September 5, 2025

Towards femtoscopy with the upgraded LHCb detector: Measurement of proton-proton correlations

Samuel Hooper

Supervisors: Dr. Thomas Pöschl and Dr. Saverio Mariani

LHCb Collaboration, Geneva, Switzerland

Abstract

The first measurement of the femtoscopic correlation function of p–p pairs with LHCb Run 3 NoBias data and simulation samples is presented. Correlations are clearly observed in data down to low relative momenta, demonstrating the feasibility of femtoscopy in this kinematic regime of the proton pair with the upgraded LHCb detector. The data follows the expected trend of enhancement at low k^* and approaches unity at larger values, while simulation exhibits significant deviations already at Monte Carlo truth level. For simulation, we find deviations from the expected behavior that are still under investigation. With the current 2024 and 2025 data sets, the precise extraction of the pp source size is deemed possible, enabling the study of rarer baryon–baryon systems.

Contents

1 Introduction

The strong interaction between pairs of baryons is a topic with many open questions. The underlying theory of quantum chromodynamics (QCD) is especially complex at low energies, due to its non-perturbative character in this regime. The perturbative techniques that work well at high energies break down, and lattice QCD or effective field theories become necessary tools for theoretical predictions. Despite decades of research, precise descriptions of baryon-baryon interactions, especially those involving strange quarks, remain incomplete.

Correlation functions provide a way to study these interactions experimentally. In heavy-ion and hadron-hadron collisions, the spatial geometry of the emission region can be accessed through femtoscopy, where two-particle correlations are sensitive to the final state interactions of the emitted baryons.

Recent ALICE measurements [1] have provided high-precision correlation functions for several baryon pairs, including p–p, p–Λ, and Λ–Λ. These results serve as benchmarks for theoretical models and are especially valuable because they are derived from heavy-ion collisions, where the system size is large and collective effects are pronounced. By comparing such measurements with theoretical predictions, one can constrain interaction potentials and test the validity of various models, such as chiral effective field theory or lattice QCD.

In contrast, femtoscopy has not yet been established at the LHCb experiment. LHCb has so far focussed on spectroscopy, where new resonances are studied by reconstructing invariant mass spectra. However, femtoscopy could provide complementary access to hadron–hadron interactions, particularly in pp collisions where LHCb has clear advantages: high luminosity, unique forward kinematic coverage, and excellent reconstruction capabilities for heavy hadrons [2].

The fixed-target configuration provided by SMOG2 [3] further extends this opportunity to lower energy collisions, characterized by different system dynamics. This opens the possibility to study baryon-baryon correlations in a new kinematic regime, complementary to the heavy-ion measurements at ALICE.

Beyond collider physics, constraints on baryon-baryon interactions are important in a wide range of fields. For example, hyperon[1]-nucleon interactions may help determine the internal composition of neutron stars [4–6]. At very high densities, it is energetically favorable for hyperons to appear. They lower the system's total energy because they help balance the Fermi pressure of neutron stars. When hyperons appear, they tend to soften the equation of state (EoS), meaning the pressure drops for a given density, making neutron star matter more compressible. However, this softened EoS cannot support very massive neutron stars as they have been observed. Similarly, nuclear structure and reaction theory, the dynamics of supernova explosions, and the formation of hypernuclei all rely on an accurate understanding of the baryon-baryon potential. Constraints on

[1]Baryons containing at least one strange quark, but no charm, bottom, or top quark, like the Λ or the Ξ.

multi-strange systems could also shed light on the possible existence of exotic states such as dibaryons or bound hyperon systems, which have been long hypothesized but remain experimentally elusive [7, 8].

This project therefore explores the use of femtoscopy at LHCb, starting from p —p correlations in Run 3. The aim is to establish the methodology in this new experimental environment and demonstrate that LHCb can provide complementary input to the global effort of understanding hadron–hadron interactions in the non-perturbative QCD regime.

2 Theory

2.1 Baryon-baryon interactions

The fundamental theory governing baryon–baryon interactions is quantum chromodynamics [9] (QCD), which is challenging to apply at hadronic scales due to non-perturbative confinement effects. Protons are the natural starting point because they are stable, abundant, and experimentally accessible, allowing for precise measurements and well-established theoretical models. One such widely used model is the Argonne v18 potential [10], which incorporates 18 operator components related to particle quantum numbers such as spin and isospin, and accurately reproduces nucleon–nucleon scattering data. These phenomenological potentials provide effective descriptions of the nuclear force by modeling meson exchanges [11] and including a strong repulsive core at short distances. However, the long-term goal is to extend studies to hyperon–nucleon and hyperon–hyperon correlations, for which the interactions are less understood but crucial for neutron star physics [4, 5].

Several theoretical approaches have been developed to tackle the complexities of baryon-baryon interactions, with two prominent methods being lattice QCD and chiral effective field theory.

Lattice QCD is a numerical approach that discretises spacetime into a finite lattice, enabling the non-perturbative solution of QCD from first principles. By simulating quarks and gluons on this lattice, it is possible to calculate hadronic properties and interactions with increasing precision. Despite its power, lattice QCD calculations remain computationally intensive, especially at physical quark masses, limiting their current applicability mostly to heavier hadrons or unphysical quark mass scenarios. Nevertheless, lattice QCD provides valuable insights into baryon interactions that are difficult to access experimentally.

Chiral effective field theory (χEFT), on the other hand, is an effective theory framework that models baryon interactions using hadrons and pions as the relevant degrees of freedom, while maintaining the symmetries of QCD, particularly chiral symmetry. This approach systematically expands the interaction potentials in terms of momenta and quark masses, controlled by low-energy constants fitted to experimental data where available. χEFT has been successfully applied to nucleon-nucleon forces and is being extended to include hyperon-nucleon interactions. Its perturbative nature allows for systematic improvement and uncertainty quantification, making it a powerful complementary tool to lattice QCD.

Together, these theoretical frameworks aim to provide a comprehensive description of baryon-baryon interactions across different energy regimes and baryon species.

2.2 Femtoscopy

Femtoscopy is a technique originally developed in heavy-ion physics to probe the space-time characteristics of particle-emitting sources. It is based on measuring the correlation between two particles emitted close in space and time, which are influenced by both quantum statistical effects and their mutual interactions in the final state. The shape of the correlation function therefore carries information about both the source and the interaction potential.

A particularly illustrative case comes from the STAR collaboration, which published results suggesting a repulsive interaction in the Λ–Λ system. This was unexpected and controversial, as no theoretical model predicted such behavior. The issue was later resolved in a re-analysis, which showed that different assumptions about residual correlations lead to the opposite conclusion: a mildly attractive interaction, consistent with theoretical expectations [12]. This example demonstrated that femtoscopic conclusions can be highly sensitive to analysis assumptions, and highlighted the need for methodological improvements if meaningful physics is to be extracted.

Femtoscopy at the LHC was motivated by the long-standing lack of experimental constraints on two-particle interactions involving strange quarks. One of the driving visions behind its use in small systems was the idea that, unlike heavy-ion collisions, these systems are less influenced by collective or in-medium effects such as those from a quark-gluon plasma. Instead, particle production is dominated by hard scatterings between quarks and gluons, which are governed by QCD. Due to its approximate flavour symmetry, QCD implies a similar space-time structure of hadron emission across different baryon species. If this symmetry holds, then a particle pair with a well-known interaction—such as proton–proton—can be used to extract the source function, which can subsequently be applied to pairs with poorly known interactions, such as p–Λ or Λ–Λ, allowing the study of their final state interactions (FSI). To contribute to this overarching goal of constraining the source function at LHCb and gaining knowledge about FSI in more complex systems, data from collider events and fixed-target in Run 3 will be used.

2.2.1 Correlation function

To extract the necessary information, the femtoscopy method uses the measured correlation function, which depends on the interplay between the source function and the FSI encoded in the two-particle wave function. This is expressed theoretically as:

$$C_{\text{theo}}(k^*) = \int \mathrm{d}^3r^*\, S(r^*)\, |\Psi(\mathbf{k}^*, \mathbf{r}^*)|^2 \ , \tag{1}$$

where

$$k^* = \frac{\sqrt{\lambda(m_{12}^2, m_1^2, m_2^2)}}{2m_{12}} \tag{2}$$

is the relative absolute momentum[2] of the pair in their rest frame with the fully symmetric Källén function

$$\lambda(a,b,c) = a^2 + b^2 + c^2 - 2ab - 2bc - 2ac \ . \tag{3}$$

Furthermore, $\mathbf{r}^*$ is their relative position, $S(r^*)$ is the emission source function, and $\Psi(\mathbf{k}^*, \mathbf{r}^*)$ is the relative wave function of the pair.

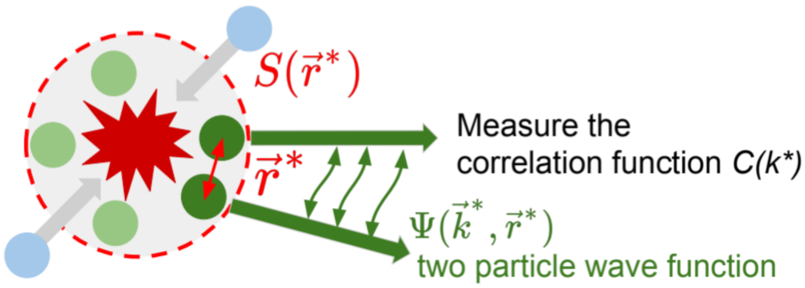

Figure 1: Principle of fentoscopy. The colliding particles (blue) create an emission source (red dashed circle), from which particles are created (green). The correlation is measured for a particle pair, and it is related to the two-particle wave function. Taken from [13].

The emission source function can be interpreted as the probability of emitting a particle pair of for a given relative momentum and position, with the normalisation condition:

$$\int \mathrm{d}^3r^* S(r^*) = 1 \ . \tag{4}$$

Furthermore, no momentum dependence of the source is included in the source function, yielding the Koonin-Pratt relation:

$$C_{\text{theo}}(k^*) \overset{k^* \to \infty}{=} 1 \ . \tag{5}$$

This links the measureable correlation function to the integral over the source and wave functions. As such, it is clear that in femtoscopic methods one of both functions should be known in order to study the other.

The experimental correlation function is given by the ratio of all the signal sample with pairs stemming from the same event, and a reference sample, where all the pairs consist of protons from separate events:

$$C_{exp}(k^*) = N(k^*) \frac{A(k^*)}{B(k^*)} \ . \tag{6}$$

[2]Vectors are denoted as bold faced letters throughout this analysis.

N is a factor to correct experimental effects. There are many considerations to take into account for the reference sample, which will be discussed in Sec. 3.3.

2.2.2 Residual correlations

In femtoscopic analyses, the measured correlation function does not only reflect the interaction between two primary particles. Instead, it often contains contributions of other origins, which are collectively referred to as *residual correlations*. These must be accounted for or minimized in order to extract meaningful physical information about the genuine interaction of interest.

There are two main sources of residual correlations:

1. **Misidentified particles**: These are particles incorrectly identified as protons (or another species of interest) due to limitations in the detector resolution or particle identification algorithms. Their contribution can be reduced through strict selection criteria, such as tight cuts on PID variables and kinematic constraints. However, a trade-off exists between purity and statistics.

2. **Feed-down from weak decays**: Many baryons observed in the detector originate from the decay of heavier particles. For example, a detected proton may stem from the weak decay of a Λ or Σ^+. These decay products still contribute to the measured correlation function, but their momenta differ from those of primary particles. Consequently, they alter the structure of the correlation signal.

Residual correlations affect the shape of the correlation function and can obscure the genuine FSI signal. To model their contribution, a formalism is used in which each pair type—both genuine and residual—is assigned a weight factor λ_{ij}. These weights are defined as:

$$\lambda_{ij} = P_i P_j f_i f_j \ , \tag{7}$$

where P_i, P_j are the purities of particles i and j (probabilities that they are correctly identified), and f_i, f_j are the feed-down fractions (probabilities that particles i and j are primary or from a specific parent species).

The full correlation function is then constructed as a weighted sum:

$$C(k^*) = 1 + \lambda_{\text{genuine}} \left[C_{\text{genuine}}(k^*) - 1\right] + \sum_{ij} \lambda_{ij} \left[C_{ij}(k^*) - 1\right] \ , \tag{8}$$

where $C_{\text{genuine}}(k^*)$ is the theoretical correlation function for primary–primary interactions, and $C_{ij}(k^*)$ are the correlation functions for each residual pair type i–j.

As an example, proton–proton correlations are considered. The measured pair sample $\{pp\}$ contains not only primary pp pairs, but also combinations of protons from decays such as $\mathit{\Lambda} \to p\pi^-$ or $\Sigma^+ \to p\pi^0$, and even misidentified particles. The full set of contributions can be written as:

$$\{pp\} = pp + p_\Lambda p + p_\Lambda p_\Lambda + p_{\Sigma^+} p + p_{\Sigma^+} p_{\Sigma^+} + p_\Lambda p_{\Sigma^+} + \tilde{p}p + \tilde{p}p_\Lambda + \tilde{p}p_{\Sigma^+} + \tilde{p}\tilde{p} \ , \tag{9}$$

where p_Λ and p_{Σ^+} denote protons from Λ and Σ^+ decays, respectively, and $\tilde{p}$ denotes a misidentified proton candidate.

