Measurement of deuterons at LHCb

Sophie Katherine Baker

Imperial College London
Department of Physics

A thesis submitted to Imperial College London
for the degree of Doctor of Philosophy

Abstract

A first measurement of deuteron production at the LHCb experiment is presented. The measurement is performed as a function of momentum, using proton-proton collision data from the 2018 data-taking period, at a centre-of-mass energy $\sqrt{s}$ =13 TeV. The simulation of deuteron production in proton-proton collisions, which is used extensively in the development of the measurement strategy, is described. The deuteron production cross-section measurement uses a neural network variable, ProbNNd, to discriminate between deuterons and other charged particles, in the momentum range $20 < p < 100\,\text{GeV}/c$, for pseudorapidities $1.89 < \eta < 4.89$. Measurements of deuterons produced in high energy collisions provide important information for the calibration of models used in indirect dark matter searches. Some of these searches use antideuteron signals in cosmic-rays to look for evidence for dark matter annihilation. The measurement at LHCb is found not to be sensitive to a deuteron signal. The steps that should be followed to compute a deuteron production cross-section at LHCb are detailed.

Declaration

The analysis presented in this thesis is the product of the research I have performed from January 2016 to September 2019. I have had the support of the Imperial College High Energy Physics group, and members of the LHCb Collaboration. All of the analysis work presented in this thesis was performed by myself, with one exception. The incorporation of the deuteron production models into the LHCb simulation software was performed by members of the LHCb Simulation Working Group: Philip Ilten, Gloria Corti, and Michal Kreps. Philip Ilten also assisted in the development of the code for the deuteron production simulation, in section 3.3.

This thesis has not been submitted for any other qualifications.

Sophie Baker, September 2019

Contents

List of figures

List of tables

Chapter 1

Introduction

This thesis presents the first measurement of the deuteron production cross-section at the LHCb experiment. The measurement is performed on data taken during 2018 in pp collisions at a centre-of-mass energy $\sqrt{s} = 13\,\text{TeV}$. The LHCb detector acceptance covers the forward region, with pseudorapidity $2 < \eta < 5$, and provides deuteron identification for momenta in the range $20 < p < 100\,\text{GeV}/c$, using a pair of Ring-Imaging Cherenkov (RICH) detectors. Therefore, the measurement at the LHCb detector is complementary to deuteron measurements at other collider experiments, such as the ALICE experiment, which has measured deuterons in a central pseudorapidity region, $|\eta| < 0.9$ and $p_{\text{T}} < 6\,\text{GeV}/c$ [1]. A novel technique for extracting the deuteron signal is used, whereby binned fits are performed to data in a neural network discriminating variable. This variable combines information from many detector elements, to provide good separation between deuterons and other charged particles.

Existing deuteron production models are not able to predict experimental results consistently, and as such, measurements at high energy experiments can be used to improve their accuracy. Knowledge of deuteron and antideuteron formation is of importance to some indirect dark matter searches [2]. These searches, such as that at AMS-02 [3], look for evidence of dark matter annihilation or decay in cosmic rays, through excesses in the observed rates of antimatter compared to those expected from standard cosmological sources. Searches have produced inconclusive results for light antimatter, such as positrons and antiprotons, and are now being extended to heavier signatures, namely antideuterons, which have very low predicted background flux rates. The backgrounds to antideuteron signals from dark matter sources will arise through antideuterons produced in cosmic ray collisions. The predicted rate for this to occur is very low, although affected by

large uncertainties. These predictions could be improved using measurments of deuteron production in controlled environments, such as at the LHCb experiment.

The structure of this thesis is as follows. In chapter 2, the Standard Model of particle physics is presented, with a focus on the theory of the strong force, quantum chromodynamics, followed by a discussion of the limitations of the Standard Model and the evidence for dark matter. In chapter 3, deuteron production models, and their implementation in simulated events are outlined. In chapter 4, the LHCb detector is reviewed. In chapter 5, the procedure used to measure the deuteron producton cross-section at the LHCb experiment is detailed. Finally, the conclusions are given in chapter 6.

Chapter 2

The Standard Model of particle physics

2.1 The Standard Model

The Standard Model (SM) of particle physics is able to describe all known particles, and almost all known forces in the universe. It provides a framework for seventeen fundamental particles and their associated interactions, and makes predictions which have been measured to high precision [4]. The particles are nominally divided into three types – fermions with spin $\frac{1}{2}$, gauge bosons with spin 1, and the Higgs boson with spin 0. The fermions constitute matter, existing either as leptons, or as quarks, which combine to form hadrons, whilst the bosons are the force carriers, providing the particle interactions. Each particle also has an antiparticle counterpart with the same mass but the opposite electromagnetic (EM) charge.

The possible interactions of particles[1] are limited to those permitted by the conservation of different symmetries. Decays and interactions permitted in the SM are governed by: EM charge, colour, baryon number, lepton number, flavour (approximate), spin, and energy, momentum, and angular momentum conservation.

The particles of the SM are shown in Fig. 2.1, with the fundamental fermions on the left, the gauge bosons on the right, and finally the Higgs boson on the top right. Interactions with the Higgs field give all massive particles their mass. As such, massless photons and gluons do not interact with the Higgs boson, but quarks, leptons, and the W and Z bosons do. Quarks and leptons each appear in three flavour doublets, as paired up in Fig. 2.1. The neutrinos are massless in the SM, though it has been seen experimentally that they have a small mass [5]. Fermions can have two chirality states, left-handed

[1] Charge conjugates implied throughout this document, unless explicitly stated.

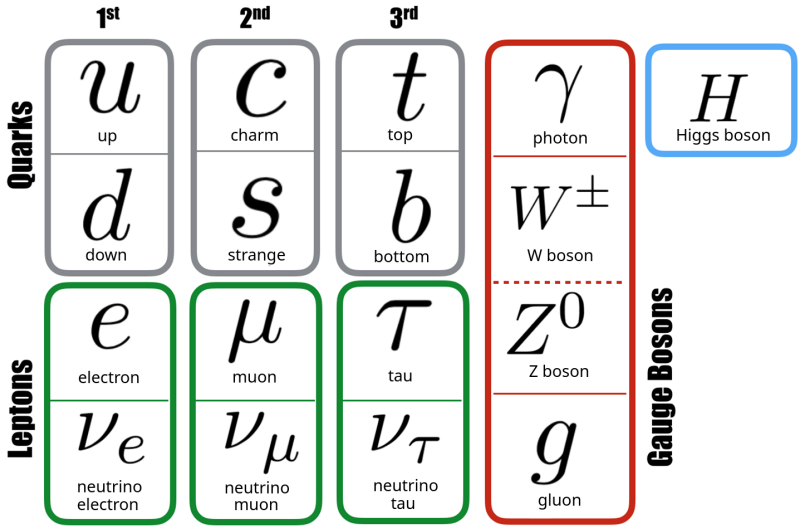


Figure 2.1: The particles of the SM, organised into generations and categories.

or right-handed, which in the massless limit, correspond to their spin pointing in the opposite or the same direction to their momentum, respectively.

The SM is a quantum field theory, invariant in the symmetry group $U(1) \times SU(2) \times SU(3)$. Each gauge symmetry has an associated charge, which is conserved under gauge transformations. The $U(1) \times SU(2)$ symmmetry group describes the weak and electromagnetic forces, whilst the strong force is described by $SU(3)$.

The charged leptons can interact with both photons, and the W and Z bosons. Neutrinos interact only through the weak W and Z bosons, and in the limit that neutrinos are massless, these interactions can only occur with left-handed neutrinos, or right-handed antineutrinos. Quark and lepton flavour can be changed through weak processes, via the W bosons [6]. The strengths of the quark couplings are given by the Cabbibo-Kobayoshi-Maskawa matrix, where quarks of the same generation have a strong coupling, and couplings across generations are weaker, and therefore such transitions occur less frequently.

Equivalently for neutrinos, the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix describes their mixing [7]. The states of neutrinos according to flavour (ν_e, ν_μ, and ν_τ) are combinations of three mass states, with the two sets of states related by the PMNS matrix. Flavour states take part in weak interactions, however their propagation over time is governed by the masss states. Consequently, a neutrino created in a pure flavour

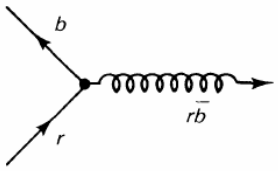


Figure 2.2: A quark-gluon vertex where the colour of the interacting quark changes ($r \to b$), and the difference is carried by the gluon ($r\bar{b}$). From Ref. [6].

state will evolve into a superposition of flavour states. This induces mixing, meaning that neutrinos are observed in different states to those in which they were created, when they are detected at a later time [5].

Quarks are also governed by the strong force, mediated by gluons. Quarks carry both EM and colour charges, and gluons carry colour charges, where the overall colour charge is conserved in strong interactions. Individually, they carry combinations of red, r, blue, b, or green, g, charges, and their anti-charges, $\overline{r}$, $\overline{b}$ and $\overline{g}$. The total colour charge of hadrons is a colour-neutral superposition of these, such as

$$\Psi_c^{q\bar{q}} = \frac{r\overline{r} + b\overline{b} + g\overline{g}}{\sqrt{3}} \tag{2.1}$$

for mesons with a quark and an antiquark, and

$$\Psi_c^{qqq} = (rbg - rgb + grb - gbr + bgr - brg)/\sqrt{6} \tag{2.2}$$

for baryons with three quarks, with gluons holding them together [7]. A colour conserving vertex is shown in Fig. 2.2. A quark and gluon meet at the vertex, where the quark changes colour, and the gluon carries away the colour difference. Gluons carry one colour and one anti-colour charge, enabling them to self-interact, and giving rise to gluon three- and four-point vertices [6].

The QED interaction can be presented using the example of an electron and a positron interacting via a photon exchange [8]. The potential for this interaction is given by the Coulomb interaction potential,

$$V_{\mathrm{QED}}(r) = -\frac{\alpha}{r}, \tag{2.3}$$

where r is the distance separating the electron and positron, and α is the fine structure constant, $\alpha = \frac{e^2}{4\pi\hbar c} \approx \frac{1}{137}$. Higher order interactions go as α^n for the n^{th} order interaction,

Figure 2.3: The tube of field lines between a colour charged and an anti-colour charged quark. From Ref. [9].

meaning that the contribution to the overall interaction, or Feynman diagram, decreases with increasing n.

A similar interaction can be presented for the strong force: the exchange of a gluon between a quark–antiquark pair, $q\overline{q}$ [8]. The difference in this situation, compared to the QED interaction, is that the quark and antiquark carry colour charge, as well as EM charge. The total effective potential experienced by a $q\overline{q}$ in a meson is

$$V_{\mathrm{QCD}}(r) = -\frac{4}{3}\frac{\alpha_s}{r} + \lambda r, \tag{2.4}$$

where $\alpha_s = g_s^2/4\pi$, for the strong coupling g_s, which is analagous to e in QED, and r is the separation between the $q\overline{q}$. The factor of $\frac{4}{3}$ in the first term is from the summation of all possible colour combinations in the $q\overline{q}$ interaction, and the second term arises from gluon self-coupling.

The first, short-range term in Eq. 2.4 describes 'asymptotic' freedom, due to single gluon exchange. The coupling in this term is in fact a 'running' coupling, which scales with the four-momentum exchange of an interaction, Q^2, with $\alpha_s(Q^2) \to 0$ as $Q^2 \to \infty$. Thus, at high energies or short distances, the coupling diminishes, and quarks within a hadron are quasi-free. This weak coupling at small distances within a hadron is asymptotic freedom.

The second, long-range term in Eq. 2.4 gives 'confinement', where the coupling becomes stronger as r increases. This is caused by increased gluon self-coupling as the $q\overline{q}$ are pulled apart, because the gluons between the $q\overline{q}$ also attract each other. Colour lines of force form in a tube between the $q\overline{q}$, with constant density as the $q\overline{q}$ are separated, as shown in Fig. 2.3. The result of this constant energy density is that quarks are never able to escape from hadrons. The effect is shown pictorially in Fig. 2.4, where a u quark is pulled away from the ud in a proton. The u quark can only be pulled so far away before a $d\overline{d}$ pair form in the middle, resulting in the formation of a pion and a neutron.

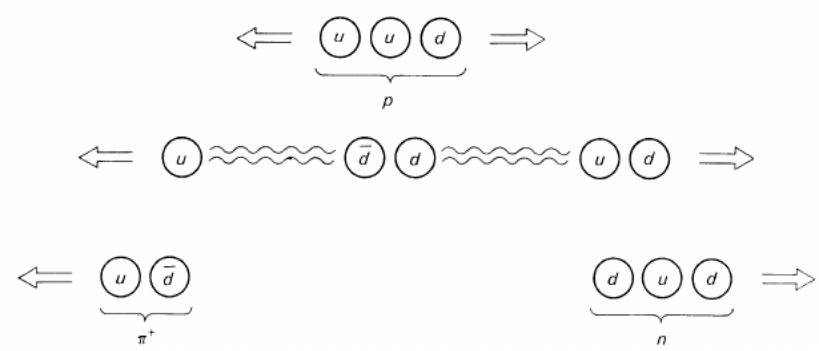


Figure 2.4: An example of hadronisation: as the u is pulled away from the u and d quarks in the proton, a pair of quarks is created in the flux tube, such that a pion and a neutron are created. From Ref. [6]

The effective potential in Eq. 2.4 works well for mesons containing b or c quarks, where the binding energy is smaller than the quark masses, and the internal motion is non-relativistic. In hadrons containing lighter quarks (u, d, s), where $\alpha_s \geq 1$ as $Q^2 \to 0$, the picture becomes more complicated. As the coupling has increased, there is a high probability that additional quarks and gluons will be created within the hadron.

This latter scenario, can be treated using the Parton Model [10], where hadrons are made up of three constituents:

- valence quarks which carry the quantum numbers of the hadron, such as charge, colour, strangeness *etc.*. These are the primary *uud* quarks in a proton, for instance;
- a 'sea' of virtual gluons, which are exchanged between quarks, and the gluons themselves;
- a 'sea' of virtual quark and antiquark pairs that are created from the vacuum polarisation of the colour field.

The likelihood of the virtual gluons and quarks forming is a function of their mass, so the probability of finding gluons is high at any energy scale, whilst, given enough energy, the virtual quarks can exist for a time interval inversely proportional to their mass, according to the uncertainty principle.

Scattering experiments are a good way to learn more about hadron structure, and deep inelastic scattering has been used for this purpose [11]. Typically, high energy electrons are scattered off nuclei in order to discover the structure of the hadrons. An electron with a few MeV of energy can be used to find the broad nuclear structure, showing peaks

for protons and their excited states. Increasing this to a few GeV results in a smooth continuum of electrons which have been scattered off quarks. This is the expected result for a target made up of free, pointlike, spin $\frac{1}{2}$, charged particles. At electron energies of a few GeV, the QCD behaviour is in the limit $Q^2 \to \infty$, such that $\alpha_s(Q^2) \to 0$, giving sensitivity to the short distance hadron structure – asymptotic freedom. Electron probes in these scattering experiments do not interact with the sea of gluons, as they do not carry colour charge.

It has been seen in deep inelastic scattering that if the electron has sufficient energy, that a quark is pushed out of the hadron, similarly to the process in Fig. 2.4. At a certain distance, it becomes energetically favourable to create a new $q\overline{q}$ pair, such that the colour tubes snap and new hadrons form. This process can continue to occur, with new mesons created after each stretch and snap of the colour tubes, until the kinetic energy of the system has dissipated. This process is known as 'hadronisation', and it is the cause of jets, or narrow cones of hadrons, in high energy collisions.

A residual strong force is felt between nucleons in atomic nuclei, which bonds them together. Quarks and gluons can be exchanged by the nucleons; namely pions, as these are the lightest colourless particle, though it is possible that heavier hadrons are exchanged, albeit at shorter ranges. A similar exchange takes place in high energy hadron scattering.

In high energy hadron scattering, the collision can be considered to occur between the partons that make up the colliding particles. If significant energy is transferred, the hadrons can fragment, and subsequently hadronise, forming jets. Most products of the interaction will be produced collinear to the colliding particles, with the cross-section of products falling off with transverse momentum, p_{T}, given by $\frac{\partial\sigma}{\partial p_{\mathrm{T}}} \sim e^{-6p_{\mathrm{T}}}$ [8]. Most products close to the interaction point will be pions, as these are the lowest mass hadrons. If a scattering process has taken place, two jets are usually produced, collinear to the incoming particles. Elastic scattering events, where neither hadron is broken up, are rare and only occur in soft collisions. For jets to be produced with high p_{T}, a scattering with a small impact parameter must have taken place, *i.e.* the colliding hadrons passed very close to each other, such that the partons are scattered at wide angles. This kind of event often results in four jets being produced: two collinear to the colliding particles, and two with high p_{T} The quarks taking part in the scattering contribute to the transverse jets, whilst the other quarks in the hadrons produce the collinear jets [8].

The information about parton density functions gained through deep inelastic scattering is invaluable for computing the processes that are initiated by the collision of

two hadrons [12]. Cross-sections from hadron collision processes are often measured as functions of p_{T}, because the longitudinal momentum of the hard subprocess is not known [4]. The cross-sections are accurately predicted using the parton density functions, but measurements at the LHC and other state-of-the-art high energy experiments stretch the theories to their limits.

Though the SM is successful at describing many aspects of what is observed of the universe, it is not able to complete the picture in detail. It does not extend to describe gravitational forces, nor predict neutrino mass, explain the matter-antimatter asymmetry that occurred after the Big Bang, or provide a candidate for dark matter.

2.2 Dark Matter

The first evidence for dark matter (DM) emerged in the early 1930s, through the study of the motion of stars [13, 14]. At the time, mass-to-luminosity ratios were often used to estimate the mass of galactic objects. It was found by Oort that the motion of stars in the Milky Way suggested that there was a much greater galactic mass than mass-to-luminosity ratios could predict [15]. In fact, the stars appeared to travel fast enough that they would escape the gravitational pull of the galaxy, if it was only formed of the luminous matter. Thus, this mysterious, additional matter was named 'dark matter'.

Since this first indication, a catalogue of evidence for the presence of DM has been compiled. Study of the rotation of galaxies shows that their motion deviates from what is expected; instead of the velocity of rotation decreasing as the distance from the centre of the galaxy increased, it was found that the rotation velocity increased with increasing radial distance, before reaching a limit and flattening out. This suggests that there is matter on the outskirts of the galaxy that cannot be seen [14].

Later, in the 1970s, evidence was mounting for the existence of DM, with gravitational lensing being used to probe the distribution of DM. This follows from Einstein's Theory of Relativity: massive objects in the universe bend spacetime, affecting the path of matter and light nearby [16]. Light originating at a distant source will be bent by a large mass between its origin and the observer, such as was observed by Walsh *et al.* when imaging a quasar, whose signal was gravitaionally lensed by an intermediate galaxy [17]. The distant object's distorted image was studied to infer the mass of the galaxy in between,

which again lead to the conclusion that there was more mass than the luminosity of the object would indicate.

Through an extensive microlensing survey of ordinary, baryonic-matter objects in the universe, it was ruled out that DM could be formed from objects such as neutron stars or black holes [18]. However, this did not go as far as completely ruling out that DM be formed of other forms of baryonic matter. Big bang nucleosynthesis (BBN) and the cosmic microwave background (CMB) provide further evidence counter to DM being baryonic [14]. The observed abundance of chemical elements in the universe can be predicted using BBN, the production of nuclei during the early phases of the universe. This study finds that only about 20% of the total matter density of the universe is formed of baryonic matter, leaving the remainder to be non-baryonic. In addition, studies of small irregularities in the CMB indicate that the majority of matter in the universe interacts with ordinary matter only through the gravitational force, and concluded too that baryonic matter only constitutes a fraction of the total matter in the universe [19].

More recently, evidence has come from the bullet cluster: where a small galaxy sub-cluster collides with a larger cluster. The two clusters pass by each other seemingly without interacting, due to the large distance between galaxies. However, the hot baryonic gas between the galaxies was compressed in the collision and shock heated, causing a large amount of X-rays to be emitted, which were observed at the Chandra X-ray Observatory [20]. The location of the radiation (*i.e.* where most of the baryonic mass of the clusters exists) was compared to the location of the majority of the total mass of the clusters, using gravitaional lensing. It was found that there was a discrepancy between the two points, showing that most of the mass of the galaxies is not baryonic [21, 22].

Seeing as we have only observed the gravitational interactions of DM, proposed DM particle candidates are electrically neutral, massive particles that do not interact strongly with other matter. It is also required that DM is stable, in order for it to exist today, and that it is 'cold', or nonrelativistic, such that structure formation is possible in the universe. It is thought that a DM candidate can be treated as an extension to the SM.

As a result of the lack of detailed knowledge about DM, the breadth of theories for its makeup is large. As such, searches for DM take multiple forms: indirect searches, direct searches, and searches at collider experiments [23]. Each of these different search-types exploits a different potential DM interaction: differences in cosmic rays produced in DM processes compared to cosmic ray signals from standard astrophysical sources [2, 24],

DM recoiling off SM-matter [25, 26], and DM being produced in high energy particle collisions [27–29], respectively.

Different experiments are built for each of these purposes, each setting limits on the possible mass of DM, and the interaction types possible. So far, no clear evidence for DM has been found using any of these, but tighter limits are set with each new iteration of a search. The DM searches most relevant to this thesis are indirect searches, in which different, complementary phase spaces of cosmic rays are observed, in the hope of finding an excess flux corresponding to DM sources. Indirect searches look for excesses of antimatter, gamma-rays or neutrinos, each motivated by a different DM model [14].

Cosmic rays are highly energetic particles that travel to earth from outer space. They are mainly protons and helium particles, but many other particles produced in the universe can also be found [30]. Cosmic rays that travel directly from their source to earth are called *primary* cosmic rays, whilst those produced in interactions in the interstellar medium (ISM) are known as *secondary* cosmic rays. The latter either come from collisions of other cosmic rays with the ISM, or unstable nuclei in the ISM decaying. These standard processes form the backgrounds to an indirect DM measurement.

Some theoretical DM models provide mechanisms for DM annihilation into standard model particles and antiparticles, $\chi\chi \rightarrow b\bar{b},\ W^+W^-,\ e^+e^-$, where χ are the DM particles, and the type of SM particles produced is dependent on the type and mass of the DM [14]. There are few standard sources of antimatter in the galaxy. Experiments exploit this to search for DM annihilation by looking for an excess in the antiparticle flux with respect to the expectation from standard astronomical processes [31]. Previous observation efforts, by PAMELA [32], AMS [33] and FermiLAT [31], were sensitive to light antimatter signals, such as positrons or antiprotons, and, though a small excess in the signal over expected astrophysical backgrounds was observed in each, the signals were not significant enough to be unambiguously attributable to DM [30]. In particular, the positron excess at AMS can be shown to have been caused by pulsars [34]. To try to enhance the significance of any potential DM signal, searches can focus instead on heavier antimatter. As the antiparticle mass increases, the standard astrophysical flux decreases, due to higher threshold energies for creation of heavier particles. Antideuterons, $\overline{d}$, are one possible signal, and the antideuteron body of research now spans many different experiments and collaborations [35, 36].

There are multiple experiments focusing on the detection of antideuterons in cosmic rays, either currently taking data, or soon to be launched [37, 38]. In all of these, only a

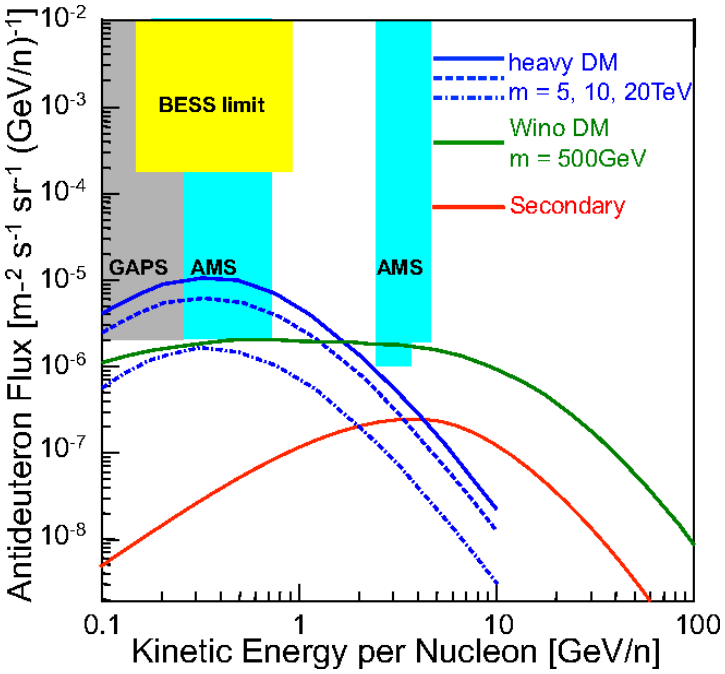


Figure 2.5: Preficted antideuteron flux as a function of kinetic energy per nucleon for annihilation of DM particles with masses $m_{DM} = 0.5, 5, 10, 20\,\text{TeV}$, and expected flux of secondary antideuterons. Also the antideuteron observation limits for experiments AMS-02, GAPS, and BESS. From Ref. [2].

very small antideuteron signal is expected, as can be seen in Fig. 2.5 from Ref. [2], which shows the predicted antideuteron fluxes as a function of kinetic energy per nucleon for DM particles of different masses. Alongside the fluxes, observation limits for different antideuteron experiments are shown, where the overlap between the observation and predicted fluxes is small. This small overlap means that any uncertainty arising from the expected background deuteron flux must be minimised in order to find significant evidence for DM. Uncertainties arise due to the lack of knowledge of both how antideuterons could be formed by the decay or annihilation of DM, as well as their production in standard processes. In addition, little is known about the propagation of antideuterons through the cosmos [2].

High energy physics experiments can produce results that help to reduce both of these uncertainties. Measurements at high energy experiments can be used to calibrate and develop deuteron formation models, and thus further the prospects of indirect DM searches [39]. The ALICE experiment at the LHC has already performed measurements of deuterons and antideuterons in both *pp* and heavy-ion collisions [1]. In Fig. 2.5 it can be seen that for heavy DM the kinetic energy per nucleon is expected to be $< 1\,\text{GeV/n}$,

whilst for secondary antideuterons from the ISM, the kinetic energy is higher, up to $\mathcal{O}(10$ GeV/n). The measurements performed at the ALICE experiment offer coverage in the lower kinetic energy per nucleon region [1]. Measurements complementary to these at other high energy particle physics experiments will help to validate these results and provide information about unexplored phase-space. Similarly to the ALICE detector, the LHCb detector records *pp* and heavy-ion collisions at the LHC, but in a different kinematic region. The measurements at the LHCb experiment will cover a higher momentum region than the measurements performed by the ALICE experiment, and will provide information that is useful for the calibration of secondary antideuterons. Therefore, a measurement of deuteron and antideuteron production at the LHCb experiment will provide new insights for the field.

Chapter 3

Deuteron formation models and simulation

3.1 Deuteron formation models

Deuterons are not an exotic particle, but the mechanism that governs their production is not well known. Owing to the low binding energy of the proton and neutron (2.23 MeV), the formation is not easily observable [4]. As a result, models for deuteron formation are not very sophisticated. The so-called *Coalescence* model has been in use since the 1960s [40, 41], and, despite its simplicity, has provided a satisfactory description of most experimental measurements of deuteron production. It is only now, when antideuterons have become such an important signature for DM [2], that a more complex model is required, which can provide more precise predictions [42]. One such model is the *Cross-section* model, proposed recently by Dal and Raklev [43]. In the following sections, details of these two deuteron formation models will be provided, as well as a brief overview of existing deuteron measurements in section 3.2, and simulation of the models in section 3.3.

3.1.1 Coalescence model

Proposed in 1962 [40], the Coalescence model suggests a simple mechanism for deuteron production. The model assumes that any pn pair, the relative momentum difference of which is below the threshold parameter p_0 will, with certainty, coalesce to form deuteron, but no other pairs will. The threshold, p_0, is determined empirically from experimental measurements. Therefore, the probability of deuteron formation takes the form of a step

function given by

$$\mathcal{P}(pn \to d|k) = \begin{cases} 1, \text{ for } k \leq p_0 \\ 0, \text{ for } k > p_0 \end{cases} \tag{3.1}$$

where k is the absolute momentum difference between p and n, given by $k = |\overrightarrow{p_p} - \overrightarrow{p_n}|$, where $\overrightarrow{p_p}$ and $\overrightarrow{p_n}$, are the vector momenta of the proton and neutron, respectively. This model was proposed when considering deuteron production in showers of nucleons produced from secondary particle beams emitted from a fixed target experiment, but has been used to model deuteron production for many different experiments since then [43].

The Coalescence model is based on an earlier model [44], which required that interactions with a nucleus also occurred, to carry away excess momentum from a coalescing pn pair, as energy and momentum cannot be conserved in this two-body to one-body process alone. In the Coalescence model, the deuteron density in momentum space is proportional to the square of the proton density,

$$\frac{d^3 n_d}{dp^3} \propto \left(\frac{d^3 n_p}{dp^3} \right)^2, \tag{3.2}$$

where p is momentum, n_d and n_p are the number of deuterons and protons respectively, and the proton and neutron densities are assumed to be the same [41]. Here, the proportionality coefficient is independent of momentum, and the energy excess from the coalescence process is transferred to the rest of the system, the details of which are not given. The full relation is given by,

$$\gamma \frac{d^3 n_d}{dp^3} = \frac{1}{2} \frac{4\pi}{3} p_0^3 \left(\gamma \frac{d^3 n_p}{dp^3} \right)^2 \tag{3.3}$$

where $\gamma d^3 n_X / dp^3$ is the relativistically invariant momentum space density for $X = d$, p, and p_0 is the momentum threshold for deuteron production [41], typically $\mathcal{O}(100 \text{ MeV}/c)$ [43].

The model seeks only to give the macroscopic deuteron yield, rather than predicting deuteron production on an event-by-event basis, or giving the infinitesimal details of this process. Though simplistic, it appears to describe the deuteron yield reasonably well for many different experiments [41]. However, p_0 appears to be non-universal, and different

values are computed from measurements with different underlying processes, and for different collision centre-of-mass energies at similar experiments [43].

3.1.2 Cross-section model

To avoid some of the issues of the Coalescence model, the so-called Cross-section model was proposed by Dal and Raklev in 2015, motivated by their interest in antideuteron as an indirect signal for dark matter [43]. This model respects energy and momentum conservation, is not deterministic, and offers a host of processes for deuteron production. This model has been cross-checked using existing data from past deuteron studies and fits well with measurements, performing at least as well as the Coalescence model.

Incorportating statistical and quantum mechanical considerations, the Cross-section model defines the probability of two nucleons binding to form deuteron as a function of the overlap of their wave functions, varying with their separation in momentum space, k. To find the probability, the cross-section corresponding to the capture of n by p is used, so the probability of deuteron formation is given by

$$\mathcal{P}(N_1N_2 \to dX|k) = \frac{\sigma_{N_1N_2\to dX}(k)}{\sigma_0}, \tag{3.4}$$

where $\sigma_{N_1N_2\to dX}(k)$ is the sum of cross-sections for two nucleons, N_1N_2, to bind to produce a deuteron, d, and additional final state particles, X; and σ_0 is a constant of proportionality to be fixed through experimental calibration, analogous to p_0 in the Coalescence model.

There are a number of possible processes with a deuteron as a final state. For processes with low k, the radiative capture process $pn \to d\gamma$ is dominant; above the pion production threshold, processes with hadronic final states, such as $pn \to d\pi^0$, are more probable. Considering charge conservation, it is also possible to have processes in which pp or nn bind to form d, such as $pp \to d\pi^+$ and $nn \to d\pi^-$, which are found to overtake pn processes at higher k values, and cannot be ignored. The cross-sections, σ, for these processes decrease as the number of final state particles increases, so only processes with up to three final state particles are considered: $pn \to d\pi^+\pi^-$, $pn \to d\pi^0\pi^0$, $pp \to d\pi^+\pi^0$ and $nn \to d\pi^-\pi^0$. The same processes can be written for the respective antiparticles to form antideuteron with the same cross-section functions, and it is assumed that the normalisation factor, σ_0, is the same for all processes.

Time invariance is used to compute the cross-section for the $pn \to d\gamma$ channel, which is the inverse of deuteron photodisintegration, $d\gamma \to pn$. As k increases, the photon momentum increases accordingly, which allows deuteron formation with higher k values than permitted by the Coalescence model, up to a few GeV/c, as seen in the cross-sections shown in Fig. 3.1a.

For the remaining two-body processes, some assumptions are made based on isospin invariance for the cross-sections,

$$\sigma_{pn \to d\pi^0} = \frac{1}{2}\sigma_{pp \to d\pi^+} \tag{3.5}$$

and

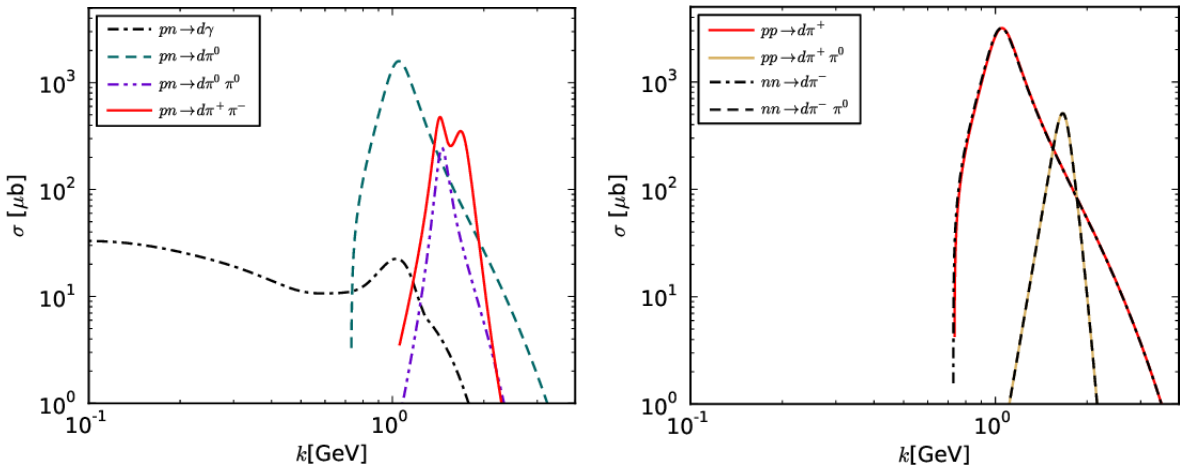
$$\sigma_{nn \to d\pi^-} = \sigma_{pp \to d\pi^+}, \tag{3.6}$$

where d is an isospin singlet with $I = 0,\ I_3 = 0$ [45]. The factor of $\frac{1}{2}$ in Eq. 3.5 is due to the pn initial state being formed of a combination of two isospin states, $I = 0$, 1 and $I_3 = 0$, each carrying a factor $\frac{1}{\sqrt{2}}$, whilst the pp initial state is a single isospin state [6]. The final state including the deuteron in each of the possible processes has $I = 1$, so only one of the two pn initial states can contribute. This leads to a relative suppression in the pn cross-section compared to the pp cross-section. These relations are not exact, as isospin symmetry is broken by the different masses of the nucleons and pions. It is, however, necessary to use them, as there is only data for the $pp \to d\pi^+$ channel, so other channels are approximated by the relations Eq. 3.5 and Eq. 3.6.

The deuteron production cross-sections are found from fits to data, and are set to zero below the kinematic thresholds for the processes. The result of these calculations is shown in Fig. 3.1, as a function of the momentum difference between the ingoing nucleons, k, with peaks induced by intermediate baryon resonances.

Also shown in Fig. 3.1 are the cross-sections for the three-body deuteron production channels, which are calculated using similar principles to the two-body channels. Again, no data is available to make fits to the $nn \to d\pi^-\pi^0$ channel, so the cross-section is found on the basis of isospin invariance and is assumed to match that of $\sigma_{pp \to d\pi^0\pi^+}$. The cross-sections are again set to zero below kinematic thresholds, and there is no angular correlation between the outgoing particles, for simplicity.

In the Cross-section model, $pn \to dX$ processes contribute only a small portion of the total deuteron yield, unlike in the Coalescence model where this is the only contribution.



(a) $pn \to dX$ **(b)** $pp \to dX$, $nn \to dX$

Figure 3.1: The calculated cross-sections for deuteron production channels in the Cross-section model, as functions of the momentum difference between the ingoing nucleons, k. Taken from Ref. [43].

The other processes, binding pp or nn, contribute significantly more to the deuteron yield than pn binding in the higher k range.

3.2 Deuteron measurements

Of the four main detectors at the LHC, only ALICE has made measurements of deuterons and antideuterons, as well as other heavier particles, such as helium [1]. In pp collisions at $\sqrt{s} = 0.9$, 2.76 and 7 TeV, an equal number of deuterons and antideuterons were observed.

