTau` and `B2KstTauTau_TauTau_SameSign` stripping lines, with the former reconstructing $B^0 \to K^{*0}\tau^+\tau^-$ candidates, and the latter reconstructing the "wrong-sign (WS)" decays $B^0 \to K^{*0}\tau^+\tau^+$, $B^0 \to K^{*0}\tau^-\tau^-$. A stripping line is simply a set of selections. The WS data sample is used in some parts of the analysis as a proxy for the combinatorial background present in the "right-sign (RS)" data sample. The versions of Stripping used for processing data (and MC) for the years 2016, 2017 and 2018 are 28r2, 29r2 and 34 respectively. DaVinci v45r6 is used to process stripped data and MC to make the ROOT tuples that are analyzed offline. The selections applied in the Stripping line are detailed in Tables 6, 7, 8. The RS and WS stripping lines consist of the same selections, they just build different decays. It is important to note that these selections are applied on an event level, and not on a candidate level (it is possible for an event to have multiple candidates in it). If an event has a candidate that passes these cuts, it is written to disk for further processing.

Table 6: Stripping selections on $\tau \to \pi^+\pi^-\pi^+$. DOCA denotes distance of closest approach. ProbNNpi is a particle identification variable pertaining to the likelihood of the track being a pion. Ghost probability is the probability of a track being fake. DIRA (direction angle) of a composite particle with respect to a vertex is defined as the cosine of the angle between the particles momentum and the line joining the particle's decay vertex to the vertex of interest.

Variable	Cut
Σp_{T} of pions	$> 800\,\mathrm{MeV}$
$m_{3\pi}$	$400 - 2100\,\mathrm{MeV}$
max DOCA of any pair of pions	$0.2\,\mathrm{mm}$
# of pions with $p_{\mathrm{T}} > 800\,\mathrm{MeV}$	$>= 1$
$\pi\ p_{\mathrm{T}}$	$> 250\,\mathrm{MeV}$
$\pi\ p$	$> 2000\,\mathrm{MeV}$
$\pi\ \chi^2_{\mathrm{IP,PV}}$	> 16
π track χ^2/ndf	< 4
π track ghost prob.	< 0.4
π ProbNNpi	> 0.55
$\tau\ p_{\mathrm{T}}$	$> 1000\,\mathrm{MeV}$
τ visible mass	$500 - 2000\,\mathrm{MeV}$
τ DIRA w.r.t PV	> 0.99
$\tau\ \chi^2_{\mathrm{vtx}}$	< 16
τ vtx. distance χ^2 from PV	> 16
τ vtx. radial distance from PV	$0.1 - 7\,\mathrm{mm}$
τ vtx. z-distance from PV	$> 5\,\mathrm{mm}$

Table 7: Stripping selections on $K^{*0} \to K^+\pi^-$. ProbNNK is a particle identification variable pertaining to the likelihood of the track being a kaon.

Variable	Cut
$\pi\ p_{\mathrm{T}}$	$> 250\,\mathrm{MeV}$
$\pi\ \chi^2_{\mathrm{IP}}$ w.r.t. PV	> 4
π ProbNNpi	> 0.5
π track χ^2/ndf	< 4
$K^+\ p_{\mathrm{T}}$	$> 250\,\mathrm{MeV}$
$K^+\ \chi^2_{\mathrm{IP}}$ w.r.t. PV	> 4
K^+ ProbNNK	> 0.2
K^+ track χ^2/ndf	< 4
$m_{K^+\pi^-}$	$700 - 1100\,\mathrm{MeV}$
$K^{*0}\ p_{\mathrm{T}}$	$> 1000\,\mathrm{MeV}$
K^{*0} vertex distance from PV	$> 3\,\mathrm{mm}$
$K^{*0}\ \chi^2_{\mathrm{vtx}}$	< 15

Table 8: Stripping selections on $B^0 \to K^{*0}\tau^+\tau^-$.

Variable	Cut
$m_{K^{*0}\tau^+}$	$< 5000\,\mathrm{MeV}$
$m_{K^{*0}\tau^+\tau^-}$	$2000 - 10000\,\mathrm{MeV}$
$B^0\ \chi^2_{\mathrm{vtx}}$	< 100
B^0 vertex distance χ^2 from PV	> 80
B^0 vertex distance from PV	$< 40\,\mathrm{mm}$

4.5.3 Offline pre-selections

In order to further reject large chunks of background at the cost of very little signal, further loose cuts are applied after the stripping selections. The reconstructed invariant mass of the K^{*0} is required to be within 200 MeV of its nominal mass of 892 MeV. Similarly the visible 3π reconstructed mass of the τ's are required to be less than 1620 MeV. The nominal mass of the τ lepton is 1777 MeV; since the neutrino (and sometimes neutral pion) from the τ decay is not reconstructed, the invariant mass distribution of the 3π combination stops well before the nominal mass. We also impose requirements on the invariant masses of two-pion combinations in the $\tau \to 3\pi X$ decay. The pions of the same sign are numbered 1 and 3, and the odd one out is numbered 2. Therefore M_{12} refers to the invariant mass of the combination of pion 1 and pion 2. We reject 3π vertices of bad quality by requiring that the χ^2/ndf of the τ vertex is less than 5. The visible mass of the B^0 (reconstructed in data as $K^{*0}(3\pi)^+(3\pi)^-$ is required to be between 3100 and 4900 MeV, its p_{T} is required to be greater than 4000 MeV and its vertex is required to have a $\chi^2 < 85$.

The cut values on all these variables were chosen by looking at their distributions in signal MC and data, and picking values which reject data without losing any significant signal. Finally, we require that the DTF fit converged successfully. A comparison of the features pertaining to the pre-selections is shown in Figs. 33 and 34 between data and signal MC before the cuts are applied. After these cuts are applied, approximately 52.1 million candidates are obtained in the data, compared to an expected signal yield of 892 ± 48 candidates at a signal branching fraction of 1×10^{-3}.

Table 9: Offline pre-selections. Cuts relevant to the τ are applied to both the τ^+ and τ^-.

Variable	Cut
$m_{K^{*0}}$	$692 - 1092\,\mathrm{MeV}$
τ visible mass	$< 1620\,\mathrm{MeV}$
$\tau\ m_{13}$	$< 1250\,\mathrm{MeV}$
$\tau\ m_{12}$	$< 1200\,\mathrm{MeV}$
$\tau\ m_{23}$	$< 1200\,\mathrm{MeV}$
τ vertex χ^2/ndf	< 5
B^0 visible mass	$3100 - 4900\,\mathrm{MeV}$
$B^0\ p_\mathrm{T}$	$> 4000\,\mathrm{MeV}$
B^0 vertex χ^2	< 85
DTF fit status	pass

4.5.4 Isolation BDT

Given the large number of final state charged hadrons in our decay of interest, we face significant backgrounds from partially reconstructed decays, where one might have a decay like $D^0 \to K^-\pi^+\pi^-\pi^+$ in an event, and the three pions are reconstructed as $\tau \to 3\pi X$ in our data sample. In order to discriminate against these backgrounds, we make use of custom made isolation variables which measure the activity of other particles in the vicinity of the tracks in our reconstructed decay tree. One aspect of this isolation is as follows: say we have a reconstructed $\tau \to 3\pi X$ candidate in data, we can take the *other* charged tracks in the event (which do not belong to our decay tree), and ask how well they vertex with our $\tau \to 3\pi$ vertex. For background cases like the one mentioned above, we would be more likely to find an extra track that is consistent with the τ vertex than for a real signal decay.

Using some of the isolation features described in the Sections 4.5.4.1, 4.5.4.2, 4.5.4.3, we train a BDT (referred to as the isolation BDT) to discriminate signal from background

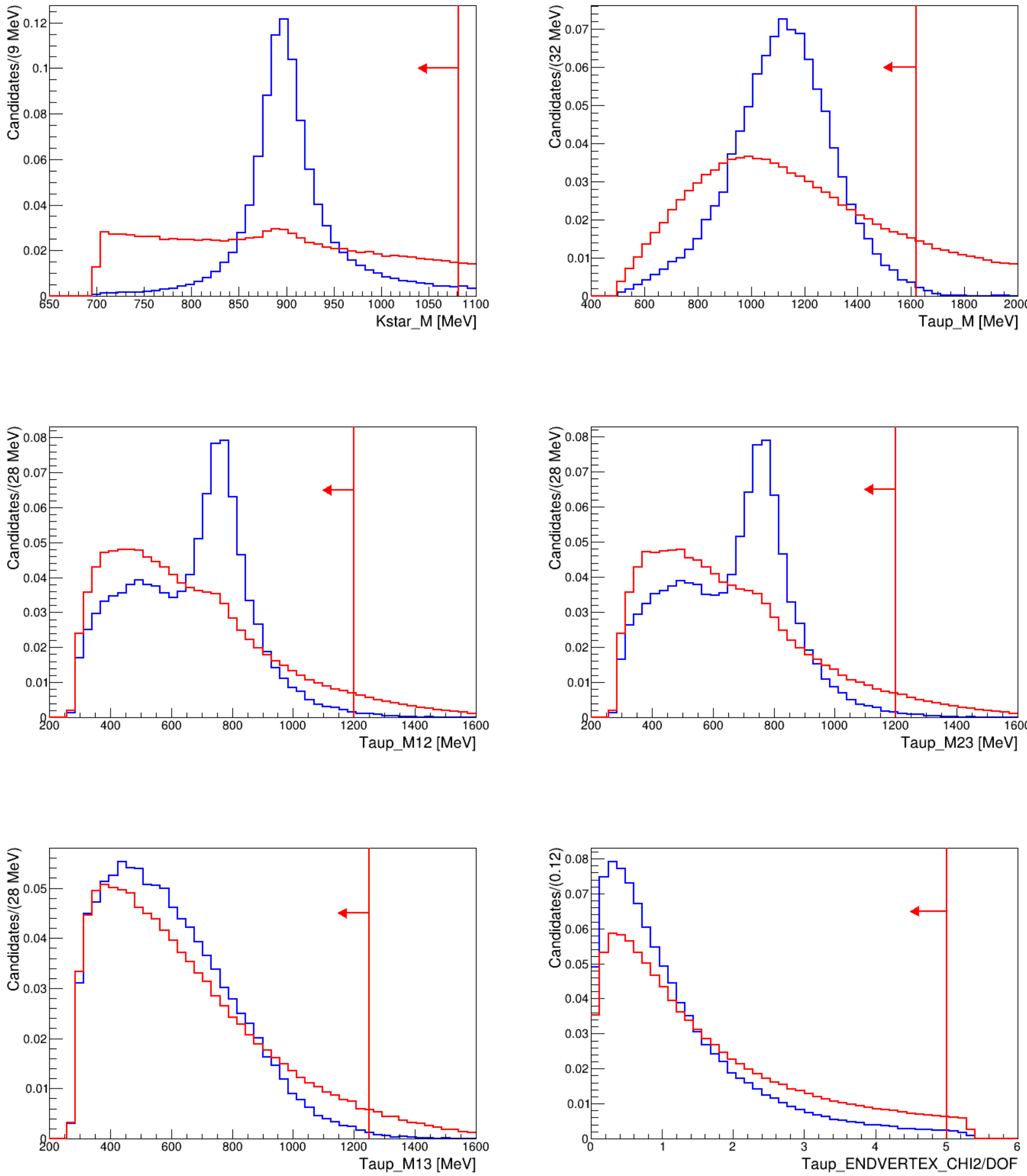


Figure 33: A comparison between signal MC (blue) and data (red), before the offline pre-selections are applied, in the variables pertaining to the pre-selections. The vertical red lines in each plot show the place where the cut is made, with the arrows showing the direction of the cut.

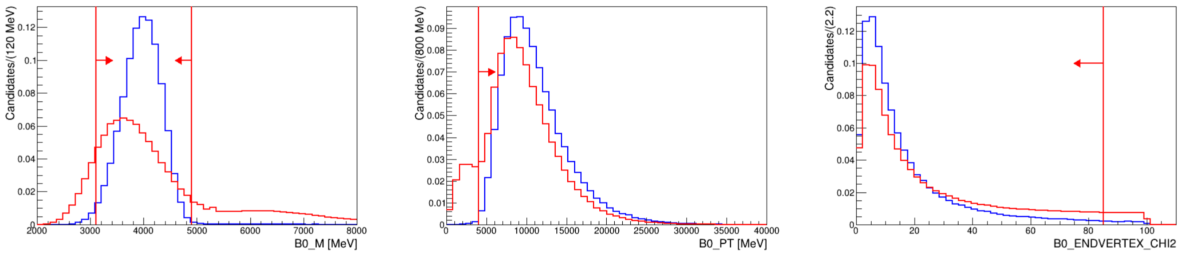


Figure 34: A comparison between signal MC (blue) and data (red), before the offline pre-selections are applied, in the variables pertaining to the pre-selections. The vertical red lines in each plot show the place where the cut is made, with the arrows showing the direction of the cut.

decays that have extra tracks as part of their decay tree. We use truth matched signal MC as the signal training sample. We employ two different approaches (which we shall denote as trainMethods 1 and 2) with regard to a choice of background training sample. The traditional approach (trainMethod1) is to use a background proxy sample that is disjoint from the data sample of interest. In this vein, the background training sample is the RS data in the upper sideband of the B DTF mass, in the region from $6.5 - 7\,\text{GeV}$. As an alternative approach (trainMethod2), we use as background training sample a subset of the actual RS data in the final fit range of $4.7 - 6.3\,\text{GeV}$. This approach has the benefit of providing the actual backgrounds that will be present in the data, and in the correct proportions. Since the signal of interest is exceedingly rare, its contamination in this background sample will be negligible, even after the predicted enhancements. To check for effects of signal bias in the background sample for trainMethod2, we train a crosscheck isolation BDT where a fraction of the signal sample is added to the background sample from trainMethod2. This background sample is used to train a the crosscheck BDT, whose response is compared with the response of the trainMethod2 isolation BDT to check for bias. This study is shown in Sec.D.

The procedure of training the BDT normally involves splitting the signal and background samples into two sets, a training set and a test set. The BDT is trained using the signal and background training samples, and its performance is evaluated using the independent testing sample. This method of using independent data to judge the performance of the BDT is used to avoid overfitting the BDT to the training samples. This usually happens when the BDT is so overtuned on the training sample that it picks up on statistical fluctuations that are present in the training sample and interprets them as general features. In doing so, it would no longer generalize to new data and its performance as evaluated on the training sample alone would be biased.

In order to make the most use of the training samples that are available, a k-folding technique is employed in the training of the BDT, with $k = 3$. This method involves splitting the available signal and background samples into k equal folds. The splitting in our case is done by using a split expression

$$(\text{eventNumber} + \text{nCandidate})\%\text{k} \tag{34}$$

where eventNumber denotes the unique number assigned to each LHCb event (be it in data or simulation) nCandidate denotes the unique number assigned to a candidate in a given event (for example if there are 3 candidates in an event, they would be numbered 0 to 2), and % denotes the modulus operation. The result of the split expression evaluated on any given candidate will be an integer number: 0, 1 or 2 (since $k = 3$). Following this, k classifiers are trained, with each using $k - 1$ of the folds as the training dataset, and the remaining one fold as the testing dataset.

The features used to train the isolation BDTs (one for trainMethod1 and one for trainMethod2) and their rank (i.e. importance in the BDT training) are shown in Table 10. The distributions of the training features for the signal and background samples are shown in Figs 62, 63 and 64 for trainMethod1, and in Figs 65, 66 and 67. The response of fold 1 of both the isolation BDTs is shown in Fig 35 for the training and testing sets of the signal and background samples. The compatibility of the responses for the training and testing sets demonstrates the absence of overtraining. The responses of the other folds are similar. We do not cut on the response of the isolation BDT right away. The selection on the isolation

Rank (trainMethod 2)	Feature	Rank (trainMethod 1)
1	τ^+ log(VtxIsoDeltaChi2TwoTrk)	2
2	τ^+ NC IT	1
3	τ^- NC IT	3
4	τ^- log(VtxIsoDeltaChi2TwoTrk)	5
5	B^0 NC IT	10
6	τ^- log(VtxIsoDeltaChi2OneTrk)	4
7	τ^- CC IT	11
8	K^{*0} log(VtxIsoDeltaChi2TwoTrk)	7
9	τ^+ CC IT	6
10	τ^+ log(VtxIsoDeltaChi2OneTrk)	9
11	K^{*0} log(VtxIsoDeltaChi2OneTrk)	8
12	B^0 CC IT	14
13	τ^+ IsoBDTSecondValue	13
14	K^{*0} IsoBDTThirdValue	12
15	τ^- IsoBDTSecondValue	15
16	τ^+ IsoBDTThirdValue	16
17	τ^- IsoBDTThirdValue	17

Table 10: Input features used to train the isolation BDT, and their ranking in terms of importance in the BDT selection. Currently, when TMVA trains BDTs with k-folding, it does not provide the ranking of the input features. The rankings shown in this table are obtained from a separate training without k-folding. Sections 4.5.4.1, 4.5.4.2, 4.5.4.3 provide the definitions of these variables. Sec. B provides the distributions of these variables for the signal and background samples.

BDT is optimized using toy studies. The choice between the two training methods is made based on which gives better sensitivity.

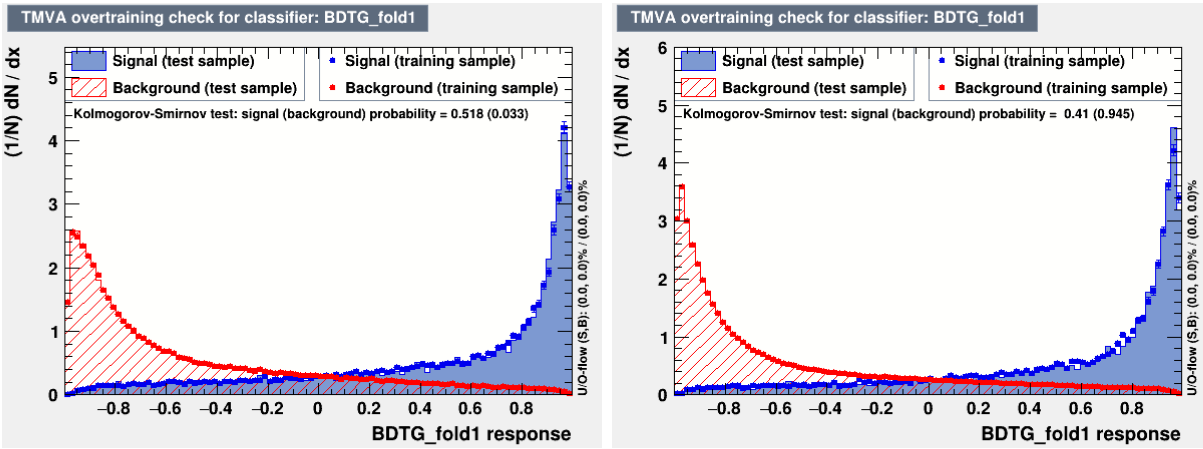


Figure 35: Response of fold 1 of the isolation BDT are shown for trainMethod1 (left) and trainMethod2 (right). The responses of the other folds are similar.

4.5.4.1 Track based isolation

Three custom track isolation variables are produced. These variables were initially developed for the $B_s^0 \to \mu^+\mu^-$ analysis [74] and then re-optimized for the $B_s^0 \to \tau^+\tau^-$ analysis. They were not specifically re-optimized for the $B^0 \to K^*\tau^+\tau^-$ analysis. These variables are derived based on the output of a BDT, trained on an inclusive $b\bar{b}$ sample to select $B_s^0 \to \tau^+\tau^-$ candidates. The values of these variables are outputted for every track of a reconstructed $B^0 \to K^*\tau^+\tau^-$ candidate, whether it is a final state charged track, or a reconstructed one, like a K^* or τ. The training of the BDT is done as follows: For a given candidate track in the event, all the other tracks (referred to as non-signal tracks) are divided into two categories, those coming from displaced B and D vertices that are part of the same true decay chain as the signal (referred to as **non-isolating tracks**), and all other tracks (referred to as **isolating tracks**) which comes either from the PV or other B and D decays in the same event. Isolating tracks are tracks which are unrelated to the signal track under consideration; they just happen to be in the same event.

The BDT is trained to separate isolating tracks (considered as the signal class) from non-isolating tracks (considered as the background class). This BDT is then applied to every long track in the event which is not a part of the signal candidate's decay chain and its response on that track is calculated. The more isolating tracks we have, the better, since that means less activity near the signal candidate.

The input variables used to train the BDT are as follows:

- The minimum χ^2_{IP} of the non-signal track with respect to any PV
- The p_{T} of the non-signal track
- The angle between the non-signal track and the signal track
- The parameter f_c defined as

$$f_c = \frac{|p_S + p_{tr}| \times \alpha}{|p_S + p_{tr}| \times \alpha + p_{T,tr}}$$

  where p_S and p_{tr} are the momenta of the signal track and non-signal tracks, $p_{T,tr}$ is the p_{T} of the non-signal track and α is the angle between the (signal + non-signal track) momentum and the vector pointing from the PV to the point $\mathcal{V}$, defined as the midpoint between the signal and non-signal tracks at their point of closest approach. A rough illustration of this is provided in Figure 36.

- The distance of closest approach between the non-signal track and signal track
- The distance between the point $\mathcal{V}$ and the B^0 decay vertex
- The distance between the point $\mathcal{V}$ and the PV.

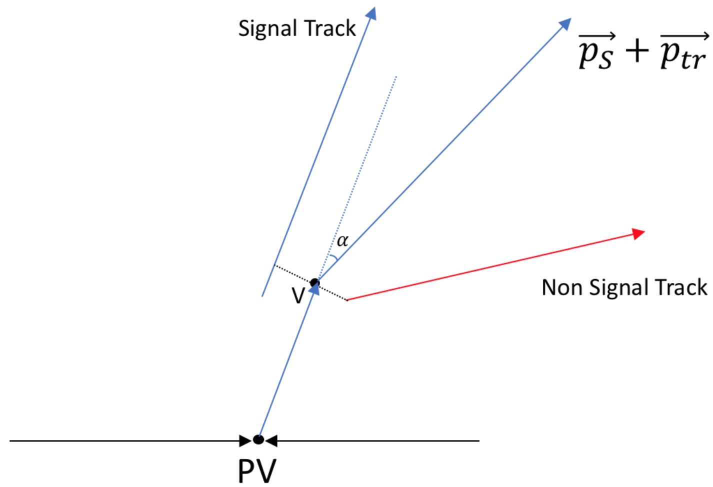


Figure 36: An illustration of the construction of the angle α.

In order to capture the "global" isolation in the event for a given signal track, a, b and c are defined as the number of non-signal long tracks in the event with BDT values less than -0.09, -0.05 and 0, respectively. The higher these numbers are, the more background like our candidate is. The three track isolation variables are then defined as:

`IsoBDTFirstValue`: This is set to $1000 \times c + 100 \times b + a$. Let us say we had 1 track with BDT less than -0.09, 2 with BDT less than -0.05 and 3 with BDT less than 0. Then `IsoBDTFirstValue` would be 3201. The higher this value is, the more background like our candidate track is.

`IsoBDTSecondValue`: This is defined as the sum of the BDT outputs for all tracks with BDT outputs less than -0.05. The lower this value is, the more background like our candidate track is.

`IsoBDTThirdValue`: This is defined as the sum of `IsoBDTSecondValue` and the minimum BDT value of all the tracks in the event. The lower this value is, the more background like our candidate track is.

4.5.4.2 Vertex based isolation

Five vertex isolation variables are produced each for the K^*, τ^+, τ^- and B^0 vertices. The vertex under consideration is referred to as $\mathcal{V}$. This vertex is combined with a single other track in the event to produce a new vertex $\mathcal{V}^*$. The vertex based isolation variables are then defined as:

`VtxIsoNumVtx`: the number of tracks in the event for which $\chi^2_{\mathcal{V}^*} < 9$.

`VtxIsoDeltaChi2OneTrk`: the smallest difference in χ^2_{vtx} between $\mathcal{V}$ and $\mathcal{V}^*$ for all the tracks in the event.

`VtxIsoDeltaChi2TwoTrk`: the smallest difference in χ^2_{vtx} between $\mathcal{V}$ and $\mathcal{V}^{**}$, which is the vertex constructed iteratively between the $\mathcal{V}^*$ with the smallest $\Delta\chi^2_{\mathrm{vtx}}$ from $\mathcal{V}$, and the other tracks in the event.

`VtxIsoDeltaChi2MassOneTrk`: the invariant mass of the tracks used to form the `VtxIsoDeltaChi2OneTrk` variable.

`VtxIsoDeltaChi2MassTwoTrk`: the invariant mass of the tracks used to form the `VtxIsoDeltaChi2TwoTrk` variable.

4.5.4.3 Cone based isolation

The cone based isolation constructs a cone of size 0.5 in both η and ϕ around the head particle (which could be a B^0 or a τ^+ or τ^-). Charged cone (`CC`) variables concern themselves with the charged tracks present in this cone, while neutral cone (`NC`) variables pertain to the neutral objects in this cone. [2] The variables are defined as follows:

`DELTAETA, DELTAPHI`: The $\Delta\eta$ and $\Delta\phi$ of the vector sum momentum of all long tracks in the cone w.r.t. the momentum of the head of the cone.

`MULT`: The multiplicity of the long tracks in the cone.

[2] The `CC` variables use tracks from the container `StdAllNoPIDsMuons`, while the `NC` variables use the container `StdLoosePhotons`.

PASYM: $\frac{p_{\text{head}}-p_{\text{cone}}}{p_{\text{head}}+p_{\text{cone}}}$, where p_{cone} refers to the 3 momentum sum of the contents of the cone.

PTASYM: $\frac{p_{T,\text{head}}-p_{T,\text{cone}}}{p_{T,\text{head}}+p_{T,\text{cone}}}$, where $p_{T,\text{cone}}$ refers to the total p_{T} of the contents of the cone.

PXASYM: $\frac{p_{x,\text{head}}-p_{x,\text{cone}}}{p_{x,\text{head}}+p_{x,\text{cone}}}$. PYASYM and PZASYM are defined analogously.

SPT: Scalar p_{T} sum of the contents of the cone.

VPT: Vector p_{T} sum of the contents of the cone.

IT: $\frac{p_{T,\text{head}}}{\sqrt{(p_{x,\text{head}}+p_{x,\text{cone}})^2+(p_{y,\text{head}}+p_{y,\text{cone}})^2}}$

4.5.5 Kinematic BDT

We now train a BDT using kinematic information (referred to as the kinematic BDT). The training methodology for the kinematic BDT follows that of the isolation BDT closely. The strategy of using two parallel training methods as well as k-folding is copied. In trainMethod 1, the WS data is used as a background sample because it does a good job of representing the distributions of the variables of interest. In trainMethod 2, we use the same background sample as was used for the isolation BDT i.e. a subset of the RS data in the final fit range of $4.7-6.3\,\text{GeV}$. In both methods, the truth matched signal MC is used as the signal training sample.

The input features used to train the kinematic BDT and their importances in the training are listed in Table 11. Here, we explain the meaning of the variables listed in the table. M_{max} and M_{min} for a given τ refer to the maximum and minimum of M_{12} and M_{23} for the pions coming from that τ [3]. These variables are defined to create a symmetrized version of M_{12} and M_{23}, since it is somewhat arbitrary as to which of the same sign pions for a τ gets called 1 and which is called 3. Similarly, it can be seen that the definitions of the features pertaining to the τ's are also symmetrized between τ^+ and τ^-.

In the RS data, there is a clear definition of what we consider to be a τ^+: it is the 3π combination that is produced positively charged along with a K^{*0} (and negatively charged

[3] M_{ij} for a $\tau^+ \to \pi_1^+\pi_2^-\pi_3^+X$ candidate denotes the invariant mass of the combination $\pi_i\pi_j$

with a $\overline{K}^{*0}$, charged conjugation in the decay is implied throughout). But in the WS data, we reconstruct $K^{*0}\tau^+\tau^+$ and $K^{*0}\tau^-\tau^-$ together, so the definition of τ^+ and τ^- becomes arbitrary in the WS data. Hence, the symmetrization of the training feature definitions between τ^+ and τ^- variables. The τ DTF DL is the decay length of the tau (measured from the origin vertex to the decay vertex) as measured by the DTF fit. $3\pi\ |\vec{p}|\ \perp \tau$ dir is the component of the 3π momentum that is perpendicular to the known τ flight direction, from the K^{*0} vertex to the 3π vertex. τ DTF ν p_{T} is the momentum of the neutrino particle added to the τ, as calculated by the DTF fit. The $\tau^+ - \tau^-$ vertex separation χ^2 is defined as

$$\chi^2 = \frac{\tau^+\tau^-\text{dist}}{|\tau_X^+ - \tau_X^-|\cdot\sqrt{\tau_{\Delta X}^{+2} + \tau_{\Delta X}^{-2}} + |\tau_Y^+ - \tau_Y^-|\cdot\sqrt{\tau_{\Delta Y}^{+2} + \tau_{\Delta Y}^{-2}} + |\tau_Z^+ - \tau_Z^-|\cdot\sqrt{\tau_{\Delta Z}^{+2} + \tau_{\Delta Z}^{-2}}} \tag{35}$$

It is not a real chi-square variable since it ignores the correlations between the X, Y and Z coordinates. τ_X^+ represents the X coordinate of the decay vertex of the τ^+ and $\tau_{\Delta X}$ represents the error on this coordinate. K^+ ProbNNK is a PID (particle identification) variable pertaining to the K^+ from $K^{*0} \to K^+\pi^-$. It is the output of a neural network trained to distinguish charged kaon tracks from other species such as pions, protons, or muons. The response of the PID variables in LHCb simulation is known to be inaccurate, and a correction procedure is applied using the `PIDCalib` package to all the MC samples in order to account for this. Fig.37 shows the comparison between the original K^+ ProbNNK response in simulation with the `PIDCalib` corrected response.