In practice, residual correlation functions $C_{ij}(k^*)$ are modeled based on known decay kinematics and theoretical input, or extracted from dedicated simulations. The overall normalisation and contribution of each component is fixed by the measured purities and feed-down fractions in data. Properly modeling and minimizing residual correlations is essential for isolating the genuine correlation signal and drawing meaningful conclusions about the strong interaction between baryons.

3 Event samples and selection

This analysis represents the first inspection of the data sets taken using pp collisions at $\sqrt{s} = 13.6\,\text{TeV}$ in Run 3 at the LHCb detector in 2024 (Block 8) with an upward magnet polarity. The analysis is performed on a minimum bias data sample, which is particularly useful as it provides an unbiased representation of the full range of pp collisions, including soft and hard processes. This ensures a comprehensive view of baryon production and correlation phenomena, and avoids potential biases introduced by specific trigger selections.

The Monte Carlo (MC) sample is simulated with the same data-taking conditions as the data, in order to permit a meaningful comparison with data. The event generation and simulation are performed using the `Gauss` framework [14], which is the official LHCb simulation toolkit. `Gauss` integrates particle generation using Pythia [15], decay modeling via EvtGen [16], and full detector simulation with Geant4 [17] [18], allowing realistic modeling of particle interactions and detector response within the LHCb apparatus.

Some selection criteria can be applied before event mixing, while others must be applied independently on the same-event and mixed-event sample, depending on the nature of the cut. Cuts that depend solely on the properties of individual tracks, such as track quality, momentum, or impact parameter, can be safely applied before mixing. Applying them at this stage reduces the dataset size and improves computational efficiency without biasing the uncorrelated reference sample. In contrast, selections that depend on the relationship between two particles, such as angular separation or shared detector hits, must be applied after mixing. Applying these cuts before mixing would remove certain track combinations from the reference sample, potentially introducing artificial correlations and distorting the experimental correlation function. By following this strategy, the mixed-event background can correctly represent the uncorrelated two-particle phase space while still allowing for efficient suppression of combinatorial contamination.

A technical complication arises from the fact that the LHCb software framework requires a well-defined decay chain to be specified, if the protons are to be paired online. Since the correlation study concerns only two final-state protons, which do not originate from a common narrow resonance, it is necessary to introduce a pseudodecay to satisfy this requirement. In practice, this is implemented by defining a fake J/ψ decay to a pair of protons:

```
fields = {
    "Jpsi": "[J/psi(1S) -> p+ p+]CC",
    "p1":  "[J/psi(1S) -> ^p+ p+]CC",
    "p2":  "[J/psi(1S) -> p+ ^p+]CC",
}
```

This construct has no physical meaning, but it provides a convenient handle for the ntupling framework to assign variables to each proton in a consistent way.

In addition to the reconstructed MC sample, a small truth-level sample is produced for testing purposes. This is obtained by running an Analysis Production with the same input files as the reconstructed MC, but with a modified configuration file that enabled truth-level information. The production is validated using `lb-ap test`, which yielded a limited dataset of generator-level protons. Although smaller in size than the reconstructed MC, this truth sample is valuable for cross-checks, such as validating correlation function inputs.

The input samples for this analysis are documented and traceable through the following Analysis Productions and MC request:

- **Data AP** – Source size AP for *pp*@SMOG2 using minBias (`!3377`). Provides the real data sample used in this study.[3]
- **MC AP** – Source size AP for *pp*@SMOG2 using minBias simulation (`!3548`). Corresponding simulated sample.[4]
- **MC request** – Ift mb pp block78 sim10f (`!1367`). Request for MC sample used as input for MC AP.[5]
- **Draft MC truth AP** – Draft AP for truth-level protons (`!3718`). Would provide MC truth with minimal cuts. Still in draft status.

3.1 Online preselection

The preselection is an essential step in the data processing pipeline, refining the dataset by applying strict selection criteria. This is done in order to suppress background contaminations and to make the data bandwidth more feasible. The goal is to study correlations of proton pairs, with as little bias through prior selections as possible. To ensure consistency and allow for meaningful comparisons, the same preselection criteria are applied to both the data and simulation sample.

The single proton candidates require kinematic cuts of $p \in [12, 100]$ GeV, as they only emit Cherenkov radiation in the RICH detectors for the lower limit and for momenta larger than the upper limit, the opening angle between the proton candidate tracks becomes too small to distinguish them. Furthermore, they require $p_T \in [0.4, 4]$ GeV for the proton candidates to be within the acceptance of the LHCb detector. As can be seen

[3] `/eos/lhcb/grid/prod/lhcb/anaprod/lhcb/LHCb/Collision24/DVNTUPLE.ROOT/00305475/0000/00305475_00000001_1.dvntuple.root`; `/eos/lhcb/grid/prod/lhcb/anaprod/lhcb/LHCb/Collision24/DVNTUPLE.ROOT/00305970/0000/00305970_00000001_1.dvntuple.root`

[4] `/eos/lhcb/grid/prod/lhcb/anaprod/lhcb/MC/2024/DVNTUPLE.ROOT/00305473/0000/00305473_00000001_1.dvntuple.root`

[5] `/MC/2024/Beam6800GeV-2024.W41.42-MagUp-Nu7.6-25ns-Pythia8/Sim10f/HLT2-2024.W40.42/30000000/HLT2.DST`

in Fig. 2, the majority of reconstructed track candidates populate the low-momentum and low-transverse-momentum region, in particular below $p = 12\,\mathrm{GeV}$ and $p_T = 0.4\,\mathrm{GeV}$. The same behaviour is observed in simulated events, shown in Fig. 16, confirming that it is not an artefact of the reconstruction but an intrinsic feature of the kinematics. This means that a large fraction of statistics is lost due to the applied cuts. However, these restrictions are unavoidable: without sufficient momentum the protons do not produce a Cherenkov signal in the RICH, and without sufficient transverse momentum they fall outside the effective tracking and particle-identification acceptance of the detector.

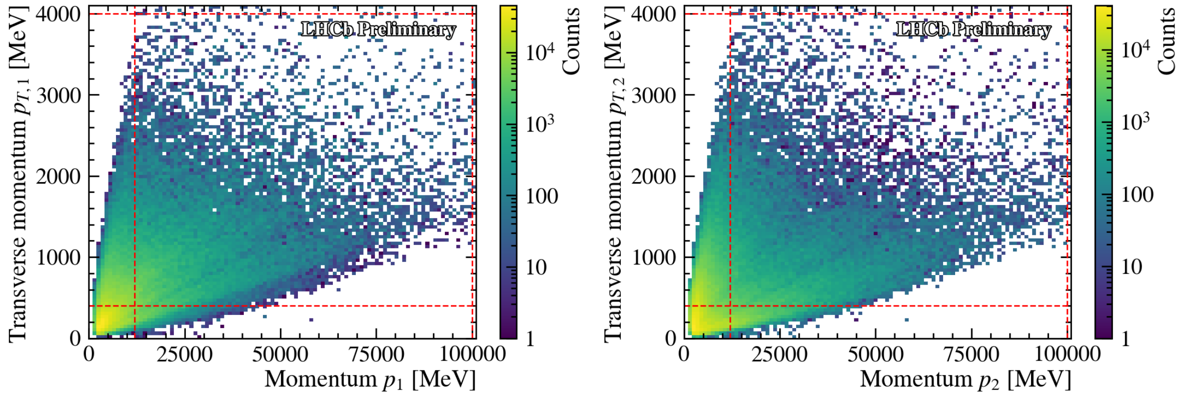

Figure 2: Transverse momentum p_T against momentum p for the first (left) and the second proton candidate (right) in the pair for data. The dashed red lines indicate the rectangular cuts.

Recognising misidentified particles is an important part of reducing the background. Due to this effect, a neural network is trained to calculate probabilities of a track belonging to a certain particle. Hence, the probability of the track being a ghost track is required to be less than 15%. Also, only tracks with electron and pion probabilities of less than 5% are kept during the selection process. For kaons, the limit is set at 10%. Figure 3 illustrates the effect of these requirements. The top row shows the distribution of proton–kaon separation, plotting `ProbNNk` versus `ProbNNghost` for the first (left) and second (right) proton candidate, with the applied cuts indicated by red dashed lines. The second row shows the analogous distributions of `ProbNN`π versus `ProbNNe`. In these plots, clear peaks are visible for kaons, electrons and pions, which are efficiently removed by the selection, whereas the ghost probability is more broadly distributed and exhibits no sharp structure. The combination of these criteria ensures that the final proton sample has minimal contamination from misidentified tracks, which is crucial for the extraction of clean correlation functions. The same behaviour can be seen for the simulation plots in Fig. 17.

Instead of also applying a cut on `ProbNNp`, the difference in log-likelihood between the proton and pion hypotheses

$$\mathrm{DLL}_{p,\pi} = \log(p \text{ likelihood}) - \log(\pi \text{ likelihood}) \tag{10}$$

is required to be greater than 10. In Fig. 4, a clear peak is visible around zero, while only the right-hand tail is kept by the selection. The same behaviour can be seen for

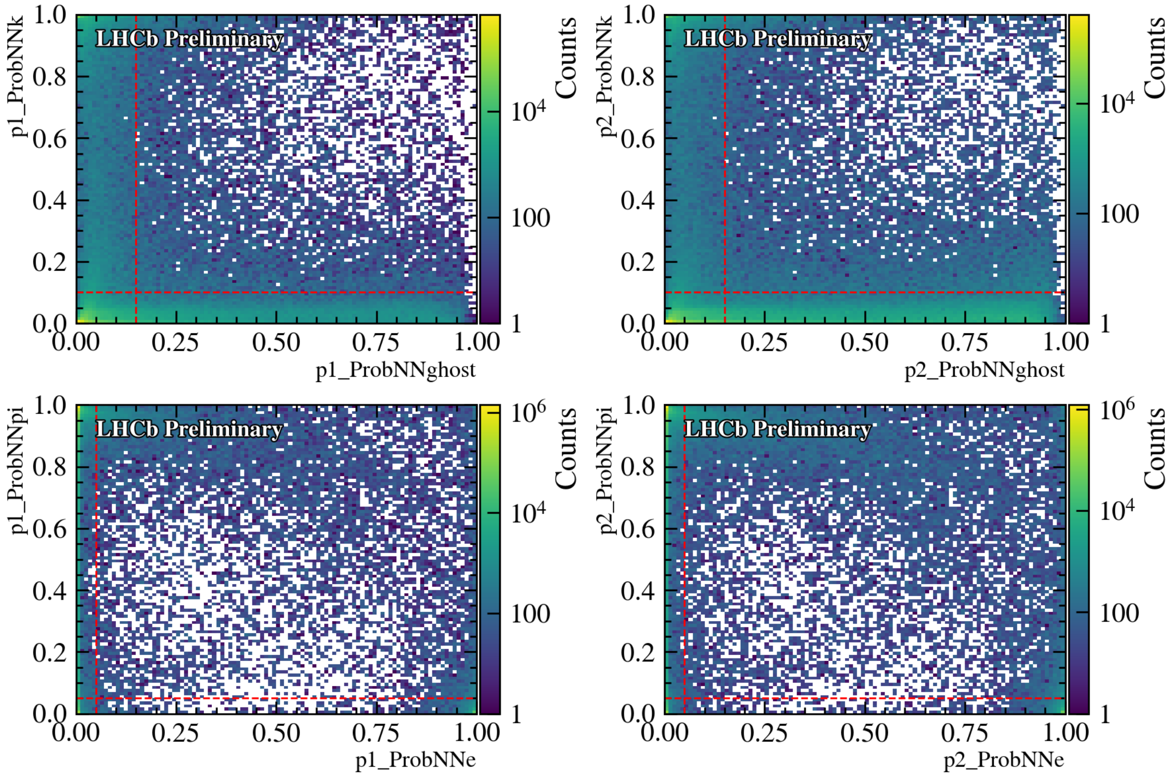

Figure 3: Particle hypothesis probability cuts in data. Top: `ProbNNk` versus `ProbNNghost` for the first (left) and second (right) proton candidate. Bottom: `ProbNN`π versus `ProbNNe` for the same candidates. The red dashed lines indicate the applied selection boundaries.

simulation in Fig. 18.

Particle identification is done using RICH information. Hence, as all final state particles are charged, they are obliged to have a positive response in the RICH detectors due to Cherenkov radiation. Besides that, the candidate is required to be in the acceptance region of either RICH1 or RICH2.