In the ALICE detector, a Time Projection Chamber (TPC) and a Time Of Flight detector (TOF) are used to identify charged particles, at central pseudorapidities, $|\eta| < 0.9$. The TPC identifies particles based on their specific ionisation energy loss, whilst the TOF is able to identify light nuclei and antinuclei with transverse momenta above 3 GeV/c using timing information. Deuterons and antideuterons can be identified in the TPC for momenta up to 1.2 GeV/c, which corresponds to a maximum transverse momentum of 1.0 GeV/c. For tracks with higher momentum, deuteron identification is possible by requiring that signals in the TOF and TPC are coincident, allowing for some spread in the energy loss measured in the TPC. The background to the measurement of deuterons comes primarily from lighter charged particles, such as pions, kaons and protons [1].

The measurements of deuterons and antideuterons in ALICE *pp* data are instructive for the measurement of deuterons and antideuterons at the LHCb detector, and parameters that were found by fitting models to the data have been used to predict the yield that is expected at the LHCb experiment.

3.3 Simulation

Both of the deuteron production models have been implemented as an extension to Pythia [46] for LHCb Monte Carlo (MC) simulation, using the Gauss package [47]. In this section, a brief overview of LHCb simulation processes will be given, followed by details specific to the simulation of the two deuteron models.

The production of particles directly from the primary *pp* collision of the LHC beams is handled by the Pythia package, and the subsequent time evolution and decay of these particles is performed by EvtGen [48]. The Pythia algorithm can be modified using plug-in tools in order to generate events in different conditions, or apply cuts to the generated events. It is in this way that each of the deuteron production models is added to the simulation. Once the particle decays have been performed by EvtGen, the particles from the event are propagated through a simulation of the LHCb detector and their interactions with the detector material are simulated using the Geant4 toolkit [49]. The simulation is then processed by the Boole package in order to reproduce the individual subdetectors' responses and their digitisation, such that LHCb simulation and data are both in the same format [50]. The MC then follows the same trigger, reconstruction and analysis procedures as are applied to data. The final simulated events contain the processes undergone in the detector, analagous to information that is available in data, as well as information about the true, generated event, which is not available in data [47].

Deuteron production is implemented after the *pp* collision has been simulated using Pythia, but before the resulting particles have been passed to the EvtGen package. The deuteron production algorithm sifts through the products of the *pp* collision to find any pairs of nucelons that may be able to bind according to the chosen deuteron production model[1].

Firstly, nucleons from the event are listed, and their order randomised. The algorithm loops through the list, comparing every nucleon to every other nucleon, and decides,

[1] Also charge conjugates.

according to the chosen deuteron production model, whether to bind each pair into a deuteron. If a pair is found to be able to bind, a vertex is added to the event with incoming nuclei, and a deuteron and other products outgoing.

In the Coalescence model, the criteria for a pair of nucleons in the event to bind is simply that they are not the same species, and that their relative momentum difference, k, is smaller than a threshold, p_0, as presented in section 3.1.1. In this implementation, the value of p_0 is taken from fits to measurements of deuterons in pp collisions at the ALICE Experiment, as this has the closest experimental conditions to the LHCb experiment. Thus, a value of $p_0 = 194.5\,\text{MeV}/c$ is used [43]. If a pn pair passes the requirement, $k < p_0$, a vertex is added to the event with incoming proton and neutron, and outgoing deuteron and photon. In many versions of the Coalescence model, the outgoing photon is not included, such that a deuteron is the only product [51]. However, because a two-to one-body interaction does not conserve momentum, a recoiling photon is added.

It was found to be very unlikely in the LHCb detector collision conditions that one proton would satisfy the condition of $k < p_0$ with more than one neutron, and vice versa, due to the fact that, in general, the nucleons have much higher momenta than the threshold, p_0. To avoid a clash in the simulation, before comparing two nucelons, it is verified that each nucleon is 'stable', and therefore has no end vertex – meaning that it neither decays, nor binds with another particle. For a clash, or double-counting, to occur at a problematic rate, the threshold, p_0, would have to be increased by multiple orders of magnitude. Verifying that the nucleons do not have an end vertex, in order to eliminate double-counting, is thought not to introduce a bias in the overall deuteron production. Therefore, there would be no advantage in using a more complex method, such as one that explores every possible binding of every nucleon.

In the Cross-section model, there are multiple channels of deuteron production, through binding pn, or also pp and nn, as presented in section 3.1.2. In the deuteron production simulation, the guidance given in the paper by Dal and Raklev was followed [43]. In a similar manner to that used with the Coalescence model, pairs of nucleons created in a pp collision are compared to each other. In this case however, their momentum difference is used to find the cross-section of possible different binding mechanisms. A random number is then thrown to determine whether one of these routes is followed, or whether instead the pair will not bind. To make this decision, the cross-sections for each of the possible processes are summed and normalised by the σ_0 parameter, found from measurement, $\sum_{i=0}^{p} \frac{\sigma_i}{\sigma_0}$, where i runs over each process, out of a total of p possible processes. A random number, r, [0,1] is generated. If $r < \sum_{i=0}^{j} \frac{\sigma_i}{\sigma_0}$ where $j \leq p$ and

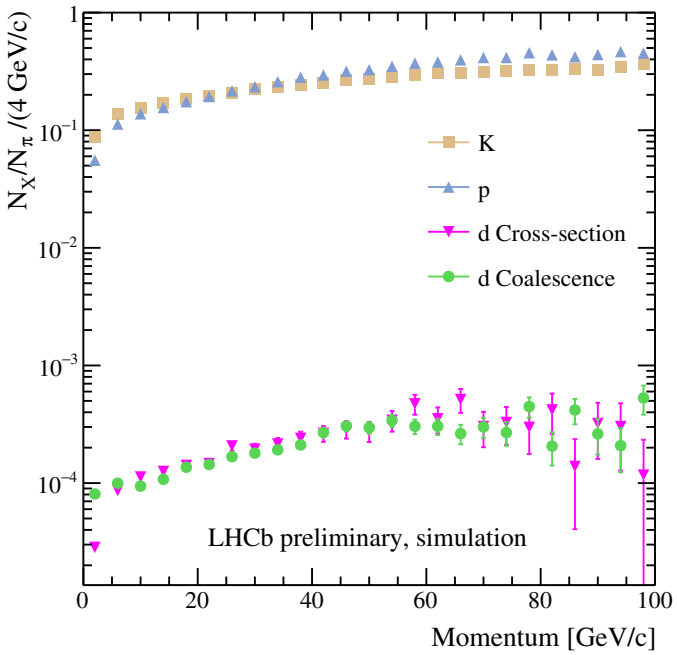


Figure 3.2: Ratios of charged particles to charged pions in bins of momentum within the LHCb detector acceptance from generator-level simulation of deuterons, using both the Coalescence and Cross-section models.

$j \geq i$, the j^{th} channel for deuteron production will be followed. If $r > \sum_{i=0}^{p} \frac{\sigma_i}{\sigma_0}$, then the nucleons will not be bound and no deuteron will be created. If it is decided that a deuteron should be created, a vertex is added to the event, with incoming nucleons, and outgoing deuteron and other products, according to the chosen process. Otherwise, the nucleons are left intact and are free to be bound with others in the event.

This is the first example of the deuteron production mechanisms being implemented in LHCb simulation, and as such is the first indication of what yield of deuterons might be expected in this environment. Fig. 3.2 shows the rates of charged particles to charged pions in bins of momentum within the LHCb detector acceptance, from simulation of pp collisions with centre of mass energy $\sqrt{s} = 13$ TeV. These tracks have not been propagated through the LHCb detector using Geant4, so it is an indication of the deuteron yield directly from the collisions, where losses due to detector effects are not taken into account. The deuterons in the plot were produced through both the Coalescence and Cross-section models, where the sample using the Cross-section model is 60% the size of the Coalescence model sample, leading to greater statistical uncertainty. It can be seen that the yield of deuterons is expected to be much lower than the yield of other

charged particles at all momenta within $0 < p < 100\,\mathrm{GeV}/c$. Across this range, there are approximately ten-thousand charged pions to every deuteron, and a thousand times as many kaons and protons as deuterons.

The values of the model parameters used in the simulation of both the Coalescence and Cross-section models are those that best describe the data recorded by the ALICE experiment [1, 43]. Therefore, it makes sense that the yields for deuterons from the two models overlap with each other in Fig. 3.2.

Chapter 4

The LHCb detector

4.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [52] is a circular collider, producing pp, Pbp and PbPb collisions at high energies. Other ion collisions, such as Xe-Xe, are also possible [53]. Four primary detectors sit along the ring: two general purpose detectors, ATLAS [54] and CMS [55]; the ALICE experiment [56], which is optimised for heavy ion physics; and the LHCb detector, which was designed to study flavour physics [57].

The LHC is situated in the tunnel previously occupied by the Large Electron-Positron (LEP) collider [52], 100 m below the ground on the Swiss-French border, near Geneva. The LHC has been in operation since 2009, with a period of collisions from 2009 to 2013 at maximum centre of mass energy $\sqrt{s} = 8\,\text{TeV}$, and then a second running period from 2015 to 2018 at energies up to $\sqrt{s} = 13\,\text{TeV}$ [53].

Protons in the LHC are accelerated around two rings of circumference $\sim 27\,\text{km}$ in opposing directions, and brought together at crossing points at each of the four detectors. The LHC protons are supplied by a chain of accelerators, which bring the proton beams close to their final intensity [52]. The protons are accelerated in stages in the CERN accelerator complex begining in a linear accelerator, Linac2, followed by three circular accelerators, the Proton Synchrotron Booster, the Proton Synchrotron, and finally the Super Proton Synchrotron, from which they are injected into the LHC with an energy of 450 GeV. The protons are grouped in 'bunches' of $\sim 10^{11}$ protons, which are organised into 'trains', with a spacing of 25 ns between each bunch. Typically, each train contains 72 bunches, followed by 12 empty spaces; a configuration which provides a gap for beams

to be dumped cleanly when necessary. At full capacity, the LHC rings can contain 2,808 bunches.

Once in the LHC beampipes, the proton bunches are accelerated to a maximum energy of 6.5 TeV each, using sixteen radio frequency chambers. A total of 1,232 dipole magnets are used to keep the bunches on a circular path, whilst 392 quadropole magnets are used to squeeze and collimate the beams. Both types of magnet are superconducting, and are cooled using superfluid helium to their operating temperature of 1.9K.

4.2 The LHCb experiment

Designed and optimised to study B-physics, the LHCb detector has a forward orientation, with a pseudorapidity acceptance $1.9 < \eta < 4.9$ [57]. This was chosen to maximise the flux of b hadrons, which are produced largely in the forward and backward directions at the LHC.

As shown in Fig. 4.1, the LHCb detector comprises several sub-detectors, each designed and optimised to track or identify different particle types [58]. The coordinate system used for the experiment has the z-axis running along the beampipe, the y-axis in the vertical direction, and the x-axis in the horizontal direction, pointing out of the page in Fig. 4.1. Starting on the left of the diagram, the pp collisions occur within the VErtex LOcator (VELO), which uses silicon modules very close to the colliding beams to find the interaction point, known as the *primary vertex* (PV), and any decay, or *secondary*, vertices which occur within the first 10.6 cm of flight [57].

Next, particles pass through the first of the RICH detectors, with the second RICH detector further downstream, after the magnet. These are both used for the identification of charged particles. They complement each other in kinematic coverage and acceptance, with RICH 2 more effective for more boosted tracks than RICH 1.

In between these two RICH detectors are the tracking stations and the dipole magnet. Charged particles interact with electrons in the tracking stations to either excite them, or ionise the medium, and lose energy in doing so. The excitation along the particle's path is detected in each tracking station, from which the trajectory is reconstructed. The magnet is used to bend the path of charged particles, where the extrapolated curvature of the track is used to measure the particle momentum. The sign of the charge of the track is also measured using the direction of curvature.

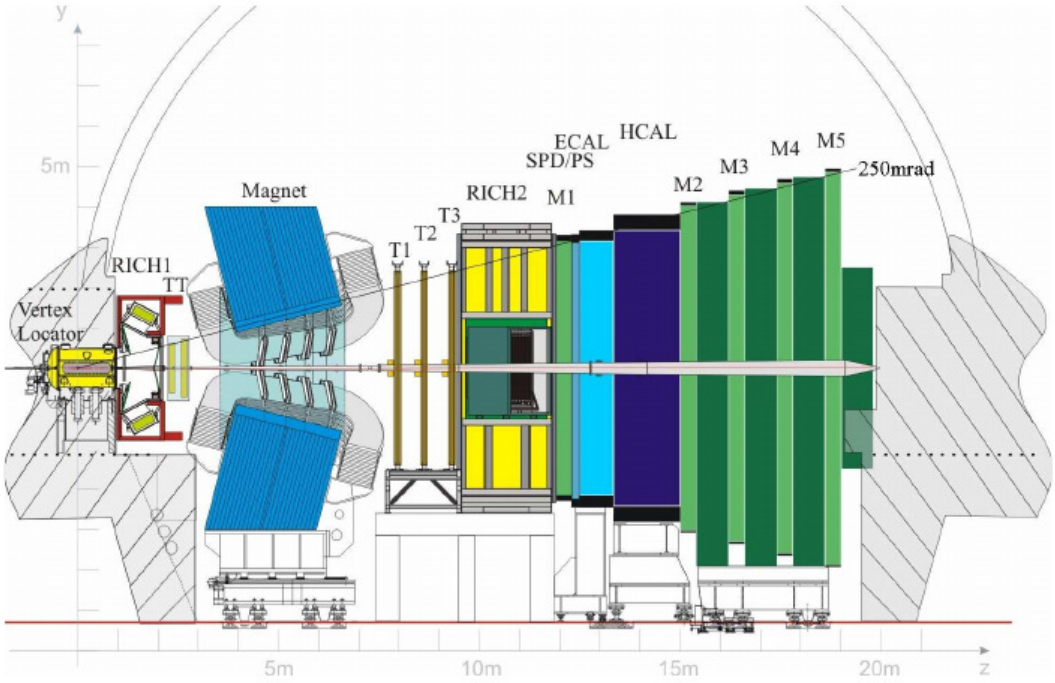


Figure 4.1: A schematic diagram of the LHCb detector, with the *pp* collision point on the left within the Vertex Locator. From Ref. [57].

The remaining sub-detectors contribute further to particle identification, starting with the preshower and scintillator pad detectors (PS/SPD), for separating electrons and pions, then the electromagnetic and hadronic calorimeters (ECAL, HCAL), and finally the muon stations, (M1 - M5). Each type of particle loses energy through different processes as it passes through the detector material, where the energy loss is used for particle identification. At high energies, electrons lose energy through Bremsstrahlung, whereby they interact with the nuclei in the detector medium and emit photons [4,59]. High energy photons lose their energy by producing an electon-positron pair, known as *pair production*. In an electromagnetic shower, both of these effects combine, with the electron losing its energy through Bremstrahlung, and the emitted Bremstrahlung photons producing pairs of electrons and positrons in turn, which then emit Bremsstrahlung photons themselves, and so forth, until the kinetic energy of the system has dissipated. Hadrons passing through the detector will interact with nuclei in the medium through the strong force; they scatter off nuclei in the material and produce secondary particles, mostly pions. The resulting hadron showers are more expansive than electromagnetic showers.

The muon stations sit at the end of the LHCb detector, to the far right of Fig. 4.1. Muons lose their energy primarily through interactions with the electrons in the detector

medium, but the energy loss is small. For typical muon energies at the LHCb experiment, muons are minimum ionising particles, meaning that their mean energy loss rates are close to the minimum, when considered as a function of their momentum. Therefore, muons will not lose much energy, and will pass all the way through the detector, and the muon stations, whilst hadrons and electrons will be stopped in the calorimeters upstream. Hits in the muon stations that correspond to tracks through the detector can be identified as muons with high efficiency.

The elements of the LHCb detector will be described in more detail in this chapter, with particular detail given to the performance of the RICH detectors, in section 4.2.3, which are crucial in the measurement of the deuteron production cross-section.

4.2.1 Tracking

Tracking is used primarily to measure the momentum of charged particles, by measuring the curvature of tracks when they are bent in the magnetic field provided by the LHCb detector magnet. To do this, the tracks must be measured both before and after the magnet. The track trajectory also provides an important input to particle identification in the RICH detectors, and to the reconstruction of vertices in the VELO. The quality of the track reconstruction can be used to distinguish between real and fake tracks.

VELO

Used to find primary and secondary vertices, the VELO is positioned a few millimetres from the interaction point during data-taking [57]. The VELO detector is distanced from the beams during injection or beam dump, when beam conditions are unstable and the hardware could be damaged if it was left closed. Semicircular silicon sensors surround the colliding beams, and measure the passage of particles in cylindrical coordinates, r and ϕ. The r and ϕ sensors are interleaved along the beam direction, as shown by the red and blue colours in Fig. 4.2. The silicon detectors are surrounded by a thin corrugated aluminium casing, so that they can be contained in a vacuum that is separate from the LHC machine vacuum. This protects the silicon detectors from picking up voltages from the beam acceleration, and protects the LHC vacuum from outgassing of the detector modules.

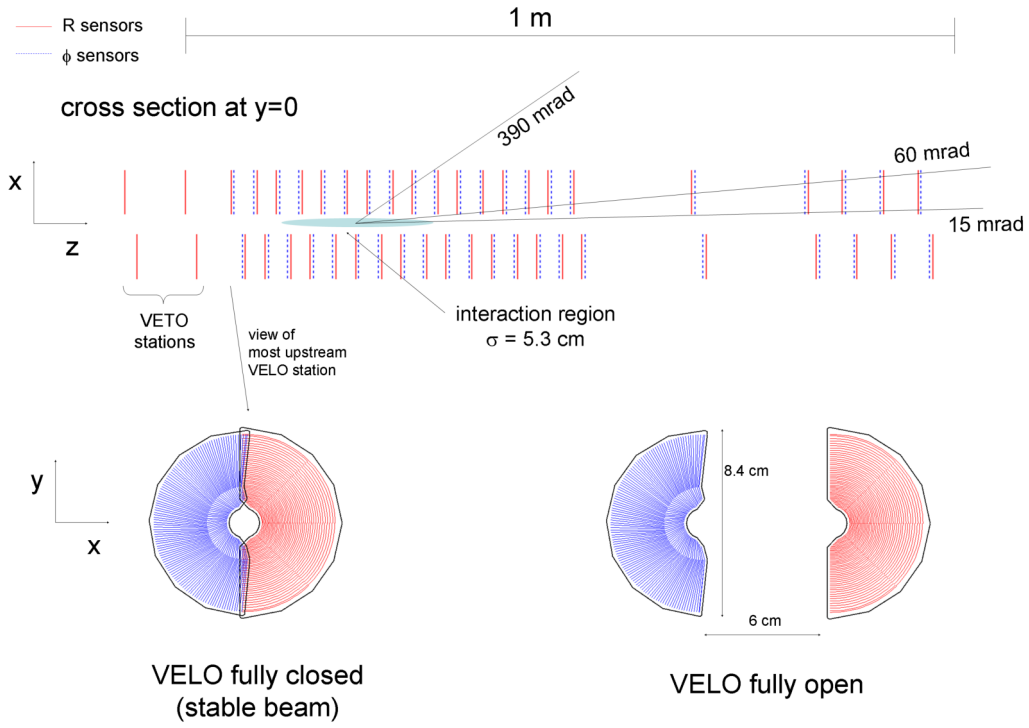


Figure 4.2: A cross-sectional diagram in the (x, z) plane of the VELO detector silicon sensors at $y = 0$, in the fully closed position. Below are the front faces of the first silicon modules, in the closed and open positions. From Ref. [57].

Measurements of tracks and vertices in the VELO are used in triggering, to select data enriched with b hadrons, and to measure decay lifetimes and impact parameters. The VELO maintains a signal to noise ratio of greater than 14, to ensure efficient trigger performance, a channel efficiency in excess of 99%, and a spatial cluster resolution of about 4 µm for a track at an angle of 100 mrad, in order to achieve an impact parameter resolution of less than 16 µm.

The VELO can detect particles that emerge within 10.6 cm of the primary vertex along the z direction, in a pseudorapidity range of $1.6 < \eta < 4.9$. A track that travels within the LHCb detector with $\theta < 300$ mrad, should cross at least three VELO silicon modules. As seen in Fig. 4.2, the modules surrounding the interaction region are more closely spaced than the modules further up- or downstream; this is to improve the resolution of the primary vertex, by reducing the exrapolation of the tracks.

Trackers

The tracking stations are placed downstream of the VELO, before and after the magnet. Two detector technologies are used in these subsystems: silicon trackers for the Tracker Turicensis (TT) and Inner Tracker (IT), and a straw tubes for the Outer Tracker (OT).

The TT and IT use silicon microstrip sensors with a thickness of $\sim 400\,\mu\text{m}$. The TT is upstream of the magnet, with a size of $1.5 \times 1.3\,\text{m}$ [57]. The IT and OT are downstream of the magnet and RICH 2. The IT covers the plane closest to the beampipe, with size $1.2 \times 0.4\,\text{m}$, at the centre of the T1, T2 and T3 tracking stations in Fig. 4.1, and the OT surrounds this, with an active area of $5.0 \times 6.0\,\text{m}$ [60].

The design of the silicon trackers was driven by considerations for good spatial resolution, hit occupancy and pile-up prevention, hit efficieny, radiation damage, material budget, and number of readout channels. A resolution of $50\,\mu\text{m}$ and a hit efficiency in excess of 99% have been achieved [58]. The average number of active channels in the tracker per event, known as *occupancy*, ranges between 0.2 and 1.9% in both the TT and the IT, with the higher occupancies closest to the beampipe.

Further away from the beampipe, the occupancy of the detector is much lower, such that a gaseous straw tube detector can be used. The inner diameter of each tube is 4.9 mm, with a length of 2.4 m, which is filled with a mixture of Argon and CO_2 gas. These provide a drift time of less than 50.0 ns [57]. The drift time resolution is 2.4 ns [61]. The drift-coordinate resolution achieved is $205\,\mu\text{m}$ [58], with an occupancy of about 10% for each tube.

4.2.2 Track reconstruction

Different types of tracks are defined by where they leave hits in the detector, as shown in Fig. 4.3, [58]:

- **Long tracks:** pass through the whole tracking system, leaving hits in the VELO, and tracking stations downstream. Because they traverse the full magnetic field, their momentum estimate is precise.
- **Upstream tracks:** with momentum too low to traverse the magnet, these tracks only pass through the VELO and TT. If their momentum is $> 1.0\,\text{GeV}/c$, they could generate a signal in RICH 1.

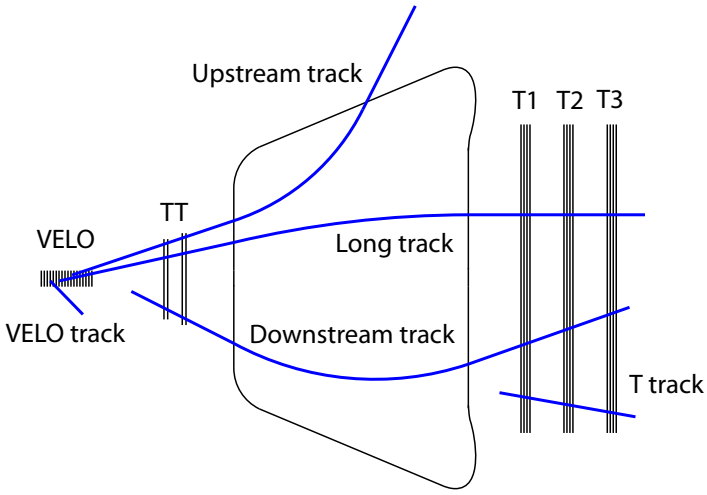


Figure 4.3: A simplified schematic diagram of the LHCb detector tracking stations and the defined track types. From Ref. [62].

- **Downstream tracks:** usually coming from long-lived particles that decay after the VELO, the tracks only appear in the TT and T stations.
- **VELO tracks:** travelling at a large angle, or backwards, these tracks only leave traces in the VELO, but are useful for identifying primary vertices.
- **T-tracks:** these pass only through the T stations, typically originating from secondary interactions.

The reconstruction of long tracks begins with the reconstruction algorithm searching for straight line trajectories in the VELO which have at least three hits in each of the r and ϕ sensors. This information from the VELO is then combined with the downstream tracking. The momentum and trajectory of a particle can be measured from a VELO track and a single hit in a tracking station. The trajectory is then used as a starting point to search for other hits in the tracking stations, in order to find the best possible combination of hits to define a long track [58].

Downstream tracks are reconstructed starting in the trackers after the magnet, and extrapolating back through the magnet to the TT, whilst upstream tracks are found by extrapolating VELO tracks to the TT. At least three TT hits are required for these to be successful.

Tracks are fitted using a Kalman filter [63], which takes into account multiple scattering and energy losses due to ionisation, both of which could affect the trajectory of tracks. The χ^2 per degree of freedom of the fit is used as an estimator of the quality of the track reconstruction.

Ghost tracks

Ghost tracks, g, are tracks that have been created in the reconstruction, but are not associated to a charged particle. These are created during the pattern recognition step in tracking, when a pseudo-random combination of hits in the detector is formed into a fake track [64], often due to a mismatch in tracks when extrapolating from the VELO to the tracking stations further downstream. In data and simulation, these will appear similar to real tracks, and can therefore skew measurements.

Ghost tracks often have attributes similar to real tracks, as they can be formed from two tracks in the detector that are falsely joined up into one track. However, compared to real tracks, ghost tracks often have some expected hits missing along their paths, and have a poor track fit quality. These two features can be use to discriminate between real tracks and ghosts in data and simulation, and are inputs to particle identification algorithms.

Ghosts tracks in simulation can be selected as those that are not associated with a particle, using Monte Carlo truth information. Therefore, the shape and behaviour of ghosts in data can be estimated using simulation. However, it is known that the levels of ghost tracks in simulation and data differ [65], so the behaviour of ghosts in simulation should be calibrated using control samples in data.

4.2.3 Particle identification

Techniques for particle identification (PID) are highly dependent on the kinematics of the particles. At low energies, time projection chambers suffice, but at higher energies, such as those found at the LHCb experiment, ring-imaging Cherenkov detectors are more effective.

In this section, the key subdetectors for particle identification will be described, focusing mainly on the RICH, which is used for charged particle identification, but also covering the muon system and calorimeters.

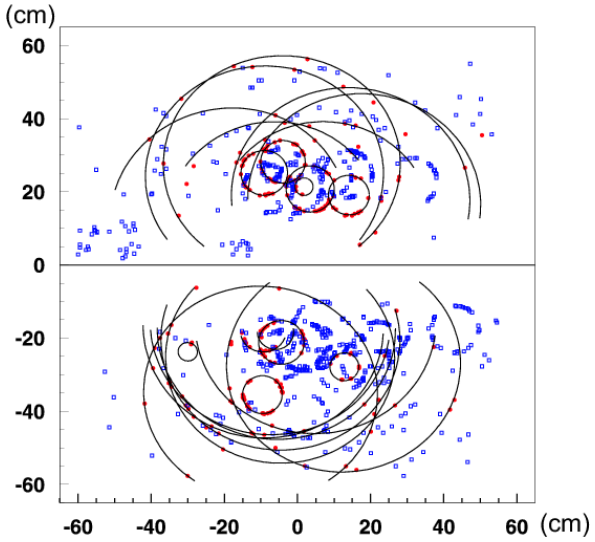


Figure 4.4: Simulated event display in RICH 1, with fitted rings superimposed. From Ref. [67].

Ring imaging Cherenkov detectors

A pair of RICH detectors with gas radiators are used to identify high energy charged particles at the LHCb experiment. A charged particle of mass m, passing through a medium faster than a photon, will emit Cherenkov photons at an angle with respect to the direction of travel, θ_C, given by

$$\cos\theta_C = \frac{\sqrt{m^2 + p^2}}{np}, \tag{4.1}$$

where p is the momentum of the particle, and n is the refractive index of the medium [66]. These photons are produced in a cone around the particle track, which can be focused to rings of photons on a plane of photon detectors, as shown for simulated events in Fig. 4.4. The photons detected in the RICH, combined with measurements of the particle momentum from other parts of the detector, can be used to separate charged particles of different mass.

The main physics programme at the LHCb experiment requires excellent separation of kaons and pions, in order to measure very rare decays of b hadrons, to test the SM [68]. Therefore, the radiators used for Cherenkov photon production are chosen to provide pion–kaon separation over a wide momentum range. The radiators, refractive indices, acceptances and momentum coverage of each of the RICH detectors are listed in Tab. 4.1 [69].

Parameter	RICH 1	RICH 2
Pion–kaon separation momenta (GeV/c)	3 – 50	15 – 100
Angular acceptance (mrad)	20 – 300	15 – 120
η coverage	1.89 – 4.39	2.8 – 4.89
Radiator	C_4F_{10}	CF_4
Refractive index	1.0014	1.0005
N_0	28	24
σ_c (mrad)	1.6	0.7

Table 4.1: Nominal parameters for the radiators of the RICH detectors at the LHCb experiment, where N_0 is the maximal number of photons emitted at the point of saturation of the Cherenkov angle, and σ_c is the angular resolution of the detector.

The number of Cherenkov photons that are emitted from high energy charged particles follows a Poisson distribution, dependent on the track momentum and the refractive index of the radiator, with mean

$$N_c = N_0 \left(\frac{\beta^2 n^2 - 1}{\beta^2 (n^2 - 1)} \right), \tag{4.2}$$

where N_0 is the maximal number of photons emitted at the point of saturation, and $\beta = v/c$, where v is the particle velocity [70]. The ring of photons emitted will have an average angle to the track direction corresponding to the expected Cherenkov angle, θ_c, given by Eq. 4.1, and a width corresponding to the angular resolution of the detector, σ_c. The values for these parameters in the radiators in the RICH detectors at the LHCb detector are listed in Tab. 4.1.

A simulated event display in RICH 1 for a typical event is shown in Fig. 4.4. For each track in the event, the RICH algorithm computes probability density functions (PDF) for photon rings with expected θ_c for different track hypotheses. These rings are shown as black circles in Fig. 4.4. The red photon hits are hits that can be associated with a projected ring, and the blue hits are those that have not been paired to a ring.

RICH hardware

A combination of spherical and flat mirrors is used to focus the Cherenkov light and reflect it out of the detector acceptance [57], with a vertical layout in RICH 1, and a

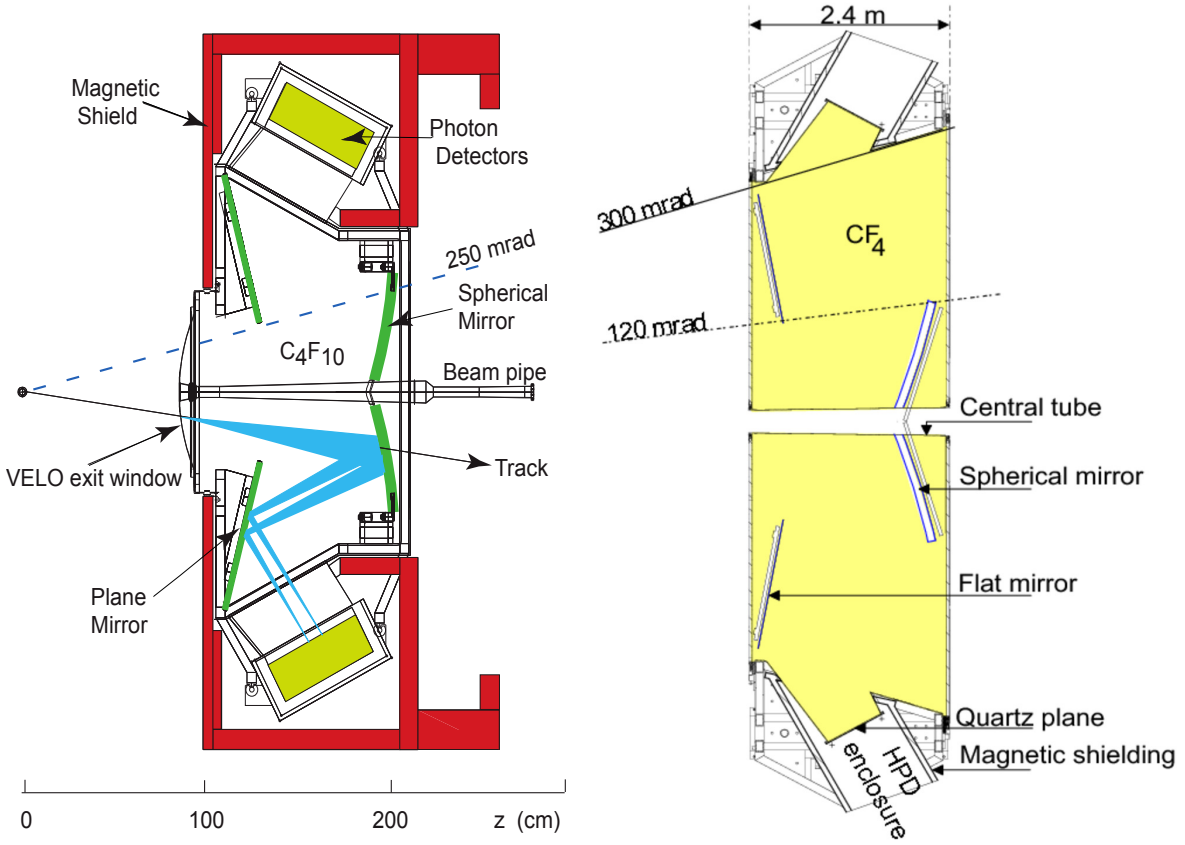


Figure 4.5: Schematics of a side view of RICH 1 (left) and top view of RICH 2 (right), from Ref. [57].

horizontal configuration in RICH 2. Cherenkov photons are detected by Hybrid Photon Detectors (HPD) [71], which are surrounded by iron shields to reduce the magnetic field in the detectors.

With optimal performance for lower momentum tracks, RICH 1 is positioned close to the interaction point, in order to limit its overall volume. The low momentum tracks are bent by the magnet further downstream, such that they travel outwards. Therefore, if RICH 1 was positioned after the magnet, it would have to be very large for the tracks to pass through its acceptance. RICH 2, which is optimised for identification of higher momentum charged particles, is placed after the magnet, because high momentum tracks are less affected by the magnet, and travel closer to the beampipe. RICH 2 is positioned after the tracking stations in order to reduce the material budget for the tracking.

Schematics of each of the RICH detectors are shown in Fig. 4.5, with a side-view of RICH 1 on the left, and a top-view of RICH 2 on the right. In both RICH detectors,

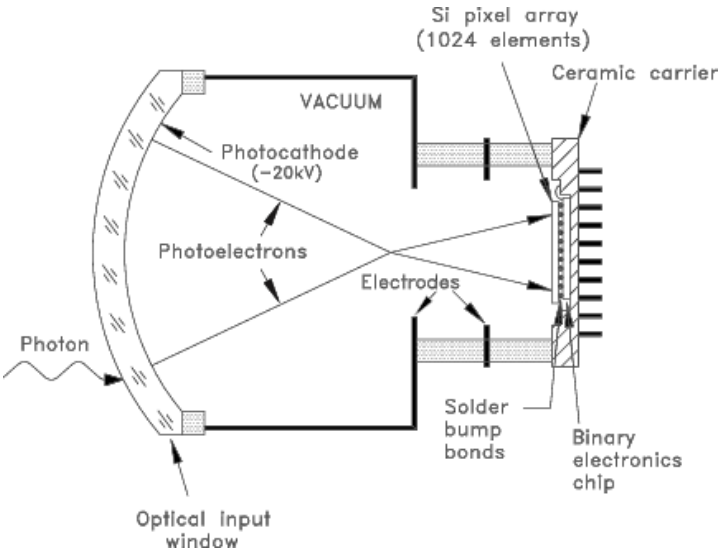


Figure 4.6: Schematic of a pixel HPD, illustrating example photoelectron trajectories, from Ref. [71].

incoming tracks pass through the radiator for the length of the detector, and then through spherical mirrors, before leaving the RICH casing and continuing further downstream. Cherenkov photons around the tracks will be focused by the spherical mirrors and reflected onto flat mirrors, from which they are reflected again into photon detectors. The flat mirrors are used to limit the length of the detectors along the beampipe direction. All mirrors have $> 90\%$ efficient reflectivity for wavelengths $200 < \lambda < 600\,\text{nm}$.

Cherenkov photons are focused into ring patterns on the photon detector planes which are outside the detector acceptance [68]. The purpose-built HPDs have vacuum tubes with a 75 mm active diameter. A schematic diagram of an HPD is shown in Fig. 4.6, where it can be seen that photons pass through the optical input window, whereupon photoelectrons are created [71]. The photoelectrons are focused onto a silicon pixel array using a voltage of $-16\,\text{kV}$. The pixel array in each tube has 1024 pixels of area $2.5 \times 2.5\,\text{mm}^2$, in a square configuration. In RICH 1 there are a total of 196 tubes, split over two planes, and 288 tubes split over two planes in RICH 2, separated from the radiator gases by quartz windows. The front-end electronics are also contained in the tubes, and bump-bonded to the silicon pixel sensors, such that they have very low noise. The tubes also have a high photon detection efficiency, of about 82 % [68].