As mentioned in the analysis strategy, we blind the mass distribution of the data in the most signal rich region. In practical terms, we choose to do this by blinding the kinematic BDT region which contains the top 30% of signal. We use the kinematic BDT to determine our blinding region as opposed to the isolation BDT because the kinematic BDT is more sensitive to signal, as can be seen in the BDT responses. For trainMethod1, the kinematic BDT region $0.98 - 1.00$ contains the top 30.4% of the signal. For trainMethod2, the kinematic

Rank (trainMethod 2)	Feature	Rank (trainMethod 1)
1	$m_{K^{*0}}$	1
2	$m_{\tau^+\tau^-}$	2
3	$\tau^+ - \tau^-$ vertex separation χ^2	4
4	$\min(\tau^\pm\ M_{\max})$	5
5	$K^{*0}\ \log(\chi^2_{\mathrm{IP,PV}})$	6
6	$\max(\tau^\pm\ M_{\max})$	3
7	$\min(\tau^\pm\ \log(\chi^2_{\mathrm{FD,PV}}))$	11
8	$\max(m_{\tau^\pm})$	9
9	$\log(\mathrm{DTF}\ \chi^2)$	10
10	K^+ ProbNNK	7
11	$\max(\tau^\pm\ M_{13})$	18
12	$\min(\tau^\pm\ \mathrm{DTF\ DL})$	8
13	$\max(\tau^\pm\ M_{\min})$	13
14	$\min(\tau^\pm\ p_{\mathrm{T}})$	14
15	$K^{*0}\ p_{\mathrm{T}}$	12
16	$\max(\log(\lvert\vec{p}_{3\pi}\rvert_{\perp\tau^\pm}))$	20
17	$K^{*0}\ K\pi\ \log(\min\ \chi^2_{\mathrm{IP}}) = \min(K\ \chi^2_{\mathrm{IP}}, \pi\ \chi^2_{\mathrm{IP}})$	24
18	B^0 DTF MERR	19
19	$B^0\ \mathrm{FD_{PV}}$	16
20	$\min(\tau^\pm\ M_{\min})$	21
21	$B^0\ \chi^2_{\mathrm{vtx}}$	15
22	$\min(\tau^\pm\ \chi^2_{\mathrm{vtx}}/\mathrm{dof})$	22
23	$\min(\log(\lvert\vec{p}_{3\pi}\rvert_{\perp\tau^\pm}))$	23
24	$\min(m_{\tau^\pm})$	17
25	$\max\ \log(\tau^\pm\ \mathrm{DTF}\ \nu\ p_{\mathrm{T}})$	25
26	$\min\ \tau\ \pi\ \log(\chi^2_{\mathrm{IP}}) = \min(\chi^2_{\mathrm{IP}}$ of all 6π from $\tau)$	26
27	$\min(\tau^\pm\ M_{13})$	27
28	$\min\ \log(\tau^\pm\ \mathrm{DTF}\ \nu\ p_{\mathrm{T}})$	28
29	$K^{*0}\ \mathrm{FD_{ORIVX}}$	29

Table 11: Input features used to train the kinematic BDT, and their ranks for trainMethod1 and trainMethod 2. Currently, when TMVA trains BDTs with k-folding, it does not provide the ranking of the input features. The rankings shown in this table are obtained from a separate training without k-folding. Sec.C provides the distributions of the input features

BDT region $0.97 - 1.00$ contains the top 33.4% of the signal. These are the regions we choose to blind.

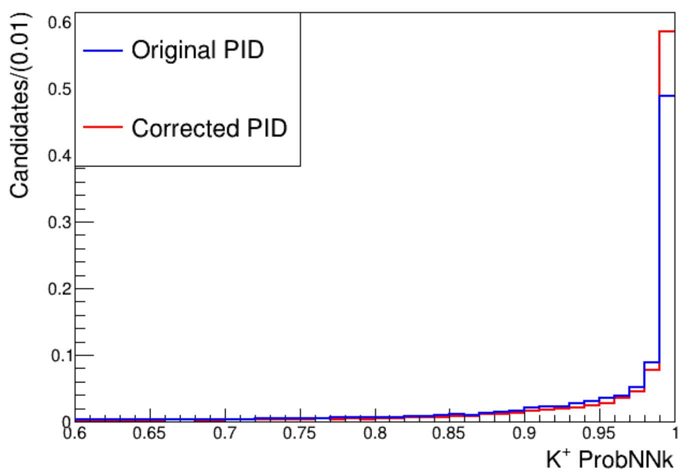


Figure 37: Comparison of original K^+ ProbNNk response in signal simulation with the `PIDCalib` corrected response.

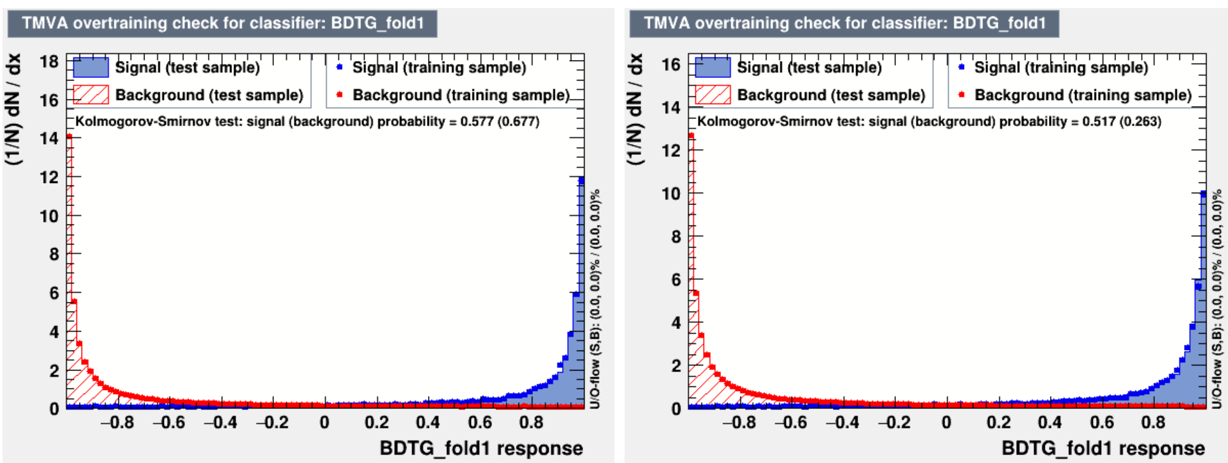


Figure 38: Response of the kinematic BDT for trainMethod1 (left) and trainMethod2 (right). Responses of the other folds are similar.

4.5.6 Best candidate selection

After the trigger and loose preselections are applied to data, and before the BDT selections are applied, there are approximately 5 candiates per event in data. This is largely a consequence of the large number of final state charged tracks, most of which are pions. Hence, one candidate in data can be misreconstructed in a large number of ways, by interchanging pions between the two τ's and interchanging pions between the K^* and one of the τ's.

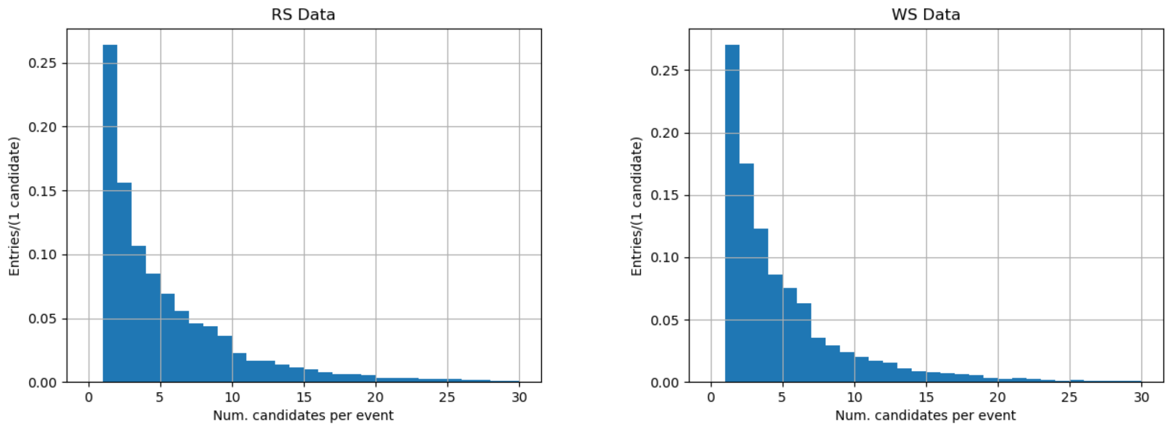


Figure 39: Distribution of candidates per event in RS data (left) and WS data (right) after the application of the offline pre-selections, and before any BDT cut. The RS data has 5.1 candidates per event on average, and the WS data has 4.6 candidates per event on average.

In signal MC too, we see around 5 candidates per event on average after the trigger selections. We can of course, pick out the correctly reconstructed signal candidate, through our truth matching selections.

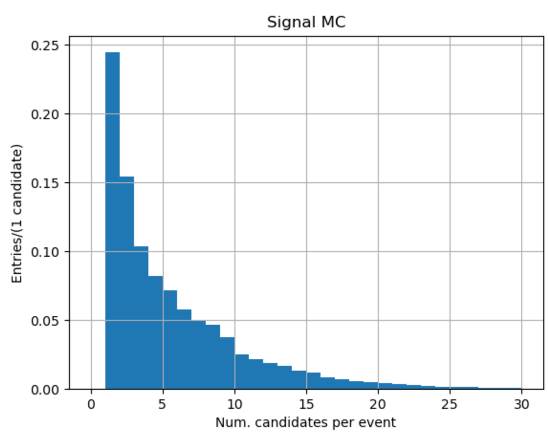


Figure 40: Distribution of candidates per event in signal MC (without any truth matching applied) after the application of the offline pre-selections, and before any BDT cut. The signal simulation has 5.2 candidates per event on average.

After applying the best candidate selection to the RS data, approximately 10 million candidates survive (compared to approximately 52 million before). The expected signal yield at a branching fraction of 1×10^{-3} is 633 ± 34 after applying the best candidate selection.

4.5.7 Check for clone tracks

During the reconstruction of tracks in LHCb, it is possible to have more than one track reconstructed for the same set of hits. These tracks are called clones in LHCb terminology. Most of these are eliminated by clone killing algorithms in the reconstruction phase, but some might survive in our data sample, especially since we have a large number of hadronic final state tracks. We investigate their presence by plotting the minimum pair angle out of all 28 pairs (8 choose 2) of final state tracks in our data events. If two tracks are clones of each other, the angle between them will be very close to zero. The presence of clones in the data should then show up as a peak near zero in this distribution of minimum pair angles. The distribution in a very small subset of data after best candidate selection, before the application of any BDT cuts is shown in Fig.41. We clearly see that there are indeed instances of events with clone tracks. Before deciding whether any explict action needs to be taken to eliminate such instances, we plot the analogous distribution with tight BDT cuts applied (kinBDT > 0.994, isoBDT > 0.97)

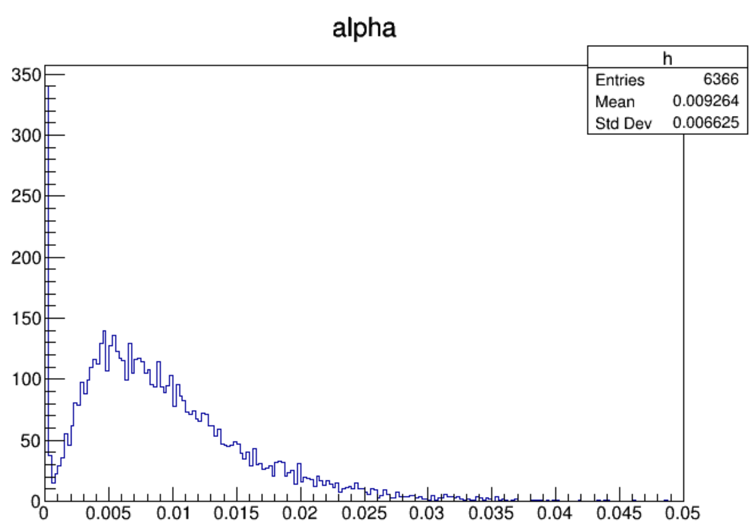


Figure 41: Clone track check: Distribution of minimum pair angle out of all 28 pairs of final state tracks is shown in a small subset of data, after best candidate selection, before any tight BDT cuts. From the peak at zero, we deduce that events with clone tracks are indeed present in the data at this stage.

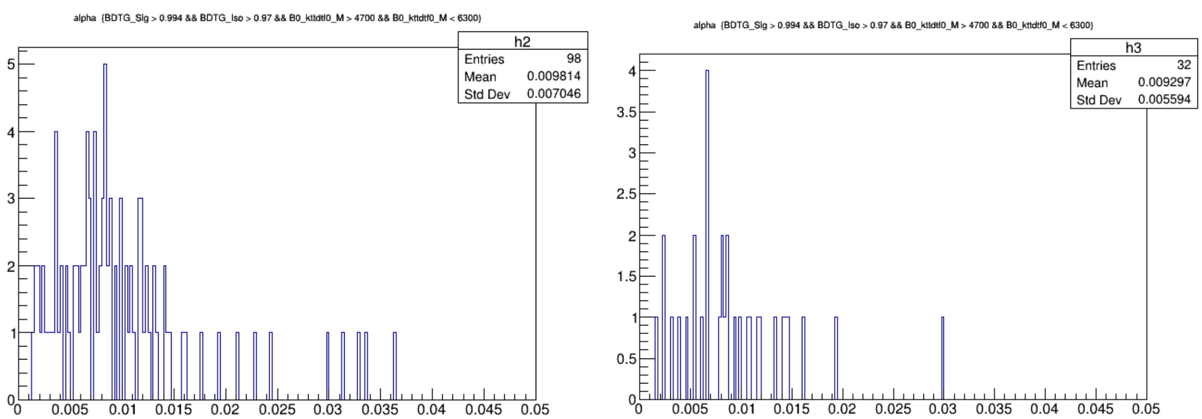


Figure 42: Clone track check: Distribution of minimum pair angle out of all 28 pairs of final state tracks is shown in a small subset of data after tight BDT cuts (kinBDT > 0.994, isoBDT > 0.97), for trainMethod1 (left) and trainMethod2(right). At this stage, we see no peak near zero, and conclude the absence of clone track events after tight BDT cuts.

4.6 Efficiencies

In this section, we detail the signal efficiencies of our various selection stages, as determined using the signal MC simulation sample. In order to reliably determine efficiencies from the signal MC sample, we must first apply a procedure known as "truth matching" to the MC sample, whereby we eliminate mis-reconstructed signal. Internally this is done by matching the detector hits of the reconstructed tracks to those left by the generated tracks. A pair of tracks is considered successfully matched if they share more than a certain threshold of hits. The truth matching follows a bottom up approach. Suppose we have a reconstructed $X \to 3\pi$ candidate where the 3 final state pions have been correctly matched to generated pion tracks, and X is some intermediate particle. X is then matched to the first common mother of the three generated tracks that the reconstructed tracks have been matched to.

We define our truth matching condition on signal as the logical OR between a tight condition where all the tracks and intermediates are correctly matched (`tightMatch`), and a condition that allows one and only one final state to fail its truth matching (`onlyOneFail`). The latter category is defined to account for the cases where a correctly reconstructed track fails the matching for some reason (one reason could be a decay in flight of a pion to a muon, or the number of shared hits being below the required threshold). The `tightMatch` category is defined by requiring all the final state tracks, the three intermediates (two τ's and one K^*) and the B^0 mother to be correctly matched. The `oneFail` category is defined as the logical OR of eight categories, each of which allow one and only one final state track (and its corresponding intermediate) to fail the matching but require the rest of the decay tree to be correctly matched. The B^0 mother will always fail the matching in the `oneFail` cases since one of the three intermediates will also fail.

Stage	Efficiency
Generator	$(2.56 \pm 0.01)\%$
Stripping	$(1.61 \pm 0.01)\%$
Reconstruction	$(5.42 \pm 0.01)\%$
L0 Trigger	$(50.43 \pm 0.19)\%$
HLT1 Trigger	$(98.97 \pm 0.43)\%$
HLT2 Trigger	$(66.25 \pm 0.32)\%$
Pre-selection	$(93.19 \pm 0.51)\%$
DTF Pass	$(63.89 \pm 0.41)\%$
Best Candidate Selection (trainMethods 1 and 2)	$(72.12 \pm 0.55)\%$
Total (trainMethods 1 and 2)	$(3.17 \pm 0.03) \times 10^{-6}$

Table 12: Signal efficiencies before BDT selections. Except for the Generator and Stripping efficiencies, all the efficiencies are quoted relative to the previous stage of selection. The uncertainties quoted are statistical only.

Stage	Efficiency
Generator	$(1.76 \pm 0.02)\%$
Stripping	$(1.53 \pm 0.01)\%$
Reconstruction	$(3.60 \pm 0.02)\%$
L0 Trigger	$(49.75 \pm 0.39)\%$
HLT1 Trigger	$(97.78 \pm 0.89)\%$
HLT2 Trigger	$(37.57 \pm 0.46)\%$
Pre-selection	$(76.18 \pm 1.22)\%$
DTF Pass	$(68.15 \pm 1.29)\%$
Best Candidate Selection (trainMethods 1 and 2)	$(69.65 \pm 1.59)\%$
Total (trainMethods 1 and 2)	$(6.38 \pm 0.14) \times 10^{-7}$

Table 13: $B^0 \to DDK^*$ background MC efficiencies before BDT selections. Except for the Generator and Stripping efficiencies, all the efficiencies are quoted relative to the previous stage of selection. The uncertainties quoted are statistical only.

Stage	Efficiency
Generator	$(1.41 \pm 0.02)\%$
Stripping	$(1.56 \pm 0.01)\%$
Reconstruction	$(2.78 \pm 0.01)\%$
L0 Trigger	$(49.67 \pm 0.43)\%$
HLT1 Trigger	$(97.62 \pm 0.98)\%$
HLT2 Trigger	$(38.34 \pm 0.52)\%$
Pre-selection	$(81.23 \pm 1.41)\%$
DTF Pass	$(68.73 \pm 1.39)\%$
Best Candidate Selection (trainMethod 1)	$(66.87 \pm 1.64)\%$
Best Candidate Selection (trainMethod 2)	$(67.47 \pm 1.65)\%$
Total (trainMethod 1)	$(4.24 \pm 0.11) \times 10^{-7}$
Total (trainMethod 2)	$(4.29 \pm 0.11) \times 10^{-7}$

Table 14: $B^+ \to DDK^*$ background MC efficiencies before BDT selections. Except for the Generator and Stripping efficiencies, all the efficiencies are quoted relative to the previous stage of selection. The uncertainties quoted are statistical only.

Stage	Efficiency
Generator	$(2.41 \pm 0.03)\%$
Stripping	$(2.66 \pm 0.01)\%$
Reconstruction	$(3.31 \pm 0.01)\%$
L0 Trigger	$(49.02 \pm 0.32)\%$
HLT1 Trigger	$(97.22 \pm 0.75)\%$
HLT2 Trigger	$(37.37 \pm 0.39)\%$
Pre-selection	$(66.42 \pm 0.94)\%$
DTF Pass	$(69.14 \pm 1.19)\%$
Best Candidate Selection (trainMethod 1)	$(57.11 \pm 1.26)\%$
Best Candidate Selection (trainMethod 2)	$(56.43 \pm 1.25)\%$
Total (trainMethod 1)	$(9.88 \pm 0.21) \times 10^{-7}$
Total (trainMethod 2)	$(9.76 \pm 0.20) \times 10^{-7}$

Table 15: $B_s^0 \to DDK^*$ background MC efficiencies before BDT selections. Except for the Generator and Stripping efficiencies, all the efficiencies are quoted relative to the previous stage of selection. The uncertainties quoted are statistical only.

The truth matching scheme defined in this section is only used to provide the efficiency breakup given in Tables 12, 13, 14 and 15. When we fit to the data after tight BDT cuts, we

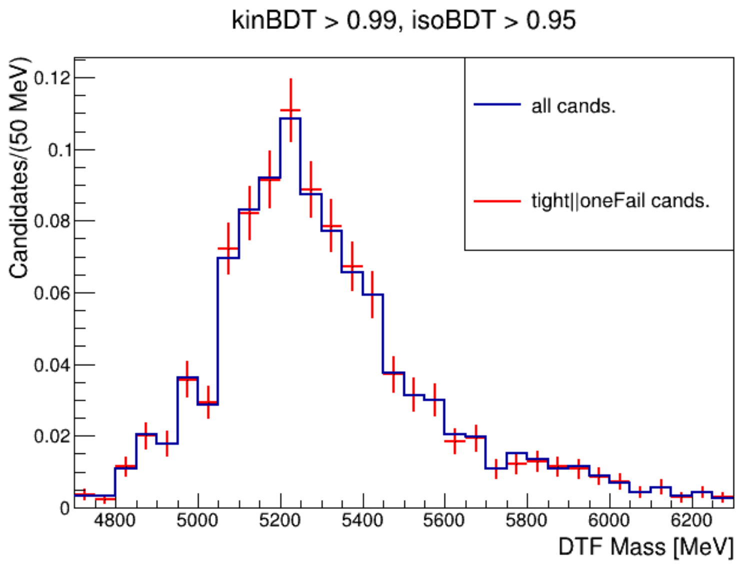


Figure 43: A comparison of the DTF mass shape in signal simulation, after the best candidate selection and tight BDT cuts (kinBDT $>$ 0.99, isoBDT $>$ 0.95, (trainMethod2)), with (red) and without (blue) truth matching applied.

do not apply truth matching to obtain the signal efficiencies and fit shape fed to the fitter. This is done based on the identical DTF mass shapes in signal after applying reasonably tight BDT cuts, with and without truth matching, as demonstrated in Fig.43. Since any mis-reconstruction happening in the simulation will also happen in data, and signal is discerned in data only from a peaking mass distribution, peaking candidates in signal must be counted in the efficiency and fit template, regardless of whether they pass the truth matching procedure (which by itself is not perfect).

4.7 Measurement of $B^0 \to D^- D^0 K^+$

This section describes the measurement of the decay $B^0 \to D^- D^0 K^+, D^- \to K^+\pi^-\pi^-, D^0 \to K^- 2\pi^+\pi^-$. The material of this section is adapted from an LHCb internal analysis note written by Harris Bernstein and Matthew Rudolph, who carried out the measurement. Only an overview is presented here, to explain how the normalization input for $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-)$ is measured.

The measurement utilizes data collected by the LHCb detector in the Run 2 years of 2016, 2017 and 2018, corresponding to the same data taking time period as the $K^*\tau^+\tau^-$ measurement. The trigger lines and trigger selections used in this measurement are the same as that used in the $K^*\tau^+\tau^-$ measurement, as described in Sec.4.5.1. This measurement uses data passing the `B02D0DKD02K3PiBeauty2CharmLine` stripping line. The data is separated into 6 disjoint samples based on the data taking year and the trigger categories `TOS` and `TISnotTOS` that were defined in Sec. 4.5.1. In the interest of brevity and clarity, the details of the offline selections applied on the data samples are not presented here, but rather we show the final mass fits, which are done separately in the 6 disjoint data samples.

Unbinned maximum likehood fits are performed to the DTF B mass (with mass constraints applied to the D meson masses, as well as a pointing constraint applied to the B^0 to ensure that it points back at the PV). The signal shape in the fit is described as the sum of two Gaussian distributions with a shared mean. The combinatorial background shape in the fit is modeled as an exponential function.

The fitted yields in each of the 6 categories, along with the corresponding signal efficiency and efficiency corrected yield are shown in Table 16.

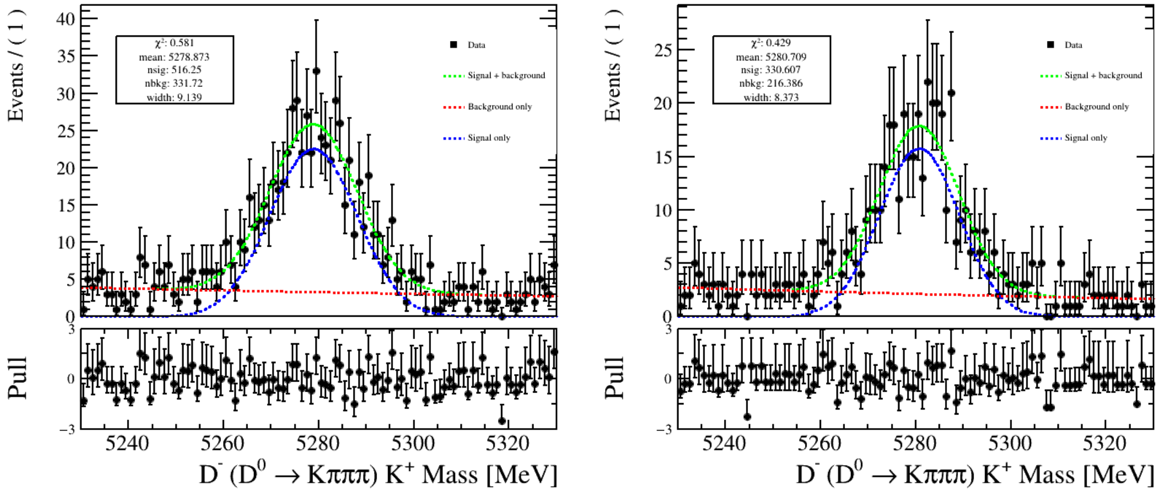


Figure 44: Fit to the $D^-D^0K^+$ mass spectrum for 2016 `TOS` data (left) and `TISnotTOS` data (right). The $B^0 \to D^-D^0K^+$ decay is clearly observed. The signal shape is shown as a dashed blue line, and the background shape as a dashed red line. The total fit is shown as a dashed green line.

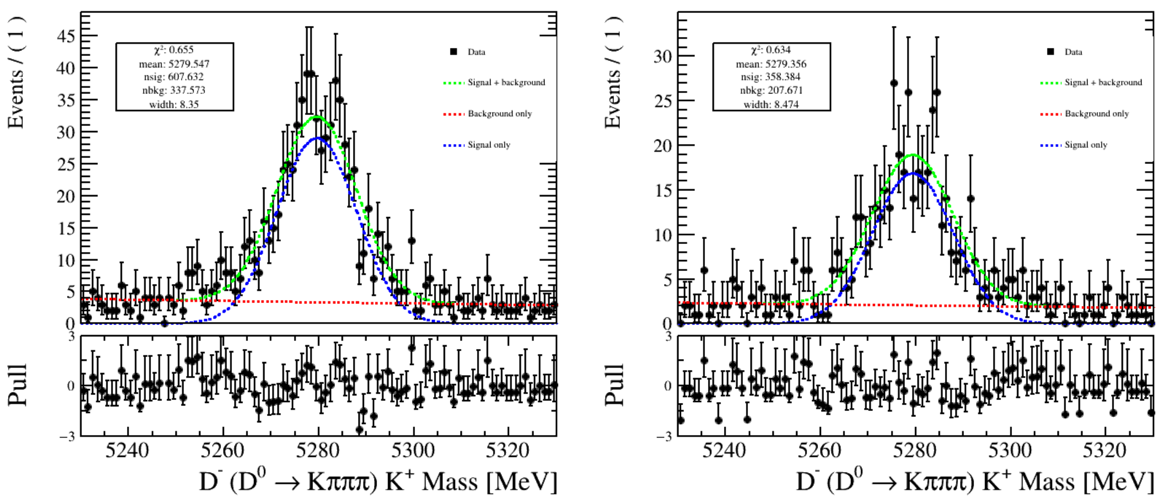


Figure 45: Fit to the $D^-D^0K^+$ mass spectrum for 2017 `TOS` data (left) and `TISnotTOS` data (right). The $B^0 \to D^-D^0K^+$ decay is clearly observed. The signal shape is shown as a dashed blue line, and the background shape as a dashed red line. The total fit is shown as a dashed green line.

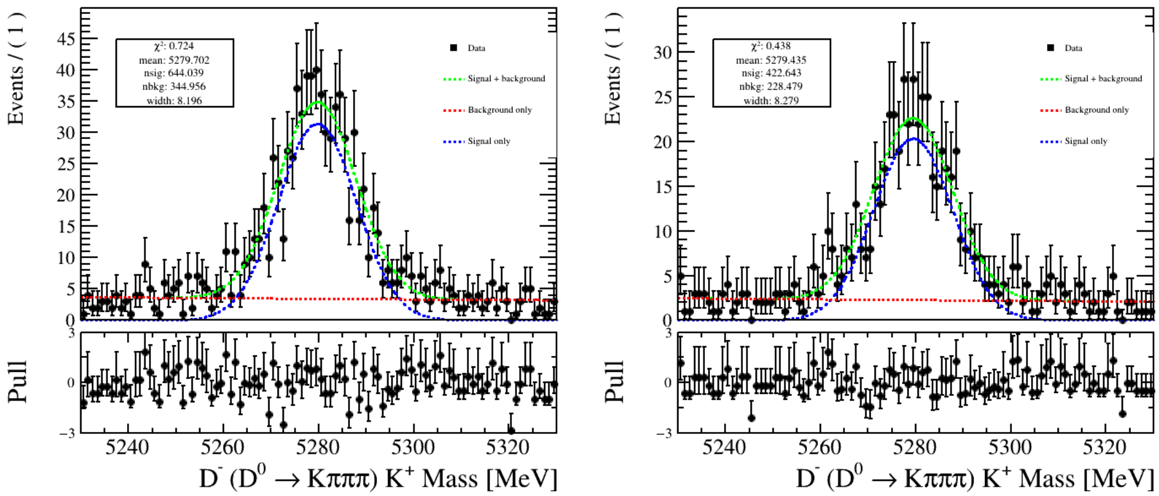


Figure 46: Fit to the $D^-D^0K^+$ mass spectrum for 2018 `TOS` data (left) and `TISnotTOS` data (right). The $B^0 \to D^-D^0K^+$ decay is clearly observed. The signal shape is shown as a dashed blue line, and the background shape as a dashed red line. The total fit is shown as a dashed green line.

Year	Trigger Category	Data Yield	Efficiency	Eff. corrected yield
2016	`TOS`	498 ± 28	$(1.29 \pm 0.05) \cdot 10^{-4}$	$(3.85 \pm 0.27) \cdot 10^{6}$
2016	`TISnotTOS`	327 ± 22	$(7.06 \pm 0.33) \cdot 10^{-5}$	$(4.63 \pm 0.38) \cdot 10^{6}$
2017	`TOS`	593 ± 31	$(1.47 \pm 0.06) \cdot 10^{-4}$	$(4.03 \pm 0.27) \cdot 10^{6}$
2017	`TISnotTOS`	358 ± 24	$(7.75 \pm 0.38) \cdot 10^{-5}$	$(4.62 \pm 0.38) \cdot 10^{6}$
2018	`TOS`	636 ± 31	$(1.21 \pm 0.06) \cdot 10^{-4}$	$(5.24 \pm 0.35) \cdot 10^{6}$
2018	`TISnotTOS`	406 ± 25	$(6.59 \pm 0.37) \cdot 10^{-5}$	$(6.16 \pm 0.52) \cdot 10^{6}$

Table 16: The fitted yields for $B^0 \to D^-D^0K^+$ in each of the 6 disjoint categories is shown, along with the corresponsing signal efficiencies and efficiency corrected yields.

4.8 Measurement of $b \to DDK^{*0}$ backgrounds

It is important in searches such as this analysis to have a good understanding of the physical backgrounds that must be accounted for. Physical backgrounds are backgrounds in the data that originate from real physics decays i.e. they are not combinatorial in nature. The physical backgrounds that will be the most dangerous in our search will be those of the form $b \to (\mathrm{Charm} \to 3\pi X)(\mathrm{Charm} \to 3\pi X)K^{*0}$, owing to their similarity to our signal. Decay modes of this form remain by and large unmeasured. Analogous modes where the K^{*0} in the final state are replaced by a K^0 or K^+ are measured with sizeable branching fractions of $O(0.1\% - 1\%)$. Therefore, it is quite possible that the K^{*0} version of the background modes are also significant.