The first two plots in Fig. 5 show the distribution of the response for the first and second proton candidate in RICH1, while the next two plot shows the same for RICH2. True values correspond to a valid Cherenkov response, whereas False indicates no response. These histograms clearly reveal that a significant fraction of tracks, roughly 15–55% depending on the detector and particle, do not produce a valid Cherenkov signal. For the MC sample, a similar behaviour is observed, as can be seen in Fig. 19. In particular, the response for the second proton in RICH1 and RICH2 is dominated by False entries, highlighting the necessity of applying the RICH response cut to remove tracks without reliable PID information.

The acceptance plots in Fig. 6 show the fraction of candidates that lie within the fiducial acceptance of each RICH detector. A value of True indicates that the proton track is within the detector acceptance, while False indicates that it is outside. These distributions

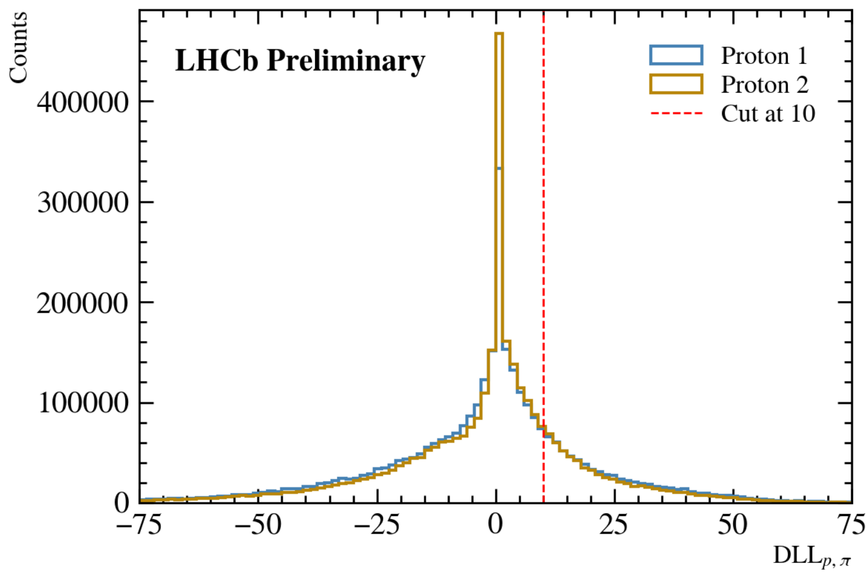

Figure 4: Distribution of the proton–pion log-likelihood difference in data, $\mathrm{DLL}_{p,\pi}$, for the first (blue) and second (orange) proton candidate. The vertical red dashed line marks the applied cut at $\mathrm{DLL}_{p,\pi} > 10$.

demonstrate that more than approximately 80% of the candidates for both protons fall outside the acceptance region. This can behaviour is can also be seen for the MC sample in Fig. 20. By applying the acceptance cut, only proton candidates that can reliably produce Cherenkov radiation within the detector volume are retained, which is crucial for ensuring the quality of the PID information and minimizing the background from poorly reconstructed tracks.

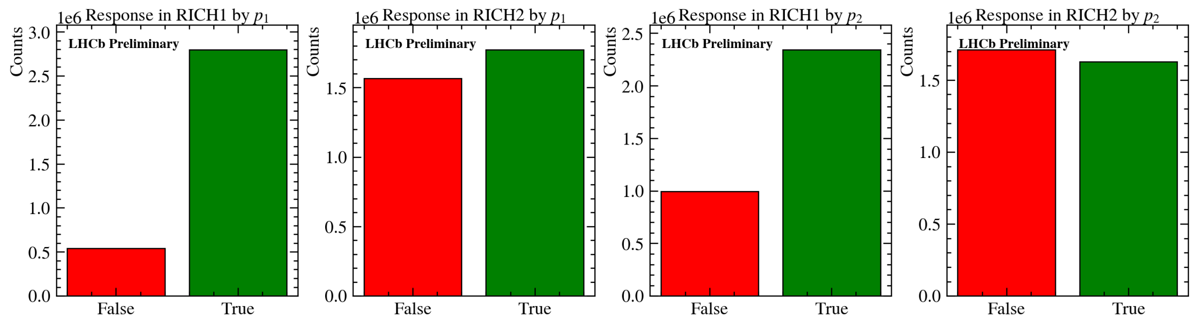

Figure 5: Binary distributions of the RICH response for proton candidates in data. For the first proton (outer left) and second proton (inner left) candidate in RICH1, as well as the first proton (inner right) and second proton (outer right) candidate in RICH2, a value of `True` indicates that the track produced a valid Cherenkov response, while `False` corresponds to no response.

The effectiveness of the PID requirements can be assessed by studying the distribution of `ProbNNp`, which serves as a measure of the proton purity in the selected sample.

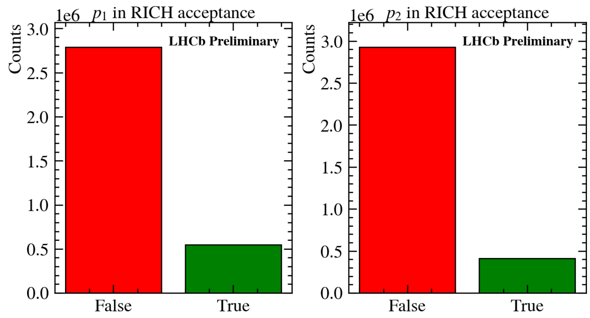

Figure 6: Binary distributions of the RICH acceptance for proton candidates in data. The value `True` indicates that the track lies within the acceptance of the corresponding RICH detector, whereas `False` indicates it falls outside the acceptance.

Before applying the preselection, the distribution exhibits a pronounced peak at low values, indicating a substantial fraction of misidentified tracks. After the cuts, this peak is suppressed by about four orders of magnitude with respect to the peak for a high probability for the proton hypothesis being true, rendering the residual contamination negligible. For comparison, prior to the selection the low-value contribution is approximately two orders of magnitude larger, highlighting the significant background from impurities. This striking reduction demonstrates both the efficiency and the necessity of the PID selection criteria.

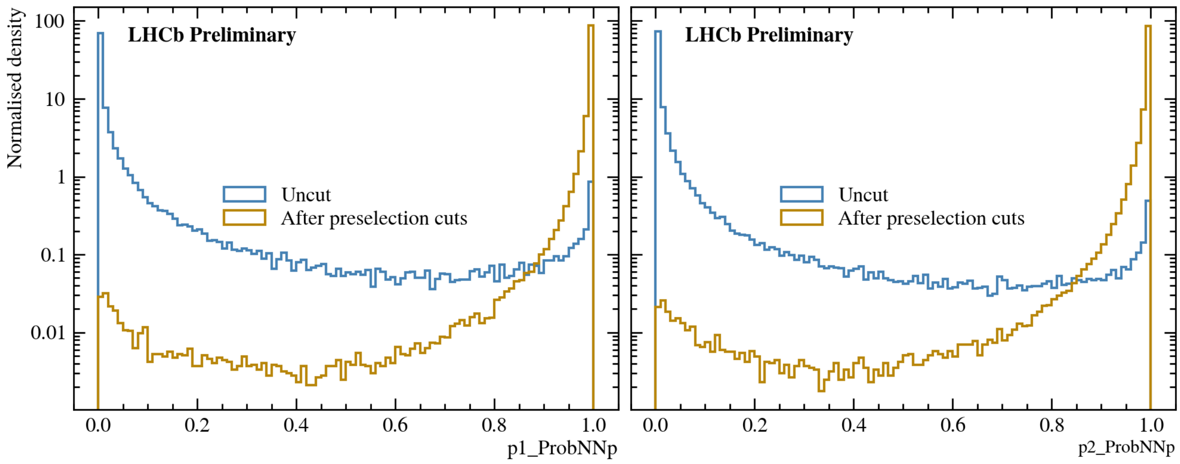

Figure 7: Normalised distributions of `ProbNNp` for the first (left) and second proton candidate (right) in data.

To ensure that the selected proton candidate pairs originate from the same primary vertex, a requirement is placed on the distance of closest approach (DOCA) between the two tracks. Since the reconstructed candidate trajectories do not necessarily intersect at an exact point, the DOCA is required to be smaller than 0.2 mm. This removes pairs with nearly overlapping tracks, which are typically of poor reconstruction quality and likely not true proton pairs. As shown in Fig. 22, a substantial fraction of the long tail towards small DOCA values is removed by this requirement, thereby improving the overall purity of the selected sample. This behaviour is also reproduced in simulation.

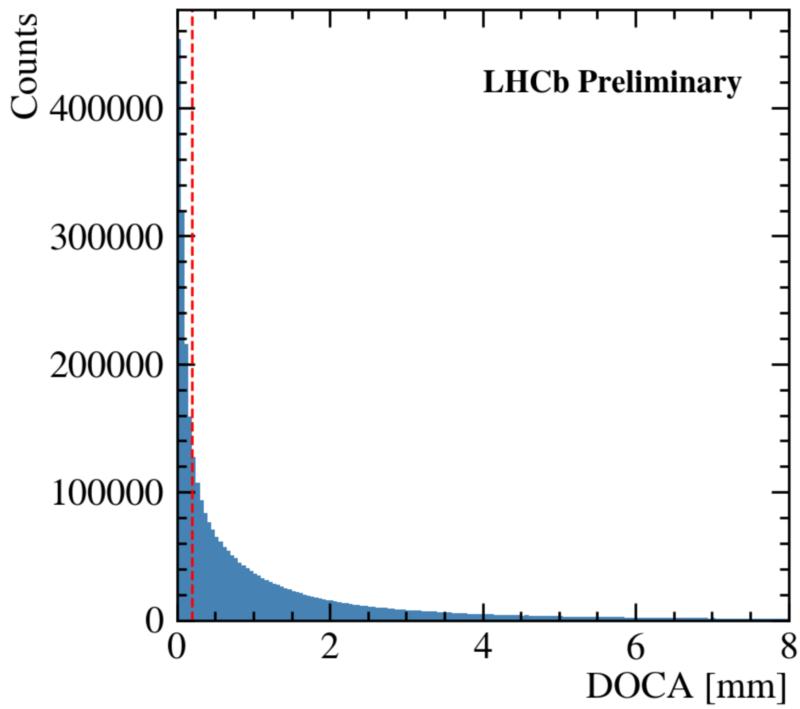

Figure 8: Distribution of the distance of closest approach (DOCA) between proton candidate tracks in data. The red dashed line indicates the applied cut at 0.2 mm.

The distribution of the reconstructed primary vertex position along the z-axis shows a clear asymmetry, with a secondary peak visible at negative z values. This feature is observed both only in data (Fig. 9), and not in simulation (Fig. 23). These vertices originate upstream of the VELO, likely from interactions with the SMOG2 fixed-target system, and are not relevant for the pp collision environment under study. To suppress this contribution and retain only vertices consistent with pp collisions inside the VELO, the best primary vertex position is required to lie within the range $z \in [-200, 200]$ mm.

A summary of the preselection is given in Tab. 1.

3.2 Offline Signal Selection

Although preselection cuts are already applied to the data sample, it is necessary to further narrow down the total number of events. All plots shown within this chapter have the preselection from the previous chapter applied to.

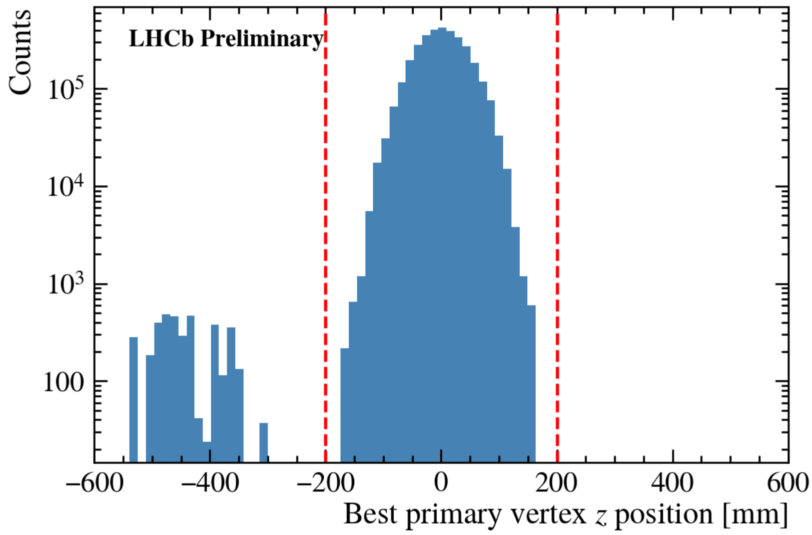

Figure 9: Distribution of the reconstructed primary vertex position z in data. The red dashed lines indicate the applied selection window restricting the z-position to the VELO region ($z \in [-200, 200]$ mm).

Table 1: Summary of online selection requirements applied to the data and MC sample with long proton candidates.