The RICH reconstruction relies on the LHCb experiment Control System (ECS) to not only operate the RICH detector components, but also to collect information about the RICH environment during operation [68]. The positioning of the RICH mirrors, and

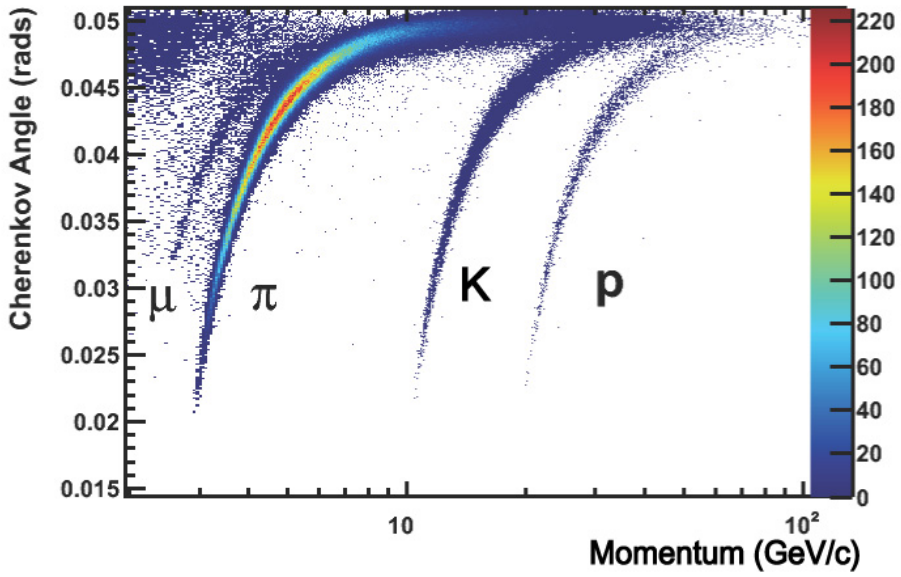


Figure 4.7: Measured Cherekov angle vs. momentum for different charged particles in RICH 1. From Ref. [68]

temperature, pressure and purity of the gas radiators are all recorded. These allow for precise recalculation of the refractive indices of the gases, which are fed automatically to the data-taking conditions database, which allows for the accurate calculation of the Cherenkov angles used in the RICH algorithm.

RICH performance

Cherenkov angles for charged particles in RICH 1 as a function of momentum, in the range $0 < p < 100\,\text{GeV}/c$, are shown in Fig. 4.7. The tracks used for the signals in this plot were selected from clean events, which are atypical during normal LHCb experiment data-taking, where there are large backgrounds present. In this plot, features due to the threshold momenta and photon emission saturation can be seen. Firstly, the momentum thresholds, *i.e.* the minimum momentum at which a given particle will undergo Cherenkov radiation when travelling through the RICH radiator. For each of the labelled particles, a sweeping signal of Cherenkov photons, going from small to larger Cherenkov angle with increasing momentum, turns-on at the corresponding threshold momentum. Secondly, the increase in the signal strength above this can be clearly seen as the distributions widen and bolden with increasing momentum, following Eq. 4.2. Lastly, all of the signals tend towards the same value of Cherenkov angle, $\approx 0.05\,\text{mrad}$,

	Threshold momentum (GeV/c)	
Particle	**RICH 1**	**RICH 2**
μ	2.0	3.3
π	2.6	4.4
K	9.3	15.6
p	17.7	29.7
d	35.4	59.3

Table 4.2: Momentum thresholds for Cherenkov photon production in each of the RICH detectors at the LHCb detector.

at high momentum, meaning that they produce identical signals at high momenta. When particles exibit this behaviour, they are said to be 'saturated' in the RICH detectors.

Table 4.2 shows the momentum thresholds for charged muons, pions, kaons, protons and deuterons passing through the two RICH detectors. The electron thresholds are not listed, as they are a fraction of 1.0 GeV/c.

The photons measured by the RICH, and the momentum measurement for each particle from the tracking, are combined by the RICH reconstruction algorithm to provide a log-likelihood for each track to correspond to a specific particle species [72]. In this procedure, every track in the event is considered simultaneously, combining information from both RICH detectors. To find the log-likelihoods, the algorithm first assumes that every track in an event is a pion. Then, the change in the overall event log-likelihood is found when changing the particle hypothesis for each track from the pion mass hypothesis to a different charged particle mass hypothesis, namely kaon, proton, muon, electron, or deuteron, whilst leaving all other tracks unchanged. The mass hypothesis that gives the biggest improvement in the event likelihood is identified, and the mass hypothesis for that track is set to its preferred value. The procedure is then repeated, until each track has been assigned its optimal hypothesis, and no further increase in the event likelihood can be found. Each of the log-likelihoods is found by computing PDFs for rings corresponding to θ_c given by Eq. 4.1 for each mass hypothesis. These modelled rings are compared to the real photon hits in the detector, as illustrated in Fig. 4.4, and a minimisation for each mass hypothesis PDF is performed. The output from the algorithm, known as DLLx, is the 'delta log-likelihood'. The change

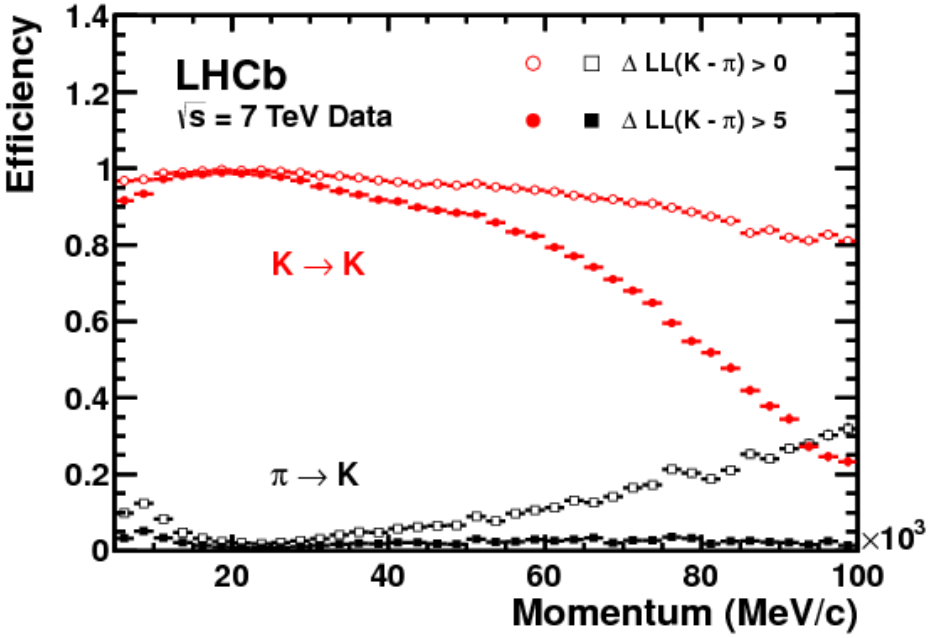


Figure 4.8: Discrimination between kaon and pion tracks for momenta $0.0 < p < 100.0\,\text{GeV}/c$ during Run 1 at the LHCb experiment. From Ref. [68].

in total event log-likelihood forms the so-called DLLx variable, where x denotes the mass hypothesis, alternative to the pion hypothesis [68].

Selections can be applied directly to the DLLx variables in order to choose samples enriched with particular particle species, to a high degree of purity. The separation between kaons and pions during Run 1 of data-taking at the LHC is shown in Fig. 4.8, as a function of momentum, for two selection criteria: DLL$K > 0$ and DLL$K > 5$ [68]. The efficiency is defined as the number of tracks passing the selection, divided by the total number of tracks in the sample. The efficiencies for pions to pass the selection, and therefore be misidentified as kaons, are shown in black, and the efficiencies for kaons to be correctly identified are shown in red. Due to the saturation of the Cherenkov signals for both pions and kaons at high momenta, their separation diminishes in this regime. A steep decline in the efficiency for kaons to be correctly identified using the DLL$K > 5$ selection is seen for tracks with momentum greater than $50\,\text{GeV}/c$, though the rate for misidentifying pions, under this selection, stays constant over this momentum range.

Due to the momentum dependence, if requirements are placed on DLLx variables to select tracks, the track kinematics should also be restricted, and different DLLx selections may be required for different momenta.

Calorimeters

The calorimeters provide identification of electrons, photons and hadrons, as well as measurement of their energy and position. This information is used in the first level of the LHCb trigger (L0) to filter events containing potentially interesting processes [57]. Prompt photons and neutral pions can be identified using the ECAL. For good background rejection, excellent resolution and shower separation are required.

For accurate photon energy measurements, which are used in the L0 trigger, showers from high energy photons must be fully contained. Therefore, the ECAL was chosen to be 25 radiation lengths in thickness. The trigger requirements on the HCAL are less demanding, so a thickness of 5.6 interaction lengths is used, due to space limitations.

All calorimeters have a higher granularity closer to the beampipe, and bigger cells towards the outside. The PS/SPD and ECAL both have three different cell sizes, and the HCAL has just two. The PS/SPD has a total of 12032 detection channels, with cells of dimensions 4×4, 6×6, and $12 \times 12\,\mathrm{cm}^2$.

The calorimeters detect particles using scintillation light that is transmitted to Photo-Multipliers (PMT) using wavelength-shifting fibres. The scintillators in the ECAL, are interleaved with layers of lead, to slow down the particles and generate showers, whilst the PS/SPD consists of one, thicker layer of lead, between scintillator pads [57]. The scintillators in the HCAL are interleaved with iron.

The PS/SPD and calorimeters provide distinction between photons, electrons and neutral pions [58]. This is done firstly by determining whether energy deposits in the calorimeters are preceded by particle tracks or not. The shape of the cluster is then used to distinguish between massive and massless candidates. It is possible for photons to convert to electrons when interacting with detector material, so this must also be taken into account. Two independent estimators are built to identify converted and non-converted candidates. Likelihoods from the PS/SPD, ECAL and HCAL are combined to determine the particle species.

Muon detectors

There are five muon stations: one before the ECAL and HCAL (M1), and four at the most downstream point of the detector (M2 - M5) [58]. The final four stations are interleaved with iron shields to select penetrating muons. Each station is divided into 276

chambers. Multi-wire proportional chambers are used for detection in the muon stations. The chambers closest to the beampipe, where there is a higher expected particle rate, use gas electron multiplier detectors instead.

Muons are identified on the basis of hits in the muon system about the extrapolated trajectory of a track reconstructed in the tracking system. A rectangular window in the x and y dimensions about each extrapolated track is searched for hits. The size of the window is chosen as a function of momentum at each muon station, and separately for each muon system region.

For a muon to traverse the calorimeters and reach the M2 and M3 stations, it must have a momentum greater than $3.0\,\text{GeV}/c$, whilst above $5.0\,\text{GeV}/c$, a muon will traverse all five muon stations [58]. For each candidate in the muon system, likelihoods for the muon and non-muon hypotheses are computed, as a function of the average squared distance between the extrapolated track points and the hits.

Incorrect identification can occur when there is a non-muon track with spurious hits in the muon stations, or if a real muon track in the muon station is associated with a non-muon track in the preceding sub-detectors, and is assigned to the non-muon track.

ProbNN variables

An additional set of variables, denoted as ProbNNx, are defined to discriminate particle species. These variables are multivariate classifiers, in this case neural networks [73], which combine information from different subdetectors and their correlations efficiently, to differentiate between the particle types. The DLLx variables are a key input to the training of these classifiers, along with tracking variables, kinematic information, and PID variables from the muon and calorimeter systems [58]. Each ProbNNx is optimised to discriminate the particle x from all other charged particles, where simulated data is used to train them. This technique takes into account correlations between the detector systems, as well as additional information, resulting in a set of variables that have better discriminating power than the DLLx variables alone. The neural net, ProbNNx, will return a value close to unity if a track is likely to be of particle species x, and a value close to zero otherwise.

Until 2018, the ProbNNd for discriminating between deuterons and other charged tracks was not available. After the inclusion of the deuteron hypothesis in the RICH algorithm in the summer of 2016, it was possible to also make a powerful discriminator

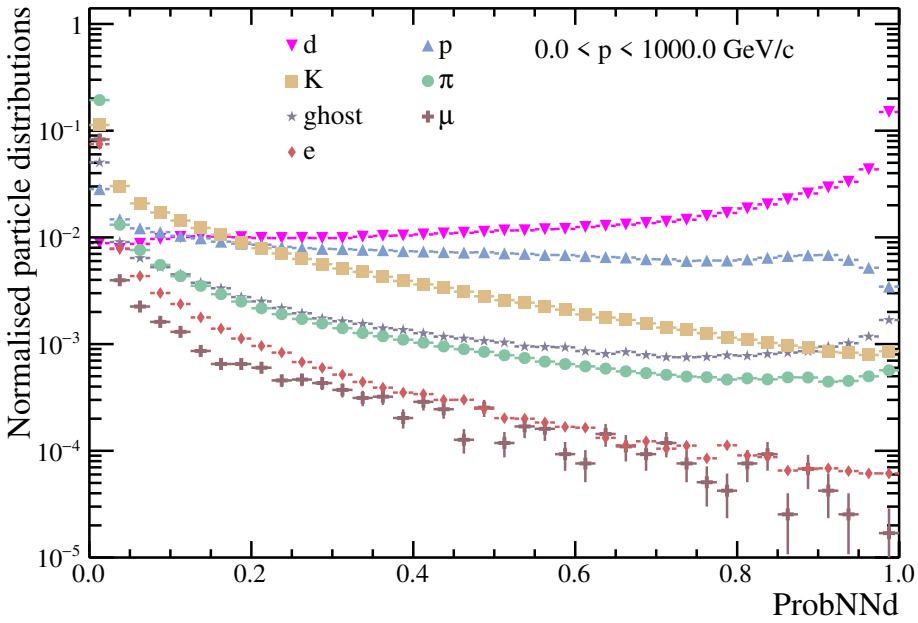


Figure 4.9: ProbNNd for different charged particles from simulation.

for deuterons. For the deuteron neural net, ProbNNd, simulated samples from a b hadron decay, $\Lambda_b^0 \to d\bar{p}$, were used for training, as these were the samples available at the time.

Distributions of ProbNNd for charged particles from simulation are shown in Fig. 4.9. It can be seen that, with the exception of deuterons, all particles yield ProbNNd values close to zero. Therefore, ProbNNd can be used to select deuterons, by choosing tracks with high ProbNNd values.

4.2.4 Data taking and triggers

If pp collisions in the LHCb detector were at the maximum luminosity that the LHC is capable of producing, the occupancy would be too high for tracking and reconstruction to occur successfully [57]. Therefore, the luminosity is reduced to a constant rate, such that on average there is a single interaction per bunch crossing, giving a rate of visible events in the LHCb detector of 40 MHz [74].

This rate is still too high for the LHCb detector to be able to process and record all of the data from every collision, so triggers are used in order to save only the interesting events. In particular, only a fraction of events result in a b hadron within the acceptance,

so only these need to be studied in order for the LHCb experiment to achieve its primary physics goals [57].

Three levels of triggers are used in order to reduce the rate from the events that are collected to one which can be fully processed [74]. The data-taking rate is limited by the bandwidth and frequency of the front-end electronics, at about 1 MHz. Field-programmable gate arrays in the L0 trigger use information from the electromagnetic and hadronic calorimeters, and the muon stations, to determine which events to process. The first stage of the High Level trigger (HLT1) then selects events based on track combinations and muons in the events through partial event reconstruction, reducing the rate to 110 kHz. Events passing the selection are then buffered to disk storage, where they are fully reconstructed in the second stage (HLT2), and a wider range of final state triggers can be applied, with a final rate of 12.5 kHz.

Many measurements at the LHCb experiment use specific triggers in order to make exclusive measurements of decays, however for the measurement of deuterons at the LHCb experiment, no such trigger will be used. Concurrently to the B-physics-driven triggers, a 'nobias' trigger runs throughout data-taking, which samples randomly from all events, at a chosen rate [74]. This means that a dataset representing all pp collisions is saved, which can be used for measurements beyond B-physics.

4.2.5 Simulation

Simulation of pp collisions, the subsequent particle decays and interactions, and the detector response are used widely in physics analyses. Simulation provides 'truth' information about particles and tracks, which can aid the understanding of detector effects seen in data.

At the LHCb experiment, Monte Carlo (MC) samples are produced centrally by the collaboration using the Gauss framework [75]. This ensures that the same models are used across the board, and that the correct detector conditions are simulated, to enable fair comparisons of results.

Pythia 8 [46] is used to simulate the proton-proton collisions, and EvtGen then models the decays [48], before Geant4 is used to simulate the particle-detector interactions [76].

Due to the very intensive demand on computing resources of MC simulations, the simulation samples can only be produced with limited statistics. To validate the performance of the simulation, calibration samples from data can be used.

4.2.6 Particle identification calibration

There is some disagreement between simulation and data in the distribution of PID variables. This is due to mismodelling in the underlying event simulation, which impacts the PID response. To overcome this, simulation can be calibrated using carefully selected data samples, to correct for the differences. The PIDCalib package has been written to do this, such that the same tools and calibration samples are be used for all physics analyses at the LHCb experiment [77]. Calibration samples are chosen in data using a tag-and-probe method to select tracks of known species. Samples are available for protons, kaons, pions, muons and electrons. The tools in the package are used primarily for computing the correct efficiencies for selecting tracks after applying PID requirements, but the samples are also accessible centrally, to be used with more flexibility.

Decays are selected in data where the parent particle can be fully reconstructed, and the daughters may be identified without use of PID information. One such decay is $D^{*+} \to D^0\pi^+$, with the subsequent decay $D^0 \to K^-\pi^+$ [78]. The vertices of these two decays are found, with the correct invariant mass reconstructed at each. The pions from the two decays are known to have the same sign charge, whilst the kaon will have the opposite charge, such that the particle species can be inferred without use of PID selections.

Once a significant number of tracks have been selected in the sample, the *sPlot* technique [79] is used to statistically subtract the background that comes from the D^{*+}. These samples of kaon and pion data are then weighted such that their kinematics match the kinematics of kaons and pions in the decay of interest. The kinematic weighting is a crucial step because the PID response of the detector is highly dependent on the momentum and pseudorapitity of tracks. In some cases the calibration sample tracks are also weighted in detector occupancy, as this can also have an effect on the PID performance.

Calibration protons can be found from the $\Lambda \to p\pi^-$ decay. Muons and electrons are both chosen from J/ψ decays: $B^\pm \to J/\psi\,(e^+e^-)K^\pm$ and $J/\psi \to \mu^+\mu^-$. For these, one of

the child candidates is required to pass strict PID requirements, whilst the other has no PID selection applied, allowing for a PID 'unbiased' sample.

Due to the fact that the calibration tracks are selected from common decays, the PIDCalib data samples have the advantage that they have high statistics, such that they can be used to high precision for stringent PID requirements.

Chapter 5

Measurement of the deuteron production cross-section

5.1 Strategy

A measurement of the deuteron production cross-section in pp collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV will be presented in this chapter. The number of deuterons produced in the collisions will be extracted from a binned fit to data, using the discriminating variable ProbNNd. From the simulation of deuteron production in the LHCb experiment conditions, described in section 3.3, it is expected that a small deuteron signal will be present amongst a large background of other charged particles.

The data used for the measurement was recorded by the LHCb detector through a minimum bias trigger during 2018. Minimum bias data is taken with minimum detector actvity requirements, unlike other common triggers, which require, for example, two high energy muons to be present in order to save an event. Conversely, the minimum bias trigger requirement at LHCb is that events should include one recorded track within the VELO [65]. The measurement is performed in bins of deuteron momentum, corresponding to regions where the particle identification response is consistent across the bin, so that no large variation in particle identification response is present.

For the deuteron measurement, the background comprises pions, kaons, protons, muons, electrons, and ghost tracks. The probability density functions for deuterons and ghost tracks used for the fit are obtained from minimum bias simulation, whilst the shapes for other charged particles are taken from calibration data samples. Minimum bias data and simulation are dominated by prompt tracks, *i.e.* produced directly from the pp

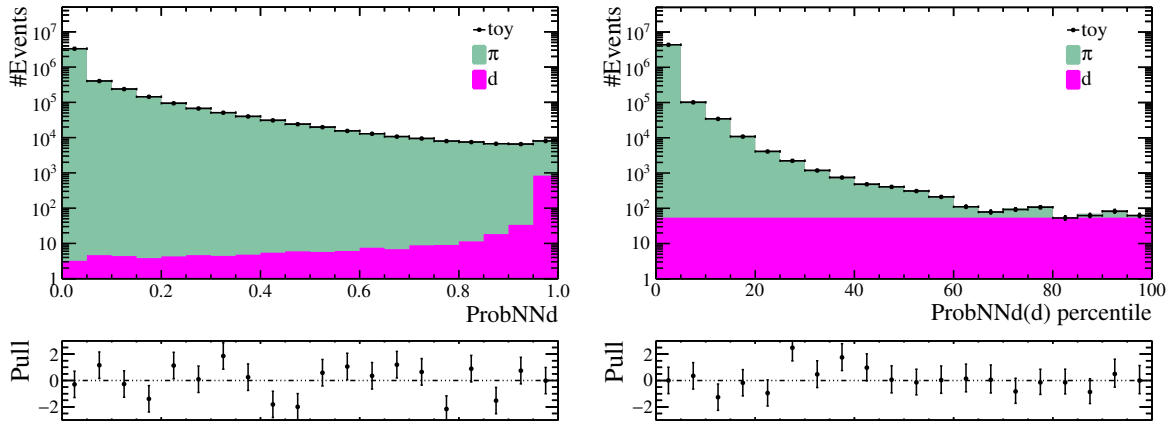


Figure 5.1: Distributions of deuterons and pions as functions of ProbNNd to illustrate the choice of binning in ProbNN$d(d)$ percentile: even width binning in ProbNNd (left) and ProbNN$d(d)$ percentile binning (right). The pion and deuteron shapes are fit to pseudo-data, made of pions and deuterons from simulation, with relative yields taken from simulation, for the momentum region $29.7 < p < 35.4\,\text{GeV}/c$.

collision point, rather than tracks from subsequent decays of promptly produced particles. In other words, particles from decays have greater impact parameters than particles produced promptly, as found in minimum bias conditions. Since the ProbNNd variables depends on kinematic quantities, such as p and η, these differences in particle origin need to be taken into account. A kinematic reweighting is applied to the calibration samples before the shapes for the different particle species are extracted.

A non-uniform binning in ProbNNd is chosen to increase the sensitivity of the fit to the deuteron signal. Instead of constant width in ProbNNd, the binning is chosen in each momentum region to make the deuteron shape flat. To achieve this configuration, the bin edges are chosen to match the i^{th} percentiles of the deuteron ProbNNd probability density function, $\mathcal{P}_i^{d\%}$, for $i = \{10, 20, \ldots, 100\}$. Furthermore, to simplify the minimisation algorithm and the visualisation of the results, $\mathcal{P}_i^{d\%}$ is used instead of ProbNNd in the fit, as it allows recovery of a binning with constant width. The shift from even-width bins to the ProbNN$d(d)$ percentile binning is illustrated in Fig. 5.1. In this simplified example, with shapes and relative yields taken from simulation, pions and deuterons are plotted on the left in even-width bins in ProbNNd, and on the right in ProbNN$d(d)$ percentile binning. On the left, the peaking structure of the deuterons is not visible in the total shape, whereas on the right, the pions are concentrated in the low ProbNN$d(d)$ percentile region, and at the highest ProbNN$d(d)$ percentile values, deuterons are the dominant component.

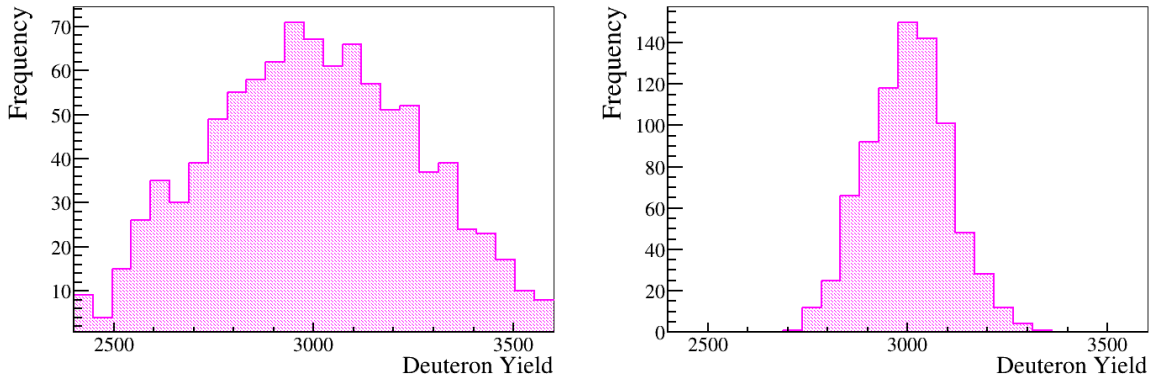


Figure 5.2: The deuteron yields returned by pseudo-experiments in the two binning schemes: even width binning in ProbNNd (left) and ProbNN$d(d)$ percentile binning (right), for the momentum region $29.7 < p < 35.4\,\text{GeV}/c$.

This binning choice improves the sensitivity of the fit to the deuteron signal, as indicated by performing simple fits to pseudo-data in the two binning schemes. The plots in Fig. 5.1 are two examples of these. The numbers of deuterons found by a thousand different fits are shown in Fig. 5.2, for the even-width binning in ProbNNd on the left, and in ProbNN$d(d)$ percentile on the right. The true number of deuterons in the pseudo-data was 3000, and the fits returned yields of $N_d^{\text{ProbNN}d} = 2999.5 \pm 253.6$ and $N_d^{\text{ProbNN}d(d)\text{ percentile}} = 3001.3 \pm 102.8$. The uncertainty on the deuteron yield is thus reduced by a factor 2.5 using the ProbNN$d(d)$ percentile binning.

The shapes of each particle species as functions of ProbNN$d(d)$ percentile from minimum bias simulation are shown in Fig. 5.3, with the distribution for each component normalised to unity. Arguments for the choice of the momentum regions for each of the plots are presented in section 5.3. Where possible, data is used to find the shapes that enter into the fit, but these plots from simulation give some indication for what their form will be. Following the definition of the binning, the deuteron shapes are flat in ProbNN$d(d)$ percentile. The shapes for all other particle species decrease with increasing percentile. The muon, electron and pion shapes are in agreement with each other in all momentum regions, though the muon and electron shapes are statistically limited, with large uncertainties on each point as a result. The shapes for each particle used in the fit to data as a function of ProbNN$d(d)$ percentile will be discussed in detail in section 5.5.

Using ProbNN$d(d)$ percentile as a discriminating variable offers the best sensitivity to a deuteron signal at the LHCb experiment, and providing that a significant yield is found in each momentum bin, a deuteron production cross-section can be computed.

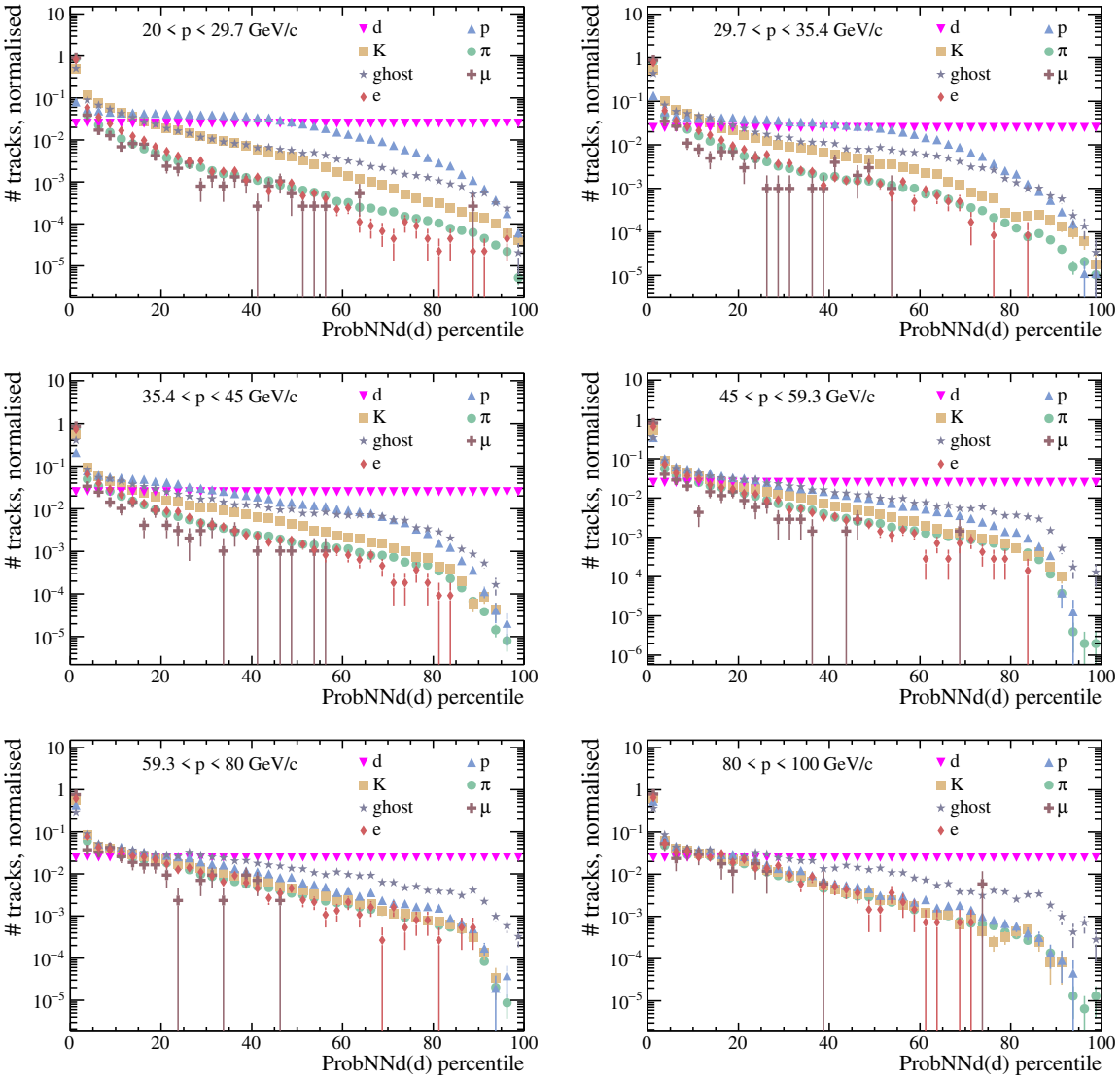


Figure 5.3: Shapes of charged particles in minimum bias simulation in ProbNN$d(d)$ percentile in different momentum regions, with the distribution for each particle species normalised to unity.

The measurement of the pp inelastic cross-section, $\sigma_{pp\to X}$, presented in Ref. [65] is used as a basis to determine the deuteron production cross-section, σ_d, as

$$\sigma_d = \sigma_{pp\to X}\frac{N_d}{N_{evt}}, \tag{5.1}$$

where N_d is the number of deuterons measured, and N_{evt} is the number of inelastic pp interactions. To compute these quantities from those measurable in data, the detection

and reconstruction efficiencies need to be taken into account, as will be discussed in section 5.8.

The full details of the steps towards the deuteron measurement are given in the rest of this chapter. First, the data and simulation samples used are summarised in section 5.2; a justification for the choice of momentum bins is given in section 5.3; the fit method using the HistFactory package is detailed in section 5.4; then the shapes and their uncertainties in ProbNN$d(d)$ percentile for each of the components entering the fit are presented in section 5.5; the fit to ProbNN$d(d)$ percentile and pseudo-experiments are described in section 5.6; the fit performed on minimum bias LHCb data in section 5.7; concluding with the method to compute the deuteron production cross-section in section 5.8;

5.2 Data and simulation samples

The data used for the measurement is the 'minimum bias' sample recorded by the LHCb detector during 2018. Events in this sample are selected by a (prescaled) random trigger, so that they constitute an unbiased representation of all pp collisions occuring in the LHCb detector throughout the data-taking period. During 2018, the centre-of-mass energy of the collisions was $\sqrt{s} = 13\,\text{TeV}$. The data, calibration samples, and simulation details are listed in Tab. 5.1.

The reason for this choice of data is driven by the calibration and simulation samples that were available during the time of the analysis being performed. The PID calibration samples are processed centrally by the PID working group in the LHCb collaboration, to ensure that all analyses use the correct and same calibration. This task usually requires long processing time, owing to the large size of the calibration samples. At the time of this analysis, the only samples that included the ProbNNd information, and were therefore of use for the deuteron measurement, were the calibration samples for the 2018 data-taking period.

Sample type	Sample size	
Minimum bias simulation	pp collisions, $\sqrt{s} = 13\,\text{TeV}$, $1.89\,\text{nb}^{-1}$	
Deuteron simulation	2 M events	
Minimum bias MC	2 M events	
PID Calibration	Pion sample:	54 M tracks
	Kaon sample:	75 M tracks
	Proton sample:	223 M tracks
	Muon sample:	30 M tracks
	Electron sample:	1.1 M tracks

Table 5.1: Data and simulation samples used in deuteron measurement, all from 2018; number of tracks in each of the PID calibration data samples.

5.3 Choice of momentum bins

The deuteron production cross-section measurement is performed in different regions of momentum, so that it can be measured as a function of momentum. The bins are chosen to optimise the sensitivity to a deuteron signal.

Charged particle identification relies heavily on the two RICH detectors, which provide a highly momentum-dependent output. This dependence is evident in Fig. 5.4, where the average ProbNNd values, $\langle\text{ProbNN}d\rangle$, for different particle species in a minimum bias simulated sample are shown as a function of momentum. The peaks in these profile histograms correspond to the Cherenkov thresholds in the RICH 1 and RICH 2 detectors for different particle types. The effects are most prominent in the proton and kaon distributions. Protons, kaons and deuterons at low momenta ($\lesssim 10\,\text{GeV}/c$) are in agreement with each other. At the kaon threshold in RICH 1 ($9.3\,\text{GeV}/c$), they start diverging as $\langle\text{ProbNN}d\rangle$ for kaons decreases rapidly, whilst the values for protons and deuterons continue to rise. This is due to the fact that kaons of this momentum passing through RICH 1 will produce Cherenkov photons, but heavier particles will not, such that the RICH algorithm is able to identify kaons, but cannot differentiate between heavier species. A similar pattern is seen at the proton threshold in RICH 1 ($17.7\,\text{GeV}/c$).

Secondary threshold effects can also be seen, for example, in the kaon distribution at the proton threshold in RICH 1, and in the proton distribution at the kaon threshold in RICH 2 ($15.6\,\text{GeV}/c$). These occur because the RICH algorithm starts to be able to discriminate between particles at these thresholds, so that it is able to discriminate

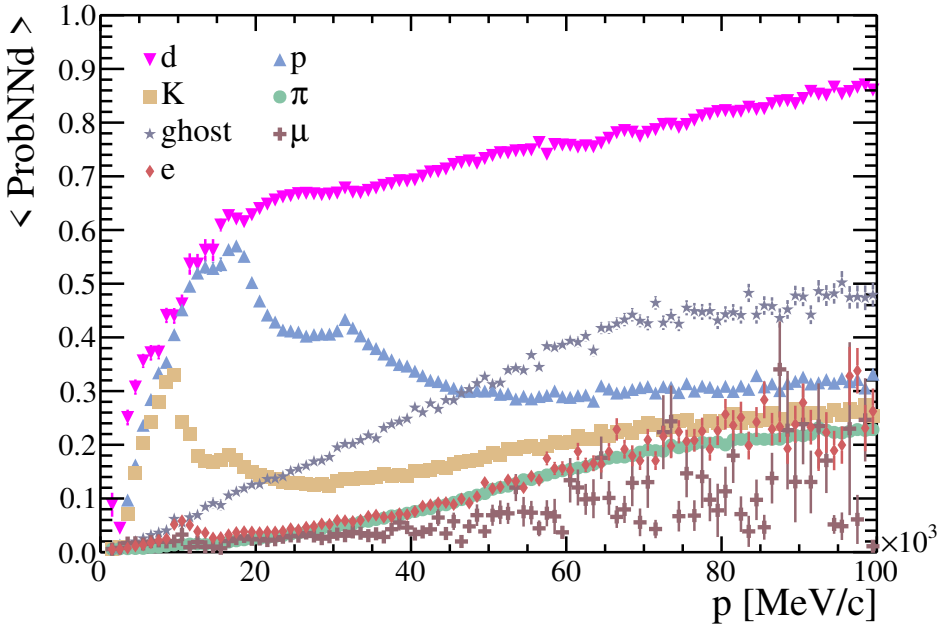


Figure 5.4: Profile histogram of ProbNNd values in bins of momentum for charged particles in minimum bias simulation. A profile shows the mean value of ProbNNd, $\langle$ProbNN$d\rangle$, for each bin of momentum, and the error bars indicate the RMS in that bin.

between particles based on 'below threshold' null signals. At higher momenta, the same effects are seen at the deuteron thresholds of 35.4 and 59.3 GeV/c, for RICH 1 and RICH 2, respectively.