This motivates a parallel analysis, conducted by Harris Bernstein and Matthew Rudolph, measuring the branching fraction of the background modes listed in Tables 17. In part of this chapter, a brief summary of the relevant aspects of the measurement is presented. The measurement utilizes data collected by the LHCb detector in the Run 2 years of 2016, 2017 and 2018, corresponding to the same data taking time period as the $K^*\tau^+\tau^-$ measurement. The trigger lines and trigger selections used in this measurement are the same as that used in the $K^*\tau^+\tau^-$ measurement, as described in Sec.4.5.1. In the interest of brevity, the details of the selections and background vetos applied in the measurement are not presented here.

The branching fractions of the 15 modes are measured by fitting to the invariant mass distributions of the following combinations:

- **Z**: $D^+D^-K^{*0}$
- **ZZ**: $\overline{D}^0D^0K^{*0}$
- **P** : $\overline{D}^0D^+K^{*0}$
- **M** : $D^-D^0K^{*0}$
- **Pst** : $\overline{D}^0D^{*+}K^{*0}$ (reconstructing soft pion from D^{*+})

Index	Decay Mode
1	$B^0 \to D^- D^+ K^{*0}$
2	$B^0 \to D^{*-} D^+ K^{*0}$
3	$B^0 \to D^- D^{*+} K^{*0}$
4	$B^0 \to D^{*-} D^{*+} K^{*0}$
5	$B^0 \to \overline{D}^0 D^0 K^{*0}$
6	$B^0 \to \overline{D}^{*0} D^0 K^{*0} + B^0 \to \overline{D}^0 D^{*0} K^{*0}$
7	$B^0 \to \overline{D}^{*0} D^{*0} K^{*0}$
8	$B^+ \to \overline{D}^0 D^+ K^{*0}$
9	$B^+ \to \overline{D}^{*0} D^+ K^{*0}$
10	$B^+ \to \overline{D}^0 D^{*+} K^{*0}$
11	$B^+ \to \overline{D}^{*0} D^{*+} K^{*0}$
12	$B_s^0 \to D_s^- D^+ K^{*0}$
13	$B_s^0 \to D_s^{*-} D^+ K^{*0}$
14	$B_s^0 \to D_s^- D^{*+} K^{*0}$
15	$B_s^0 \to D_s^{*-} D^{*+} K^{*0}$

Table 17: A list of the $b \to DDK^*$ modes that are measured in the parallel analysis. D^0 are reconstructed as $K^-\pi^+$, D^+ as $K^-\pi^+\pi^+$, D_s^+ as $K^+K^-\pi^+$, K^{*0} as $K^+\pi^-$.

- **Zs** : $D_s^- D^+ K^{*0}$

The **Z**, **ZZ**, **P**, **M** and **Pst** mass distributions are fit to simultaneously, while the **Zs** mass distribution is fit to separately. In each of these mass spectra, two or three peaks are observed in the data. For example, in the $D^+D^-K^{*0}$ spectrum three peaks are observed. The rightmost peak is corresponds to the fully reconstructed $B^0 \to D^+D^-K^{*0}$ decay. The middle peak is composed of $B^0 \to D^{*+}D^-K^{*0}$ and $B^0 \to D^+D^{*-}K^{*0}$, with $D^{*+} \to D^+\pi^0$ and the neutral pion not being reconstructed causing the peak to be shifted down from the nominal B^0 mass. The leftmost peak corresponds to the decay with two missing pions: $B^0 \to D^{*+}D^{*-}K^{*0}$. All six mass spectra can be seen in Fig. 47. It is important to note that a mass peak can have contributions from more than one decay, and that a decay can contribute to more than one mass peak, necessitating the simultaneous fit of the first five spectra. An example of the former is the lowest mass peak in the $\overline{D}^0D^0K^{*0}$ (**ZZ**) spectrum, which contains contributions from $B^0 \to (D^{*-} \to \overline{D}^0\pi^-)(D^{*+} \to D^0\pi^+)K^{*0}$, $B^0 \to (\overline{D}^{*0} \to \overline{D}^0(\pi^0/\gamma))(D^* \to D^0(\pi^0/\gamma))K^{*0}$

and $B^+ \to (\overline{D}^{*0} \to \overline{D}^0(\pi^0/\gamma))(D^{*+} \to D^0\pi^+)K^{*0}$. An example of the latter is the decay $B^0 \to (D^{*-} \to \overline{D}^0\pi^-)(D^{*+} \to D^0\pi^+)K^{*0}$ which contributes to the lowest mass peak in the $\overline{D}^0 D^0 K^{*0}$ (**ZZ**) spectrum as well as to the lower mass peak in the $\overline{D}^0 D^{*+} K^{*0}$ (**Pst**) spectrum.

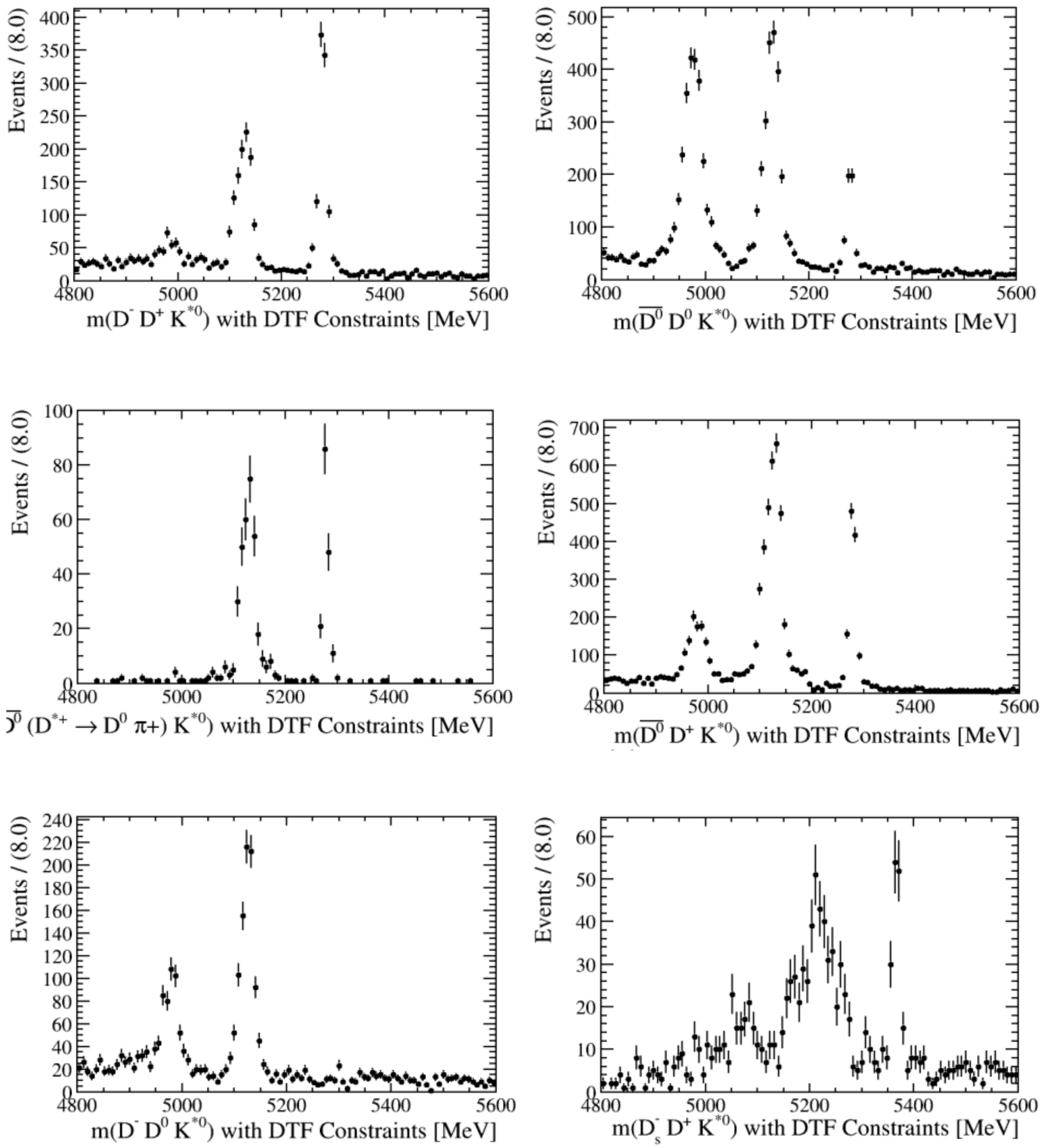


Figure 47: The mass spectra for $D^-D^+K^*$ (top left), $\overline{D}^0D^0K^*$ (top right), $\overline{D}^0D^{*+}K^*$ (middle left), $\overline{D}^0D^+K^*$ (middle right), $D^-D^0K^*$ (bottom left) and $D_s^+D^+K^*$ (bottom right) are shown.

Two known normalization modes are used in the measurement of the branching fractions: the 7 track $B^+ \to (\overline{D}^0 \to K^+\pi^-)(D^0 \to K^-\pi^+\pi^+\pi^-)K^{*0}$ (**norm7**) mode and the 8 track $B^0 \to (D^- \to K^+\pi^-\pi^-)(D^0 \to K^-\pi^+\pi^+\pi^-)K^+$ (**norm8**) mode. The **norm8** mode is used for the **Z** and **Zs** spectra, while the **norm7** mode is used for all the rest.

The formalism behind the full simultaneous fit is somewhat non trivial, but simply put, the non-background contribution in each data peak is described as a sum over contributing components as

$$N_{\text{data yield}} = \sum_{\text{sig}\in\text{components}} \frac{B_{\text{sig}}}{B_{\text{norm}}} \sum_{i\in\texttt{categories}} \left(\epsilon_{\text{sig, i}} \cdot \frac{N_{\text{norm, i}}}{\epsilon_{\text{norm, i}}} \right) \tag{36}$$

In the above equation, norm denotes the appropriate normalization mode for the signal mode under consideration, ϵ denotes the efficiency, `categories` refers to the disjoint (year, trigger) regions. Branching fractions for the intermediate resonance decays have been merged into the terms B_{sig} and B_{norm}. Efficiencies and fit shapes are obtained from signal MC samples generated for each decay mode. `DecayTreeFitter` is used to constrain the masses of the intermediate D mesons, as well as to constrain the momentum of the b mother to point at the PV. The fits to the mass spectra are shown in Fig. 48. The fit to the $D_s^- D^+ K^*$ (**Zs**) spectrum does not describe the data well, but this turns out to not be a problem since the final background normalization (as shown in Tables 20, and 21) for $B_s^0 \to DDK^*$ turn out to be negligible. The measured yields for each of the decay modes, along with the signal efficiencies, charm branching fraction product and corrected yields are given in Table 18.

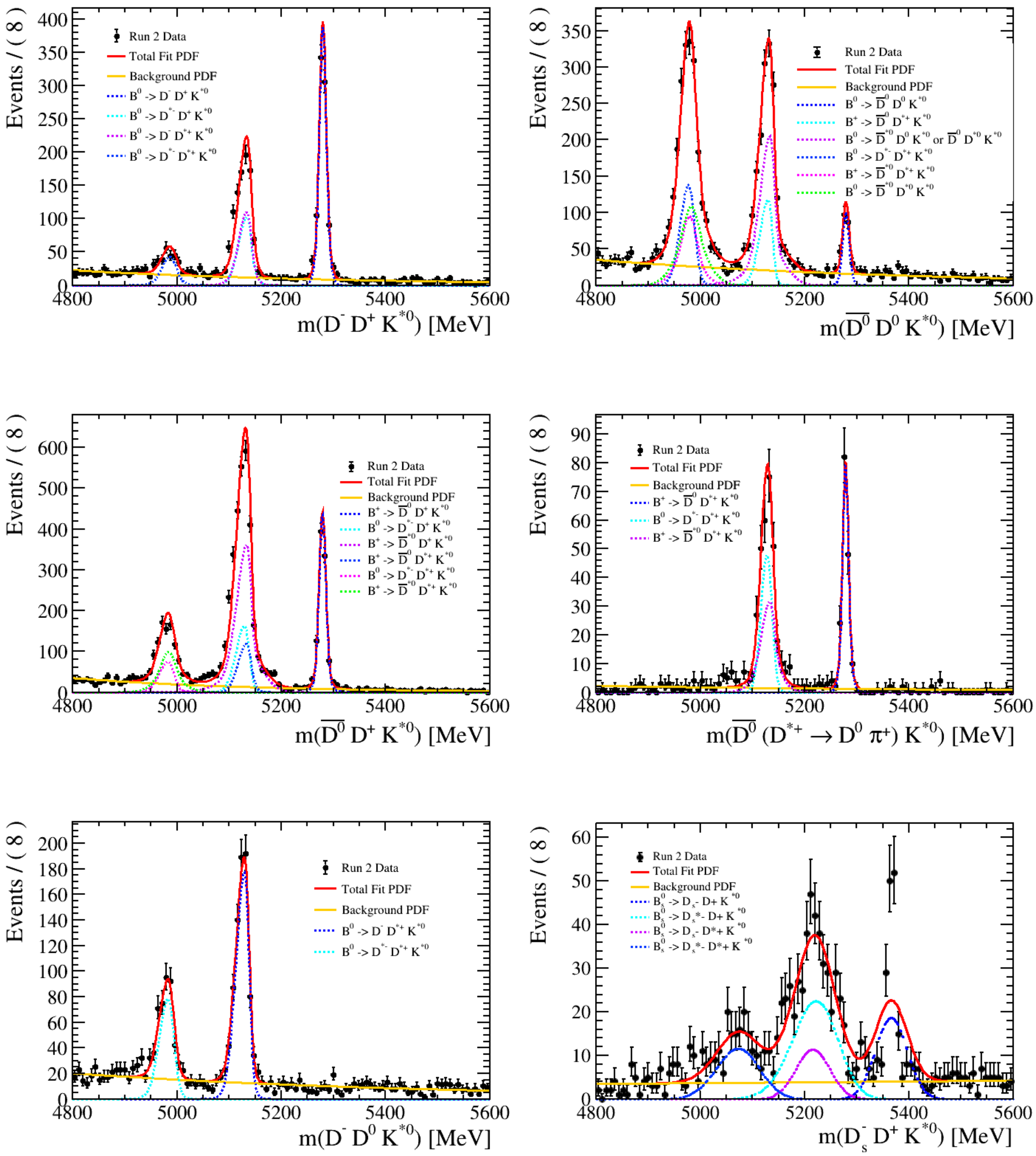


Figure 48: The fit to the $D^-D^+K^*$ (top left), $\overline{D}^0D^0K^*$ (top right), $\overline{D}^0D^+K^*$ (middle left), $\overline{D}^0D^{*+}K^*$ (middle right), $D^-D^0K^*$ (bottom left) and $D_s^-D^+K^*$ (bottom right) spectra is shown. The fit to the $D_s^-D^+K^*$ spectrum does not describe the data well, but for the purposes of our measurement this turns out to not be a problem (see text).

Decay Mode	Data Yield	D BF Product	Efficiency	Corrected yield
$B^0 \to D^- D^+ K^{*0}$	929 ± 30	$(8.8 \pm 0.3) \cdot 10^{-3}$	$(1.41 \pm 0.03) \cdot 10^{-4}$	$(7.5 \pm 0.4) \cdot 10^{8}$
$B^0 \to D^{*-} D^+ K^{*0}$	460 ± 30	$(2.8 \pm 0.1) \cdot 10^{-3}$	$(1.22 \pm 0.03) \cdot 10^{-4}$	$(1.3 \pm 0.1) \cdot 10^{9}$
$B^0 \to D^- D^{*+} K^{*0}$	514 ± 18	$(2.8 \pm 0.1) \cdot 10^{-3}$	$(1.22 \pm 0.03) \cdot 10^{-4}$	$(7.0 \pm 0.4) \cdot 10^{9}$
$B^0 \to D^{*-} D^{*+} K^{*0}$	232 ± 10	$(9.2 \pm 0.4) \cdot 10^{-4}$	$(9.28 \pm 0.24) \cdot 10^{-5}$	$(2.7 \pm 0.2) \cdot 10^{9}$
$B^0 \to \overline{D}^0 D^0 K^{*0}$	215 ± 16	$(1.56 \pm 0.02) \cdot 10^{-3}$	$(6.08 \pm 0.11) \cdot 10^{-4}$	$(2.3 \pm 0.2) \cdot 10^{8}$
$B^0 \to \overline{D}^{*0} D^0 K^{*0} + B^0 \to \overline{D}^0 D^{*0} K^{*0}$	1189 ± 52	$(1.56 \pm 0.02) \cdot 10^{-3}$	$(5.72 \pm 0.11) \cdot 10^{-4}$	$(1.33 \pm 0.06) \cdot 10^{9}$
$B^0 \to \overline{D}^{*0} D^{*0} K^{*0}$	1069 ± 55	$(1.56 \pm 0.02) \cdot 10^{-3}$	$(5.04 \pm 0.09) \cdot 10^{-4}$	$(1.4 \pm 0.1) \cdot 10^{9}$
$B^+ \to \overline{D}^0 D^+ K^{*0}$	1013 ± 31	$(3.71 \pm 0.07) \cdot 10^{-3}$	$(3.13 \pm 0.06) \cdot 10^{-4}$	$(8.8 \pm 0.4) \cdot 10^{8}$
$B^+ \to \overline{D}^{*0} D^+ K^{*0}$	2060 ± 75	$(3.71 \pm 0.07) \cdot 10^{-3}$	$(2.66 \pm 0.6) \cdot 10^{-4}$	$(2.1 \pm 0.1) \cdot 10^{9}$
$B^+ \to \overline{D}^0 D^{*+} K^{*0}$	611 ± 35	$(1.20 \pm 0.03) \cdot 10^{-3}$	$(2.70 \pm 0.06) \cdot 10^{-4}$	$(1.9 \pm 0.1) \cdot 10^{9}$
$B^+ \to \overline{D}^{*0} D^{*+} K^{*0}$	677 ± 34	$(1.20 \pm 0.03) \cdot 10^{-3}$	$(2.51 \pm 0.05) \cdot 10^{-4}$	$(2.3 \pm 0.1) \cdot 10^{9}$
$B_s^0 \to D_s^- D^+ K^{*0}$	175 ± 35	$(5.1 \pm 0.2) \cdot 10^{-3}$	$(1.46 \pm 0.04) \cdot 10^{-4}$	$(2.3 \pm 0.5) \cdot 10^{8}$
$B_s^0 \to D_s^{*-} D^+ K^{*0}$	304 ± 268	$(5.1 \pm 0.2) \cdot 10^{-3}$	$(1.23 \pm 0.03) \cdot 10^{-4}$	$(4.9 \pm 4.3) \cdot 10^{8}$
$B_s^0 \to D_s^- D^{*+} K^{*0}$	114 ± 211	$(1.63 \pm 0.06) \cdot 10^{-3}$	$(1.35 \pm 0.03) \cdot 10^{-4}$	$(5.1 \pm 9.5) \cdot 10^{8}$
$B_s^0 \to D_s^{*-} D^{*+} K^{*0}$	148 ± 51	$(1.63 \pm 0.06) \cdot 10^{-3}$	$(1.04 \pm 0.03) \cdot 10^{-4}$	$(8.7 \pm 3.0) \cdot 10^{8}$

Table 18: Measurements of the $b \to DDK^*$ decay modes. The fitted yield in data, along with the product charm branching fraction, signal efficiency and corrected yield is shown. The corrected yield is the fitted data yield divided by the efficiency and the charm product branching fraction.

Next, we propagate the results of these measurements into an estimate of the $b \to DDK^{*0}$ background present in our $K^{*0}\tau^+\tau^-$ data sample, after a given set of tight cuts.

Sum Branching Fraction	Value
$\sum \mathcal{B}(D^+ \to 3\pi X)$	10.4%
$\sum \mathcal{B}(D^0 \to 3\pi X)$	10.4%
$\sum \mathcal{B}(D^0 \to 2\pi X)$	37.4%
$\sum \mathcal{B}(D^{*+} \to 3\pi X)$	$(0.677 \times \mathcal{B}(D^0 \to 2\pi X)) + (0.323 \times \mathcal{B}(D^+ \to 3\pi X)) = 28.7\%$
$\sum \mathcal{B}(D^{*0} \to 3\pi X)$	10.4%
$\sum \mathcal{B}(D_s^+ \to 3\pi X)$	17.4%
$\sum \mathcal{B}(D_s^{*+} \to 3\pi X)$	17.4%

Table 19: Cumulative branching fractions for charm intermediates going to $3\pi X$ final states. D^{*+} decays to $D^0\pi^+$ 67.7% of the time, and when the resulting D^0 decays to $2\pi X$, it is possible for the two D^0 daughter pions to be combined with the bachelor pion from the D^{*+} to form a 3π combination. This is more likely to occur when the D^0 decay time is very small.

We do this in the following manner

$$N_{\text{mode}_i, K^*\tau\tau} = \frac{N_{\text{mode}_i}}{\epsilon_{\text{avg, i}} \cdot \mathcal{B}(D_1)\mathcal{B}(D_2)} \cdot \mathcal{B}(D_1 \to 3\pi X) \cdot \mathcal{B}(D_2 \to 3\pi X) \cdot \epsilon_{\text{mode}_i} \tag{37}$$

In this method, we start with efficiency and charm branching fraction corrected yields measured for each $b \to DDK^*$ mode. For example, if the decay mode is $B^0 \to D^- D^+ K^{*0}$, the charm branching fractions in the denominator are $\mathcal{B}(D^+ \to K^- \pi^+ \pi^+)^2$. The efficiency in the denominator is the average of those across the different `categories`. This efficiency and charm branching fraction corrected yield then gets multiplied by the cumulative Charm $\to 3\pi X$ branching fractions (see Table 19) and the overall efficiency of the $K^{*0}\tau^+\tau^-$ selections on the corresponding background cocktail MC(ϵ_{mode_i}). Since in this method we do not calculate the efficiency ratio by split into the disjoint categories (due to insufficient MC statistics in the cocktail samples after tight BDT cuts), the central value of the left hand side might be slightly off because we average over the changes in trigger thresholds. The results of calculating the $b \to DDK^*$ background contribution to our $K^*\tau^+\tau^-$ data with this method are shown in Tables 20 and 21, for some representative tight BDT cuts. Note that the background contribution from $B^0 \to DDK^*$ completely overwhelms that from $B^+ \to DDK^*$ and $B_s^0 \to DDK^*$. The fact that the contribution from $B_s^0 \to DDK^{*0}$ is almost zero shows that the failed fit to the **Zs** spectrum in the parallel analysis is not a problem. The DDK^* background normalization calculated in this manner for a given set of selections is fed to the fitter as a constraint on the DDK^* background fit yield.

BDT cuts	Data Yield	$B^0 \to DDK^{*0}$	$B^+ \to DDK^{*0}$	$B_s^0 \to DDK^{*0}$	DDK^* total	Exp. signal at $\mathcal{B} = 10^{-3}$
$\mathrm{kinBDT} > 0.98, \mathrm{isoBDT} > 0.95$	1035	77 ± 8	16 ± 2	6 ± 1	99 ± 8	57 ± 3
$\mathrm{kinBDT} > 0.98, \mathrm{isoBDT} > 0.97$	385	45 ± 6	5 ± 1	3 ± 1	52 ± 6	34 ± 2
$\mathrm{kinBDT} > 0.99, \mathrm{isoBDT} > 0.95$	490	44 ± 5	9 ± 2	3 ± 1	57 ± 6	41 ± 2
$\mathrm{kinBDT} > 0.99, \mathrm{isoBDT} > 0.97$	189	26 ± 4	3 ± 1	2 ± 1	31 ± 4	24 ± 2

Table 20: Results of the DDK^* background normalization for some representative cuts with trainMethod1. Yields shown in final fit range of $4700 - 6300\,\mathrm{MeV}$ in DTF mass. Only statistical uncertainties are quoted.

BDT cuts	Data Yield	$B^0 \to DDK^{*0}$	$B^+ \to DDK^{*0}$	$B_s^0 \to DDK^{*0}$	DDK^* total	Exp. signal at $\mathcal{B} = 10^{-3}$
$\mathrm{kinBDT} > 0.98, \mathrm{isoBDT} > 0.95$	619	50 ± 6	11 ± 2	4 ± 1	65 ± 6	51 ± 3
$\mathrm{kinBDT} > 0.98, \mathrm{isoBDT} > 0.97$	205	25 ± 4	4 ± 1	2 ± 1	29 ± 4	27 ± 2
$\mathrm{kinBDT} > 0.99, \mathrm{isoBDT} > 0.95$	262	32 ± 5	5 ± 1	2 ± 1	39 ± 5	31 ± 2
$\mathrm{kinBDT} > 0.99, \mathrm{isoBDT} > 0.97$	88	15 ± 3	2 ± 1	0.6 ± 0.3	18 ± 3	17 ± 2

Table 21: Results of the DDK^* background normalization for some representative cuts with trainMethod2. Yields shown in final fit range of $4700 - 6300\,\mathrm{MeV}$ in DTF mass. Only statistical uncertainties are quoted.

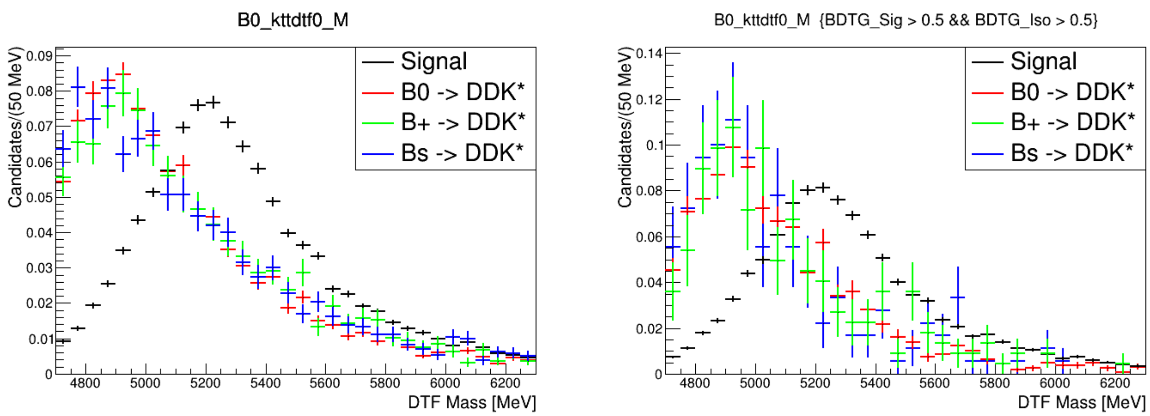


Figure 49: A comparison between the DTF mass shape for signal and various $b \to DDK^*$ backgrounds is shown. The left figure shows the comparison after best candidate selection (with trainMethod1), and the right figure shows the same, but with loose cuts applied on the isolation and kinematic BDT responses. The large error bars for the background shapes in the right figure reflect the scarcity of statistics in these MC samples.

4.9 Fitter

In order to determine the branching fraction of the signal, the main ingredient that needs to be extracted from data is the yield of the potential signal. We do this by performing a binned maximum likelihood template fit to the B^0 DTF mass distribution in data. Nominally, our fit has three components, one for signal, one for the DDK^* background, and one for combinatorial background. The shapes for signal and the DDK^* background are described with templates extracted from the appropriate simulation simples. The combinatorial background shape is extracted from data using a background dominated BDT slice, adjacent to the signal rich region. We use a 50 MeV bin size in our fit, with a fit range from $4700 - 6300$ MeV in the DTF mass. The choices bin size and fit range are driven by the width of the signal shape and its span, respectively.

A simple equation for describing the signal yield is

$$n_{\text{sig}} = \frac{\mathcal{B}(B^0 \to K^*\tau^+\tau^-) \cdot \mathcal{B}(K^* \to K^+\pi^-) \cdot \mathcal{B}(\tau \to 3\pi(\pi^0)\nu_\tau)^2}{\mathcal{B}(B^0 \to D^-D^0K^+) \cdot \mathcal{B}(D^0 \to K3\pi) \cdot \mathcal{B}(D^- \to K2\pi)} \cdot \frac{N_{\text{norm}}}{\epsilon_{\text{norm}}} \cdot \epsilon_{\text{sig}} \tag{38}$$

where B_{sig} and B_{norm} include the branching fractions of the intermediate resonance decays. We refine this a step further by splitting up the normalization data and simulation, as well as the signal simulation, into the disjoint (trigger, year) `categories`. In doing so, we account for the changes in LHCb's trigger thresholds from one year of data taking to the other.

$$n_{\text{sig}} = \frac{\mathcal{B}(B^0 \to K^*\tau^+\tau^-) \cdot \mathcal{B}(K^* \to K^+\pi^-) \cdot \mathcal{B}(\tau \to 3\pi(\pi^0)\nu_\tau)^2}{\mathcal{B}(B^0 \to D^-D^0K^+) \cdot \mathcal{B}(D^0 \to K3\pi) \cdot \mathcal{B}(D^- \to K2\pi)} \cdot \sum_{i \in \texttt{categories}} \left[\frac{N_{\text{norm, i}}}{\epsilon_{\text{norm, i}}} \cdot \epsilon_{\text{sig, i}} \right] \tag{39}$$

All the parameters on the right hand side that are not the signal branching fraction are collected into a single term, which we refer to as α.

$$n_{\text{sig}} = \mathcal{B}(B^0 \to K^*\tau^+\tau^-) \cdot \alpha \tag{40}$$

$$\alpha = \frac{\mathcal{B}(K^* \to K^+\pi^-) \cdot \mathcal{B}(\tau \to 3\pi(\pi^0)\nu_\tau)^2}{\mathcal{B}(B^0 \to D^- D^0 K^+) \cdot \mathcal{B}(D^0 \to K3\pi) \cdot \mathcal{B}(D^- \to K2\pi)} \cdot \sum_{i \in \texttt{categories}} \left[\frac{N_{\text{norm, i}}}{\epsilon_{\text{norm, i}}} \cdot \epsilon_{\text{sig, i}} \right] \quad (41)$$

The layout of this section is as follows: we start by presenting the formalism of the binned likelihood fit in `HistFactory` [75], which is the ROOT framework we use to develop the fitter. We then present the Beeston Barlow method for handling the template uncertainties in the fit. Following this, we discuss the manner in which we extract the templates for the combinatorial and DDK^* backgrounds, and the subtleties involved.