Description	Requirement
$p^{\pm}$	
Momentum	$p \in [12, 100]$ GeV
Transverse momentum	$p_T \in [0.4, 4]$ GeV
Difference in log-likelihood	$\mathrm{DLL}_{p,\pi} > 10$
Probability of track belonging to a ghost	ProbNNghost< 0.15
Probability of track belonging to a π/e	ProbNN{pi,e}< 0.05
Probability of track belonging to a K	ProbNNk< 0.1
RICH1 or RICH2 in acceptance region	True
RICH response	True
$p^{\pm}p^{\pm}$ pair	
Distance of closest approach	DOCA < 0.2 mm
Best primary vertex position in z	$\lvert\mathrm{BPV}_z\rvert < 200$ mm

3.2.1 Residual structures

In addition to the genuine physical correlations, various detector-related effects can leave residual structures in the two-proton-candidate sample. Examples include non-uniform acceptance, efficiency variations across kinematic regions, and small distortions due to the event reconstruction procedure. Such effects typically influence both the same-event distribution $A(k^*)$ and the mixed-event reference $B(k^*)$ in a similar way. As the experimental correlation function is constructed as the ratio $C(k^*) = A(k^*)/B(k^*)$, these contributions cancel to first order, ensuring that $C(k^*)$ remains unaffected by global acceptance and efficiency biases.

However, not all sources of artificial correlations are removed in this way. Effects that are present only in the same-event sample—such as track reconstruction artifacts, particle misidentification, or contributions from weak decays—do not have an analogue in the mixed-event sample and therefore survive in the ratio. These require dedicated treatments and will be discussed in the following sections.

3.2.2 Clone contamination

One particularly important source of residual correlations in this analysis is the presence of clone tracks. Clones are artificial duplicates of a single true particle trajectory that the tracking algorithm reconstructs as two (or more) separate tracks. While in some experiments (e.g. ALICE) the dominant issue arises from track merging due to detector occupancy, at LHCb the main challenge stems from the extremely high spatial resolution of the tracking detectors. This precision makes the reconstruction sensitive to subtle ambiguities in hit assignment, which can result in clone tracks being formed.

Clone contamination is problematic because it introduces fake proton pairs into the sample. These clones are not independent particles, but rather two reconstructions of the same proton, and hence they cluster in regions of low relative momentum. In practice, this can be observed in two-dimensional histograms of k^* versus the opening angle θ. Clone pairs accumulate in the region of small k^* and small θ. An additional complication is that clone tracks only contaminate the same-event sample. Since both reconstructions of a proton candidate belong to the same collision event, they can only appear as pairs in this category. In the mixed-event sample, where tracks from different collisions are paired, clones cannot be combined and thus do not contribute. As a result, the numerator of the correlation function (eq. (6)) is artificially enhanced at small k^*, while the denominator remains unaffected. This creates a peak in $C(k^*)$ that mimics the very femtoscopic signal under study, since this is where strong final-state interaction effects produce the characteristic enhancement. If not corrected for, such contamination would lead to an overestimation of the correlation strength at low k^* and consequently to biased source size and interaction parameters. Therefore, clone suppression is a crucial step to ensure the physical reliability of the extracted correlation function. Unfortunately, the clones cannot simply be removed by cutting on k^*, because the correlation function $C(k^*)$ itself is defined as a function of this variable. Other analyses sometimes discard the problematic low k^* region altogether, but for our purposes this would remove the signal we wish to study.

The current strategy to mitigate clone contamination is based on identifying track overlap in each of the LHCb tracking subdetectors. For every pair of tracks, the fraction of detector hits that are shared between them can be calculated. A high degree of overlap provides strong evidence that the two tracks are in fact clones of the same underlying particle. As shown in Fig. 10 for data and in Fig. 24 for simulation, the distribution of overlap weights exhibits a pronounced excess at low k^*, reflecting the strong contribution of clone pairs in this region. To suppress this contamination, a cut at 90% overlap is applied, which effectively removes clone candidates while retaining genuine proton pairs.

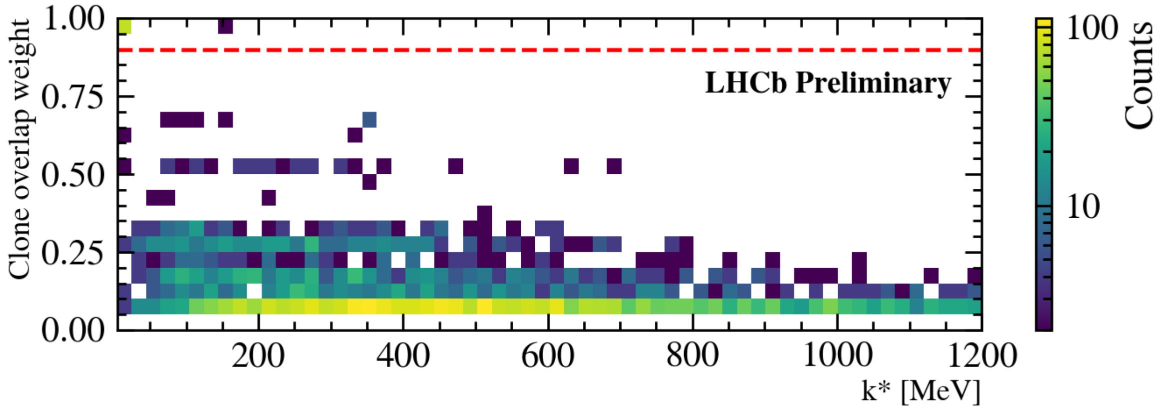

Figure 10: Distribution of the track-overlap weight for proton candidate pairs in data as a function of k^*. The red dashed line indicates the applied cut at 90%.

The implementation of the overlap-based clone suppression in the Analysis Production relies on the `LHCbIDOverlapRelationTable` algorithm, which constructs relation tables between pairs of selected proton candidates. Separate relation tables are created for each subdetector — the VELO Pixel (VP), Upstream Tracker (UT), and Forward Tracker (FT) — by enabling or disabling the respective `IncludeVP`, `IncludeUT`, and `IncludeFT` options. For each track pair, the relation table records whether they share detector hits and assigns a corresponding weight that quantifies the degree of overlap. These relation tables are then mapped into the FunTuple output via dedicated functors. For every subdetector, variables are stored that provide the number of matched tracks, the overlap weight, the particle ID of the matched partner, and the number of hits in the relevant subdetector. This mapping ensures that overlap information can be directly accessed at the ntuple level for both proton candidates, making it possible to apply cuts such as the 90% overlap removal described above. Finally, the relation tables and associated functors are integrated into the Analysis Production configuration, alongside the usual event and particle-level variables.

It is important to note that this cut is applied before the event-mixing procedure, because the overlap information is intrinsic to the individual tracks. Since these weights are defined at the track level and do not mix between events, they must be applied prior to constructing the mixed-event sample to ensure that the reference distribution only contains genuine, non-clone tracks.

In the future, a more robust method will be available through the use of `LHCbIDs`, which are the unique identifiers of hits associated with each reconstructed track. By explicitly comparing the sets of `LHCbIDs` of two tracks, one can unambiguously determine whether they share detector information and quantify the degree of overlap. This approach provides a precise and detector-independent way of identifying clone tracks. However, the necessary tools are still under development and will only become available once they are merged into the master reconstruction framework.

In addition to the overlap criterion, we further require a minimum angular separation between the two tracks of

$$R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} > 0.03 \ , \tag{11}$$

where $\Delta\eta$ and $\Delta\phi$ are the differences in pseudorapidity and azimuthal angle, respectively. This requirement suppresses pairs that are reconstructed with nearly identical trajectories, which are highly likely to be clones. As shown in Fig. 11 for data and in Fig. 25 for simulation, the clone contamination manifests itself as a sharp peak at small $\Delta\eta$ and $\Delta\phi$ in the same-event sample. In contrast, no such structure appears in the mixed-event sample, since mixing by construction pairs tracks from different events and therefore cannot produce clone pairs. The value $R > 0.03$ is chosen as a compromise: it is large enough to reject the majority of clone candidates clustered at very small angular separation, while still being small compared to the typical separations of genuine proton pairs in the kinematic region under study. Together with the track-overlap cut, this ensures that the final sample is largely free of clone contamination, especially in the low-k^* region that is most relevant for femtoscopic correlations.

This cut is applied after mixing because it depends on the relative kinematics of the two particles in the pair, which only exist once the pair is formed. Unlike the track-overlap weight, which is defined for individual tracks and can be applied before mixing, the angular separation requires knowledge of both particles' momenta. Applying it after mixing ensures that the mixed-event background and same-event distributions are treated consistently, while effectively removing spurious pairs that would otherwise distort the correlation function at low k^*.

3.2.3 Non-prompt proton contamination

Another source of background arises from non-prompt proton candidates, which are protons not originating directly from the primary pp collision vertex. Non-prompt proton candidates can stem from weak decays of hyperons or secondary interactions in the detector material. Including these particles would introduce artificial correlations unrelated to the genuine proton-proton femtoscopic signal, thereby biasing the measurement of the correlation function. To ensure that only prompt proton candidates are selected, a requirement on the impact parameter with respect to the primary vertex is applied.

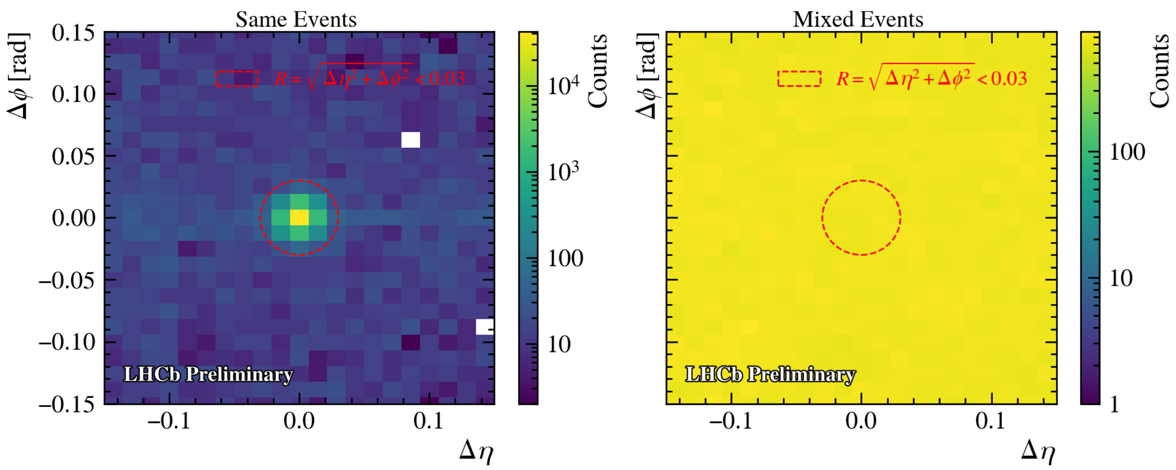

Figure 11: Two-dimensional distributions of $\Delta\eta$ versus $\Delta\phi$ for proton candidate pairs in data, restricted to the region $[-0.15, 0.15]$ (mrad) in both variables. The left plot shows the same-event distribution, where clone contamination is visible as a pronounced peak at the origin. The right plot shows the mixed-event distribution, which is free of clones by construction. The red dashed circle illustrates the applied cut of $R > 0.03$.

The impact parameter significance for the best primary vertex is quantified using the BPVIPχ^2 of each proton. Fig. 12 shows a two-dimensional histogram of BPVIPχ^2 for the first proton candidate on the x-axis versus the second proton candidate on the y-axis in data. A strong peak is visible at the origin, corresponding to proton candidates consistent with the primary vertex. Higher values, indicative of non-prompt proton candidates, are rejected by applying a cut of BPVIP$\chi^2 < 3$ for both tracks, as indicated by the red dashed lines in the figure. This selection effectively removes secondary proton candidates while retaining the prompt sample necessary for a clean correlation analysis. The same behaviour can be seen for simulation in Fig. 26.

3.2.4 Primary vertex consistency

To ensure that the two selected proton candidates originate from the same primary vertex, a consistency requirement on their reconstructed positions is applied. Each proton candidate is associated with a primary vertex candidate, and the differences in the x, y, and z coordinates of the two vertices are computed:

$$\Delta x = x_1 - x_2 \ , \quad \Delta y = y_1 - y_2 \ , \quad \Delta z = z_1 - z_2 \ . \tag{12}$$

Only proton candidate pairs for which all three differences are smaller than 5×10^{-5} mm are accepted. This selection guarantees that the pair is kinematically consistent with originating from the same collision, removing contributions from proton candidates produced in different interactions within the same bunch crossing.