For momenta greater than 60.0 GeV/c, the distributions for all real, charged particles, except deuterons, are in agreement with each other. Significantly above the corresponding RICH momentum threshold, the angle at which different types of particles radiate Cherenkov photons saturates, *i.e.* the Cherenkov angle becomes independent of the particle mass (see Fig. 4.7). Above 60.0 GeV/c, all charged particles have reached their saturation point, with the exception of deuterons and ghost tracks. Ghosts are reconstructed from combinations of unrelated track fragments, and they are not associated with clean signals in the RICH detectors, and therefore they are not interpreted well by the RICH algorithm. Consequently, the ghost $\langle$ProbNN$d\rangle$ values are higher than the values for real particles, but the trend in increasing $\langle$ProbNN$d\rangle$ for ghosts decreases at momenta greater than 60.0 GeV/c, such that the values level off. The deuteron signal does not saturate until momenta greater than 100.0 GeV/c, and as such their $\langle$ProbNN$d\rangle$ are much higher than those for other particles so it is possible for them to be separated from the other particles on this basis.

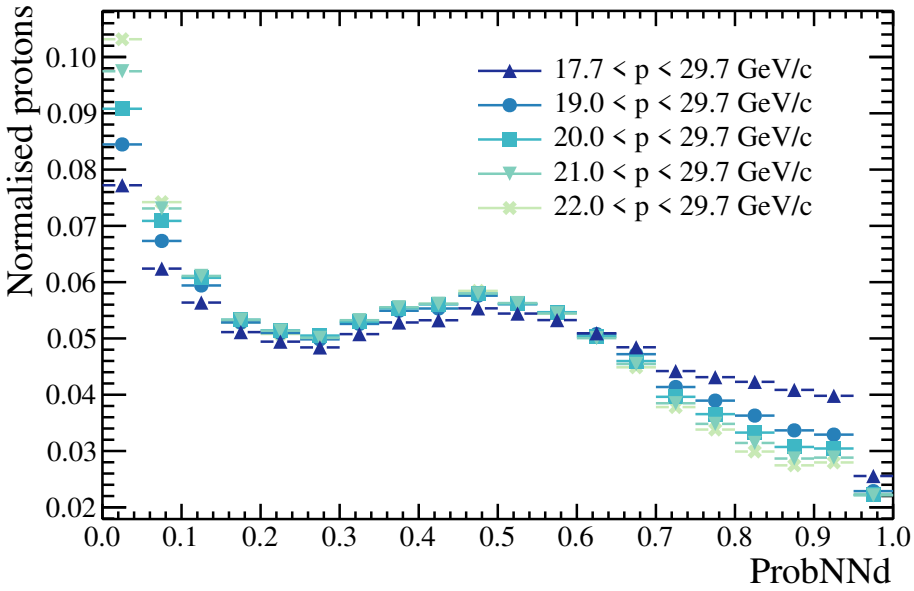


Figure 5.5: ProbNNd shape for protons from minimum bias simulation for different ranges of momentum, above the proton threshold in RICH 1 (17.7 GeV/c), and below the deuteron threshold in RICH 1 (29.7 GeV/c).

Based on what is presented in Fig. 5.4, it is clear that only tracks with momenta above the proton threshold in RICH 1 should be considered for the deuteron measurement, because there is no separation between protons and deuterons at momenta below this.

Very close to the proton threshold in RICH 1, the number of Cherenkov photons produced by a proton track in RICH 1 is very low, according to Eq. 4.2, so that the signal is weak. Therefore, the RICH algorithm has difficulty separating protons and deuterons here. As such, when choosing the lowest momentum boundary for the measurement, instead of using exactly the threshold, a slightly higher value was chosen, where a stronger Cherenkov signal will be produced by protons. The choice should be made to optimise the proton–deuteron separation. To choose the lower bound, the ProbNNd disributions for protons with momenta close to the threshold were compared. Figure 5.5 shows these distributions for protons with $p_{min} < p < 29.7\,\text{GeV}/c$, for various values of p_{min}, as extracted from a minimum bias simulation sample. For a lower bound of 17.7 GeV/c, the shape does not decrease steadily towards ProbNNd values of 1.0. If the lower bound is increased to 20.0 GeV/c, the shape decreases more steeply. Little would be gained if it were further increased, as the shapes with lower bounds of 21.0 GeV/c and 22.0 GeV/c agree with the shape for 20.0 GeV/c at high ProbNNd values. A lower bound of 20.0 GeV/c is chosen, as increasing the bound would also lead to a loss of statistics.

Momentum bin edge (GeV/c)	Attribute
20.0	Above proton RICH 1 threshold
29.7	Deuteron RICH 1 threshold
35.4	Proton RICH 2 threshold
45.0	Interim bin edge
59.3	Deuteron RICH 2 threshold
80.0	Interim bin edge
100.0	Upper momentum limit

Table 5.2: Momentum bin boundaries used for the measurement of deuterons at the LHCb experiment and the justification of the choice of each one.

The signal significance can be characterised as $S/\sqrt{S+B}$, where S is the number of signal events, and B is the number of background events. To find the expected significance of the deuteron signal, S is the expected number of deuterons, and B is the expected number of other charged particles. Using expected rates of particles from minimum bias simulation, with the total number of tracks $(S+B)$ equal to the number of tracks in data, the expected signal significance can be calculated for the lower momentum bin boundaries, 20.0, 21.0, and 22.0 GeV/c, with upper bound 29.7 GeV/c. For a lower bound of 20.0 GeV/c, the significance is found to be 0.70, for 21.0 GeV/c it is 0.65, and for 22.0 GeV/c it is 0.60. Therefore, increasing the lower bound from 20.0 GeV/c to 21.0 GeV/c, would result in a 7% reduction in signal significance.

The remaining momentum bins are chosen such that they align with higher RICH detector momentum thresholds, to a maximum of 100.0 GeV/c, which is the point beyond which charged particles are no longer centrally calibrated by the LHCb collaboration. No complications due to threshold effects, such as that for protons illustrated in Fig. 5.5, were seen at the other RICH thresholds, so the momentum bin boundaries sit at the exact thresholds. The bin boundaries, along with the corresponding thresholds, are summarised in Tab. 5.2. The 'interim' bins are chosen such that the momentum bins are of similar widths, in order to produce a measurement of the deuteron production cross-section as a function of momentum.

Shapes in ProbNNd for different charged particles from minimum bias simulation are plotted in Fig. 5.6, in each momentum bin. In every momentum bin, the deuterons peak strongly at the high ProbNNd values, whereas the trends of the other particle types change through the bins. The pion, muon and electron shapes are largely in agreement

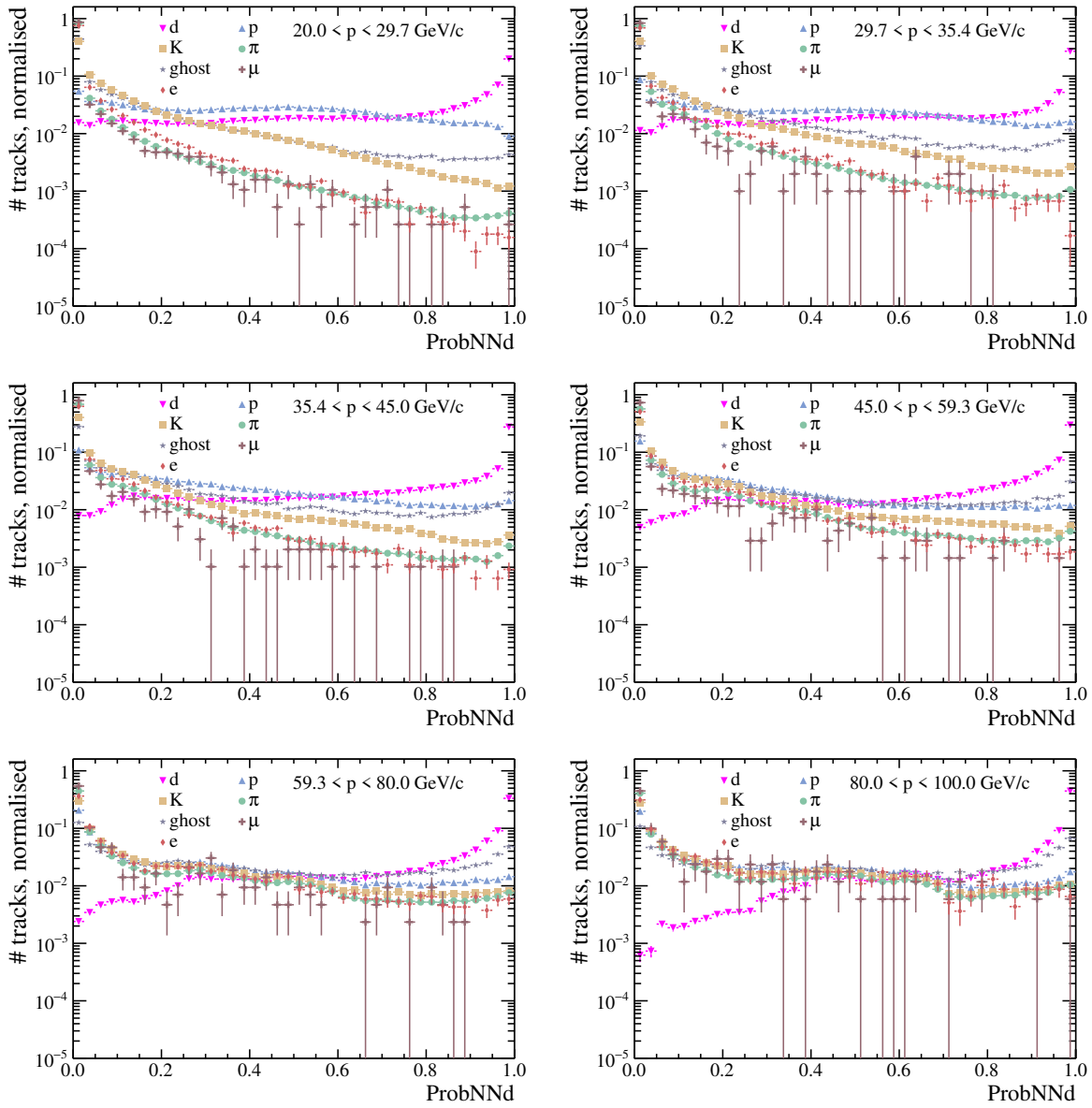


Figure 5.6: Shapes in ProbNNd of charged particles in minimum bias simulation for each bin of momentum used in the analysis. Histogram for each component is normalised to unity.

with each other in every momentum bin, due to the fact that they are well above their RICH momentum thresholds. The muons and electrons have large errors for most values of ProbNNd, because their disributions are concentrated in the bins at the low ProbNNd values, leaving few entries for the upper bins. The same effect is not seen for pions, because pion statistics are higher than the muon and electron statistics.

The kaon and pion shapes agree with each other increasingly as momentum increases; they disagree most in the first momentum bin, $20.0 < p < 29.7\,\text{GeV}/c$. This is likely due to the fact that kaons are not far from their RICH 2 momentum threshold ($15.6\,\text{GeV}/c$) in this bin, meaning that their Cherenkov angle in RICH 2 will not have saturated, whilst the pions are saturated in RICH 1 and RICH 2 at momenta $p > 20.0\,\text{GeV}/c$. Therefore, kaons and pions will not appear equal in the RICH algorithm, hence they have different distributions in ProbNN variables too.

The proton shape in the first momentum bin, $20.0 < p < 29.7\,\text{GeV}/c$, is seen to decrease rapidly at high ProbNNd values. In this bin, the proton is above its momentum threshold in RICH 1, but deuterons are below threshold, such that protons will cause a signal in RICH 1, but deuterons will not. This difference provides strong disrimination between the two. In higher momentum bins, the proton shape does not turn downward as strongly as in the first momentum bin.

The ghost shapes in every momentum bin are seen to turn upward at the highest ProbNNd values, where the shape gets closer to the deuteron shape as momentum increases. There is better distinction between deuterons and ghosts in the first two momentum bins, $20.0 < p < 29.7\,\text{GeV}/c$ and $29.7 < p < 35.4\,\text{GeV}/c$, where deuterons are below threshold in both RICH detectors, so will not produce a signal in either. If ghosts are formed from segments of unrelated tracks, wrongly reconstructed as single tracks, and these segments originate from pions, or other light particles, then it is likely that the ghosts are associated with signals in the RICH detectors. Therefore, ghosts and deuterons can be distinguised using the lack of signal from deuterons in this low momentum region. At higher momenta, where the deuterons produce Cherenkov photons, the same cannot be said, and as such, the separation is reduced.

5.3.1 Low momentum deuterons

Given that the deuteron measurement presented here is performed only for tracks with momentum $p > 20.0\,\text{GeV}/c$, it is worth noting that this analysis is not sensitive to the many deuterons that are produced with lower momenta. The deuteron yield falls off exponentially as a function of momentum, so the bulk of deuterons exist at low momentum, in the region where the ProbNNd variable offers no discrimination, as seen in Fig. 5.4.

In the current versions of ProbNN variables, the main particle identification inputs are from the RICH detectors and the calorimeters. However, there is scope to add further variables to the training of the discriminators, which could improve the identification of deuterons at low momenta. An in-depth study of these prospects is beyond the reach of this thesis, but a brief overview of the variables that were studied is given here, with new insights into the prospective performance for deuteron identification, as a result of work done during this PhD.

The VELO and the Outer Tracker (OT) each have some discriminating power for low momentum tracks based on the time of flight of the particles. The performance of the OT was improved in Run 2 of the LHC by making the time and space alignment of the subdetector more precise, allowing for the time measured in the OT to provide some PID discrimination; a brief description of this process is presented here, but full details of the OT are given in Ref. [61]. The use of VELO measurements of energy loss, $\mathrm{d}E/\mathrm{d}x$, for PID was also studied at the start of Run 1 of the LHC. These studies were however not taken any further because it was found that the VELO information is only effective for very low momentum tracks, which get swept out of the LHCb detector acceptance by the magnet [80]. The PID information offered by the VELO will be revisited here, with the inclusion of deuterons.

OT track time

The arrival time of signals in the OT is measured with respect to the LHC clock [61], and represents the time between a pp collision and a signal being processed in the electronics in the OT. The arrival time is therefore the sum of the time taken for a track to travel from the primary vertex to the OT, t_{tof}, the drift time in the OT staw tube, t_{drift}, the propagation time of the signal along the wire at the centre of the straw tube, t_{prop}, and the lag in the front end electronics, t_{FE}. In addition, differences in the alignment of the LHCb experiment and LHC clocks must be considered. The total measured time between a bunch crossing and a signal reaching the front end electronics in the OT, t_{TDC}, is given by

$$t_{TDC} = (T_{\mathrm{collision}} - T_{\mathrm{clk}}^{\mathrm{FE}}) + t_{\mathrm{tof}} + t_{\mathrm{drift}} + t_{\mathrm{prop}} + t_{\mathrm{FE}}, \tag{5.2}$$

where $T_{\mathrm{collision}}$ is the time of the pp collision according to the LHC clock, $T_{\mathrm{clk}}^{\mathrm{FE}}$ is the phase of the clock used by the OT. The sum $(T_{\mathrm{collision}} - T_{\mathrm{clk}}^{\mathrm{FE}}) + t_{\mathrm{FE}}$, can be rewitten as $(T_{\mathrm{collision}} - T_{\mathrm{clk}}) + t_0$, where t_0 is the offset between the global LHCb clock and the OT

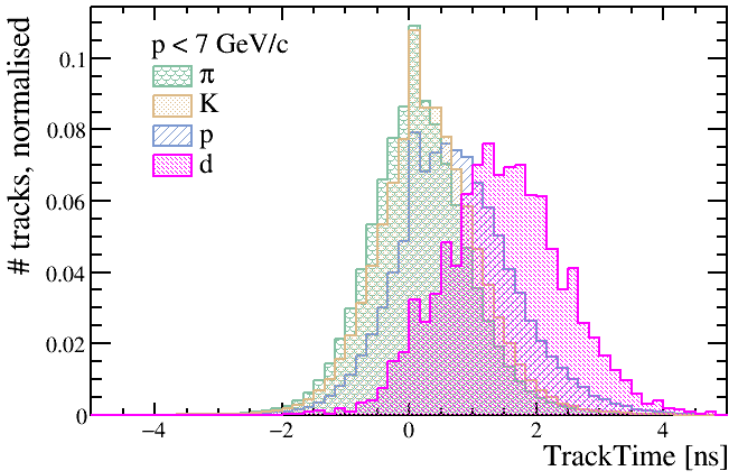


Figure 5.7: OT track time distributions normalised to unity for π, K, p and d with $p < 7.0\,\text{GeV}/c$ from minimum bias simulation.

front end electronics clock, and T_{clk}, is the time according to the LHCb clock. During Run 2 of the LHC, the calibration of this offset was improved by running it in real-time.

The expected drift time, $t(r)$, where r is the distance between a track and the wire at the centre of the closest straw tube, can be calculated using the 'distance to drift time relation', which includes the effects of t_{tof}, t_{drift} and t_{prop}. This is the expected time for the track itself, without the effects caused by the misalignment of the LHCb and LHC clocks, or the lag in the front end electronics. The measured time, t_{TDC}, can be compared to $t(r)$, where the difference between the two can be fitted with a Gaussian, with mean t_0, known as the drift time residual. The residual has been used primarily to monitor the alignment of the OT, LHCb and LHC clocks, but also provides some PID discrimination at low momenta where the difference in flight time for a particle at a given momentum as a function of the particle mass is distinguishable by the OT resolution.

A track that traverses the whole OT will typically leave 22 hits in the OT stations. For each individual hit the residual, t_0, can be calculated. The sum of all the individual residuals, weighted by their errors, is known as the 'track time'. This is effectively the difference in the measured arrival time of a particle at the OT, compared to the expected arrival time for a particle travelling at the speed of light. Low momentum, heavy particles have a sufficiently low velocity that their measured arrival time is significantly later than their expected arrival time.

For pions, kaons, protons and deuterons with momentum $p < 7.0\,\text{GeV}/c$ from minimum bias simulation, the track time distributions are plotted in Fig. 5.7, each normalised to unity. There is a clear shift towards higher values of track time in the deuteron distribution, compared to the distributions for lighter particles, *i.e.* the difference in the measured arrival time for deuterons in the OT compared to the expected arrival time is greater than for lighter particles. The mean track time for deuterons is $t_d = 1.48 \pm 0.94\,\text{ns}$, compared to $t_\pi = 0.15 \pm 0.80\,\text{ns}$ for pions.

VELO energy loss

At the LHCb experiment, it is possible to employ the VELO for PID of low momentum tracks, using measurements of the energy deposited in the silicon pixels of the VELO modules [80] [1]. Each of the 21 silicon modules in the VELO has 2048 silicon strip sensors, in which an electron signal is produced by a passing track. The gain of the signal in each strip sensor is the ratio of the output of the strip to the deposited input charge from the passing track. The gain varies from sensor-to-sensor, due to differences in strip capacitance and electronics performance. This variation leads to significant and asymmetric fluctuations in the measured energy loss of particles, and contributes to the Landau distribution that is found for the $\text{d}E/\text{d}x$ of particles. The units of the gain in the VELO are Analogue-to-digital Converter (ADC) counts, where one ADC unit corresponds to a signal of 380 electrons.

In order to find the Most Probable Value (MPV) of energy loss, a truncated mean is found, in place of the mean of the full distribution. A truncated mean is simply the mean of the distribution, with a fraction of the entries removed from the higher end of the distribution. So, the 40% truncated mean removes the top 40% of entries, where the percentage removed can be varied to give other truncated means. This method removes the highest outliers in the Landau distribution, and gives a mean value that is much closer to the peak of the distribution than if the full set of entries had been used.

Figure 5.8 shows the MPV found from the 40% truncated mean ADC count vs. momentum for long deuteron, proton, kaon and pion tracks in minimum bias simulation with $p < 5.0\,\text{GeV}/c$. There is a depletion of entries for $p < 1.3\,\text{GeV}/c$, because tracks with those momenta are swept out of the LHCb detector acceptance by the magnet. It can be

[1] The study in Ref. [80] concluded that this was not a viable PID method, because the performance was significantly better in simulated data, where MC truth information was used, instead of momentum measurements with limited resolution in data.

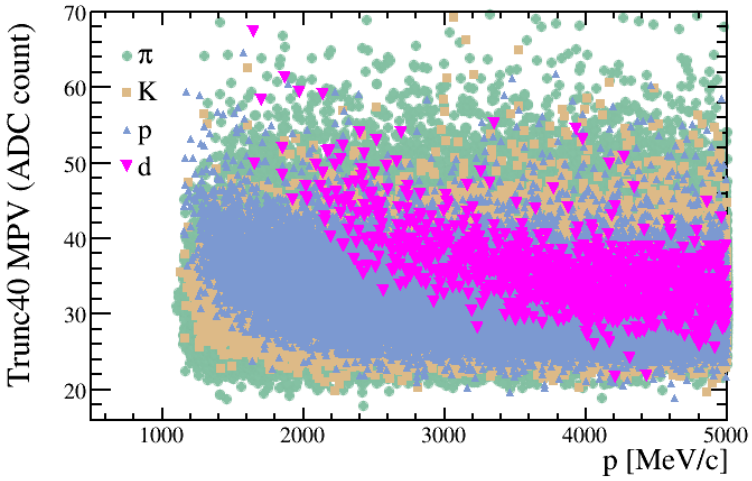


Figure 5.8: VELO time vs. momentum for π, K, p and d with $p < 5.0\,\text{GeV}/c$, relative yields of particles taken from minimum bias simulation.

distinguished that pions and kaons both form wide bands in MPV for all momenta, whilst the proton and deuteon distrbutions show a curving shape towards higher MPV values for lower momenta, which can be used to discriminate between protons and deuterons, and lighter particles. The large width of the bands in truncated mean for each particle, which results in limited particle identification, is due to the large variation in the ADC count for the particles. This is due to the variation in the capacitance of the silicon strips which make up the VELO detector, namely because of their differing size, state of depletion, and variation in the lengths of electronic links to the circuit boards for data collection [81].

It was found that deuterons can be best distinguished from other tracks in a narrow momentum range of $1.8 < p < 2.8\,\text{GeV}/c$, which corresponds to where there is the most curvature for deuterons in Fig. 5.8. The MPV of deuteron, proton, pion and kaon tracks from minimum bias simulation are shown in Fig. 5.9, where the distribution for each particle species is normalised to unity. It is clear here that the values for deuterons are shifted to higher MPV values than the other, lighter particles.

Though the OT track time and VELO $\mathrm{d}E/\mathrm{d}x$ variables provide some separation between deuterons and other charged particles at low momenta, the study has not been extended any further for the purposes of this thesis, because, due to the low deuteron expected yield compared to other particles, these variables will not provide enough discrimination to measure a significant deuteron signal. In the future, if these variables were included in neural networks, such as ProbNN, it may be found that it is possible to

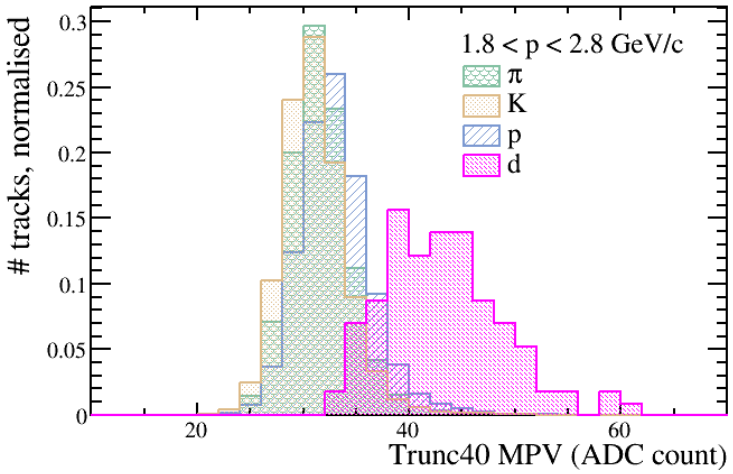


Figure 5.9: VELO time distributions normalised to unity for π, K, p and d (bottom).

improve the discrimination and perform a measurement of deuterons at low momentum at the LHCb experiment. It may also be worthwhile to add other variables, such as measures of the spread of the $\mathrm{d}E/\mathrm{d}x$ signals, to the neural network to improve the performance [82].

5.4 Fit to extract the deuteron yield

To extract the deuteron yield in each momentum bin, each corresponding data histogram is fitted using a set of binned models, known as *templates*. Histograms with forty bins in ProbNN$d(d)$ percentile are made from data and simulation for each momentum bin. The templates used to model deuterons and ghosts are taken from simulation; for other charged particles they are extracted from calibration data samples. With the deuteron yield as the parameter of interest, the model templates are combined to fit the data using the HistFactory package [83]. The use of binned models and data rather than continuous functions results in uncertainties due to the statistics of the models, in addition to the data. These extra uncertainties can be handled using the Beeston-Barlow method [84]. The shapes of the models entering the fit also have associated systematic uncertainties that alter their form. These systematic uncertainties are also taken into account in the fit.

The method used by the HistFactory package for fitting histograms in a template fit is outlined in section 5.4.1; followed by explanations of the Beeston-Barlow method in section 5.4.2, and the systematic shape variation that is included in the fit in section 5.4.3.

5.4.1 HistFactory package

The template fit used to find the deuteron yield is performed using the HistFactory [83] package within the RooFit framework [85]. ROOT histograms are taken as fixed template histograms to build parametrised probability density functions [86].

To understand how HistFactory works, it is sufficient to consider a simple case with one signal channel and one background for the discriminating variable x [83]. In the case of the deuteron measurement, the deuterons form the signal, all other charged particles form the background, and the discriminating variable is ProbNN$d(d)$ percentile. The number of signal events is denoted as S, and the background as B, where each has the continuous shape in x of $f_S(x)$ and $f_B(x)$. The signal strength, μ, is the parameter of interest, where $\mu = 0$ is true for a background-only hypothesis, and $\mu = 1$ is for the nominal signal plus background hypothesis.

Considering a dataset containing n events, with values of x for each event denoted x_e, a likelihood function can be written as

$$\mathcal{L}(\mu) = \mathcal{P}\left(\{x_1 \dots x_n\}|\mu\right) = \text{Pois}\left(n|\mu S + B\right) \left[\prod_{e=1}^{n} \frac{\mu S f_S(x_e) + B f_B(x_e)}{\mu S + B}\right], \tag{5.3}$$

where $\text{Pois}\left(n|\mu S + B\right) = (\mu S + B)^n \frac{e^{-(\mu S+B)}}{n\,!}$.

The negative log-likelihood function, which is minimised in the extended maximum likelihood fit, is found by taking the logarithm of Eq. 5.3, to get

$$-\ln \mathcal{L}(\mu) = -n \ln\left(\mu S + B\right) + (\mu S + B) + \ln n\,! - \sum_{e=1}^{n} \ln\left[\frac{\mu S f_S(x_e) + B f_B(x_e)}{\mu S + B}\right] \tag{5.4}$$

$$= (\mu S + B) + \ln n\,! - \sum_{e=1}^{n} \ln\left[\mu S f_S(x_e) + B f_B(x_e)\right]. \tag{5.5}$$

When using histograms as the signal and background shapes, a binned version of Eq. 5.4 is needed. This can be found by replacing the continuous shapes, with the histograms for signal and background, denoted ν_b^{sig} and ν_b^{bkg}, respectively, where b is the

bin index, and the contents are the number of events expected in the data. Then the continuous shapes can be rewritten as

$$f_S(x_e) = \frac{\nu^{sig}_{b_e}}{S\Delta_{b_e}},\ f_B(x_e) = \frac{\nu^{bkg}_{b_e}}{B\Delta_{b_e}}, \tag{5.6}$$

where b_e is the bin containing entry x_e, and Δ_{b_e} is the bin width. In Eq. 5.6, both $f(x)$ are normalised to unity, since $S = \sum_b \nu^{sig}_b$ and $B = \sum_b \nu^{bkg}_b$.

In terms of individual bins, the likelihood function can be written

$$\mathcal{L}(\mu) = \mathcal{P}(n|\mu) = \text{Pois}(n|\mu S + B) \left[\prod_{b \in bins} \frac{\mu\nu^{sig}_b + \nu^{bkg}_b}{\mu S + B} \right] \tag{5.7}$$

$$= \mathcal{N}_{comb} \prod_{b \in bins} \text{Pois}(n_b|\mu S + B), \tag{5.8}$$

where n_b is the number of events in bin b in data, and $\mathcal{N}_{comb}$ is a combinatorial factor that is constant and can therefore be ignored in the minimisation [83].

5.4.2 Beeston-Barlow method for finite model statistics

To consider the statistical limitations of the models used in the fit method, the Beeston and Barlow method in Ref. [84] will be presented. The method is included in HistFactory by default in a light form, which can be optionally extended to the full version, where the full version takes into account the statistics of every component model in every bin, whilst the light form has one statistical factor per bin for the model statistics.

Consider the binned data sample $\{n_1, n_2, \ldots n_{bins}\}$, where n_b is the number of events in real data that fall into bin b, and the predicted number of events in each bin is $g_b(P_1, P_2, \ldots P_m)$, where P_j are the proportions of different components, m, in the data, to be determined by the fit. In the case presented above for HistFactory, $m = 2$, where the components are the signal and background, with $P_{sig} = \mu S/n$ and $P_{bkg} = B/n$.

The observed number of events in each model in each bin, $\{a_{jb}\}$, can be generated as a Poisson distribution about the unknown, expected number of events in each model in each bin, A_{jb}. In the Beeston-Barlow method, the expected number, instead of the observed number, is used to predict the number of events in each bin. Thus, the prediction for the

number of events in bin, b, in the data is

$$g_b = n \sum_{j=1}^{m} \frac{P_j}{N_j} A_{jb}, \tag{5.9}$$

where n is the total observed events, and N_j is the total number of events from component j. The total likelihood to be maximised is a combination of the probabilities of the observed numbers of events $\{n_b\}$, and the observed number of events in each model $\{a_{jb}\}$ in each bin,

$$\ln \mathcal{L} = \sum_{b=1}^{bins} (n_b \ln g_b - g_b) + \sum_{b=1}^{bins} \sum_{j=1}^{m} \left(a_{jb} \ln A_{jb} - A_{jb}\right). \tag{5.10}$$

The first sum in Eq. 5.10 appears the same as the result for the case without the statistical uncertainties for each of the models in the fit, in the case where a_{jb} is used in place of A_{jb} in Eq. 5.9; this result alone is often used for binned maximum likelihood fits. The second sum provides the Beeston-Barlow treatment for the statistics of the models, comparing the observed to the expected number of events from each model. This method unfortunately results in a maximisation with $m \times (b+1)$ unknowns. This could be as many as 7×41 unknowns in the case of the deuteron measurement.

HistFactory uses a lighter version of the Beeston-Barlow method, which instead of considering a nuisance parameter per bin per component, has only one nuisance parameter per bin [83]. This is analagous to taking one A_b per bin, instead of one for each component, A_{jb}. This takes into account the total estimate from the models, and therefore the total statistical uncertainty in that bin. Consequently, the parameter for each bin will be driven by the model with the largest statistics in the bin.

Including the statistical uncertainty in each bin, following from Eq. 5.7, the likelihood can be written as

$$\mathcal{L} \propto \prod_{b \in bins} \text{Pois}(n_b | \nu_b(\boldsymbol{\beta}) + \gamma_b \Delta_b(\boldsymbol{\beta})), \tag{5.11}$$

where $\boldsymbol{\beta}$ is the set of parameters giving the relative amounts of the various components, similar to the signal strength μ in the two-component example in section 5.4.1, γ_b is the nuisance parameter providing the statistical variation of each bin, $\nu_b(\boldsymbol{\beta})$ is the number of events in each bin, as predicted by the total model for certain values of $\boldsymbol{\beta}$, and $\Delta_b(\boldsymbol{\beta})$

is the difference between the true number of events, and the estimate given by the statistically limited models, analagous to $\sum_j a_{jb} - A_b$.

The light configuration of the Beeston-Barlow treatment in HistFactory is used in the fits to data in ProbNN$d(d)$ percentile for the deuteron production cross-section measurement, which will be described in section 5.7.

5.4.3 Shape variation

In HistFactory, the `HistoSys` element can be used to introduce coherent, systematic shape variations to each component entering the fit [83]. The fit interpolates between the nominal and alternative templates. The interpolation is controlled by a set of parameters that are allowed to float in the fit [87]. Two alternative histograms provide systematic variation for each component, representing the upward and downward variations in the shape, when the corresponding model parameter is varied $\pm\sigma$. This results in extra parameters in the fit: the shape systematic variables, α_{jp}, for component j, and the systematic effect p, which is altered by a factory $\pm\sigma$.

HistFactory interpolates between the three shapes in order to find the shape systematic variable for a given component, where a value of $\alpha_{jp} = 0$ returns the nominal shape for component j, and $\alpha_{jp} = \pm 1$ give the $\pm\sigma$ variations.

For each of the components in the fit to data to extract the deuteron yield, shape variations are provided according to systematic effects on the particle identification response. Details of the shapes for each component are given in the next section (5.5).

5.5 ProbNNd models

In this section, the shapes in ProbNNd for each of the charged particles types will be shown, with accompanying descriptions of how they were aquired and the systematic shape variations for each of them.

The nominal shapes for the deuteron signal are taken from simulation of the deuteron production models in the LHCb detector environment. The background shapes for charged particles and ghost tracks are taken from calibration samples and minimum bias simulation, respectively. The distinct charged particle backgrounds that are considered are pions, kaons, protons, muons and electrons, each of which are present in data in much

higher numbers than deuterons, such that their misidentification as deuterons may be significant. It will be shown that, due to statistical limitations in the calibration samples for muons and electrons, the pion shape is used to model them.

5.5.1 Deuteron models

The ProbNN$d(d)$ percentile shape for deuterons comes from a minimum bias simulation sample with prompt deuteron production using the Cross-section model included [43], as detailed in section 3.1.2. This is the best available shape that can be used for deuterons, as there are no calibration data samples of deuterons to use in place of simulation. Nonetheless, data can be used to indicate what the agreement of the simulated and data shapes could have been if they were both available, and as such, the uncertainty that should be placed on the shape from simulation when applying it to data.

The binning in ProbNNd for the analysis has been chosen using deuterons from the simulated sample, such that the shapes of the deuterons are flat in each momentum bin. Therefore, the shapes for every momentum bin in ProbNN$d(d)$ percentile are identical, as seen in the plots in Fig. 5.10.

Protons are the most similar particles to deuterons in the LHCb detector, with respect to their detection and interactions, because they are stable and closest in mass to deuterons. Therefore, simulation and data for protons are compared and the difference between the two can be used as a model for how different deuterons in data may be to deuterons in simulation. Firstly, for a given momentum bin, protons in simulation are used to find a binning in ProbNNp that gives a flat shape for protons, *i.e.* finding the bin boundaries for ProbNN$p(p)$ percentile binning. Secondly, protons from simulation and from data calibration samples are each plotted in this binning, and each distribution is normalised to unity. The fractional difference between the two normalised shapes in the ProbNN$p(p)$ percentile binning gives the difference between protons in simulation and data and is indicative of the equivalent difference for deuterons in simulation and data as a function of ProbNN$d(d)$ percentile.