4.9.1 HistFactory

This subsection described the binned maximum likelihood formalism used by HistFactory. In a toy fit to a single channel that has one signal and one background contribution, and no systematics, the probability model for obtaining n events in data where the discriminating variable for event e has the value x_e can be written as

$$\mathcal{P}(x_1, x_2, ...x_n|\mu) = \text{Pois}(n|\mu S + B) \cdot \left[\prod_{e=1}^{n} \frac{\mu S f_S(x_e) + B f_B(x_e)}{\mu S + B} \right]. \quad (42)$$

This expression is known as the marked Poisson model. μ is a signal strength parameter that is zero in the background only hypothesis and one in the nominal signal+background hypothesis. The first term on the right hand side denotes the probability for obtaining n events in data when $\mu S + B$ events are expected. The second term is the probability of getting the value x_e for event e given the relative mixture of the signal and background PDFs $f_S(x_e), f_B(x_e)$ for a given value of μ, combined over all the events. This expression is interpreted as the likelihood when the data is taken to be fixed, making it a function of μ, the parameter of interest, $L(\mu)$. The usual parameter estimation scheme is to maximize the likelihood, or equivalently, to minimize its negative logarithm

$$-\ln L(\mu) = (\mu S + B) + \ln n! - \sum_{e=1}^{n} \ln \left[\mu S f_S(x_e) + B f_B(x_e)\right]. \quad (43)$$

The binned maximum likelihood is a binned equivalent of this formalism, where continuous signal and background shapes are replaced by histogram templates ν_b^{sig} and ν_b^{bkg}, where b denotes the bin index, and the histogram contents represent the number of events expected in data. These are related to the shapes $f(x)$ via

$$f_S(x_e) = \frac{\nu_{b_e}^{\text{sig}}}{S\Delta_{b_e}} \tag{44}$$

$$f_B(x_e) = \frac{\nu_{b_e}^{\text{bkg}}}{B\Delta_{b_e}} \tag{45}$$

where b_e denotes the bin index of the bin containing event e and Δ_{b_e} denotes the width of that bin. Naturally, we have $S = \sum_b \nu_b^{\text{sig}}$ and $B = \sum_b \nu_b^{\text{bkg}}$

The likelihood now takes the form of a product of Poisson distributions in each bin

$$L = \prod_b \text{Pois}(n_b|\mu\nu_b^{\text{sig}} + \nu_b^{\text{bkg}}) \tag{46}$$

This is the general form of the likelihood that is maximized by `HistFactory`. But there is a slight deficiency in this formalism. It takes the bin counts in the supplied templates as the expected values for the corresponding source (signal or some kind of background). It does not account for the statistical fluctuations that will be present in these templates, which would throw off the expected values. The Beeston Barlow method [76] is used to account for this.

4.9.2 Beeston Barlow method

Ideally, we would like to have MC templates with very large statistics, in order to minimize the fluctuations. Typically, a rule of thumb is that if the MC templates contain ten times the statistics present in the data being fit to, the effect of fluctuations would not be significant. But the generation of MC samples is a computationally expensive process and we typically do not have this luxury. So, we must deal with the statistical fluctuations present in the MC sample.

Instead of framing the example as just one signal and one background source, we can consider some m sources in the fit, each with their own MC template histograms. Let these histograms contain n bins, with a_{ji} MC events generated for the j-th source in the i-th bin. The total generated size of the j-th MC sample is $N_j = \sum_{i=1}^{n} a_{ji}$. Based on this, the expected number of events in the i-th bin of the data is

$$f_i = N_D \cdot \sum_{j=1}^{m} \frac{P_j a_{ji}}{N_j} \tag{47}$$

where N_D is the total size of the data sample, and P_j is the strength of the component j, i.e. its fit fraction, such that $\sum_j P_j = 1$. We can see that fluctuation in the MC templates would affect the a_{ji} leading to a mismatch between the expected data and observed data. (It can also be seen that these fluctuations would be damped down by a factor of $\frac{N_D}{N_j}$, proving our intuition that the effect of fluctuations decreases with large MC samples.)

So for each source, there is some unknown expected number of events A_{ji}, and the observed MC events a_{ji} follows a Poisson distribution with A_{ji} as the expected value. If we denote the observed number of events in bin i of the data as d_i, then the total likelihood to be maximized is the combined probability of observing $\{d_i\}$ and $\{a_i\}$

$$L = \left(\prod_i \text{Pois}(d_i | f_i) \right) \cdot \left(\prod_i \prod_j \text{Pois}(a_{ji} | A_{ji}) \right) \tag{48}$$

The Beeston Barlow method is the maximization of this log likelihood

$$\text{log}L = \sum_i d_i \text{log}(f_i) - f_i + \sum_i \sum_j a_{ji} \text{log} A_{ji} - A_{ji} \tag{49}$$

The maximization of this likelihood will yield the optimal values for the m P_j parameters (which we are interested in), and the $m \times n$ A_{ji} parameters(which we do not really care about). Computationally, this becomes a maximization problem in $m \times (n+1)$ parameters.

In `HistFactory`, to make the problem computationally tractable, a single nuisance parameter is assigned per bin, shared by all the sources, instead of having one nuisance parameter bin source per bin. Let us denote the bin index as b, the observed number of events in bin b as n_b, the part of the expected yield on which we don't need a statistical uncertainty (either because it is data driven or because the MC sample is large) as ν_b, the part of the expected yield on which we do need a statistical uncertainty as ν_b^{MC}. If the total statistical uncertainty in bin b (from all sources combined) is δ_b, the relative statistical uncertainty is $\frac{\delta_b}{\nu_b^{\mathrm{MC}}}$. This corresponds to a total MC sample in bin b of size $m_b = \left(\frac{\nu_b^{\mathrm{MC}}}{\delta_b}\right)^2$. Let us denote by γ_b the nuisance parameter in bin b reflecting how much the true rate n_b differs from the expectation ν_b^{MC}.

The contribution to the likelihood now is the factor

$$L_{\mathrm{BB}} = \mathrm{Pois}(n_b|\nu_b + \gamma_b\nu_b^{\mathrm{MC}}) \cdot \mathrm{Pois}(m_b|\gamma_b m_b) \tag{50}$$

where $m_b = \left(\frac{\nu_b^{\mathrm{MC}}}{\delta_b}\right)^2$ is used to define an total equivalent expected yield in bin b, using δ_b as the total statistical uncertainty on ν_b^{MC}. m_b is simply the yield that has the same statistical uncertainty as the combination of all the MC yields in bin b. It is defined based on the logic that the relative uncertainty in bin b goes as $\frac{1}{\sqrt{m_b}}$. In other words m_b is a "sum-of-sources" equivalent of the a_{ji} in the original Beeston Barlow formalism.

Now

$$L_{\mathrm{BB}} = \frac{e^{-(\nu_b+\gamma_b\nu_b^{\mathrm{MC}})} \cdot (\nu_b + \gamma_b\nu_b^{\mathrm{MC}})^{n_b}}{n_b!} \cdot \frac{e^{-(\gamma_b m_b)} \cdot (\gamma_b m_b)^{m_b}}{m_b!} \tag{51}$$

We can analytically find the solutions $\hat{\gamma_b}$ by setting $\frac{\partial(-\log L_{BB})}{\partial\gamma_b} = 0$. After some algebra this leads to the quadratic equation

$$0 = \hat{\gamma_b}^2 \cdot [(\nu_b^{\mathrm{MC}})^2 + m_b\nu_b^{\mathrm{MC}}] + \hat{\gamma_b} \cdot [\nu_b^{\mathrm{MC}}\nu_b - n_b\nu_b^{\mathrm{MC}} + \nu_b m_b - m_b\nu_b^{\mathrm{MC}}] - m_b\nu_b \tag{52}$$

which can be solved simply

$$\hat{\gamma_b} = \frac{-B + \sqrt{(B^2 - 4AC)}}{2A} \tag{53}$$

with

$$A = (\nu_b^{\mathrm{MC}})^2 + m_b \nu_b^{\mathrm{MC}}$$

$$B = \nu_b^{\mathrm{MC}} \nu_b - n_b \nu_b^{\mathrm{MC}} + \nu_b m_b - m_b \nu_b^{\mathrm{MC}}$$

$$C = -m_b \nu_b$$

It is important to note here that both ν_b and ν_b^{MC} are functions of the other fit parameters, specifically the fit fractions of the different components. So, in the minimization procedure for the main fit, at each step for values of the fit fractions, the γ_b values are calculated analytically as above.

4.9.3 Combinatorial background shape

In order to describe the shape of the combinatorial background in the signal region (with whatever isoBDT and kinBDT selections are chosen by the toy studies), we look to a background dominated BDT region of the RS data, keeping in mind that the signal region(kinBDT $0.98 - 1.00$ for trainMethod 1 and kinBDT $0.97 - 1.00$ for trainMethod 2) is blinded.

We choose a nominal background region of kinBDT $0.90 - 0.95$, isoBDT > 0.9 (for both trainMethods). Despite what looks like tight selections on the BDT responses, this region in data is still background dominated. At an assumed signal branching fraction of 1×10^{-3}, with trainMethod1, 77 ± 9 signal candidates are expected in this region and 4673 data candidates are observed. Similarly, for trainMethod2, 93 ± 10 signal candidates are expected in this region and 4261 data candidates are observed. This tells us that this region is overwhelmingly dominated by combinatorial background. In order to have a smooth combinatorial background shape input to the fitter, we fit to the mass distribution in the nominal background region with a third order polynomial. A histogram constructed from the fitted curve is the combinatorial background shape input to the fitter.

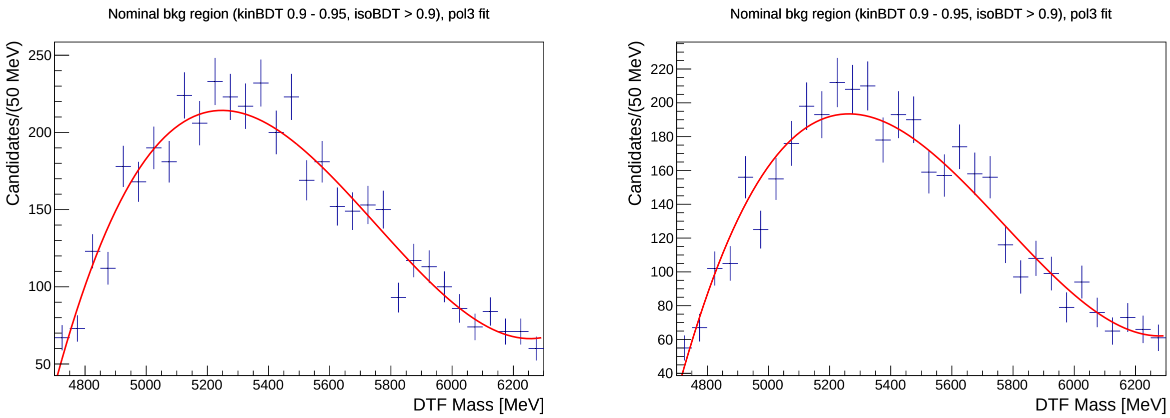


Figure 50: DTF mass shape for the RS data in the nominal background region (kinBDT 0.9-0.95, isoBDT > 0.9) for trainMethod1 (left) and trainMethod2 (right). χ^2/ndf for left fit is 46/28, and for the right fit is 39/28.

We know that the final kinBDT and isoBDT selections we apply on the signal region will be tighter. It is therefore important to understand how much the combinatorial background shape in the signal region could vary from that seen in the nominal background region. First, we examine the evolution of the DTF mass shape in slices of the kinematic BDT, with no cut on the isolation BDT, to see how much of an effect the kinematic BDT has on the shape. Fig.51 shows that there isn't a drastic effect on the shape of the mass distribution (which is almost completely combinatorial background in these regions) in tightening kinBDT slices.

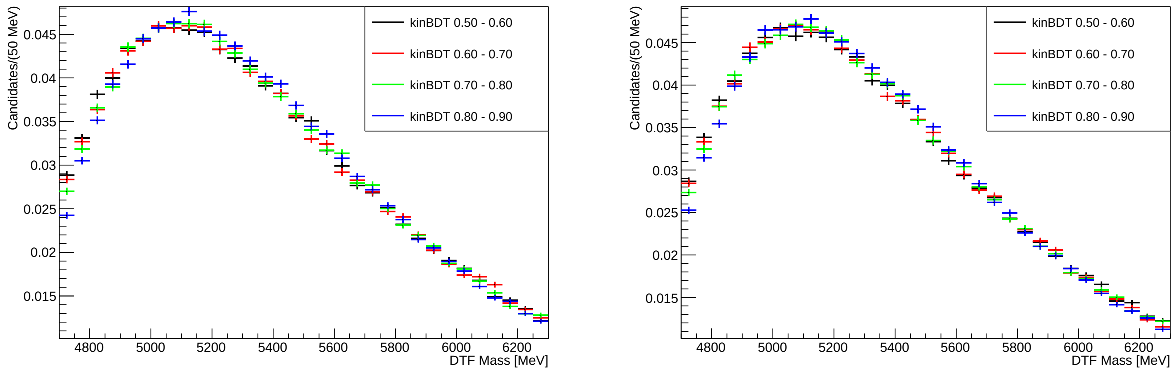


Figure 51: DTF mass shape for the RS data in bins of the kinematic BDT for trainMethod1 (left) and trainMethod2 (right). The histograms are all normalized to unit area. No significant variation in shape is seen in this region as we move to higher kinBDT slices.

We then investigate the evolution of the mass shape in a given kinBDT slice, over increasingly tight cuts on the isoBDT. We see in Fig.52 that cutting on the isoBDT doesn't have a drastic effect on the mass shape.

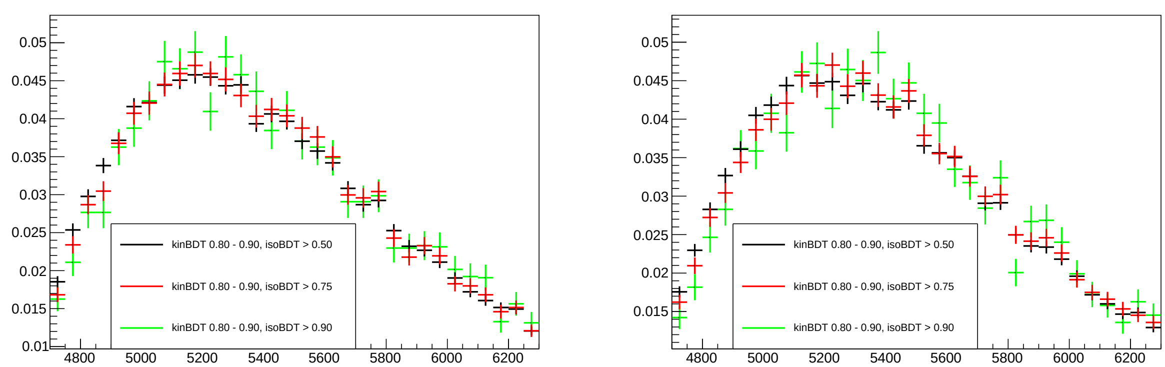


Figure 52: DTF mass shape for the RS data in bins of the isolation BDT for trainMethod1 (left) and trainMethod2 (right). The histograms are all normalized to unit area. No significant variation in shape is seen as the cut on isoBDT is tightened.

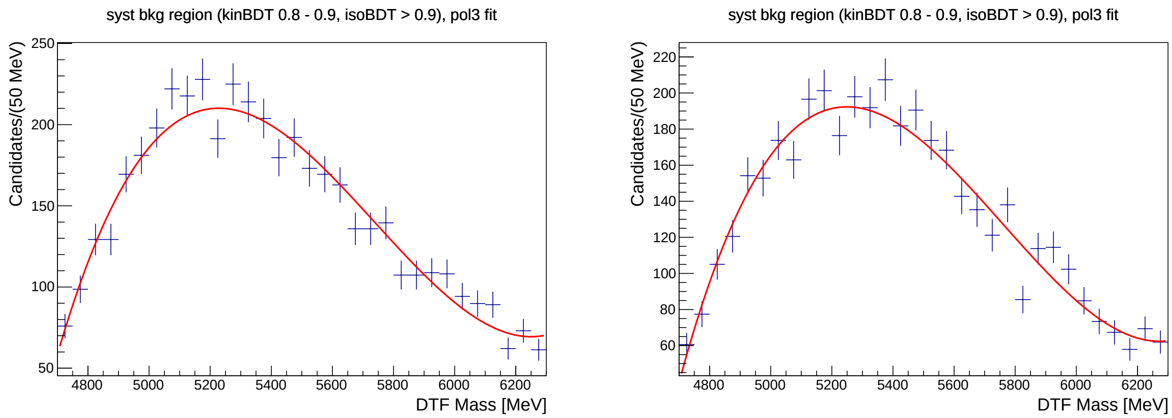


Figure 53: DTF mass shape for the RS data in the `syst` background region (kinBDT 0.8-0.9, isoBDT > 0.9) for trainMethod1 (left) and trainMethod2 (right). χ^2/ndf for left fit is 36/28, and for the right fit is 46/28.

With the knowledge that there is not going to be a drastic change in the shape of the combinatorial background from the nominal background region to the signal region, we assign a systematic shape uncertainty to the nominal combinatorial background shape. In order to get a systematic uncertainty on the nominal background shape, we examine the mass distribution in the region kinBDT $0.8-0.9$, isoBDT > 0.9, which we will refer to as the `syst` background region. As with the nominal background region, we fit to the mass distribution in this region with a third order polynomial. This shape is treated as a "downward" (or -1σ) variation on the nominal shape. The upward (or $+1\sigma$) variation is obtained by performing a bin-by-bin reflection of the downward variation around the nominal shape. Fig.54 shows the comparison of the nominal combinatorial background shape with the $\pm1\sigma$ variations.

In HistFactory, the systematic variations on a fit template are handled via a `HistoSys` (as in, histogram systematic). A parameter α is created with a value of 0 for the nominal shape, and ±1 for the $\pm1\sigma$ variations. This is one of the fitted parameters, and the background shape is interpolated linearly between the two variations as α varies. If the template histogram is denoted by σ (which is now thought of as a function of α), the interpolation can be expressed simply as:

$$\sigma(\alpha) = \sigma(0) + I_{\text{lin}}(\alpha; \sigma(0), \sigma(1), \sigma(-1)) \tag{54}$$

$$I_{\text{lin}}(\alpha; I^0, I^+, I^-) = \begin{cases} \alpha \cdot (I^+ - I^0) & \alpha \geq 0 \\ \alpha \cdot (I^0 - I^-) & \alpha < 0 \end{cases} \tag{55}$$

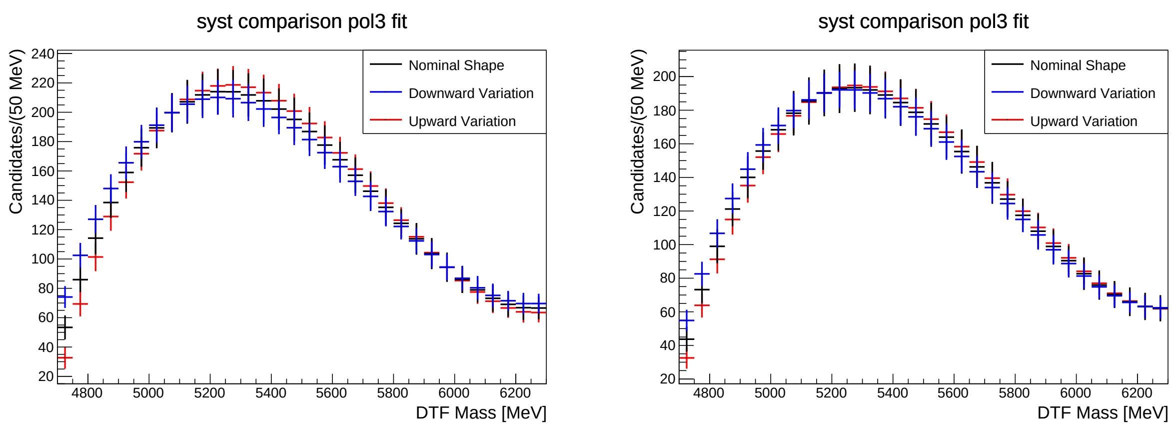


Figure 54: The nominal combinatorial background shape overlaid with the $\pm 1\sigma$ systematic variations for trainMethod1 (left) and trainMethod2 (right). All histograms have been normalized to the area of the nominal histogram.

4.9.4 DDK^* background shape

As we saw from the calculations in Sec. 4.8, the background contribution from $B^0 \to DDK^*$ overwhelms that from $B^+ \to DDK^*$ and $B_s^0 \to DDK^*$. The total DDK^* background contribution after tight BDT selections is a non-negligible fraction of the total data yield. As a result, it is imperative to include a shape in the fit to describe this background. Given the fact that the DTF mass shapes of these backgrounds are almost identical (see Fig. 49), and that the total background contribution is dominated by $B^0 \to DDK^*$, we use only the $B^0 \to DDK^*$ cocktail background MC sample to extract the shape for the total $b \to DDK^*$ background.

As was mentioned before, the background MC samples are scarce in statistics after tight BDT cuts. In order to have enough statistics to describe the shape, we loosen the isolation and kinematic BDT cuts to isoBDT > 0, kinBDT > 0. To be able to do this meaningfully, we show in Fig. 55 that loosening these cuts does not affect the shape of this background.

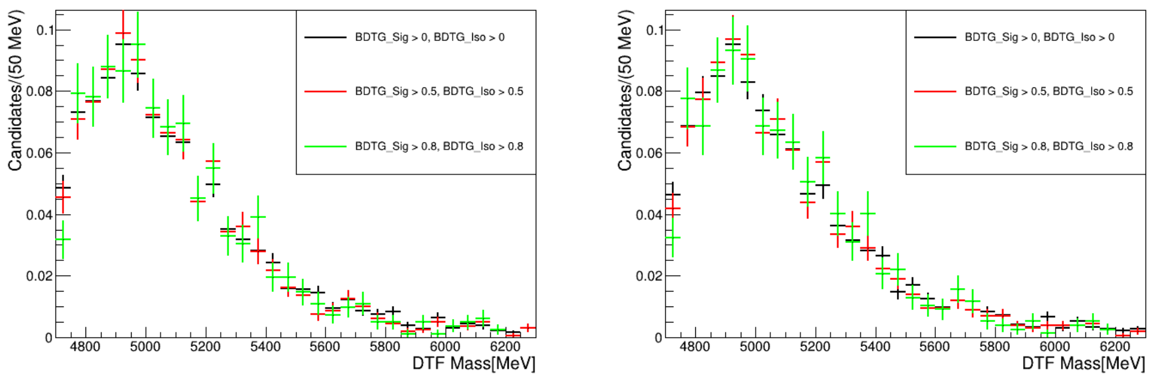


Figure 55: Study of the effect of tightening kinBDT and isoBDT cuts on the shape of the DDK^* background for trainMethod1 (left) and trainMethod2 (right). No significant variation of shape is seen as we tighten the BDT cuts.

4.9.5 Constraints

Based on external knowledge, we impose constraints in the fit. Generally speaking, if we have an auxiliary measurement a_p, associated to a parameter in the fit model α_p (unrelated to our α), a constraint term $f(a_p|\alpha_p)$ multiplying the likelihood can be defined, to incorporate the auxiliary knowledge into the fit. The Gaussian constraint for an auxiliary measurement with uncertainty σ_p takes the form

$$G(a_p|\alpha_p, \sigma_p) = \frac{1}{\sqrt{2\pi\sigma_p^2}} \exp\left[-\frac{(a_p - \alpha_p)^2}{2\sigma_p^2} \right] \tag{56}$$

For a given set of selections, we are able to calculate a central value and an error for the parameter α that was previously defined. This is the number that multiplies the signal branching fraction to give the signal yield. A Gaussian constraint is imposed on α in the fit. Further, following the procedure outlined in Sec. 4.8, we are able to calculate a central estimate and error for the DDK^* background contribution in the data, n_{DDK^*}, and thus impose a Gaussian constraint on its yield in the fit.

For a given set of BDT selections, we obtain central values and errors for α_{norm} and n_{DDK^*}. These numbers are input to the fitter in the form of a Gaussian constraint which multiplies the likelihood PDF.

4.10 Sensitivity Measurement

Now that we have set up our fitter, we perform toy studies to accomplish the following:

- Ensure that fitter is doing what it is supposed to: i.e. when fitting to background only data, it picks up zero signal on average, and when fitting to data that has signal injected in it, the fitter is able to capture (on average) the correct amount of signal.

- Study the evolution of the expected sensitivity to signal as a function of the cuts on trainMethod, isoBDT and kinBDT; and by doing so pick the optimal trainMethod and BDT cuts as the ones that maximize our expected sensitivity to the signal

- Study the dependence of the fitter on the calculated normalization of the DDK^* background.

To perform the toy studies, we first obtain the yield of the RS data for the selections of interest (which comprise the fit window $4700 - 6300\,\mathrm{MeV}$, isoBDT cut and kinBDT cut). Let us denote this number as n_{data}. Using the procedure outlined in Sec.4.8, we calculate the normalization of the DDK^* backgrounds, n_{DDK^*}. When performing background only toy studies, the amount of combinatorial background in the data is taken to be $n_{\mathrm{comb}} = n_{\mathrm{data}} - n_{DDK^*}$. If signal injection toy studies are being performed, and $n_{\mathrm{sigInject}}$ is the amount of signal injected in the data, $n_{\mathrm{comb}} = n_{\mathrm{data}} - n_{DDK^*} - n_{\mathrm{sigInject}}$ (Therefore, we cannot inject more signal than $n_{\mathrm{data}} - n_{DDK^*}$).

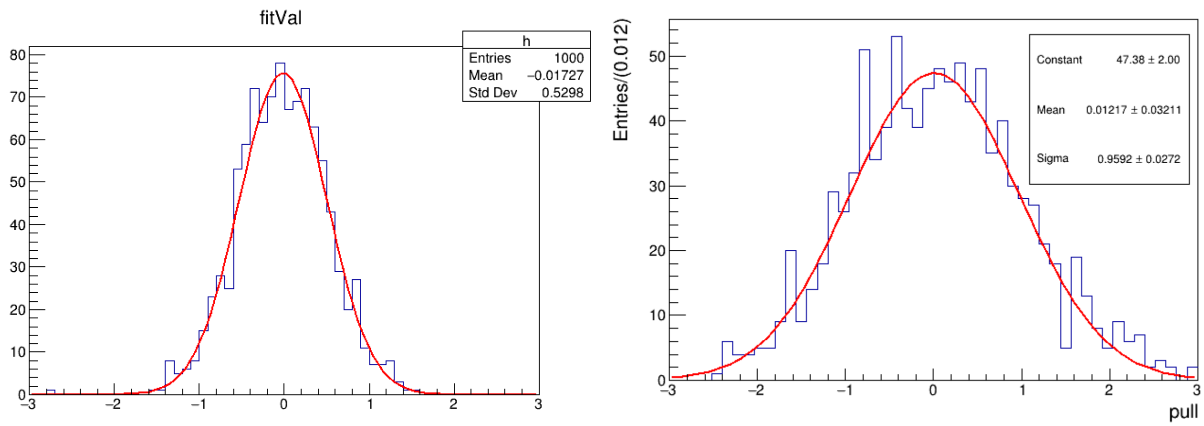
Using the templates for the DDK^* and combinatorial backgrounds (see Sections 4.9.4 and 4.9.3), we randomly generate toy histograms for each of these components with yields of n_{DDK^*} and n_{comb} respectively. Let these histograms be denoted as H_{DDK^*} and H_{comb}. If signal injection toy studies are being performed, the signal template is used to generate an injected signal histogram $H_{\mathrm{sigInject}}$ with a yield of $n_{\mathrm{sigInject}}$. The toy data histogram, H_{toyData}, is then $H_{DDK^*} + H_{\mathrm{comb}}$ for background only toy studies, and $H_{DDK^*} + H_{\mathrm{comb}} + H_{\mathrm{sigInject}}$ for signal injection studies. Either way, the total size of H_{toyData} is equal to n_{data}.

Fit Parameter	Fitted Value
B_{sig}	$(2.9 \pm 6.8) \times 10^{-4}$
α_{combBkg}	-0.09 ± 0.99
α_{norm}	$(-2.2 \times 10^{-5} \pm 0.99)$
α_{nDDK^*}	0.06 ± 0.94
n_{combBkg}	19 ± 9

Table 22: Results of one instance of a background only toy fit with $\text{kinBDT} > 0.994, \text{isoBDT} > 0.970$ (trainMethod 2). The fitted values of the 32 Beeston Barlow parameters are not shown. They are all close to 1, with errors around 0.05. α_{combBkg} is the Gaussian interpolation parameter on the combinatorial background shape. α_{norm} and α_{nDDK^*} are the Gaussian constraint parameters on the branching fraction normalization (defined in Eq. 40) and the DDK^* background yield. All three α's denote how much the fit pulls the constrained parameters from its mean, in units of the constraint width. The corresponding fit is shown in Fig. 57.

The fit has 37 free parameters: 1 POI, 1 yield for the combinatorial background, two Gaussian constraint parameters pertaining to the gaussian constraints on α and n_{DDK^*} (the constraint parameters denote how much the fitted value deviates from the mean of the constraint, in units of the width of the constraint), one interpolation parameter for the shape systematic on the combinatorial background and 32 Beeston Barlow nuisance parameters, one for each bin in the fit.

We first discuss the background only toy studies. For each background toy that is generated, we fit to the background only toy data with our nominal fit model, and record the resulting fitted value of the signal branching fraction. These toy studies are conducted across a range of selections on isoBDT and kinBDT, for both trainMethod. Our first concern is to make sure that the fitter is functioning properly in background only situations. At $\text{kinBDT} > 0.994, \text{isoBDT} > 0.970$, we generate and fit to 1000 background only toys. An instance of the background only toy fit along with the corresponding upper limit scan is shown in Fig.57. The distribution of the fitted signal branching fraction values from these toys is shown in Fig. 56a. The resulting distribution is fit with a Gaussian PDF, whose fitted mean is consistent with zero. This demonstrates that the fitter does not pick up signal, on average, when fitting to background only data.



(a) (b)

Figure 56: Distribution of fitted signal branching fraction values (in units of 10^{-3}) (left) and pull values(right) from 1000 background only toy fits with $\mathrm{kinBDT} > 0.994, \mathrm{isoBDT} > 0.970$ (trainMethod 2), shown as a blue histogram, along with a Gaussian PDF fitted to it, shown as the red lineshape.

For each toy study instance, the pull of the fitted branching fraction is also captured, using the asymmetric errors calculated by MINUIT [77]. The pull is defined as

$$\mathrm{pull} = \begin{cases} \frac{\mathrm{fitVal}}{\mathrm{fitErr\_Low}} & \mathrm{fitVal} > 0 \\ -\frac{\mathrm{fitVal}}{\mathrm{fitErr\_High}} & \mathrm{fitVal} < 0 \end{cases} \tag{57}$$

The distribution of the pull values from background only toy studies is shown in Fig.56b, and follows a standard Gaussian distribution as expected.