Fig. 13 illustrates the distributions of Δx, Δy, and Δz in three panels for data. The left panel shows all three differences over a zoomed-in region, demonstrating that the main peak near zero is preserved. The dashed red lines indicate the applied selection window,

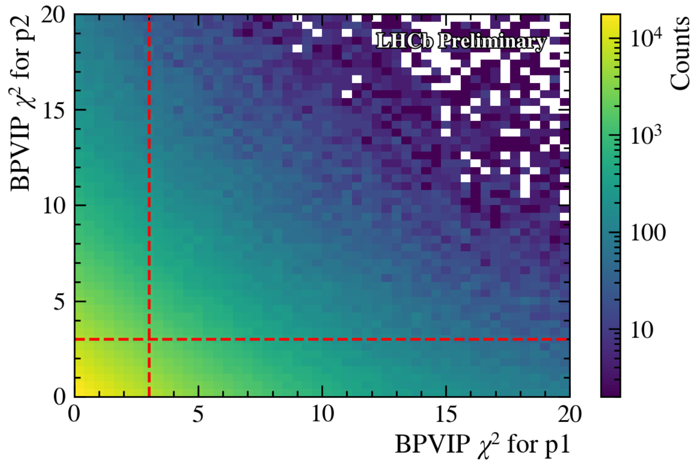

Figure 12: Two-dimensional distribution of the impact parameter significance for the best primary vertex (BPVIPχ^2) of the first proton candidate against the second proton candidate in data. The red dashed lines indicate the applied selection of BPVIP$\chi^2 < 3$ for both tracks.

which lies in a region without bins. The middle panel displays Δx and Δy over the full range, highlighting that most pairs are clustered close to zero while the tails correspond to mismatched vertices. The right panel shows the Δz distribution, which is broader but still features a clear peak near zero. This visualisation confirms that the selection retains genuine same-vertex pairs while rejecting an extremely large amount of mismatched pairs, which is essential for maintaining the integrity of the correlation measurement. The same behaviour can be observed in Fig. 27 for simulation.

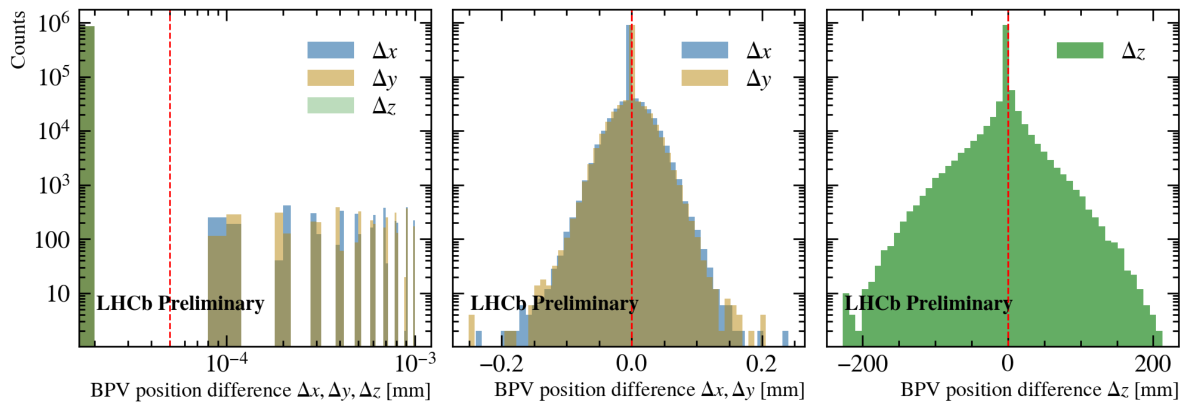

Figure 13: Differences in primary vertex coordinates (Δx, Δy, Δz) for all proton candidate pair in the data sample. Left: zoomed-in view showing the main peak near zero for all three directions, with the red dashed lines indicating the selection window ($|\Delta x|, |\Delta y|, |\Delta z| < 5 \times 10^{-5}$ mm). Middle: full-range distributions for Δx and Δy, showing that most pairs are close to zero. Right: full-range distribution of Δz, which is broader but still exhibits a clear peak near zero.

The offline selection is able to completely remove a residual contamination which is visible for $\Delta\phi = 0$ for data (Fig. 28), and almost completely for simulation (Fig 29). It is important to note that the leftover structures visible in these plots after the offline selection do not contaminate the correlation function if they are present in both same-event and mixed-event sample, as they cancel in the ratio. The complete offline selection is summarised in Tab. 3 and it is combined with the online selection in Tab. **??**.

Table 2: Summary of offline selection requirements applied to the data and MC sample with long proton candidates.

Description	Requirement
$p^{\pm}$	
Best primary vertex impact parameter χ^2	$\chi^2 < 2$
$p^{\pm}p^{\pm}$ pair	
Best primary vertex position difference	$\lvert \mathrm{BPV}\{\mathrm{X,Y,Z}\}_{p_1} - \mathrm{BPV}\{\mathrm{X,Y,Z}\}_{p_2} \rvert < 50\,\mu\mathrm{m}$
Angular separation	$R = \sqrt{\Delta\eta^2 + \Delta\phi^2} > 0.03$
Relative track overlap	Weight < 0.9

3.3 Mixed event sample

In order to compute the experimental correlation function introduced in eq. (6), it is necessary to create a reference sample $B(k^*)$ where the two proton candidates in a pair originate from two separate events. This step is performed chronologically before the event selection, as to avoid introducing artificial correlations from the selection procedure and to guarantee that the reference distribution represents only the uncorrelated phase-space of two independent particles.

The construction of the mixed-event sample must ensure that the two proton candidates in a pair are uncorrelated, while still preserving the global event characteristics that influence the single-particle spectra. Two events are considered uncorrelated if they are sufficiently separated in the data-taking sequence, since correlations only occur within the same collision. This holds equally for the real data and the simulation sample.

To achieve this, the mixing procedure is implemented by pairing each proton candidate from a given event with proton candidates from other events in the dataset that are separated by a fixed offset in the event ordering. The offset is chosen randomly within a defined window of event separation, such that

$$10 < n_{\mathrm{shift}} < 110, \qquad n_{\mathrm{shift}} \in \mathbb{N}\ , \tag{13}$$

where n_{shift} is the number of events between the original and the mixed partner event. This choice suppresses any residual short-range correlations that might arise if events too close in sequence are used, while at the same time avoiding very large separations that could introduce biases due to slow variations in detector conditions or beam parameters.

In addition, the mixing is restricted to events with similar global properties such as reconstructed primary vertex position. The reason for this restriction is that the detector acceptance and reconstruction efficiency depend on the geometry of the collision, which in turn is influenced by the primary vertex position. If pairs are formed from events with very different vertex locations, the mixed-event background would no longer reproduce the acceptance effects present in the real distribution. This could introduce spurious structures into the correlation function, which would mimic or distort genuine physical signals. By ensuring that only events with comparable global characteristics are mixed, the resulting reference distribution reflects the same detector response as the data while containing no intrinsic correlations. Thus, the mixed-event sample provides a realistic representation of the uncorrelated two-particle phase space and is well suited for the proper normalisation of the experimental correlation function.

As part of future improvements, the construction of track-multiplicity dependent samples is also planned. This is motivated by the fact that reconstruction efficiency may vary significantly between high-occupancy and low-occupancy events. By stratifying the mixed-event reference sample according to event multiplicity, it will be possible to further reduce residual acceptance differences and ensure that the background description is robust across the full range of collision environments.

The entire event-mixing procedure is implemented in a single Python program, which handles both the construction of mixed-event pairs and the application of post-mixing selections to both same-event and mixed-event samples. Several measures are taken to optimise performance and reduce memory usage. The program is parallelized to take advantage of multiple CPU cores, and for MC truth samples a chunked processing approach is employed to handle large datasets efficiently. Instead of storing the full event-level DataFrames, only the relevant histograms and correlation function outputs are saved, minimizing storage requirements while preserving all essential information for the analysis. This design ensures both computational efficiency and reproducibility of the correlation function calculations.

4 Results

The correlation function extracted from the Run 3 proton–proton data set is shown in Fig. 14, together with a comparison to the corresponding measurement from ALICE. The data exhibit the expected qualitative trend: a strong enhancement at low k^* values, followed by a plateau around unity at intermediate k^*, and a small upward deviation at larger relative momenta. This behaviour is consistent with the presence of strong final-state interactions in the low-k^* region, while at higher values the distribution approaches the uncorrelated baseline.

In contrast, the MC sample shows only a very modest deviation from unity at low k^*, and a pronounced undershooting beginning at intermediate and increasing for high k^* values. This is in disagreement with expectations, as the simulation should in principle yield a flat distribution at unity across the entire k^* range, given that no genuine femtoscopic

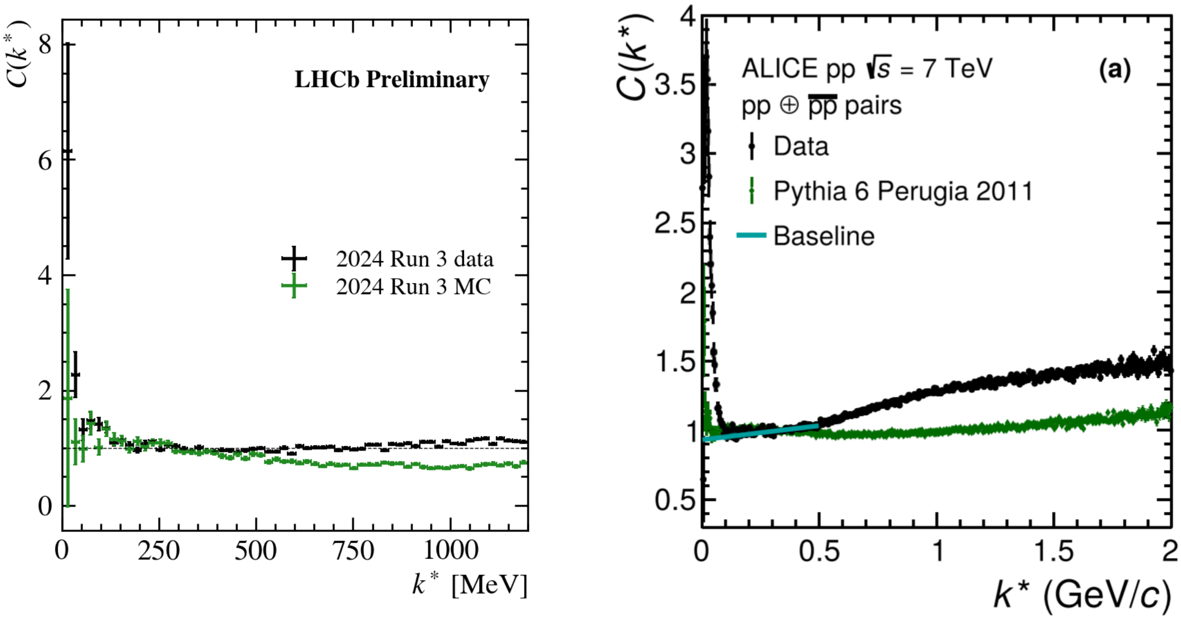

Figure 14: Comparison of the p–p correlation function obtained in this analysis (left) with the corresponding measurement from ALICE (right, taken from [1]).

correlations are included in the simulation. A similar though less severe behaviour is observed in the ALICE study, where the MC deviated slightly towards values greater than unity for high k^*. The discrepancy observed in this analysis is much more significant, and its origin is not yet understood.

The correlation functions are normalised to unity in the range $k^* \in [240, 340]$ MeV, following the procedure of the previous LHCb study by Himbert [19]. It should be noted that residual correlations and possible systematic biases are not yet corrected for in this preliminary result, which would contribute to deviations in both data and simulation.

To investigate the origin of the undershooting observed in simulation, studies are performed at the MC truth particle level. In this approach, all track types are considered, without restricting to long tracks. The selection is limited to basic requirements on particle type, promptness, and kinematics: only protons and antiprotons are kept, they are required to originate promptly from the primary interaction, and their longitudinal momentum is restricted to $|p_z| < 6$ TeV. The latter excludes spectator protons which have a characteristically high energy.

Fig. 15 shows the resulting correlation functions for the three online selection requirements: a cut on the total momentum p (blue), on the transverse momentum p_T (green), and on the reconstructed primary vertex consistency (BPV{X,Y,Z}, orange). In all cases, deviations from the expected flat behaviour around unity are visible. At very low k^*, all three selections exhibit a sharp drop to nearly zero, the origin of which is not yet understood. At intermediate k^*, the correlation functions remain close to unity, consistent with the applied normalisation. At higher k^*, different trends emerge: the p selection remains close to one, the p_T selection shows a mild enhancement, while the BPVXYZ requirement leads to a slight suppression down to about 0.9 at $k^* \approx 1200\,\text{MeV}$. While the combined average of all cuts is around unity with these statistics, the individual samples have deviations which need to be studied further. To enable this, a larger sample created by an Analysis Production would be necessary, as the large fluctuations in this plot are due to the limited size of the MC truth sample.