It has been shown in section 5.3 that the choice in momentum region being analysed has an impact on the shape and discriminating power of the ProbNNd variable. This is also true for ProbNNp, though the trends in the case of ProbNNp are different to those in ProbNNd. For the proton discriminator, ProbNNp, the proton RICH momentum thresholds play a bigger role than the deuteron RICH momentum thresholds. To account

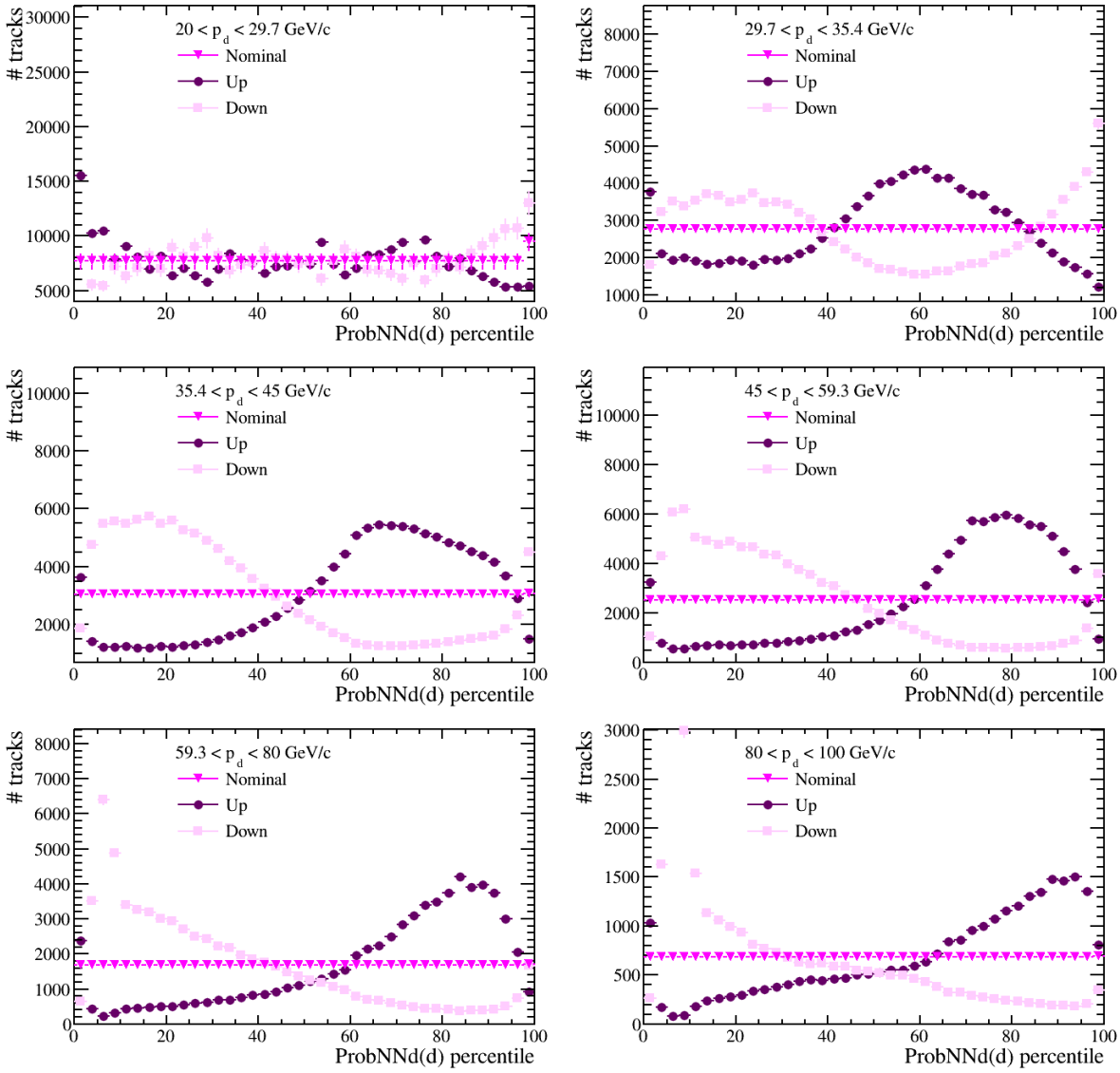


Figure 5.10: Deuteron systematic shape variation in each momentum bin: nominal shape taken from minimum bias simulation with prompt deuteron production, upwards and downwards variations obtained from protons in calibration samples and minimum bias simulation.

for this difference, the fractional differences in shapes between simulation and data found for protons of certain momenta, must be applied to deuterons at different momenta.

Deuterons in the momentum bin $35.4 < p < 45.0\,\text{GeV}/c$ are above the deuteron threshold in RICH 1, but below the deuteron threshold in RICH 2. Therefore, protons above their RICH 1 momentum threshold and below their RICH 2 threshold, corresponding to $17.7 < p < 29.7\,\text{GeV}/c$, must be used to find the systematic shape difference here.

Proton bin (GeV/c)	**Deuteron bin (GeV/c)**
9.3 – 15.6	20.0 – 29.7
16.0 – 17.7	29.7 – 35.4
17.7 – 29.7	35.4 – 45.0
29.7 – 35.4	45.0 – 59.3
35.4 – 45.0	59.3 – 80.0
45.0 – 59.3	80.0 – 100.0

Table 5.3: Momentum bins for shape differences found between protons in data and simulation, which are translated to shape variations for deuterons in their corresponding momentum bins.

The momentum bins for protons and deuterons that are used in this way are listed in Tab. 5.3, with the momentum bin used to find the difference in shape between simulation and data for protons in ProbNN$p(p)$ percentile on the left, and the momentum bin to which this shift is applied for deuterons in ProbNN$d(d)$ percentile on the right.

The resultant shapes for the deuteron shape systematics are shown in Fig. 5.10. The nominal shape is the flat deuteron shape in ProbNN$d(d)$ percentile that enters the fit. The 'up' and 'down' shapes are the upward and downward shape variations that are given to the HistFactory fit, where the nominal shape in the fit is allowed to vary between them. In the lowest momentum bin, on the top left of Fig. 5.10, there is no overarching structure to the 'up' and 'down' shapes compared to the flat deuteron shape, indicating that the difference between the proton data and simulation shapes was small. In the remaining momentum bins, a trend can be seen in the 'up' and 'down' shapes. The 'up' shape is consistently shifted to the higher ProbNN$d(d)$ percentile values than the nominal shape, indicating that the protons in data tended to have higher ProbNN$p(p)$ percentile values than the protons in simulation. That being said, a steep decline in the 'up' shape is seen at the highest ProbNN$d(d)$ percentile values in all of these momentum bins, showing that the protons in data are less likely to have the highest values of ProbNN$p(p)$ percentile than protons in simulation. By design, the 'down' shapes show the opposite behaviour to the 'up' shapes.

5.5.2 Models for charged particles

For pions, protons, kaons, muons and electrons, the ProbNN$d(d)$ percentile shapes are derived from datasets of pure particle species, made by the LHCb PID Working Group for the PID Calib package [77]. As detailed in section 4.2.6, these particles are selected using common decays of particles, where one decay product is identified and the other is known without use of any PID information. The resulting calibration samples from 2018 data taking are used to find the shapes for each particle type in ProbNN$d(d)$ percentile.

This section will begin with a short comment on the muon and electron shapes, followed by the pion, kaon and proton shapes that enter into the fit.

Muon and electron models

As was seen in Fig. 5.6, the muon, electron and pion shapes in ProbNNd agree with each other for every bin of momentum. The same is true in ProbNN$d(d)$ percentile, in Fig. 5.3, albeit with limited statistics for muons and electrons in these plots.

Distributions in ProbNN$d(d)$ percentile of muons and electrons from calibration data samples are shown in bins of momentum in Fig. 5.11, where the effects of the limited statistics are evident. In all momentum bins, the distributions for both muons and electrons are discontinuous, with some empty bins in higher ProbNNd (d) percentile bins. The electron shapes are particularly unreliable, with accordingly large errors.

The reason for the statistical limitation is both because the calibration samples are not large, particularly in the case of electrons, as seen in Tab. 5.1, but also because of the trend in muons and electrons in ProbNNd. The neural network offers good discrimination between muons, electrons and deuterons, because the muons and electrons have much smaller masses than deuterons, and therefore appear different to the RICH algorithm. Consequently, most muons and electrons will have ProbNNd values within the first few ProbNN$d(d)$ percentile bins, leaving very few muons and electrons to provide the shape in the upper ProbNN$d(d)$ percentile bins. This means that muons and electrons are not well parametrised for the full ProbNN$d(d)$ percentile range, and the sparse shapes can cause problems with the fit procedure.

In the case when a template has low statistics, spikes and empty bins are likely to appear in the shape. If these shapes are included in the generation of the pseudo-data in order to perform fits to validate the fit procedure, then the same spikes or dips will

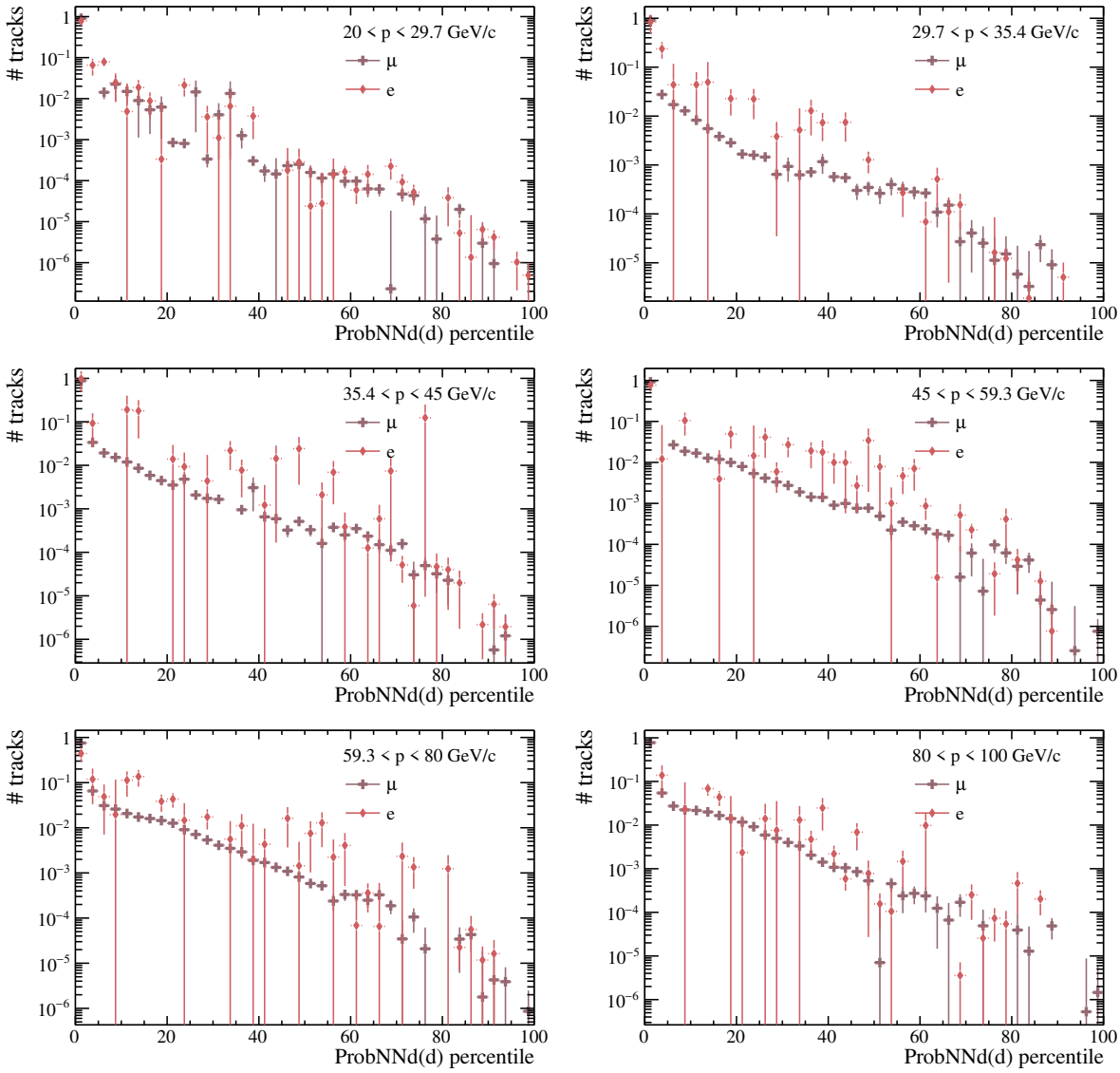


Figure 5.11: Shapes in ProbNN$d(d)$ percentile of muons and electrons from calibration data samples in bins of momentum, normalised to unity.

appear in the pseudo-data. When a fit is performed to this pseudo-data, with these templates, the fit will over-perform, as it is able to latch on to the distinct features in the shape of the pseudo-data. This leads to errors in the fit parameters being underestimated compared to what will happen when the fit is performed with the same templates on data, where the distinct features do not exist [87].

It is evident in the plots in Fig. 5.11 that the muon and electron shapes agree with each other. Therefore, instead of using the electron shape from the calibration data samples, the muon shape is used as an estimation of the electron shape. This provides a

smoother shape, with fewer empty bins in the histograms, because the calibration sample for muons has higher statistics than the sample for electrons. In pseudo-experiments in section 5.6, the muon shape is used to represent both the muon and the electron components in pseudo-data.

In addition, after studying the performance of the fit with pseudo-experiments, it was decided not to float the muon and electron yields individually. Instead the muon and electron shapes are added as variations to the pion shape in the fit, as the pion, muon and electron shapes are in agreement for each momentum bin in ProbNNd(d) percentile, as seen in Fig. 5.3 from simulation. The percentage of tracks from each particle type in each momentum bin is listed in Tab. 5.7. At the relevant momenta, according to minimum bias simulation, muons constitute about 0.1% of tracks, and electrons make up $< 2\%$ of tracks, compared to pions which constitute about 70% of tracks. The muon and electron yields are both higher than the expected deuteron yield, which is at about 0.02% of tracks. Thus, the muon and electron shapes are added as shape variations to the pion component, according to these relative yields from simulation. This approach was seen not to bias the result of the deuteron signal in fits to pseudo-data.

The electron and muon yields in the fits are included in the pion yield, and a variation in the pion template to account for the addition of the muons and electrons is detailed in the next section.

Pion, kaon and proton models

The pion, kaon and proton shapes for the fit are taken from data. Whilst this averts issues due to data–simulation disagreements, the calibration data used has different kinematic coverage to the minimum bias data to which it is applied. Therefore, the shapes from calibration cannot be used without accounting for the kinematic differences first.

When using calibration data in a measurement, it is important to consider the kinematic coverage of the calibration samples compared to the kinematics of the signal one is trying to measure [77]. In the case of the deuteron measurement, the calibration samples, in which all particles come from decays in triggered data, are used to find shapes that are fit to minimum bias data. Minimum bias data is by design not biased by a trigger, and the majority of tracks in it will come directly from the pp interaction, instead of from decays. To find the kinematic distributions of pions, kaons and protons in minimum bias data, minimum bias simulation is used. Momentum and pseudorapidity

RICH coverage	η bin edges
RICH 1 only	1.890, 2.005, 2.120, 2.235, 2.350, 2.465, 2.580, 2.695
RICH 1 and RICH 2	2.810, 2.908, 3.006, 3.104, 3.203, 3.301, 3.399, 3.497, 3.595, 3.693, 3.791, 3.889, 3.988, 4.086, 4.184, 4.282, 4.380
RICH 2 only	4.444, 4.508, 4.571, 4.635, 4.699, 4.763, 4.826, 4.890

Table 5.4: Pseudorapidity bins used to find the distributions of calibration tracks and tracks in minimum bias simulation, and the RICH coverage corresponding to the pseudorapidity regions.

are used to parametrise the kinematics of the tracks in data and simulation, as these have the biggest effect on PID performance.

Normalised, two-dimensional distributions in momentum, p, and pseudorapidity, η, for charged particles from minimum bias simulation and calibration data samples are shown for protons in Fig. 5.12 and 5.13, for kaons in Fig. 5.14 and 5.15, and for pions in Fig. 5.16 and 5.17. The distributions from simulation are on the left, with distributions from data on the right, for each momentum bin used for the deuteron measurement.

The bin widths of the histograms in (p, η), in Fig. 5.12 to 5.17, are chosen to be large enough that the histograms do not become statistically limited, whilst having high enough granulatity for the PID response across each bin to be uniform, such that uncertainties arising from the choice in binning are minimised [77]. For the momentum bins, consideration has already been made in section 5.3 for the RICH momentum thresholds that effect the PID response, so the bin boundaries are of even width between the lower and upper bound for each momentum region of the measurement. The pseudorapidity bins are chosen to be of even width between the edges of the RICH detector acceptances, at $\eta = 1.89$, 2.81, 4.38 and 4.89. The pseudorapidity bin boundaries used are listed in Tab. 5.4.

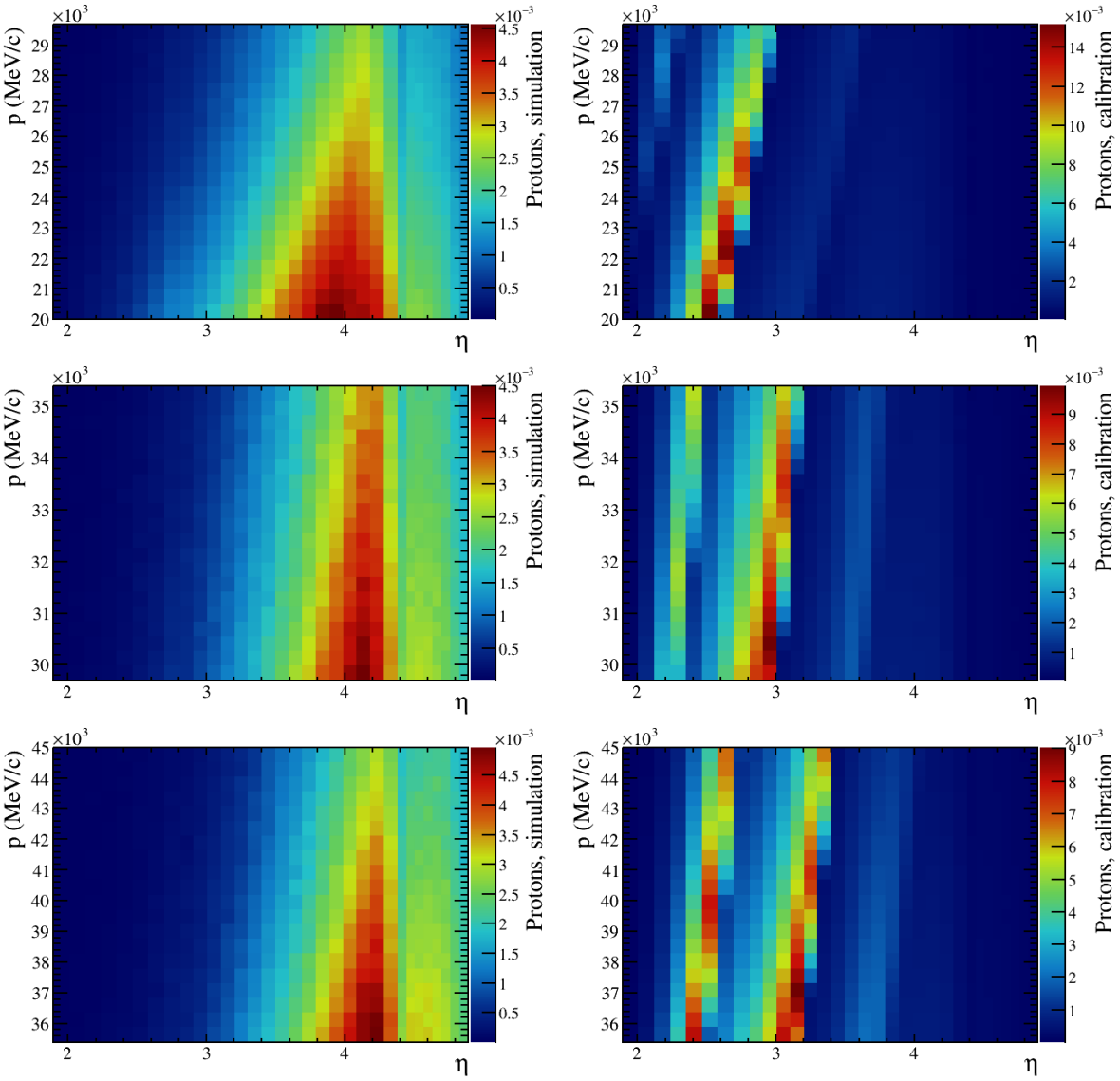


Figure 5.12: Distributions of momentum vs. pseudorapidity for protons from minimum bias simulation (left) and from calibration data samples (right) in bins of momentum.

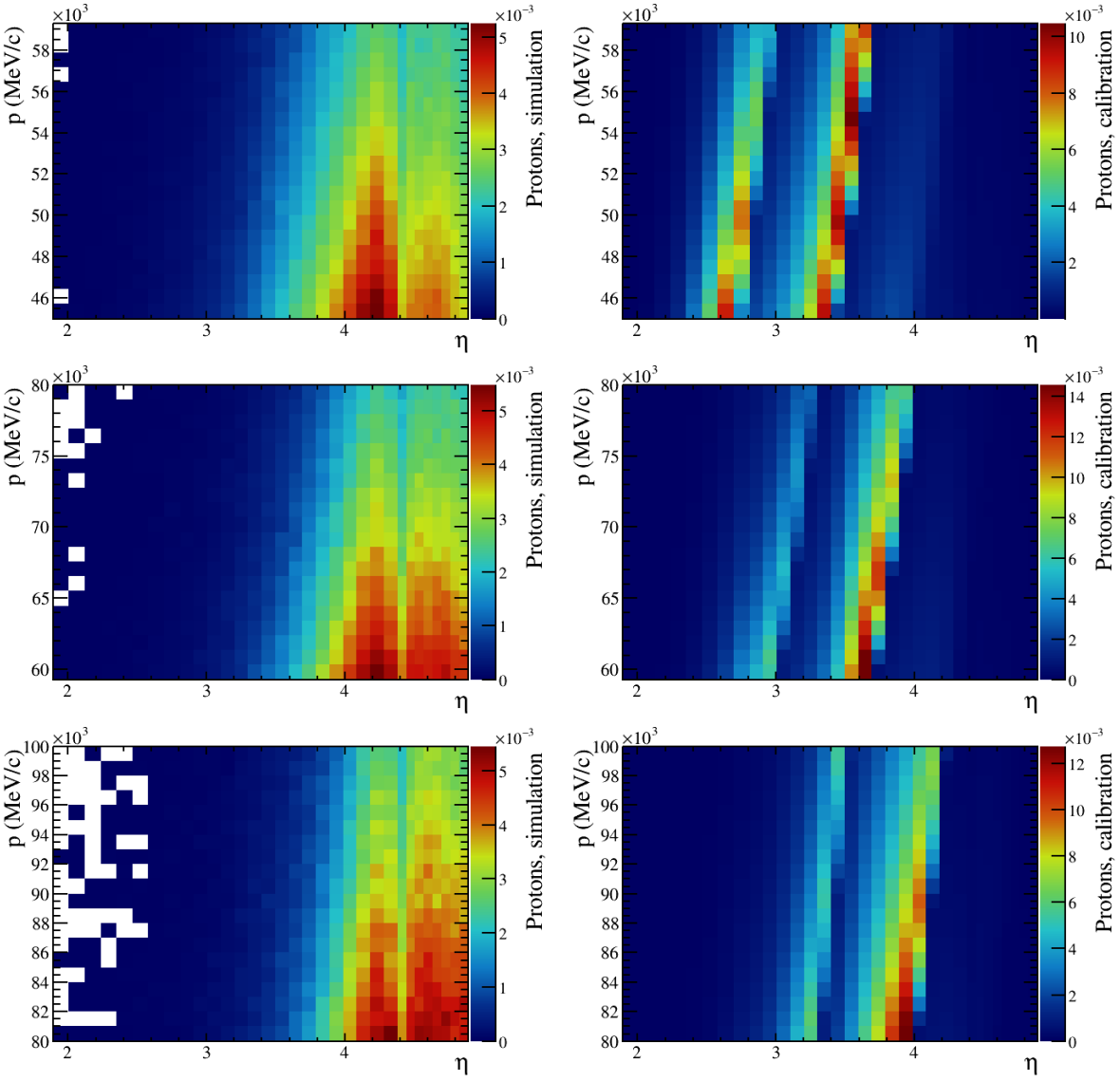


Figure 5.13: Distributions of momentum vs. pseudorapidity for protons from minimum bias simulation (left) and from calibration data samples (right) in bins of momentum, normalised to unity.

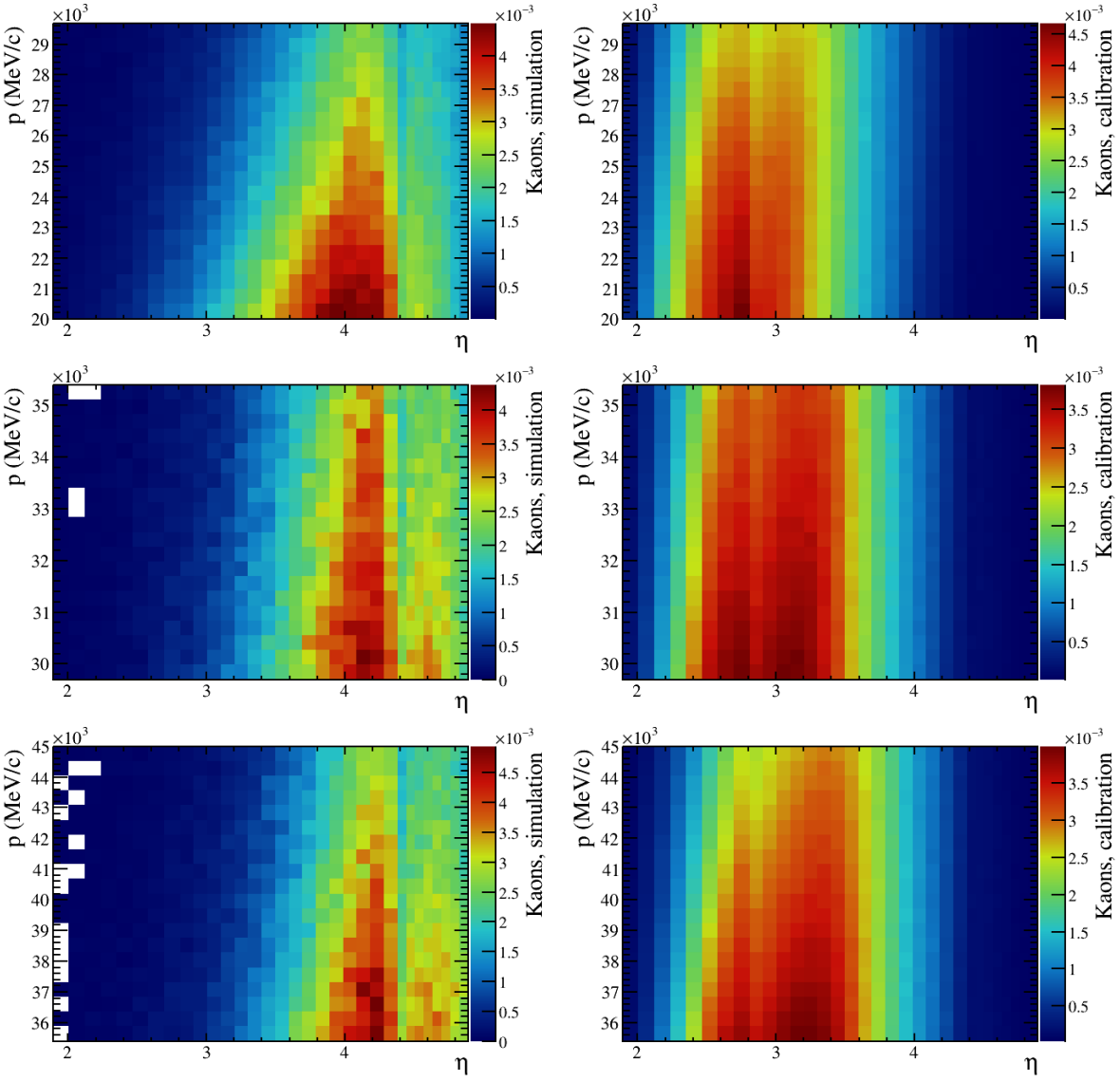


Figure 5.14: Distributions of momentum vs. pseudorapidity for kaons from minimum bias simulation (left) and from calibration data samples (right) in bins of momentum, normalised to unity.

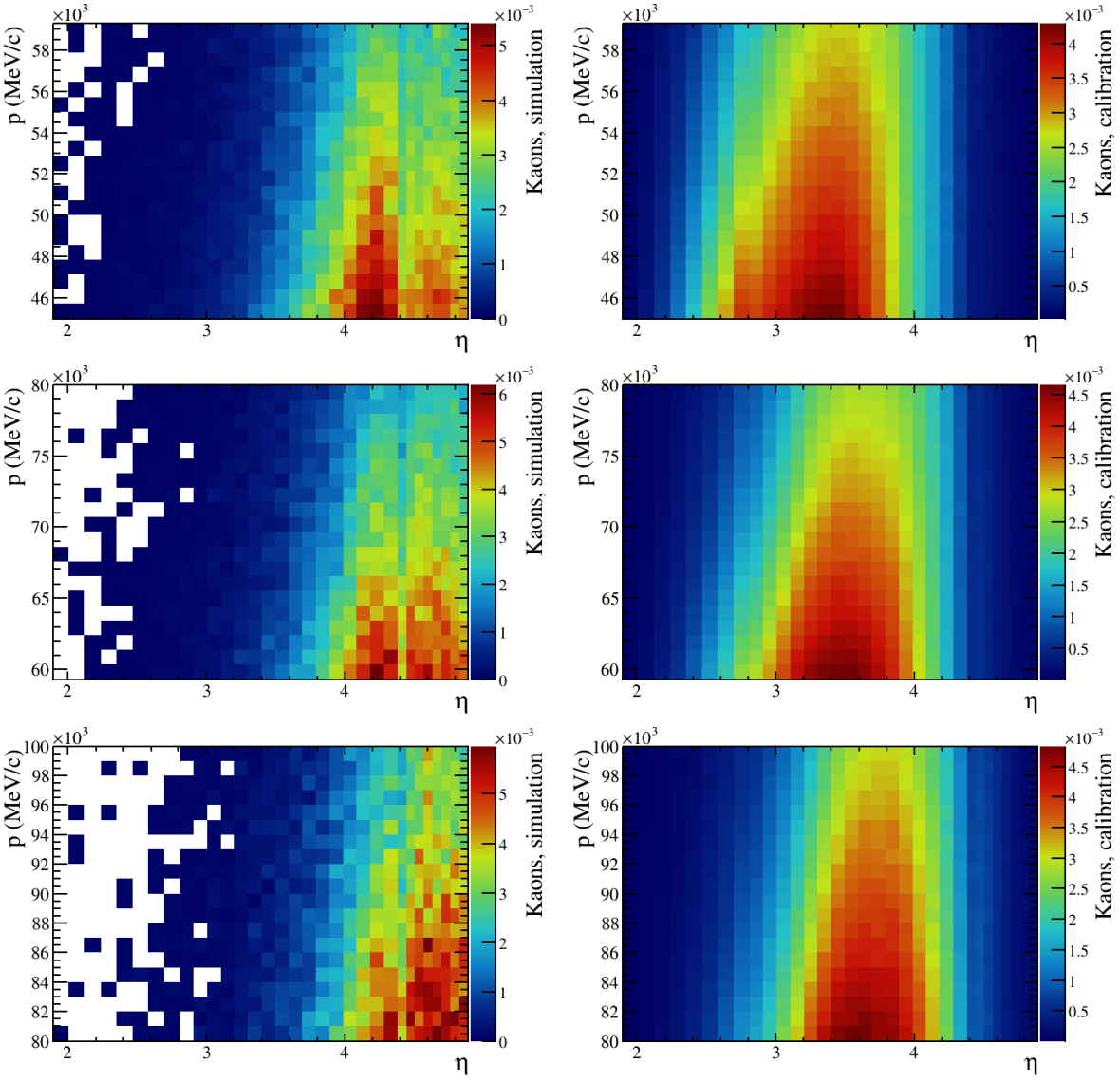


Figure 5.15: Distributions of momentum vs. pseudorapidity for kaons from minimum bias simulation (left) and from calibration data samples (right) in bins of momentum, normalised to unity.

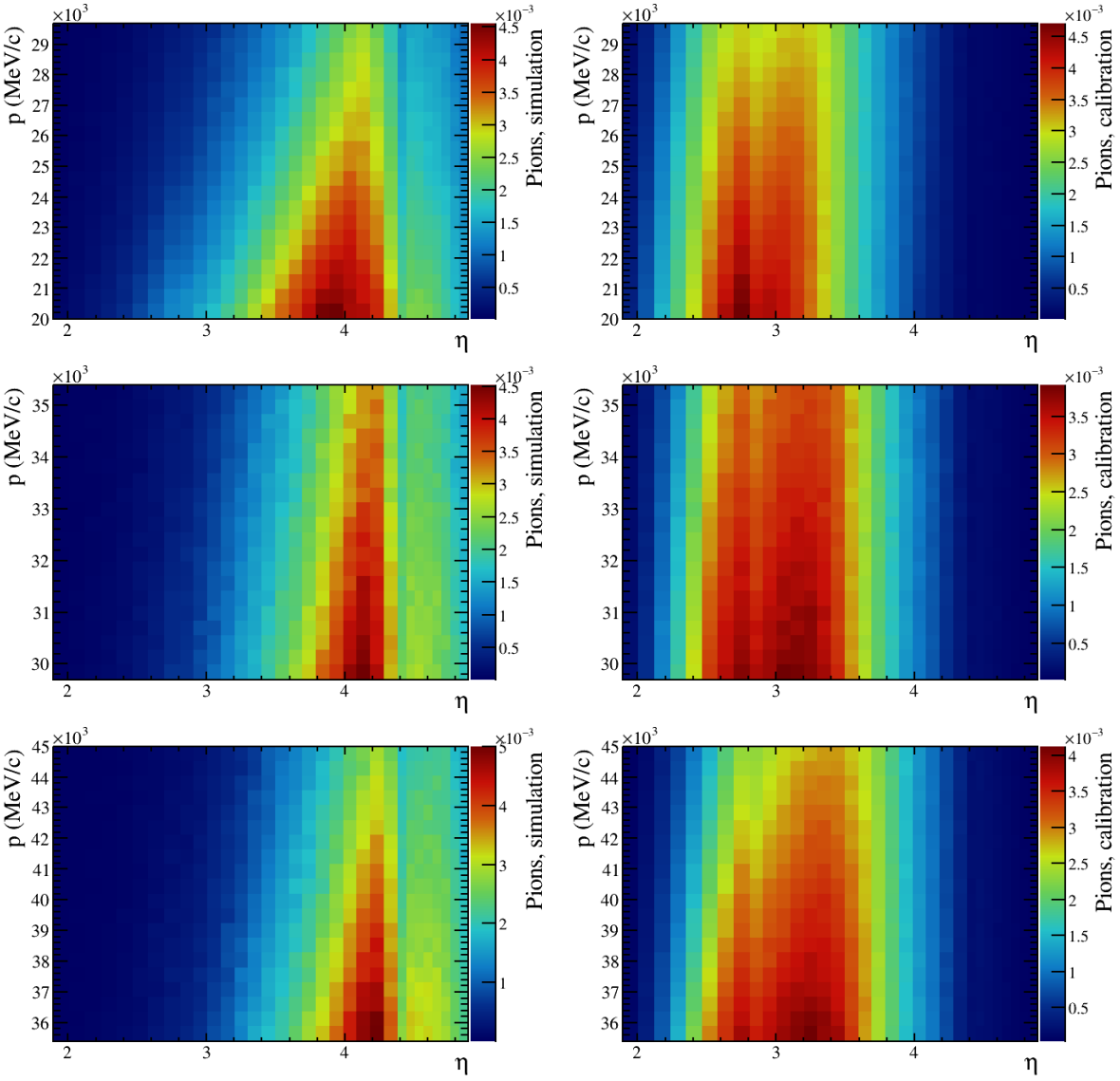


Figure 5.16: Distributions of momentum vs. pseudorapidity for pions from minimum bias simulation (left) and from calibration data samples (right) in bins of momentum, normalised to unity.