The result of the fit is also converted into an upper limit at 95% C.L. on the signal branching fraction, using the CLs method. The upper limit scan as a function of the signal branching fraction is shown in Fig.58. The upper limit calculation provides an expected upper limit (that is calculated under the background-only hypothesis, and is independent of the actual distribution of the data), as well as an observed upper limit (which is calculated based on the observed data). The fitted value of the signal branching fraction corresponding to this instance is $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) = (0.3 \pm 0.7) \times 10^{-3}$. We optimize the cuts on the BDT selections based on the expected limit, so as to not be influenced by fluctuations in the toy data. The variation of the expected limit as a function of the BDT cuts is shown in Fig.59 for both BDT training methods. We see that the best sensitivity is obtained with trainMethod2, with kinBDT > 0.994, isoBDT > 0.970, with an upper limit of

$$\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) < 1.51 \times 10^{-3} \text{ at } 95\%\text{C.L.} \tag{58}$$

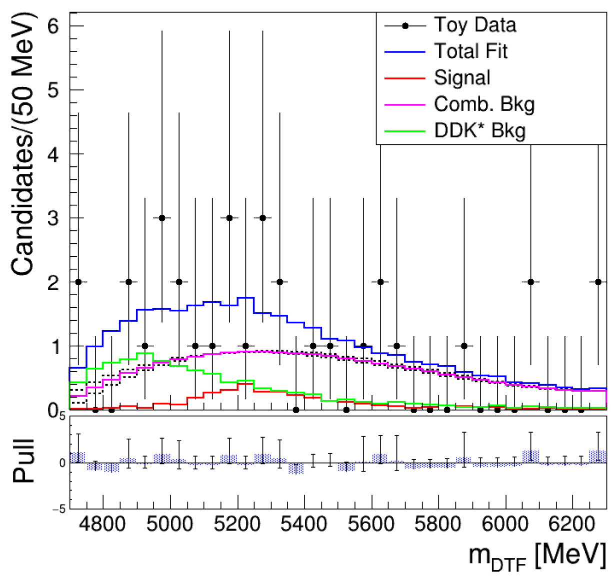


Figure 57: A fit to background only toy data is shown, corresponding to the selections kinBDT $>$ 0.994, isoBDT > 0.97 (trainMethod2). The dotted black lines around the combinatorial background fit shape shows the $\pm 2\sigma$ shape uncertainty envelope on that component.

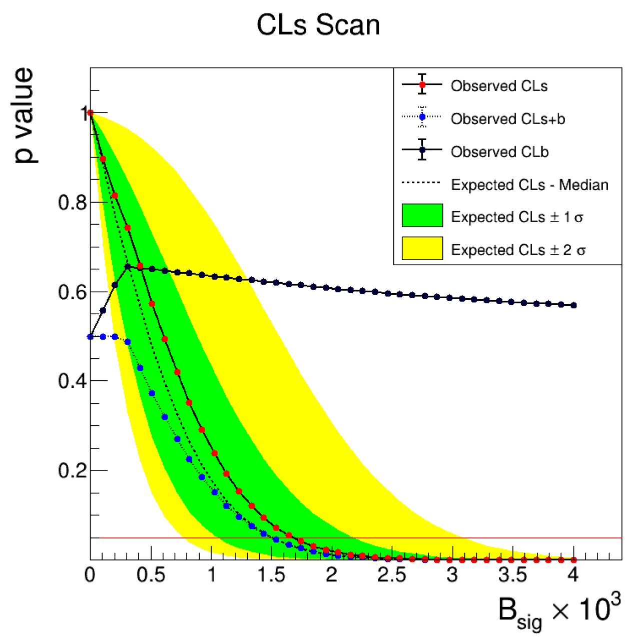


Figure 58: The asymptotic upper limit scan for a background only fit (kinBDT > 0.994, isoBDT > 0.97, trainMethod 2)) is shown, as a function of the signal branching fraction. The expected limit at 95% C.L. is the value of the signal branching fraction at which the expected CLs curve falls below 0.05 in p-value.

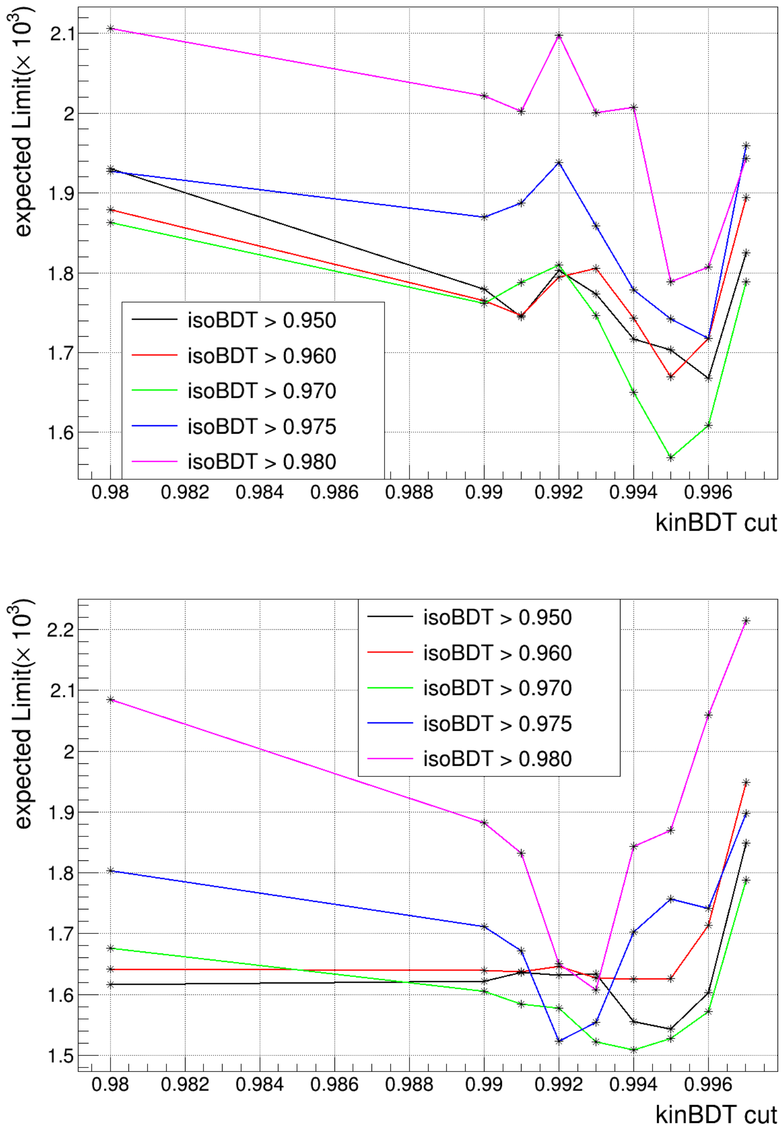


Figure 59: The variation of the expected upper limit (at 95% C.L.) on the signal branching fraction, as calculated from background only toy studies, as a function of cuts on the kinBDT and isoBDT responses, is shown for trainMethod1 (top) and trainMethod2 (bottom). The best sensitivity obtained is 1.51×10^{-3} for trainMethod2 with kinBDT > 0.994, isoBDT > 0.970.

We also perform signal injection toy studies, to verify that the fitter is capable of capturing potential signal present in the data. An instance of a fit to a signal injection toy, with signal injected at $\mathcal{B}(B^0 \to K^*\tau^+\tau^-) = 2 \times 10^{-3}$, for kinBDT > 0.990, isoBDT > 0.97 with trainMethod 2, is shown in Fig.60 (left), along with the distribution of the fitted signal branching fraction over 1000 toys. We see that on average, the fitter is able to accurately capture the amount of signal present in the toy data. The fitted value for the signal branching fraction for the instance shown in Fig.60 (left) is $\mathcal{B}(B^0 \to K^* \tau^+ \tau^-) = (2.9 \pm 0.9) \times 10^{-3}$.

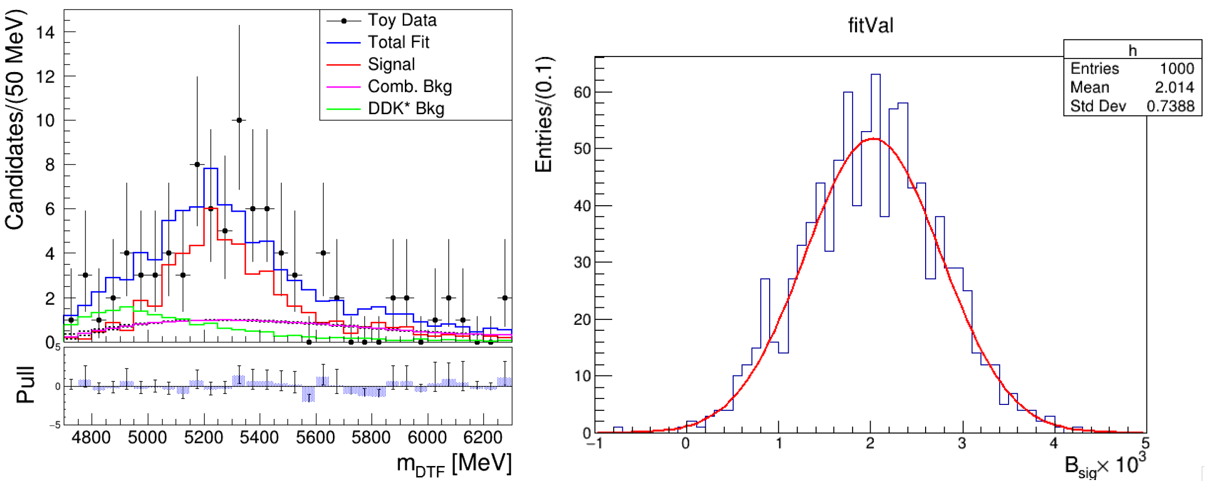


Figure 60: (Left) A signal injection toy fit (kinBDT > 0.990, isoBDT > 0.97, trainMethod 2) is shown, with signal injected at $\mathcal{B}(B^0 \to K^* \tau^+ \tau^-) = 2 \times 10^{-3}$. The dotted black lines around the combinatorial background fit shape shows the $\pm 2\sigma$ shape uncertainty envelope on that component. (Right) The distribution of the fitted signal branching fraction in 1000 signal injection toy fits. The mean of the distribution is consistent with the injected signal branching fraction.

Finally, we perform a crosscheck to study our the dependence of the fitter on the calculated DDK^* background normalization. Since we are reliant on the known $D \to 3\pi X$ modes, it is possible that we are undercounting the amount of this background present in the data. Further, the calculation of this background normalization is somewhat non trivial, and relies on external inputs. Hence, we generate and fit to 1000 signal injection toys (with signal injected at $B_{\text{sig}} = 2 \times 10^{-3}$) at kinBDT > 0.990, isoBDT > 0.97 with trainMethod 2, while increasing the DDK^* background yield in the toy data by 50%, but keeping the constraint in the fit at its original value. By doing so, we simulate the effect of more DDK^* background in the data than we anticipated. To find out if this biases the fitted signal captured by the fitter, we plot the distribution of the fitted signal branching fractions across all 1000 toys in Fig.61, and fit to it with a Gaussian PDF. We see that there is no significant bias on the distribution of the signal branching fraction, since the mean is consistent with the injected signal.

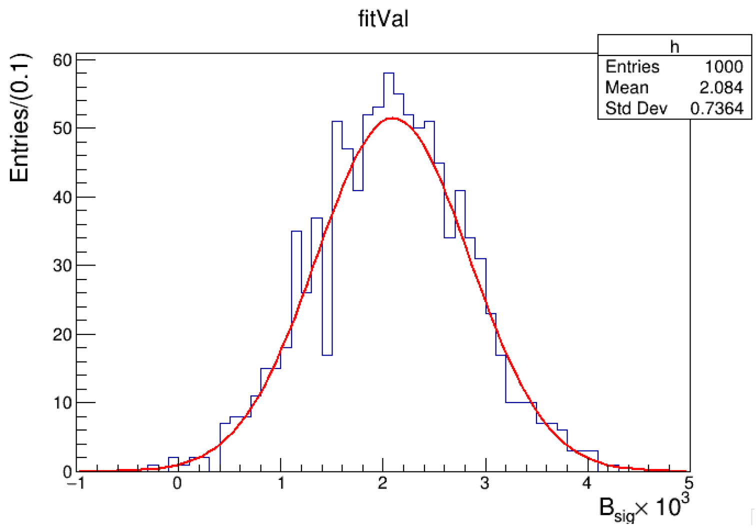


Figure 61: Distribution of fitted signal branching fraction values (in units of 10^{-3}) from 1000 signal injection toy fits with $\text{kinBDT} > 0.990, \text{isoBDT} > 0.970$ (trainMethod 2), shown as a blue histogram, along with a Gaussian PDF fitted to it, shown as the red lineshape. In these toys, the amount of DDK^* background in the data is artificially increased by 50%. No biasing effect is seen on the signal.

4.10.1 Data-MC differences

Systematic uncertainties and effects affect the final sensitivity mainly through the signal efficiency, as seen in Equation 38. The similarity of the signal and normalization modes ensure that a majority of systematic uncertainties cancel out in the ratio of efficiencies. Any residual uncertainty (not correction) has little to no effect on the final sensitivity since the fit is statistically limited.

On the other hand, systematic corrections to the central value of the efficiency (due to the MC being an imperfect approximation of the data) will linearly affect the estimated sensitivity. The estimation of such corrections has not been a part of the work done in this thesis (besides the correction of the PID response). From a practical point of view, this is not a big problem since the predicted enhancements for $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-)$ stop an order of magnitude below our estimated sensitivity. But in the future, one possible way of estimating such corrections could be to reconstruct $B^0 \to D^- D^+ K^{*0}$, $D^+ \to K^0\pi^-\pi^+\pi^+$ through the $K^{*0}\tau^+\tau^-$ selection stream. If the signal is significantly observed in data, the background subtract data could be used to derive data-MC corrections.

5 Conclusions

Anomalies observed in neutral current $b \to sl^+l^-$ and charged current $b \to cl\nu_l$ data, hinting at the violation of lepton flavour universality, have been connected to large enhancements in the rate of $b \to s\tau^+\tau^-$ currents by effective field theory analyses. Using LHCb data collected in 2016 - 2018, we estimate a sensitivity of $\mathcal{B}(B^0 \to K^{*0}\tau^+\tau^-) < 1.5 \times 10^{-3}$ at 95% C.L., using the three prong $\tau^+ \to \pi^+\pi^-\pi^+(\pi^0)\overline{\nu}_\tau$ decay mode. The data in the signal region remains blinded for the moment. The analysis would need to go through two stages of review and approval within the LHCb collaboration before it can be unblinded. This work provides an important benchmark of the LHCb experiment's sensitivity to this rare decay mode.

Appendices

A DDK^* cocktail weighting

Decay Mode	Weight
$B^0 \to D^- D^+ K^{*0}$	8.1
$B^0 \to D^{*-} D^{*+} K^{*0}$	11.8
$B^0 \to D^- D^{*+} K^{*0}$	18.7
$B^0 \to D^{*-} D^{*+} K^{*0}$	34.2
$B^0 \to \overline{D}^0 D^0 K^{*0}$	3.8
$B^0 \to \overline{D}^{*0} D^0 K^{*0}$	8.8
$B^0 \to \overline{D}^0 D^{*0} K^{*0}$	8.8
$B^0 \to \overline{D}^{*0} D^{*0} K^{*0}$	13.7

Table 23: Weighting of $B^0 \to DDK^{*0}$ decay modes in the background cocktail MC samples. These weights are in accordance with measurements of the BFs of these decays in a parallel analysis

Decay Mode	Weight
$B^+ \to \overline{D}^0 D^+ K^{*0}$	15.2
$B^+ \to \overline{D}^{*0} D^+ K^{*0}$	40.3
$B^+ \to \overline{D}^0 D^{*+} K^{*0}$	18.7
$B^+ \to \overline{D}^{*0} D^{*+} K^{*0}$	25.4

Table 24: Weighting of $B^+ \to DDK^{*0}$ decay modes in the background cocktail MC samples. These weights are in accordance with measurements of the BFs of these decays in a parallel analysis

Decay Mode	Weight
$B_s^0 \to D_s^- D^+ K^{*0}$	1.6
$B_s^0 \to D_s^{*-} D^+ K^{*0}$	3.1
$B_s^0 \to D_s^- D^{*+} K^{*0}$	6.9
$B_s^0 \to D_s^{*-} D^{*+} K^{*0}$	8.9

Table 25: Weighting of $B_s^0 \to DDK^{*0}$ decay modes in the background cocktail MC samples. These weights are in accordance with measurements of the BFs of these decays in a parallel analysis

Decay Mode	Weight
$D^{*+} \rightarrow D^0\pi^+$	0.677
$D^{*+} \rightarrow D^+\pi^0$	0.307
$D^{*+} \rightarrow D^+\gamma$	0.016

Table 26: Relevant branching fractions for D^{*+} decays in the background cocktail MC samples, obtained from the PDG

Decay Mode	Weight
$D^{*0} \rightarrow D^0\pi^0$	0.647
$D^{*0} \rightarrow D^0\gamma$	0.353

Table 27: Relevant branching fractions for D^{*0} decays in the background cocktail MC samples, obtained from the PDG

Decay Mode	Weight
$D_s^{*+} \rightarrow D_s^+\gamma$	93.5
$D_s^{*+} \rightarrow D_s^+\pi^0$	5.8
$D_s^{*+} \rightarrow D_s^+e^+e^-$	0.67

Table 28: Relevant branching fractions for D_s^{*+} decays in the background cocktail MC samples, obtained from the PDG

Decay Mode	PDG Value (%)	PDG ID	K^0 factor	Resonance BF	Weight	% of tot
$D^+ \to K^0\pi^+\eta_{2\pi x}$	1.31	$\Gamma_{71}(2021)$	2	$\mathcal{B}(\eta \to 2\pi X) = 0.271$	0.71	6.8
$D^+ \to K^0\pi^+\eta'_{2\pi x}$	0.19	$\Gamma_{72}(2021)$	2	$\mathcal{B}(\eta' \to 2\pi X) = 0.432$	0.16	1.5
$D^+ \to {K^0}_s 2\pi^+\pi^-$	3.1	$\Gamma_{74}(2021)$	2	-	6.2	-
$D^+ \to a_1^+ K^0$	1.8	$\Gamma_{68}(2008)$	2	-	3.6	34.5
$D^+ \to \overline{K}_1(1400)\pi^+$	-	[Γ_{74}(21) - Γ_{68}(08) - Γ_{73}(08)]×0.5	2	-	0.94	9.0
$D^+ \to K^{*-}\pi^+\pi^-$	-	[Γ_{74}(21) - Γ_{68}(08) - Γ_{73}(08)]×0.5	2	-	0.94	9.07
$D^+ \to K^0\pi^+\pi^+\pi^-$	0.36	$\Gamma_{73}(2008)$	2	-	0.72	6.9
$D^+ \to K^-\pi^+\pi^+\eta_{2\pi x}$	0.135	$\Gamma_{75}(2021)$	-	$\mathcal{B}(\eta \to 1\pi X) = 0.271$	0.036	0.34
$D^+ \to K^0\pi^+\pi^0\eta_{2\pi x}$	0.122	$\Gamma_{76}(2021)$	2	$\mathcal{B}(\eta \to 2\pi X) = 0.271$	0.066	0.63
$D^+ \to K^- 3\pi^+\pi^-$	0.57	$\Gamma_{77}(2021)$	-	-	0.57	-
$D^+ \to \overline{K}^{*0}\pi^+\pi^+\pi^-$	0.12	$\Gamma_{78}(2021)$	-	-	0.123	1.12
$D^+ \to a_1^+\overline{K}^{*0}$	0.23	$\Gamma_{80}(2021)$	-	-	0.235	2.25
$D^+ \to K^-\rho^0\pi^+\pi^+$	0.172	$\Gamma_{82}(2021)$	-	-	0.176	1.69
$D^+ \to K^-\pi^+\pi^+\pi^+\pi^-$	0.04	$\Gamma_{83}(2021)$	-	-	0.041	0.39
$D^+ \to a_1^+\pi^0$	1.16× 0.5	$\Gamma_{101}(2021)$×0.5	-	-	0.58	5.56
$D^+ \to \rho^+\rho^0$	1.16× 0.5	$\Gamma_{101}(2021)$×0.5	-	-	0.58	5.56
$D^+ \to 3\pi^+ 2\pi^-$	0.166	$\Gamma_{102}(2021)$	-	-	0.166	1.59
$D^+ \to \eta_{2\pi x}\pi^+$	0.377	$\Gamma_{103}(2021)$	-	$\mathcal{B}(\eta \to 2\pi X) = 0.271$	0.102	0.98
$D^+ \to \eta_{2\pi x}\pi^+\pi^0$	0.205	$\Gamma_{104}(2021)$	-	$\mathcal{B}(\eta \to 2\pi X) = 0.271$	0.056	0.54
$D^+ \to \eta\pi^+\pi^+\pi^-$	0.341	$\Gamma_{105}(2021)$	-	-	0.341	3.27
$D^+ \to \eta_{2\pi x}\pi^+\pi^0\pi^0$	0.320	$\Gamma_{106}(2021)$	-	$\mathcal{B}(\eta \to 2\pi X) = 0.271$	0.087	0.83
$D^+ \to \eta_{2\pi x}\eta_{2\pi x}\pi^+$	0.296	$\Gamma_{107}(2021)$	-	$2 \times \mathcal{B}(\eta \to 2\pi X) = 0.542$	0.160	1.53
$D^+ \to \omega_{2\pi x}\pi^+\pi^0$	0.390	$\Gamma_{109}(2021)$	-	$\mathcal{B}(\omega \to 2\pi X) = 0.908$	0.354	3.39
$D^+ \to \eta'_{2\pi x}\pi^+$	0.497	$\Gamma_{110}(2021)$	-	$\mathcal{B}(\eta' \to 2\pi X) = 0.432$	0.214	2.05
$D^+ \to \eta'_{2\pi x}\pi^+\pi^0$	0.16	$\Gamma_{111}(2021)$	-	$\mathcal{B}(\eta' \to 2\pi X) = 0.432$	0.069	0.66
Total $\mathcal{B}(D^+ \to 3\pi X)$	-	-	-	-	10.4%	-

Table 29: Branching fractions for the known ways in which D^+ decays to 3 charged pions in the background cocktail MC samples, obtained from the PDG. a_1^+ always decays according to $a_1^+ \to \rho^0\pi^+, \rho^0 \to \pi^+\pi^-$.

Decay Mode	PDG Value (%)	PDG ID	K^0 factor	Resonance BF	Weight	% of $2\pi X$	% of $3\pi X$
$D^0 \to \rho^0 K^0$	0.63	$\Gamma_{40}(2021)$	2	-	1.26	3.3	-
$D^0 \to K^0 f_0$	0.12	$\Gamma_{43}(2021)$	2	-	0.24	0.6	-
$D^0 \to K^0 f_0(1370)$	0.28	$\Gamma_{43}(2021)$	2	-	0.56	1.5	-
$D^0 \to K^{*-}\pi^+$	1.64	$\Gamma_{46}(2021)$	2	-	1.64	4.3	-
$D^0 \to K_0^*(1430)\pi^+$	0.267	$\Gamma_{47}(2021)$	2	-	0.534	1.4	-
$D^0 \to K^-2\pi^+\pi^{-0\dagger}$	8.23	$\Gamma_{71}(2021)$	-	-	8.23	-	-
$D^0 \to K^-\pi^+\rho^{0\dagger}$	0.61	$\Gamma_{73}(2021)$	-	-	0.63 (R)	1.6	4.0
$D^0 \to \overline{K}^{*0}\rho^{0\dagger}$	1.01	$\Gamma_{74}(2021)$	-	-	1.13 (R)	2.9	7.2
$D^0 \to a_1^+K^{-\dagger}$	4.32	$\Gamma_{76}(2021)$	-	-	4.47 (R)	11.6	28.4
$D^0 \to K_1^-(1270)\pi^{+\dagger}$	0.39	$\Gamma_{77}(2021)$	-	-	0.35 (R)	0.9	2.2
$D^0 \to K^-\pi^+\pi^+\pi^{-\dagger}$	1.81	$\Gamma_{81}(2021)$	-	-	1.64 (R)	4.3	10.4
$D^0 \to K^0\pi^+\pi^-\pi^0$	5.2	$\Gamma_{82}(2021)$	2	-	10.4	-	-
$D^0 \to \eta K^0, \eta \to \pi^+\pi^-\pi^0$	0.117	$\Gamma_{83}(2021)$	2	-	0.254 (R)	0.7	-
$D^0 \to \omega K^0, \omega \to \pi^+\pi^-\pi^0$	0.99	$\Gamma_{84}(2021)$	2	-	2.16 (R)	5.6	-
$D^0 \to K^{*-}\rho^+, K^{*-} \to K^0\pi^-$	2.1	$\Gamma_{71}(2008)$	2	-	4.58 (R)	11.9	-
$D^0 \to K_1(1270)^-\pi^+, K_1^- \to K^0\pi^-\pi^0$	0.22	$\Gamma_{72}(2008)$	2	-	0.48 (R)	1.2	-
$D^0 \to K^{*0}\pi^+\pi^-, K^{*0} \to K^0\pi^0$	0.24	$\Gamma_{73}(2008)$	2	-	0.52 (R)	1.4	-
$D^0 \to K^0\pi^+\pi^-\pi^0$ NR	1.1	$\Gamma_{74}(2008)$	2	-	2.4(R)	6.3	-
$D^0 \to K^-2\pi^+\pi^-\pi^{0\dagger}$	4.3	$\Gamma_{86}(2021)$	-	-	4.3	-	-
$D^0 \to (\overline{K}^{*0} \to K^-\pi^+)\pi^+\pi^-\pi^{0\dagger}$	1.3	$\Gamma_{87}(2021)$	-	-	1.3	3.4	8.3
$D^0 \to K^-\pi^+\omega, \omega \to \pi^+\pi^-\pi^{0\dagger}$	2.8-0.65	$\Gamma_{88} - \Gamma_{89}(2021)$	-	-	2.15	5.6	13.7
$D^0 \to (\overline{K}^{*0} \to K^-\pi^+)(\omega \to \pi^+\pi^-\pi^{0\dagger})$	0.65	$\Gamma_{89}(2021)$	-	-	0.65	1.7	4.1
$D^0 \to K^-\pi^+\pi^+\pi^-\pi^{0\dagger}$	-	-	-	-	0.2	0.5	1.3
$D^0 \to K^0\eta_{2\pi x}\pi^0$	1.01	$\Gamma_{90}(2021)$	2	$\mathcal{B}(\eta \to 2\pi X) = 0.271$	0.54	1.4	-
$D^0 \to K^-\pi^+\eta_{2\pi x}^\dagger$	1.88	$\Gamma_{93}(2021)$	-	$\mathcal{B}(\eta \to 2\pi X) = 0.271$	0.51	1.3	3.2
$D^0 \to K^-\pi^+\pi^0\eta_{2\pi x}^\dagger$	0.449	$\Gamma_{97}(2021)$	-	$\mathcal{B}(\eta \to 2\pi X) = 0.271$	0.12	0.3	0.8
$D^0 \to K^0\pi^+\pi^-\eta_{2\pi x}^\dagger$	0.28	$\Gamma_{98}(2021)$	2	$\mathcal{B}(\eta \to 2\pi X) = 0.271$	0.15	0.4	0.9
$D^0 \to K^0\pi^+\pi^-\eta_{not2\pi x}$	0.28	$\Gamma_{98}(2021)$	2	$\mathcal{B}(\eta! \to 2\pi X) = 0.729$	0.41	1.1	-
$D^0 \to K^0\rho^0\pi^+\pi^{-\dagger}$	0.11	$\Gamma_{101}(2021)$	2	-	0.22	0.6	1.4
$D^0 \to (K^{*-} \to K^0\pi^-)\rho^0\pi^{+\dagger}$	0.16	$\Gamma_{103}(2021)$	2	-	0.32	0.83	2.0
$D^0 \to K^-\pi^+\eta'^{\dagger}{}_{2\pi x}$	0.643	$\Gamma_{114}(2021)$	-	$\mathcal{B}(\eta' \to 2\pi X) = 0.432$	0.28	0.7	1.8
$D^0 \to K^0\eta'_{2\pi x}\pi^0$	0.252	$\Gamma_{115}(2021)$	2	$\mathcal{B}(\eta' \to 2\pi X) = 0.432$	0.22	0.6	-
$D^0 \to (\rho^+ \to \pi^+\pi^0)\pi^-$	1.01	$\Gamma_{134}(2021)$	-	-	1.01	2.6	-
$D^0 \to (\rho^0 \to \pi^+\pi^-)\pi^-$	0.386	$\Gamma_{135}(2021)$	-	-	0.386	1.0	-
$D^0 \to (\rho^- \to \pi^-\pi^0)\pi^+$	0.515	$\Gamma_{136}(2021)$	-	-	0.515	1.3	-
$D^0 \to 2\pi^+2\pi^{-\dagger}$	0.117	$\Gamma_{151} - \Gamma_{152} - \Gamma_{164}(2021)$	-	-	0.117	0.3	0.7
$D^0 \to a_1^+\pi^{-\dagger}$	0.454	$\Gamma_{152}(2021)$	-	-	0.454	1.2	2.9
$D^0 \to 2(\rho^0 \to \pi^+\pi^-)^\dagger$	0.185	$\Gamma_{164}(2021)$	-	-	0.185	0.5	1.2
$D^0 \to \pi^+\pi^-2\pi^0$	1.02	$\Gamma_{178}(2021)$	-	-	1.02	2.7	-
$D^0 \to 2\pi^+2\pi^-\pi^{0\dagger}$	0.42	$\Gamma_{182}(2021)$	-	-	0.42	1.1	2.7
$D^0 \to \pi^+\pi^-\pi^0\eta_{2\pi x}^\dagger$	0.323	$\Gamma_{187}(2021)$	-	$\mathcal{B}(\eta \to 2\pi X) = 0.271$	0.09	0.2	0.6
$D^0 \to \pi^+\pi^-\pi^0\eta_{not2\pi x}$	0.323	$\Gamma_{187}(2021)$	-	$\mathcal{B}(\eta! \to 2\pi X) = 0.729$	0.23	0.6	-
$D^0 \to K^+K^-\pi^+\pi^-$	0.247	$\Gamma_{230}(2021)$	-	-	0.247	0.6	-
$D^0 \to K^0K^0\pi^+\pi^-$	0.053	$\Gamma_{257}(2021)$	4	-	0.212	0.6	-
$D^0 \to K^+K^-\pi^+\pi^-\pi^0$	0.31	$\Gamma_{261}(2021)$	-	-	0.261	0.7	-
Total $\mathcal{B}(D^0 \to 2\pi X)$	-	-	-	-	37.4%	-	-
Total $\mathcal{B}(D^0 \to 3\pi X)$	-	-	-	-	15.4%	-	-

Table 30: Branching fractions for the known ways in which D^+ decays to 2 or 3 charged pions in the background cocktail MC samples, obtained from the PDG. a_1^+ always decays according to $a_1^+ \to \rho^0\pi^+, \rho^0 \to \pi^+\pi^-$. Modes marked with a † are $D^0 \to 3\pi X$ modes. Others are $D^0 \to 2\pi X$ modes. (R) indicates a rescaling applied to the PDG value, in order to have sub-modes of a decay sum to the inclusive branching fraction.