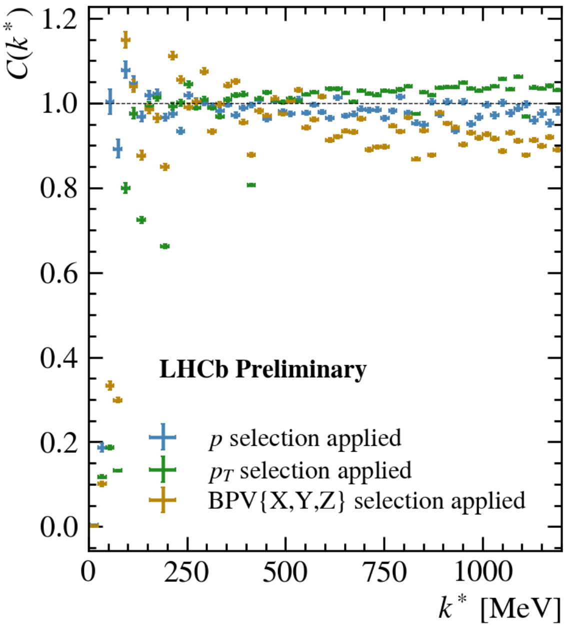

Figure 15: Correlation functions at MC truth level for different additional selection criteria. Blue: total momentum cut, green: transverse momentum cut, orange: BPV{X,Y,Z} cut. All three show an unexpected dip at low k^*, approach unity at intermediate k^* due to normalisation, and diverge at high k^* with different characteristic behaviours.

5 Conclusion

In this work, the femtoscopic correlation function of p–p pairs is measured using the LHCb Run 3 NoBias data set for pp collisions. The analysis demonstrated that such correlations can indeed be extracted with the currently available statistics, even in the low-k^* region that is most sensitive to quantum statistical and final-state interaction effects. This is not obvious at the outset, and establishes the feasibility of femtoscopy with the upgraded detector and trigger infrastructure.

The results show good qualitative agreement between data and expectations, with the anticipated enhancement at low relative momentum, a return to unity at intermediate k^*, and residual deviations at larger values. However, persistent discrepancies are identified in the simulation sample. These deviations are already present at the MC truth level, even when only minimal selection criteria are applied, pointing to a more fundamental issue in the modelling of correlations or event structure. Understanding and resolving this behaviour will be crucial for establishing a reliable baseline against which genuine femtoscopic effects can be compared.

Looking forward, a substantial increase in statistical precision is expected as more data become available. The full 2024 data set will provide roughly an order of magnitude more events, with further increases anticipated from the 2025 data set. This will enable a more detailed study of the correlation function and allow for systematic investigations of the residual effects that are not yet included in the present analysis. In addition, the increase in statistics will make the study of more complex baryon–baryon systems feasible. While p–p pairs are abundant in the current NoBias data set due to the large production cross section, heavier baryons and hyperons are produced much less frequently. Their pair yields are therefore limited in the present sample. With the larger statistics expected in future data sets, sufficient numbers of such rare pairs will become available to allow meaningful femtoscopic correlation studies beyond the p–p case.

An important next step will be to measure the correlation function in bins of event multiplicity. The particle density in the final state directly affects the space–time extent of the emitting source, and multiplicity-dependent femtoscopy has been a key tool in other experiments to disentangle collective effects from intrinsic two-particle interactions. Adding this dimension at LHCb will provide a more complete characterisation of the collision dynamics.

Subsequently, it will be necessary to refine the correlation function by accounting for residual correlations[6] and to perform a full systematic uncertainty evaluation. Once a precise p–p correlation function is established, it will become possible to extract the emission source size for pp collisions in the unique kinematic environment of the LHCb detector. Beyond the proton system, femtoscopic studies can also be extended to other baryon–baryon pairs, as well as to heavy-ion collisions such as Pb–Pb, where larger source sizes and stronger collective effects are expected. For rarer systems, dedicated trigger lines may be required to ensure sufficient statistics.

In summary, this analysis provides the first demonstration that femtoscopic p–p correlations are accessible at LHCb with Run 3 data. It lays the groundwork for a broad future programme of femtoscopy, exploiting the unique capabilities of the experiment to study both abundant and rare baryonic systems with unprecedented precision.

Acknowledgements

I would like to thank both of my supervisors, Thomas and Saverio, for creating a supportive and motivating environment throughout my summer student project. Thomas always made time to answer my questions and provided clear guidance whenever I needed it. Saverio's friendly manner and thoughtful advice were also greatly appreciated. Together, you both contributed significantly to making this a rewarding and enriching experience.

[6] The procedure is explained in Sec. 2.2.2.

I am also grateful to my colleagues in the office and other summer students for their support, discussions, and the enjoyable atmosphere they created during the project. In particular, I thank my new friends Guillermo, Idan, Reccta, Gift, Angel, Mateo, Eliot and Uday, for their encouragement outside of work, which made the overall experience even more memorable.

Appendices

A Event samples and selection

A.1 Online preselection

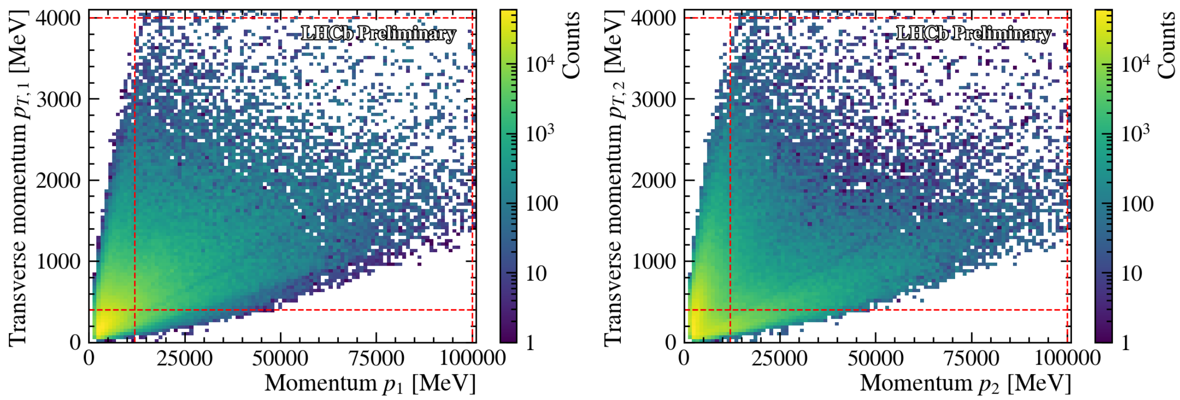

Figure 16: Transverse momentum p_T against momentum p for the first (left) and the second proton in the proton pair (right) for MC with no selection applied. The dashed red lines indicate the rectangular cuts.

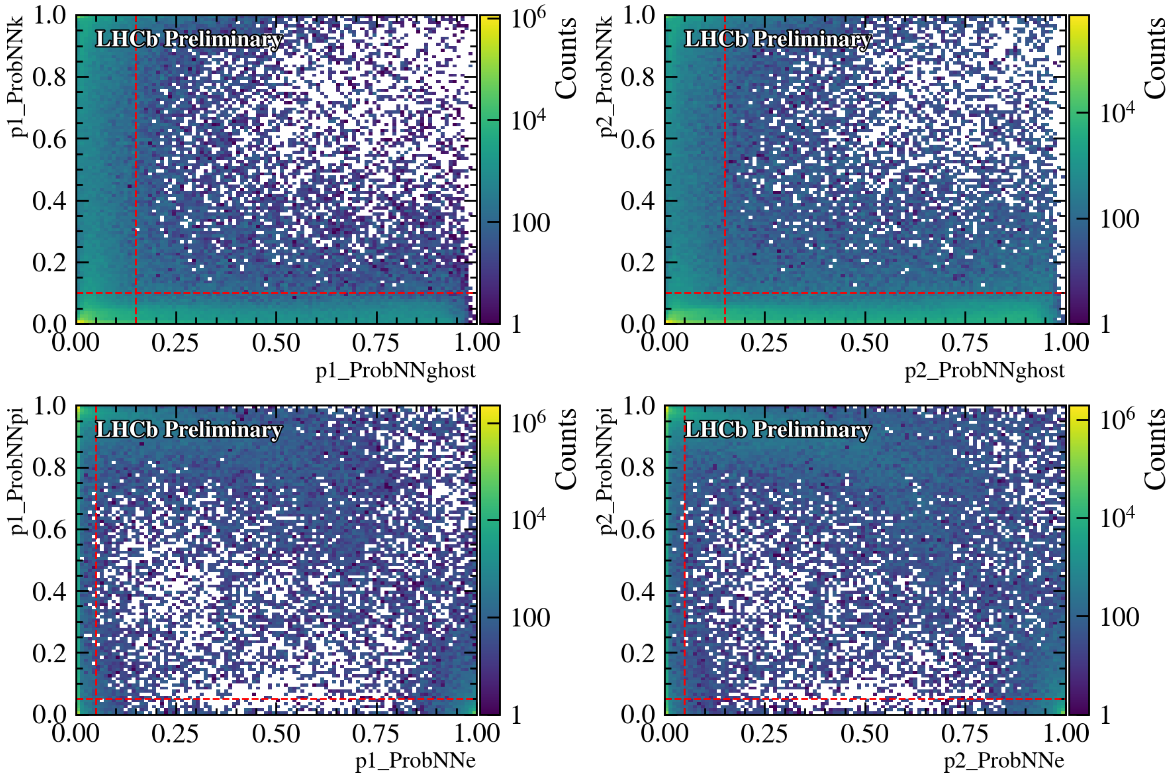

Figure 17: Particle hypothesis probability cuts in MC. Top: `ProbNNk` versus `ProbNNghost` for the first (left) and second (right) proton candidate. Bottom: `ProbNN`π versus `ProbNNe` for the same candidates. The red dashed lines indicate the applied selection boundaries.

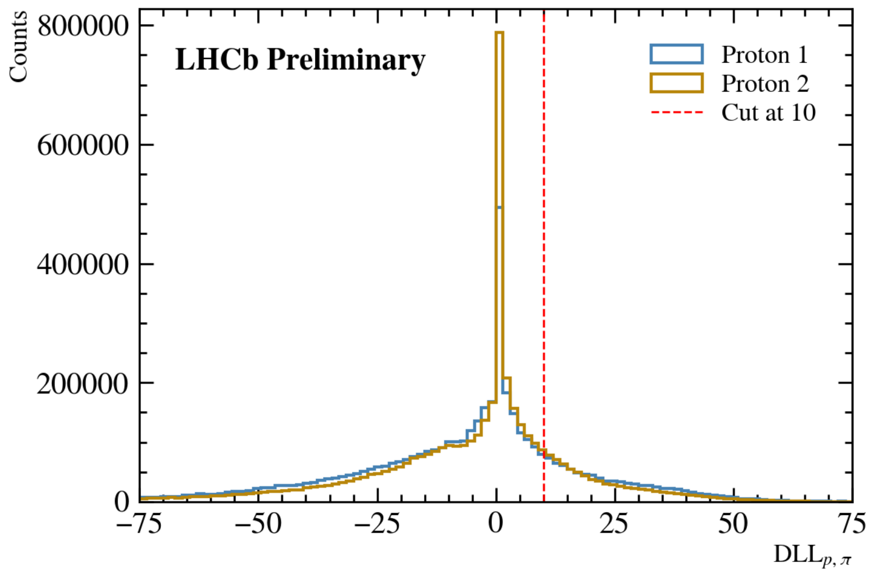

Figure 18: Distribution of the proton–pion log-likelihood difference in MC, $\mathrm{DLL}_{p,\pi}$, for the first (blue) and second (orange) proton candidate. The vertical red dashed line marks the applied cut at $\mathrm{DLL}_{p,\pi} > 10$.

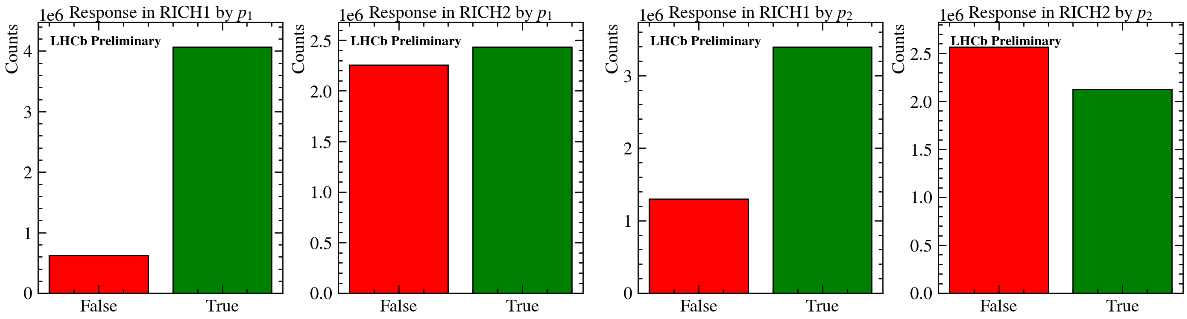

Figure 19: Binary distributions of the RICH response for proton candidates in simulation. Top row: first proton (left) and second proton (right) in RICH1; bottom row: first proton (left) and second proton (right) in RICH2. A value of `True` indicates that the track produced a valid Cherenkov response, while `False` corresponds to no response.