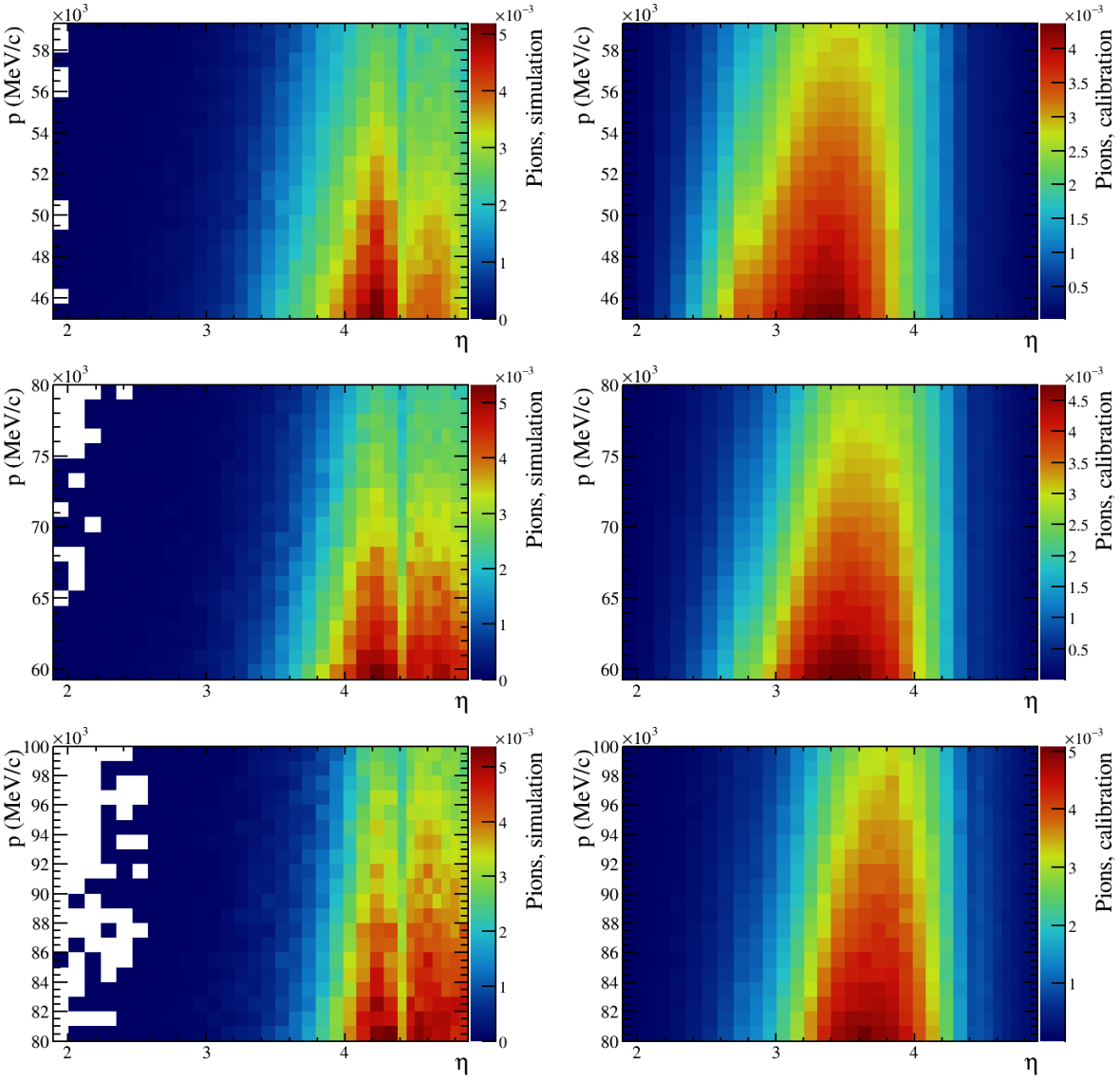


Figure 5.17: Distributions of momentum vs. pseudorapidity for pions from minimum bias simulation (left) and from calibration data samples (right) in bins of momentum, normalised to unity.

Most strikingly in the plots in Fig. 5.12 to 5.17, the tracks from minimum bias simulation on the left are, in general, more forward than those from calibration data, such that the distributions are shifted towards higher η values. This is because the calibration tracks are selected from decays, whilst the minimum bias simulation samples are dominated by prompt tracks. The prompt tracks travel preferentially along the beampipe direction, as they are boosted by the colliding protons, whilst the tracks from decays are more likely to travel in a more outward direction, because they are produced back-to-back with other decay products from the decay vertex. It is also the case that

decays are more easily selected in LHCb data and simulation if they are more central, as observation of the decay vertex requires the outgoing tracks to be separated with respect to each other [81]. This results in particles from discernable decays having a lower pseudorapidity than prompt particles, due to the bias in detection.

In Fig. 5.12 and 5.13 for protons, the plots to the right, from calibration data, are not smooth distributions in momentum and pseudorapidity. This is because the proton calibration data samples for protons have transverse momentum selections applied to them to achieve a good overall momentum coverage. Protons in data are selected from $\Lambda \to p\pi^-$ decays, which have a very high cross-section at the LHCb experiment. In order to improve the kinematic coverage that these tracks will provide, the decay is selected for different p_{T} ranges, with a prescaling factor applied to the lower momentum tracks, to suppress their contribution. The selections can be seen in Fig. 5.18, where a red line is drawn for each of the p_{T} regions for illustration. The p_{T} regions are at 1.5, 3.0 and 6.0 GeV/c, where a prescale is applied to the tracks selected with the lowest p_{T} limit, and lesser prescales applied to the subsequent regions, such that a more even distribution of tracks in momentum is achieved. These different regions appear as features that look like stripes in Fig. 5.12 and 5.13. No such prescale is applied to the other charged particles.

Visible in all distributions from minimum bias simulation in Fig. 5.12 to 5.17 is a feature at approximately $\eta = 4.4$ where there is a depletion in tracks for all momentum values. This break in the distributions is caused by the insulation surrounding the beampipe [57], which interrupts tracks travelling at this pseudorapidity. The effect is also present in data, but is not visible in the samples presented here, because the pseudorapidity coverage of the tracks does not give sensitivity in this forward range.

For each particle species, the pseudorapidity distribution tends towards to higher, more forward values as momentum increases, such that the areas of concentration in the figures appear to drift to the right. This is because high momentum tracks are produced preferentially along the beampipe direction, or z-axis, because they are boosted by the momentum of the colliding protons. In all plots, the tracks are most concentrated at the bottom of the 2D distribution, at the lower momentum edge of each momentum bin, because more tracks are produced with low momentum than high momentum. This effect is masked in the plots of protons from the calibration samples, to the right of Fig. 5.12 and 5.13, which have been prescaled to prevent tracks with low p_{T} dominating the distributions and offer better overall momentum coverage.

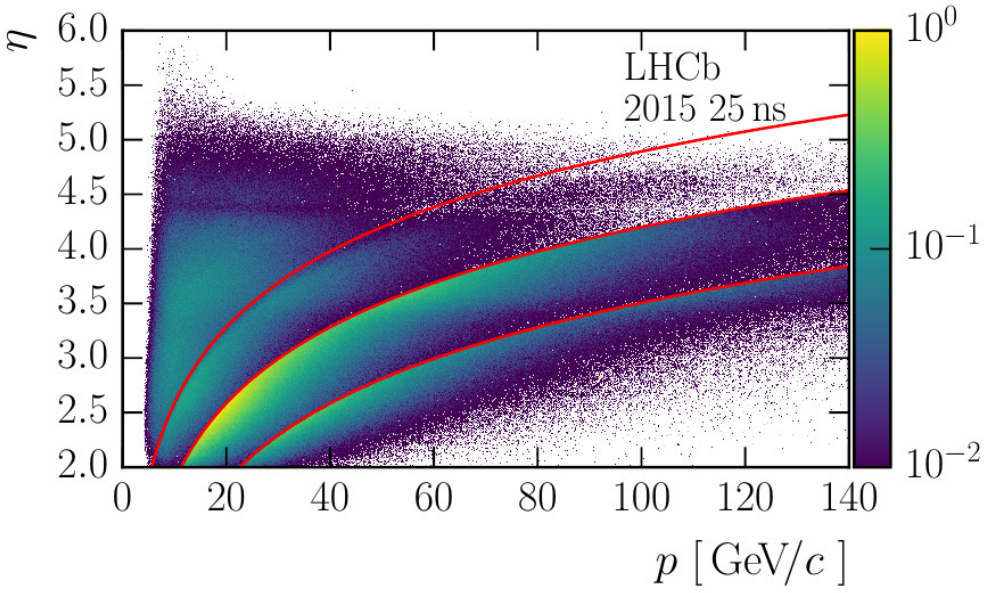


Figure 5.18: Distribution of proton calibration sample track pseudorapidity vs. momentum for the sample collected during the 2015 data-taking period, from Ref. [77]. The red lines correspond to p_{T} thresholds of 1.5, 3.0 and 6.0 GeV/c. The z-axis units are arbitrary. The same p_{T} selections were applied to the calibration samples of protons collected in 2018.

Due to the differences in coverage between tracks from minimum bias simulation and the calibration samples, the tracks in the calibration samples are weighted, so that their kinematic distribution matches that of minimum bias tracks. The shapes for the fit in ProbNN$d(d)$ percentile for each particle species are found by weighting each histogram entry.

To perform the weighting, two-dimensional histograms in momentum and pseudorapidity for each particle species from the data calibration samples and the minimum bias simulation are made, for each momentum bin of the measurement. These histograms are written as $h_i^{Sim.}(p,\eta)$ from simulation, and $h_i^{Calib.}(p,\eta)$ from the calibration data, where i denotes the momentum bin index. These histograms are each normalised to unity, written as $n_i^{Sim.}(p,\eta)$ and $n_i^{Calib.}(p,\eta)$, as shown in Fig. 5.12 to 5.17. A two-dimensional histogram of weights $W_i(p,\eta)$ is found by dividing them

$$W_i(p,\eta) = \frac{n_i^{Sim.}(p,\eta)}{n_i^{Calib.}(p,\eta)}, \tag{5.12}$$

which gives the relative distribution of tracks in simulation compared to the distribution in data. When using the calibration sample, for each particle in the sample, the weight

corresponding to its momentum and pseudorapidity is found, and it is multiplied by the weight when it is added to the shape in ProbNN$d(d)$ percentile, such that the recovered shape is representative of what is expected of tracks with minimum bias kinematics.

From Fig. 5.12 to 5.17 it would appear that the calibration samples, shown on the right, do not offer coverage in the forward pseudorapidity region, as will be needed to reweight the simulation tracks. However, the calibration samples have high statistics, meaning that even in the forward, depleted region, there are sufficient entries in the 2D-histograms to find meaningful weights. The numbers of tracks in each sample are shown in Tab. 5.1.

The calibration data tracks are plotted alongside the same tracks with the weights applied in ProbNN$d(d)$ percentile in bins of momentum for protons in Fig. 5.19, for kaons in Fig. 5.20, and for pions in Fig. 5.21. In all cases, the points for the weighted tracks fall off at a steeper rate at high percentile values, compared to the tracks taken directly from the calibration data samples. Note the logarithmic scale on the y-axis of these plots.

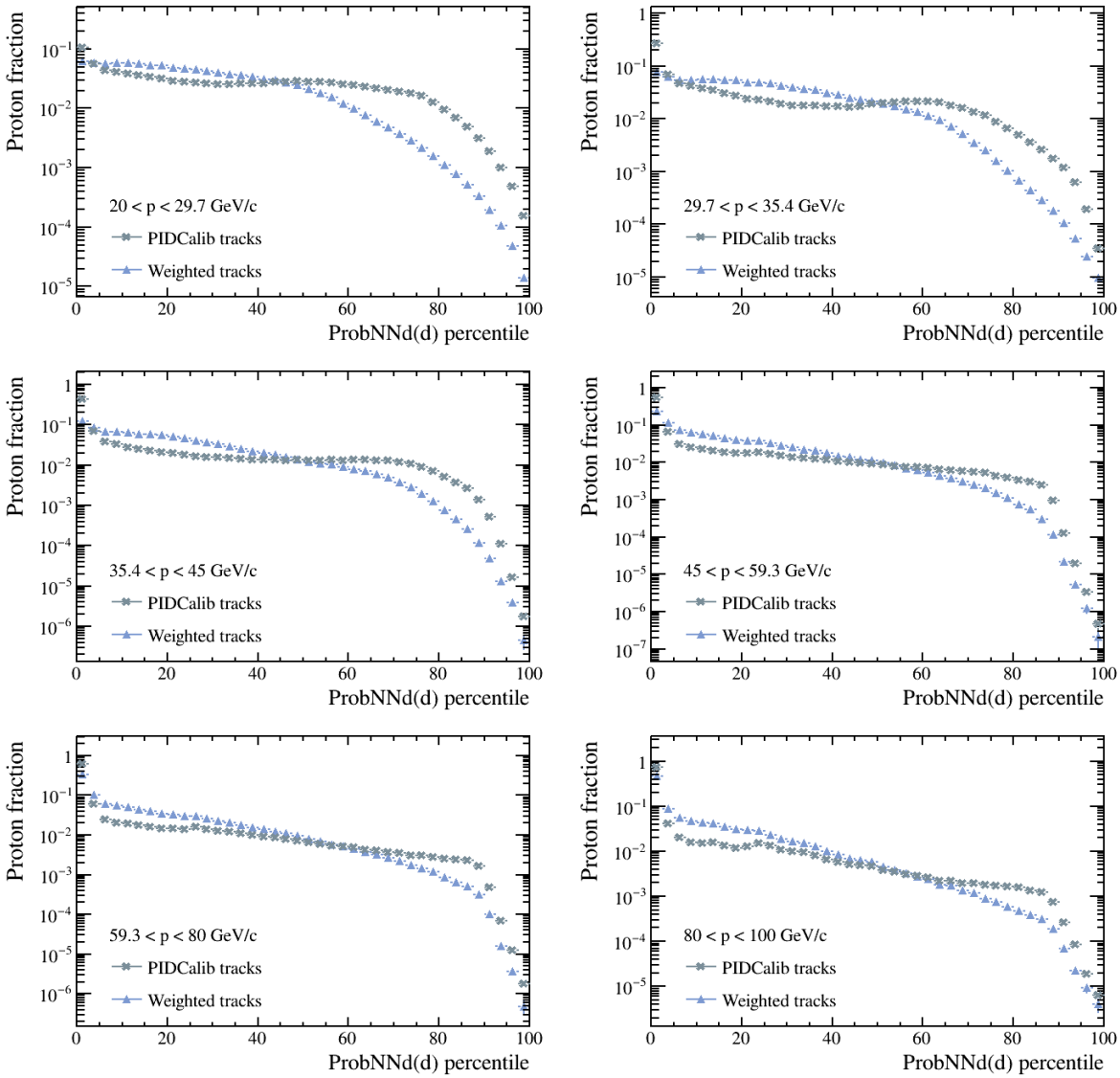


Figure 5.19: Proton distributions in ProbNN$d(d)$ percentile of tracks from calibration samples, and the tracks after weighting them to match minimum bias kinematics, in bins of momentum, normalised to unity.

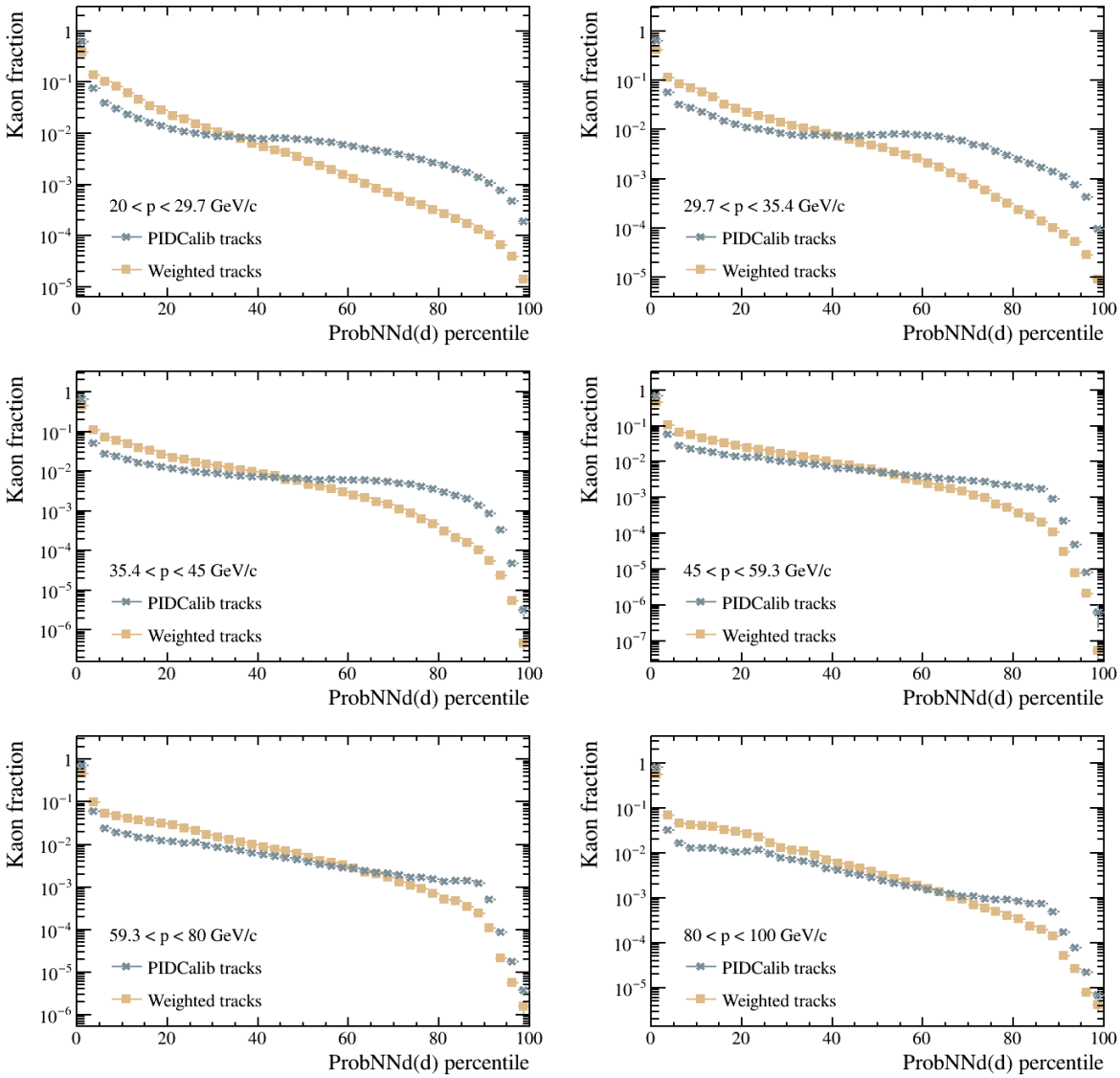


Figure 5.20: Kaon distributions in ProbNN$d(d)$ percentile of tracks from calibration samples, and the tracks after weighting them to match minimum bias kinematics, in bins of momentum, normalised to unity.

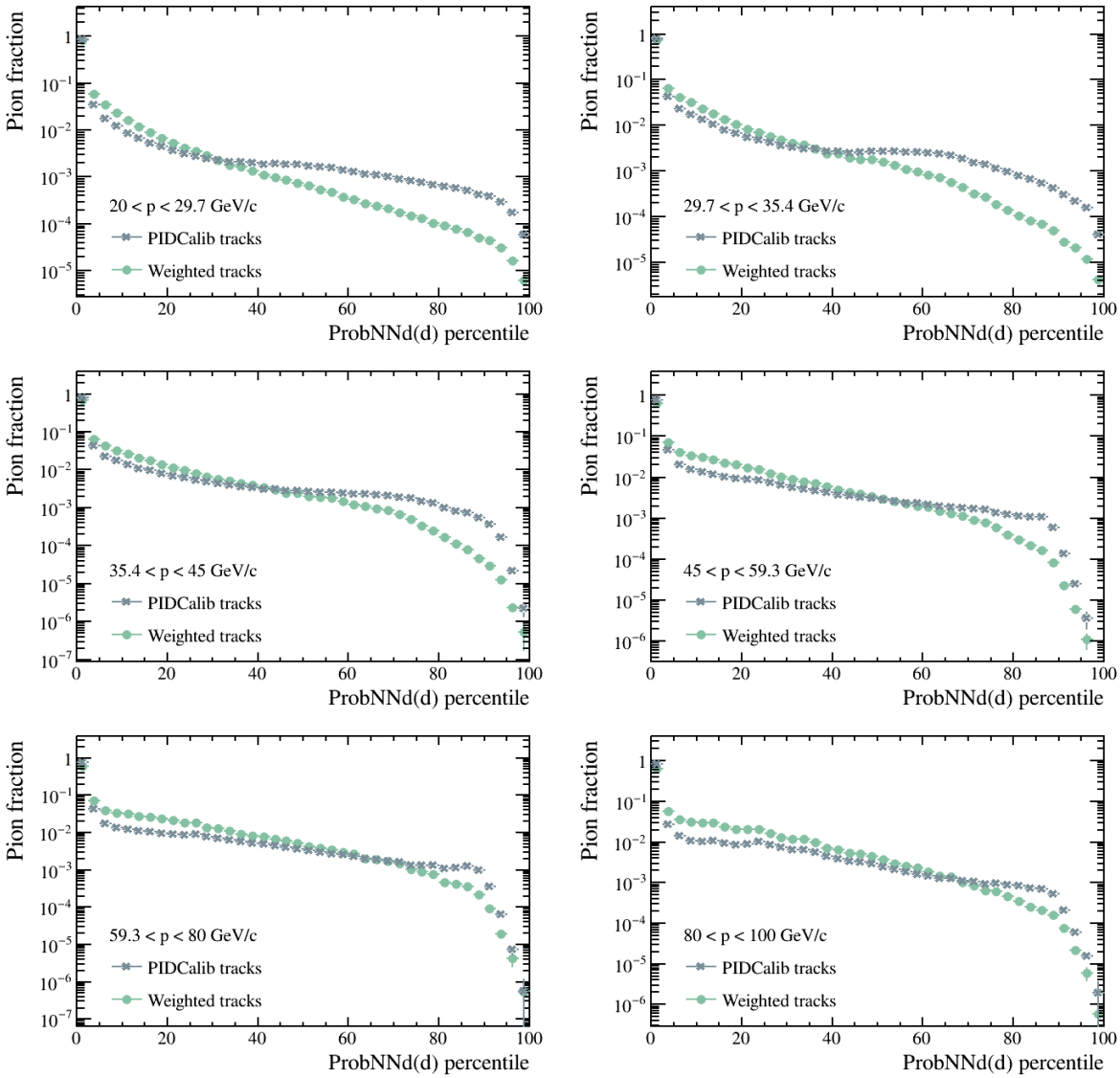


Figure 5.21: Pion distributions in ProbNN$d(d)$ percentile of tracks from calibration samples, and the tracks after weighting them to match minimum bias kinematics, in bins of momentum, normalised to unity.

		Requirements			
Particle	**Decay**	p **(GeV/c)**	p_{T} **(GeV/c)**	χ^2_{IP}	χ^2**/ndf**
p	$\Lambda \to p\pi^-$	> 2.0	$> 0.0,\ 1.5,\ 3.0,\ 6.0$	> 36	< 4
K	$D^* \to D^0\pi^+,\ D^0 \to K^-\pi^+$	> 2.0	> 0.25	> 16	< 3
π	$D^* \to D^0\pi^+,\ D^0 \to K^-\pi^+$	> 2.0	> 0.25	> 16	< 3
μ	$J/\psi \to \mu^+\mu^-$	> 3.0	> 0.5	> 9	< 5
e	$J/\psi \to e^+e^-$	> 3.0	> 0.5	> 9	< 5

Table 5.5: Requirements placed on tracks when selecting the pure calibration samples from the stated particle decays; full details are in Ref. [77].

Not only are the kinematics of the minimum bias and calibration sample tracks different; the tracks in the calibration samples also have some requirements placed on them. These requirements are listed in Tab. 5.5.

The requirements placed on p and p_{T} of the calibration sample tracks have a small effect on the kinematic distributions of tracks in the momentum range being used for the deuteron measurement. The lowest momentum bound in this analysis, 20.0 GeV/c, is much greater than the requirements placed on calibration data tracks, which are all below 6.0 GeV/c, such that all tracks considered in this analysis would pass the requirement. Other requirements on the tracks in the calibration samples are typical of those applied when measuring decays of long-lived particles, as they help to eliminate backgrounds from prompt tracks. In the case of the deuteron measurement, where the signal is prompt, no such selections are applied to the data. Consequently, even after the calibration, where tracks have been weighted to match minimum bias kinematics, some residual effect may exist due to the track selections. To account for this, a shape variation is allowed in the proton, pion and kaon templates, corresponding to the effect the track selections have.

Distributions in ProbNN$d(d)$ percentile with and without the PIDCalib selections applied are shown in Fig. 5.22 for protons, Fig. 5.23 for kaons, and Fig. 5.24 for pions, from minimum bias simulation in bins of momentum. These comparisons are limited by the statistics of the simulation samples, as is evident by the large vertical error bars in some ProbNN$d(d)$ percentile bins for the histograms with the cuts applied. However, in most cases, the histograms with and without cuts applied are seen to agree with each other. Two notable deviations from this are the shapes for protons and kaons in the first momentum bin, $20.0 < p < 29.7$ GeV/c, where the shapes with the PIDCalib selections applied do not follow the same pattern as the shapes without any selections applied.

The fractional changes to the minimum bias shapes when the cuts are applied compared to without the cuts are given as systematic shape variations to the pion, kaon and proton shapes in the template fits.

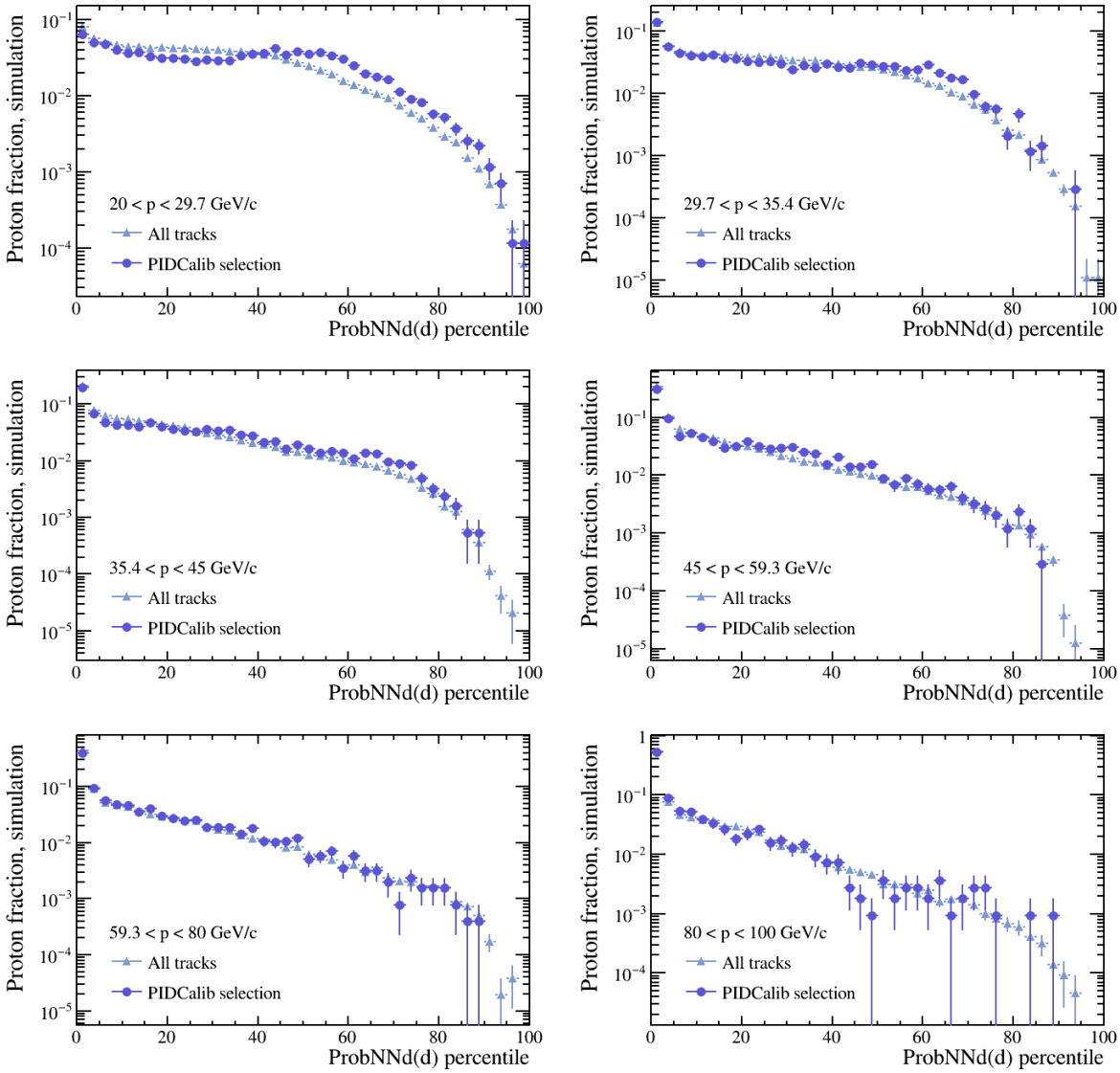


Figure 5.22: Comparison of shapes in ProbNN$d(d)$ percentile of protons from minimum bias simulation with and without PIDCalib cuts applied: $p_{\mathrm{T}} > 2.0\,\mathrm{GeV}/c$, $\chi^2_{\mathrm{IP}} > 36$, and $\chi^2/\mathrm{ndf} < 4$.

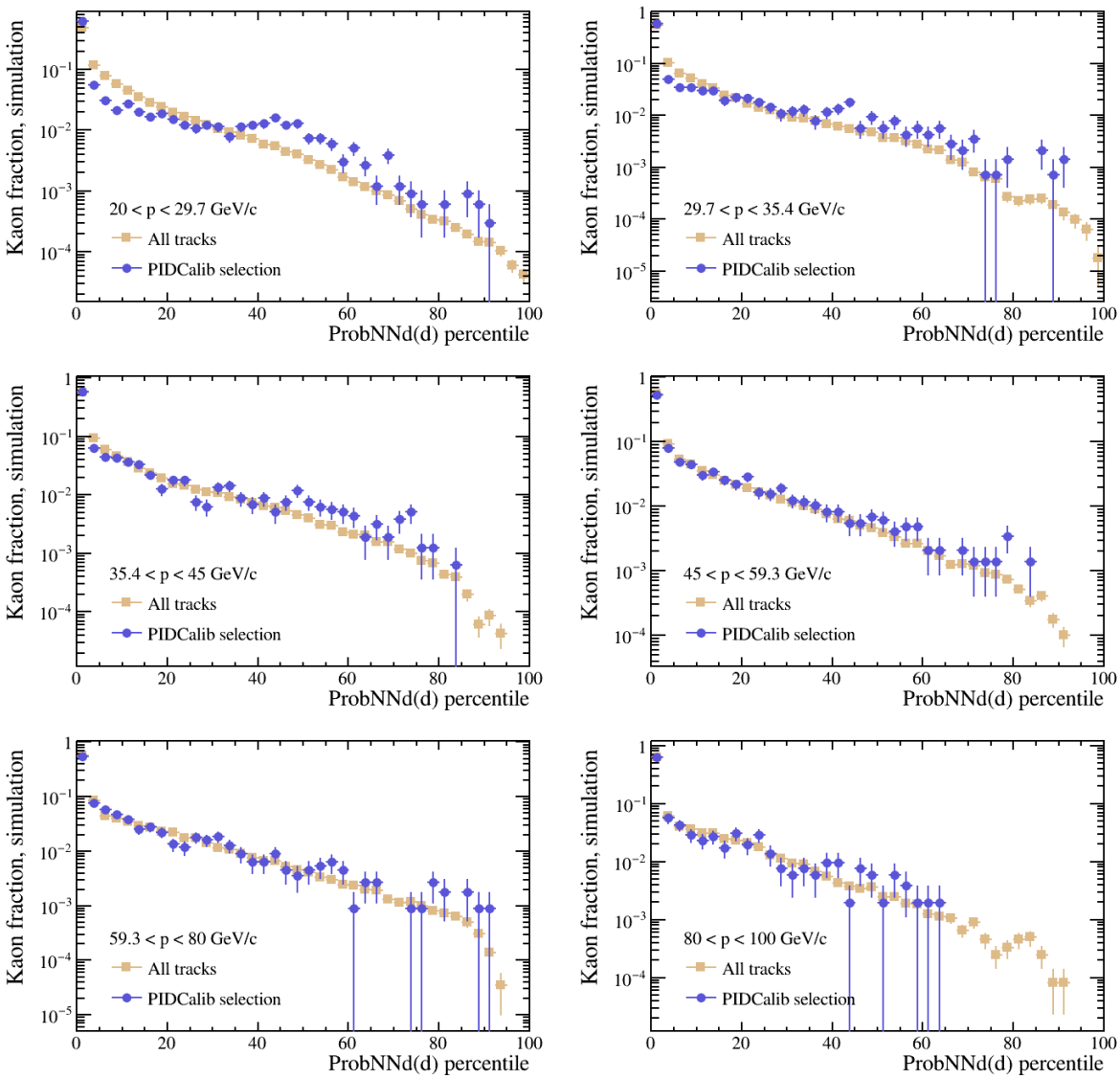


Figure 5.23: Comparison of shapes in ProbNN$d(d)$ percentile of kaons from minimum bias simulation with and without PIDCalib cuts applied: $p > 2.0\,\text{GeV}/c$, $p_\text{T} > 0.25\,\text{GeV}/c$, $\chi^2_\text{IP} > 36$, and $\chi^2/\text{ndf} < 3$.

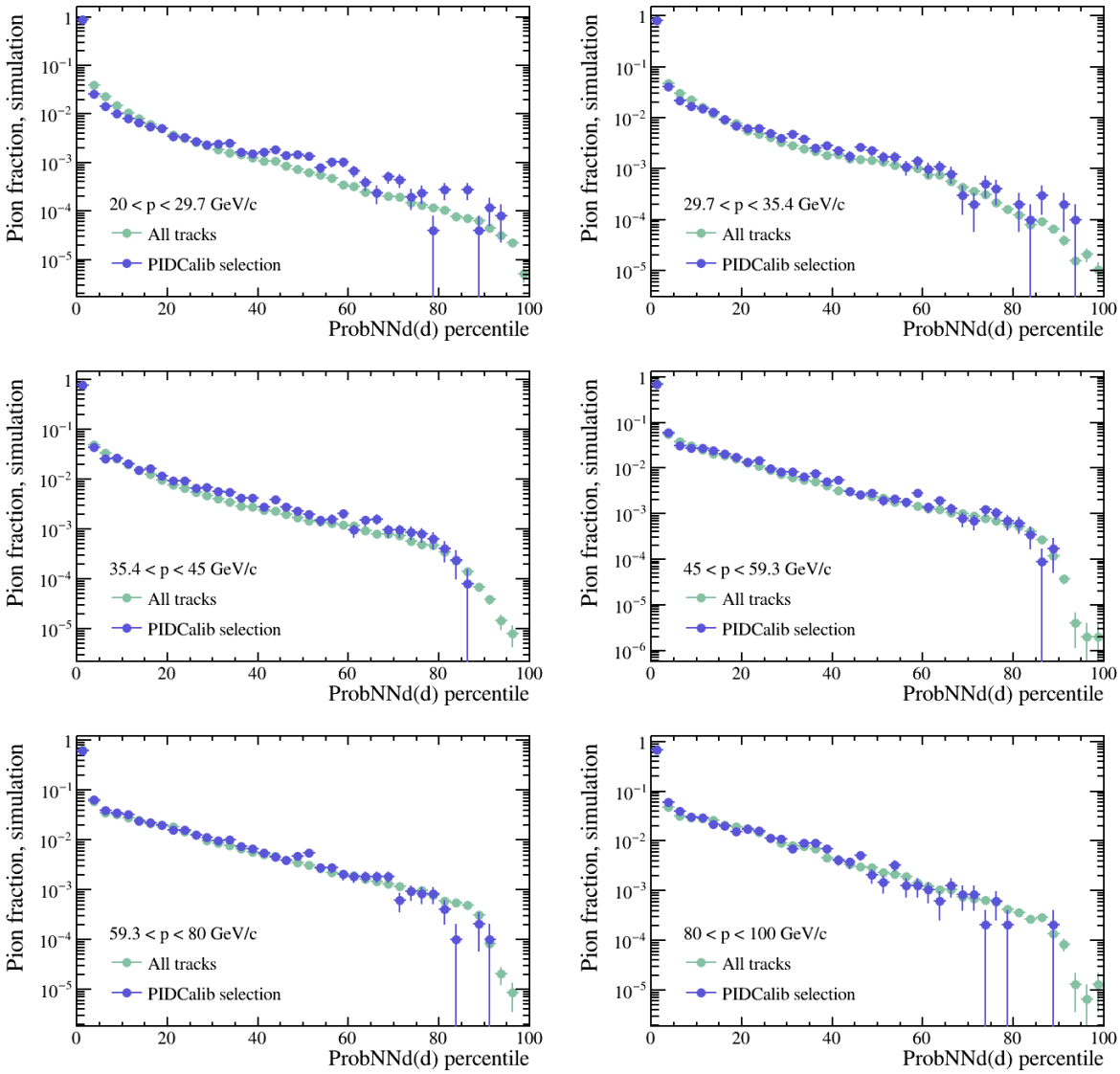


Figure 5.24: Comparison of shapes in ProbNN$d(d)$ percentile of pions from minimum bias simulation with and without PIDCalib cuts applied: $p > 2.0\,\text{GeV}/c$, $p_\text{T} > 1\,\text{GeV}/c$, $\chi^2_\text{IP} > 36$, and $\chi^2/\text{ndf} < 3$.

5.5.3 Models for ghosts tracks

The shape for the ProbNN$d(d)$ percentile of ghost tracks is found from minimum bias simulation. Though it is known that this simulation is not a perfect representation of the data, there is no good alternative for finding ghosts in data, as they cannot be selected in the same way as charged particles from decays.