Decay Mode	PDG Value (%)	PDG ID	K^0 factor	Resonance BF	Weight
$D_s^+ \to a_1^+ \phi$	0.86	$\Gamma_{58}(2021)$	-	-	0.86
$D_s^+ \to 2K^0 2\pi^+\pi^-$	0.084	$\Gamma_{62}(2021)$	4	-	0.336
$D_s^+ \to \eta_{2\pi x}\pi^+$	1.68	$\Gamma_{74}(2021)$	-	$\mathcal{B}(\eta \to 2\pi X) = 0.271$	0.46
$D_s^+ \to \omega_{2\pi x}\pi^+$	0.192	$\Gamma_{75}(2021)$	-	$\mathcal{B}(\omega \to 2\pi X) = 0.908$	0.17
$D_s^+ \to 3\pi^+ 2\pi^-$	0.79	$\Gamma_{76}(2021)$	-	-	0.79
$D_s^+ \to \rho^+ \eta_{2\pi x}$	8.15	$\Gamma_{78}(2021)$ mod	-	$\mathcal{B}(\eta \to 2\pi X) = 0.271$	2.21
$D_s^+ \to (a_0^+ \to \eta_{2\pi x}\pi^+)\pi^0$	2.2×0.5	$\Gamma_{81}(2021)$ $\times 0.5$	-	$\mathcal{B}(\eta \to 2\pi X) = 0.271$	0.3
$D_s^+ \to (a_0^0 \to \eta_{2\pi x}\pi^z)\pi^+$	2.2×0.5	$\Gamma_{81}(2021)$ $\times 0.5$	-	$\mathcal{B}(\eta \to 2\pi X) = 0.271$	0.3
$D_s^+ \to \omega_{2\pi x}\pi^+\pi^0$	2.8	$\Gamma_{82}(2021)$	-	$\mathcal{B}(\omega \to 2\pi X) = 0.908$	2.54
$D_s^+ \to 3\pi^+ 2\pi^- \pi^0$	-	$\Gamma_{83} - \Gamma_{84} \times \mathcal{B}(\omega \to \pi^+\pi^-\pi^0) - \Gamma_{85} \times \mathcal{B}(\eta' \to \pi^+\pi^-\eta) \times \mathcal{B}(\eta \to \pi^+\pi^-\pi^0)(2021)$	-	-	3.09
$D_s^+ \to \omega 2\pi^+\pi^-$	1.6	$\Gamma_{84}(2021)$	-	-	1.6
$D_s^+ \to \eta'_{2\pi x}\pi^+$	3.94	$\Gamma_{85}(2021)$	-	$\mathcal{B}(\eta' \to 2\pi X) = 0.432$	1.7
$D_s^+ \to \rho^+ \eta'_{2\pi x}$	5.8	$\Gamma_{88}(2021)$	-	$\mathcal{B}(\eta' \to 2\pi X) = 0.432$	2.5
$D_s^+ \to K^0\pi^+\pi^-$	0.3	$\Gamma_{104}(2021)$	2	-	0.6
Total $\mathcal{B}(D_s^+ \to 3\pi X)$	-	-	-	-	17.4%

Table 31: Branching fractions for the known ways in which D_s^+ decays to 3 charged pions in the background cocktail MC samples, obtained from the PDG.

B Isolation BDT input distributions

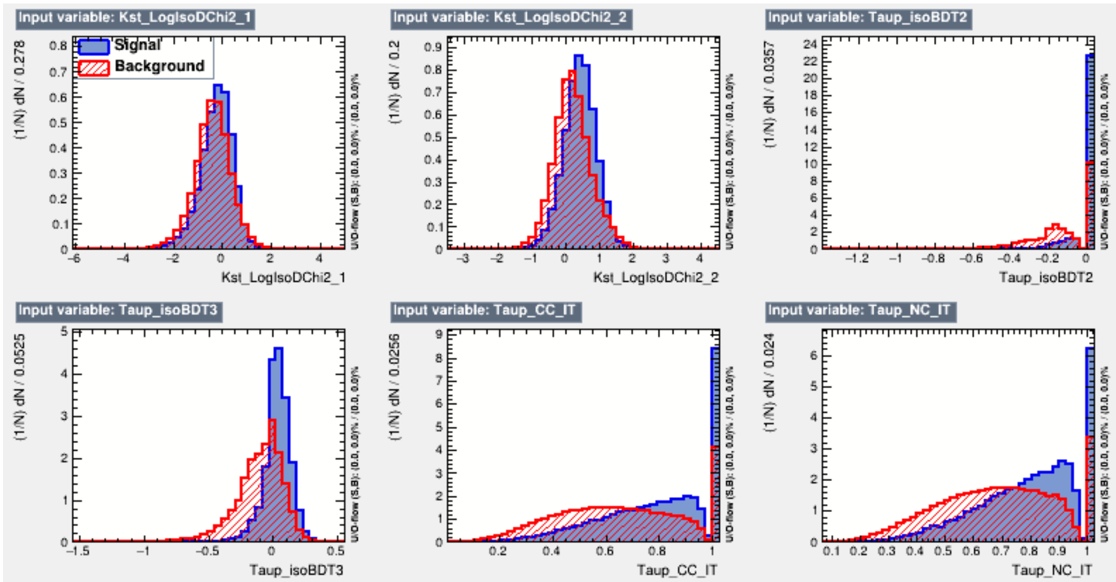


Figure 62: Distributions of the features (part 1 of 3) used to train the isolation BDT with trainMethod1.

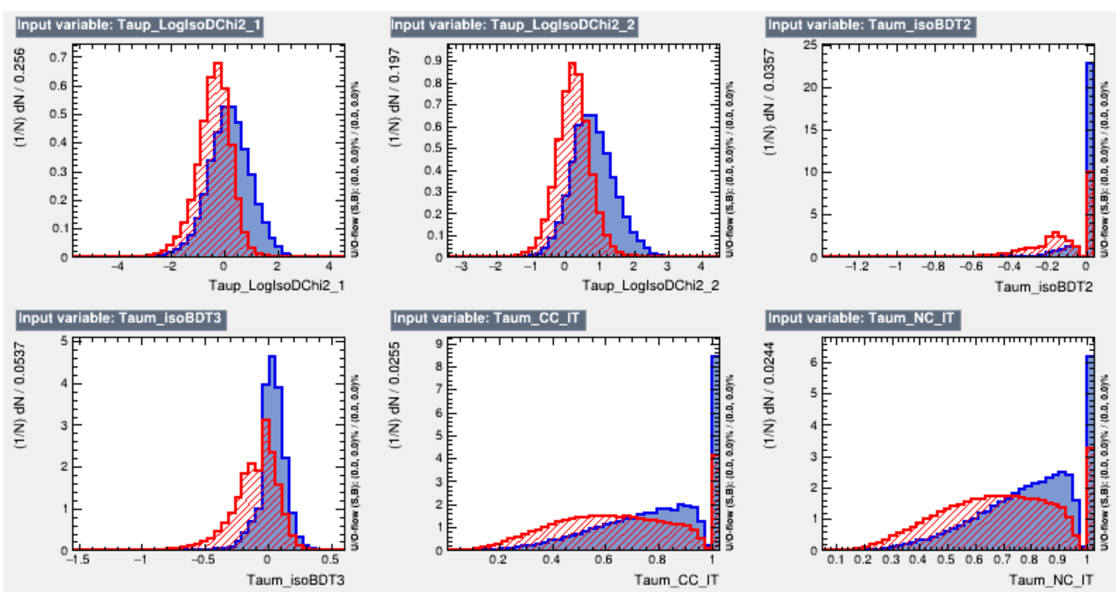


Figure 63: Distributions of the features (part 2 of 3) used to train the isolation BDT with trainMethod1.

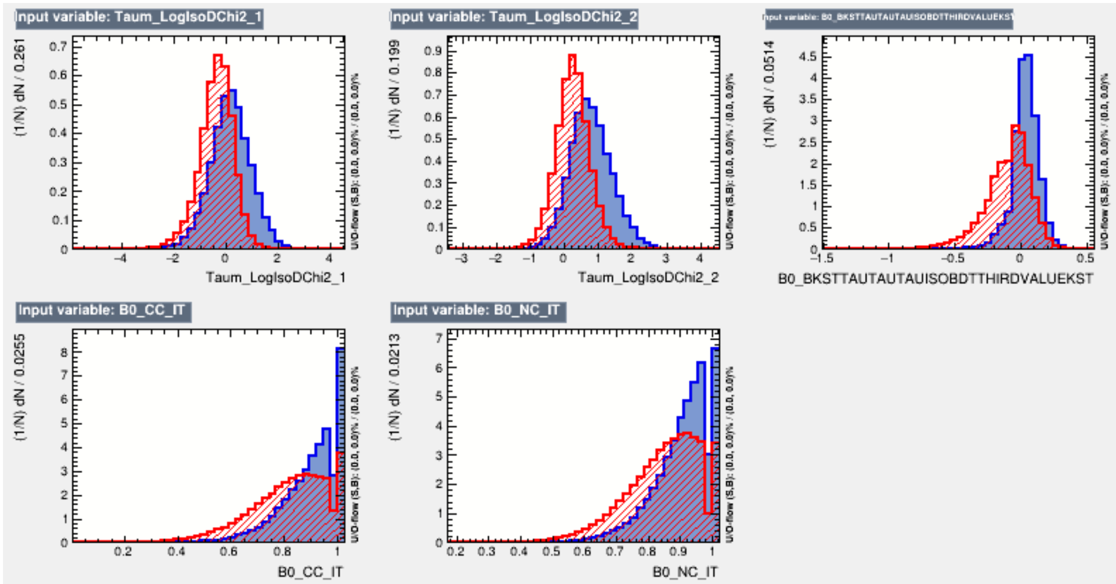


Figure 64: Distributions of the features (part 3 of 3) used to train the isolation BDT with trainMethod1.

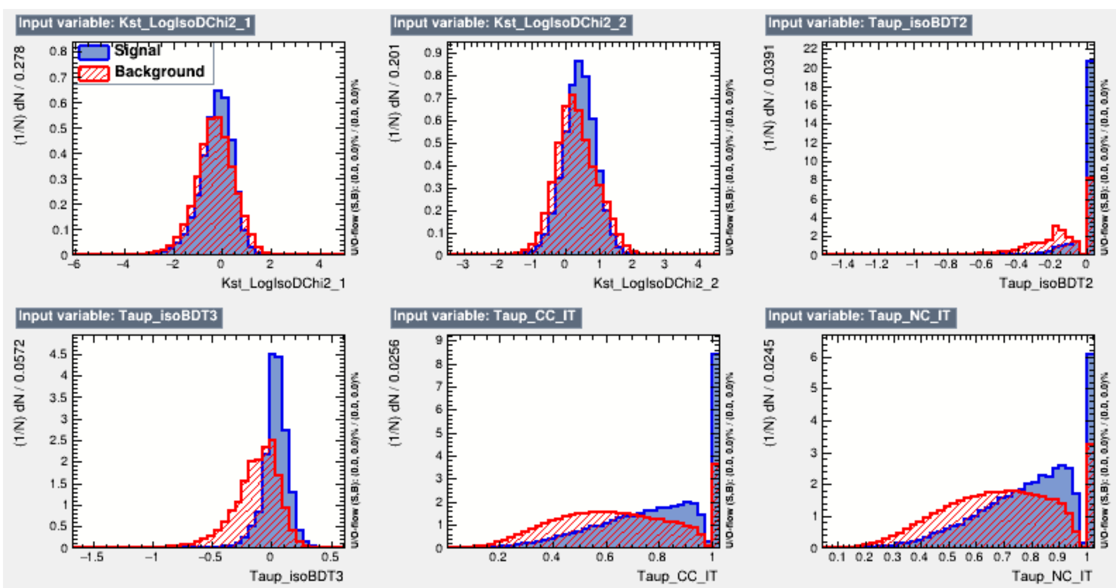


Figure 65: Distributions of the features (part 1 of 3) used to train the isolation BDT with trainMethod2.

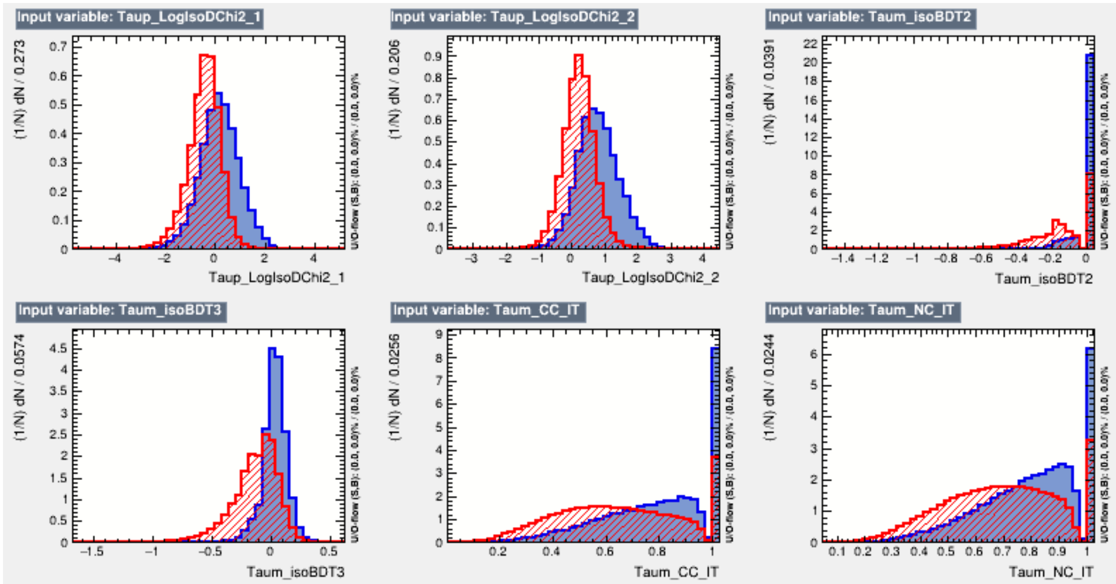


Figure 66: Distributions of the features (part 2 of 3) used to train the isolation BDT with trainMethod2.

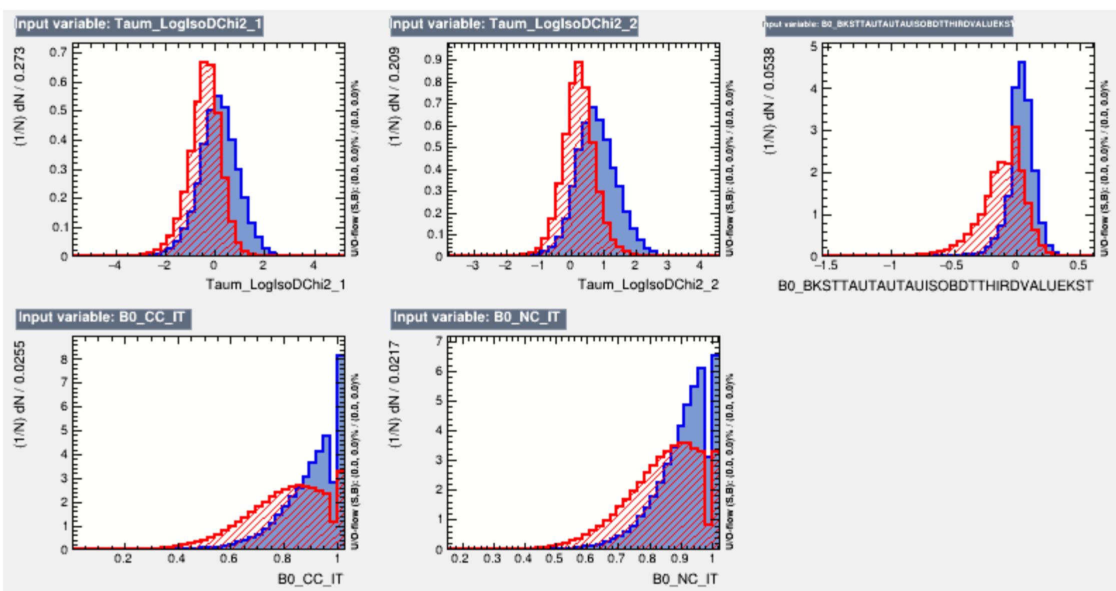


Figure 67: Distributions of the features (part 3 of 3) used to train the isolation BDT with trainMethod2.

C Kinematic BDT input distributions

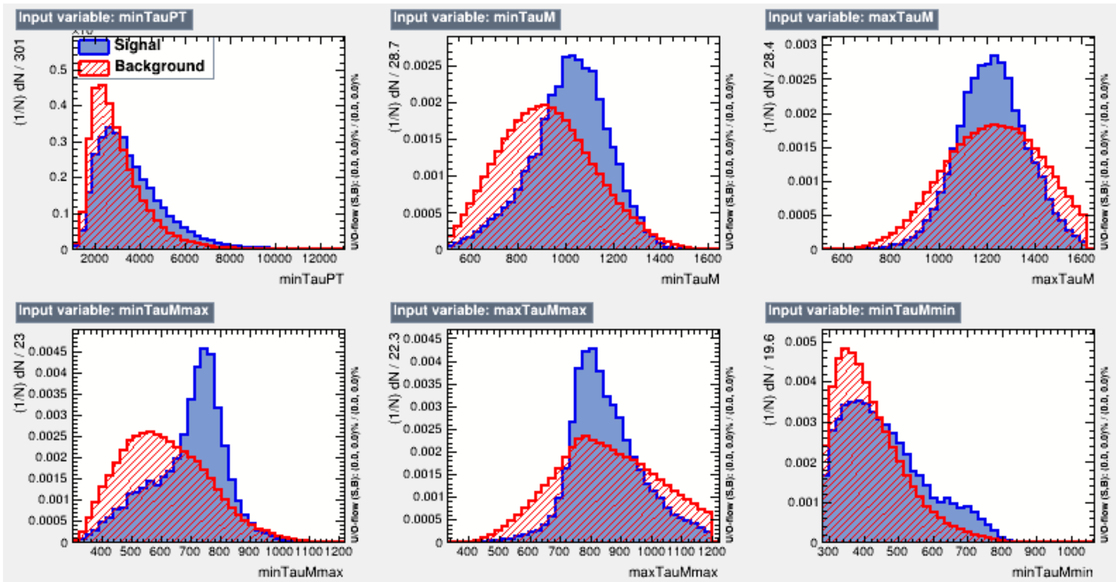


Figure 68: Distributions of the features (part 1 of 5) used to train the kinematic BDT with trainMethod1.

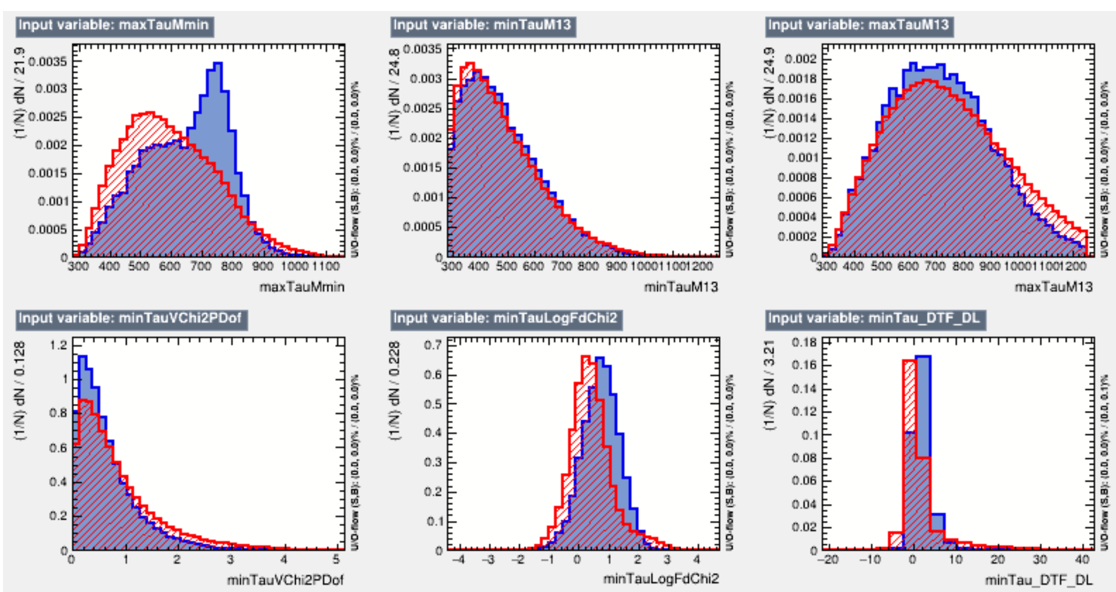


Figure 69: Distributions of the features (part 2 of 5) used to train the kinematic BDT with trainMethod1.

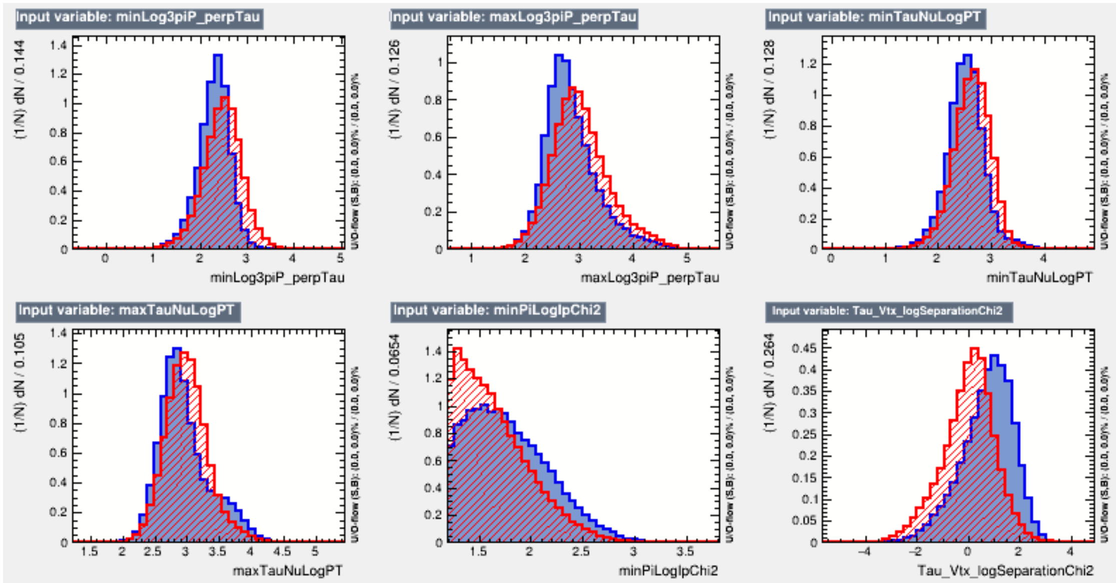


Figure 70: Distributions of the features (part 3 of 5) used to train the kinematic BDT with trainMethod1.

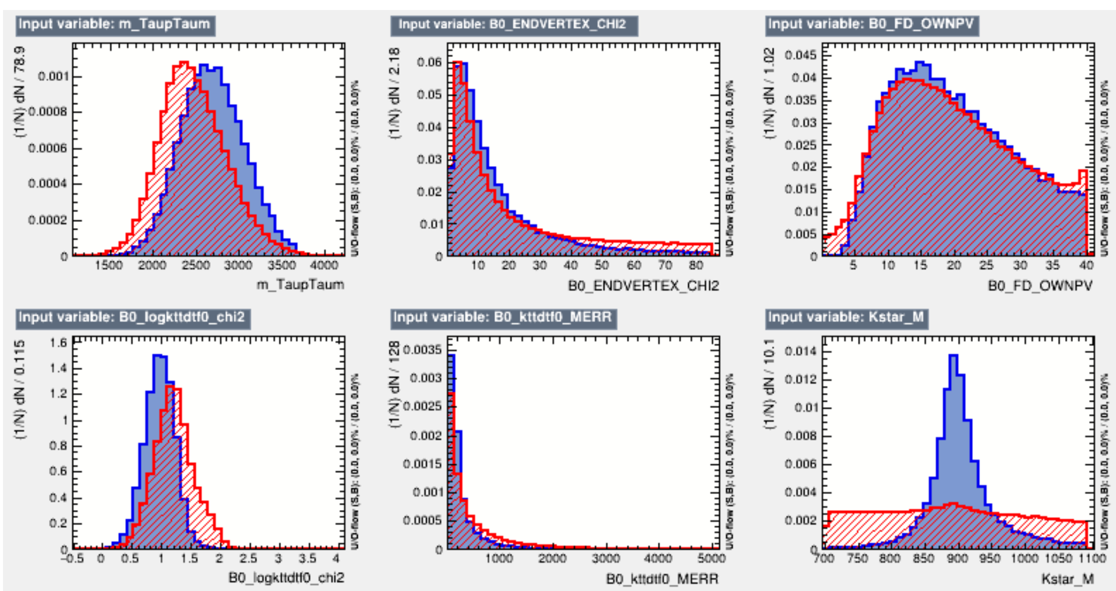


Figure 71: Distributions of the features (part 4 of 5) used to train the kinematic BDT with trainMethod1.

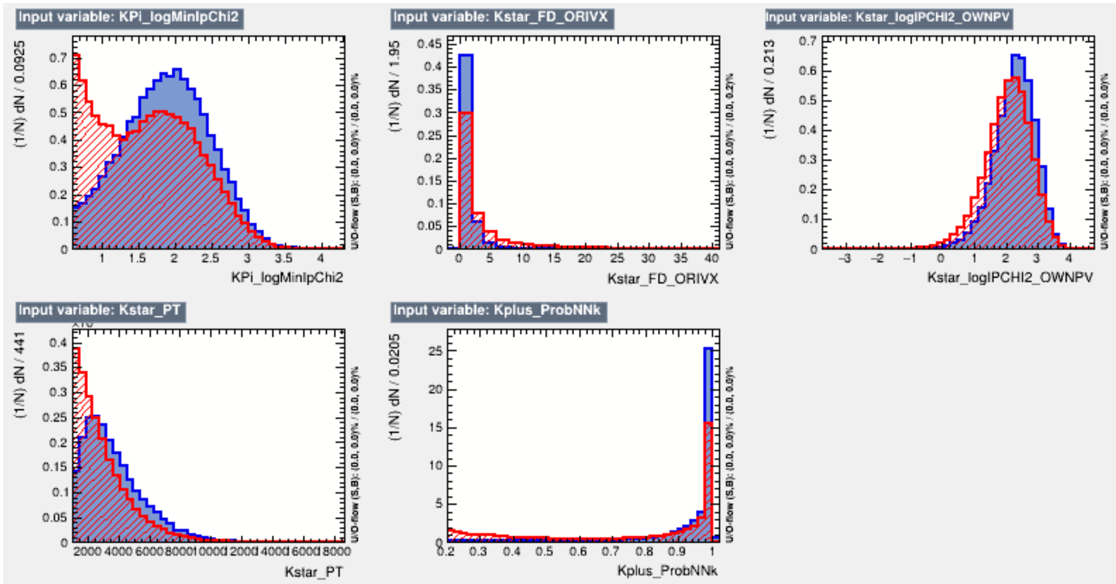


Figure 72: Distributions of the features (part 5 of 5) used to train the kinematic BDT with trainMethod1.

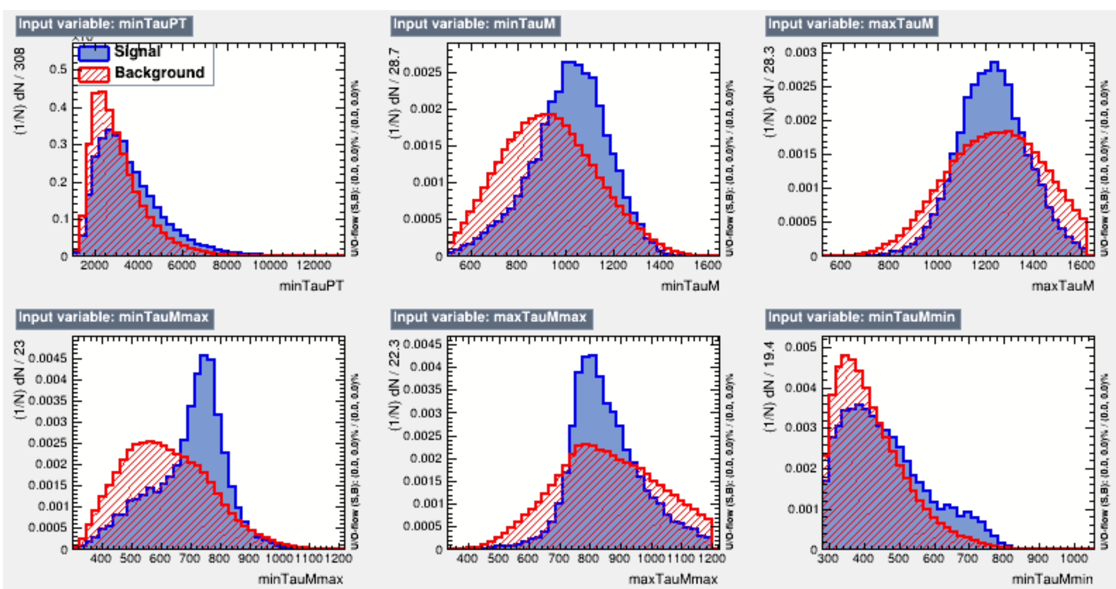


Figure 73: Distributions of the features (part 1 of 5) used to train the kinematic BDT with trainMethod2.

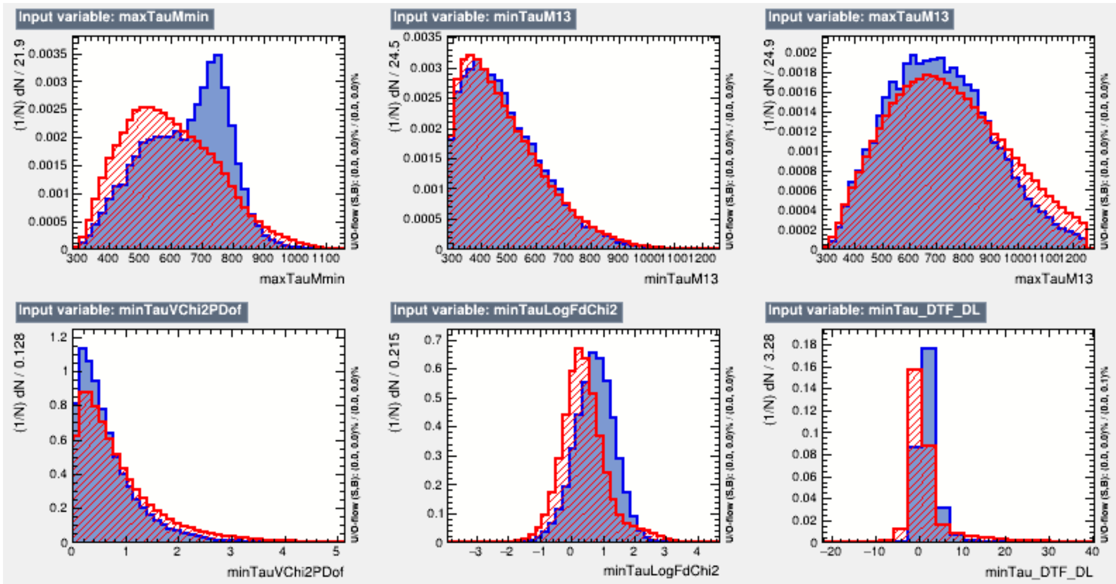


Figure 74: Distributions of the features (part 2 of 5) used to train the kinematic BDT with trainMethod2.