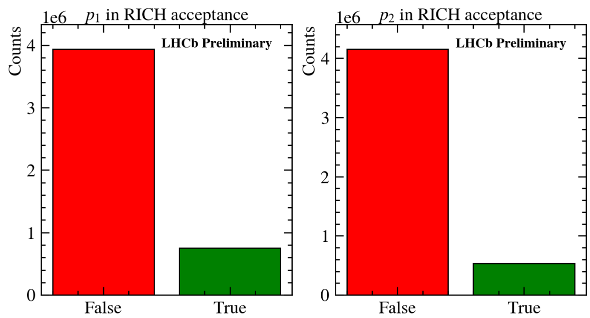

Figure 20: Binary distributions of the RICH acceptance for proton candidates in simulation. The value `True` indicates that the track lies within the acceptance of the corresponding RICH detector, whereas `False` indicates it falls outside the acceptance.

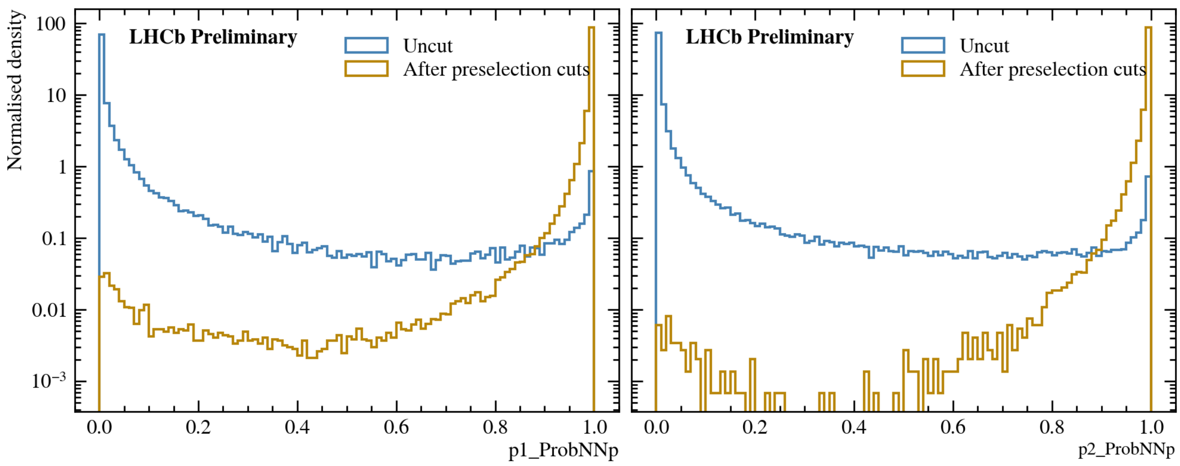

Figure 21: Normalised distributions of `ProbNNp` for the first (left) and second proton candidate (right) in simulation.

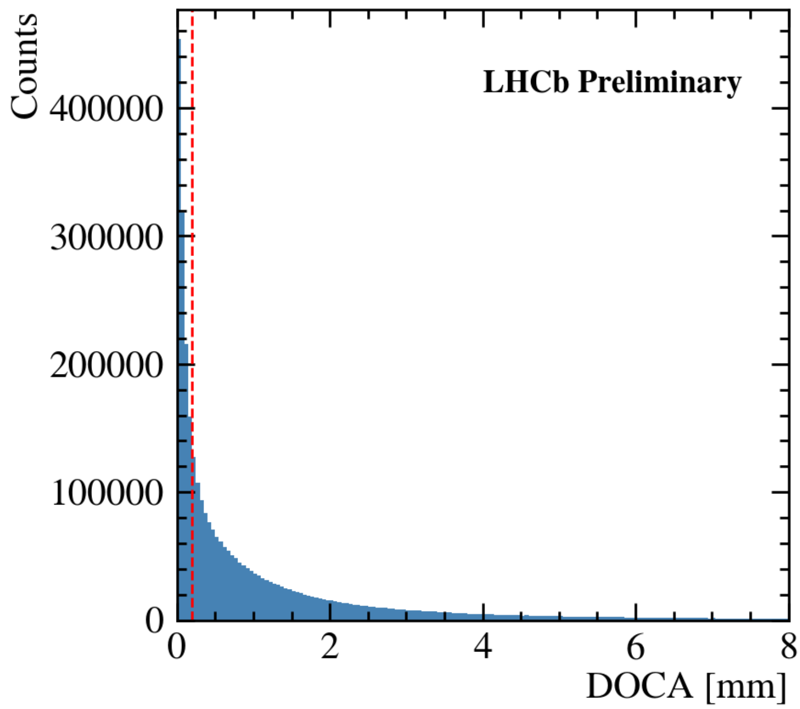

Figure 22: Distribution of the distance of closest approach (DOCA) between proton tracks in simulation. The red dashed line indicates the applied cut at 0.2 mm.

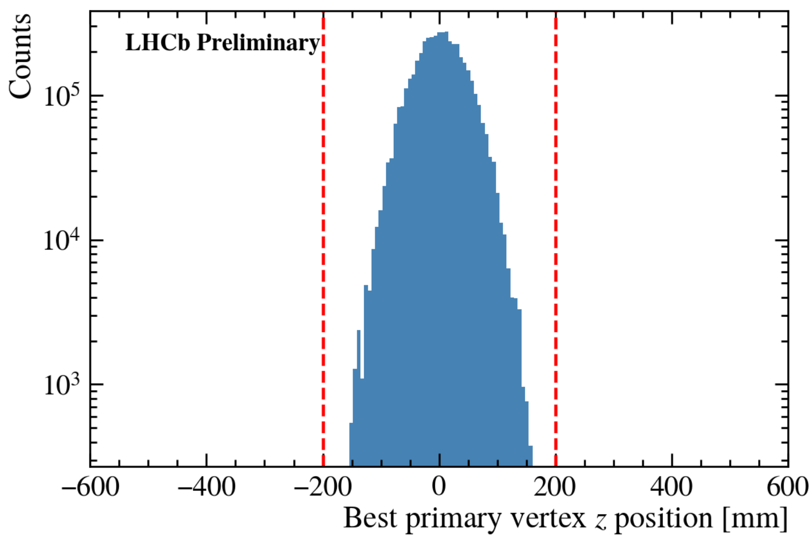

Figure 23: Distribution of the reconstructed primary vertex position z in simulation. The red dashed lines indicate the applied selection window restricting the z-position to the VELO region ($z \in [-200, 200]$ mm).

A.2 Offline event selection

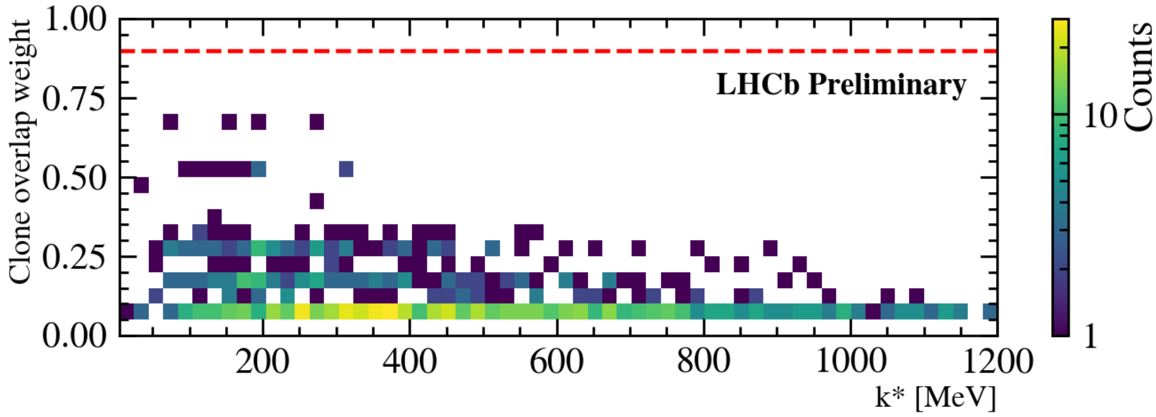

Figure 24: Distribution of the track-overlap weight for proton pairs in simulation as a function of k^*. The red dashed line indicates the applied cut at 90%.

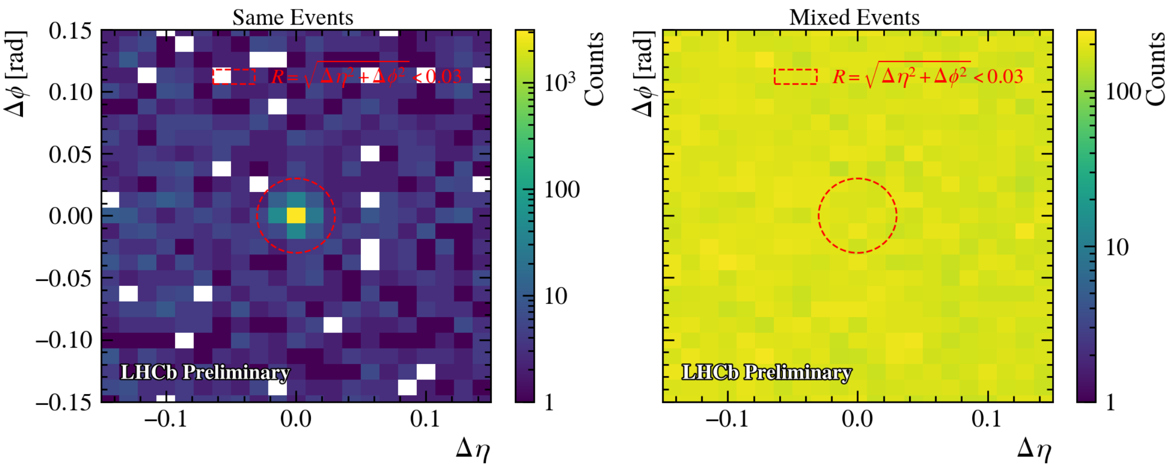

Figure 25: Two-dimensional distributions of $\Delta\eta$ versus $\Delta\phi$ for proton pairs in simulation, restricted to the region $[-0.15, 0.15]$ (mrad) in both variables. The left plot shows the same-event distribution, where clone contamination is visible as a pronounced peak at the origin. The right plot shows the mixed-event distribution, which is free of clones by construction. The red dashed circle illustrates the applied cut of $R > 0.03$.

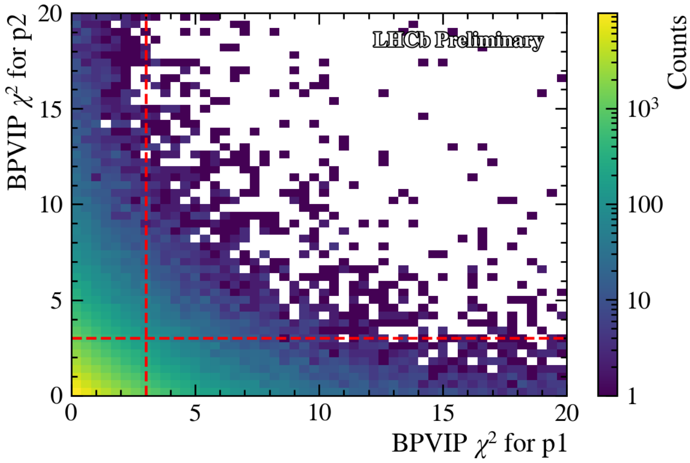

Figure 26: Two-dimensional distribution of the impact parameter significance for the best primary vertex (BPVIPχ^2) of the first proton against the second proton in simulation. The red dashed lines indicate the applied selection of BPVIP$\chi^2 < 3$ for both tracks.

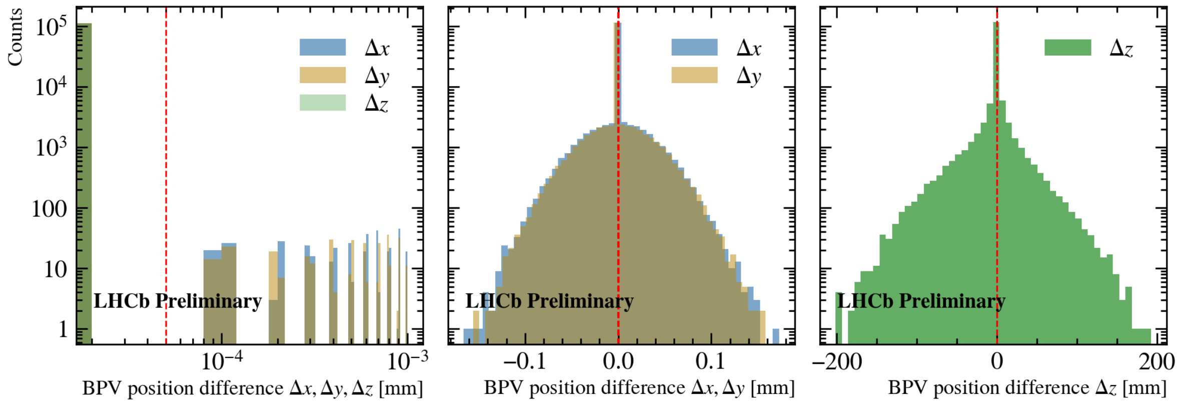

Figure 27: Differences in primary vertex coordinates (Δx, Δy, Δz) for all proton pairs in the simulation sample. Left: zoomed-in view showing the main peak near zero for all three directions, with the red dashed lines indicating the selection window ($|\Delta x|, |\Delta y|, |\Delta z| < 5 \times 10^{-5}$ mm). Middle: full-range distributions for Δx and Δy, showing that most pairs are close to zero. Right: full-range distribution of Δz, which is broader but still exhibits a clear peak near zero.