Previous studies of ghosts at the LHCb experiment have considered the differences between ghosts in data and simulation, though they focus on the difference in the abundance of ghosts in each, rather than the misrepresentation of the kinematics or distributions of PID parameters of ghosts in simulation [65]. It was found that ghosts are more abundant in data than in simulation, by as much as a factor of two. In the measurement of deuterons using a template fit to data in ProbNN$d(d)$ percentile, the shapes of the components entering the fit must be well known, and the yield of each component is left free in the fit. Therefore, in the case of this analysis, the shape of ghosts in ProbNNd (d) percentile is of more importance than knowledge of the abundance of ghosts. As such, the shape in ProbNN$d(d)$ percentile of ghosts in simulation is compared to a best-estimate of ghost tracks in data. In the fit, the shape from simulation is used as the nominal shape, with the shape difference in simulation compared to data as the permitted shape variation.

The systematic shape difference between data and simulation for ghosts is found using pions from the PID Calib package data samples that are ghost-like, and comparing these to ghost-like pions in simulation, where ghost-like tracks are defined as those with ProbNNghost > 0.5. Similar to ProbNNd for deuterons, ProbNNghost is a neural net discriminator, trained to discriminate between ghost tracks and real tracks using PID and tracking information. This information includes variables from both RICH detectors, as well as from the calorimeters and the muon stations. This method allows for a systematic to be applied that accounts for differences between ghosts in simulation and ghosts in data. This extraction assumes that the ghost-like pions are a good representation of all ghosts, which could be problematic if ghost-like pions are very different to real ghosts, or if the ghost-like pions do not provide coverage at the high values of ProbNNd.

Comparisons in simulation of ghost and real tracks are shown in Fig. 5.25. Here entries labelled as 'ghostlike tracks' are protons, kaons and pions with ProbNNghost $>$ 0.5, and 'ghostlike ghosts' and 'all ghosts' are true ghosts in simulation with ProbNNghost $>$ 0.5 and for all ProbNNghost values, respectively. The distributions have been normalised to unity, such that their shapes can be compared to each other. In all momentum

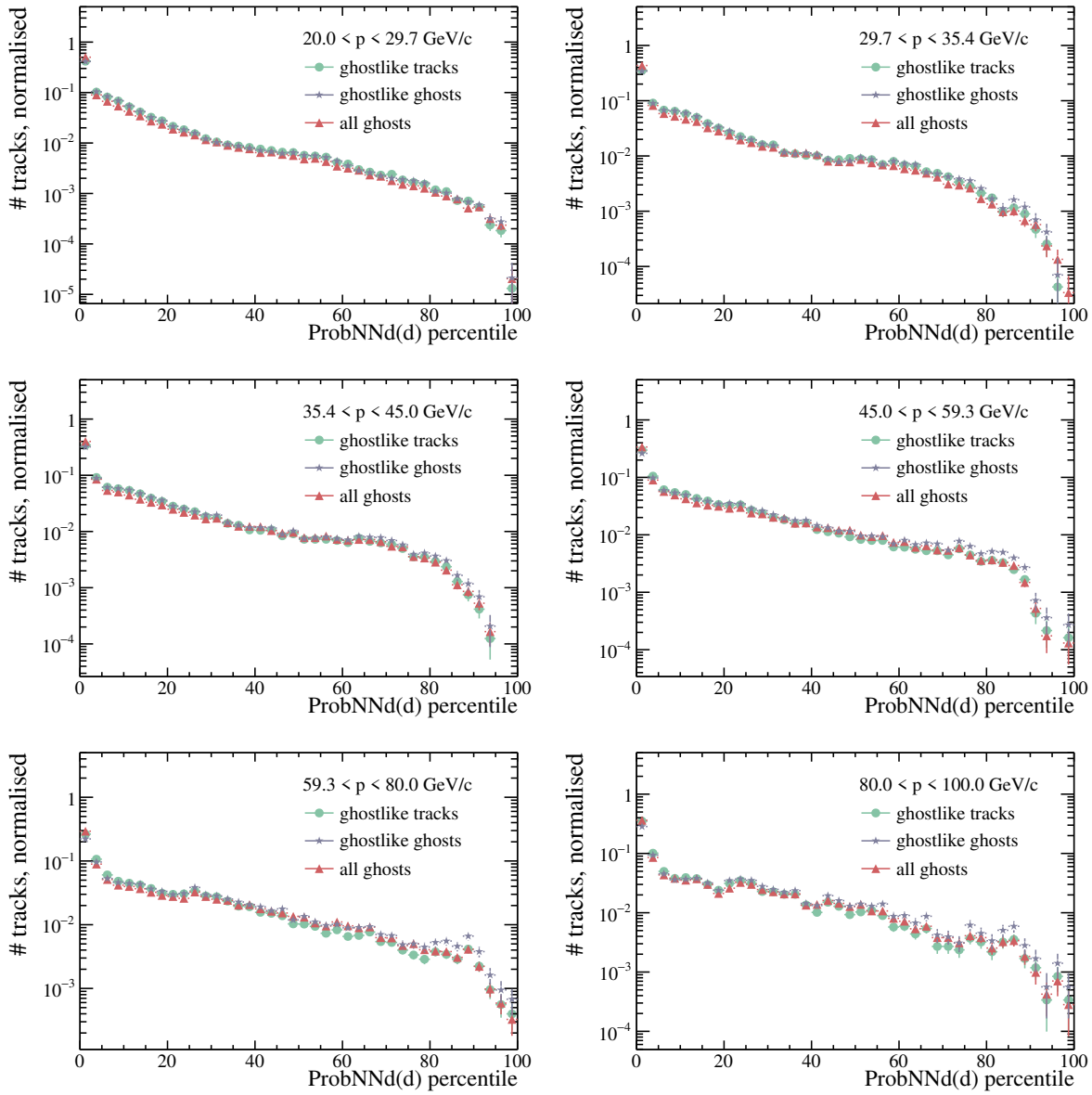


Figure 5.25: Comparison of ghosts and charged particles in ProbNN$d(d)$ percentile from simulation, where 'ghostlike' denotes tracks passing a selection of ProbNNghost > 0.5, and tracks comprise pions, protons and kaon. All distributions normalised to unity.

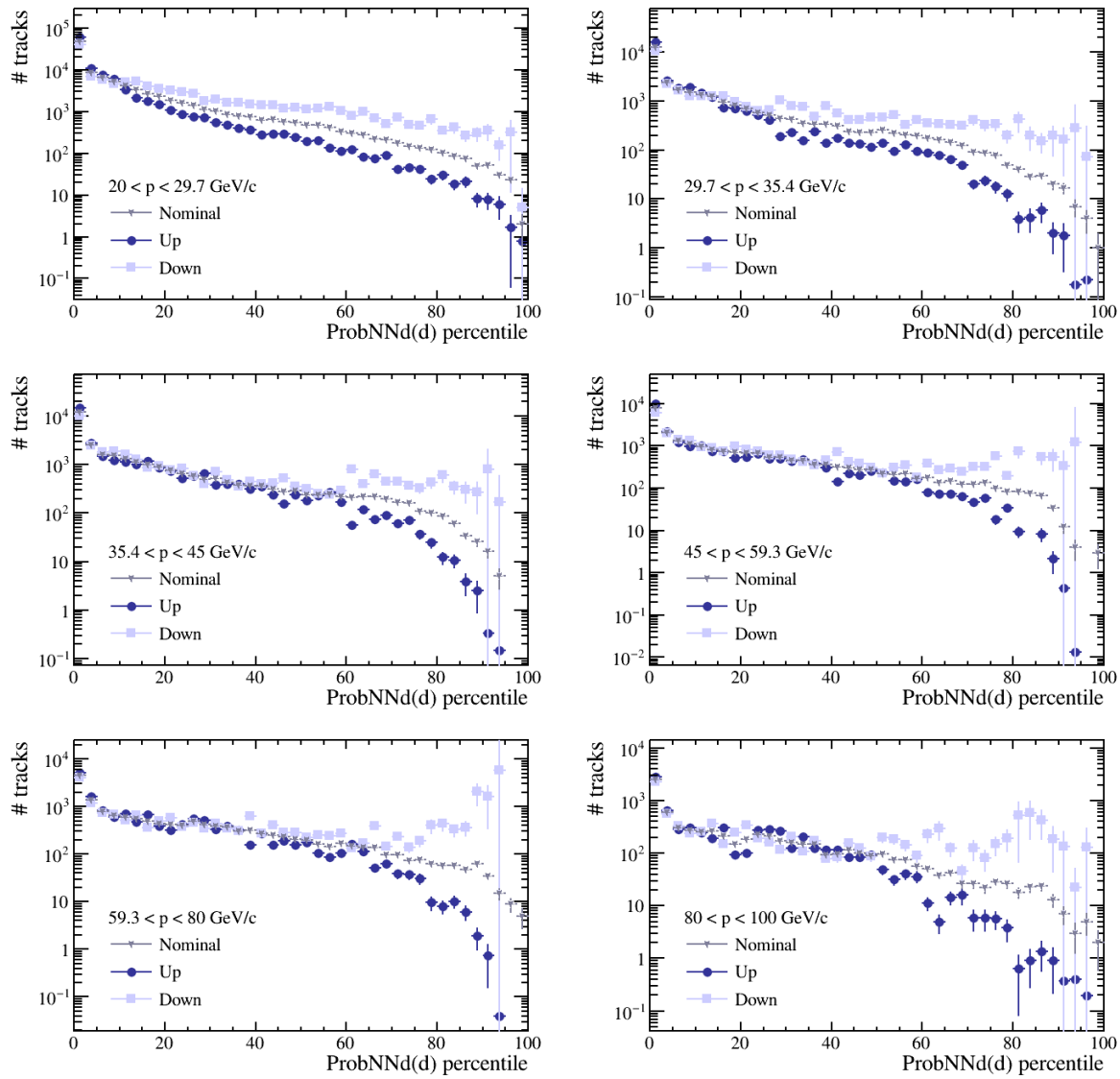


Figure 5.26: Ghost systematic shape variation in ProbNN$d(d)$ percentile in each momentum bin.

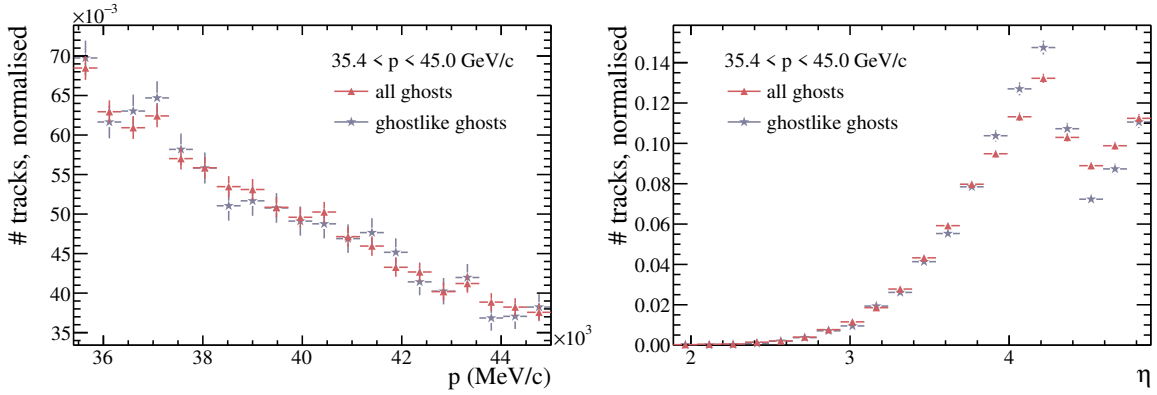


Figure 5.27: Momentum and pseudorapidity distributions of ghosts, and ghostlike ghosts for the analysis momentum bin of $35.4 < p < 45.0\,\text{GeV}/c$ from minimum bias simulation. All distributions normalised to unity.

bins, the distributions for ghostlike ghosts and ghostlike tracks agree with each other. When compared to all ghosts, there is some disagreement; most notably at low values of ProbNN$d(d)$ percentile.

The agreement between ghosts and ghostlike tracks in simulation suggests that ghostlike tracks in data can be used as a proxy for ghosts in data. To estimate the difference in the shape of ghosts in simulation and data, the shape difference between ghostlike pions in simulation and data is found. This shape difference is used as the variation in the ghost shape that is allowed in the fit. The shape of ghosts in the template fit is allowed to vary according to this shape difference, which compensates for the fact that the ghost shape in the fit is taken from simulation instead of data. The ghost systematic shape variations are shown in Fig. 5.26 for each momentum bin.

The assignment of this allowed shape variation assumes that ghostlike ghosts are representative of all ghosts. The shapes of ghostlike ghosts and all ghosts in ProbNN$d(d)$ percentile are shown in Fig. 5.25, where they are in agreement. In minimum bias simulation, it is also true that the distributions of momentum and pseudorapidity agree with each other for ghostlike ghosts and all ghosts. Therefore the samples do not need to be weighted to account for kinematic differences when extrapolating the shape for ghostlike ghosts to all ghosts. In Fig. 5.27 the distributions in momentum and pseudorapidity for ghosts and ghostlike ghosts in the momentum bin $35.4 < p < 45.0\,\text{GeV}/c$ obtained from minimum bias simulation are shown. It can be seen that there is no shift in coverage when ghostlike ghosts are selected, compared to the coverage of all ghosts.

5.6 Fit validation

In order to evaluate the quality of the fit to data, pseudo-experiments, termed 'toys', are performed. For these toys, pseudo-datasets with known constituent yields are generated and then fitted with the nominal model using the RooFit [85] and HistFactory packages [83].

For each toy, the yields generated for each of the fit components are allowed to fluctuate, following a Poisson distribution of mean T_X, where $X = \pi$, K, p, d, g, for $g =$ ghosts. The component shapes used to generate the toys are fixed to the nominal shapes. Each toy is then fitted with the nominal model, and the only floating parameters are the yields of the different components, and the parameters controlling the systematic shape variations. The results of these pseudo-experiments provide a way to estimate the precision that can be achieved in measuring the deuteron yield.

The components included in the pseudo-data are deuterons, pions, protons, kaons, muons and ghosts. Electrons are not added explicitly, because the shapes taken from calibration data are statistically limited, such that unrealistic spikes appear in the toys if they are included. Muons are included in the pseudo-data, but the muon yield is absorbed by the pion shape when the fit is performed, because the muon shapes are statistically limited, and the pion and muon template shapes are similar, due to the similarity of their masses.

The pseudo-data are generated assuming a relative yield between each of the components found from simulation, and with a total integral of the number of tracks that are found in the data. The number of tracks found in data in each momentum bin is listed in Tab. 5.6 to the nearest million. The percentages of charged particles made up of each track species, taken from minimum bias simulation in bins of momentum are listed in Tab. 5.7. The ratios of charged particles to the number of pion tracks in each momentum bin are shown in Fig. 5.28. The K:π and p:π ratios are almost flat across the momentum range, whilst the other charged particles exhibit a decrease in abundance, compared to pions, as the momentum increases. However, the proportion of deuterons increases compared to pions, with increasing momentum. The number of deuterons to pions, d:π, is $\approx 10^{-4}$ for all momenta.

The results of these pseudo-experiments assume that the modelling of deuteron production in the minimum bias simulation with the Cross-section model for deuteron formation is accurate [43]. The parameter used in the modelling, taken from measurements

p (GeV/c)	Tracks (millions)
20.0 – 29.7	48
29.7 – 35.4	15
35.4 – 45.0	15
45.0 – 59.3	11
59.3 – 80.0	6
80.0 – 100.0	2

Table 5.6: Number of tracks in 2018 minimum bias data in each momentum bin.

p (GeV/c)	20.0 – 29.7	29.7 – 35.4	35.4 – 45.0	45.0 – 59.3	59.3 – 80.0	80.0 – 100.0
Particle	Percentage of tracks (%)					
π	69.90	70.26	70.94	71.84	72.75	73.73
d	0.01	0.02	0.02	0.02	0.02	0.02
p	10.66	10.94	11.02	11.06	10.94	10.53
K	13.62	13.61	13.25	12.74	12.19	11.63
μ	0.15	0.12	0.11	0.10	0.09	0.08
e	1.81	1.44	1.24	0.99	0.79	0.66
g	3.95	3.60	3.43	3.24	3.22	3.37

Table 5.7: Percentages of tracks made up of each charged particle species in bins of momentum from minimum bias simulation.

of deuterons at the ALICE detector in pp collisions with $\sqrt{s} = 7\,\text{TeV}$, determines the rate of deuteron production in the model [1]. This rate of deuteron production was used to find T_d for each momentum bin in the pseudo-data.

As mentioned in section 5.5.3, the rate of ghost tracks in simulation is not the same as the rate of ghost tracks in data. Though the simulation performs a very complex imitation of the detector, it does not model the detector performance perfectly, nor the noise and occupancy that lead to ghost tracks. The ghost rate in simulation may be as much as a factor two smaller than the data ghost rate [65].

The results of pseudo-experiments in the momentum bin $35.4 < p < 45\,\text{GeV}/c$ are given in Tab. 5.8. The mean yield, N_X, is the number of each component returned by the fit, whilst the number included in the pseudo-datasets is T_X. The mean error on the yield returned by the fit is σ_X, and the pull, P_X is defined as $(N_X - T_X)/\sigma_X$. The variation in

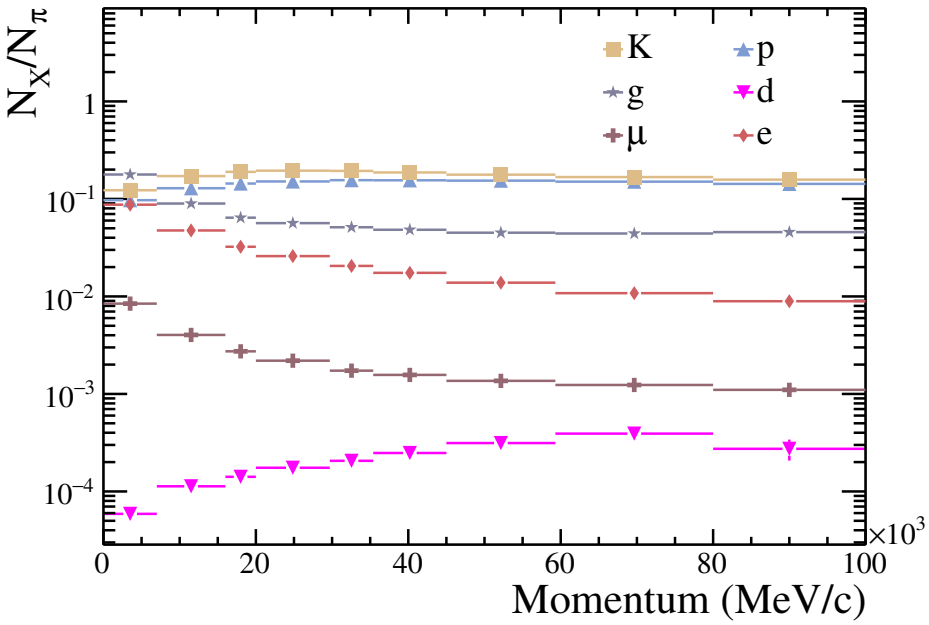


Figure 5.28: Ratios of charged particles to charged pions in minimum bias simulation in bins of momentum.

the true values is found by $\sqrt{T_X}$, which corresponds to the expected variation in each of the yields when the pseudo-datasets are generated. The floating parameters in the fit, N_X and α_X, were unconstrained.

There is an interplay between the background components, due to the degeneracy between their shapes in ProbNN$d(d)$ percentile. The shapes for pions, kaons, protons and ghosts all peak strongly at low ProbNN$d(d)$ percentile values, such that the fit will have difficulty in separating them. The similarity in the shapes of the background components can be seen in Fig. 5.29.

In Fig. 5.30 the charged particle yields, N_d, N_π, N_p, N_K and N_g, returned by the fits to pseudo-data with momentum $35.4 < p < 45.0\,\text{GeV}/c$, are shown as two-dimensional histograms. The trends in these plots show the interplay between the components entering the fits. The background yields all show a significant correlation, whilst the comparison of yields returned for deuterons and pions shows no correlation. The N_g and N_K are both anticorrelated with N_π, whilst N_p is correlated with N_π. The anticorrelation shows that the pion template is taking events that the kaon and ghost shapes would otherwise contain. The opposite occurs for protons, however, which could suggest that when the pion template absorbs entries from either kaons or ghosts, a space is left in

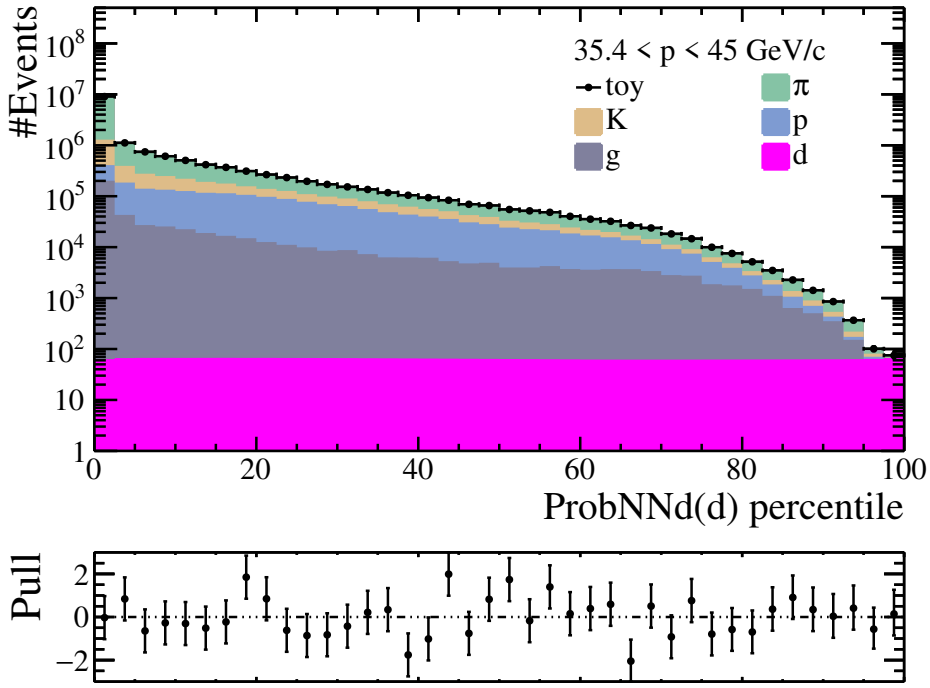


Figure 5.29: Projection of fit to pseudo-data of deuteron, pion, kaon, proton and ghost shapes for the momentum bin $35.4 < p < 45.0\,\text{GeV}/c$.

ProbNN$d(d)$ percentile for the proton shape. The same behaviour is seen in the yields for other momentum bins, where the deuteron signal is indepentent of the background yields, but the background components show some correlation. Therefore, the individual yields that the fit finds for the background components, and their interplay, do not prevent the fit finding an unbiased value for the deuteron signal.

The values for the systematic shape variation, α_X, in Tab. 5.8 are the result of the interpolation for each component between the nominal shape, and the upper and lower shape variations that are provided in the fit. In the case of this measurement, the variations allow for differences between data and simulated data for deuterons and ghosts, and for the differences between the data samples used in the calibration data and minimum bias data for pions, kaons and protons. For pions, there is an additional shape parameter, $\alpha_{\pi,\mu,e}$, which gives the amount of shift in the pion template due to the systematic for the muon and electron shapes in the pion shape. In all cases, α_X is consistent with zero, meaning that, on average, the fit does not shift the templates far from the nominal shapes.

Parameter		Fit value
Deuteron	N_d	$2.72 \times 10^{3} \pm 3.03 \times 10^{2}$
	σ_d	$3.16 \times 10^{2} \pm 4.14 \times 10^{1}$
	P_d	$1.06 \times 10^{-1} \pm 9.50 \times 10^{-1}$
	T_d	$2.67 \times 10^{3} \pm 5.17 \times 10^{1}$
	α_d	$2.07 \times 10^{-2} \pm 2.77 \times 10^{-1}$
Pion	N_π	$1.10 \times 10^{7} \pm 2.88 \times 10^{4}$
	σ_π	$3.46 \times 10^{4} \pm 1.40 \times 10^{4}$
	P_π	$-1.19 \times 10^{-2} \pm 9.37 \times 10^{-1}$
	T_π	$1.10 \times 10^{7} \pm 3.31 \times 10^{3}$
	α_π	$2.01 \times 10^{-4} \pm 6.26 \times 10^{-3}$
	$\alpha_{\pi,\mu,e}$	$9.81 \times 10^{-2} \pm 2.43 \times 10^{-2}$
Kaon	N_K	$2.01 \times 10^{6} \pm 3.43 \times 10^{4}$
	σ_K	$4.17 \times 10^{4} \pm 2.16 \times 10^{4}$
	P_K	$5.50 \times 10^{-2} \pm 9.49 \times 10^{-1}$
	T_K	$2.01 \times 10^{6} \pm 1.42 \times 10^{3}$
	α_K	$3.95 \times 10^{-4} \pm 8.99 \times 10^{-3}$
Proton	N_p	$1.67 \times 10^{6} \pm 1.63 \times 10^{4}$
	σ_p	$1.83 \times 10^{4} \pm 5.88 \times 10^{3}$
	P_p	$-3.52 \times 10^{-2} \pm 9.55 \times 10^{-1}$
	T_p	$1.67 \times 10^{6} \pm 1.29 \times 10^{3}$
	α_p	$2.00 \times 10^{-4} \pm 1.19 \times 10^{-2}$
Ghost	N_g	$5.19 \times 10^{5} \pm 1.62 \times 10^{4}$
	σ_g	$1.86 \times 10^{4} \pm 8.23 \times 10^{3}$
	P_g	$-4.50 \times 10^{-2} \pm 9.25 \times 10^{-1}$
	T_g	$5.20 \times 10^{5} \pm 7.21 \times 10^{2}$
	α_g	$2.74 \times 10^{-4} \pm 5.96 \times 10^{-3}$

Table 5.8: Results from pseudo-experiments for template fits in the momentum bin $35.4 < p < 45.0\,\text{GeV}/c$, with deuteron, pion, kaon, proton and ghost shapes in the fit. The mean value from the pseudo-experiments, and the spread in the mean, are given for each of the yields, N_X, the errors on the yields, σ_X, the pulls in the yield P_X, the generated numbers in the pseudo-data, T_X, and the shape variation parameters, α_X.

The deuteron shape is flat and dominates the bins at high ProbNNd (d) percentile values, as seen in Fig. 5.29, providing the fit with the ability to find the correct deuteron yield.

Momentum (GeV/c)	**Parameter**	**Fit value**
$20.0 < p < 29.7$	N_d	6162.6 ± 1722.2
	σ_d	2636.0 ± 1391.0
	P_d	-0.023 ± 0.626
	T_d	6064.3 ± 77.9
	α_d	0.061 ± 0.464
	Sig.	$2.3\ \sigma$
$29.7 < p < 35.4$	N_d	2593.4 ± 1175.5
	σ_d	1341.2 ± 542.7
	P_d	0.122 ± 0.625
	T_d	2224.5 ± 47.2
	α_d	0.117 ± 0.451
	Sig.	$1.9\ \sigma$
$35.4 < p < 45.0$	N_d	2680.9 ± 301.5
	σ_d	321.3 ± 8.4
	P_d	-0.091 ± 0.946
	T_d	2705.1 ± 52.0
	α_d	-0.001 ± 0.289
	Sig.	$8.4\ \sigma$
$45.0 < p < 59.3$	N_d	2510.9 ± 259.8
	σ_d	252.1 ± 5.9
	P_d	-0.061 ± 1.033
	T_d	2522.4 ± 50.2
	α_d	-0.013 ± 0.205
	Sig.	$10.0\ \sigma$
$59.3 < p < 80.0$	N_d	1864.9 ± 425.7
	σ_d	430.7 ± 23.0
	P_d	-0.080 ± 1.004
	T_d	1890.2 ± 43.5
	α_d	-0.022 ± 0.3
	Sig.	$4.3\ \sigma$
$80.0 < p < 100.0$	N_d	560.3 ± 255.8
	σ_d	272.7 ± 47.1
	P_d	0.072 ± 0.942
	T_d	516.5 ± 22.7
	α_d	-0.015 ± 0.411
	Sig.	$2.1\ \sigma$

Table 5.9: Deuteron values returned from fits to pseudo-experiments in each momentum bin, with their associated errors. Sig. is the signficance of the signal in each momentum bin, defined as N_d/σ_d.

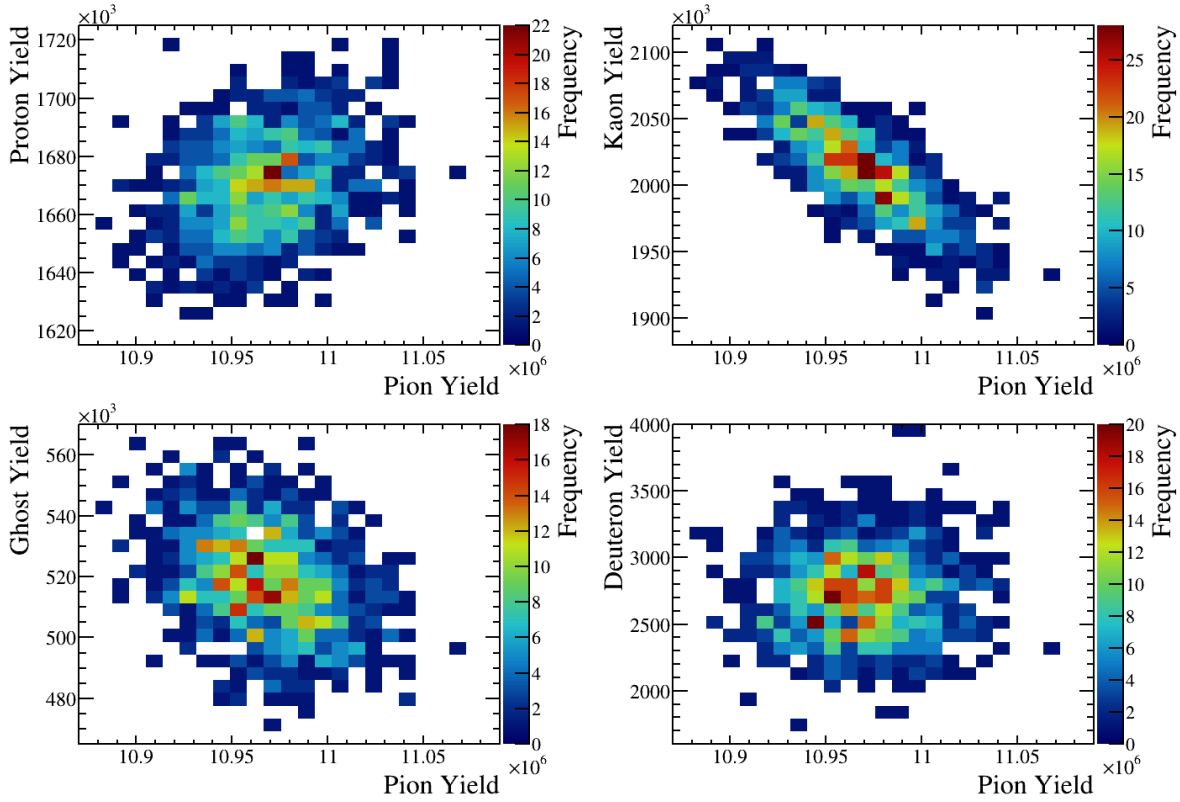


Figure 5.30: Two-dimensional plots showing the correlation between the yields of particles returned from pseudo-experiments for the momentum bin $35.4 < p < 45.0\,\text{GeV}/c$, for pion–proton (top left), pion–kaon (top right), pion–ghost (bottom left), and pion–deuteron (bottom right).

The deuteron signal results from pseudo-experiments for all of the momentum bins are listed in Tab. 5.9. For all momentum bins, P_d and α_d are consistent with zero, which suggests that there is no bias in the results of the fits. The RMS in pulls, P_d, for the first two momentum bins are, at $\sim$ 0.6, considerably below unity, which suggests that the errors the fit computes for the deuteron yield are conservative. The significance of the yields, denoted 'Sig.', is a representation of the sensitivity to the deuteron signal in each bin, according to the assumed true number of deuterons from the Cross-section model. In the first two momentum bins, the significances of the signals are low, in spite of the relatively high T_d values in these regions. The deuteron yields in these bins have large errors, likely owing to the fact that the deuteron is below its Cherenkov momentum thresholds, such that the particle discrimination using ProbNNd is not good enough here to isolate a deuteron signal. As momentum increases, the significance of the signal increases, before decreasing again in the highest momentum bin, where the statistics are lower. At high momenta, the deuteron and background Cherenkov signals are saturated,

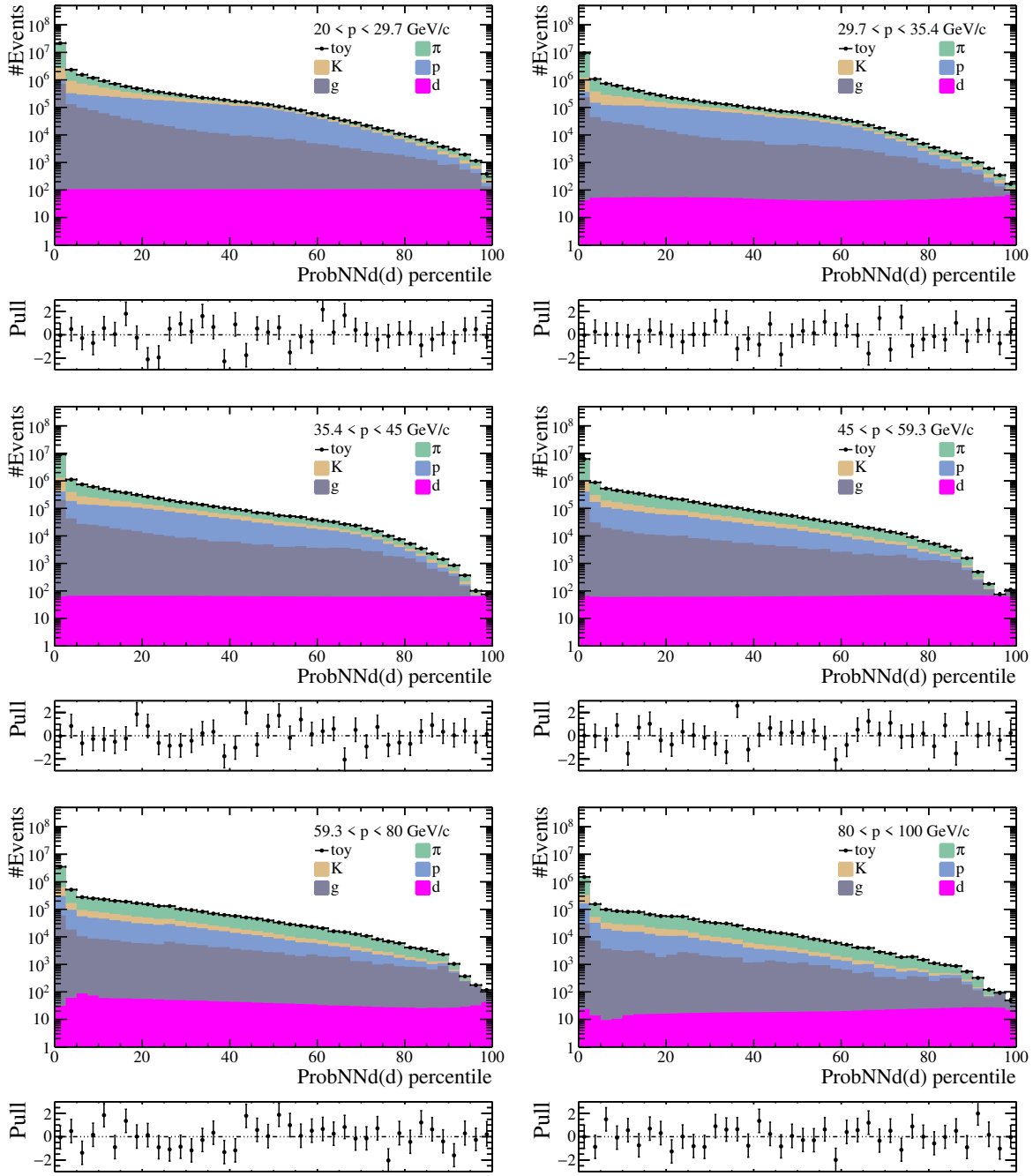


Figure 5.31: Fits to pseudo-data of deuteron, pion, kaon, proton and ghost shapes for each momentum bin.

so that all signals appear the same in the RICH algorithm, which decreases the particle discrimination. Examples of pseudo-experiments for all momentum bins are shown in Fig. 5.31.

Taking the results presented in Tab. 5.9 into consideration, it is clear that some momentum bins used in the measurement will likely not yield a significant signal, assuming that the models used to predict the deuteron yield in each momentum bin are accurate. This is due to lack of separation in ProbNN$d(d)$ percentile for signal and background in the first two momentum bins, and because of lack of statistics, coupled with saturating Cherenkov signals, in the highest momentum bin.