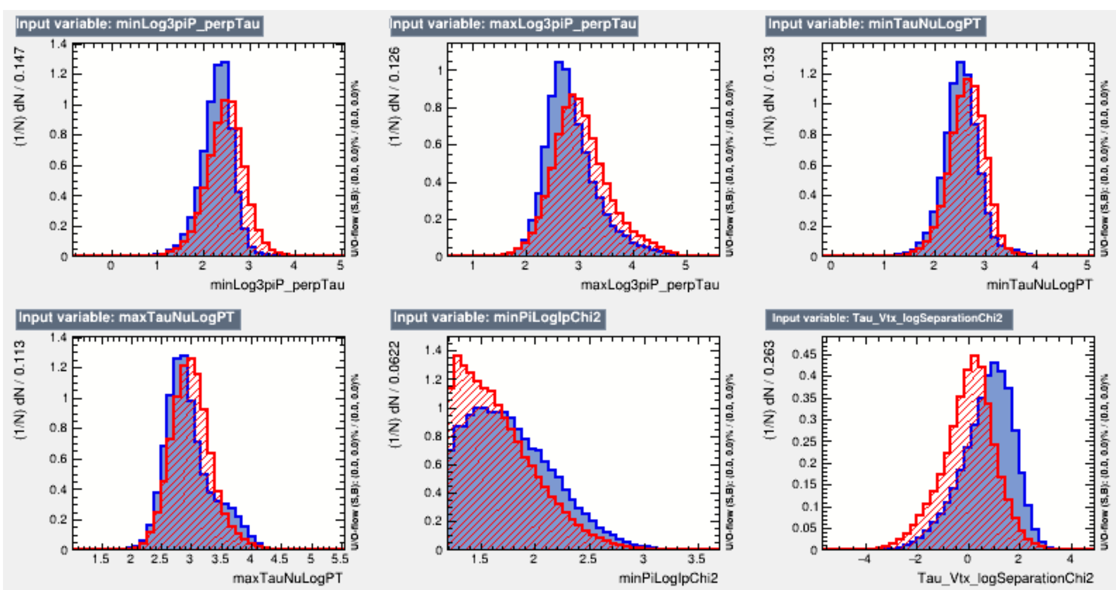


Figure 75: Distributions of the features (part 3 of 5) used to train the kinematic BDT with trainMethod2.

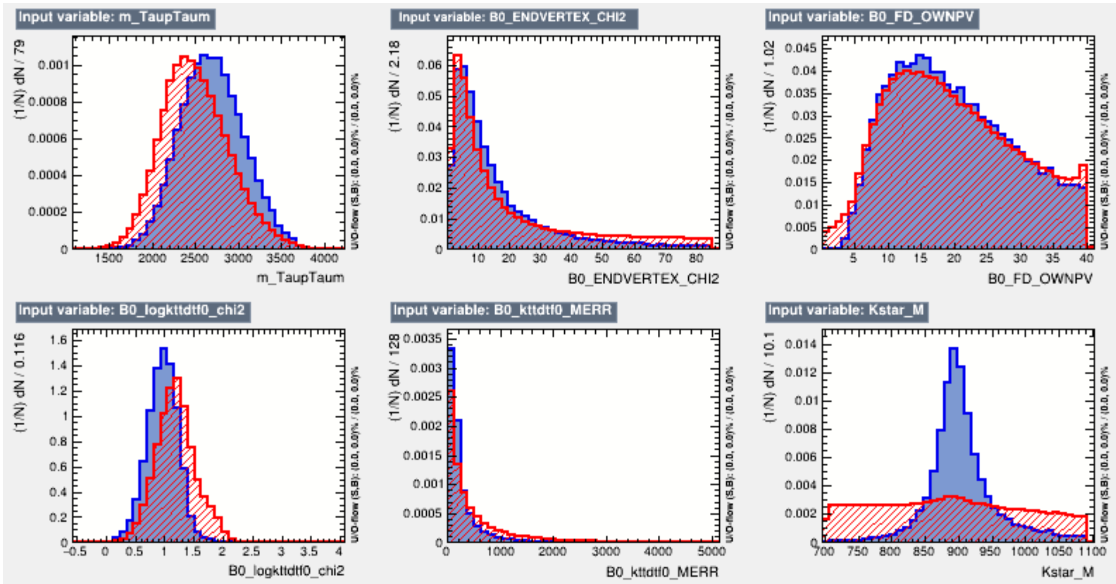


Figure 76: Distributions of the features (part 4 of 5) used to train the kinematic BDT with trainMethod2.

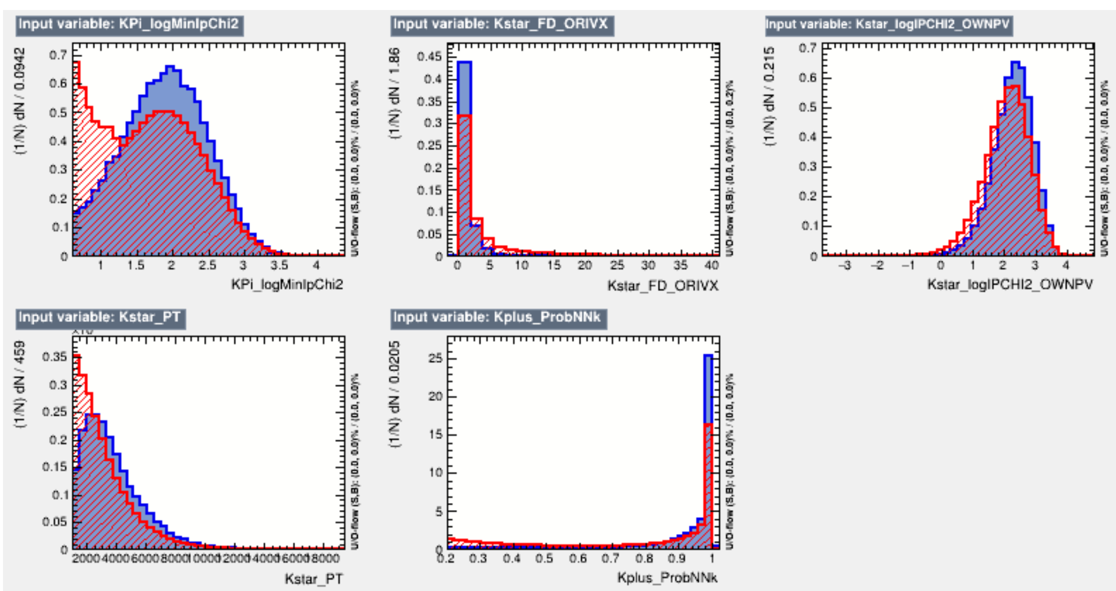


Figure 77: Distributions of the features (part 5 of 5) used to train the kinematic BDT with trainMethod2.

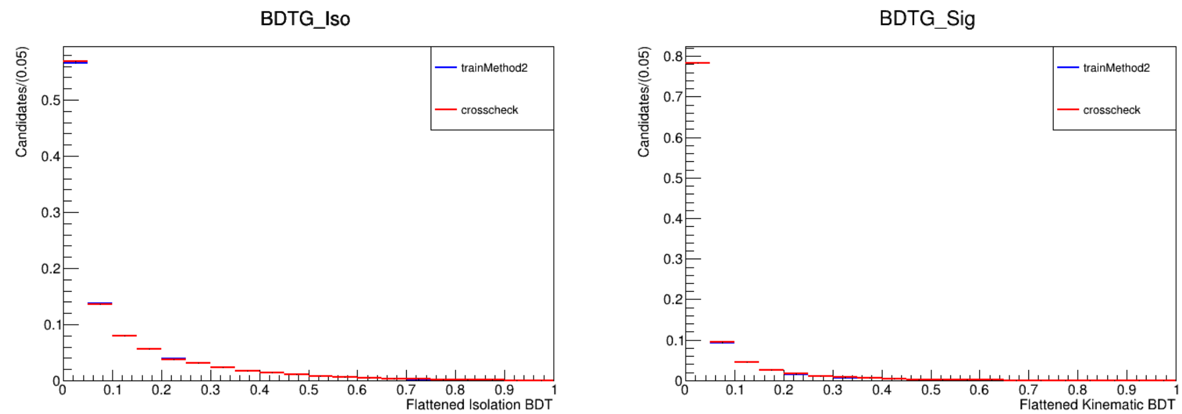


Figure 78: A comparison (between trainMethod2 [blue] and the crosscheck BDT [red]) of the flatted BDT response is shown for the isolation BDT (left) and kinematic BDT(right). The compatibility of the responses demonstrates that the trainMethod2 BDT is not biased by any potential signal that might be present in its background training sample.

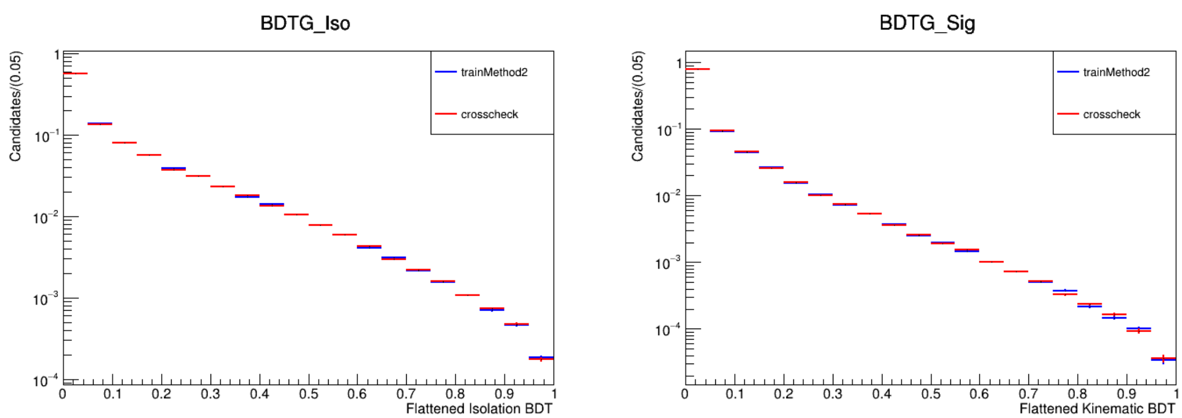


Figure 79: A comparison (between trainMethod2 [blue] and the crosscheck BDT [red]) of the flatted BDT response is shown for the isolation BDT (left) and kinematic BDT(right), with the Y axis in log scale. The compatibility of the responses demonstrates that the trainMethod2 BDT is not biased by any potential signal that might be present in its background training sample.

D BDT flattening and comparison

We wish to compare the response of the isolation and kinematic BDTs trained with trainMethod2 with the responses of the crosscheck versions of the BDTs (which have been trained with 5% of the signal sample artificially added to the background sample, to emulate the effect of signal contamination in the background training data). The BDT response by itself it not a meaningful distribution to compare for a single species.

In order to make a meaningful comparison of the BDT responses, we apply a "flattenning" transformation to them, whereby the BDT response is transformed such that the signal response is flat. This is done by finding the boundaries in the original BDT signal response (distributed between -1 and 1) which divide the signal equally into 5% chunks. With this information in hand, we can take the BDT response on a candidate in data, query which two boundaries it falls between, and assign it to the appropriate flattenned BDT bin. For example, if it fell within the first 5% BDT chunk, it would be assigned a "flattenned" BDT value of 0.025 so that it fell in the $0 - 0.05$ bin of the flattenned BDT response, which is now distributed between 0 and 1. The BDT response on data, transformed in this manner, is one that can be meaningfully compared.

E Isospin amplitudes in $\mathit{\Lambda}_b^0 \to J/\psi\Lambda(\Sigma^0)$ and $\mathit{\Xi}_b \to J/\psi\Xi(\Lambda)$

This section is taken verbatim from a paper [78] published by the LHCb collaboration, based on research conducted by the author with Sheldon Stone and Michael Wilkinson. It is provided here as a record of work done by the author in graudate school.

Measurements of ratios of isospin amplitudes A_i (i denotes the final state isospin) in hadronic weak decays are a sensitive way to probe the interplay between strong and weak interactions. Such ratios can also reveal the presence of non-Standard Model amplitudes. For example, in $K \to \pi\pi$ decays the experimentally determined ratio $|A_0/A_2| \approx 22.5$ has not been understood for over 50 years [79]. Recent models of the strong dynamics [80] and lattice gauge calculations [81] for these decays give only partial explanations. Determinations of isospin amplitudes from $D \to \pi\pi$ and $B \to \pi\pi$ decays, using input from other two-body decays into light hadrons, found $|A_0/A_2| \approx 2.5$ [82], and $|A_0/A_2| \approx 1.0$ [83], respectively.

In this Letter, we investigate $\Lambda_b^0 \to J/\psi\Lambda(\Sigma^0)$ and $\Xi_b^0 \to J/\psi\Xi^0(\Lambda)$ decays. (Mention of a specific decay implies the use of its charge-conjugate as well.) The leading order Feynman diagrams for all four processes are shown in Fig. 80. The isospins of the J/ψ meson and Λ baryon are zero, and that of the Σ^0 baryon is one. The isospin of the Λ_b^0 baryon is predicted by the quark model to be zero. Since the $b \to c\bar{c}s$ weak operator involves no isospin change, if this prediction is correct, we expect a dominant A_0 amplitude and a preference for the $J/\psi\Lambda$ final state over $J/\psi\Sigma^0$, which proceeds via the A_1 amplitude. Isospin breaking effects are possible due to the difference in mass and charge of the u and d quarks and can also be induced by QED, electroweak-penguin, or new physics processes [84]. If the Λ_b^0 baryon comprises a ud diquark such effects should be small. Mixing of the Λ and Σ^0 baryons is also predicted to be small, $\sim 1^\circ$, and could contribute ~ 0.01 to the $|A_1/A_0|$ amplitude ratio [85]. A severely suppressed $J/\psi\Sigma^0$ final state would determine the isospin of the Λ_b^0 baryon to be zero. Some previous LHCb analyses of Λ_b^0 decays made assumptions concerning isospin amplitudes. For instance, the pentaquark analysis, using the $\Lambda_b^0 \to J/\psi K^- p$ channel [86], assumed that the A_0 amplitude was dominant, and in the measurement of $|V_{ub}/V_{cb}|$ using $\Lambda_b^0 \to p\mu^-\overline{\nu}$ decays [87] the $A_{3/2}$ amplitude was assumed to be much smaller than the $A_{1/2}$ amplitude.

In $\Xi_b^0 \to J/\psi\Xi^0(\Lambda)$ decays, taking the Ξ_b isospin as 1/2, the final state results from an isospin change of zero (1/2) and has $A_i = A_{1/2}$ (A_0). In the reaction resulting in a final state Λ baryon, the weak transition changes isospin due to the $b \to c\bar{c}d$ rather than the $b \to c\bar{c}s$ transition. Here we investigate if the larger isospin change is suppressed, or if the decay amplitude is independent of the isospin change. Note that we measure the decay $\Xi_b^- \to J/\psi\Xi^-$ for two purposes: as a proxy for $\Xi_b^0 \to J/\psi\Xi^0$, which is difficult for us to measure, and to determine the background in $J/\psi\Lambda$ mass spectrum from these decays where $\Xi \to \Lambda\pi$.

The LHCb detector is a single-arm forward spectrometer covering the pseudorapidity range $2 < \eta < 5$, described in detail in Refs. [88, 89]. The trigger [90] consists of a hardware stage, based on information from the calorimeter and muon systems, followed by a software stage, which reconstructs charged particles. Natural units are used here with $c = \hbar = 1$. We use data collected by the LHCb detector, corresponding to 1.0 fb^{-1} of integrated luminosity in 7 TeV pp collisions, 2.0 fb^{-1} at 8 TeV, and 5.5 fb^{-1} collected at 13 TeV. Hereafter, the data recorded at 7 and 8 TeV is referred to as Run 1 and the data recorded at 13 TeV is referred to as Run 2.

Simulation is required to model the effects of the detector acceptance and selection requirements. We generate pp collisions using PYTHIA [64]with a specific LHCb configuration [65]. Decays of unstable particles are described by EVTGEN [91], where final-state radiation is generated using PHOTOS [92]. The interaction of the particles with the detector, and its response, are implemented using the GEANT4 toolkit [68]as described in Ref. [63]. The lifetimes for the Λ_b^0 and Ξ_b^- baryons are taken as 1.473 and 1.572 ps [93], respectively. All simulations are performed separately for Run 1 and Run 2.

Our strategy is to fully reconstruct the $J/\psi\Lambda$ final state and partially reconstruct the $J/\psi\Sigma^0$ mode by ignoring the photon from the $\Sigma^0 \to \gamma\Lambda$ decay, because of the low efficiency of the calorimeter at small photon energies. For these decays the $J/\psi\Lambda$ mass distribution is almost uniform in the mass range 5350–5620 MeV. We simulate its shape and then fit the mass distribution to ascertain its size. The J/ψ meson is reconstructed through the $J/\psi \to \mu^+\mu^-$ decay. Candidates are formed by combining two oppositely charged tracks identified as muons, with transverse momentum $p_{\mathrm{T}} > 550$ MeV. Each of the two muons are required to have a maximal χ^2 of distance of closest approach of 30 and are also required to form a vertex with $\chi^2_{\mathrm{vtx}} < 16$. The J/ψ candidate is required to have a decay length significance from every

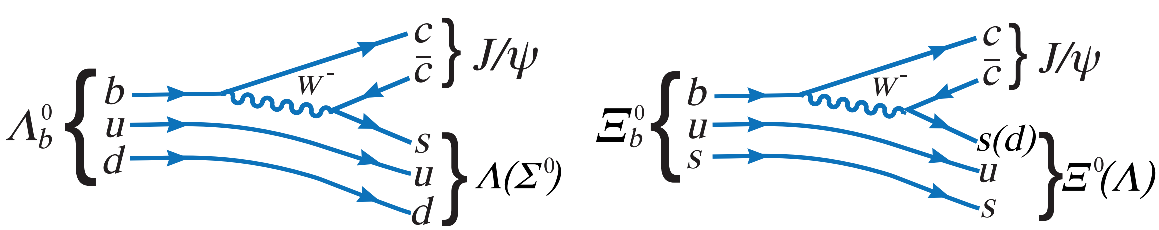


Figure 80: Leading order Feynman diagrams for $\Lambda_b^0 \to J/\psi\Lambda(\Sigma^0)$ and $\Xi_b^0 \to J/\psi\Xi^0(\Lambda)$ decays.

primary vertex, PV, of greater than 3 and a mass in the range 3049–3140 MeV.

Candidate Λ baryons are formed from a pair of identified proton and π^- particles, each with momentum greater than 2 GeV. Due to their long lifetime and high boost, a majority of the Λ baryons decay after the vertex detector. However, we use only putative decays that occur inside the vertex detector. Each of the two tracks must be inconsistent with having originated from a PV, have a maximal χ^2 of distance of closest approach of 30, form a vertex with $\chi^2_{\text{vtx}} < 12$ that is separated from that PV by more than 3 standard deviations, and have a mass between 1105 and 1124 MeV. In addition, we eliminate candidates that when interpreted as $\pi^+\pi^-$ fall within 7.5 MeV of the known K^0_S mass. Candidate $\Xi^- \to \Lambda\pi^-$ decays are reconstructed using the criteria in Ref. [94], with the additional requirement that the Ξ^- decays in the LHCb vertex detector. These are combined with selected J/ψ mesons to form candidate Ξ_b^- baryons.

We improve the $J/\psi\Lambda$ mass resolution by constraining the J/ψ and Λ candidates to their known masses and their decay products to originate from each of the relevant decay vertices; we also constrain the J/ψ and the Λ candidates to come from the same decay point [95].

After these selections, we use two boosted decision trees (BDT) [96, 97] implemented in the TMVA toolkit [98] to further separate signal from background. The first BDT is trained to reject generic $b \to J/\psi X$ decays where X contains one or more charged tracks. We train this "isolation" BDT using the following information: the χ^2_{IP} of additional charged tracks with respect to the J/ψ vertex, where χ^2_{IP} is defined as the difference in the χ^2_{vtx} of the J/ψ vertex reconstructed with and without the track being considered; the χ^2_{vtx} of the vertex formed by the J/ψ plus each additional track; the minimum χ^2_{IP} of the additional track with respect to any PV; and the p_{T} of the additional track. For the isolation BDT training, we use samples of $\Lambda_b^0 \to J/\psi\Lambda$ and $B^- \to J/\psi K^-$ candidates for the signal and background models, respectively. Both samples are background subtracted using the *sPlot* technique [99]. The output of the isolation BDT is used as an input variable in the final BDT.

The twenty discrimination variables used in the final BDT are listed in the Supplemental material. These mostly exploit the topology of the decay using the vertexing properties of the J/ψ, Λ, and Λ_b^0 candidates, and particle identification of their decay products. The signal sample again is background-subtracted $\Lambda_b^0 \to J/\psi\Lambda$ combinations. For background training we use candidates in the upper sideband with $J/\psi\Lambda$ masses between $5.7-6.0\,\mathrm{GeV}$, excluding events in $5.77-5.81\,\mathrm{GeV}$ to avoid including $\Xi_b^0 \to J/\psi\Lambda$ decays in the background sample. We use k-folding cross validation with five folds in both BDTs, to avoid any possible bias [100]. The final BDT selection is optimized to maximize the Punzi figure of merit, $\epsilon_s/(\sqrt{B}+1.5)$ [101], where ϵ_s is the efficiency of the final BDT selection on simulated $\Lambda_b^0 \to J/\psi\Sigma^0$ decays and B is the number of background candidates in the above defined sideband that pass the BDT requirement, scaled to the width of the $J/\psi\Sigma^0$ signal window. The analysis is performed separately on Run 1 and Run 2 data. The resulting $J/\psi\Lambda$ mass spectrum for Run 2 data is shown in Fig. 81. The Run 1 mass distribution is similar and is shown in the Supplemental material.

There are two signal peaks evident in the mass distribution in Fig. 81. The larger is due to $\Lambda_b^0 \to J/\psi\Lambda$ decays, and the smaller corresponds to $\Xi_b^0 \to J/\psi\Lambda$ decays. The latter is a heretofore unobserved Cabibbo-suppressed decay. The Run 1 and 2 mass distribution data are fit jointly to determine the $\Lambda_b^0 \to J/\psi\Lambda$, $\Lambda_b^0 \to J/\psi\Sigma^0$ and $\Xi_b^0 \to J/\psi\Lambda$ yields. The $\Lambda_b^0 \to J/\psi\Sigma^0$ signal is modeled using a Gaussian kernel [102] shape fit to simulation. The $\Lambda_b^0 \to J/\psi\Lambda$ signal is described by a Hypatia function, whose tail parameters are fixed from simulation, with the mass and width allowed to vary in the fit to the data. [103]. The $\Xi_b^0 \to J/\psi\Lambda$ peak is fit to the same shape but with its mean constrained to the fitted Λ_b^0 mass plus the known $\Xi_b^0 - \Lambda_b^0$ mass difference of $172.5\,\mathrm{MeV}$ [93].

While most of the candidates above the Λ_b^0 peak are the result of combinatoric background, those below are due to additional sources. One is due to $\Lambda_b^0 \to J/\psi\Lambda^*$ decays, with $\Lambda^* \to \Sigma^0\pi^0$ and $\Sigma^0 \to \gamma\Lambda$. Here, Λ^* denotes strange-baryon resonances ranging from 1405 MeV to 2350 MeV in mass. Another source comprises partially reconstructed $\Lambda_b^0 \to \psi(2S)\Lambda$ decays, where $\psi(2S) \to \pi\pi J/\psi$. These decays mainly populate masses lower than the $\Lambda_b^0 \to J/\psi\Sigma^0$ signal,

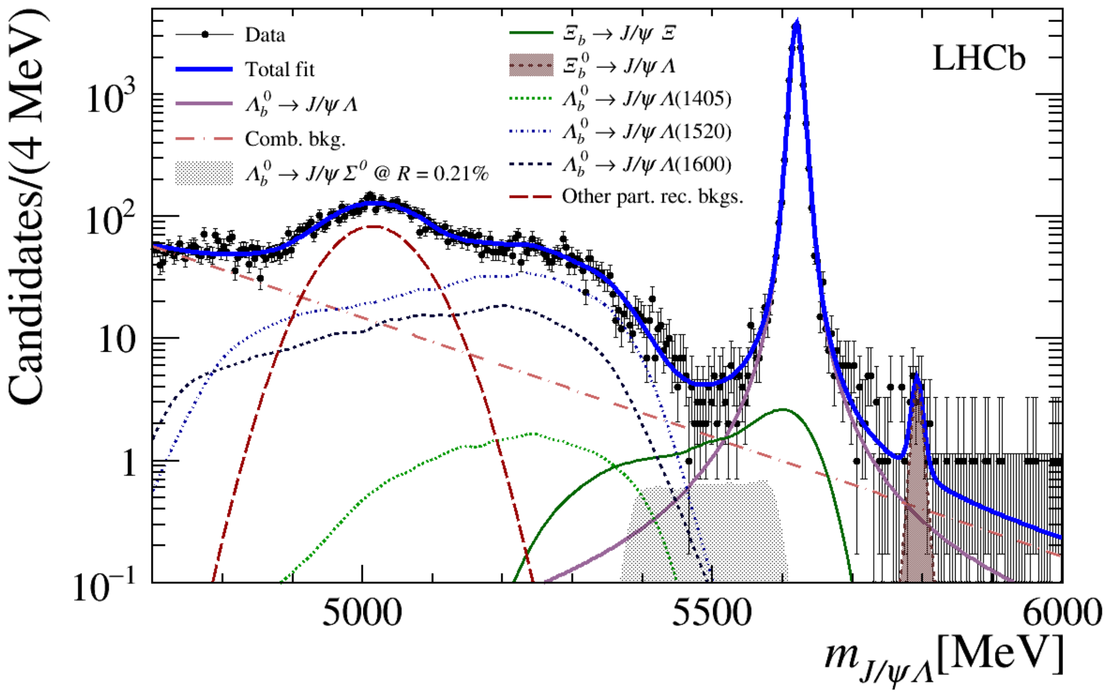


Figure 81: Distribution of the $J/\psi\Lambda$ mass for Run 2 data. Error bars without data points indicate empty bins. Also shown is the projection of the joint fit to the data. The thick (blue) solid curve shows the total fit. For illustrative purposes, the $\Lambda_b^0 \to J/\psi\Sigma^0$ signal component is artificially scaled to its measured upper limit. The shapes are identified in the legend.

but need to be included to accurately model the combinatoric background. The existence of the $\Lambda_b^0 \to J/\psi\Lambda^*$ channels was demonstrated in a study of $\Lambda_b^0 \to J/\psi K^- p$ decays [86]. We can model the resulting $J/\psi\Lambda$ mass shapes of the different $\Lambda_b^0 \to J/\psi\Lambda^*$ backgrounds, although we do not know their yields due to lack of knowledge of the relative $\Lambda^* \to \Sigma^0\pi^0$ branching fractions. We use separate shapes in the fit for the backgrounds corresponding to the $\Lambda(1405)$, $\Lambda(1520)$ and $\Lambda(1600)$ resonances. These backgrounds are simulated, processed through the event selections and fit using Gaussian kernel shapes. We collectively model the sum of the remaining Λ^* and $\psi(2S)$ backgrounds in the fit using a Gaussian shape. Note that our aim here is not to accurately disentangle each source of background, but only to model their collective sum.

A third background source arises from $\Xi_b \to J/\psi\Xi$ decays, where $\Xi \to \Lambda\pi$, when the pion from the Ξ decay is not reconstructed. This background is modeled by a Gaussian kernel shape fit to simulated $\Xi_b^- \to J/\psi \Xi^-$ decays, which are partially reconstructed as $J/\psi \Lambda$. The normalization of this background is determined by fully reconstructing $\Xi_b^- \to J/\psi \Xi^-$ decays in data and simulation to obtain an efficiency-corrected yield. The reconstruction uses the criteria in Ref. [94]. The reconstructed $J/\psi \Xi^-$ mass distribution in data is shown in the Supplemental material. The efficiency-corrected yield is multiplied by the relative efficiency of reconstructing $\Xi_b^- \to J/\psi \Xi^-$, as $J/\psi \Lambda$, and then more than doubled to account for $\Xi_b^0 \to J/\psi \Xi^0$ decays. The production rates are unequal mostly because the $\Xi_b'(5935)^0$ state is too light to decay into $\Xi_b^- \pi^+$ so it always decays into the Ξ_b^0 baryon [104]. In addition, we incorporate the production measurements of other excited Ξ_b resonances [105] to determine the inclusive production ratio of $\Xi_b^0/\Xi_b^- = 1.37 \pm 0.09$, where the uncertainty arises mainly from the production fraction measurements of the excited states. We further corrected for the lifetime ratio $\tau_{\Xi_b^-}/\tau_{\Xi_b^0} = 1.08 \pm 0.04$ [106]. This normalization is introduced into the final fit as a Gaussian constraint, and done separately for Run 1 and Run 2 data, as the detection efficiencies differ.

The remaining background comes mostly from random combinations of real J/ψ and Λ, which contribute both above and below the $\Lambda_b^0 \to J/\psi \Lambda$ mass peak. This combinatoric background is modeled using an exponential function.

The Run 1 and Run 2 mass distribution data are fit simultaneously, using a binned extended maximum-likelihood fit, where the efficiency-corrected relative yields of the $\Lambda_b^0 \to J/\psi \Sigma^0$ signal, and those of the three $\Lambda_b^0 \to J/\psi \Lambda^*$ decays, with respect to the $\Lambda_b^0 \to J/\psi \Lambda$ signal, are constrained to be the same in the two data sets. We define

$$\mathcal{R} \equiv \frac{|A_1|^2}{|A_0|^2} = \frac{\mathcal{B}(\Lambda_b^0 \to J/\psi \Sigma^0)}{\mathcal{B}(\Lambda_b^0 \to J/\psi \Lambda)} \cdot \Phi_{\Lambda_b^0} = \frac{N_{\Lambda_b^0 \to J/\psi \Sigma}}{N_{\Lambda_b^0 \to J/\psi \Lambda}} \cdot \frac{\epsilon_{\Lambda_b^0 \to J/\psi \Lambda}}{\epsilon_{\Lambda_b^0 \to J/\psi \Sigma}} \cdot \Phi_{\Lambda_b^0}, \tag{59}$$

where $N_{\Lambda_b^0 \to J/\psi \Sigma}$ and $N_{\Lambda_b^0 \to J/\psi \Lambda}$ are the yields of the $\Lambda_b^0 \to J/\psi \Sigma$ and $\Lambda_b^0 \to J/\psi \Lambda$ decays; $\epsilon_{\Lambda_b^0 \to J/\psi \Sigma}$ and $\epsilon_{\Lambda_b^0 \to J/\psi \Lambda}$ are their respective efficiencies, as estimated from simulation; the phase space correction factor, $\Phi_{\Lambda_b^0}$, is 1.058. The free parameters of interest in the fit are $\mathcal{R}$,

$N_{\Lambda_b^0 \to J/\psi \Lambda}$, and $N_{\Xi_b^0 \to J/\psi \Lambda}$; $N_{\Lambda_b^0 \to J/\psi \Sigma}$ can be calculated from these. Systematic uncertainties are folded into the fit components as Gaussian constraints. These include uncertainties on the simulated ratios of efficiencies for the different Λ_b^0 final states with respect to the $J/\psi \Lambda$ final state, which range from 1.4 to 2.4%. The uncertainty on the relative normalization of the $\Xi_b \to J/\psi \Xi$ background is estimated to be 12.1% for Run 1 and 9.8% for Run 2. This has contributions from the fit yield of the fully reconstructed $\Xi_b^- \to J/\psi \Xi^-$ decay, the reconstruction and efficiency of finding the $\Xi^- \to \Lambda \pi^-$ decay, and the Ξ_b^- / Ξ_b^0 lifetime ratio.