A.3 Uniformity in $\Delta\phi$

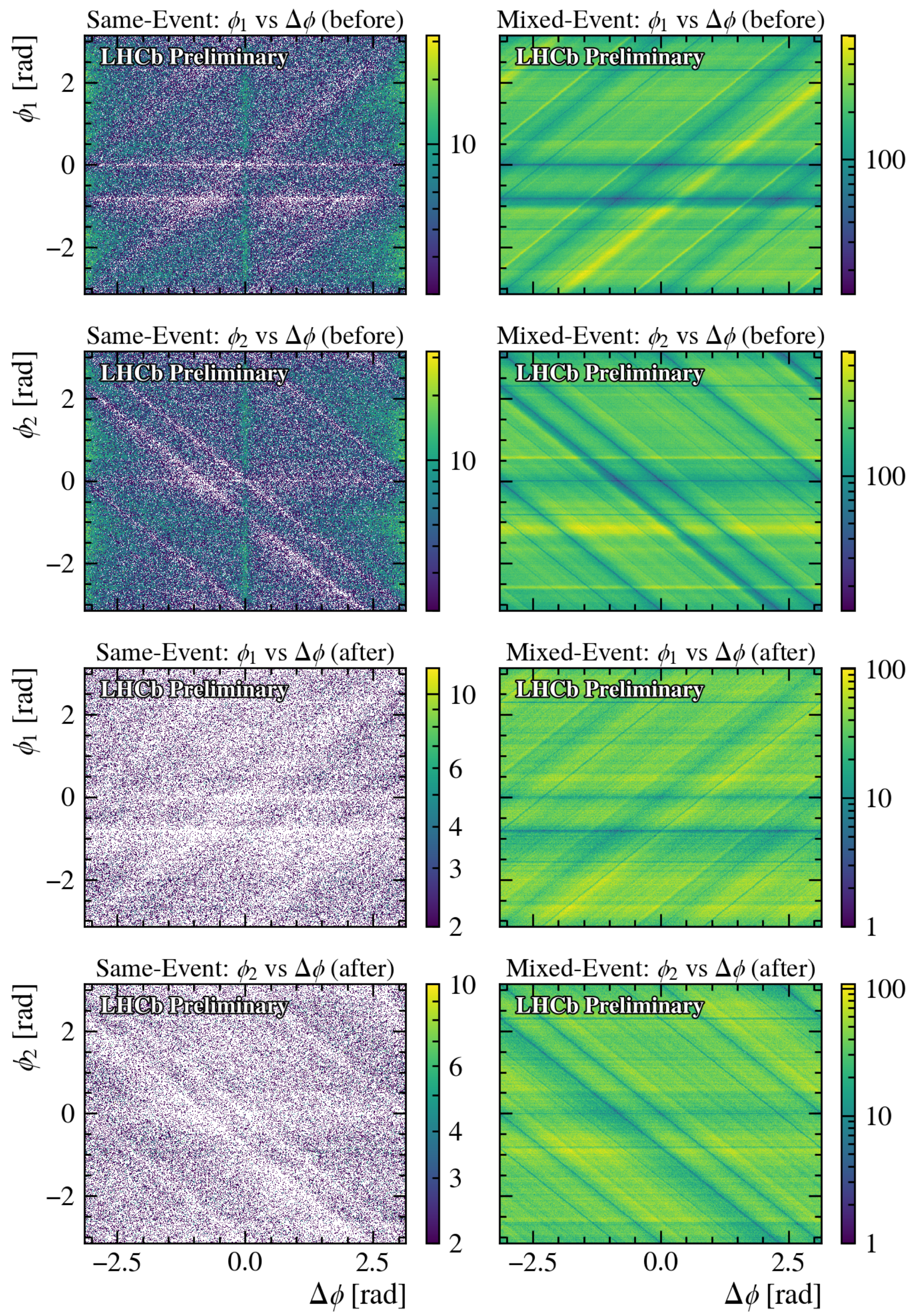

Figure 28: Two-dimensional distributions of the azimuthal angle difference $\Delta\phi$ with respect to the absolute azimuthal angle ϕ_1 (rows 1 and 3) and ϕ_2 (rows 2 and 4), shown separately for same-event (left column) and mixed-event (right column) pairs in data. The top two rows correspond to the sample before the offline selection, while the bottom two rows correspond to the sample after applying the selection cuts.

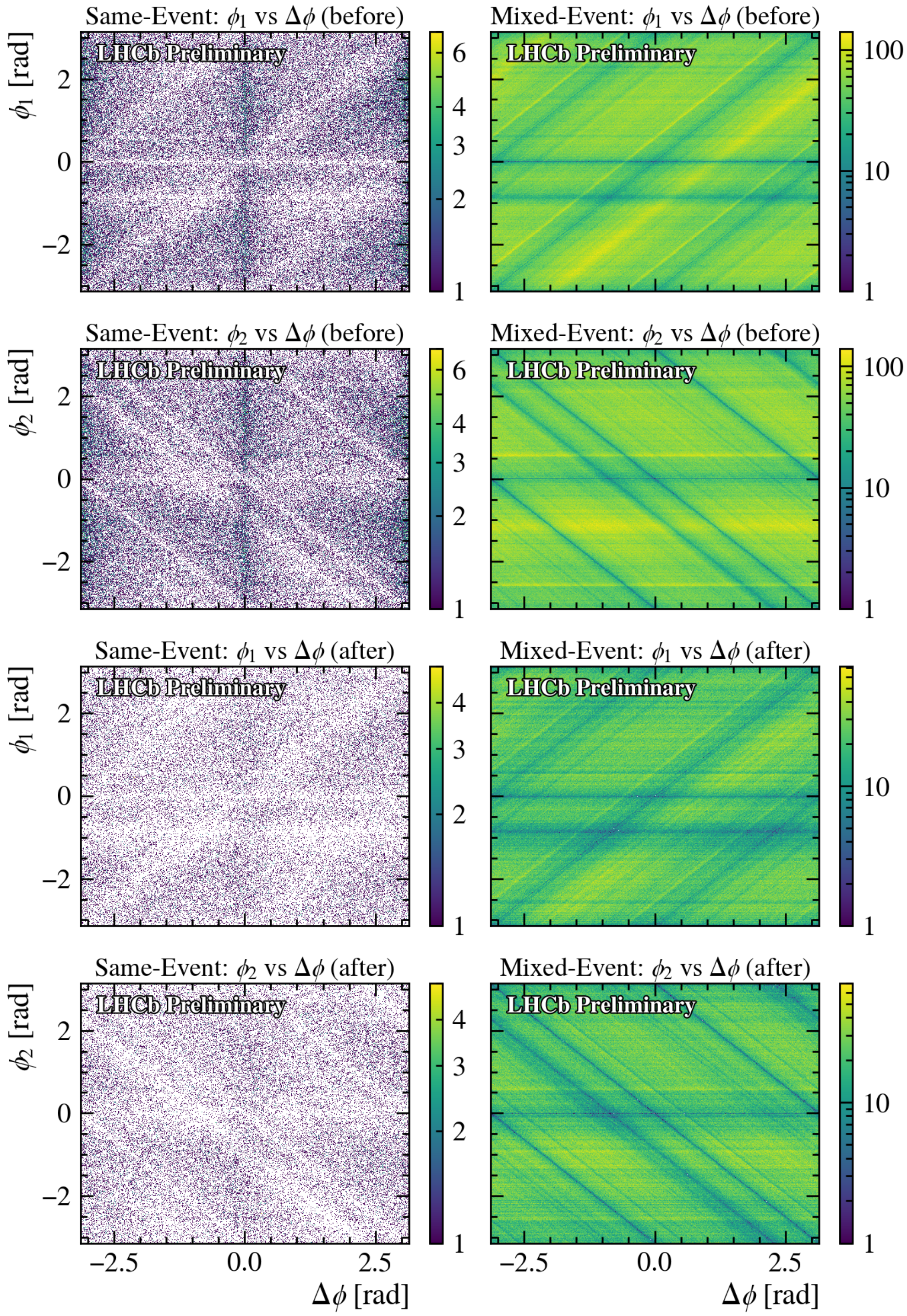

Figure 29: Two-dimensional distributions of the azimuthal angle difference $\Delta\phi$ with respect to the absolute azimuthal angle ϕ_1 (rows 1 and 3) and ϕ_2 (rows 2 and 4), shown separately for same-event (left column) and mixed-event (right column) pairs in simulation. The top two rows correspond to the sample before the offline selection, while the bottom two rows correspond to the sample after applying the selection cuts.

A.4 Complete event selection

Table 3: Summary of online and offline selection requirements applied to the data and MC sample with long proton candidates.

Description	Requirement
$p^{\pm}$	
Momentum	$p \in [12, 100]\,\mathrm{GeV}$
Transverse momentum	$p_T \in [0.4, 4]\,\mathrm{GeV}$
Difference in log-likelihood	$\mathrm{DLL}_{p,\pi} > 10$
Probability of track belonging to a ghost	ProbNNghost< 0.15
Probability of track belonging to a π/e	ProbNN{pi,e}< 0.05
Probability of track belonging to a K	ProbNNk< 0.1
RICH1 or RICH2 in acceptance region	True
RICH response	True
Best primary vertex impact parameter χ^2	$\chi^2 < 2$
$p^{\pm}p^{\pm}$ pair	
Distance of closest approach	$\mathrm{DOCA} > 0.2\,\mathrm{mm}$
Best primary vertex position in z	$\lvert\mathrm{BPV}_z\rvert < 200\,\mathrm{mm}$
Best primary vertex position difference	$\lvert\mathrm{BPV\{X,Y,Z\}}_{p_1} - \mathrm{BPV\{X,Y,Z\}}_{p_2}\rvert < 50\,\mu\mathrm{m}$
Angular separation	$R = \sqrt{\Delta\eta^2 + \Delta\phi^2} > 0.03$
Relative track overlap	Weight < 0.9

References

[1] ALICE Collaboration, *p–p, p–Λ, and Λ–Λ correlations studied via femtoscopy in pp reactions at* $\sqrt{s} = 7$ *TeV*, Physical Review C **99** (2019) .

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

[3] LHCb collaboration, *LHCb SMOG Upgrade*, CERN-LHCC-2019-005, 2019.

[4] S. Weissenborn, D. Chatterjee, and J. Schaffner-Bielich, *Hyperons and massive neutron stars: the role of hyperon potentials*, Nuclear Physics A **881** (2012) 62.

[5] D. Lonardoni, A. Lovato, S. Gandolfi, and F. Pederiva, *Hyperon puzzle: Hints from quantum monte carlo calculations*, Physical Review Letters **114** (2015) 092301.

[6] D. Gerstung, N. Kaiser, and W. Weise, *Hyperon–nucleon interactions and neutron star matter*, European Physical Journal A **56** (2020) 175.

[7] HADES Collaboration, J. Adamczewski-Musch *et al.*, *Search for a deeply bound* k^-pp *state in p+p reactions at 3.5 gev*, Physical Review C **94** (2016) 025201.

[8] HADES Collaboration, G. Agakishiev *et al.*, *Hyperon–nucleon interaction studies in proton–nucleus collisions*, Physical Review C **82** (2010) 021901.

[9] M. A. Shifman, *Quark-hadron duality*, Foundations of Physics **30** (2000) 455.

[10] R. B. Wiringa, V. G. J. Stoks, and R. Schiavilla, *Accurate nucleon-nucleon potential with charge-independence breaking*, Phys. Rev. C **51** (1995) 38.

[11] R. Machleidt, *The meson theory of nuclear forces and nuclear structure*, Advances in Nuclear Physics **19** (1989) 189.

[12] STAR Collaboration, *-– correlation function in Au+Au collisions at* $\sqrt{s_{nn}} = 200$ *gev*, journal = Phys. Rev. Lett., volume = 114, pages = 022301, year = 2015, doi = 10.1103/PhysRevLett.114.022301, note = Later re-analysis with residual correlation considerations found the interaction to be mildly attractive., .

[13] D. L. Mihaylov, *Analysis techniques for femtoscopy and correlation studies in small collision systems and their applications to the investigation of p–Λ and Λ–Λ interactions with ALICE*, phd thesis, TU München, 2024.

[14] I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, Journal of Physics: Conference Series **331** (2011) 032047.

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

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

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

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

[19] J. N. Himbert, *p–Ω^- and p–Ξ^- interaction with two-particle correlation femtoscopy at LHCb* technical report, CERN, 2024.