The three momentum bins, in the range $35.4 < p < 80.0\,\text{GeV}/c$, offer good separation of deuterons from other charged particles, using the rate of production in the deuteron simulation from ALICE measurements. In this momentum range, deuterons are above their Cherenkov momentum threshold in RICH 1, such that the discrimination power is improved.

5.7 Fits to the data

The final fits to data for each momentum bin are shown in Fig. 5.32. For each momentum bin, the templates, for the deuteron, pion, proton, kaon, and ghost components are plotted as a stack, alongside the data points. The data is the minimum bias data collected by the LHCb detector, with no selections applied to the tracks.

Disagreement between the data and the combination of component templates in the fits is visible in the pull distributions at the bottom of each of the plots in Fig. 5.32. For all but the final plot ($80.0 < p < 100.0\,\text{GeV}/c$), the pulls, giving the difference in the data and the fit for each bin, have a large variation. The units of the pulls for each bin are the number of standard deviations by which the data and fit disagree. Therefore, any value of pull greater than two indicates a significant disagreement. As a result of this disagreement, the χ^2/ndf, used as a measure of the goodness-of-fit, are in excess of one hundred, for all but the highest momentum bin.

For all momentum bins in Fig. 5.32, a steeply falling background of pions, protons, kaons, and ghost tracks is seen, above a roughly flat signal of deuterons. The relative amounts of each of the background tracks change significantly through the momentum bins. For the middle two momentum bins, $35.4 < p < 45.0\,\text{GeV}/c$ and $45.0 < p < 59.3\,\text{GeV}/c$, the relative amounts of background tracks look similar to what was simulated for the pseudo-experiments. In the other momentum bins, there appears to be variation in their relative amounts. For $29.7 < p < 35.4\,\text{GeV}/c$, there are comparatively few ghosts, and

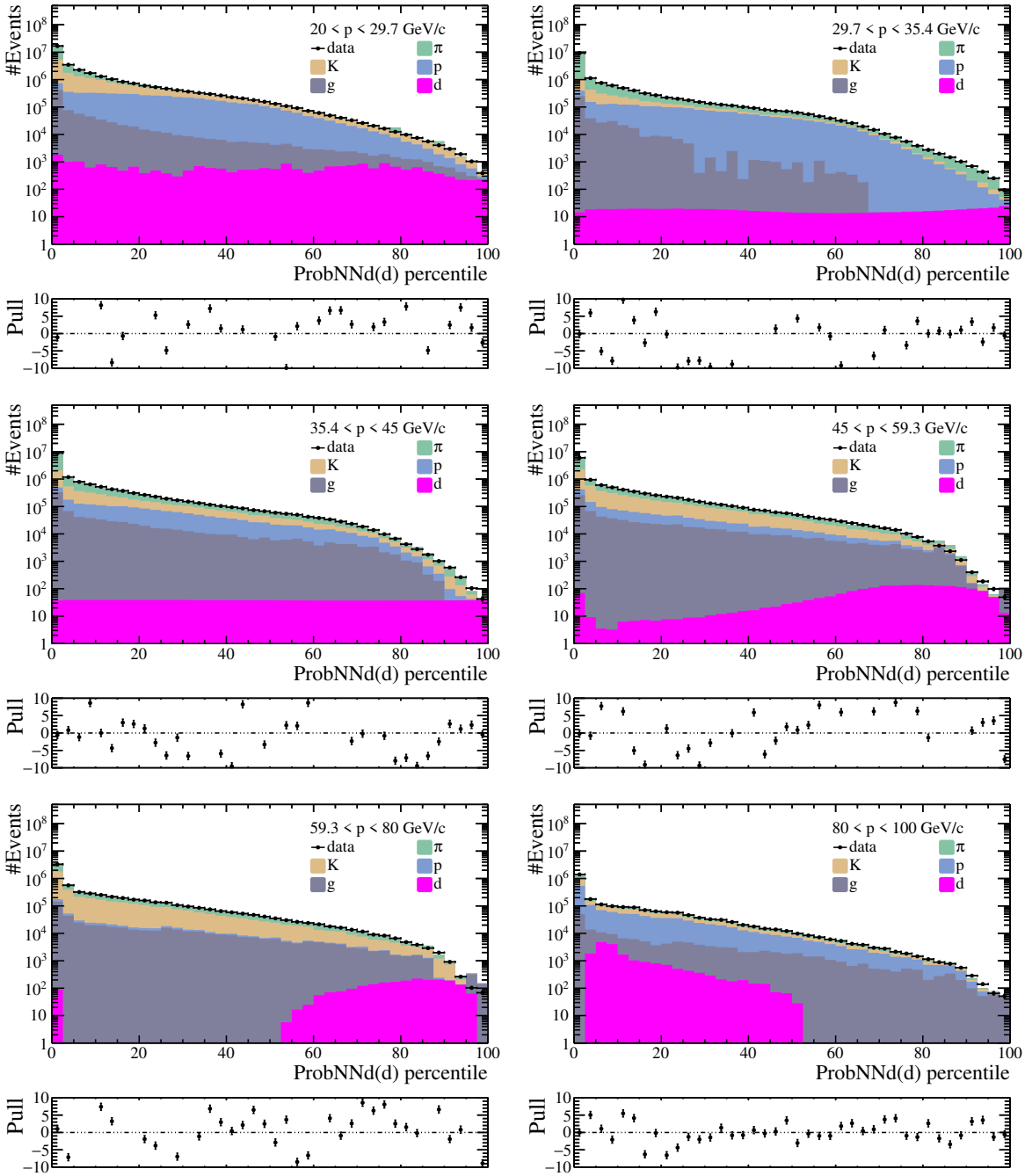


Figure 5.32: Fits to data for each momentum bin of the measurement.

in the plot for $80.0 < p < 100.0\,\text{GeV}/c$, the yield of pions is very low, compared to the other background tracks. This variation is attributed to the correlation between the background components. However, as seen in the pseudo-experiments in Fig. 5.30, the correlations between the background shapes does not bias the deuteron signal yield.

Momentum (GeV/c)	Parameter	Fit value
$20.0 < p < 29.7$	N_d	$23\,909.0 \pm 12\,900.0$
	Sig.	$1.9\ \sigma$
	α_d	2.048 ± 0.739
$29.7 < p < 35.4$	N_d	682.8 ± 315.0
	Sig.	$2.2\ \sigma$
	α_d	-0.442 ± 0.615
$35.4 < p < 45.0$	N_d	1517.3 ± 198.0
	Sig.	$7.7\ \sigma$
	α_d	-0.034 ± 0.244
$45.0 < p < 59.3$	N_d	2090.4 ± 203.0
	Sig.	$10.3\ \sigma$
	α_d	1.192 ± 0.304
$59.3 < p < 80.0$	N_d	1390.3 ± 194.0
	Sig.	$7.2\ \sigma$
	α_d	3.595 ± 0.572
$80.0 < p < 100.0$	N_d	$11\,936.0 \pm 1800.0$
	Sig.	$6.6\ \sigma$
	α_d	-3.717 ± 0.584

Table 5.10: Deuteron values returned from fits to data in each momentum bin. Sig. is the significance of the signal in each bin, defined as N_d/σ_d

Some shape variation from the flat, nominal deuteron shapes can be seen in the fit for every momentum bin, in Fig. 5.32. For the first three momentum bins, the deuteron shape remains fairly flat, but for $p > 45.0\,\text{GeV}/c$, there is significant curvature in the shapes. The values of the parameters, α_d, that give a measure of the amount the nominal shape has been allowed to vary are given in Tab. 5.10. According to the HistFactory package, $\alpha \pm 1.0$ corresponds to variations in the nominal shapes of $\pm\sigma$, for a given systematic shape variation. The values for α_d from the fit are mostly out of this range, such that the variations needed in the deuteron shapes for the fit to converge were greater than the expected variation. The shapes the fits found for deuterons are extreme cases of the deuteron systematic shape variations shown in Fig. 5.10, where $\alpha_d > 0$ corresponds to the 'up' shape, and $\alpha_d < 0$ corresponds to the 'down' shape in each momentum bin.

Though not easily visible in the plots of the fits (Fig. 5.32), there is also a lot of variation in the shapes of the background components in the fits to data. The values

Momentum (GeV/c)	**Parameter**	**Fit value**
$20.0 < p < 29.7$	α_π	$-1.65 \pm 4.30 \times 10^{-3}$
	$\alpha_{\pi,\mu,e}$	$-5.28 \times 10^{-1} \pm 9.74 \times 10^{-3}$
	α_p	$-4.11 \times 10^{-1} \pm 7.99 \times 10^{-3}$
	α_K	$1.83 \times 10^{-1} \pm 1.58 \times 10^{-3}$
	α_g	$1.07 \times 10^{-1} \pm 8.89 \times 10^{-2}$
$29.7 < p < 35.4$	α_π	$9.56 \times 10^{-2} \pm 6.50 \times 10^{-3}$
	$\alpha_{\pi,\mu,e}$	$9.74 \times 10^{-4} \pm 1.11 \times 10^{-2}$
	α_p	$-1.64 \times 10^{-2} \pm 6.04 \times 10^{-3}$
	α_K	$-4.64 \times 10^{-1} \pm 2.00 \times 10^{-2}$
	α_g	$1.63 \pm 1.82 \times 10^{-2}$
$35.4 < p < 45.0$	α_π	$2.83 \times 10^{-1} \pm 1.37 \times 10^{-2}$
	$\alpha_{\pi,\mu,e}$	$6.55 \times 10^{-2} \pm 4.75 \times 10^{-3}$
	α_p	$1.36 \times 10^{-1} \pm 1.73 \times 10^{-2}$
	α_K	$1.82 \times 10^{-1} \pm 3.00 \times 10^{-2}$
	α_g	$-2.02 \times 10^{-1} \pm 7.28$
$45.0 < p < 59.3$	α_π	$-2.02 \times 10^{-1} \pm 1.07 \times 10^{-2}$
	$\alpha_{\pi,\mu,e}$	$9.05 \times 10^{-1} \pm 5.11 \times 10^{-2}$
	α_p	$-3.13 \times 10^{-1} \pm 3.18 \times 10^{-2}$
	α_K	$-4.40 \times 10^{-2} \pm 8.55 \times 10^{-3}$
	α_g	$8.53 \times 10^{-2} \pm 5.14 \times 10^{-3}$
$59.3 < p < 80.0$	α_π	$1.59 \times 10^{-1} \pm 1.95 \times 10^{-2}$
	$\alpha_{\pi,\mu,e}$	$4.16 \times 10^{-1} \pm 7.72 \times 10^{-2}$
	α_p	$-1.16 \pm 4.41 \times 10^{-1}$
	α_K	$1.28 \times 10^{-1} \pm 4.21 \times 10^{-3}$
	α_g	$1.75 \times 10^{-1} \pm 1.97 \times 10^{-2}$
$80.0 < p < 100.0$	α_π	$4.61 \times 10^{-2} \pm 6.95 \times 10^{-3}$
	$\alpha_{\pi,\mu,e}$	$4.66 \times 10^{-2} \pm 1.36 \times 10^{-2}$
	α_p	$3.41 \times 10^{-2} \pm 7.46 \times 10^{-3}$
	α_K	4.53 ± 1.02
	α_g	$9.35 \times 10^{-2} \pm 2.75 \times 10^{-2}$

Table 5.11: Shape systematic parameters, α_X, for the background components in the fits to data.

of the shape parameters, α_X, are summarised in Tab. 5.11, which comprises all of the systematics involved in the fits. In all but the third and fourth momentum bins, there are values of $|\alpha_X| > 1.0$. This shows that the assigned alternative shapes, provided to HistFactory along with the nominal templates, did not provide sufficient variation for the fit to be able to use them within one standard deviation, $|\alpha_X| < 1.0$.

This combination of extreme values of systematic shape parameters, α, and poor goodness-of-fit, indicates that the fits have not been able to describe the data accurately, and therefore, the deuteron yields from the fits are not reliable.

Suggestions for improving the fit

The inability of the fit to find a significant deuteron yield in data, due to poor modelling of the shape in ProbNN$d(d)$ percentile, suggests that an oversight has been made in the fit method. This could be caused by a number of factors: performing the fit over a too large range of ProbNNd, including too many fit components, or using an unsuitable binning.

It is possible that the fit could be improved by limiting the range in ProbNNd over which it is performed. Removing the low regions in ProbNNd of the data distribution, and the component shapes, could improve performance of the fits. In the pull distribution for each fit in Fig. 5.32, the pull in the first bin in ProbNN$d(d)$ percentile is consistently in agreement with zero. Therefore, in every fit, the data and model components agree in the first bin, before diverging in higher ProbNN$d(d)$ percentile bins. The first bin is dominated by the background components, and if they are constrained by this bin, they will have limited flexibility in other bins, which may be of detriment to the goodness-of-fit and the deuteron signal.

On the other hand, extracting the deuteron yield after removing a portion of data in ProbNNd means that the deuteron efficiency for surviving this removal must be known, *i.e.* the fraction of deuterons that are not removed, before being able to compute the associated cross-section. This efficiency would have to be taken from simulation. In order to assess the uncertainty arising from using simulation to find this, instead of data, a similar strategy could be employed as was used to find the deuteron systematic shape variation, in section 5.5.1. The analagous selection could be made for protons in ProbNNp in simulation and in calibration data samples, and the difference between the two assigned as the systematic uncertainty on the deuteron efficiency.

Alternatively, the components included in the fit could be changed, to simplify the fit. It has been seen in pseudo-experiments, that the deuteron yield found by the fits was not biased when the fit included only pion, ghost track, and deuteron components, whilst the pseudo-data included the same particle constituents as the fits in section 5.6. Combining all the background components in to a single template, with systematic uncertainties corresponding to each of the components associated with it, could limit the number of model configurations compatible with the data, and help the minimisation process that the fit performs.

A different binning choice to the ProbNN$d(d)$ percentile binning used in this measurement could be used. It has not been assessed what the reliance of the binning in ProbNN$d(d)$ percentile has on the specific simulation sample that is used to define the bin boundaries. With forty bins in this fit, some bins correspond to a change in ProbNNd of as little as $\Delta\text{ProbNN}d = 0.01$. It is possible that a change in the sample of deuterons used to find these bins would shift the bin boundaries by a factor greater than $\Delta\text{ProbNN}d = 0.01$. If this is true when comparing deuterons in data and simulation, then the shape found in simulation will not be able to model data well. An additional systematic shape variation could also be added to each component to account for this.

Finally, foregoing measuring a deuteron production cross-section in some of the less sensitive momentum bins, the fit procedure could be improved using data, before applying it to the momentum bins where a significant yield is expected. Table 5.9 shows the fit results for pseudo-experiments, which were based on the expected deuteron yield for each momentum bin from minimum bias simulation with deuteron production included. It can be seen that the significance of the deuteron yield in the two lowest momentum bins is low – less than 2.5 σ in each. Therefore, it can not be expected that fits to data would perform better than this, so a significant deuteron yield cannot be found in either of these bins. Consequently, the momentum region $20.0 < p < 35.4\,\text{GeV}/c$ could be eliminated from the measurement, and instead these bins could be used to improve the fit strategy. The changes discussed above, such as different binning choices or fit components, could be trialled in fits in the two lowest momentum bins, in order to find the procedure that gives the most stable behaviour and best deuteron significance.

5.8 Deuteron production cross-section

Given a significant deuteron signal in each bin of momentum, the deuteron production cross-section as a function of momentum within the LHCb detector, $\sigma_d(p)$, can be found. In this section, the method to find $\sigma_d(p)$ is outlined, starting with the measurement of the inelastic pp cross-section in the LHCb detector, $\sigma_{pp\to X}$, which forms the basis of $\sigma_d(p)$.

The cross-section $\sigma_d(p)$, is found from the momentum-dependent form of Eq. 5.1. $\sigma_{pp\to X}$. It is given as

$$\sigma_d(p) = \sigma_{pp\to X}\frac{N_d(p)}{N_{evt}}. \tag{5.13}$$

where $N_d(p)$ is the number of deuterons observed per momentum region, the number of events in the corresponding dataset is N_{evt}, and the cross-section for those events is $\sigma_{pp\to X}$.

A measurement of the inelastic pp cross-section at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ was performed in 2016 at the LHCb experiment [65, 88]. Defining an inelastic event as one that contains at least one prompt, long-lived, charged particle within the LHCb detector acceptance, inelastic events are selected by requiring that there is a track with $p > 2.0\,\text{GeV}/c$ and $2 < \eta < 5$. The cross-section is found to be $\sigma_{acc} = 62.2 \pm 0.2 \pm 2.5_{lumi}\,\text{mb}$, where the first uncertainty is statistical, and the second is the systematic uncertainty, dominated by the luminosity measurement for the data, as well as smaller contributions due to the beam energy uncertainty, tracking efficiency uncertainties, and uncertainties due to ghost probability calibration between data and simulation. To eliminate background events, where the track detected is either a ghost track or a track from a secondary vertex, it is also required that the tracks have Ghost Probability < 0.3, and that the chi-squared per degree of freedom for the track fit is low, *i.e.* Track $\chi^2/\text{ndf} < 3$.

During each data-taking period at the LHCb experiment, the triggers used to select events are updated to meet the demands of the LHCb Collaboration. The inelastic cross-section measurement published in 2016 used data from 2015, when one set of triggers was in place, many of which were changed by 2018, which is when the data for the deuteron measurement was taken. In 2015, instead of minimum bias data being taken, the data collected without specific triggers was called 'no bias' data, and the events included in the data sample are all 'leading bunch collisions'. These are the collisions of the first

bunches in the trains of bunches, as opposed to events occurring consecutively after other bunch–bunch collisions in the train, or empty–bunch, bunch–empty, or empty–empty events; all of which are included in the 2018 minimum bias data sample. These leading bunch collisions also differ from other typical events in the LHC because they will not experience any *spillover* from previous events, given that the previous events will have been empty–empty crossings. Spillover is caused by the finite duration of analogue signals in electronics in the detector. If the amplitude of this protracted signal is not negligible, it will be recorded falsely in the successive event.

In 2018, the minimum bias data was taken using a randomised trigger that recorded events from any kind of bunch crossing, at a specified, average rate. Therefore, this data includes events from leading bunch crossings, as well as empty events and events from consecutive bunch crossings, which may have some spillover effects. Leading bunches have a higher intensity than other bunches in a train, and therefore a higher luminosity. In addition, the 2018 minimum bias data was taken with a 25 ns bunch-spacing, whereas the 2015 data was taken with a bunch-spacing of 50 ns. Due to these differences in data-taking, it is likely that the uncertainty on the luminosity for the 2018 data sample will be bigger than the uncertainty on the 2015 luminosity measurement.

Inelastic *pp* cross-section

In Ref. [65], the time-integrated inelastic *pp* cross-section is calculated as follows. The cross-section is the number of interactions, N_{int}, divided by the integrated luminosity of the data sample, $\mathcal{L}$,

$$\sigma_{pp\to X} = \frac{N_{int}}{\mathcal{L}} = \frac{N_{Bx}\mu}{\mathcal{L}}, \tag{5.14}$$

where the N_{Bx} is the number of bunch crossings, and μ is the average number of interactions per bunch crossing. Due to inefficiencies in the detector, μ cannot be found directly from measurement. The number of visible interactions follows a Poisson distribution, where the probability of n interactions occuring in a single bunch crossing is

$$P_n = e^{-\mu}\frac{\mu^n}{n!}. \tag{5.15}$$

The most straightforward way to find μ is from the probability of having a bunch crossing that has no visible tracks,

$$\mu = -\ln P_0 = -\ln\left(1 - \frac{N_{vis} - N_{bkg}}{\epsilon_{evt} N_{Bx}}\right), \tag{5.16}$$

where N_{vis} is the number of visible events, the efficiency for detecting the event is ϵ_{evt}, and the number of background events is N_{bkg}. Background events are defined as either events that do not come from a *pp* interaction (empty-empty, beam-empty, or empty-beam crossings), or where the visible track in the event did not originate from a *pp* collision, *i.e.* a ghost or a secondary track.

The efficiency, ϵ_{evt}, takes into account the efficiency for detecting visible events and is a function of the single track reconstruction efficiency, ϵ. It is assumed that all final state particles within the LHCb detector acceptance are reconstructed independently, with the same single track efficiency, ϵ. The single track reconstruction efficiency can be found from simulation, as

$$\epsilon = \epsilon^{Acc} \times \epsilon^{Reco} \times \epsilon^{Cuts} \times \rho^{Tracking}. \tag{5.17}$$

where ϵ^{Acc} is the acceptance efficiency, defined as the fraction of signal particles that reach the end of the tracking stations, over total signal; ϵ^{Reco} is the reconstruction efficiency, defined as the number of tracks surviving the track selections of Ghost Probability < 0.3 divided the total number of long tracks, found from simulation; ϵ^{Cuts} is the efficiency of offline selection cuts, and is the ratio of reconstructed tracks passing all cuts in an analysis, and the reconstructed tracks passing Ghost Probability < 0.3; and $\rho^{Tracking}$ is the data–MC correction factor, that is provided by the tracking group in the LHCb Collaboration, and is the average of the ratio of reconstruction efficiencies in data and simulation. In the case of the deuteron production cross-section measurement, no cuts are applied to the tracks, so $\epsilon^{Cuts} = 1.0$.

The number of background events is estimated from fractions of visible events originating from beam-empty (be), empty-beam (eb), and empty-empty (ee) bunch crossings. The fraction of background interactions in the sample is

$$f_{bkg} = q_0^{bb}\left(f_{bkg}^{be} + f_{bkg}^{eb} - f_{bkg}^{ee}\right), \tag{5.18}$$

where q_0^{bb} is the probability of finding an event from a beam-beam interaction with no tracks detected, and f_{bkg}^{ij} are the fractions of visible be, eb, and ee events. The rates

of additional background events from ghost or secondary tracks are estimated using simulation.

Computing deuteron production cross-section

To be able to use the inelastic pp cross-section summarised above, a similar treatment of collisions must take place in order to correctly compute the deuteron production cross-section. Restating Eq. 5.13,

$$\sigma_d(p) = \sigma_{pp\to X} \frac{N_d(p)}{N_{evt}}, \tag{5.19}$$

where $N_d(p)$ is found from the measurement in section 5.7, $\sigma_{pp\to X}$ is the result in Ref. [65], and N_{evt} should be found in a similar way to what is presented above for the $\sigma_{pp\to X}$ measurement, $N_{evt} = N_{Bx}\mu$.

Chapter 6

Conclusions and future outlook

6.1 Conclusions

Template fits to data in the variable ProbNN$d(d)$ percentile are performed in order to measure the deuteron yield at the LHCb experiment as a function of momentum, with a view to computing the deuteron production cross-section in pp collisions at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. The fit was found not to be sensitive to a deuteron signal, due to the template shapes for the components in the fit not being able to model the data accurately, which resulted in poor fit quality.

The deuteron signal was modelled using simulation, whilst the background components of pions, kaons, proton, muons and electrons were modelled using calibration samples from data. Ghost tracks were modelled using simulation. A small signal of deuterons against a large background of other charged particles was expected, based on simulation of deuteron production models. Pseudo-experiments for fit validation suggested that the ProbNN$d(d)$ percentile parameter the fit was performed in would provide ample discrimination, assuming that the components were modelled accurately. However, the expected performance was not achieved when fitting the data, suggesting that there is some mismodelling of the signal and background components.

Whilst no measurement of deuteron production has been made in data taken during 2018 at the LHCb experiment, it does not preclude other measurements of deuterons, either using different methods, or in different datasets.

6.2 Future deuteron measurements at LHCb

The method presented in this thesis, using ProbNN$d(d)$ percentile as the discriminating variable for a template fit in data to extract a deuteron signal, can be further developed, in order to perform a measurement using other data. The deuteron hypothesis was also included in the RICH algorithm in data taken during 2016 and 2017, so these datasets can be used.

In addition to this, other deuteron measurements at the LHCb detector can be made, such as searches for deuterons in b hadron decays, or searches for deuterons in other collision data, such as pHe [89].

Bibliography

[1] ALICE, J. Adam *et al.*, *Production of light nuclei and anti-nuclei in pp and Pb-Pb collisions at energies available at the CERN Large Hadron Collider*, Phys. Rev. **C93** (2016) 024917, arXiv:1506.08951.

[2] T. Aramaki *et al.*, *Review of the theoretical and experimental status of dark matter identification with cosmic-ray antideuterons*, Phys. Rept. **618** (2016) 1, arXiv:1505.07785.

[3] AMS, L. Accardo *et al.*, *High Statistics Measurement of the Positron Fraction in Primary Cosmic Rays of 0.5–500 GeV with the Alpha Magnetic Spectrometer on the International Space Station*, Phys. Rev. Lett. **113** (2014) 121101.

[4] Particle Data Group, M. Tanabashi *et al.*, *Review of particle physics*, Phys. Rev. D **98** (2018) 030001.

[5] Super-Kamiokande, Y. Fukuda *et al.*, *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett. **81** (1998) 1562, arXiv:hep-ex/9807003.

[6] D. Griffiths, *Introduction to elementary particles*, Wiley-VCH, Weinheim, Germany, 2008.

[7] M. Thomson, *Modern particle physics*, Cambridge University Press, New York, 2013.

[8] P. D. B. Collins and A. D. Martin, *Hadron Reaction Mechanisms*, Rept. Prog. Phys. **45** (1982) 335.

[9] P. G. Ratcliffe, *Introduction to Elementary Particle Phenomenology*, IOPP, 2014.

[10] J. D. Bjorken and E. A. Paschos, *Inelastic Electron Proton and gamma Proton Scattering, and the Structure of the Nucleon*, Phys. Rev. **185** (1969) 1975.

[11] H. Abramowicz *et al.*, *Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data*, Eur. Phys. J. **C75** (2015)

580, arXiv:1506.06042.

[12] G. Altarelli, *Collider Physics within the Standard Model: a Primer*, arXiv:1303.2842.

[13] J. L. Feng, *Dark Matter Candidates from Particle Physics and Methods of Detection*, Ann. Rev. Astron. Astrophys. **48** (2010) 495, arXiv:1003.0904.

[14] K. Garrett and G. Duda, *Dark Matter: A Primer*, Adv. Astron. **2011** (2011) 968283, arXiv:1006.2483.

[15] J. H. Oort, *The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems*, Bull. Astron. Inst. Netherlands **6** (1932) 249.

[16] A. Einstein, *Die Grundlage der allgeminen Relativitätstheorie*, Annalen der Physik **354** (1916) 769.

[17] D. Walsh, R. F. Carswell, and R. J. Weymann, *0957 + 561 A, B - Twin quasistellar objects or gravitational lens*, Nature **279** (1979) 381.

[18] MACHO, C. Alcock *et al.*, *The MACHO project: Microlensing results from 5.7 years of LMC observations*, Astrophys. J. **542** (2000) 281, arXiv:astro-ph/0001272.

[19] WMAP, N. Jarosik *et al.*, *Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Sky Maps, Systematic Errors, and Basic Results*, Astrophys. J. Suppl. **192** (2011) 14, arXiv:1001.4744.

[20] M. Markevitch *et al.*, *A Textbook example of a bow shock in the merging galaxy cluster 1E0657-56*, Astrophys. J. **567** (2002) L27, arXiv:astro-ph/0110468.

[21] M. Markevitch *et al.*, *Direct constraints on the dark matter self-interaction cross-section from the merging galaxy cluster 1E0657-56*, Astrophys. J. **606** (2004) 819, arXiv:astro-ph/0309303.

[22] D. Clowe, A. Gonzalez, and M. Markevitch, *Weak lensing mass reconstruction of the interacting cluster 1E0657-558: Direct evidence for the existence of dark matter*, Astrophys. J. **604** (2004) 596, arXiv:astro-ph/0312273.

[23] D. Bauer *et al.*, *Dark Matter in the Coming Decade: Complementary Paths to Discovery and Beyond*, Phys. Dark Univ. **7-8** (2015) 16, arXiv:1305.1605.

[24] J. Carr, G. Lamanna, and J. Lavalle, *Indirect detection of dark matter*, Rept. Prog.

Phys. **69** (2006) 2475.

[25] T. Marrodán Undagoitia and L. Rauch, *Dark matter direct-detection experiments*, J. Phys. **G43** (2016) 013001, arXiv:1509.08767.

[26] D. S. Akerib *et al.*, *Projected WIMP Sensitivity of the LUX-ZEPLIN (LZ) Dark Matter Experiment*, FERMILAB-PUB-18-054-AE-PPD, 2018.

[27] P. J. Fox, R. Harnik, J. Kopp, and Y. Tsai, *Missing Energy Signatures of Dark Matter at the LHC*, Phys. Rev. **D85** (2012) 056011, arXiv:1109.4398.

[28] CMS, A. M. Sirunyan *et al.*, *Search for dark matter in events with energetic, hadronically decaying top quarks and missing transverse momentum at* $\sqrt{s} = 13$ *TeV*, JHEP **06** (2018) 027, arXiv:1801.08427.

[29] ATLAS, M. Aaboud *et al.*, *Search for new phenomena in final states with an energetic jet and large missing transverse momentum in pp collisions at* $\sqrt{s} = 13$ *TeV using the ATLAS detector*, Phys. Rev. **D94** (2016) 032005, arXiv:1604.07773.

[30] D. M. Gomez Coral, *Deuteron and antideuteron production in galactic cosmic-rays.*, PhD thesis, UNAM, 2019.

[31] M. Ackermann *et al.*, *Searching for Dark Matter Annihilation from Milky Way Dwarf Spheroidal Galaxies with Six Years of Fermi-LAT Data*, Physical Review Letters **115** (2015) 1, arXiv:1503.02641.

[32] O. Adriani *et al.*, *Observation of an anomalous positron abundance in the cosmic radiation*, Nature **458** (2009) 607, 0810.4995v.

[33] J. Kopp, *Constraints on dark matter annihilation from AMS-02 results*, Physical Review D **88** (2013) 1, arXiv:1304.1184.

[34] I. Cholis and D. Hooper, *Dark Matter and Pulsar Origins of the Rising Cosmic Ray Positron Fraction in Light of New Data From AMS*, Phys. Rev. **D88** (2013) 023013, arXiv:1304.1840.

[35] F. Donato, N. Fornengo, and P. Salati, *Anti-deuterons as a signature of supersymmetric dark matter*, Phys. Rev. **D62** (2000) 043003, arXiv:hep-ph/9904481.

[36] P. von Doetinchem *et al.*, *Status of cosmic-ray antideuteron searches*, PoS **ICRC2015** (2016) 1218, arXiv:1507.02712.

[37] A. Ibarra and S. Wild, *Prospects of antideuteron detection from dark matter annihi-*

lations or decays at AMS-02 and GAPS, JCAP **1302** (2013) 021, arXiv:1209.5539.

[38] R. A. Ong *et al.*, *The GAPS Experiment to Search for Dark Matter using Low-energy Antimatter*, PoS **ICRC2017** (2018) 914, arXiv:1710.00452.

[39] D.-M. Gomez-Coral *et al.*, *Deuteron and Antideuteron Production Simulation in Cosmic-Ray Interactions*, Phys. Rev. **D98** (2018) 023012, arXiv:1806.09303.

[40] A. Schwarzschild and C. Zupancic, *Production of Tritons, Deuterons, Nucleons, and Mesons by 30-GeV Protons on A-1, Be, and Fe Targets*, Phys. Rev. **129** (1963) 854.

[41] J. I. Kapusta, *Mechanisms for deuteron production in relativistic nuclear collisions*, Phys. Rev. **C21** (1980) 1301.

[42] A. Ibarra and S. Wild, *Determination of the Cosmic Antideuteron Flux in a Monte Carlo approach*, Phys. Rev. **D88** (2013) 023014, arXiv:1301.3820.

[43] L. A. Dal and A. R. Raklev, *Alternative formation model for antideuterons from dark matter*, Phys. Rev. **D91** (2015) 123536, arXiv:1504.07242.

[44] S. T. Butler and C. A. Pearson, *Deuterons from High-Energy Proton Bombardment of Matter*, Phys. Rev. **129** (1963) 836.

[45] J. Bystricky *et al.*, *Energy dependence of nucleon-nucleon inelastic total cross-sections*, .

[46] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820.

[47] S. Miglioranzi *et al.*, *The LHCb Simulation Application, Gauss: Design, Evolution and Experience*, LHCb-PROC-2011-006, 2011.

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

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

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

[51] M. Buescher, A. A. Sibirtsev, and K. Sistemich, *Fast deuteron production in proton nucleus interactions*, Z. Phys. **A350** (1994) 161.

[52] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.

[53] J. Wenninger, *Operation and Configuration of the LHC in Run 2*, CERN-ACC-NOTE-2019-0007, 2019.

[54] ATLAS, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[55] CMS, S. Chatrchyan *et al.*, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004.

[56] ALICE, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

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

[58] LHCb, R. Aaij *et al.*, *LHCb Detector Performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, arXiv:1412.6352.

[59] T. Humair, *Testing lepton universality in penguin decays of beauty mesons using the LHCb detector*, PhD thesis, Imperial College London, 2019, Presented 02 May 2019.

[60] R. Arink *et al.*, *Performance of the LHCb Outer Tracker*, JINST **9** (2014) P01002, arXiv:1311.3893.

[61] P. d'Argent *et al.*, *Improved performance of the LHCb Outer Tracker in LHC Run 2*, JINST **12** (2017) P11016, arXiv:1708.00819.

[62] E. Rodrigues, *Tracking definitions*, LHCb-2007-006, 2007.

[63] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. **A262** (1987) 444.

[64] M. Needham, *Identification of Ghost Tracks using a Likelihood Method*, CERN-LHCB-2008-026, LPHE-2008-04, 2008.

[65] A. Dosil Suarez *et al.*, *Measurement of the inelastic pp cross-section at centre-of-mass energies* $\sqrt{s} = 13TeV$, LHCb-ANA-2016-077, 2016.

[66] P. A. Cerenkov, *Visible radiation produced by electrons moving in a medium with velocities exceeding that of light*, Phys. Rev. **52** (1937) 378.

[67] A. Contu, *The Measurement of the Production Cross Section Ratio of Identified*

Hadrons and the Calibration of the Magnetic Distortion in RICH1 at LHCb, PhD thesis, Oxford U., 2012.

[68] M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. **C73** (2013) 2431, arXiv:1211.6759.

[69] A. Papanestis and C. D'Ambrosio, *Performance of the LHCb RICH detectors during the LHC Run II*, Nucl. Instrum. Meth. **A876** (2017) 221, arXiv:1703.08152.

[70] LHCb, R. Aaij *et al.*, *Search for long-lived heavy charged particles using a ring imaging Cherenkov technique at LHCb*, Eur. Phys. J. **C75** (2015) 595, arXiv:1506.09173.

[71] M. Adinolfi *et al.*, *Performance of the LHCb RICH photodetectors in a charged particle beam*, Nucl. Instrum. Meth. **A574** (2007) 39.

[72] R. Forty, *RICH pattern recognition for LHCb*, Nucl. Instrum. Meth. **A433** (1999) 257.

[73] D. Derkach *et al.*, *Machine-Learning-based global particle-identification algorithms at the LHCb experiment*, J. Phys. Conf. Ser. **1085** (2018) 042038.

[74] R. Aaij *et al.*, *Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14** (2019) P04013, arXiv:1812.10790.

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

[76] GEANT4, S. Agostinelli *et al.*, *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[77] L. Anderlini *et al.*, *The PIDCalib package*, LHCb-PUB-2016-021, 2016.

[78] LHCb, R. Aaij *et al.*, *Observation of CP violation in $B^{\pm} \to DK^{\pm}$ decays*, Phys. Lett. **B712** (2012) 203, arXiv:1203.3662.

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

[80] G. D. McGregor, *Calibration of the LHCb VELO Detector and Study of the Decay Mode $D^0 \to K^-\mu^+\nu_\mu$*, PhD thesis, Manchester U., 2011.

[81] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) P09007, arXiv:1405.7808.

[82] H. Patton and B. Nachman, *The Optimal Use of Silicon Pixel Charge Information for Particle Identification*, Nucl. Instrum. Meth. **A913** (2019) 91, arXiv:1803.08974.

[83] K. Cranmer *et al.*, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, CERN-OPEN-2012-016, 2012.

[84] R. J. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Comput. Phys. Commun. **77** (1993) 219.

[85] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, eConf **C0303241** (2003) MOLT007, arXiv:physics/0306116.

[86] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. **A389** (1997) 81.

[87] J. Wimberley, *Semitauonic B_c^+ Decays and Quark Flavor Identification Methods*, PhD thesis, Maryland U., 2017.

[88] R. Aaij *et al.*, *Measurement of the inelastic pp cross-section at a centre-of-mass energy of 13 TeV*, JHEP **06** (2018) 100. 18 p.

[89] R. Aaij *et al.*, *Measurement of Antiproton Production in p-He Collisions at $\sqrt{s_{NN}} =$ 110 GeV*, Phys. Rev. Lett. **121** (2018) 222001, arXiv:1808.06127.