The results of the fit are shown in Fig. 81, and reported in Table 32. The fitted value for $\mathcal{R}$, is consistent with zero. In Fig. 81, we illustrate what this component would look like if observed at the upper limit on $\mathcal{R}$. We do not quote the yields of the $\Lambda_b^0 \to J/\psi \Lambda^*$ decays as these are highly correlated.

To set an upper limit on $\mathcal{R}$ we use the CLs method [107]. The variation of the observed and expected CLs versus $\mathcal{R}$ is scanned from 0 to 0.005 and shown in Fig. 82. Our observed upper limit on $\mathcal{R}$ is

$$\mathcal{R} < 0.0021 \text{ at } 95\% \text{ CL}.$$

Systematic uncertainties are incorporated in the fit and included in this limit. Further consistency checks include changing the fit range, eliminating the $\Lambda_b^0 \to J/\psi \Lambda^*$ background components one at a time, and fitting the $\Lambda_b^0 \to J/\psi \Lambda$ peak with different functions. These change the upper limit only by small amounts.

The Run 1 and Run 2 signal yields for $\Xi_b^0 \to J/\psi \Lambda$ are listed in Table 32. The statistical

Table 32: Results from the fit to the $J/\psi \Lambda$ mass distribution. The fitted yields are indicated by N. Note $N_{\Xi_b \to J/\psi \Xi}$ indicates the sum of Ξ_b^- and Ξ_b^0 decays.

Parameter	Shared value	Run 1 value	Run 2 value
$\mathcal{R}$	$(0 \pm 5.3) \cdot 10^{-4}$	–	–
$N_{\Lambda_b^0 \to J/\psi \Lambda}$	–	4417 ± 66	$16\,970 \pm 130$
$N_{\Xi_b \to J/\psi \Xi}$	–	23.3 ± 5.7	139.7 ± 21.9
$N_{\Xi_b^0 \to J/\psi \Lambda}$	–	6.2 ± 3.0	17.8 ± 5.1

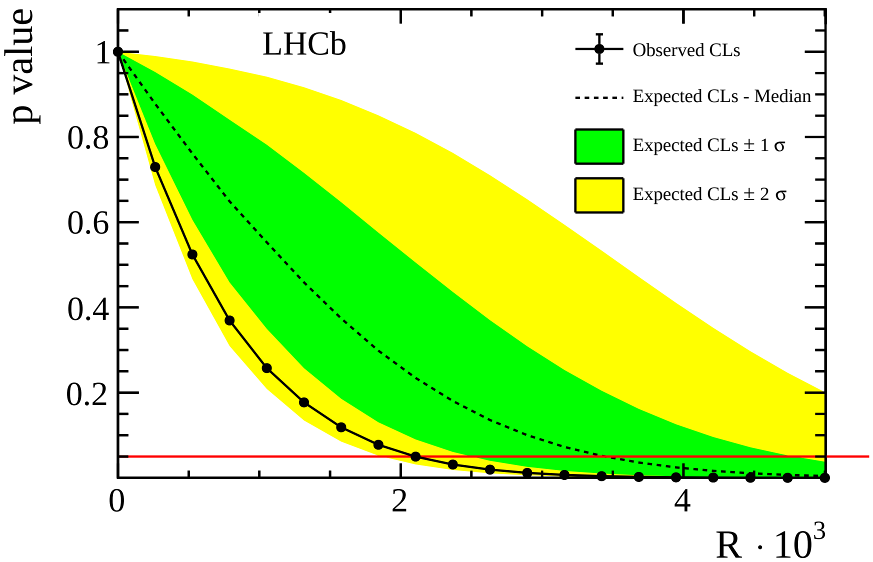


Figure 82: Result of the hypothesis tests conducted using the CLs method by varying $\mathcal{R}$ is shown. The observed CLs distribution is shown by the round (black) points. The expected CLs distribution (based on the background only hypothesis) is shown by the dashed line (black), with 1 and 2σ uncertainty bands depicted in dark shaded (green) and light shaded (yellow) bands. The observed and expected upper limits are obtained by seeing where the bands cross the p-value of 0.05 shown as the horizontal (red) line.

significance of the $\Xi_b^0 \to J/\psi \Lambda$ signal is 5.6 standard deviations, obtained using Wilks' theorem [108] and includes both the statistical and systematic uncertainties. The branching fraction ratio $\mathcal{B}(\Xi_b^0 \to J/\psi \Lambda)/\mathcal{B}(\Xi_b^0 \to J/\psi \Xi^0)$ is determined using the fully reconstructed $\Xi_b^- \to J/\psi \Xi^-$ sample described above. To determine the branching fraction of $\mathcal{B}(\Xi_b^0 \to J/\psi \Xi^0)$, we assume equal decay widths for the two different $\Xi_b \to J/\psi \Xi$ charge states and correct for the different neutral and charged Ξ_b production rates as described above. We use the measured lifetime ratio [106] to translate the decay width equality into the needed branching fraction. The Run 1 and Run 2 results are consistent. Combining the two, we find

$$R_{\Xi_b} \equiv \frac{\mathcal{B}(\Xi_b^0 \to J/\psi \Lambda)}{\mathcal{B}(\Xi_b^0 \to J/\psi \Xi^0)} = (8.2 \pm 2.1 \pm 0.9) \cdot 10^{-3},$$

where the first uncertainty is statistical the second is systematic, where the leading source is the systematic uncertainty in the $\Xi_b^- \to J/\psi \Xi^-$ fit yield.

We convert R_{Ξ_b} into a measurement of the amplitude ratio

$$\left|\frac{A_0}{A_{1/2}}\right| = \frac{1}{\lambda}\sqrt{\frac{R_{\Xi_b}}{\Phi_{\Xi_b}}} = 0.37 \pm 0.06 \pm 0.02$$

where $\Phi_{\Xi_b} = 1.15$ is the relative phase space factor, and $\lambda = 0.231$ is the relative Cabibbo suppression $|V_{cd}|/|V_{cs}|$, which is assumed equal to $|V_{us}|/|V_{ud}|$ [93]. Taking the s and u quarks in the Ξ_b^0 baryon to be a diquark state with isospin 1/2 and combining with the null isospin of the s quark from the b quark decay, leads to isopsin 1/2 for the $J/\psi\Xi^0$ final state. On the other hand, for the Cabibbo suppressed transition with the isospin 1/2 d quark, we have either isospin 0 or 1 final states. The former corresponds to $J/\psi\Lambda$, with the latter to $J/\psi\Sigma^0$, which we cannot currently measure. In order to predict the expected ratio of isospin amplitudes the SU(3) flavor [109] b-baryon couplings must be taken into account [110]. Then, if there are no other amplitudes, the theoretically predicted ratio corresponding to no preference between isospin 0 and 1/2 amplitudes is $|A_0/A_{1/2}|$ equal to $1/\sqrt{6}$ (≈ 0.41). Therefore, our result is consistent with no suppression of the isospin changing amplitude. These results are not precise enough to see the effects of SU(3) flavor symmetry breaking.

In conclusion, we set an upper limit in $\Lambda_b^0 \to J/\psi\Lambda(\Sigma^0)$ decays on the isospin amplitude ratio

$$|A_1/A_0| = \sqrt{\mathcal{R}} < 1/21.8 \text{ at } 95\% \text{ CL}.$$

This limit is stringent and rules out isospin violation at a $\sim$1% rate. Isospin violation has been seen at this level, for example, in $\rho-\omega$ mixing in $\overline{B}^0 \to J/\psi\pi^+\pi^-$ decays [111]. Our limit is consistent with the Λ_b^0 being formed of a b quark and a ud diquark. This measurement also constrains non-Standard Model A_1 amplitudes contributing to Λ_b^0 decays. Furthermore, our results support the quark model prediction of the Λ_b^0 being an isosinglet. Assumptions of isospin suppression in $\Lambda_b^0 \to J/\psi X$ decays made in past analyses are shown to be justified. Finally, we report the discovery of the Cabibbo suppressed decay $\Xi_b^0 \to J/\psi\Lambda$ and measure its branching fraction relative to $\Xi_b^0 \to J/\psi\Xi^0$ to be $(8.2 \pm 2.1 \pm 0.9) \cdot 10^{-3}$. We see no evidence for the preference of either isospin amplitude in the ratio $|A_0/A_{1/2}| = 0.37 \pm 0.06 \pm 0.02$, as the prediction for the equality of isospin amplitudes is $1/\sqrt{6}$.

References

[1] *Standard model of elementary particles.*, `https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg`. Accessed: 2022-02-21.

[2] *Quark structure of proton.*, `https://en.wikipedia.org/wiki/Proton#/media/File:Quark_structure_proton.svg`. Accessed: 2022-02-21.

[3] S. Bifani, S. Descotes-Genon, A. Romero Vidal, and M.-H. Schune, *Review of Lepton Universality tests in B decays*, J. Phys. G **46** (2019) 023001, `arXiv:1809.06229`.

[4] *Cern accelerator complex*, `https://cds.cern.ch/images/CERN-GRAPHICS-2019-002-1`. Accessed: 2022-02-17.

[5] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.

[6] *Lhcb facts*, `https://twiki.cern.ch/twiki/bin/view/Main/LHCb-Facts`. Accessed: 2022-02-17.

[7] LHCb RICH Group, M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. C **73** (2013) 2431, `arXiv:1211.6759`.

[8] LHCb collaboration, R. Aaij *et al.*, *Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity*, JHEP **02** (2016) 104, `arXiv:1512.04442`.

[9] Belle collaboration, A. Abdesselam *et al.*, *Angular analysis of $B^0 \to K^*(892)^0\ell^+\ell^-$*, in *LHC Ski 2016: A First Discussion of 13 TeV Results*, 2016, `arXiv:1604.04042`.

[10] ATLAS collaboration, M. Aaboud *et al.*, *Angular analysis of $B^0_d \to K^*\mu^+\mu^-$ decays in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, JHEP **10** (2018) 047, `arXiv:1805.04000`.

[11] CMS collaboration, A. M. Sirunyan *et al.*, *Measurement of angular parameters from the decay* $\mathrm{B}^0 \to \mathrm{K}^{*0}\mu^+\mu^-$ *in proton-proton collisions at* $\sqrt{s} = 8$ *TeV*, Phys. Lett. B **781** (2018) 517, `arXiv:1710.02846`.

[12] S. Descotes-Genon, L. Hofer, J. Matias, and J. Virto, *On the impact of power corrections in the prediction of* $B \to K^*\mu^+\mu^-$ *observables*, JHEP **12** (2014) 125, `arXiv:1407.8526`.

[13] LHCb collaboration, R. Aaij *et al.*, *Measurement of the* $B_s^0 \to \mu^+\mu^-$ *branching fraction and effective lifetime and search for* $B^0 \to \mu^+\mu^-$ *decays*, Phys. Rev. Lett. **118** (2017) 191801, `arXiv:1703.05747`.

[14] ATLAS collaboration, M. Aaboud *et al.*, *Study of the rare decays of* B_s^0 *and* B^0 *mesons into muon pairs using data collected during 2015 and 2016 with the ATLAS detector*, JHEP **04** (2019) 098, `arXiv:1812.03017`.

[15] J. Aebischer *et al.*, *B-decay discrepancies after moriond 2019*, The European Physical Journal C **80** (2020) .

[16] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality in beauty-quark decays*, `arXiv:2103.11769`.

[17] BaBar collaboration, J. P. Lees *et al.*, *Measurement of Branching Fractions and Rate Asymmetries in the Rare Decays* $B \to K^{(*)}l^+l^-$, Phys. Rev. D **86** (2012) 032012, `arXiv:1204.3933`.

[18] BELLE collaboration, S. Choudhury *et al.*, *Test of lepton flavor universality and search for lepton flavor violation in* $B \to K\ell\ell$ *decays*, JHEP **03** (2021) 105, `arXiv:1908.01848`.

[19] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality with* $B^0 \to K^{*0}\ell^+\ell^-$ *decays*, JHEP **08** (2017) 055, `arXiv:1705.05802`.

[20] M. Bordone, G. Isidori, and A. Pattori, *On the Standard Model predictions for* R_K *and* R_{K^*}, Eur. Phys. J. C **76** (2016) 440, `arXiv:1605.07633`.

[21] B. Capdevila, S. Descotes-Genon, J. Matias, and J. Virto, *Assessing lepton-flavour non-universality from $B \to K^*\ell\ell$ angular analyses*, JHEP **10** (2016) 075, arXiv:1605.03156.

[22] N. Serra, R. Silva Coutinho, and D. van Dyk, *Measuring the breaking of lepton flavor universality in $B \to K^*\ell^+\ell^-$*, Phys. Rev. D **95** (2017) 035029, arXiv:1610.08761.

[23] A. Bharucha, D. M. Straub, and R. Zwicky, *$B \to V\ell^+\ell^-$ in the Standard Model from light-cone sum rules*, JHEP **08** (2016) 098, arXiv:1503.05534.

[24] S. Jäger and J. Martin Camalich, *Reassessing the discovery potential of the $B \to K^*\ell^+\ell^-$ decays in the large-recoil region: SM challenges and BSM opportunities*, Phys. Rev. D **93** (2016) 014028, arXiv:1412.3183.

[25] Belle collaboration, J.-T. Wei *et al.*, *Measurement of the Differential Branching Fraction and Forward-Backward Asymmetry for $B \to K^{(*)}\ell^+\ell^-$*, Phys. Rev. Lett. **103** (2009) 171801, arXiv:0904.0770.

[26] BaBar collaboration, J. P. Lees *et al.*, *Measurement of an Excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ Decays and Implications for Charged Higgs Bosons*, Phys. Rev. D **88** (2013) 072012, arXiv:1303.0571.

[27] Belle collaboration, M. Huschle *et al.*, *Measurement of the branching ratio of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ relative to $\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle*, Phys. Rev. D **92** (2015) 072014, arXiv:1507.03233.

[28] Belle collaboration, S. Hirose *et al.*, *Measurement of the τ lepton polarization and $R(D^*)$ in the decay $\bar{B} \to D^*\tau^-\bar{\nu}_\tau$ with one-prong hadronic τ decays at Belle*, Phys. Rev. D **97** (2018) 012004, arXiv:1709.00129.

[29] Belle collaboration, G. Caria *et al.*, *Measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ with a semileptonic tagging method*, Phys. Rev. Lett. **124** (2020) 161803, arXiv:1910.05864.

[30] LHCb collaboration, R. Aaij *et al.*, *Measurement of the ratio of branching fractions $\mathcal{B}(\bar{B}^0 \to D^{*+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \to D^{*+}\mu^-\bar{\nu}_\mu)$*, Phys. Rev. Lett. **115** (2015) 111803, `arXiv:1506.08614`, [Erratum: Phys.Rev.Lett. 115, 159901 (2015)].

[31] LHCb collaboration, R. Aaij *et al.*, *Test of Lepton Flavor Universality by the measurement of the $B^0 \to D^{*-}\tau^+\nu_\tau$ branching fraction using three-prong τ decays*, Phys. Rev. D **97** (2018) 072013, `arXiv:1711.02505`.

[32] D. Bigi and P. Gambino, *Revisiting $B \to D\ell\nu$*, Phys. Rev. D **94** (2016) 094008, `arXiv:1606.08030`.

[33] P. Gambino, M. Jung, and S. Schacht, *The V_{cb} puzzle: An update*, Phys. Lett. B **795** (2019) 386, `arXiv:1905.08209`.

[34] M. Bordone, M. Jung, and D. van Dyk, *Theory determination of $\bar{B} \to D^{(*)}\ell^-\bar{\nu}$ form factors at $\mathcal{O}(1/m_c^2)$*, Eur. Phys. J. C **80** (2020) 74, `arXiv:1908.09398`.

[35] B. Capdevila *et al.*, *Searching for New Physics with $b \to s\tau^+\tau^-$ processes*, Phys. Rev. Lett. **120** (2018) 181802, `arXiv:1712.01919`.

[36] R. Alonso, B. Grinstein, and J. Martin Camalich, *Lepton universality violation and lepton flavor conservation in B-meson decays*, JHEP **10** (2015) 184, `arXiv:1505.05164`.

[37] *Tmva users guide*, `https://root.cern.ch/download/doc/tmva/TMVAUsersGuide.pdf`. Accessed: 2022-03-01.

[38] G. A. Cowan, D. C. Craik, and M. D. Needham, *RapidSim: an application for the fast simulation of heavy-quark hadron decays*, Comput. Phys. Commun. **214** (2017) 239, `arXiv:1612.07489`.

[39] *The cls method: information for conference speakers*, `http://www.pp.rhul.ac.uk/~cowan/stat/cls/CLsInfo.pdf`. Accessed: 2022-03-21.

[40] A. N. Taylor *et al.*, *Gravitational lens magnification and the mass of abell 1689*, The Astrophysical Journal **501** (1998) 539–553.

[41] K. C. Freeman, *On the disks of spiral and SO Galaxies*, Astrophys. J. **160** (1970) 811.

[42] L. Maiani, *The GIM Mechanism: origin, predictions and recent uses*, in *48th Rencontres de Moriond on Electroweak Interactions and Unified Theories*, 3–16, 2013, `arXiv:1303.6154`.

[43] S. Descotes-Genon, L. Hofer, J. Matias, and J. Virto, *Global analysis of $b \to s\ell\ell$ anomalies*, JHEP **06** (2016) 092, `arXiv:1510.04239`.

[44] V. V. Gligorov, C. Thomas, and M. Williams, *The HLT inclusive B triggers* , CERN, Geneva, 2011. LHCb-INT-2011-030.

[45] *Moore*, `https://lhcbdoc.web.cern.ch/lhcbdoc/moore/master/index.html`. Accessed: 2022-02-21.

[46] G. Barrand *et al.*, *GAUDI - A software architecture and framework for building HEP data processing applications*, Comput. Phys. Commun. **140** (2001) 45.

[47] G. Corti *et al.*, *Software for the LHCb experiment*, IEEE Trans. Nucl. Sci. **53** (2006) 1323.

[48] LHCb collaboration, R. Aaij *et al.*, *Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays*, JHEP **08** (2017) 055, `arXiv:1705.05802`.

[49] LHCb collaboration, R. Aaij *et al.*, *Measurement of the ratio of the $B^0 \to D^{*-}\tau^+\nu_\tau$ and $B^0 \to D^{*-}\mu^+\nu_\mu$ branching fractions using three-prong τ-lepton decays*, Phys. Rev. Lett. **120** (2018) 171802, `arXiv:1708.08856`.

[50] A. Crivellin, D. Müller, and T. Ota, *Simultaneous explanation of $R(D^{(*)})$ and $b{\to}s\mu^+\mu^-$: the last scalar leptoquarks standing*, JHEP **09** (2017) 040, `arXiv:1703.09226`.

[51] D. A. Faroughy, A. Greljo, and J. F. Kamenik, *Confronting lepton flavor universality violation in B decays with high-p_T tau lepton searches at LHC*, Phys. Lett. B **764** (2017) 126, arXiv:1609.07138.

[52] F. Feruglio, P. Paradisi, and A. Pattori, *Revisiting Lepton Flavor Universality in B Decays*, Phys. Rev. Lett. **118** (2017) 011801, arXiv:1606.00524.

[53] A. J. Buras, J. Girrbach-Noe, C. Niehoff, and D. M. Straub, *$B \to K^{(*)}\nu\overline{\nu}$ decays in the Standard Model and beyond*, JHEP **02** (2015) 184, arXiv:1409.4557.

[54] LHCb collaboration, R. Aaij *et al.*, *Search for the decays $B_s^0 \to \tau^+\tau^-$ and $B^0 \to \tau^+\tau^-$*, Phys. Rev. Lett. **118** (2017) 251802, arXiv:1703.02508.

[55] BaBar collaboration, J. P. Lees *et al.*, *Search for $B^+ \to K^+\tau^+\tau^-$ at the BaBar experiment*, Phys. Rev. Lett. **118** (2017) 031802, arXiv:1605.09637.

[56] Belle collaboration, T. V. Dong *et al.*, *Search for the decay $B^0 \to K^{*0}\tau^+\tau^-$ at the Belle experiment*, arXiv:2110.03871.

[57] Particle Data Group, M. Tanabashi *et al.*, *Review of particle physics*, Phys. Rev. **D98** (2018) 030001, and 2019 update.

[58] Particle Data Group, P. A. Zyla *et al.*, *Review of particle physics*, Prog. Theor. Exp. Phys. **2020** (2020) 083C01.

[59] W. D. Hulsbergen, *Decay chain fitting with a kalman filter*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **552** (2005) 566–575.

[60] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1554, arXiv:1007.1727, [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[61] A. Wald, *Tests of statistical hypotheses concerning several parameters when the number of observations is large*, Transactions of the American Mathematical Society **54** (1943) 426.

[62] A. L. Read, *Modified frequentist analysis of search results (The CL(s) method)*, in *Workshop on Confidence Limits*, 81–101, 2000.

[63] M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023.

[64] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820.

[65] I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, J. Phys. Conf. Ser. **331** (2011) 032047.

[66] N. Davidson, T. Przedzinski, and Z. Was, *PHOTOS interface in C++: Technical and physics documentation*, Comp. Phys. Comm. **199** (2016) 86, arXiv:1011.0937.

[67] Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[68] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270.

[69] D. Müller, M. Clemencic, G. Corti, and M. Gersabeck, *ReDecay: A novel approach to speed up the simulation at LHCb*, Eur. Phys. J. C **78** (2018) 1009, arXiv:1810.10362.

[70] I. M. Nugent *et al.*, *Resonance chiral Lagrangian currents and experimental data for* $\tau^- \to \pi^-\pi^-\pi^+\nu_\tau$, Phys. Rev. D **88** (2013) 093012, arXiv:1310.1053.

[71] BaBar collaboration, I. M. Nugent, *Invariant mass spectra of* $\tau^- \to h^-h^-h^+\nu_\tau$ *decays*, Nucl. Phys. B Proc. Suppl. **253-255** (2014) 38, arXiv:1301.7105.

[72] A. Gulin, I. Kuralenok, and D. Pavlov, *Winning the transfer learning track of yahoo!'s learning to rank challenge with yetirank*, in *Proceedings of the Learning to Rank Challenge* (O. Chapelle, Y. Chang, and T.-Y. Liu, eds.), **14**, (Haifa, Israel), 63–76, PMLR, 2011.

[73] V. V. Gligorov and M. Williams, *Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree*, JINST **8** (2013) P02013, `arXiv:1210.6861`.

[74] G. Mancinelli and J. Serrano, *Study of Muon Isolation in the $B_s^0 \to \mu^+\mu^-$ Channel* , CERN, Geneva, 2010.

[75] ROOT collaboration, K. Cranmer *et al.*, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, .

[76] R. Barlow and C. Beeston, *Fitting using finite monte carlo samples*, Computer Physics Communications **77** (1993) 219.

[77] *Minuit reference manual*, `https://root.cern.ch/download/minuit.pdf`. Accessed: 2022-03-22.

[78] LHCb collaboration, R. Aaij *et al.*, *Isospin amplitudes in $\Lambda_b^0 \to J/\psi\Lambda(\Sigma^0)$ and $\Xi_b^0 \to J/\psi\Xi^0(\Lambda)$ decays*, Phys. Rev. Lett. **124** (2020) 111802, `arXiv:1912.02110`.

[79] H.-Y. Cheng, *Status of the $\Delta I = 1/2$ rule in kaon decay*, Int. J. Mod. Phys. **A4** (1989) 495.

[80] A. J. Buras, J.-M. Gérard, and W. A. Bardeen, *Large N approach to kaon decays and mixing 28 years later: $\Delta I = 1/2$ Rule, $\hat{B}_K$ and ΔM_K*, Eur. Phys. J. **C74** (2014) 2871, `arXiv:1401.1385`.

[81] RBC and UKQCD collaborations, P. A. Boyle *et al.*, *Emerging understanding of the $\Delta I = 1/2$ rule from lattice QCD*, Phys. Rev. Lett. **110** (2013) 152001, `arXiv:1212.1474`; T. Blum *et al.*, *$K \to \pi\pi$ $\Delta I = 3/2$ decay amplitude in the continuum limit*, Phys. Rev.

D91 (2015) 074502, `arXiv:1502.00263`; N. Ishizuka, K.-I. Ishikawa, A. Ukawa, and T. Yoshié, *Calculation of $K \to \pi\pi$ decay amplitudes with improved Wilson fermion action in lattice QCD*, Phys. Rev. **D92** (2015) 074503, `arXiv:1505.05289`.

[82] E. Franco, S. Mishima, and L. Silvestrini, *The Standard Model confronts CP violation in $D^0 \to \pi^+\pi^-$ and $D^0 \to K^+K^-$*, JHEP **05** (2012) 140, `arXiv:1203.3131`.

[83] B. Grinstein, D. Pirtskhalava, D. Stone, and P. Uttayarat, *B decays to two pseudoscalars and a generalized $\Delta I = \frac{1}{2}$ rule*, Phys. Rev. **D89** (2014) 114014, `arXiv:1402.1164`.

[84] Y. Grossman, M. Neubert, and A. L. Kagan, *Trojan penguins and isospin violation in hadronic B decays*, JHEP **10** (1999) 029, `arXiv:hep-ph/9909297`.

[85] S. R. Coleman and S. L. Glashow, *Electrodynamic properties of baryons in the unitary symmetry scheme*, Phys. Rev. Lett. **6** (1961) 423; R. H. Dalitz and F. Von Hippel, *Electromagnetic $\Lambda - \Sigma^0$ mixing and charge symmetry for the Λ-hyperon*, Phys. Lett. **10** (1964) 153; CSSM/QCDSF/UKQCD collaboration, Z. R. Kordov *et al.*, *Electromagnetic contribution to Σ-Λ mixing using lattice QCD+QED*, `arXiv:1911.02186`.

[86] LHCb collaboration, R. Aaij *et al.*, *Observation of $J/\psi p$ resonances consistent with pentaquark states in $\Lambda_b^0 \to J/\psi K^- p$ decays*, Phys. Rev. Lett. **115** (2015) 072001, `arXiv:1507.03414`.

[87] LHCb collaboration, R. Aaij *et al.*, *Determination of the quark coupling strength $|V_{ub}|$ using baryonic decays*, Nature Phys. **11** (2015) 743, `arXiv:1504.01568`.

[88] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.

[89] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, `arXiv:1412.6352`.

[90] R. Aaij *et al.*, *The LHCb trigger and its performance in 2011*, JINST **8** (2013) P04022, `arXiv:1211.3055`.

[91] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.

[92] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A precision tool for QED corrections in Z and W decays*, Eur. Phys. J. **C45** (2006) 97, `arXiv:hep-ph/0506026`.

[93] Particle Data Group, M. Tanabashi *et al.*, *Review of particle physics*, Phys. Rev. **D98** (2018) 030001, and 2019 update.

[94] LHCb collaboration, R. Aaij *et al.*, *Measurement of the mass and production rate of Ξ_b^- baryons*, Phys. Rev. **D99** (2019) 052006, `arXiv:1901.07075`.

[95] W. D. Hulsbergen, *Decay chain fitting with a Kalman filter*, Nucl. Instrum. Meth. **A552** (2005) 566, `arXiv:physics/0503191`.

[96] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, Wadsworth international group, Belmont, California, USA, 1984.

[97] Y. Freund and R. E. Schapire, *A decision-theoretic generalization of on-line learning and an application to boosting*, J. Comput. Syst. Sci. **55** (1997) 119.

[98] H. Voss, A. Hoecker, J. Stelzer, and F. Tegenfeldt, *TMVA - Toolkit for Multivariate Data Analysis with ROOT*, PoS **ACAT** (2007) 040; J. Stelzer, A. Hocker, P. Speckmayer, and H. Voss, *Current developments in TMVA: An outlook to TMVA4*, PoS **ACAT08** (2008) 063.

[99] M. Pivk and F. R. Le Diberder, *sPlot: A statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555** (2005) 356, `arXiv:physics/0402083`.

[100] A. Bagoly *et al.*, *Machine learning developments in ROOT*, J. Phys. Conf. Ser. **898** (2017) 072046.

[101] G. Punzi, *Sensitivity of searches for new signals and its optimization*, eConf **C030908** (2003) MODT002, `arXiv:physics/0308063`.

[102] K. S. Cranmer, *Kernel estimation in high-energy physics*, Comput. Phys. Commun. **136** (2001) 198, `arXiv:hep-ex/0011057`.

[103] D. Martínez Santos and F. Dupertuis, *Mass distributions marginalized over per-event errors*, Nucl. Instrum. Meth. **A764** (2014) 150, `arXiv:1312.5000`.

[104] LHCb collaboration, R. Aaij *et al.*, *Observation of two new Ξ_b^- baryon resonances*, Phys. Rev. Lett. **114** (2015) 062004, `arXiv:1411.4849`.

[105] LHCb collaboration, R. Aaij *et al.*, *Measurement of the properties of the Ξ_b^{*0} baryon*, JHEP **05** (2016) 161, `arXiv:1604.03896`.

[106] LHCb collaboration, R. Aaij *et al.*, *Precision measurement of the mass and lifetime of the Ξ_b^- baryon*, Phys. Rev. Lett. **113** (2014) 242002, `arXiv:1409.8568`.

[107] A. L. Read, *Presentation of search results: The CL_s technique*, J. Phys. **G28** (2002) 2693.

[108] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, Ann. Math. Statist. **9** (1938) 60.

[109] G. Hiller, M. Jung, and S. Schacht, *SU(3)$_F$ in nonleptonic charm decays*, PoS **EPS-HEP2013** (2013) 371, `arXiv:1311.3883`.

[110] A. Dery, M. Ghosh, Y. Grossman, and S. Schacht, *SU(3)$_F$ analysis for beauty baryon decays*, `arXiv:2001.05397`.

[111] LHCb collaboration, R. Aaij *et al.*, *Measurement of the CP-violating phase β in $\overline{B}^0 \to J/\psi\pi^+\pi^-$ decays and limits on penguin effects*, Phys. Lett. **B742** (2015) 38, `arXiv:1411.1634`.

Vita

NAME OF AUTHOR: Aravindhan Venkateswaran

PLACE OF BIRTH: Chennai, India

DATE OF BIRTH: May 3, 1992

DEGREES AWARDED: Bachelors of Technology in Engineering and Masters of Technology in Engineering from the Indian Institute of Technology, Madras

SELECTED PUBLICATIONS:

- The LHCb collaboration, *Search for weakly decaying **b**-flavored pentaquarks*, Phys. Rev. D97 (2018) 3, 032010, arXiv:1712.08086.
- The LHCb collaboration, *Isospin amplitudes in* $\Lambda_b^0 \to J/\psi\Lambda(\Sigma^0)$ *and* $\Xi_b^0 \to J/\psi\Xi^0(\Lambda)$ *decays*, Phys. Rev. Lett. 124 (2020) 111802, arXiv:1912.02110.